INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA INFORMÁTICA

Sistema inteligente de gestão de energia em edifícios

ARIA JOZI

Julho de 2019

POLITÉCNICO
DO PORTO

INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA INFORMÁTICA

Intelligent energy management system in buildings

ARIA JOZI

Julho de 2019

POLITÉCNICO
DO PORTO

Intelligent energy management system in buildings

Aria Jozi

Dissertation to obtain a master's degree in

Computer Engineering, Information and Knowledge Systems

Advisor: Doctor Goreti Marreiros

Co- advisors: Doctor Tiago Pinto

Eng. Luís Gomes

Porto, July 2019

To my family and my friends

Acknowledgments

I would like to express my special appreciation and thanks to my advisors: Prof. Doctor Goreti Marreiros, Doctor Tiago Pinto and Eng. Luís Gomes. You have been a tremendous mentor for me. I would like to thank you for all your advice and helps. It is a pleasure to work with you.

A special thanx to my family. Words cannot express how grateful I am to have you. Thank you for all the sacrifices that you have made on my behalf.

This work could not be finalized so straightforwardly without the help of all my friends and colleagues in GECAD research center. Thank you for your help and supports.

Abstract

Energy management systems have become one of the most significant concepts in the power energy area, due to the dependency of nowadays human's lifestyle on electrical appliances and increment of energy demand during the past decades. From a general perspective, the total energy consumption by humans can be divided into three main economic sectors, namely industry, transportation, and buildings. Based on recent studies, the buildings present the largest share of consumption, standing for approximately 40% of the total consumption. This fact makes buildings energy management the most important component of energy management. On another hand, according to the variety of different types of buildings and several existing consumption appliances, the management of energy consumption in the building becomes a challenging problem. The main goal of a building energy management system is to control the energy consumption of the building by considering several facts, such as current and estimated consumption and generation, the energy price and comfort of the users. Due to the complexity of this management and limitations of available information, most of the existing systems focus on optimizing the consumption value and the cost of the energy with less consideration of the comforts and habits of the users. Moreover, the context of decision-making is also not sufficiently explored. However, the energy management in the building can be designed based on an intelligent system which has the knowledge to estimate the comforts and needs of the users and acts based on this awareness.

This work studies and develops an intelligent energy management system for buildings energy consumption. This system receives the historical data of the building and uses a set of artificial intelligence techniques as well as several designed rulesets and acts as a recommender system. The goal of the generated recommendations by this system is to attune the usage of the electrical appliances of the building by comforts and habits of the residents while considering the price of the electricity market and the current context. Results show that the system enables users to obtain a comfortable environment in the building in the most affordable way.

Keywords: Building energy management systems, Data mining techniques, Energy consumption, Intelligent systems, Recommender systems.

Resumo

Nas últimas décadas, a dependência do estilo de vida na elevada utilização de dispositivos elétricos e grande consumo energético, faz com que os sistemas de gestão de energia sejam um dos conceitos mais relevantes no setor energético. Numa perspetiva geral, o total da energia consumida divide-se essencialmente em três setores económicos: industrial, transporte e edifícios. Os edifícios têm a maior representatividade, correspondendo aproximadamente a 40% do consumo total. Assim, a gestão energética em edifícios é a componente com maior importância nesta área. Por outro lado, devido à variedade dos diferentes tipos de edifícios e dispositivos de consumo, a gestão do consumo de energia nos edifícios apresenta desafios. O objetivo principal de um sistema de gestão energética em edifícios consiste em controlar o consumo energético no edifício, considerando diversos fatores, tais como o consumo e produção atuais, a sua estimativa, o preço de mercado e conforto dos seus utilizadores. Perante a complexidade desta gestão e das limitações da informação disponível, a maioria dos sistemas tem foco na otimização do consumo e os seus custos, tendo em menor consideração o conforto e hábito dos utilizadores. Além disso, o contexto da tomada de decisão não é devidamente explorado, enquanto a gestão energética em edifícios pode ser baseada num sistema inteligente, cujo conhecimento pode estimar o conforto e necessidades dos seus utilizadores, e assim atuar com base nessa consciência.

Este trabalho estuda e desenvolve um sistema inteligente para a gestão do consumo de energia em edifícios. O sistema recebe o histórico de dados de um edifício, e utiliza um conjunto de técnicas de inteligência artificial e conjuntos de regras, funcionando como um sistema de recomendações. O objetivo das recomendações geradas pelo sistema é adaptar os dispositivos elétricos do edifício ao conforto e hábitos dos utilizadores enquanto são considerados o preço de mercado e o contexto atual. Os resultados demonstram que o sistema permite aos utilizadores obter um ambiente confortável no edifício, da forma mais económica possível.

Palavras-chave: Consumo energético, Sistemas de gestão energética em edifícios, Sistemas de recomendação, Sistemas inteligentes, Técnicas de Data Mining

.

چکیده

امروزه سیستم های مدیریت انرژی به دلیل وابستگی شیوه ی زندگی انسانی به لوازم الکتریکی و افزایش تقاضای انرژی در دهه های گذشته، به یکی از مهم ترین مفهوم ها در زمینه ی انرژی برق تبدیل شده اند.به طور کل مصرف انرژی انسان ها میتواند به سه بخش اصلی اقتصادی تقسیم شود: صنعت، حمل و نقل و ساختمان. بنا بر تحقیقات اخیر ساختمان ها با تقریبا ۴۰٪ مصرف کل انرژی، بیشترین سهم را در مصرف انرژی دارند. این امر باعث میشود که مدیریت انرژی ساختمانی مهمترین مولفه ی مدیریت انرژی باشد. از طرفی دیگر، با توجه به میزان گوناگونی ساختمان ها و دستگاه های برقی مصرفی مختلف، مدیریت انرژی مصرفی ساختمان ها تبدیل به مشکل چالش برانگیزی شده است.
هدف اصلی سیستم مدیریت انرژی ساختمان کنترل انرژی مصرفی ساختمان از طریق در نظر گرفتن چند نکته از جمله تولید و مصرف فعلی و تخمین زده شده، هزینه انرژی و آسایش مصرف کنندگان است. به دلیل پیچیدگی این مدیریت و محدود بودن اطلاعات در دسترس، تمرکز اکثر سیستم های موجود بیشتر بر بهینه سازی ارزش مصرف و هزینه انرژی در تصمیم گیری به حد کافی است تا در نظر گرفتن آسایش و عادات مصرف کنندگان. علاوه بر این نیز به بافت آگاه پرداخته نشده است.این در حالی است که مدیریت انرژی ساختمان میتواند بر اساس یک سیستم هوشمند که دانش تخمین زدن نیاز و میزان آسایش مصرف کنندگان را داشته باشد و بر اساس این آگاهی ها عمل کند، باشد.

این پروژه سیستم مدیریت انرژی هوشمند برای مصرف انرژی ساختمان ها را مطالعه کرده و توسعه میدهد. این سیستم تاریخچه اطلاعات ساختمان ها را دریافت کرده و از مجموعه ای از تکنیک های هوش مصنوعی و تعدادی از قوانین طراحی شده استفاده می کند و همانند یک سیستم پیشنهاد دهنده عمل میکند. هدف پیشنهاد های تولید شده از این سیستم ، هماهنگ کردن استفاده لوازم برقی ساختمان با عادات و آسایش ساکنین همراه با در نظر گرفتن قیمت بازار و بافت جاری است. نتایج نشان میدهد که این سیستم کاربران را قادر میسازد تا محیطی پر آسایش را در ساختمان به نحو مقرون به صرفه ای بدست بیاورند.

کلید واژه ها : ساختمان سیستم های مدیریت انرژی، تکنیک های هوش مصنوعی ، مصرف انرژی، سیستم های هوشمند، سیستم های توصیهگر.

Index

Figure Index

Table Index

Notation

ANN	Artificial Neural Networks
AC	Air-conditioning systems
Act	Forecasted Activity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Auto Regressive and Moving Average models
AvgP	Average price
BCI	Brightness Comfort Interval
BCR	Brightness Comfort basis Rate
C2C	Click to Control
CARS	Context-Aware Recommender Systems
CB	Current Brightness
CBR	case-based reasoning
CF	Consumption forecast
CP	Current electricity market price
CS	Consumption State
CT	Current temperature
DT	Decision Trees
EA	evolutionary algorithms
FB	Forecasted ideal Brightness
FFE	Fuzzy Front End
FRBS	Fuzzy Rule Base Systems
FT	Forecasted ideal temperature
GECAD	Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
GF	Generation forecast
GFRB	Generated Fuzzy Rule Bases
GFS.FR.MOGUL	Genetic Fuzzy Systems for Fuzzy Rule learning based on the MOGUL methodology
HyFIS	Hybrid Neural Fuzzy Inference System
IRL	rule learning approach
k-NN	k-Nearest Neighbor
MAPE	Mean Absolute Percentage Error
MLP	multilayered perceptron
PLC	Programmable Logic Controller
RS	Recommender systems
SCV	Support Vector Classification
SVM	Supporter Vector Machine
TCI	Temperature Comfort Interval
TCR	Temperature Comfort basis Rate
USA	United States of America
WCSS	Within-Cluster Sum of Squares
WM	Wang and Mendel’s Method

1 Introduction

1.1 Motivation

The increment of the renewable energy usage, auto consumption systems and importance of the amount of the energy bills during the past decades, made the energy management systems one the most important and considered concepts in the energy market. Industry, transportation, and buildings are the three main economic sectors with the highest amount of energy consumption, where buildings present the most significant portion (Khosravani, Castilla, Berenguel, Ruano, & Ferreira, 2016). Most of the people spend 90% of their time inside the buildings and relying on mechanical heating and air conditioning systems, buildings become the largest energy consumers. According to the past studies, approximately, 40% of total energy consumption worldwide, have been consumed by the buildings (Cao, Dai, & Liu, 2016). Due to this amount of energy consumption in the building, the control and management of this consumption take a vital role to reduce the unnecessary energy consumptions and control the energy bills.

The energy management systems for buildings change and adapt the consumption of the building by considering the price of the market, expected production, estimated consumption and the comfort of the users (Jamil & Mittal, 2017). Most of the recognized building types are office, residential, and engineering buildings, varying from small rooms to big estates (H. X. Zhao & Magoulès, 2012). The energy system in the building can be complex, depending on the types of building and types of included energy consumers. The consumption of a building is mostly divided into the three-consumer type, namely, air conditioning systems, lights and electrical sockets. This variety requires the consideration of many different factors to have a clear and trustable recognition analysis of the consumption behavior on the building. On another hand, the limitation of the available data made it difficult for this kind of systems to create a suitable adaptation of the consumption profile which considers all the important aspects of the market and corresponds to all of the needs and comforts of the users. Moreover, so that this management can be done in an intelligent way and taking into account the needs and preferences of the users in each moment, it is essential that the decision making be done in a way that depends on the contexts of use. Intelligent, contextual-dependent decision support is not yet properly exploited, hindering development in the area.

1.2 Objectives

The main objective of this work is *to study and develop an intelligent energy management system for the energy consumption of the building* to have better control on the energy consumption, reduce the unnecessary consumptions and adapt the usage of the electrical appliances of the building to the demand and needs of the users. This system should act as a recommender system where the recommendations are based on the studies and analyses of the existing data of the building; namely, the collected data from the sensors, analyzers and energy meters of the building. These data should pass a preparation phase to make sure that the data have the right structure and format to be used for the intelligent data analysis techniques as well as forecasting, classification, and clustering methodologies. The systems should use different data analysis services for each different type of data and needed information and based on the results of these analyses make recommendations to control the consuming devices of the building.

In order to achieve the main objective of this work the following specific objectives have been proposed:

- Study the current state of the art:
  - Analyze the principal approaches for energy management in the building
  - Study and analyze the existed recommender systems for energy management
  - Study and analyze the existed data mining and data analyses techniques in the context of energy management.
- Identify the best approaches and techniques to implement the intended system.
- Create and design the architecture of the system, considering:
  - The different data analysis techniques;
  - The recommender model;
  - Data preparation and pre-processing methodologies;
- Develop the energy management system in buildings, considering the defined architecture and the designed models.
- Integrate the developed system with the building N from Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development

(GECAD) facilities, using the collected real data of this building and update the recommender model with the existing device control system of the building.

- Experimentation of the real case studies to validate and test the implemented system.

1.3 Contributions

The purpose of this work is to develop an intelligent energy management system in buildings to control the usage of energy consuming appliances of the building. The proposed system uses e set of data mining techniques to predict the demands and comforts of the residents of the building based on the recorded data from past days. Based on the predicted conditions and by considering the consumption cost of the building recommends the state of the electrical appliances to meet the predicted conditions in the most affordable way.

Figure 1 presents a brief presentation of the structure of this system. The system in the first place receives the historical data of the building from the database of the building. These data include all the available variables from the building which has an influence on energy consumption or usage of the electrical appliances. The system receives this data and starts a data cleaning process. As the recorded data in the database might have some failures, at this phase the system makes sure that the received data set is completed without any unreal data. The available data in the database mostly are recorded in short time intervals. This way, the next step is to aggregate the received data into the intended time intervals by the system. This time interval is related to how often a new recommendation should be created. After the aggregation phase, the system uses a clustering method to divide the data into several groups based on different contexts. These contexts are created according to the type of variables and objective of the recommendations. The data from the intended context will be used as the training data in the forecasting processes. In the forecasting phase, the system uses the selected data from the clustering process to train the methods. The system includes four forecasting and one classification methods and based on the type of the variable one of these data mining techniques will be used to predict the target value. If forecasting is the case, the system uses all four algorithms and predicts four values. The predicted values will be stored in the created database for the system. To generate the recommendations the rule engine component receives the current actual data of the building from the source database and the predicted values from the database of the system. At this point, the system analyses the errors of each forecasting method during the last iterations and sends the results of the methods with less errors to the rule engine. The system includes several rulesets according to the objective of the recommendations. These rulesets will be executed by the received actual and estimated data and will generate recommendations about the state of the electrical appliances of the building. The created recommendations by the system will be also stored into the database of the system to be accessible for the users.

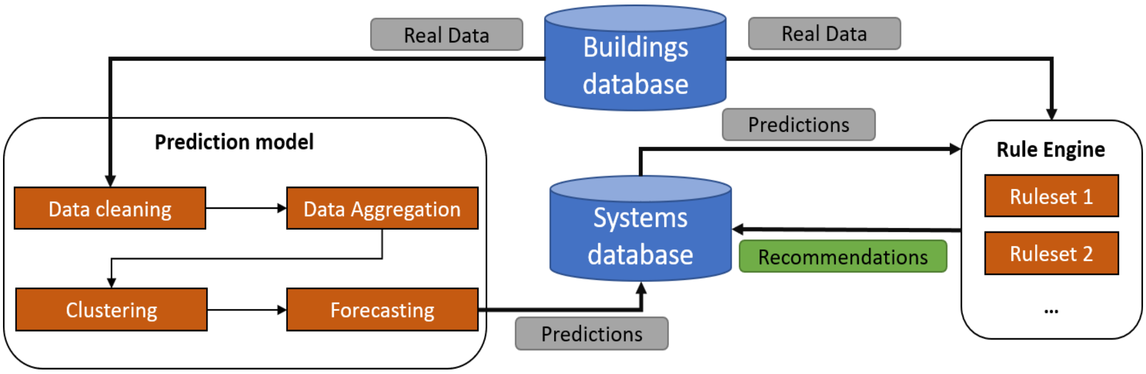

Figure 1 – A brief presentation of the structure of the proposed system

The designed system has been implemented for the building N of the GECAD facilities located in Porto, Portugal. This building is an office building which includes several office rooms and laboratories where many different energy consuming appliances are used. The building also has a set of solar panels that provide an amount of power generation to be used in the building. The implemented system considers the Air Conditioning system (AC) and the lights of this building as the target of the recommendations. The proposed system has been successfully implemented for the mentioned building, and the performance and results of the system have been analyzed and evaluated by several case studies which prove the efficiency of the system.

The developed work and studies of this project have resulted in the publication of six scientific papers, including three in internationals journals and 3 in conference proceedings. All of them have been accepted, and three articles are already published online.

- A.Jozi, T.Pinto, I.Praça, Z.Vale, "Decision Support Application for Energy Consumption Forecasting." Appl. Sci. 2019, 9, 699, doi: 10.3390/app9040699
- L.Gomes, C.Ramos, A.Jozi, B.Serra, L.Paiva, Z.Vale, "IoH: A Platform for the Intelligence of Home with a Context Awareness and Ambient Intelligence Approach." Future Internet 2019, 11, 58. doi:10.3390/fi11030058
- A. Jozi, T. Pinto, I. Praça, F. Silva, B. Teixeira, Z. Vale, "Genetic fuzzy rule-based system using MOGUL learning methodology for energy consumption forecasting," ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 2019, (Accepted June 2019)
- A. Jozi, T. Pinto, I. Praça, and Z. Vale, "Day-ahead forecasting approach for energy consumption of an office building using support vector machines," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1620-1625. doi: 10.1109/SSCI.2018.8628734

- A. Jozi, T. Pinto, G. Marreiros, Z. Vale, "Electricity consumption forecasting in office buildings: an artificial intelligence approach", IEEE PowerTech 2019, Milano, Italy, June 2019
- A. Jozi, D. Ramos, L. Gomes, P. Faria, T. Pinto, Z. Vale, "Demonstration of an Energy Consumption Forecasting System for Energy Management in Buildings," 2019 EPIA Conference on Artificial Intelligence, Vila Real, Portugal (Accepted)

More papers using the proposed forecasting strategies to predict the load consumption of several devices such as refrigerators to use in the Demand Response program are under process and will be submitted during the following months.

1.4 Document structure

This document is divided into six chapters, which are: Introduction, context and the State of Art, Proposed solution, Implementation of the proposed model for building N, Case studies and Conclusions.

The second chapter reviews the state of the art of the energy management system in buildings, recommender systems and several techniques of data mining. The chapter focuses on the importance of the energy manage in the buildings and the proposed solutions for this matter during the past years. The recommender systems in the buildings are presented and discussed as well as the main data mining techniques namely forecasting, classification and clustering. Several methods of every technique are reviewed and most recent studies using these techniques are presented.

In the third chapter the proposed model is presented including the detailed explanation of every phase of the system. The description of the data mining method, as well as data treatment process from data cleaning to the forecasting process are presented. Also, the designed rule sets for Consumption state, Air Conditioning system (AC) and lights of a building is included in this chapter.

The fourth chapter introduces the implemented proposed system for the building N of GECAD facilities. In this chapter all the used variables, data structures and calculations in every parts of the system are explained and presented by examples. The structure of the created databases is are also included as well as the presentation of a created page in the existed energy management system of the building N to monitories the results and recommendations of the system.

A set of different case studies are presented in the fifth chapter. These case studies are designed to evaluate the performances of the data mining methods and the accuracy of the predicted values as well as rationality of the generated recommendations by the rulesets of the system.

The sixth chapter presents the main contributions of this project as well as a perspective for future works and some limitations of the implementation.

2 Context and the State of Art

This chapter presents a review of the energy management systems, Recommender Systems, data mining, and data analyses techniques in the concept of energy systems as well as the value analysis for the proposed system. Data mining techniques include Forecasting Classification and Clustering techniques.

2.1 Energy management systems

2.1.1 Energy management in buildings

Buildings account for a substantial part of the energy consumption, and it is where people spend most of their time, whether for housing, working or leisure. Due to the relationship between productivity and comfort at work, the operation costs of an office building are directly linked to the workers' income (Soares et al., 2017). In this sense, energy consumption and the conditions of environmental comfort are, in most cases, in conflict with each other (Dounis & Caraiscos, 2009). Smart building is, therefore, a pivotal technology to contribute to improving social wellbeing and economic growth while at the same time, reduce fossil fuels dependence and the emissions footprint.

To attain those goals, smart buildings require not only adequate hardware but also efficient computational tools. Artificial intelligence can play a significant role in this sense by introducing learning capabilities and real-time consumption control (Manic, Wijayasekara, Amarasinghe, & Rodriguez-Andina, 2016). Some recent studies advocate the use of advanced learning techniques, such as case-based reasoning (CBR) to predict patterns and energy consumption in smart buildings (Platon, Dehkordi, & Martel, 2015)(Z. Wang & Srinivasan, 2017). In (Corchado et al., 2017) a novel Case-Based Reasoning (CBR) application for intelligent management of energy resources in residential buildings has been proposed. In this work, a mathematic optimization and clustering have been used to identify similar past cases and optimize the choice of the variables that define each case. Results suggest that satisfactory results are obtained without compromising the comfort of the users. An interesting possibility in smart buildings is to use CBR to select the most adequate algorithm for the energy resource management depending on the context of use and execution time versus quality of results requirements. The work in (J. Zhao, Lasternas, Lam, Yun, & Loftness, 2014) develops a practical data mining approach using the data from power consuming appliance of an office to learn the occupant "passive" behavior. The consumption of heating, ventilation, and air conditioning is studied. Results capture diversified individual behavior in using office appliances.

Many published studies can be found, which presents a new approach to manage different type of building or diverse applicants' types. In (Fernandes, Morais, Vale, & Ramos, 2014), the author proposes an innovative method to manage the appliances on the house during a demand response event. The main contribution of this work is to include time constraints in resources management, and the context evaluation to ensure the required comfort levels. The published study in (Santos et al., 2017), presents a new agent-based home energy management approach, using ontologies to enable semantic communications between heterogeneous multi-agent entities. The main goal is to support efficient energy management of end consumers in the context of microgrids, obtaining scheduling for both real and virtual resources. In (Doukas, Patlitzianas, Iatropoulos, & Psarras, 2007) an intelligent decision support model using rulesets based has been proposed for a typical building energy management system.

The rising demand of intelligent data analysis to manage the energy consumption in the building is giving way to a variety of new intelligent approaches, namely intelligent data analysis and management, using combinations of different artificial intelligence approaches. Agent-based systems, meta-heuristic optimization, machine learning, game theory, and data mining approaches, in particular, are promising solutions for the most relevant challenges in the field.

2.1.2 Recommender Systems

Recommender Systems (RS) are software tools and techniques, providing suggestions for items to be of use to a user (Robin Burke, 2007)(Mahmood & Ricci, 2009). A recommender system consists of a filtering system that can predict the preference that a particular user gives to an item. This type of systems can be applied based on two different approaches: content-based or personality-based. Filtering process uses information about past events to recommend a new decision or item. These models are also able to recommend a new item based on the properties that it shares with other items and the interest the user has in, expressed by a query (Talib & Elshaiekh, 2014). Collaborative filtering has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating recommendations (Jonathan, Joseph, Loren, & John, 2004).

Traditional recommender systems mostly focus on recommending the most relevant items to individual users and do not take into consideration any contextual information. Such as the ones based on content-based and collaborative filtering, which tend to use relatively simple user models (Adomavicius & Tuzhilin, 2015). This type of recommender system deal with situations that only two types are entities are used, users and items, and these data are not ordered in contexts to be used to provide the recommendations. Meanwhile, considering the contextual information clearly will improve the efficiency of the recommendations for every user. This way, the Context-Aware Recommender Systems (CARS) addresses the concept of the recommender system, which considers the different contexts of the input data and based their recommendations on this context. For example, in (Hosseinzadeh Aghdam, 2019) presents a novel hierarchical hidden Markov model to identify changes in the user's preferences over time by modeling the latent context of users. Or in (Marreiros, Santos, Ramos, & Neves, 2010) where has been developed a context-aware emotion-based model to design intelligent agents for group decision-making processes. Also, (Zheng & Jose, 2019) propose a novel recommender for context-aware recommendations in which estimates the user preferences by sequential predictions. In (Rodriguez-Fernandez et al., 2019) has been introduced a new model to estimate the expected prices that can be achieved in bilateral contracts under a specific context, enabling adequate risk management in the negotiation process.

Many energy management systems applied RSs to solve different energy issues. The presented study in (Ballenger, Herath, Caceres, Venayagamoorthy, & Corchado, 2017) develops an interface that includes a recommender system which proposes recommendations about the human behavior in order to save money and energy in a domestic environment. In (Talib & Elshaiekh, 2014), a mobile recommender system for energy management has been presented. In (S. Wang et al., 2017), a recommender system has been created in order to manage the energy consumption of the appliances of a smart grid residential users.

2.2 Data mining techniques

This section presents a review on the Clustering, Clustering and Forecasting techniques and their usage in the energy management systems.

2.2.1 Clustering techniques

The clustering process is used to discover and identify natural groups, also known as an unsupervised learning task. This process creates groups where the members of a particular cluster are more similar to each other than the members of a different one (Vale, Morais, Ramos, Soares, & Faria, 2011). The process of clustering can be divided into the following steps. First, the most characteristic features from the original data set are extracted and selected. Then, the clustering algorithm must be chosen according to the characteristics of the problem. After that, a clustering evaluation is made to judge the validity of the algorithm. Finally, the practical explanation for the clustering results is given.

Clustering algorithms are based on a distance criterion (i.e., reflecting similarity or dissimilarity properties) to arrange the data in groups. Usually, distance is used to recognize the relationship between data. For quantitative[1] data, the following distance functions are typically used: Minkowski Distance, Standardized Euclidean Distance, Cosine distance, Pearson Correlation Distance and Mahalanobis Distance. For qualitative[2] features, the following functions are used: Jaccard similarity and Hamming similarity (Xu & Tian, 2015).

Several works in the Power and Energy Systems domain use clustering technics. The work in (Ramos, Duarte, Soares, Vale, & Duarte, 2012), presents a comparison between clustering methods for identifying load profiles in smart grids context. In the publication (Ribeiro, Pinto, Silva, Ramos, & Vale, 2015), the authors studied where the clustering algorithms have been used to create data sub-groups according to their correlations. This work is carried out in the scope of the remuneration and tariff of VPP. In some cases, following a type of existing measurement in literature is not necessary for the clustering processes. Also, in (Spinola, Faria, & Vale, 2017), the processes of clustering are performed by considering the optimization of the resources scheduling.

2.2.1.1 Evolutionary clustering

Evolutionary Clustering algorithms are methods that use Evolutionary Algorithms (EA) in their frameworks. EAs will optimize an objective function, which may depend on the intra-cluster, connectivity, or other criteria. Evolutionary clustering can be used to determine the number of clusters of a data set automatically or to provide a Pareto front with solutions that satisfice more than one criterion (Hruschka, Campello, Freitas, & de Carvalho, 2009).

Regarding applications in the energy domain, in (Labeeuw, Stragier, & Deconinck, 2015), (Kwac, Flora, & Rajagopal, 2014) and (Chicco, Napoli, & Piglione, 2006), various clustering algorithms (namely, hierarchical clustering and K-Means) are used to segment customers with particular consumption behaviors. In (Faria, Spinola, & Vale, 2016), the aggregation of consumers using hierarchical clustering is done in order to define the best DR programs to apply to each group (cluster). The evolutionary clustering is involved in (Lezama, Rodriguez-Gonzalez, & de Cote, 2016), with the application of the Differential Evolution algorithm, for determining the load pattern energy consumption of customers. This information can be used to find niche clients and design tariff and remuneration strategies.

---

[1] Data that focuses on numbers and mathematical calculations and can be calculated and computed.
[2] Data concerned with descriptions, which can be observed but cannot be computed.

2.2.1.2 Hierarchical clustering

In hierarchical clustering, the basic idea is to construct the hierarchical relationship among data in a sequential manner. In a first step, let's suppose that each point in the data group represents an individual cluster. In the next iteration, the two neighbor's clusters join and form a new cluster. The process repeats itself until there is only one cluster (Xu & Tian, 2015).

2.2.1.3 K-Means and G-Means

The K-Means and G-Means are a kind of clustering algorithm based on partition. K-Means is the most used algorithm nowadays (Xu & Tian, 2015). The fundamental idea of how K-Means works is to verify a high intra-cluster similarity and a low inter-cluster similarity. The clusters similarity is measured regarding the mean value of objects in the cluster. First, the user needs to select K initial clusters centers. The algorithm makes a random distribution of the data by the number of pre-defined clusters, calculating the intra-cluster, and inter-cluster measurements. The process repeats itself iteratively until an arrangement that meets the stop criterion is found, and the solution is returned (Xu & Tian, 2015).

On the other hand, the operation of the G-Means is based on the K-Means, although it is not necessary to define the number of clusters. The G-means starts its aggregation with 1 cluster (using K-Means or any other clustering algorithm). After it finishes, a statistical test is performed in which it is tested if the intra-cluster distance follows a Gaussian distribution. If yes, the algorithm returns the number of ideal clusters. Otherwise, the process repeats itself increasing the number of clusters until the condition is verified (Elkan & Hamerly, 2003).

2.2.2 Classification techniques

The classification process consists in examining objects (defined by a set of features) and attributing them labels corresponding to one of the predefined classes. This problem is a supervised learning task where the output information is a discrete classification. In general, the process of classification consists of the construction of a classification model that can be applied to unclassified data. All the classification techniques use a set of training data to be able to generate the classification model. After that, when a new input is inserted, it will be attached to one of the existing classes of the classification model.

In (POPEANGA, 2015), the authors describe classification techniques that can be used for determining consumers profiles based on different variables, determining the possibility of purchase and install specialized equipment for renewable energy generation.

2.2.2.1 Artificial Neural Networks

Artificial Neural Networks (ANN) are a set of connected input-output networks where a determined weight is associated with each connection. This technique combines one input layer linked to one or more intermediate layer, which in turn are connected to one output layer. The learning process of an ANN is performed by adjusting the weights of each connection (B M Wilamowski & Hao Yu, 2010).

ANNs have been used in (Macedo, Galo, de Almeida, & de C. Lima, 2015) for the successful classification of the consumer load curves to create patterns of consumption.

2.2.2.2 Decision Trees

Decision Trees (DT) can transform a set of data into a set of rules, finding patterns in the data. The created rules by this process, divide the instances into separated classes in the best way to form the tree. Usually, the construction of a decision tree can have as many nodes as existing variables in the training data. If the variable presents a single standard without variation, it can be discarded. The creation of the rules for dividing the data is done iteratively, forming new smaller subgroups in each node. The division ends when a particular criterion is met (Kotsiantis, 2013).

In (Yu, Haghighat, Fung, & Yoshino, 2010), decision trees are used for classifying and accurately predicted building energy demand levels.

2.2.2.3 K-Nearest Neighbor

The k-Nearest Neighbor (k-NN) is a method used for classification and regression (Denoeux, 1995). The input consists of <u>k</u> closest training examples data in the feature space (with <u>k</u> as a positive integer). k-NN uses a metric (usually the Euclidean distance) to perform its search. When k-NN is used for classification, the existing training dataset is divided into classes. Therefore, for a new object, the output is a class membership determined by a majority vote of its <u>k</u> neighbors, with the object being assigned to the class most common among its k nearest neighbors.

2.2.2.4 Support Vector Machines

The Support Vector Machines (SVM) perform a nonlinear mapping (using kernel functions) of the data into a high dimensional space. The SVM is based on the idea of finding a hyperplane that divides a data set into two or more classes.

In (Jindal et al., 2016), an SVM was used for identifying the fraudulent consumers involved in electrical theft and classify the consumers as normal or malicious.

2.2.3 Forecasting techniques

One of the essential tasks of the energy operators is to control energy consumption and also have readiness for the amount of consumption in the following hours or days. Having a trustable estimated perspective of energy consumption is the most necessary information to reach this task. Many forecasting techniques can be used to predict values. Each of these techniques can forecast an acceptable value in different situations of energy consumption. Also, many different facts can help in order to have better forecasts value for energy consumption, such as available data with different variables and available historical data. Related to how the energy has been consumed, using different variables in the process of the prediction, such as environmental temperature, humidity, solar radiation and number of the consumers, can reduce the error of the forecasted value. Several forecasting techniques have been proposed, namely as, Time series, Artificial Neural Networks (ANN), Fuzzy Rule Base Systems (FRBS) and Support vector machines. The following sections explain the structure of these forecasting techniques.

2.2.3.1 Time Series Analysis

This type of methods takes a pattern of past events as a signal and uses it to forecast future events. In these techniques, the next events are only a function of the previous events. The ARMA (Auto-Regressive and Moving Average models) models are the best example of this technique which assumes that the future load at any particular time can be estimated by a linear combination of a few previous times. Or ARIMA (Autoregressive Integrated Moving Average), which is a generalization of an ARMA. In (Ediger & Akar, 2007) has been presented a case study using ARIMA to estimate the future primary energy demand of Turkey from 2005 to 2020. The ARIMA forecasting of the total primary energy demand appears to be more reliable than the summation of the individual forecasts. The results have shown that the average annual growth rates of individual energy sources and total primary energy will decrease in most of the cases.

2.2.3.2 Artificial Neural Networks (ANN)

This technique is the most common technique to forecast energy consumption. ANNs are inspired by the human brain and their number of neurons with high interconnectivity. ANNs are several combined nodes or neurons, divided into different levels, and interconnected by numeric weights. They resemble the human brain in two fundamental points: the knowledge being acquired from the surrounding environment, through a learning process; and the network's nodes being interconnected by weights (synaptic weights), used to store the knowledge. Every neuron implements a simple operation, the weighted sum of its input connections, that generate the exit signal that it sends to the other neurons. The network learns by setting up the connection weight in order to produce the desired output - the output layer values (Schaefer, Udluft, & Zimmermann, 2007). According to a large number of correct examples, ANN is able to change their connections until they generate outputs that are coincident with the right values. This way, ANN can extract basic rules from data (Bogdan M. Wilamowski & Yu, 2010).

Many studies have been published using ANN’s in order to predict different variables. For example, in (Vinagre, Gomes, & Vale, 2015) the Autor uses the ANN’s in order to forecast the energy consumption of an office building for the next 24 hours. in this work the energy consumption has been divided in to three values that each of them corresponds for the consumed value by a specific type of consumers. This way, ANN forecast three values for each hour and the sum of these values presents the final forecasted value. Also related to the type of consumers, some different variables are used in order to train the method. For example, the environmental variable is used to train the method to forecast the consumption of the air-conditioning systems, which has a positive influence on the accuracy of the final predicted value.

2.2.3.3 Fuzzy Rule Base Systems (FRBS)

Fuzzy rule-based systems (FRBSs) are based on the combinations of fuzzy logic with different techniques. Fuzzy logic has been proposed by Zadeh in 1965 (Zadeh, 1965). This logic represents the reasoning of human experts in production rules (a set of IF-THEN rules) to handle real-life problems from domains such as control, prediction and inference, data mining, bioinformatics data processing, robotics, and speech recognition. FRBSs are also known as fuzzy inference systems and fuzzy models (Riza, Bergmeir, Herrera, & Benítez, 2015).

As having the required information in the right format from human experts can be difficult, acquiring the knowledge by generating the fuzzy IF-THEN rules automatically from the numerical training data can be an alternative and effective way. There are two critical processes that must be considered while modelling an FRBS, which are structure identification and parameter estimation. Structure identification is a process to find appropriate fuzzy IF-THEN rules and to determine the overall number of rules, and parameter estimation is applied to tune parameters of membership functions. The FRBS is the combination of fuzzy logic and other techniques. These systems can be divided into five groups:

1. ***FRBS based on space partition***
2. ***FRBS based on neural networks***
3. ***FRBS based on a clustering approach***
4. ***FRBS based on genetic algorithms***
5. ***FRBS based on the gradient descent method***

There are several forecasting methods implemented based on one of these approaches. For example, Hybrid Neural Fuzzy Inference System (HyFIS) which is a combination of neural networks and fuzzy systems. This method is one of the most trustable methods to forecast energy consumption. For example, in (Jozi, Pinto, Praca, et al., 2016), has presented a case study which uses HyFIS for an hour-ahead energy consumption forecasting. In this work, the historical data of the energy consumption and the environmental temperature have been used to train the forecasting method. This method has a high capacity to take advantage of using various variables to predict energy consumption. Wang and Mendel’s Method (WM), which has been widely known because of being simple and having an excellent performance (X. H. Yang, Guo, Zhang, & Yang, 2008). An hour-ahead energy consumption forecasting study has been published in (Jozi, Pinto, Praça, et al., 2016) using this method. Also, Genetic Fuzzy

Systems for Fuzzy Rule learning based on the MOGUL methodology (GFS.FR.MOGUL) which is a combination of the fuzzy systems and genetic algorithms. In (Jozi, Pinto, Praca, et al., 2017), This method also has been used for an hour-ahead energy consumption forecasting and presents a more reliable result than the other FRBS's.

2.2.3.4 Support vector machines (SVM)

R. A. Fisher created the first algorithm for pattern recognition in 1936. The first running kernel of SVM only for classification problems was implemented by a generalization of the nonlinear algorithm Generalized Portrait that has been created by Vapnik and Lerner in the sequence of (A. L. V. Vapnik, 1963). This approach is implemented based on statistical learning theory (V. N. Vapnik, 1995) and the concept of the SVM can be tracked to the time that statistical learning theory was developed further with Vapnik, in 1979.

During the last decades, several applications of SVM can be found, both for classification and for regression problems. Some examples are: pattern recognition, image recognition, classification and regression analysis, text categorization, medical science, classification of proteins, weather forecast, wind speed prediction, energy prices forecast, among other practical applications (X. Yang, Tan, & He, 2014)(Chen, Shao, Deng, & Feng, 2014).

Several works use SVM in order to forecast the energy consumption of different places. The work presented in (Vinagre, Pinto, Ramos, Vale, & Corchado, 2016), implement the SVM method to forecast the energy consumption and analyzes the performance and accuracy this algorithm when it is implemented in different Frameworks.

2.3 Summary

This chapter presented the review on energy management systems in the building and the importance of this manage mange as well as the published studies in literature in recent years in this area. Different recommendation systems in the power energy area have been discussed the concept of context awareness recommender system has been introduced. Also, a brief review of the three principal data mining techniques has been included, considering different approaches in each case and several published studies using these techniques. These reviews prove that all these concepts have been explored and studies during past decades, however, few studies can be found which considers the combination of them all. Most of the energy management systems in building pay more attention to the cost and optimization of the energy and have less consideration on users comforts and the current context of the building. Meanwhile, the AI techniques introduce several essential approaches to bring enough intelligence to the system to be able to estimate the comforts and act based on this knowledge. Also, on another hand, context-aware systems present a great potential to improve the performance of a recommender system and AI techniques can make it possible to recognize this context. This way, this study proposes an intelligent energy management system in the buildings which uses several data mining techniques to estimate the comforts of the users and create different contexts. The system based on this generated information created recommendations about the state of the electrical appliances of the building.

3 Proposed solution

The objective of this project is to create an intelligent energy management system for a building which considers the comforts and habits of the residents as well as the energy consumption cost of the building. This system acts as a recommender system to comply with the comforts of the residents with the minimum consumption cost possible. This way, a sequence of data preparation and data mining techniques is proposed to generate the necessary estimation values. Additionally, a set of statistical rules have been created to receive the predicted data and make recommendations about the energy consuming appliances of the building. Figure 2 demonstrates a general perspective of this sequence.

the system starts by extracting the required data from the main database of the building. As these data might have some failures or the sensors may have measured some unreal data, the system uses a cleaning process to make sure the received data are complete and clean. After this phase, the collected data should be aggregated into specific time intervals that the system considers. The aggregated data will enter a clustering phase in order to be divided into groups based on different contexts. The created contexts will be used to choose the training data sets for the forecasting algorithms. The forecasting algorithms receive the selected data and generate the required estimated data to execute the created rulesets. As the forecasting process includes more than one forecasting method, at the end of the forecasting process, many predicted results are calculated. Automatic forecast selection component is responsible for receiving the forecasted data and finding the most trustable results to be used as the best-forecasted results. After selecting the predicted results, this data will be used to execute the created rulesets of the system to generate the recommendations. In the last phase, these recommendations will be inserted into the database of the system.

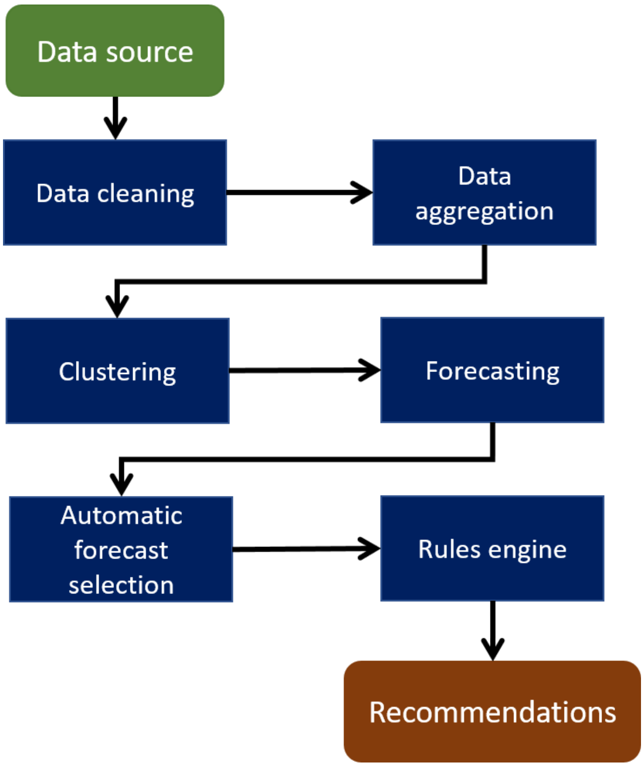

Figure 2 – The general aspect of the process to meet the recommendations

Figure 3 presents the created domain model for the proposed system. This domain model presents all the concepts and the relations of the system. Every time a new set of data needs to be forecasted, a forecasting and a clustering process will be created. The clustering process requires aggregated data sets which receive from the data aggregator and creates the clustering input tables based on the objective of the clustering and intended contexts. This component executes the K-Means algorithm and gets the results. The forecasting process receives the results from the clustering process and creates the required data tables to run the forecasting algorithms. In this phase, the structured data will be sent to all four forecasting methods of the system, namely as SVM, HyFIS, WM, and GFS.FR.MOGUL. The results of the methods will be stored in the database of the system. Before executing the rulesets, the best-forecasted results by the methods will be selected. The rules engine component receives the selected forecasted results and creates the recommendations based on the designed rulesets of the system and stores them into the created database of the system.

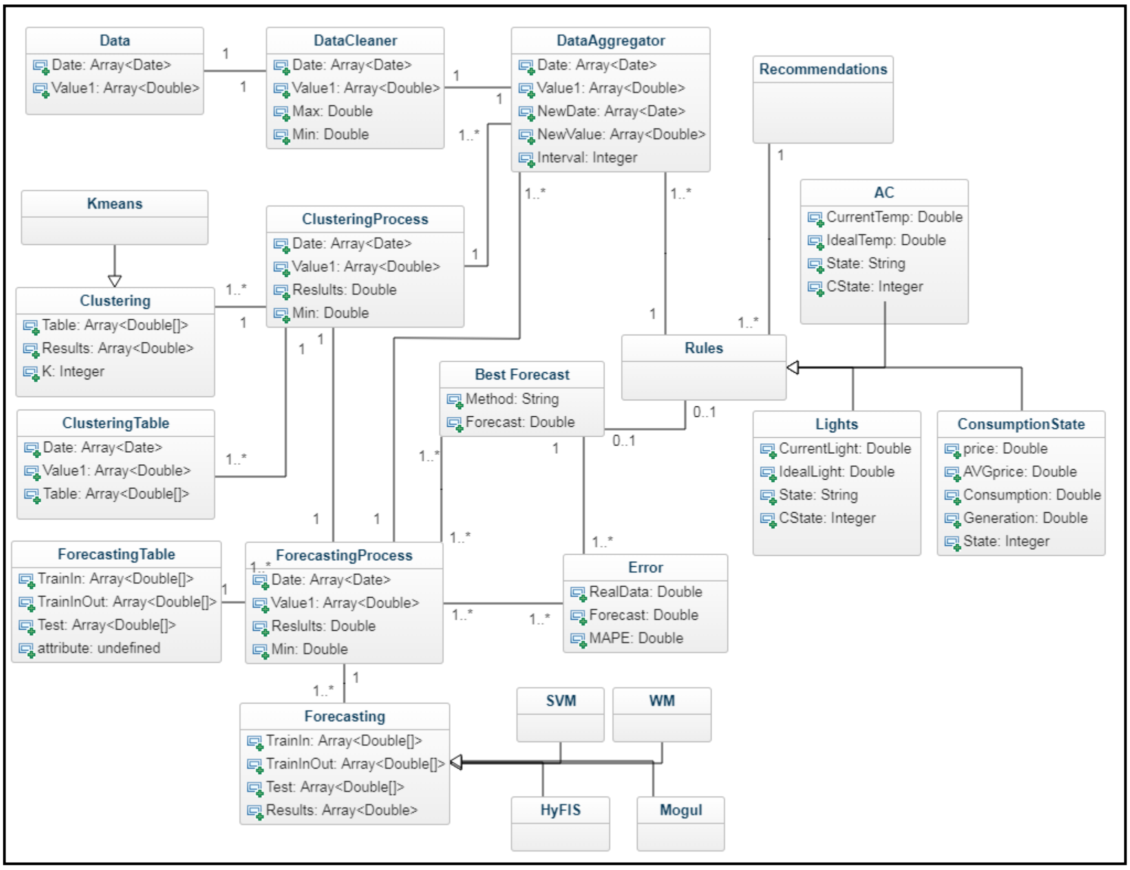

Figure 3 – Domain Model

The following sections present a detailed description of every phase of this system.

3.1 Data Cleaning

The model proposed in this work is conceived so as to be implemented for a building that includes sensors and energy meters to provide the necessary data for the system. These energy meters usually record the total power consumption of the building or power consumption of every room or zone separately. Also, the sensors of the building can record different types of information, such as movement, internal temperature, humidity, solar radiation, brightness, etc. These data should be stored in a central database in which the energy management system has permissions to extract the data. The stored data in this database mostly are received directly from the sensors or energy meters, which means that these data might have many failures or have recorded unreal data. For example, related to the accuracy of the sensors and the position that these sensors are mounted, it is possible that the sensor record a much higher or lower value than the real values. Even sometimes in the phase of receiving data from sensors to the database, some errors might happen, and the recorded data can be unreal. This way, the system includes a data cleaning process before using the received data. As has been explained, two types of problems can appear with data:

data failure during a time interval and unreal recorded data. In the case of failure data, the system recognizes the failed data and calculate the average of the last recorded value before and first value after the failure and uses this average as the value of the missed time interval. As can be been in equation (1), where V Is the value and n present the failure time interval:

$$IF(\,V_n = Null), \qquad then \;\; V_n = \frac{V_{n-1} + V_{n+1}}{2} \tag{1}$$

For the case of unreal data, the system has a maximum and minimum limit for every type of data that receives. In case that the received value is more than the maximum limit or less than the minimum limit, the system recognizes this value as an unreal data. The equation (2) presents this calculation:

$$IF\;(\,V_n < Min\;or\;V_n > Max), \qquad then \;\; V_n = Null \tag{2}$$

In this situation, the system deletes the unreal value and delete it from the data set. Then the system acts as same as a failure data situation to feel this row and uses the equation (1) to complete the data set.

3.2 Data aggregation

The received data from the database of the building can be recorded by any time interval, which is usually very short interval such as 1 to 10 seconds. The objective of the system is to generate recommendations for a much longer time interval such as 10, 15 or 20 minutes. Thus, these data need to be aggregated into a longer time interval. As the clustering or forecasting processes always call the data aggregation process, in every call it receives an integer number which indicates the duration of the target interval per minutes. Data aggregator component receives the cleaned data and creates a new data set with the received time interval. The way that the new values are calculated is related to the type of collected data. For example, for variables such as power consumption, temperature or brightness, the average of the values during each interval is used as the final value. Or in the case of the movement sensor, only if one action has been recorded during the time interval, the final value will present detected movement during the interval.

3.3 Clustering

The principal objective of using clustering technique in this system is to create data contexts that correspond to different power consumption situations in the building. This component includes the creation of specific data structures as the input of the clustering process as well as the implementation of the K-Means method. In order to be able to allocate the application of the clustering method in a different server, a web service based on Java programming language has been created, which includes the implementation of K-means clustering algorithm in R programming language. This way the algorithm is accessible for different purposes and also is able to receive many requests at the same time. Figure 4 presents an overview of the clustering component.

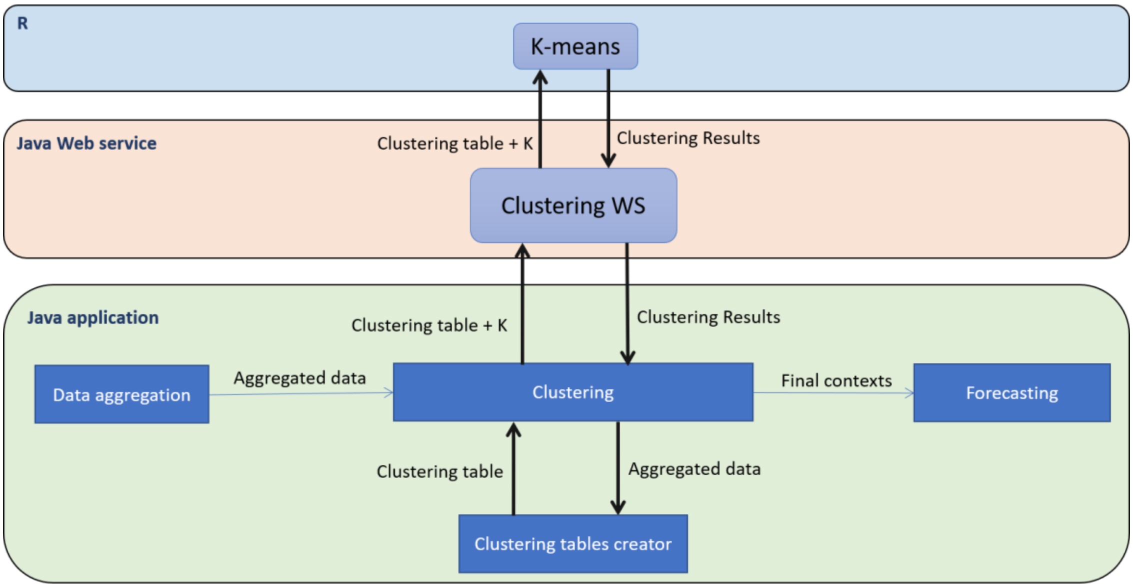

Figure 4 - Overview of the clustering component

As one can see in Figure 4, the received aggregated data is firstly sent to the clustering table creator. In this part, the data will be selected based on the aim of the clustering and structured into a format that is acceptable for the created web service. Then the created set of data will be sent to the web service as well as the number of the intended clusters (K). The web service executes the K-Means algorithm in R and returns the results.

The system is able to create data groups based on any types of contexts according to the trends of the data, such as days or hours with high temperature; high activity, etc. The selected contexts and number of clusters in each case is relative to the forecasting target and the available data. The system creates data groups based on the considered contexts and the data associated to the same context as the context of the forecasting target will be sent to the forecasting component to be used as the training data.

To implement the K-Means method the system uses the "cluster" library of R language (Maechler, Rousseeuw, Struyf, & Hubert, M. Hornik, 2014). This method is executed through a Java-based web server, developed for this purpose. This web service reserves the clustering input table in a JSON format and runes the K-Means method in R using these input data.

The K-Means clustering methodology considers a set of observations ($x_1, x_2, ..., x_n$), where each observation is a d-dimensional real vector, and n is the number of considered observations.

The clustering process aims at partitioning the n observations into k (≤ n) clusters $C = \{C_1, C_2, \ldots, C_k\}$ so that the Within-Cluster Sum of Squares (WCSS) is minimized (3).

$$\min \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \tag{3}$$

With the objective of minimizing equation (3), the clustering process executes an iterative process between two steps: (i) the assignment step, where each observation x_p is assigned to the cluster $C^{(t)}$ whose mean value yields the minimum WCSS in iteration t, as presented in (4); and (ii) the update step, where the new means of each cluster are calculated, considering the newly assigned observations, determining the new centroid μ_i of each cluster, as in (5).

$$C_i^{(t)} = \{x_p : ||x_p - \mu_\mathrm{i}^{(\mathrm{t})}||^2 \leq ||x_p - \mu_\mathrm{j}^{(\mathrm{t})}||^2 \ \forall j, 1 \leq j \leq k\} \tag{4}$$

$$\mu_\mathrm{i}^{(\mathrm{t+1})} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j \tag{5}$$

The execution of the algorithm stops when the convergence process is completed, i.e. when the assignments of observations to different clusters no longer change. By minimizing the WCSS objective, in equation (3), the K-Means clustering methodology assigns observations to the nearest cluster by distance. It means that each subject will be grouped in the same cluster as the more similar ones.

3.4 Forecasting

In a forecasting process, the training data selection is one of the most important fact to have a trustable result; however, it not the only point. To be able to achieve the best possible results, many different facts such as forecasting method selection, input data structure and the configuration of the methods should be considered. This way, four forecasting and one classification algorithms have been chosen to be implemented in this system namely Hybrid Neural Fuzzy Inference System (HyFIS), Wang and Mendel's Method (WM) and Genetic Fuzzy Systems for Fuzzy Rule learning based on the MOGUL methodology (GFS.FR.MOGUL) for forecasting and Support Vector Machine (SVM) for forecasting and classification. These methods have been chosen to be able to take advantage of using different forecasting technique and use the most exact predicted values. As same as the clustering method, forecasting methods are also implemented in R programming language, and two Java-based web services have been developed to execute the forecasting and classification process. Figure 5 presents a perspective of the forecasting component in this system.

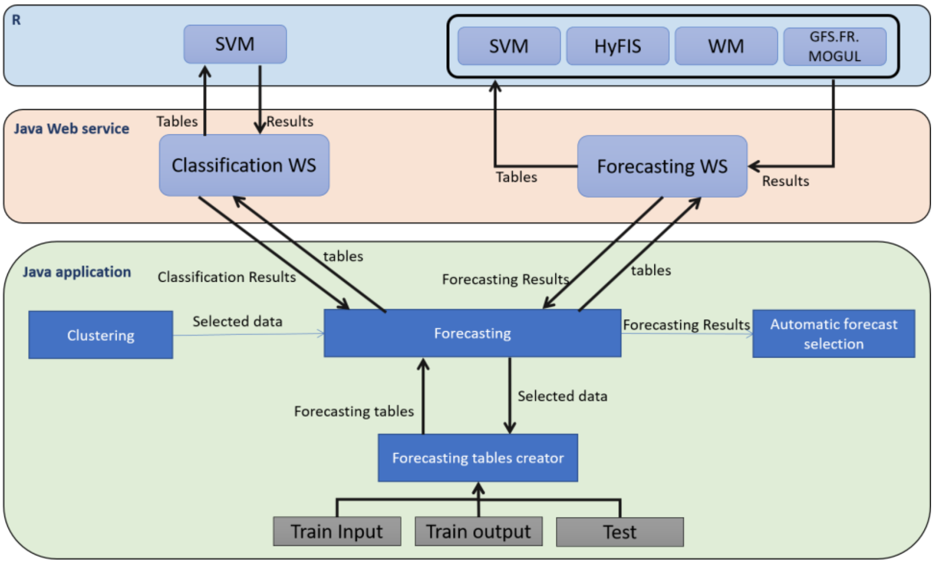

Figure 5 - A brief perspective of the forecasting component

As can be seen in Figure 5, to run the forecasting methods, three data sets are required: Train input, Train output and Test. Train input and Train output tables include the training data for the methods. These data are extracted from the received selected data from the clustering component. The test table is the main input of the methods and includes the information related to the target value of forecasting. The classification method also uses the same data structure and tables to estimate the final value. Related to the type of the target variable the system decides to use the forecasting or classification technique. The results will be sent to the automatic forecast selection component to be stored, analyzed and finally be used to execute the rulesets.

The forecasting and classification web services of this implementation are developed on Java programming language and connected to the R programming language. These web services use the available R libraries to implement the following forecasting and classification methods.

3.4.1 Support Vector Machines (SVM)

Support Vector Machine (SVM) techniques are one of the most popular classifications and regression methods. SVM application using Kernels are wildly used for reasons such as: often concentrating on convex problems; allowing many linear algebra techniques to be used in a non-linear way; have shown robustness in many application domains and spend fewer resources and half the time of artificial neural networks.

The input data for the SVM method must follow this structure (Pinto, Sousa, Praça, Vale, & Morais, 2016):

$$(y_1, \mathbf{x}_1), \ldots, (y_l, \mathbf{x}_l), x \in R^n, y \in R, \tag{6}$$

In this formula, each example x_i is a space vector example, and y_i is corresponding value for x_i and n is the size of training data. For classification: y_i assumes finite values; in binary classifications: $y_i \in \{+1,-1\}$; in digit recognition : $y_i \in \{1,2,3,4,5,6,7,8,9,0\}$; and for regression purposes, y_i is a real number ($y_i \in$ R).

The implementation of SVM requires considering some critical aspects, namely:

Feature Space: the method that can be used to construct a mapping into a high dimensional feature space by the use of reproducing kernels. The idea of the kernel function is to enable operations to be performed in the input space rather than the potentially high dimensional feature space. Hence the inner product does not need to be evaluated in the feature space. This provides a way of addressing the curse of dimensionality. However, the computation is still highly dependent on the number of training patterns, and good data distribution for a high dimensional problem generally requires large training sets.

Loss Functions: In statistics, the decision theory and machine learning, the loss function is a function that maps an event to a real number, representing some "costs" associated with the difference between the estimated and the actual data for an occasion. The purpose of this function is to modulate the input data, when applied to a training set, and then forecasting the values (or sorting). The loss function uses the forecast values and compares how much they deviate from the actual values, quantifying the deviation.

Kernel Functions: The kernel functions, in general, are a set of algorithms for pattern examination. The main task is to find patterns and study the type of associations, in a particular pattern (e.g, groups, classifications, major components, correlations, classifications) for general types of data (such as sequences, text documents, sets of points vectors, images, etc). The kernel function approaches the problem by mapping the data to a dimensional space, where each coordinate corresponds to a characteristic of each input value transforming the data into a set of points in Euclidean space. Some examples of kernels are Polynomial, Gaussian Radial Basis Function, Exponential Radial Basis Function, Multi-Layer Perceptron, Splines, B splines (Brereton & Lloyd, 2010).

Non-linear Regression: Likewise, to classification problems, a non-linear model is typically required to adequately model data. In the same method as the non-linear Support Vector Classification (SVC) approach, a nonlinear mapping can be used to map the data into a high dimensional feature space where linear regression is executed. The kernel approach is again employed to address the curse of dimensionality.

In this system, an R based library called “e1071” is used to implement the SVM method for bouts classification and forecasting. (Meyer, Evgenia Dimitriadou, Kurt Hornik, Leisch, Chang, & Lin, 2019)

3.4.2 Hybrid Neural Fuzzy Inference System (HyFIS)

HyFIS is one of the most known fuzzy rule-based systems. The learning method in HyFIS includes two phases, which are described as follows (Kim & Kasabov, 1999):

- The first phase concerns the structure learning, i.e. finding the rules by using the knowledge acquisition module;
- The second phase regards the parameter learning phase for tuning fuzzy membership functions to achieve a desired level of performance (Pedrycz, Ekel, & Parreiras, 2010).

One of the advantages of this approach is that the fuzzy rule base can be easily updated when there is new available data (L. X. Wang & Mendel, 1992). When there is a new available pair data, a rule is created for this data, and the fuzzy rule base is updated by this new rule, as shown in Figure 6.

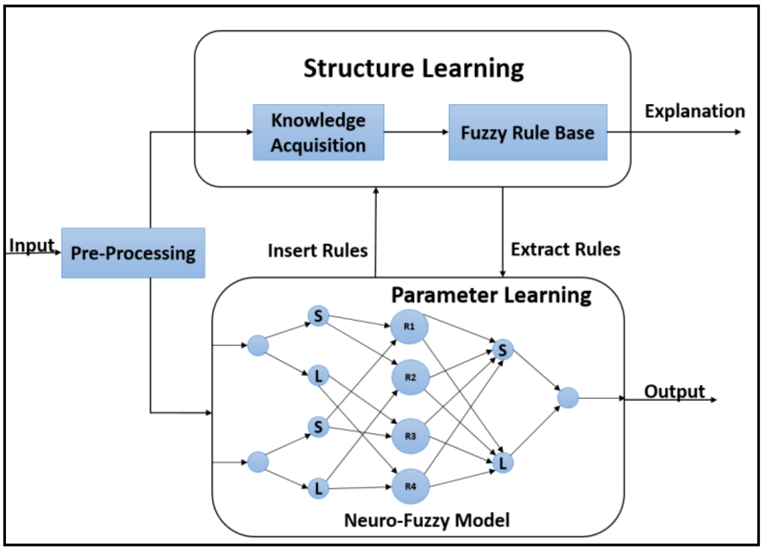

Figure 6 - General schematic diagram of the HyFIS (Jozi, Pinto, Praça, et al., 2017)

In the learning phase, the neuro-fuzzy model in the HyFIS uses a multilayered perceptron (MLP) network based on a gradient descent learning algorithm for adapting the parameters of the fuzzy model (Rudd, Muro, & Ferrari, 2014). The architecture simplifies learning from data and approximate reasoning, as well as knowledge acquisition. It allows using the combination of both numerical data and fuzzy rules, thus producing the synergistic benefits associated with the two sources.

The proposed neuro-fuzzy model in the HyFIS is a multilayer ANN, based on a combination with fuzzy systems. The system includes five layers, as shown in figure 7. In this structure, the input and output nodes are the input state and output control/decision signals. The nodes in the hidden layers detain the responsibility of representing the membership functions and rules.

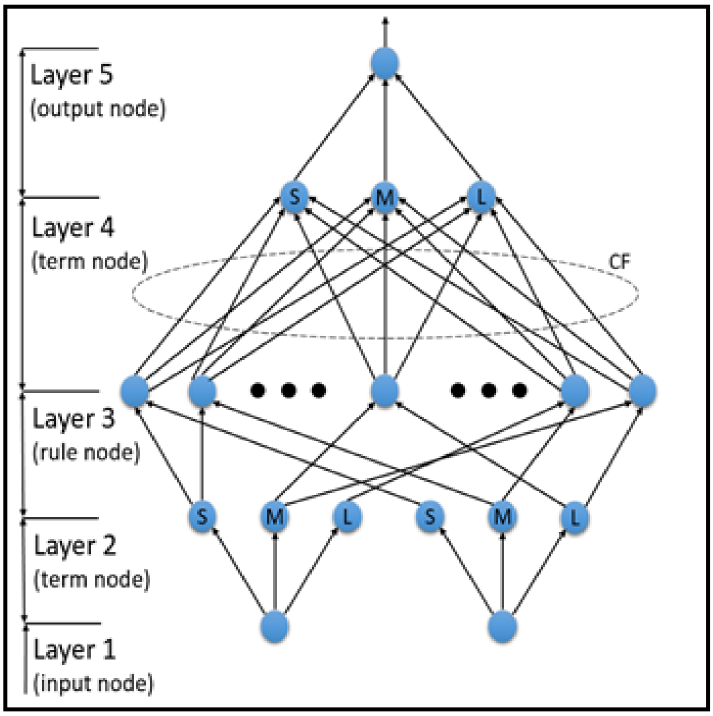

Figure 7 – The structure of the Neuro-Fuzzy model from the HyFIS architecture (Jozi, Pinto, Praça, et al., 2017).

In the first layer, the nodes are the inputs that transmit input signals to the next layer. In the second and fourth layers, nodes are the term nodes. These nodes act as membership functions to express the input-output fuzzy linguistic variables. In these layers, the fuzzy sets defined for the input-output variables are represented as large (L), medium (M), and small (S). However, for some applications or specific cases, these can be more specific and represented as, e.g. large positive (LP), small positive (SP), zero (ZE), small negative (SN), and large negative (LN). In the third layer, each node is a rule node and represents one fuzzy rule. The connection weights between the third and fourth layer represent certainty factors of the associated rules, i.e. each rule is activated to a certain degree controlled by the weight values. Finally, the fifth layer detains the node that represents the output of the system.

3.4.3 Wang and Mendel's Fuzzy Rule Learning Method (WM)

The WM model (L. X. Wang & Mendel, 1992) has been widely known because of being simple and having a good performance (X. H. Yang et al., 2008). This method is based on working with an input-output data set, as in equation (7).

$$E = \{e_1, \dots, e_p\}, e_1 = \{x_1^l, \dots, x_n^l, y^l\} \tag{7}$$

The generation of the fuzzy rules bases is put into effect using the following steps (Casillas, J.; Córdon, O.; Herrera, 2000):

i. Divide the Input and Output Spaces into Fuzzy Regions

It may be obtained from the expert information (if it is available) or by a normalization process. If the latter is the case, perform a fuzzy partition of the input variable spaces dividing each universe of discourse into several equal or unequal partitions, select a kind of membership function and assign one fuzzy set to each subspace. In our case, we will work with symmetrical fuzzy partitions of triangular membership functions (Figure 8).

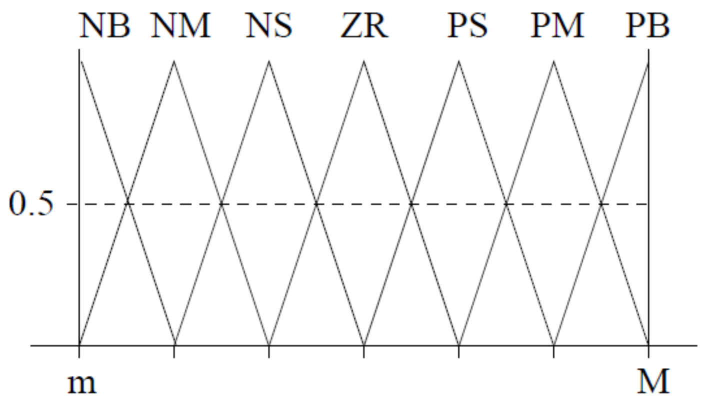

Figure 8 - Graphical overview of a uniform fuzzy separation (X. H. Yang et al., 2008).

ii. Generate Fuzzy Rules from Given Data Pairs

In this step, the generation of fuzzy rules includes the training data, using the database from Step 1. First, the degrees of the membership functions are calculated for all values in the training data. For each instance in the training data, a linguistic term having a maximum degree in each variable will be determined. Moreover, the process will be repeated for every instance in the training data to construct fuzzy rules covering the training data (Riza et al., 2015).

iii. Assign a degree to each rule

Determine a degree for each rule. The degrees of each rule are determined by aggregating the degree of membership functions in the antecedent and consequent parts. In this case, we are using the product aggregation operators. If $R_1 = IF\ x_1\ is\ A_1\ and \dots and\ x_n\ is\ A_n\ THEN\ y\ is\ B$, be the linguistic rule generated from the example $e_l, l = 1, \dots, p$. The importance degree associated with it will be obtained as equation (8):

$$G(R_l) = \mu A_1(x_1^l), \dots, \mu A_1(x_n^l). \mu B(y^l) \tag{8}$$

iv. Create a Combined Fuzzy Rule Base

The ρ candidate rules are first divided into g different groups, each one of these groups composed of all the candidate rules presenting the same antecedent. It will note by R_{ji} , the *j*-

th rule in the i-th group. To produce the final rules, the rule with the highest importance degree is chosen in each group i, $i = 1, .., g$. Hence, g will be both the number of different antecedent combinations in the candidate ruleset and the number of linguistic rules in the final generated rule base.

Interpolative reasoning that FRBS develops is one of the exciting characteristics of this method, which plays a key role in the high performance of FRBSs. It is a consequence of the cooperation among the fuzzy rules composing the Knowledge Base. The FRBS output is not usually because of a single fuzzy rule, and due to the cooperative action of several fuzzy rules that have been exported, they match the system input to any degree. Regarding the rule with the best performance in each subspace and, the covering degree, WM Method hooks up the example data set into the fuzzy subspaces (the antecedent combinations mentioned in step 4 of the algorithm). Consequently, the global interaction among the rules of the Knowledge Base is not mentioned. This causes that the obtained ruleset, do not cooperate appropriately. Therefore, these rules make the method more sensitive to noise due to local processing.

3.4.4 Genetic Fuzzy Systems for Fuzzy Rule learning based on the MOGUL methodology (GFS.FR.MOGUL)

GFS.FR.MOGUL is a forecasting method that implements a genetic algorithm determining the structure of the fuzzy IF-THEN rules and the membership function parameters. Two general types of fuzzy IF-THEN rules are considered:

- ***Descriptive rules;***
- ***Approximate/free semantic approaches.***

In the first type, the linguistic labels illustrate a real-world semantic and are uniformly defined for all rules. In contrast, in the approximate approach, there is no associated linguistic label.

Modelling a fuzzy IF-THEN rule on a chromosome which consists of the parameter values of the membership function shows that every rule has the own membership function values. A population contains many such generated chromosomes, based on the iterative rule learning approach (IRL). IRL means that the chromosomes will be produced one by one, taking into account the fitness value and covering factor, until there are sufficient chromosomes in the population. After having obtained the population, the genetic algorithm is started, using the genetic operators' selection, mutation, and crossover (Riza et al., 2015).

Many important statistical properties must be considered by the Fuzzy Rule Base (FRB) to obtain an FRB System FRBS presenting good behavior (Dimiter Driankov, Hans Hellendoorn , Michael Reinfrank, L. Ljung, R. Palm , B. Graham, 1996). In the Generated Fuzzy Rule Bases (GFRB) obtained from MOGUL, there will be considered the satisfaction of two of these statistical properties, which are:

- ***completeness***
- ***consistency***

As an inductive approach to building GFRBSs is considered, both features will be based on the existence of a training data set, Ep, composed of p numerical input-output problem variable pairs. These examples will present the following structure Equation (9):

$$e_l = \left(ex_1^l, ex_2^l, \dots, ex_n^l, ey^l\right), \quad l = 1, \dots, p \tag{9}$$

i. Completeness of a Fuzzy Rule Base

As it is explained in (Casillas, J.; Córdon, O.; Herrera, 2000), it is clear that an FRBS most be always able to infer a proper output for all possible system input. This property is called τ-completeness in the field of inductive learning, and it may be mathematically formulated using a real value t by means of the following expressions, equation (10), (11) and (12) (Herrera, Lozano, & Verdegay, 1995):

$$C_R(e_l) = \bigcup_{i=1\dots t} R_i(e_l) \geq \tau, \quad l = 1, \dots, p \tag{10}$$

$$\left[R_i(e_l) = * \left(A_i\left(ex^l\right), B_i(ey^l)\right)\right] \tag{11}$$

$$A_i\left(ex^l\right) = * \left(A_{i1}\left(ex_1^l\right), \dots, A_{in}(ex_n^l)\right) \tag{12}$$

Where * is a t-norm, and R_i (e_l) is the *compatibility degree* between the rule R_i and the example e_l .Given an FRB composed of *T* fuzzy rules R_i, the *covering value* of an example $e_l \in E_p$ is defined as equation (13):

$$CV_R(e_l) = \sum_{i=1}^{T} R_i(e_l) \tag{13}$$

And also, the following condition equation (14) is required:

$$CV_R(e_l) \geq \epsilon, \quad l = 1, \dots, p \tag{14}$$

A good FRB must satisfy both the conditions presented above, to verify the τ-completeness property and to achieve an appropriate final covering value.

ii. Consistency of a Fuzzy Rule Base

A generic set of IF-THEN rules is *consistent* if it does not contain contradictions. It is necessary to relax the consistency property to consider the fuzzy rule bases. And it will be done using the positive and negative examples concepts (Lezama et al., 2016). Equations (15) and (16) are examples of the positive and negative set for the rule R_i.

- ***Positive:***

$$E^+(R_i) = \left\{ e_l \in \frac{E_p}{R_i(e_l)} \geq 0 \right\} \tag{15}$$

- ***Negative:***

$$E^-(R_i) = \left\{ e_l \in \frac{E_p}{R_i(e_l)} = 0 \; and \; A_i(ex^l) > 0 \right\} \tag{16}$$

And by giving the value k $[0,1]$ and equations (17) and (18):

$$n_{R_i}^+ = |E^+(R_i)| \tag{17}$$

$$n_{R_i}^- = |E^-(R_i)| \tag{18}$$

We get that:

$$R_i \; is \; k - consistent \; when \; n_{R_i}^- \leq k . n_{R_i}^+ \tag{19}$$

So, the way to incorporate the satisfaction of this property in the designed GFRBSs is to encourage the generation of *k*-consistent rules. Those rules not verifying this property will be penalized so as not to allow them to be in the FRB finally generated.

3.5 Automatic Forecast Selection

This component of the system is responsible for receiving the forecasting result, storing these results, calculating the errors, saving the errors and selecting the most trustable forecasted value to send to the rule engine. The proposed system includes a PostgreSQL database server to store the results. This server has two databases, one for the forecasting results and recommendations and one for the errors of the forecasted values. First, the system receives the forecasted results and store this results in Result DB. At the end of every period when the real data for the forecasted value is available, the system receives the actual data and calculates the Mean Absolute Percentage Error (MAPE) of the forecasted values and inserts this error into the Error DB. The MAPE error is calculated by the equation (20), where A_t is the real value and F_t presents the forecasted value.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| * 100 \tag{20}$$

Therefore, all of the forecasted values and the errors along the time will be stored in these databases. Figure 9 shows the structure of this phase of the system.

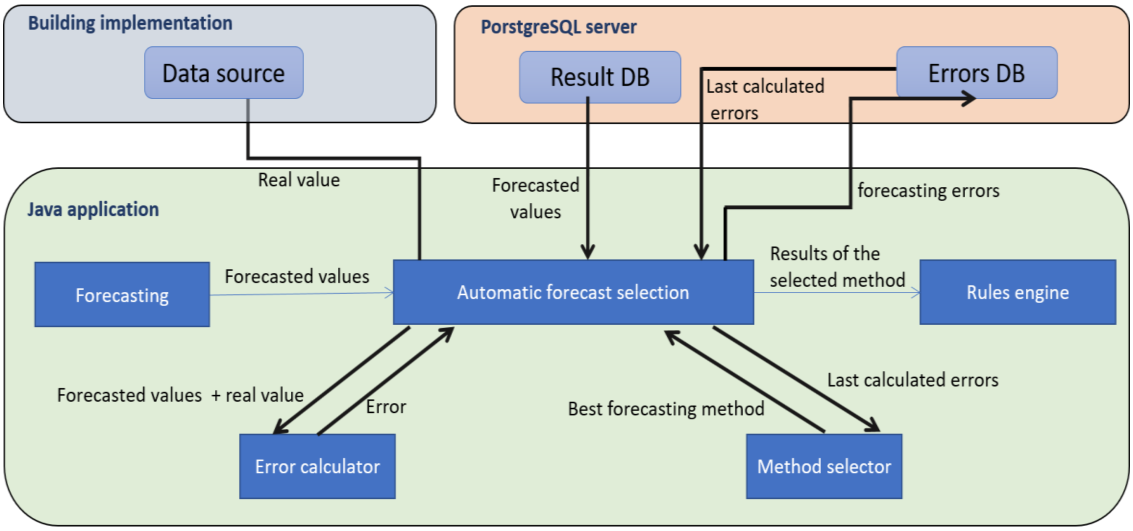

Figure 9 - A brief perspective of the automatic forecast selection component

The system includes four forecasting methods and generates four estimated values for every variable. In this phase, the system must decide to use one of these values to execute the rules. This way, the system selects the last 20 recorded errors of each forecasting method and calculates the average of every method. The method selection will be based on the equations (21), (22) and (23):

$$E_{m,n} = \{error_{m,n}\}$$

m=$\{HyFIS, WM, GFS.FR.Mogul, SVM\}$, n= {1, …, 20} (21)

$$AvgE_m = \frac{\sum_{n=1}^{20} E_{m,n}}{20}, m = \{HyFIS, WM, GFS.FR.Mogul, SVM\} \tag{22}$$

$$best_m = \min(AvgE_m) \tag{23}$$

Where E Includes the last 20 errors of every forecasting method and $AvgE$ is the average of these 20 errors for every method. The method that has the lowest $AvgE$ is recognized as the most trustable method at the current iteration, and the forecasted value by this method will be sent to the Rules engine.

3.6 Rules engine

The final step of the system is to use the generated data form the forecasting methods as well as the available real data to create recommendations about the usage of energy consuming devices of the building. The forecasted data, which are used to create the recommendations, are based on the created contexts by the clustering algorithms. So, the generation process of the recommendations considers the different contexts of the input data, which makes this system a Context-Aware Recommender System (CARS) (Adomavicius & Tuzhilin, 2015).

The generated recommendations can be about the state of the power consumption systems of the building, electrical appliances or any controllable equipment in the building. For any types of devices, if the information and data of the effective variables on the usage of the device are available, it is possible to create a new ruleset and generate new recommendations for the intended device. Related to the needed information to run the rules, the rest of the system is configured to generate this data. The rules and the recommendations are meant to adapt the actions of the building according to the comfort and habits of the residents. But also, these recommendations must consider the state of the power consumption in the building and the cost that the recommended action brings. This way, the first ruleset of the system is developed to recognize the consumption state of the building. The detailed explanation of this ruleset can be found in the following sections. The consumption state ruleset is the only ruleset of the system that is not related to the type of target recommendations. For any building and different types of devices, the system must have this ruleset.

As has been explained, the system is able to include a ruleset for every type of consumption system or devices in the building. In this work, the system considers the air conditioning devices and lights of a building as the targets of the recommendations and includes a ruleset for each one of these energy consumers. The following sections explain the structure and development of these three rulesets.

3.6.1 Power consumption state ruleset

The power consumption state of the building is a relative variable that represents the feasibility of the building to consume more energy. This variable is corresponding to the estimated consumption and generation of the building during the next period and the current electricity price. When this value is higher, it means that at the moment an additional consumption has a higher cost. Table 1 presents the input variables of this ruleset.

Table 1 – Input variables for power consumption state ruleset

Variable	Description
Consumption forecast (CF)	Forecasted power consumption in the building during the next time period
Generation forecast (GF)	Forecasted power generation in the building during the next time period
Current price (CP)	Current electricity market price received from
Average price (AvgP)	Daily average electricity market price received from

The electricity market price is considered a high price when it is more than the daily average price and low when it is equal or lower than the average price. The rules of this ruleset are presented in table 2 and can be explained as:

- ***If the forecasted generation is higher than the forecasted consumption, then Consumption State is 1.***
- ***If the forecasted consumption is higher than the forecasted generation and the current price is low, then Consumption State is 2.***
- ***If the forecasted consumption is higher than the forecasted generation and the current price is high, then Consumption State is 3.***

Table 2 – Consumption state ruleset

Conditions	Consumption State (CS)
(GC > FC)	1
(GC <= FC) & (CP <= AvgP)	2
(GC <= FC) & (CP > AvgP)	3

Based on the created rules the consumption state of the building is selected at each moment. This value will be used in the other rulesets to calculate the comfort radius in the case of any device.

3.6.2 Air conditioning system ruleset

For the air conditioning (AC) ruleset the objective is to recommend the state of the AC (ON or OFF) for the next period. The system first needs to predict the activity in the rooms and the ideal temperature for the users during the upcoming period. These values are predicted by the previous phases of the system, using the historical data of the room. Table 3 presents the list of the input variable for this ruleset.

Table 3 – List of input variables for AC ruleset

Variable	Description
Consumption State (CS)	The power consumption state of the building calculated by Consumption State ruleset
Forecasted Activity (Act)	The forecasted activity in the room for the next time interval. It is a YES or NO value.
Forecasted ideal temperature (FT)	The forecasted ideal temperature for the users in the next period
Current temperature (CT)	The current temperature of the room
Temperature Comfort basis Rate (TCR)	Basis rate at which the temperature comfort interval will be variated
Temperature Comfort Interval (TCI)	The Temperature value interval that the system recognizes as the comfort temperature for the users

The first step is to calculate the Temperature Comfort Interval (TCI). This interval presents the temperature values that the system recommends as the ideal temperature. This interval is based on the predicted ideal temperature, Consumption State of the building and TCR value, which is considered as 1ºC. When the cost of additional consumption in the building is high, the TCI must be a larger interval. So, the possibility of recommending to turning on the AC will increase. The equation (24) calculates the TCI.

$$\mathrm{TCI} = [\mathrm{FT} - (\mathrm{TCR} * \mathrm{CS})\,, \mathrm{FT} + (\mathrm{TCR} * \mathrm{CS})] \tag{24}$$

Figure 10 shows the TCI for different values for Consumption State.

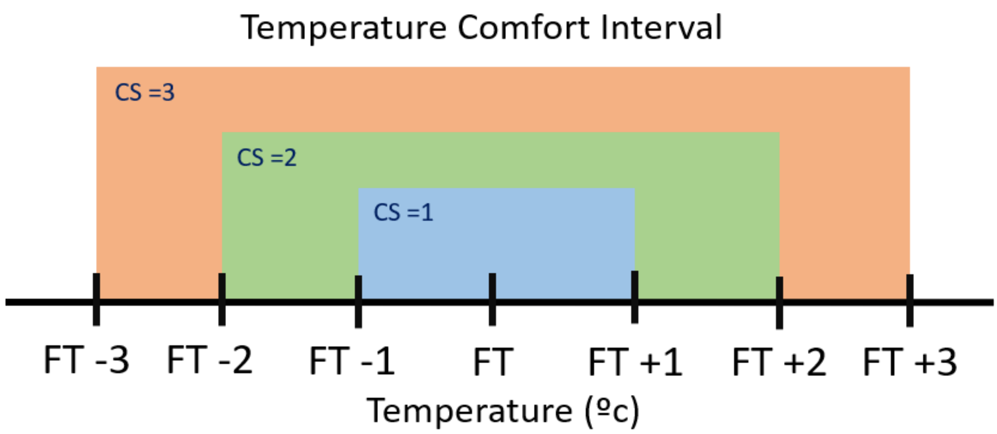

Figure 10 - Temperature Comfort Interval in case of every Consumption State

After detecting the Temperature Comfort Interval, the system uses the following rules to make the recommendation.

- ***If the forecasted activity is NO, then the AC state should be "OFF".***
- ***If the forecasted activity is YES and the current temperature is in temperature comfort interval, then the AC state should be "OFF".***
- ***If the forecasted activity is YES and the current temperature is NOT in temperature comfort Interval, then the AC state should be "ON".***

Figure 11 presents the decision tree that represents these rules.

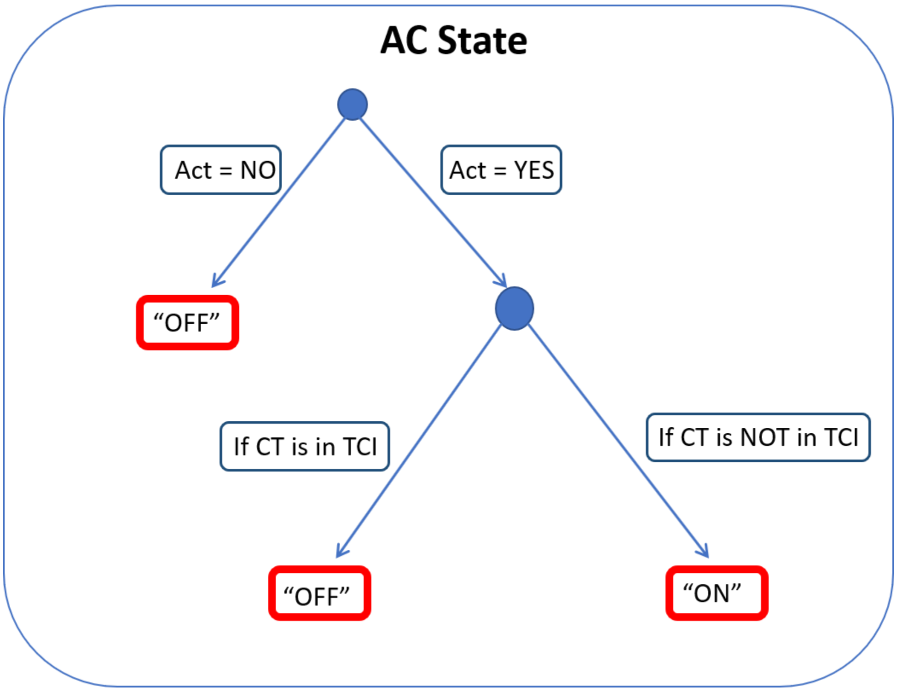

Figure 11 – AC ruleset decision tree

Based on these rules the system only recommends turning on the AC when the current temperature of the room is not in the TCI. In the end, the final recommended AC state, Ideal Temperature (FT), Maximum and Minimum value of TCI will be inserted into the Result DB and available to be activated.

3.6.3 Brightness ruleset

As same as the AC ruleset, the objective of the brightness ruleset is to recommend the state of the lights in the building or rooms. For this ruleset, it has been considered that the intensity of the lights in the target building is controllable, and the residents are able to select the intensity of the light between 0% to 100%. Therefore, the recommendations for the intensity of the lights will recommend one of possible four states: OFF, REDUCE, KEEP and INCREASE. Before executing this ruleset there are a set of variables that must be generated by the forecasting component of the system such as estimated ideal brightness and the activity of the room during the next period. The list of all input variables for this ruleset is presented in Table 4.

Table 4 - List of input variables for the AC ruleset

Variable	Description
Consumption State (CS)	Power consumption state of the building calculated by Consumption State ruleset
Forecasted Activity (Act)	Forecasted activity in the room for the next time interval. It is a YES or NO value.
Forecasted ideal Brightness (FB)	Forecasted ideal brightness in the room during next time period
Current Brightness (CB)	Current brightness in the room
Brightness Comfort basis Rate (BCR)	Basis rate at which the brightness comfort interval will be variated
Brightness Comfort Interval (BCI)	Brightness percentage interval that the system recognizes as the comfort percentage for the users

The Brightness Comfort Interval (BCI) calculation in this ruleset uses the same approach as the AC ruleset to calculate the minimum value. However, when considering the maximum value, the ideal forecasted brightness is always considered. This occurs because, if the current brightness is higher than the ideal brightness and even if the generation of the building is more than its consumption and the power consumption cost of the building is zero, there is no reason to consume energy to have this higher brightness; so the light intensity of the room should be reduced. Consumption State (CS) of the building is used to calculate the minimum value of this interval as well as the Brightness Comfort Rate (BCR) which is considered as 5%. BCI calculation formula is as presented in (25).

$$\mathrm{BCI} = [\mathrm{FB} - (\mathrm{BCR} * \mathrm{CS})\,, \mathrm{FB}] \tag{25}$$

The possible BCIs for different consumption states is shown in Figure 12.

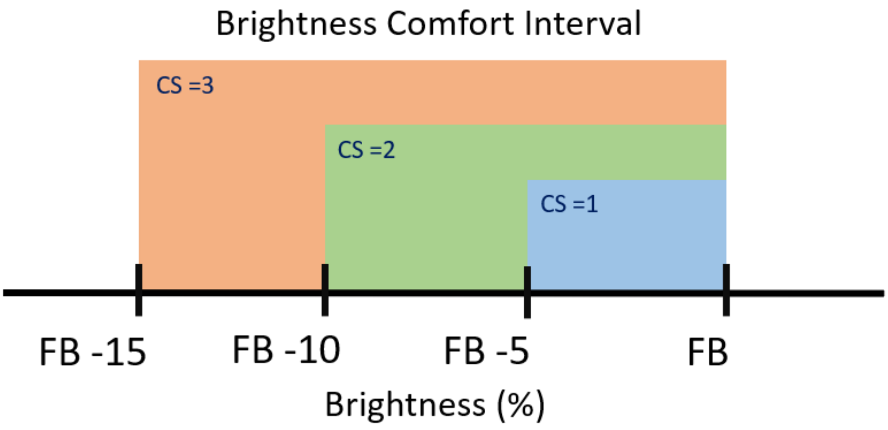

Figure 12 - Brightness Comfort Interval in case of every Consumption State

The following step is to use the calculated BCI to run the rules and find the best state for the Lights. The following rules have been created for this purpose.

- ***If the forecasted activity is NO, then the state of the lights should be "OFF".***
- ***If the forecasted activity is YES and the current brightness value is higher than the maximum value of Brightness Comfort Interval, then the state of the lights should be "REDUCE".***
- ***If the forecasted activity is YES and the current brightness value is the Brightness Comfort Interval, then the state of the lights should be "KEEP".***
- ***If the forecasted activity is YES and the current brightness value is lower than the minimum value of Brightness Comfort Interval, then the state of the lights should be "INCREASE".***

Figure 13 presents the decision tree that leads to these rules.

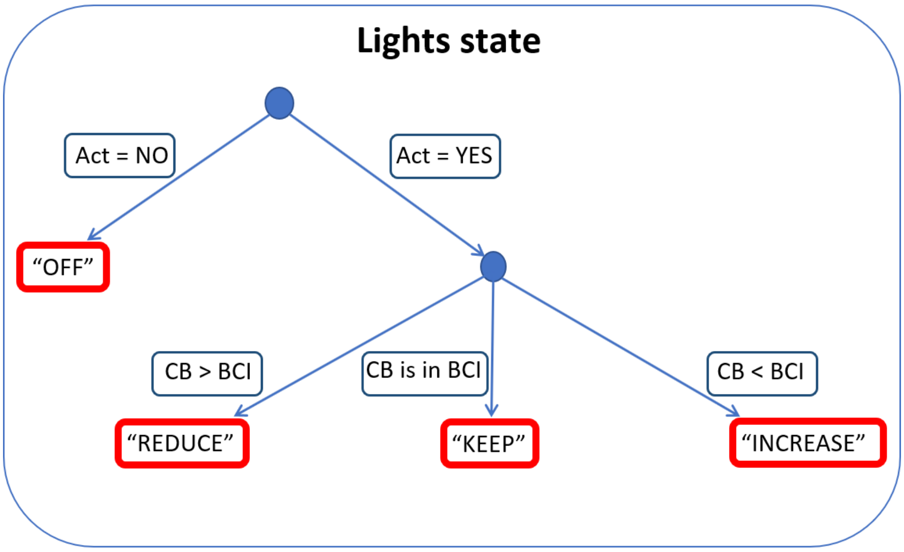

Figure 13 - Brightness ruleset decision tree

At the final step, the recommended state, ideal brightness percentage and minimum value of the brightness comfort interval will be inserted to the Result DB to be available to monetarize and activate in the building.

3.7 Summary

The third section includes a detailed description of the proposed intelligent energy management system for buildings. The main objective of this system is to generate recommendations about the state of the electrical appliances of the building based on calculated predictions. The system starts by receiving the available data from the database of the building and uses a sequence of data analysis to predicts the required values namely as data cleaning, data aggregation, clustering, and forecasting. The system uses clustering techniques to create different contexts of data and base on the associated data to the target context, trains the forecasting methods to predict the target values. The description of all implemented data mining methods are included as well in this chapter. A set of rulesets are created in order to recommend the state of AC system and lights of a building based on the predicted and current values as well as the current price of the market. The next chapter presents the explanations of this system implemented for building N of GECAD facilities.

4 Implementation of the proposed model for GECAD/ISEP building N

This work proposes an intelligent energy management system in the building, which uses Artificial Intelligence techniques to generate recommendations in order to adapt the usage of the electrical appliance of the building with comfort and habits of the residents. In this work, the building N of GECAD facilities located in ISEP, Porto, Portugal, has been chosen as the target building to implement the system. The building N is an office building which includes 12 office rooms, one server room, two WCs, two laboratories, and one kitchen. The map of this building can be found in Figure 14.

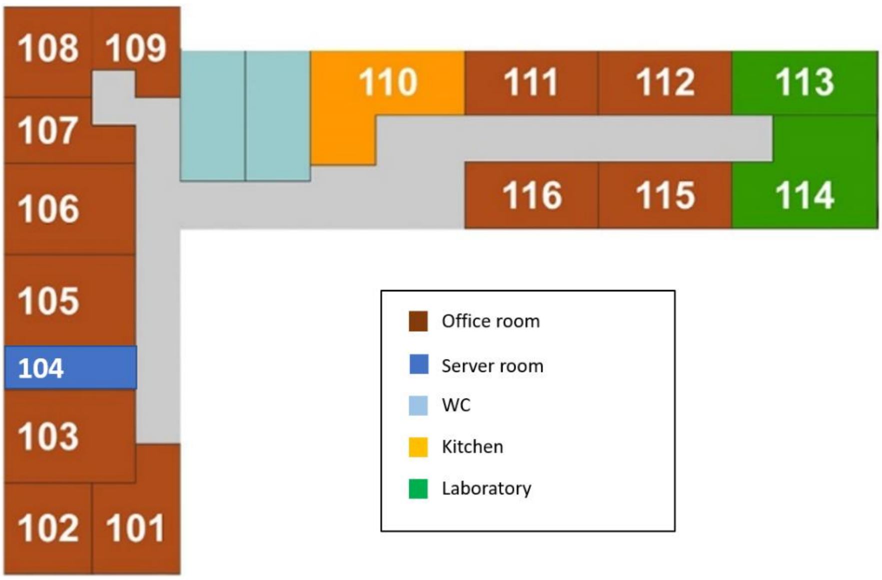

Figure 14 – Map of GECAD/ISEP building N

Every room in this building has an AC system, electrical sockets, and lights. The building also includes a set of solar panels located on the roof of the building, which provides energy generation for the building. The data management of the building and implementation of the proposed model can be found in the following sections.

4.1 Data management in building N

For a system that uses learning methods to generate new information, one of the most important fact to have a good performance is to have the necessary data from the building. The historical consumption and generation information of the building is the primary data that this system requires to obtain acceptable performance. Besides that, further available data such as occupancy, internal and external temperature, solar radiation or any type of information that influences the power consumption of the building is an advantage for the system. The building N includes many energy meters, Programmable Logic Controllers (PLC) and different sensors that record several types of information from the building. This building has a SQL server that stores all of this information in various databases.

The power consumption of the building is measured by five energy meters. Every energy meter is responsible for a specific zone of the building and records the power consumption of the intended zone divided into three values: Air conditioning system, electrical sockets, and lights. This energy meters record these values by a time interval of 10 seconds and store all of these data in a database located in the SQL server of the building. Also, as the building has solar energy generation, the value of this generation is stored in the same database by the same time interval.

In order to store the meteorological data, the data server of the building uses the generated data by the meteorological station of the ISEP (ISEP meteo, 2018). This station provides information such as external temperature, wind direction, wind speed, humidity, solar radiation, etc. The received data from this station are stored in the weather database of the building by a time interval of 5 minutes.

Every room in the building includes a set of sensors that record different types of internal information of the room, such as internal temperature, internal humidity, CO_2, movement, brightness, etc. These data are also recorded and stored in a database by 10 seconds time interval.

As the objective of the proposed system is to generate recommendations about AC system and Lights of the rooms, the room number 103 has been chosen as the target room for experimentation, and the following variables have been used as the selected values from the SQL data server of the building:

- ***Total power consumption of the building (W)***

- ***Total power generation of the building (W)***
- ***The external temperature of the building (W)***
- ***Solar radiation of the related place (W/m^2)***
- ***External humidity of the building (%)***
- ***Total power consumption of room 103 (W)***
- ***Movement of room 103 (Binary)***
- ***The brightness of room 103 (%)***
- ***The internal temperature of room 103 (ºC)***

4.2 Process management

As has been explained the previous chapter, this system has a unit implementation, and all of the implemented parts of the system are located in the same server, except the forecasting, classification, and clustering web services that can be located in a different server. The different phases of the system should be executed one followed by another, but it cannot be a single execution process. Different phases of the system must happen at different time steps, and it requires different processes to complete the whole sequence of the system.

The first step to create these processes is to decide about how often the users want a new recommendation from the system; in other words, the periodicity of the recommendation generation. In this implementation, a 15 minutes time interval has been considered for this periodicity. It means that at every 15 minutes, the system must present a new recommendation. This way, the given processes in Table 5 should be created to complete the sequence of the system.

As the periodicity of the recommendations in this implementation has been considered as 15 minutes, the P_Forecast process should be executed in the following cycles: 11, 26, 41 and 56 minutes of each hour, to generate forecasted data for the following periods (in minutes): 15-30; 30-45; 45-60 and 0-15 minutes. This situation is the same for the other two processes; P_Rules must be executed at each 14,29, 44, and 59 minutes of each hour. And P_Errors should be executed in: 0, 15,30 and 45 minutes and 30 seconds.

Table 5 – Execution processes of the implemented system

Name	Execution time	Description
P_Forecast	4 minutes before the need for a recommendation	Responsible for receiving data, clean and aggregate data, creating context and forecast and store the needed values to run the rules.
P_Rules	1 minute before the need for a recommendation	Responsible for receiving the last forecasted data, selecting the best values, executing the rules and store the results in the database.
P_Errors	30 second after each recommendation	Responsible for receiving the forecasted data for the past duration, collect the real data, calculate the errors and store the errors in the errors database.

4.3 Data cleaning implementation

Data cleaning phase of the system will be executed by P_Forecast process. In this phase, the system receives the data from the SQL database of the building makes sure that the revised data set is completed and has no unreal data. The system searches these data and detects the time intervals, which no data has been recorded and also detects the recorded data that are exaggeratedly too high or too low. In the case of data failure, the system acts the same way for any type of data and calculates the average between the last recorded value before the failure and the first recorded value after the failure and uses this value as the value of the failed interval. For the case of unreal data, the system uses a maximum and a minimum value for every type of data that the system receives. If the received data is more than the maximum or less than the minimum value, the system detects this data as unreal data. The maximum and minimum values for every received data from the building database are presented in Table 6.

All received data will be verified and the values that are detected as unreal data will be deleted from the data set. At this point, the same calculation as the failure data will be done to fill the eliminated rows. After the cleaning phase, these data will be sent to the aggregation phase.

Table 6 – Maximum and Minimum value for data cleaning

Data	Maximum	Minimum
Total consumption(W)	20000	0
Total generation(W)	20000	0
External temperature(ºC)	45	-15
Solar radiation(W/m^2)	1500	0
External humidity (%)	100	0
Power consumption of 103(W)	10000	0
Temperature of 103(ºC)	45	-15
Brightness of 103(%)	100	0

4.4 Data Aggregation implementation

The data aggregation starts with the selection of the period of the recommendation. As in this implementation has been decided to have a new recommendation at every 15 minutes then the data should be aggregated into time intervals of 15 minutes. First, the system needs to find the most recent time interval in the data set, so the system selects the most recent data row and considers the last time step before this row as the beginning of the most recent time interval. For example, in this implementation, the period is 15 minutes the time steps will be at 00, 15, 30 and 45 minutes of each hour. If the most recent data row is at 12:20 so the last time step is 12:15 and the most recent time interval for this data set will be from 12:15 to 12:30. After this, the system goes back 15 to 15 minutes and calculate the aggregated value for every received variable until the oldest received value. The calculation of a single value for 15 minutes is related to the type of the variable. In this implementation for all of the variables except movement data, the average of the recorded values during the target 15 minutes is considered as the value of the time interval. The movement data is a binary value recorded at every 10 seconds, where 0 means no recorded movement and 1 means recorded movement. If during the time interval more than four rows are recorded by value 1, the movement for this time interval will be 1. Otherwise, it will be 0.

4.5 Clustering Implementation

The objective of clustering analysis in this work is to create four contexts of the input data and use the data from the target context as the training data of the forecasting methods in the next phase. For every forecasting target, the clustering component creates four contexts based on the term of the prediction. Two clustering analyses will be done to create these contexts. The first clustering analyze divides the data into two groups as Active days/Non-Active days, and the second analyze divides each one of these groups into two new groups which are Active times/Non-Active times. In the second clustering, the time is related to the term of the forecasting target. In this case, as the forecasting target is to predict the value of a variable for the next 15 minutes, then the data will be aggregated into intervals of 15 minutes. The clustering method receives these data, and the clustering results will correspond for the intervals of 15 minutes. By these analyses, every row of the collected aggregated data will be placed into one of the presented contexts in Figure 15.

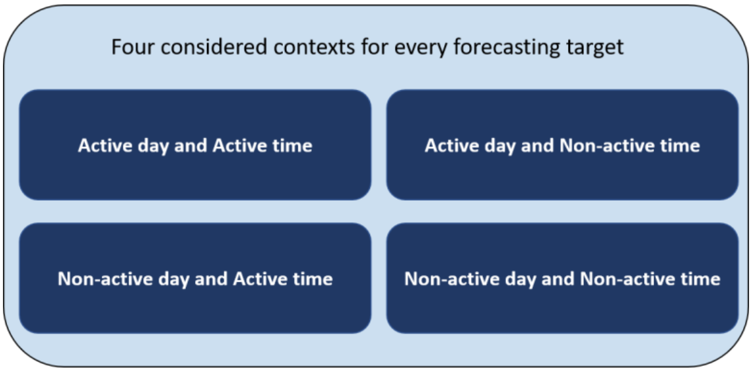

Figure 15 - Considered contexts of the system

In the next step, the data that belong to the same context as the context of the forecasting target, will be selected and sent to the forecasting component.

Related to the objective of the forecasting process, the concept of these considered contexts can be different. Due to the designed rulesets, five different values must be forecasted to be able to execute the rules:

- ***Total power consumption of the building during the next period***
- ***Total power generation of the building during the next period***
- ***Activity in room 103 during the next period***
- ***The ideal temperature in room 103 during the next period***
- ***Ideal brightness in room 103 during the next period***

Based on these five required forecasting values, three different groups of contexts are needed. The first group is the contexts which consider the total consumption of the building as the target variable. In this group of contexts, Active and Non-active days are estimated as official days and weekends or holidays. Also, Active and Non-active periods should represent the working periods and other times of the day. The next contexts group will be created to forecast the generation value of the building. In these contexts, solar radiation is considered as the main input data, and the concepts of the created contexts are pointed at sunny or cloudy days and times. The last group of contexts is created for room 103. These contexts receive the related data to this room and detect Active and Non-active days and times in this room. The next sections present the used data structures to create these contexts. To run the clustering methods, the selected data will be formatted into a Jason structure and sent to the clustering web service.

4.5.1 Context creation for total power consumption

To create the considered contexts, the system needs to run two clustering analyses. One to find the Active and Non-active days and one to detect the Active and Non-active periods. As the chosen periodicity in this implementation is 15 minutes, the system receives the total power consumption, external temperature and external humidity data aggregated in intervals of 15 minutes from past ten days and uses this data to create the contexts.

For the daily clustering, the average values of the received data for every day will be calculated and this data, as well as the number of the day of the week, will be used as the input of the clustering method. Table 7 presents an example of daily clustering data structure.

As can be seen in Table 7, the consumption value is used twice in each row. As the consumption value is the target value of the forecasting, using this value twice in the clustering data set helps the K-Means method to associate a larger weight to this value and achieve more exact results. To find the daily contexts, these data should be sent to the clustering method with K=2. This way, the K-Means method divides this data set into two groups. The group which has a higher power consumption value represents the working days, and the other group presents the weekends or holidays. At this point, the system needs to select one of these groups to use in the next clustering process. If the target forecasting value is on a working day (Monday to Friday), the groups with the higher consumption average will be sent to the next phase, and if the target day is weekend (Saturday and Sunday), the other group will be selected.

Table 7 - Daily clustering input data table to forecast power consumption

Date	**Consumption**	**Consumption**	**Ext.Temperature**	**Ext.Humidity**	**Day of the week**
10-5-2019	3545.20	3545.20	15.85	89.17	5
9-5-2019	3914.55	3914.55	15.04	91.88	4
8-5-2019	3509.65	3509.65	15.20	89.02	3
7-5-2019	3589.61	3589.61	15.71	89.07	2
6-5-2019	3452.07	3452.07	15.16	75.44	1
5-5-2019	2911.65	2911.65	16.99	69.91	7
4-5-2019	2979.10	2979.10	19.02	55.17	6
3-5-2019	3567.97	3567.97	20.00	54.89	5
2-5-2019	3569.15	3569.15	19.66	54.90	4
1-5-2019	2829.94	2829.94	16.36	78.37	3

To create Active or Non-active period contexts, the system also uses consumption, external temperature, humidity and day of the week values as well as number of the rooms with activity, hour and minutes from the selected days aggregated in time intervals of 15 minutes. Table 8 presents a summarized example of selected data to send to the clustering method in this case.

Table 8 – Periodical clustering input data table to forecast power consumption

Date Hour	**Consumption**		**Ext. Temperature**	**Ext.Humidity**	**Active rooms**	**Day of the week**	**Hour**	**Minutes**
2019-05-10 15:45	4876.4	4876.4	17.7	79	4	5	15	45
2019-05-10 15:30	4465.5	4465.5	17.2	78.3	5	5	15	30
…	…	…	…	…	…	…	…	…
2019-05-02 00:00	2048.5	2048.5	15.35	64.5	0	4	00	00

The active rooms value in the table above presents the number of rooms in the building that have recorded activity during the period. Currently, in the building N, nine rooms have movement sensors so the active rooms variable always will be a number between 0 to 9. As can been seen, the last value of the table is from 2/5/2019 while in the daily clustering data 1/5/2019 was also included. The target forecasting value for this data set is the consumption of 16:00 of 10/5/2019, which is a working day (Friday), so only the working days should be selected for the second clustering process. The first of May of 2019 is Wednesday, but this day is a holiday in Portugal, so the system detects this day as a Non-active day and the data from this day are not included in the second clustering data set.

The second data set will also be sent to the clustering method with K=2. So, the results of the K-Means method divide these data into two groups. To find out which group should be sent to the forecasting methods, the system selects the chosen cluster for the same period as the target but from past selected days. If the same periods in the past selected days have been detected as Active times, the data from Active period context will be sent to the forecasting process. Otherwise, the data from the Non-active context will be sent. As an example, for the presented database the target time period is 16:00 to 16:15 of 10/5/2019, the system selects the associated clusters to this time period from the past selected days and if these rows are more associated to cluster “A” than the data from cluster “A” will be sent to the forecasting process. As 16:00 is a working hour, it is expected that the cluster “A” presents the Active time's context.

The selected data at the end of this phase will be sent to the power consumption forecasting process.

4.5.2 Context creation for total power generation

As the building N uses a set of solar panels to generate energy, then power generation in this building is directly related to the solar radiation of the place. This way, the daily clustering in this process is used to divide the past ten days into days with the high and low generation or from another perspective, sunny and cloudy days. The average values of generation and solar radiation during every day is used as the input of the clustering method to create these two groups. Table 9 present an example of an input table for this daily clustering.

Table 9 - Daily clustering input data table to forecast power generation

Date	**Generation**	**Generation**	**Solar Radiation**
10-5-2019	948.63	948.63	179.91
9-5-2019	237.08	237.08	39.76
…	…	…	…
1-5-2019	1332.85	1332.85	284.66

The K-Means method divides these days into two groups and the days that are in the same cluster as the target day will be used for the second clustering analysis. For the second clustering process, the system uses the same variables as the daily clustering plus hour and minutes of every period. Table 10 presents an example of the used data table as the input of periodical clustering to for power generation of the building.

Table 10- Periodical clustering input data table to forecast power generation

Date Hour	**Generation**	**Generation**	**Solar Radiation**	**Hour**	**Minute**
2019-05-10 15:45	2944.32	2944.32	719.00	15	45
2019-05-10 15:30	1782.32	1782.32	377.33	15	30
…	…	…	…	…	…
2019-05-01 16:00	1959.73	1959.73	703.33	16	00

The presented data in Table 10 will be sent to the K-Means method with K=2. To find that the data from which cluster should be used for forecasting, the system uses the same strategy as consumption data clustering. It selects the associated cluster to the same period but from the selected days and based on this result chooses the context. For the presented values, 16:00 is a sunny time of the day and the selected cluster for these hours shows the Active/ high generation context of the data. The data from the selected context will be sent to the power generation forecasting process.

4.5.3 Context creation for room 103

As the room 103 is a portion of the building N, the results of the daily clustering analysis for power consumption of the building N will also be used for this room. The objective of daily clustering for this room is to find the working and Non-working days, which is the same objective as the daily clustering of total consumption data. This way, the clustering process for room 103 receives the selected days form total consumption daily clustering and uses the recorded data by the sensors of this room to create the input table for periodical clustering. To create this data set the system uses movement, internal temperature, and brightness of the room, day of the week, hour, and minute, aggregated into time intervals of 15 minutes from the selected days. A brief example of this data set can be found in Table 11.

Table 11 - Periodical clustering input data table for room 103

Date Hour	Movement		Int.Temperature	Brightness	Day of the week	Hour	Minutes
2019-05-10 15:45	100.0	100.0	26.19	63.84	5	15	45
2019-05-10 15:30	100.0	100.0	25.44	62.80	5	15	30
…	…	…	…	…	…	…	…
2019-05-02 00:00	0	0	24.9	4.0	4	00	00

The movement value in the presented table is a 0 or 100 value that indicates if there has been any recorded movement during every period. As the movement value presents the activity of the room, it has been used twice in the data set to be more effective on the results. As same as other periodical clustering processes, to find the target context, the associated cluster to the same period from the past days before the target period is considered as the target context and the data from this cluster will be sent to the forecasting process.

4.6 Forecasting Implementation

As has been mentioned in the previous sections, the designed rulesets for the building N in this work required five forecasted values to make the target recommendations. These values are total power consumption and total power generation of the building N and the activity, ideal internal temperature and ideal brightness of the 103. All these values should be predicted for the future period which for this implementation is 15 minutes.

This implementation takes advantage of using four forecastings and one classification methods, namely as SVM, HyFIS, WM, and GFS.FS.MOGUL as forecasting methods and SVM for classification. These methods are implemented in the R programming language and are accessible through two developed web services. Every forecasting or classification method needs a set of three data table to predict the final values. These tables are Train Input, Train Output, and Test. The Train Input and Train Output table are used to train the methods and create the prediction model. The method uses these two data set to learn the influence of every input variable on the target value. The Train Input table includes historical data from the available variables. The Train Output table has the values of the target variable from past periods where every value corresponds as the output of one row in the Train Input table. The Test table is the primary input of the forecasting. The data in this table are related data to the target value with the same structure as the Train Input table. Having enough available data is one of the most important facts to have a good forecasting result. However, the way that these data are selected and structured in these three forecasting input tables plays an essential role to obtain better results. The following sections present the structure of the input tables for every needed forecasting value in this work.

4.6.1 Total power consumption forecasting

The process starts with receiving the selected data from the power consumption clustering process. The system receives these data and based on the selected periods creates the forecasting input tables. The input variables in Train Input and Test tables are the day of the week, hour, minutes, power consumption from the past three periods, the number of active rooms and external temperature from the past period. Table 12 and 13 present examples of Train Input and Test tables to forecast the energy consumption of 16:00 of 10/5 2019.

Table 12 – Train Input table to forecast the power consumption at 16:00 of 10/5/2019
**DoW = Day of the Week - ** Cons= Consumption, T= Time period*

Date Hour	DoW*	Hour	Minutes	Cons -1t**	Cons -2t	Cons -3t	Active rooms	Ext.Temp
2019-05-10 15:45	5	15	45	4465	4875	4670	5	15
2019-05-10 15:30	5	15	30	4875	4670	5261	5	15
…	…	…	…	…	…	…	…	…
2019-05-02 10:00	4	10	00	3593	3309	3432	2	19.2

Table 13 - Test table to forecast the power consumption at 16:00 of 10/5/2019

Date Hour	DoW	Hour	Minutes	Cons -1t	Cons -2t	Cons -3t	Active rooms	Ext.Temp
2019-05-10 16:00	5	16	00	4876	4465	4875	5	15

The train output table includes the real consumption value for every row in the Train Input table. Table 14 presents the Train output table for the given case on the tables above.

Table 14 - Train output table to forecast the power consumption at 16:00 of 10/5/2019

Date and ***Hour***	**power consumption**
2019-05-10 15:45	4876
2019-05-10 15:30	4465
…	…
2019-05-02 10:00	4962

These data tables will be sent to the four forecasting methods of this implementation, and every method predicts the value of the power consumption from 16:00 to 16:15 and then these results will be sent to the Automatic Forecast Selection component.

4.6.2 Total power generation forecasting

This forecasting process uses the results of the power generation clustering to create forecasting input tables. The system selects the day of the week, hour, minute, generation of past three periods and solar radiation of the past period and uses this data set to create the Train Input and Test tables. Table 15 presents an example of Train Input table created to forecast the power generation of 16:00 of 10/5/2019.

As same as the power consumption forecasting tables, the Test table for this forecasting process includes the same data set but from the 16:00. And the Train Output table includes the generation values of every row in the Train Input table. The four forecasting methods of the system will be trained by this data and will forecast the value of the power generation for the interval of 16:00 to 16:15.

Table 15 - Train Input table to forecast the power generation at 16:00 of 10/5/2019
**DoW = Day of the Week - ** Gen= Generation, T= Time period*

Date Hour	**DoW***	**Hour**	**Minutes**	**Gen-1t****	**Gen -2t**	**Gen -3t**	**Solar radiation**
2019-05-10 15:45	5	15	45	1782	1730	1833	519
2019-05-10 14:45	5	14	45	3245	3349	2597	720
…	…	…	…	…	…	…	…
2019-05-02 10:00	4	8	00	2721	1494	1057	598

4.6.3 Room activity prediction

The activity prediction is a classification analysis because it is a yes or no (0 or 1) value. A classification process needs the same types of data tables as the forecasting processes. This way, the system receives the result of daily consumption clustering that separates the working days and Non-working days and uses the selected days to create the classification data tables. The system uses five variables to create the Train Input table; namely activity and brightness of the room during the past period, the day of the week, hour and minutes. A brief example of classification Train Input table can be found in Table 16.

Table 16 - Train Input table to predict the activity in room 103 at 16:00 of 10/5/2019

Date Hour	**Activity**	**Brightness**	**Day of the week**	**Hour**	**Minutes**
2019-05-10 15:45	100.0	62.81	5	15	45
2019-05-10 15:30	100.0	62.49	5	15	30
…	…	…	…	…	…
2019-05-02 00:00	00	10.28	4	00	00

The Train Input of this classification process includes the activity values (0 or 100) of every row in the Train Input table, and the Test table has the five selected variables in Train Input table but from the target period. This system uses the SVM method to make this classification process. This method is implemented in the classification web service, mentioned in the previous chapter. The result of this prediction will also be sent to the Automatic Forecast Selection component to be stored and used in the rulesets.

4.6.4 The temperature of the room forecasting

For this forecasting process, the system uses the result of room 103 data clustering to create the forecasting tables. This clustering process selects the periods with the same type of activities in the room as the target period, and base on this selection the system creates the Train Input which includes the activity and temperature of the room as well as day of the week, hour and minutes of the selected periods. An example of this table is presented in Table 17.

Table 17 - Train Input table to forecast the temperature of room 103 at 16:00 of 10/5/2019

Date Hour	Activity	Temperature	Day of the week	Hour	Minutes
2019-05-10 15:45	100.0	26.19	5	15	45
2019-05-10 15:30	100.0	25.73	5	15	30
…	…	…	…	…	…
2019-05-02 10:00	100.0	23.07	4	10	00

The Test and Train Output tables are created in the same way as the other forecasting processes. The Train Output table includes the temperature values for every row in the Train Input table, and Test table has the same variables as the Train Input table but from the 16:00. It is considered that the residents of the building always use the air conditioning system when they feel that it is needed, and they always put their favorite temperature. Based on this consideration, the historical temperature data of the room presents the comfortable temperatures for the residents, and as this data train the forecasting methods, so the forecasted temperature in this process can be considered as the comfortable temperature for the residents during the next period.

4.6.5 The brightness of the room forecasting

To create the input tables for brightness forecasting the same data structure as temperature forecasting is used, and the only difference is that instead of temperature value in tables, the brightness percentage is used. Table 18 presents an example of Train Input table for brightness forecasting.

Table 18 - Train Input table to forecast the brightness of room 103 at 16:00 of 10/5/2019

Date Hour	Activity	Brightness	Day of the week	Hour	Minutes
2019-05-10 15:45	100.0	62.81	5	15	45
2019-05-10 15:30	100.0	62.49	5	15	30
…	…	…	…	…	…
2019-05-02 10:00	100.0	69.4	4	10	00

The Train Output table has the brightness value of the room for every period in Train Input table, and Test table includes the same value from 16:00 hour. As same as the other forecasting and classification results, the forecasted values for brightness also will be sent to the Automatic Forecast Selection component. The forecasted brightness in this process is considered as the comfortable brightness for the residents. Similarly, to the historical data of the temperature, for the brightness the system also considers that the residents always maintain their favorite brightness in the room, so the recorded data during the past days represents the comfort brightness for the users.

4.7 Automatic Forecast Selection implementation

Automatic forecast selection component is responsible for receiving the forecasting data, storing these data, calculating the MAPE error of every forecasted value, storing these errors and select the most trustable forecasting method based on the recent forecasted values. The three principal execution processes of this implementation use this component to complete the mentioned actions.

P_Forecast process after forecasting every value uses this component to store the results into Result DB. This database is a PostgreSQL based database that has been designed to save the forecasting results of the system as well as the generated recommendations. Figure 16 presents the created tables in this database.

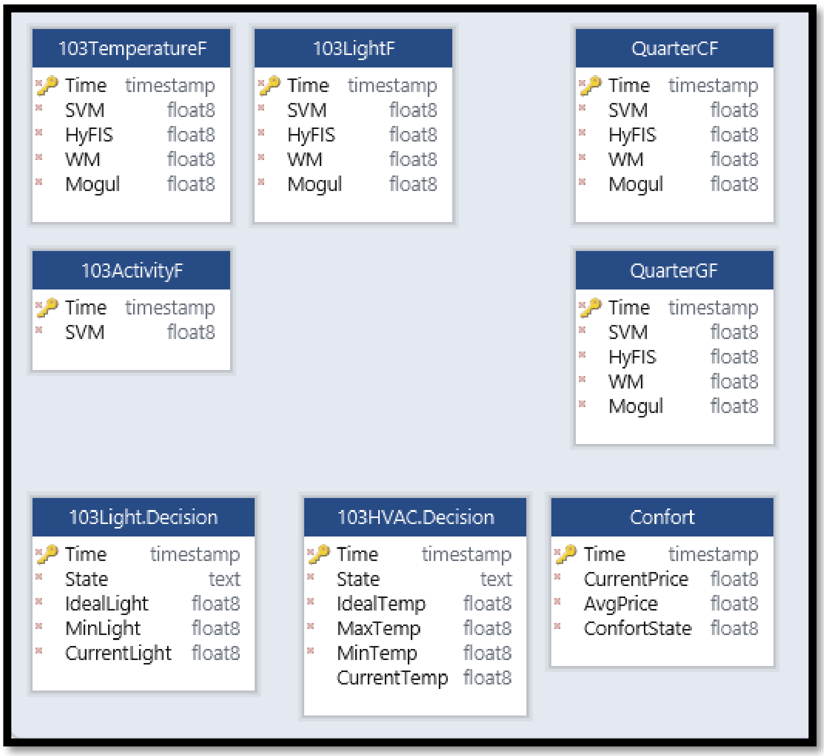

Figure 16 – Result DB structure

The three tables of *103TempratueF, 103LightF,* and *103ActivityF* are created to store the forecasted values for room 103. *QuarterCF* and *QuarterGF* are responsible for storing the consumption and generation forecasting. Finally, the *103Light.Decision, 103HVAC.Decision* and *Confort* Tables are used to store the results of the rulesets, which are the final results of the system.

The P_Errors start at 30 seconds after every iteration and is responsible for calculating the MAPE error of the forecasted values for the past period and store these errors in the Error database. The system selects the real values from the SQL server of the building and receives the forecasted values from Result DB. It calculates the errors based on the MAPE formula and inserts them into the Errors DB. As well as the Results DB, the Errors DB is a PostgreSQL database that is designed to store the errors of every forecasted value by the system. Figure 17 presents the structure of this database.

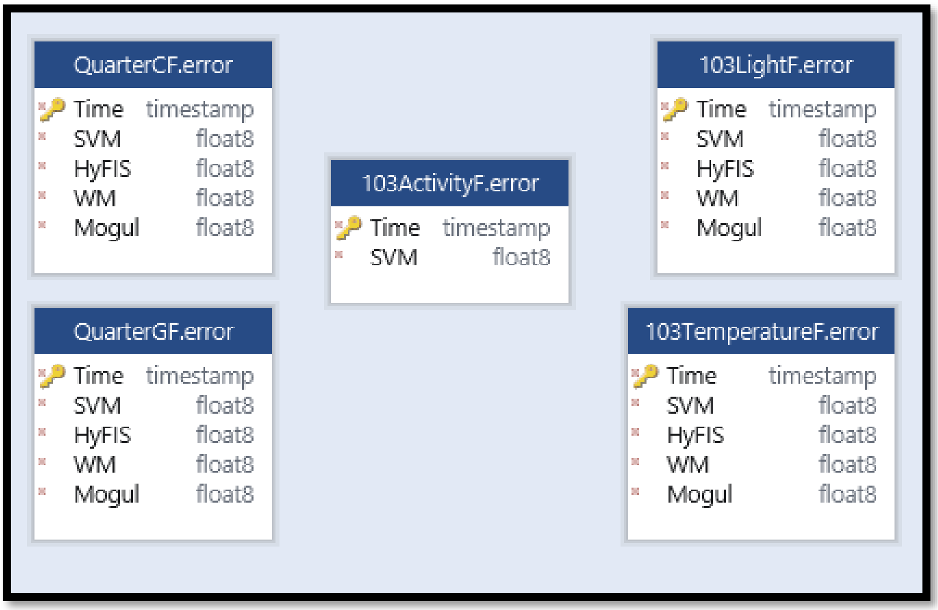

Figure 17 - Result DB structure

As one can see in Figure 17, every forecasting variable has a table in this database where the results of all forecasting methods of the system will be stored.

The P_Rules process starts by finding the best-forecasted values for the next period. To find the best results, the system for every forecasting variable selects the last 20 recorded error from the Error DB and calculates the average error of every forecasting method. The forecasting method with the lower average error is recognized as the most trustable method at the current iteration, and the forecasted value by this method will be sent to the Rule engine component.

4.8 Rules engine implementation

The presented rulesets in section 6 of chapter 3 are implemented in Java programming langue and are executed by P_Rules process. The system needs three types of data to run the rules: forecasted values, current data of the room and price of the electricity market. As has been explained, the system receives the forecasted values form the automatic forecast selection component. The real current data are selected from the main SQL data server of the building, and the electricity market price is received from an API, supported by OMIE web site ("OMIE," 2019). This API provides the hourly market price in Portugal from the current day. The values from the beginning of the day until the current hour are real data and the values from the current hour until the end of the day are forecasted prices by this web site. This way, the system receives the current market price and calculates the daily average of the present day.

The designed rulesets in this work will be executed based on these selected data, and the recommendations about the states of the air conditioning system and lights of room 103 will be inserted to the Results DB.

4.9 Recommendations monitorization

The building N of GECAD facilities has an implemented energy management system called C2C (Click to Control) that makes it possible for the users to monetarize the current state of different variables of the building such as energy consumption and generation, lights intensity, etc. And also, make some actions like opening the doors or change the intensity of the lights in every room. This way has been created a new page in this system where the user is able to see the last results of the system (the forecasted value and recommendations). To make the communications between the C2C system and Result DB of this implementation, a Java-based API has been developed which is connected to the Result DB. When the C2C system sends a request to this API, it selects the most recent rows of every table in Results DB and returns this selected data in a JSON structure. Figures 18 and 19 present this page in C2C system.

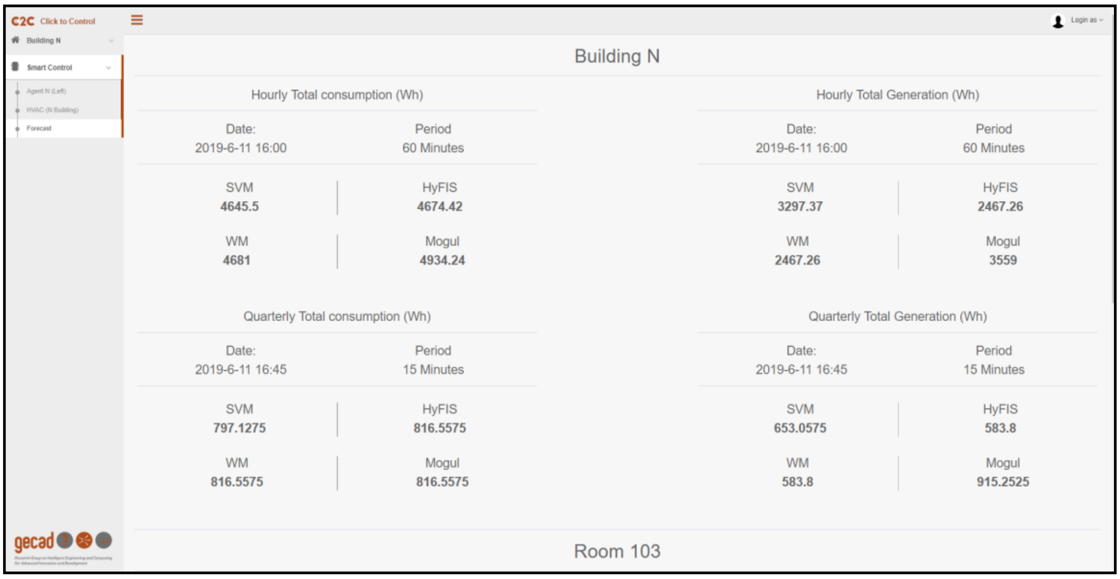

Figure 18 - Result monitorization in C2C system (1)

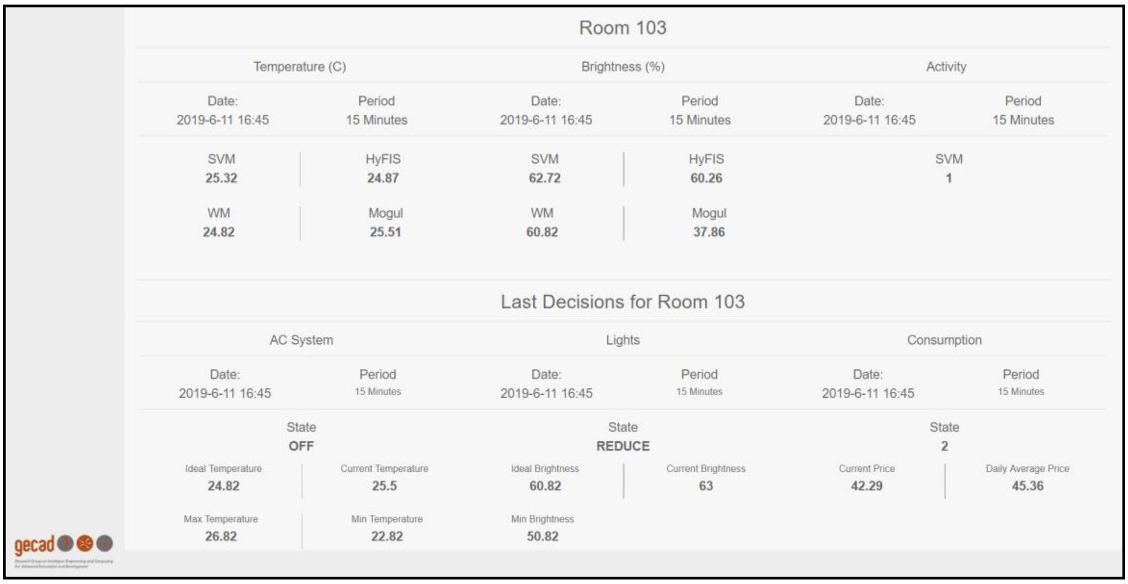

Figure 19 - Result monitorization in C2C system (2)

As can be seen in Figure 18, The forecasted quarterly values for energy consumption and generation of the building are monetarized as well as hourly forecasted values for these variables. The hourly forecasted values are also calculated for this building based on the proposed model to be included in this page. These values are not used in the rulesets; however as many different works and studies have been published and implemented in order to forecast these hourly values, we decided to also include these forecasts in this implementation to compare the results to the previous works and evaluate the performance of the forecasting model of this system.

4.10 Summary

This chapter presented the details of the implementation of the proposed system for building N of GECAD facilities. All of the used variables in every phase as well as an example of each data structure have been included in this chapter. For this implementation, according to the type of available variables, the system considers creation of four contexts for each forecasting value which are based on two divisions: Active or Non-active days and Active or Non-active time intervals. As has been decided to have a new recommendation at every 15 minutes, the received data will be aggregated into time intervals of 15 minutes and the forecasted values are also 15 minutes ahead values. The system has been implemented successfully with all of the expected functionalities. A result monitorization page have been added into the energy management system of the building N to present the last forecasted values and generated recommendations as well.

5 Case Studies

In order to evaluate the performance of the system, reliability of the forecasted results and recommended actions, this chapter presents several case studies that are created to evaluate the behavior of the forecasting methods and accuracy of the forecasted results as well as reliability and rationality of the recommended actions. In the following sections the proposed forecasting model and data mining methods are evaluated by analyzing the obtained results and errors. Also, the comparison between the results of this work and the other published studies is included. After the evaluation of the forecasting process, the subsequent section presents different case studies for different situations in the room and building to assess the recommendations made by this system.

5.1 Forecasting evaluation

This section analyzes the performances of the proposed forecasting strategies in the case of every forecasting variable in this implementation. The proposed forecasting strategy in this work includes the context creation (clustering) and prediction of the required values based on the created contexts. In the following sections, the clustering processes are analyzed by evaluating the chosen number of the clusters in each case of clustering. For the next step the results of four forecasting methods used to predict the power consumption and generation of the building, internal temperature and brightness of room 103 are presented as well as the results of activity classification will be presented and discussed.

5.1.1 The best number of clusters

In order to analyze the chosen K (number of clusters) for the K-Means method in every clustering processes, the Within-Cluster Sum of Squares (WCSS) calculation has been selected to verify the clustering input data in this work (Thinsungnoen, Kaoungku, Durongdumronchai, Kerdprasop, & Kerdprasop, 2015). The WCSS is a measure of the variability of the observations within each cluster. In general, a cluster that has a small sum of squares is more compact than a cluster that has a large sum of squares. Clusters that have higher values exhibit greater variability of the observations within the cluster("Interpret all statistics and graphs for Cluster K-Means," 2018).

As much as the WCSS value is lower for a certain number of clusters, the error of clustering will be lower. But on another hand, to have enough values in each context to train the forecasting methods, the smallest possible number of clusters must be selected. Considering these two facts, for a clustering process, the lowest possible k number with the lowest error is the best number of clusters.

Daily and periodical consumption, periodical generation and periodical state of room 103 clustering, are the 5 clustering processes used in this implementation. For all of them, it has been decided to use two clusters. Figures 20 and 21 present the WCSS value for every chosen number of clusters in each case of these clustering processes.

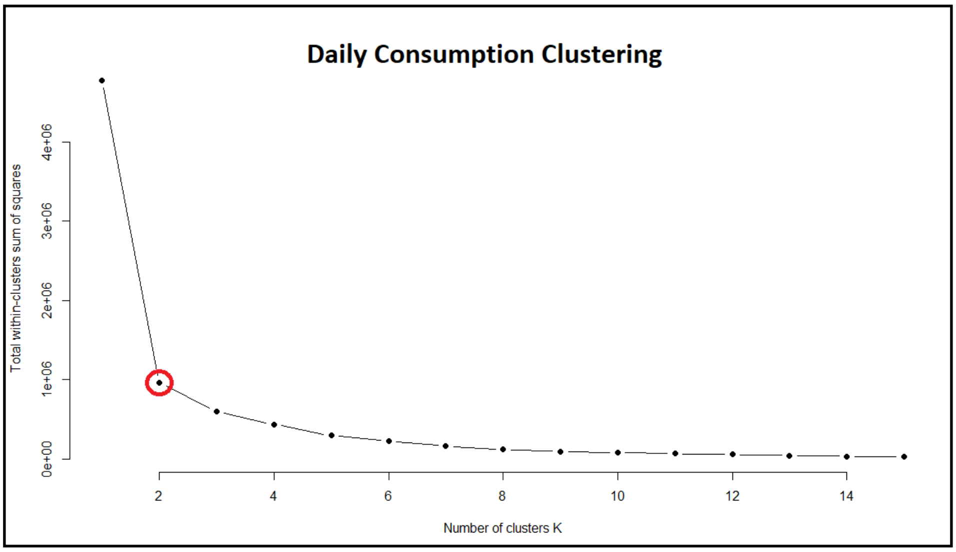

Figure 20 - Within-cluster sum of squares for daily consumption clustering

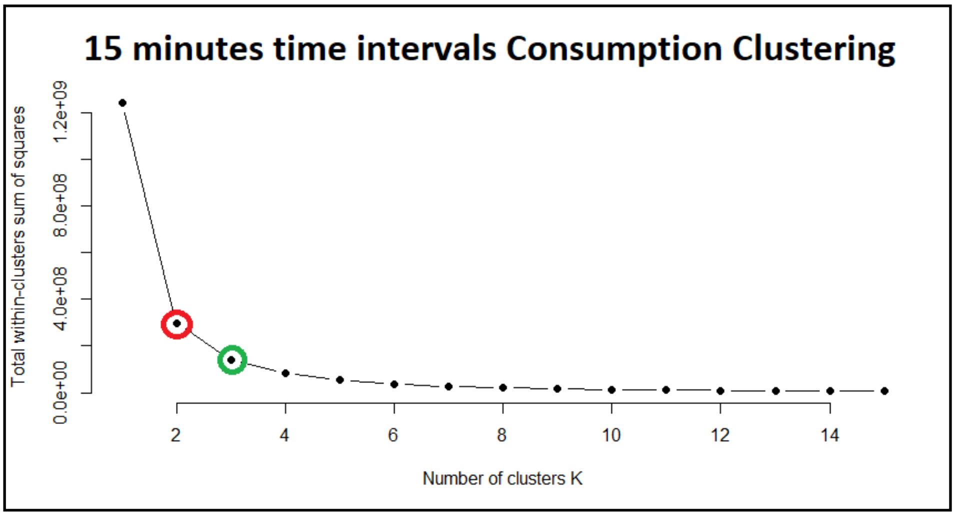

Figure 21 - Within-cluster sum of squares for 15 minutes time intervals consumption clustering

As can be seen in Figure 20, the best number of clusters for a daily consumption clustering data set is 2. As the curve of the presented diagram starts to be more stable after 2 clusters, it can be concluded that 2 is the best number. For the periodical consumption clustering, the situation is different. The best number of clusters can be chosen between 2 or 3. However, if 3 clusters had been chosen, the data associated to the third cluster would be few rows and not enough to train the forecasting methods this way the best number of clusters, in this case, it is also 2.

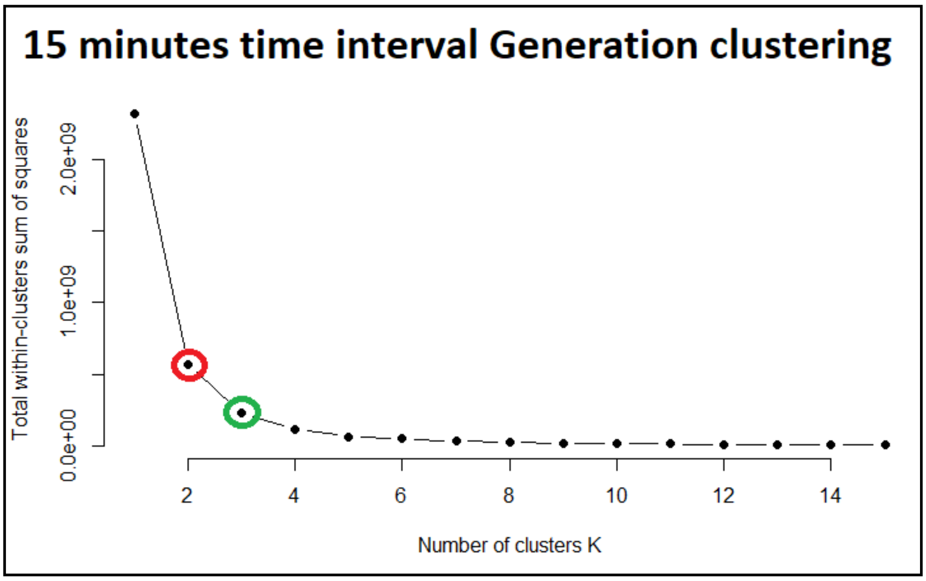

Figure 22 - Within-cluster sum of squares for 15 minutes time intervals generation clustering

Figure 22 presents the WCSS values for different numbers of clusters for periodical generation clustering. For the periodical generation, the same situation as periodical consumption appears. The best number of clusters should be chosen between 2 and 3. But with 3 clusters, one cluster will have a few data and not enough to train the forecasting methods, so the best number of clusters for this data set is also 2.

Figure 23 demonstrates the WCSS value for different numbers of clusters for the periodical state of room 103 clustering in this implementation.

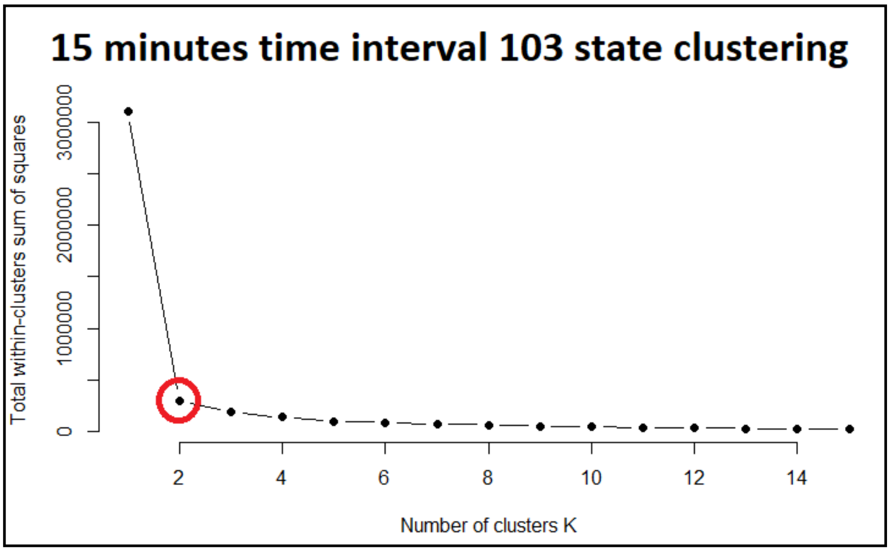

Figure 23 - Within-cluster sum of squares for 15 minutes time intervals for the stare of room 103 clustering

As it is visible in Figure 23, in the case of the used data set of state of room 103 clustering, the best number for clusters is 2.

5.1.2 Power consumption forecasting case studies

This section includes the presentation of the predicted results for power consumption of the building N by four forecasting methods, namely as SVM, HyFIS, WM, GFS.FR.MOGUL. As has been mentioned, the recommendations of this system consider the next 15 minutes as their periodicity. This way, the system forecasts the power consumption value for intervals of 15 minutes. Form another hand; the system also forecasts the hourly power consumption of the building using these methods. The hourly predicted values are not used in the rulesets of the system, but the compression of these values with the published works can approve the improvement that the presented forecasting model brings for the system. Thus, this section presents the results and errors of hourly and periodical power consumption forecasting.

To evaluate the accuracy of the hourly forecasted consumptions, the forecasted values for 24 hours of an official day has been chosen to be presented. Figure 24 demonstrates the hourly forecasted consumption values by four forecasting methods for 24 hours of 11/06/2019 as well as the real consumption value of each hour.

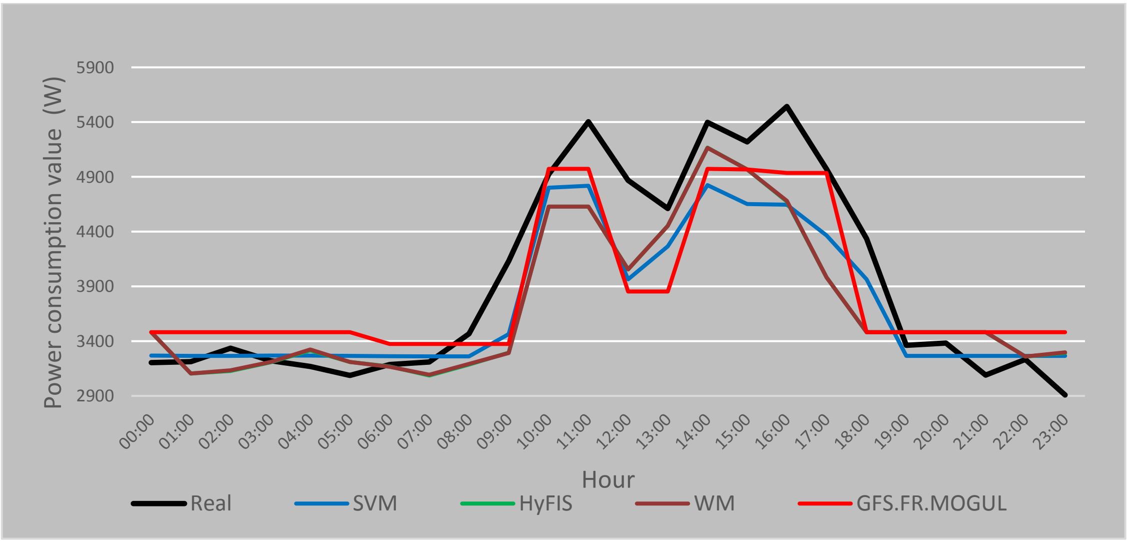

Figure 24 – Forecasted consumption values for 24 hours of 11/6/2019

As can be seen in the presented chart, during the peak hours of consumption the methods are able to estimate the consumption pattern. However, the largest errors also have appeared during these hours due to the instability of the consumption value in this duration. The HyFIS and WM methods, in this case, are presenting a similar performance. The errors of each forecasted value in Figure 24 can be found in table 19 as well as the average error of each method.

Table 19 – MAPE errors of hourly forecasted consumption values for 11/6/2019

	SVM	**HyFIS**	**WM**	**GFS.FR.MOGUL**
0:00	1.965%	8.672%	8.672%	8.672%
1:00	1.626%	3.346%	3.321%	8.413%
2:00	2.139%	6.246%	6.024%	4.397%
3:00	1.473%	0.396%	0.360%	8.148%
4:00	3.071%	4.583%	4.848%	9.851%
5:00	5.779%	3.998%	3.967%	12.843%
6:00	2.448%	0.469%	0.580%	5.937%
7:00	1.593%	3.843%	3.624%	5.125%
8:00	5.957%	8.140%	7.968%	2.687%
9:00	16.168%	20.249%	20.300%	18.319%
10:00	2.488%	5.997%	5.997%	1.021%
11:00	10.829%	14.346%	14.346%	7.951%
12:00	18.571%	16.742%	16.742%	20.850%
13:00	7.493%	3.392%	3.392%	16.407%
14:00	10.620%	4.286%	4.318%	7.873%
15:00	10.886%	4.858%	4.858%	4.858%
16:00	16.184%	15.662%	15.543%	10.974%
17:00	12.180%	19.865%	19.865%	0.720%
18:00	8.540%	19.659%	19.659%	19.659%
19:00	2.914%	3.570%	3.570%	3.570%
20:00	3.521%	2.922%	2.922%	2.922%
21:00	5.701%	12.760%	12.760%	12.760%
22:00	0.952%	0.785%	0.768%	7.694%
23:00	12.245%	13.188%	13.406%	19.741%
AVG	**6.889%**	**8.249%**	**8.242%**	**9.225%**

As can be seen in this table, between the four forecasted methods, the SVM by the average error of 6.89%, presents the best performance. Also, the largest error has been obtained by GFS.FR.MOGUL method at 6:00 by 20.85%. in order have a better perspective of the accuracy of the forecasting methods figure 25 shows the average error of every method for 15 days from 30/5/2019 to 13/6/2019.

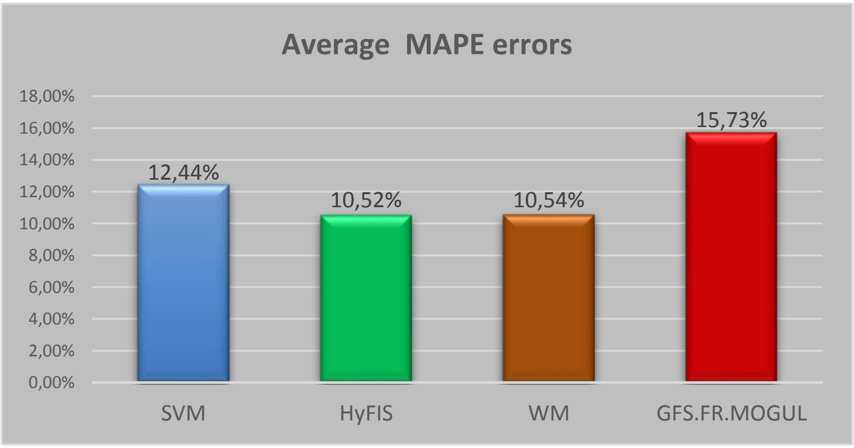

Figure 25 – Average errors of hourly forecasted consumption value for 15 days

As the average errors of these 15 days prove, unlike the presented results for 11/6/2019, the HyFIS and WM methods are predicting more reliable values than the SVM during these days. This is the main reason why the system uses four forecasting methods and before every recommendation selects the best method based on the accuracy of the last forecasted values. For a fixed and particular data set, it is easy to find the best method, but while the system is working automatically along the time, it must be considered that the performances of the methods in different times and situations may be different, so the system should be able to use the best possible results at every iteration.

For the 15 minutes ahead forecasting, the results of the same day have been chosen to be presented, Figure 26 presents 15 minutes ahead forecasted consumption values for 11/6/2019.

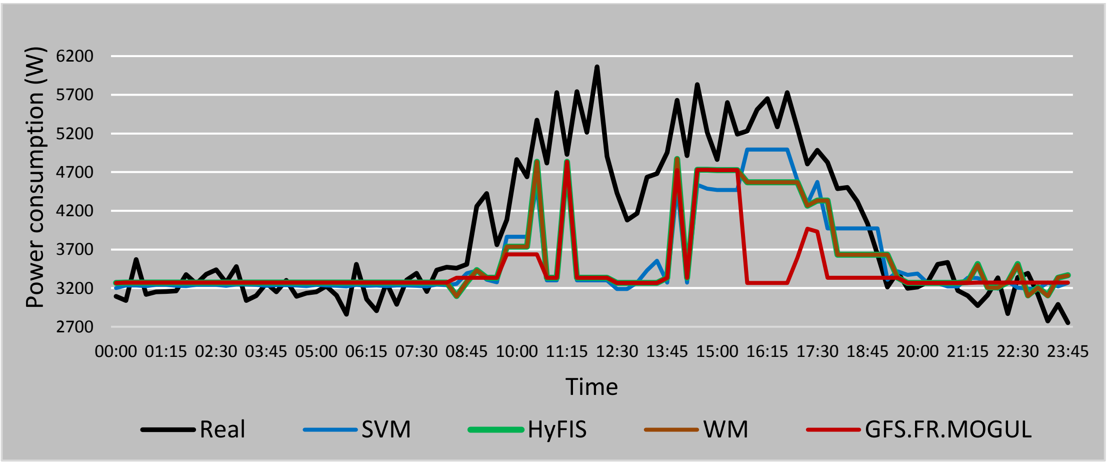

Figure 26 – 15 minutes ahead forecasted consumption values for 11/6/2019

As can be seen in Figure 26, while the forecasting period is shorter, the real values present more variety. For a shorter interval, the possibility of having the same consumption value for the next period increases because the state of the consuming devices can be the same. But form another hand, when the interval is shorter, devices with high consumption peak can easily bring an unexpected consumption value. For example, if a microwave is being used in the building even if it is used just for 3 minutes. Its power consumption is between 1200 W to 1500 W. This consumption might not affect the hourly consumption value but when the considered period is 15 minutes, the average consumption for this period will have an additional unexpected value comparing to the past periods. This is the main reason why power consumption forecasting can be more challenging when it is done for a shorter period. The most challenging fact for the methods in this case of forecasting is to be able to detect these peak periods and model a reliably estimated consumption profile. Figure 27 presents the errors of these forecasted values.

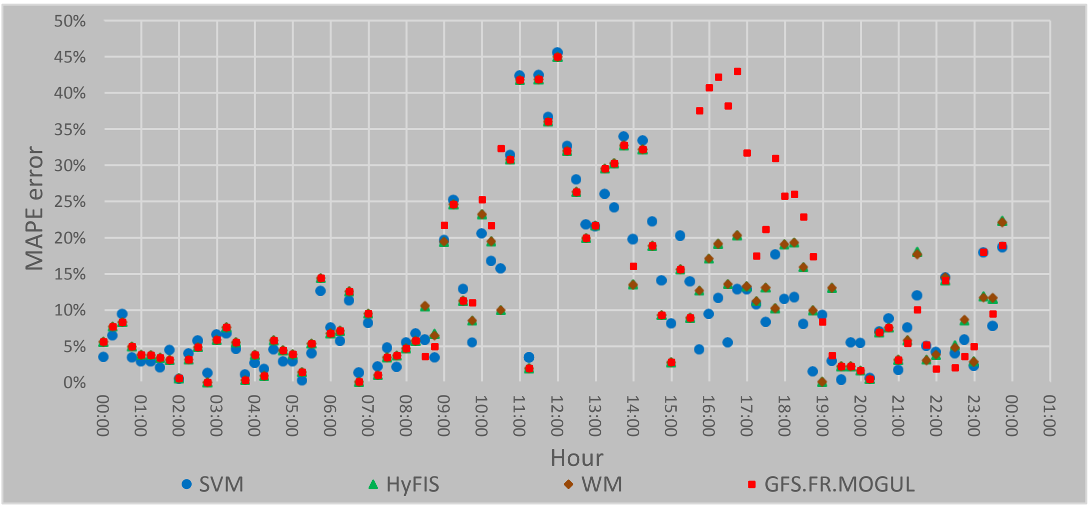

Figure 27 – Errors of 15 minutes ahead consumption forecasting for 11/6/2019

AS can be seen in the given figure, the errors of the methods are higher during the active hours of the day. Between these four methods, the SVM by the average error of 11.1% presents the most trustable results followed by WM and HyFIS by 11.72% and 11.73. Finally, the GFS.FR.MOGUL method has the highest average error during this day by the error of 13.92%. To have a wider perspective of the accuracy of the methods, Figure 28 presents the average errors of consumption forecasted values from 30/5/2019 to 13/6/2019.

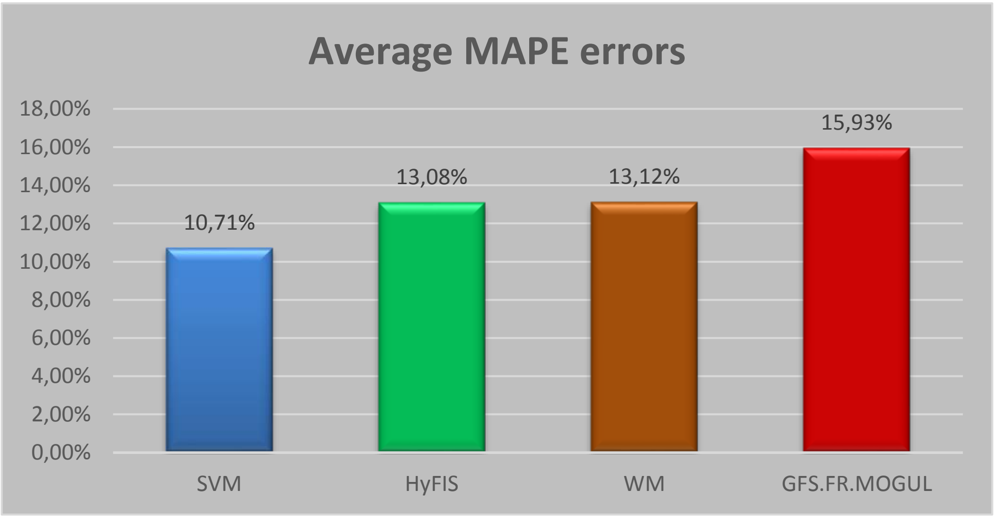

Figure 28 - Average errors of periodically forecasted consumption value for 15 days

The presented average errors also introduce the SVM as the best methods between these four. Many published studies can be found in the area of load consumption forecasting. In the (Jozi, Pinto, Praca, et al., 2017), the HyFIS, WM and GFS.FR.MOGUl methods have been used to predict the hourly load consumption of the building and the average MAPE error of these methods in this work are 12.42% for HyFIS, 18.41% for WM and 9.54% for GFS.FR.MOGUL. Also, the presented work in (Vinagre et al., 2015), Uses the ANN forecasting method for the same purpose and proposes two strategies to forecast the consumption value. The final results in this paper are 15.0% and 13.6%. The comparison of the presented results of this implementation and the mentioned published studies proves that the performance of the proposed forecasting strategies for power consumption forecasting is acceptable and improved. The most important fact that should be considered in this comparison is that the presented studies from the published papers use a specific data set and the configuration of the methods and data preparation are set to obtain the best results for the intended data sat. While, in such an implementation, it is a real-time forecasting system and the system at every 15 minutes receives a different data set and predicts many more values along the time. Due to this nature, it is rational that having a low and acceptable average error from the system for 15 days (or longer periods) is a more challenging task. According to this difference, it can be concluded that the obtained results of the power consumption forecasting in this implementation are improved and more reliable.

5.1.3 Power generation forecasting case study

The Power consumption forecasting in this implementation is based on the historical data from past periods plus the current solar radiation of the place. As the selected period for the building N is 15 minutes, the system needs 15 minutes ahead generation predicted values. As the generation is through the solar panels of the building, the value for the night hours of the day is zero. In this case study the forecasted values from 7:00 to 20:00 hours of an official day by four proposed forecasting methods of this implementation will be discussed. These hours have been chosen due to the sunny hours of the day in Porto and the time of the year. It is clear that the generation hours and generated power in winter days are less than summer days. Figure 29 presents the forecasted generation values for these hours as well as the real generation values by 15 minutes time interval.

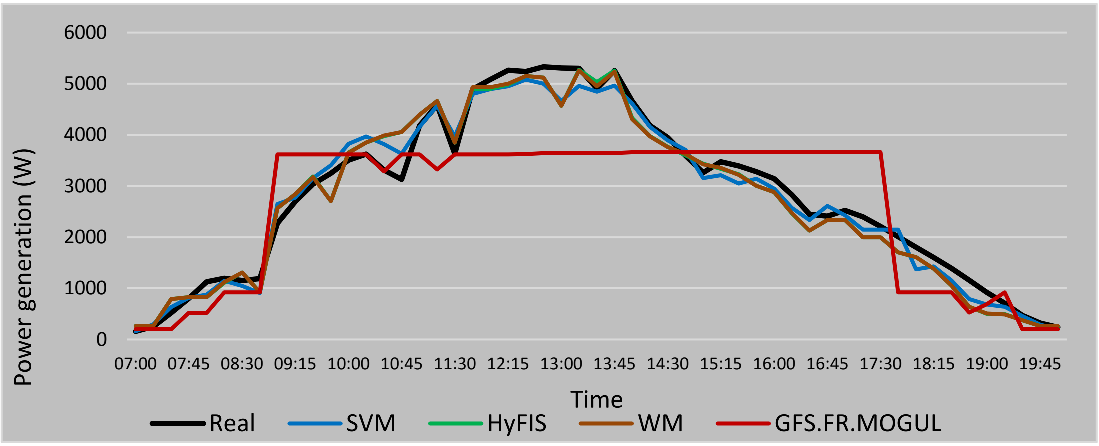

Figure 29 - 15 minutes ahead generation forecasted values for 11/6/2019

The presented results by GFS.FR.MOGUL method, show that this method in the case of power generation forecasting is not able to recognize the pattern of the generation and during most of the hours of the day the results of this method are the same. From another hand, it is visible that the other three methods found the right pattern of the generation and at most of the times the predicted values by these methods are close to the real generated power value. Figure 30 presents the MAPE errors of the presented values. As can be seen in this figure the GFS.FR.MOGUL method has the highest errors. The average error of SVM in this figure is 8.83 %, which is the lowest average error. The HyFIS and WM have an average error of 12.78% and 12.87%, and the GFS.FR.MOGUL by the average error of 28.41% has the highest average error.

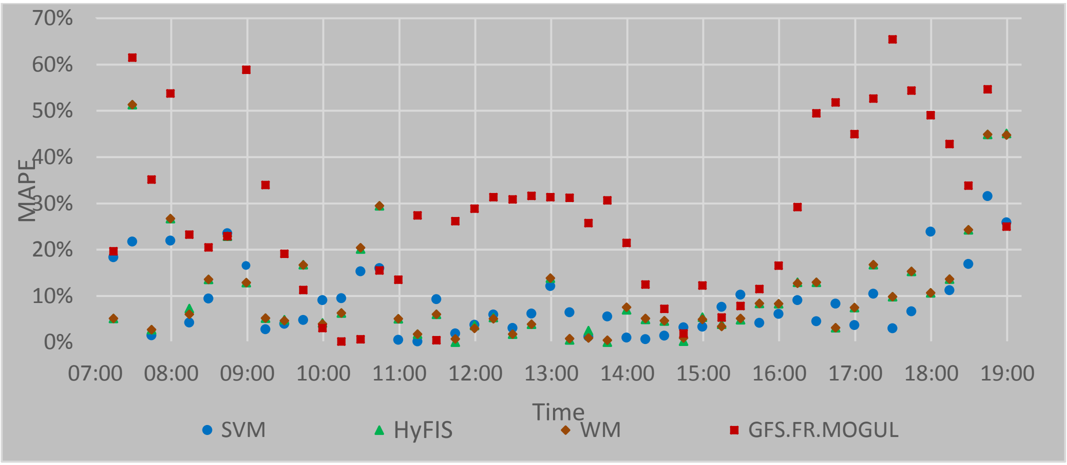

Figure 30 - Errors of 15 minutes ahead generation forecasting for 11/6/2019

In the case of power generation as the same as power consumption, the SVM method presented the best results for the selected periods. However, figure 31 presents the average errors of these four methods for 15 days from 30/5/2019 to 13/6/2019 to be able to evaluate the reaction of the methods while different situations of generation.

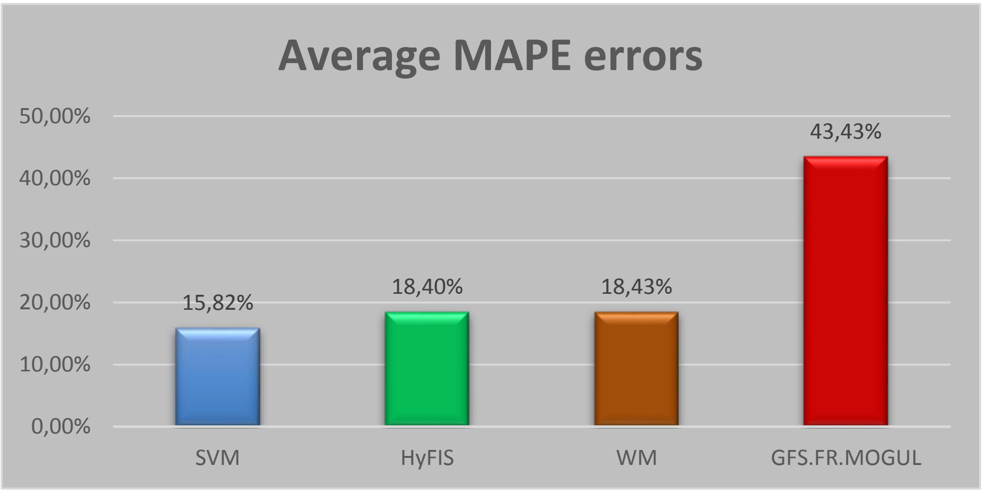

Figure 31 - Average errors of forecasted generation values for 15 days

As it was expected, the average errors for the selected 15 days are higher than the average errors in the presented case study. Yet, 15.82% by SVM, 18.4% by HyFIS and 18.43% by WM are acceptable average errors for this case of forecasting and the system can trust the forecasted generation values based on these averages.

5.1.4 Activity prediction case study

The activity estimation in this work uses a classification technique based on SVM method to detect the activity of the room during the next 15 minutes. As this value is a binary value where 1 means recorded activity (movement) and 0 means no activity, a set of statistical analysis have been chosen to evaluate the performance of this classification process, namely as Confusion or Error Matrix, Accuracy, Recall and Precision(Georgios Drakos, 2018).

The confusion matrix is a table to describe the performance of a classification model on a set of input data which the real values are available(Lewis & Brown, 2001). This table has four values, described as:

- True Positives (TP): the number of times that the predicted value is 1 (active) and the actual value is also 1.
- True Negatives (TN): the number of times that the predicted value is 0 (Non-active) and the real value is also 0.
- False Positive (FP): the number of times that the predicted value is 1 and the real value is 0.
- False Negative (FN): the number of times that the predicted value is 0 and the real value is 1.

To evaluate the results of the SVM method for activity prediction of room 103, the predicted values by this method from the beginning of 30/5/2019 until the end of 13/6/2019 are selected to be analyzed. Table 20 presents the confusion matrix of these results.

Table 20 - The confusion matrix of the predicted values for the activity of room 103

		Predicted		
		Active	Non-Active	Total
Actual	Active	TP= 417	FN= 32	449
	Non-Active	FP= 39	TN= 952	991
	Total	456	984	1440

As can be seen in this table from a total of 449 active periods, the system in 417 times predict the right Incident, and for 991 Non-active times, the system in 952 cases predicted the correct value.

The Accuracy is e percentage that points at the cases that correct result is predicted by the method, comparing to total cases. The Accuracy is calculated by the equation (26).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} * 100 \tag{26}$$

The Accuracy of the presented results in this case study is 95.06%. It means that in 95.06% of times the method is predicting the right event.

The recall is the percentage of the time that the actual value is 1, and the method predicts it correctly. This value is calculated based in (27):

$$\text{Recall} = \frac{TP}{TP + FN} * 100 \tag{27}$$

The Recall value for the presented results is 92.28%. It implicates that in 92.28% of periods that there is activity in the room, the method predicts this activity correctly.

The Precision is the certainty of the method when it predicts activity in the room, which is calculated equation (28).

$$\text{Precision} = \frac{TP}{TP + FP} * 100 \tag{28}$$

The Precision for this case study is 91.44%, which means that when the method predicts activity in the room, in 91.44% of times, it is a correct decision.

According to the presented results and analyses, it has been shown that the SVM method has an acceptable performance to predict the activity of room 103. Based on this, the results of the classification process are trustable enough for the system to use them as a base to create the recommendations.

5.1.5 Ideal temperature forecasting case study

To forecast the ideal temperature of the room, the system uses the historical data of the temperature to train the methods. As it is considered that the residents of the building always use the AC system and these historical data is based on their favorite temperature, then the forecasted value by the methods that were trained by these data can be used as the ideal temperature for the next period. The forecasted temperature values during 24 hours of an official day are presented in Figure 32.

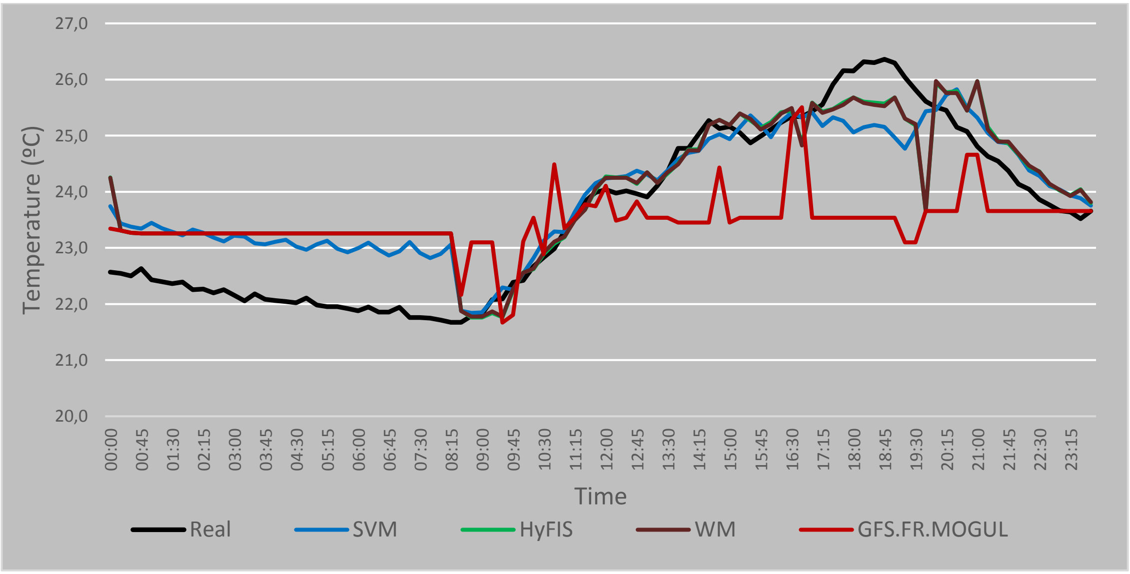

Figure 32 - 15 minutes ahead temperature forecasted values for 11/6/2019

The SVM, HyFIS and WM methods seem to be able to find the right pattern of the temperature during the day, special in the hours of activity. As the values of ideal temperature only will be used is the rulesets when there are activities in the room, the accuracy of the methods at these times is more important. As the room 103 is an office room, the activity hours in this room are between 10:00 to 7:00. It is visible that the methods in the active hours are presenting a better performance than the other hours. The calculated MAPE errors for the presented results are demonstrated in figure 33.

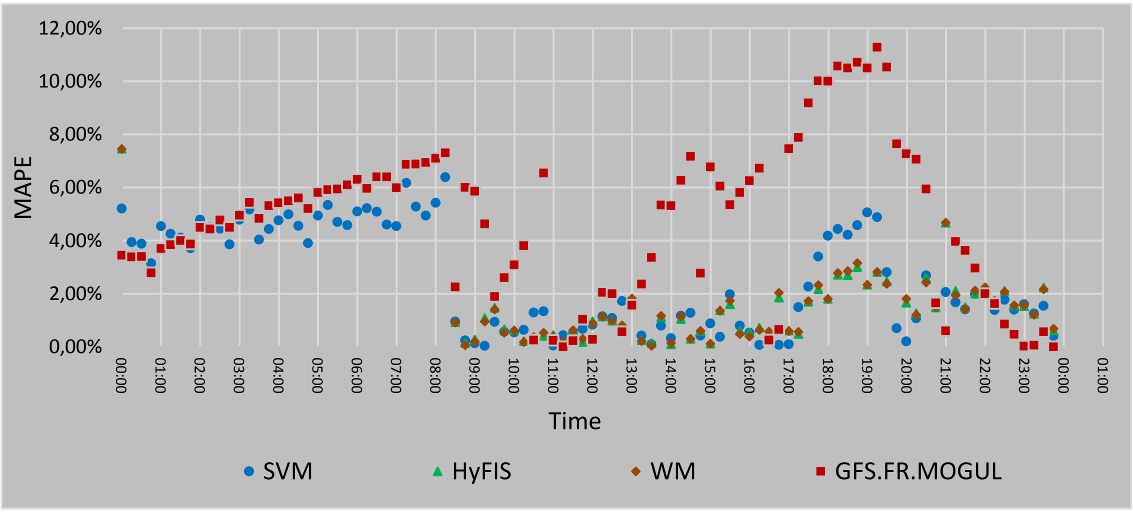

Figure 33 - Errors of 15 minutes ahead temperature forecasting for 11/6/2019

According to the presented errors, the forecasting methods during the hours of activity in the room have much lower errors than during the night. The average of the errors for these 96 periods is 2.56% for SVM, 2.79% for HyFIS and WM, and 4.68% for GFS.FR.MOGUL. This is while the average MAPE errors of the active hours (from 10:00 to 7:00) are 1.38% for SVM, 0.99% for HyFIS, 1.03% for WM and 4.84% for GFS.FR.MOGUL.

Figure 34 presents the average MAPE errors for 15 days from 30/5/2019 to 13/6/2019. This figure includes the average of all the recorded errors as well as the average of the record errors during active hours.

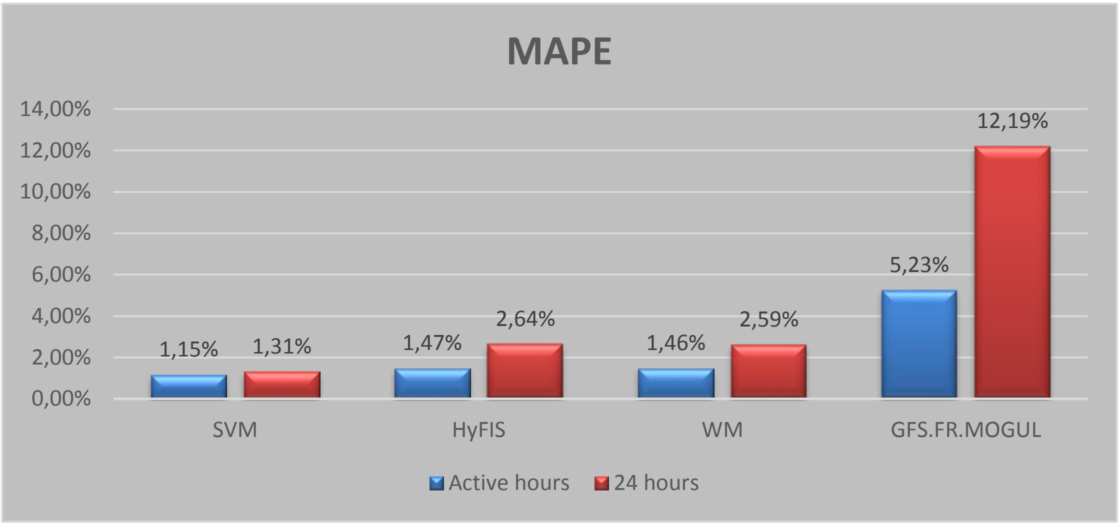

Figure 34 - Average errors of forecasted temperature values for 15 days

The average errors for these 15 days prove that the SVM, HyFIS, and WM are predicting trustable temperature values for room 103, especially during the active hours, which are more important. The GFS.FR.MOGUL in the case of temperature forecasting is not able to predict reliable values comparing to the other three methods. However, as the system always selects the one method with the best performance in the past periods, the final used ideal temperature value from the rulesets is trustable.

5.1.6 Ideal brightness forecasting case study

The ideal brightness forecasting has the same situation as ideal temperature forecasting. The historical brightness data is used to train the methods, and the results of the methods are considered as the ideal temperature. Also, the results for hours with activity in the room are important for the system because based on the designed rulesets, when no activity is predicted in the room the system recommends turning off the lights and the forecasted brightness is not used. Figure 35 monitors the ideal forecasted brightness values for 24 hours of 11/6/2109 by a time interval of 15 minutes.

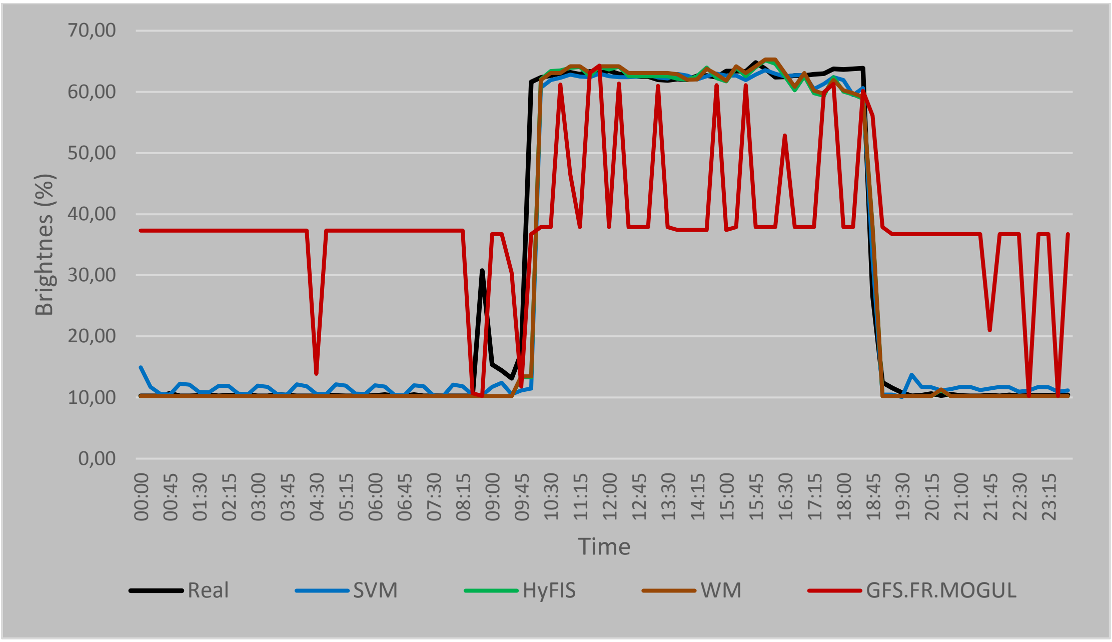

Figure 35 - 15 minutes ahead brightness forecasted values for 11/6/2019

As it is visible the obtained results by GFS.FR.MOGUL are so far from the real values at most of the periods. From another hand, the other three methods can keep up with the variation of the brightness and present reasonable values. The MAPE errors of the presented values can be seen in figure 36. The errors of the GFS.FR.MOGUL are not presented in this figure. The obtained results by this method are so far apart from the real values that it is concluded that this method is not suitable for the forecasting of the ideal brightness.

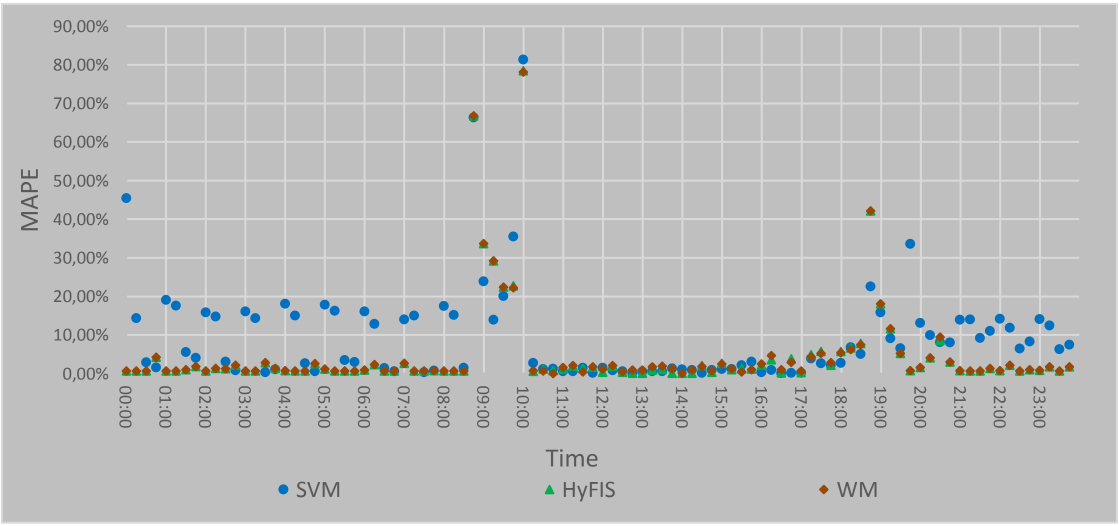

Figure 36 - Errors of 15 minutes ahead brightness forecasting for 11/6/2019

The SVM method in the initial hours of the day has higher errors, but during the active hours, all the presented three methods have forecasted values with low errors. The average error of the methods by these results is 9.31% for SVM, 4.84% for HyFIS and 4.89% for WM. For all 96 periods of the day, the HyFIS presents the best results but calculating the average of the errors during the active hours show that for these hours the best performance is obtained by SVM. The average errors during these hours are 4.56% for SVM, 5.48% for HyFIS and 5.60% for WM. Also, the average errors of GFS.FR.MOGUL in the active hours is 35.27%.

To have a more general perspective of performances of the forecasting methods for brightness prediction, Figure 37 presents the average error of the methods for 15 days from 30/5/2019 to 13/6/2019.

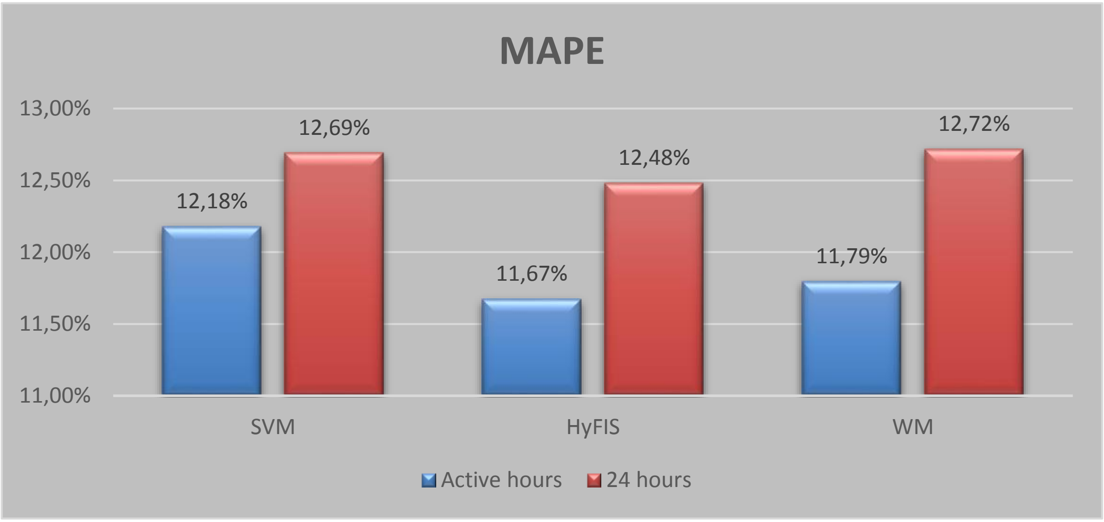

Figure 37 - Average errors of forecasted brightness values for 15 days

As same as the other forecasting cases, the average errors of these 15 days are higher but still acceptable. As it became visible by all the presented forecasted results, it is not possible to select one of the methods as the best one and use it of all periods and variables. Every method in different situations presented a different performance, and it proves that having several forecasting methods and using the results of the best one in each case will increase the accuracy of the final selected predictions at every iteration.

5.2 Rulesets evaluation

In this section, the generated recommendations by the system are discussed. These recommendations are created based on three rules sets: Consumption state, Air conditioning system and Brightness ruleset. The three following sections present the recommended actions by these two rulesets and aim the rationality of them.

5.2.1 Consumption State ruleset case study

The Consumption state ruleset has been created to indicate a number between 1,2 or 3 which presents the state of the energy consumption in the building. As has been mentioned in section 3.6.1, this value is relative to estimated energy consumption and generation during the next time period and the current electricity market price. The consumption state value and consumption cost of the building are directly related, a higher consumption state value means a higher consumption cost for the building. To evaluate the results of this rule set Figure 38 presents the obtained consumption state values by this ruleset as well as the electricity market price during 11/6/2019.

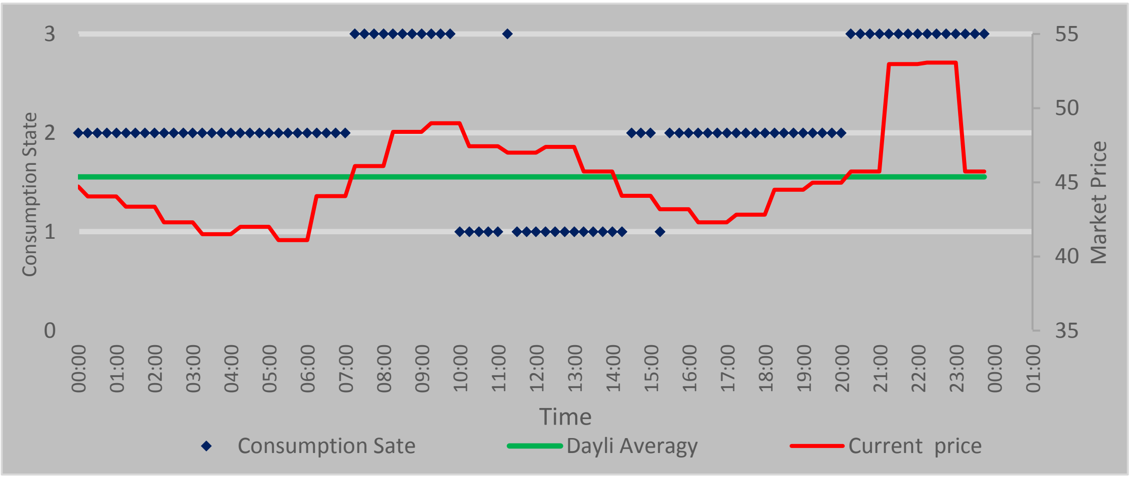

Figure 38 – Consumption state ruleset results for 11/6/2019

During the initial hours of the day as the market price is lower than the daily average, the CS is 2. Form the 7:00 to 14:00 the market price is above the average and the CS should be 3. But in another hand from 10:00 to 14:00 the estimated generation is higher than the estimated consumption and that is the reason why the CS is 1 during these hours. For the rest of the day the estimated consumption is always higher than the estimated generation, so the CS is 2 while the price is less than the average and 3 when its higher. The following sections evaluate the performance of other two rulesets which used the generated CS value by this ruleset. The influence of this value is more visible in these evaluations.

5.2.2 Brightness ruleset case study

The objective of the Brightness ruleset is to recommend the state of the lights for the next period which in this case study is next 15 minutes. The state of the lights should be selected between four options: OFF, REDUCE, KEEP and INCREASE. In order to evaluate the performance of this ruleset, Figure 39 presents the recorded recommendations for lights during 24 hours of 11/6/2019, registered in the database of the system.

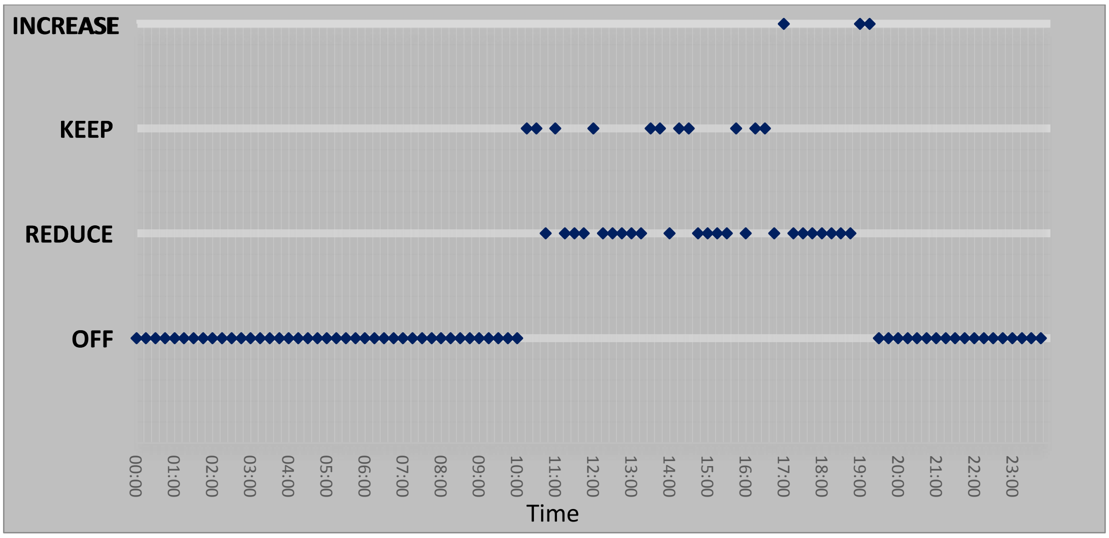

Figure 39 - Brightness ruleset results for 11/6/2019

As can be seen in the presented results, the recommendations during the night until 10:15 are all "OFF" because the system predicts no activity for these hours, which is correct. As it is an office room usually everyone come to work between 10:00 to 10:30 and that is the reason that from 10:15 the system recommends turning on the lights. The commands KEEP, REDUCE and INCREASE are recommended when the lights are on or should be turned on. During the working hours, the system mostly recommends reducing the lights because the ideal forecasted brightness is less than the current brightness. As these recommendations are not being activated in the room, the situation stays the same for the next periods. If after the first REDUCE command the lights had been reduced, the next recommendations would be KEEP.

To have a more exact verification on these recommendations, the presented cases in table 21 have been chosen to be analyzed with more details.

Table 21 – Selected results of the Brightness ruleset

	Time	State	Current brightness (%)	Ideal brightness (%)	Consumption state
Case 1	14:00	REDUCE	62.9	61.98	1
Case 2	10:30	KEEP	61.2	63.06	1
Case 3	17:00	INCREASE	13.1	60.43	2

In the first case, the current brightness is higher than the predicted ideal brightness and, in this case, even though the consumption state is 1 and the generation of the building is more than its consumption, but the system recommends to reduce the intensity of the lights to cut the unnecessary extra consumption. Figure 40 presents the lights state decision tree for this case.

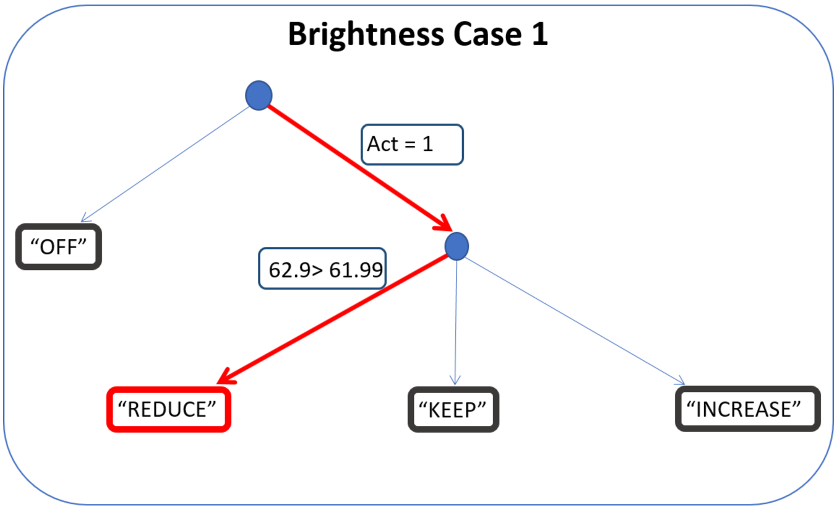

Figure 40 - lights state decision tree for the first brightness case

In the second case, the system recommends Keeping the current intensity of the lights. This situation happens when the current brightness in less than the ideal brightness and the difference between them is less than 5*Consumption state. In this case, the consumption state is 1, and the difference between current and ideal brightness is 1.86%. Based on these conditions, the final recommended sate will be KEEP. The flow of this decision is presented in Figure 41.

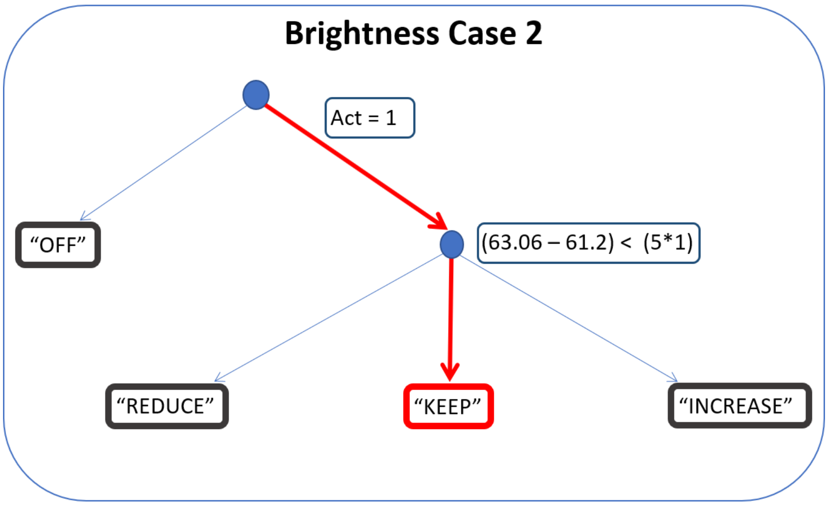

Figure 41 - lights state decision tree for the second brightness case

In the third case, the recommended action is "INCREASE". It means that the current brightness is less than the forecasted ideal value, and the difference between these two values is more than 5*Consumption state. At this period the current brightness is 13.1%, which means that probably for any reason the lights were turns off at this moment. As the system predicts positive activity for this period and the ideal forecasted brightness is 60.43%, then the recommendation of the system will be “INCREASE”. Figure 42 shows the decision tree in this case.

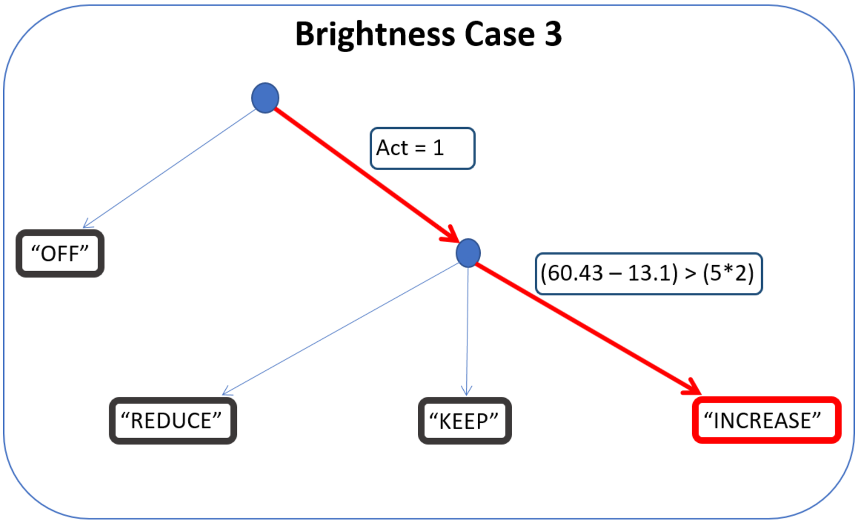

Figure 42 - - lights state decision tree for the third brightness case

5.2.3 Air conditioning system ruleset case study

The goal of this ruleset is to recommend the state of the AC system of the room which can be selected between two options: “OFF” and “ON”. For this objective, the system estimates the ideal temperature for the room and receives the currents temperature form SQL data server of the building N as well. Based on the difference between the current and ideal temperature and by considering the consumption state of the building, the ruleset decides about the final command. The 15 minutes ahead recommended actions by this ruleset during the 11/6/2019 have been presented in Figure 43.

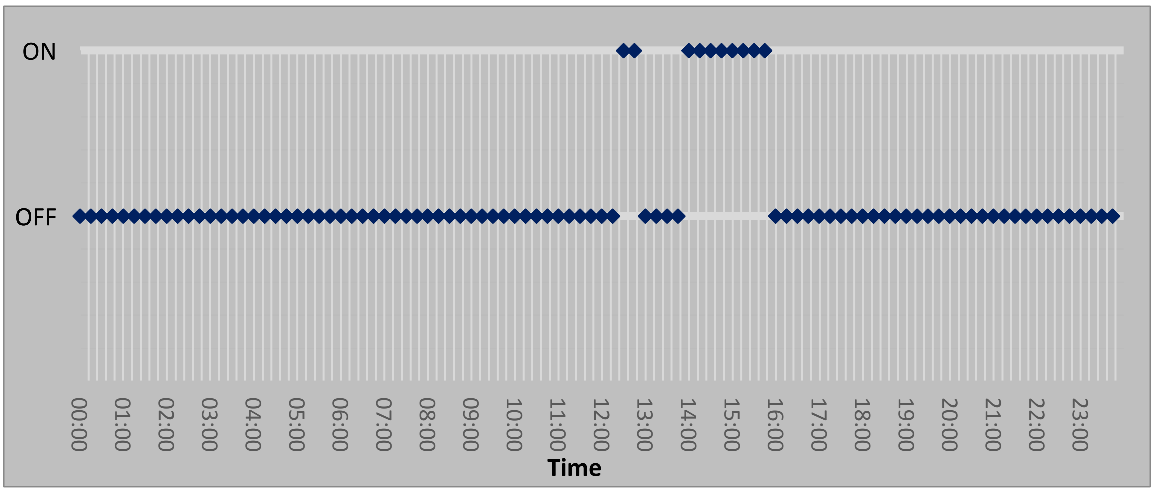

Figure 43 – AC ruleset results for 11/6/2019

During the night, the system prediction for activity in the room is zero activity, and the recommendation for AC state is always OFF when there is no activity in the room. As this is an office room, the activity hours in this room are usually from 10:00 to 19:00. During these hours the system calculates the difference between the ideal and current temperature, and if this difference is more than 1*consumption state value, the system recommendation will be “ON”. Otherwise, it will be “OFF”.

Table 22 includes two chosen case from the presented result. These cases have been selected to be studied more precisely.

Table 22 - Selected results of Air Conditioning system ruleset

	Time	State	Current temperature	Ideal temperature	Consumption state
Case 1	11:15	OFF	24.7	22.64	3
Case 2	12:30	ON	24.2	22.80	1

In the first selected case, the current temperature is 24.7ºC, and the ideal predicted temperature is 22.64 ºC. The difference between these two values is 2.06. As can be seen in Table 22, the consumption state for this period is 3, which means that the consumption of the building is more than its generation and the electricity market price is high. So, the usage of the AC system has a high cost at this moment. In such a situation, the system only recommends using the AC, if the difference between the current and ideal temperature is more than 3. In this case, as this difference is 2.06, the recommendation for AC state is “OFF”. Figure 44 presents the AC state decision tree for this case.

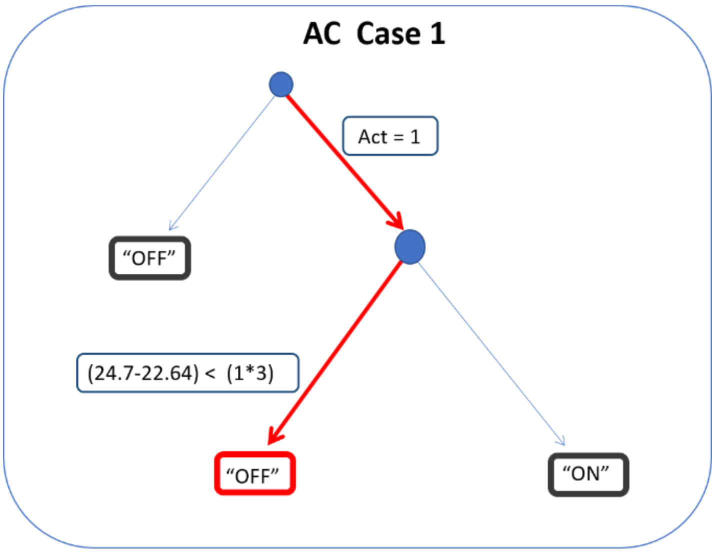

Figure 44 - AC state decision tree for case 1

In the second case, the recommendation of the system is "ON". In this case, the difference between the current and ideal temperature is less than this difference in the first case but the reason that for this period the system recommends to turning on the AC is that the consumption state is 1. It means that at this moment the generation of the building is more than its consumption, so to use the AC system, it is not necessary to buy any energy. The decision three for this case is presented in Figure 45.

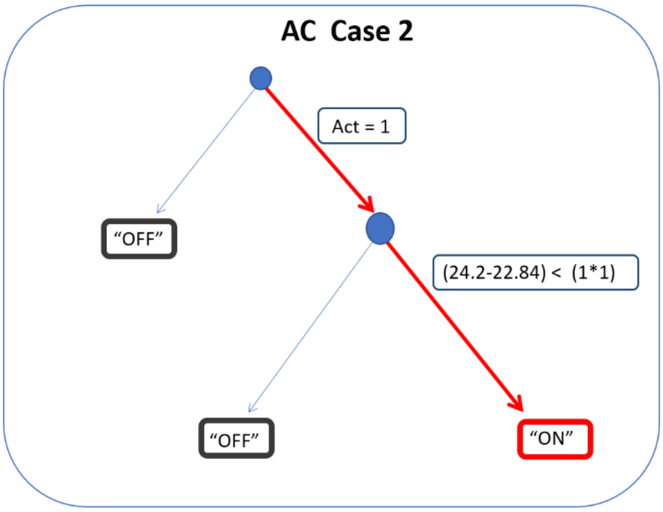

Figure 45 - AC state decision tree for case 2

As one can see in the presented decision tree, the difference between the current and ideal temperature, in this case, is 1.36. And as it is higher than the 1*consumption state, the recommendation of the system for the state of the AC is "ON".

The presented results of both rulesets of the system prove that the designed rulesets can create rational recommendations based on the forecasted values by the system and current state of the consumption. According to the presented case studies and analyses, the system has achieved its goal. The implemented system is able to recognize the comfort and habits of the residents of the building and is able to estimate their wishes by acceptable errors. And based on these recognized wishes creates recommendations that consider the cost of the possible actions and recommends the best and affordable actions for the upcoming period.

5.3 Summary

The fifth chapter presented several case studies to evaluate the performance of the proposed data mining approaches, the accuracy of the forecasted values and rationality of the recommended actions. By analyzing the used data sets for clustering methods, it became visible that for all of the clustering cases the best number of clusters is 2. In the case of the forecasting methods, in most of the times, the SVM has presented the best performance with the most accurate values, but this fact is not guaranteed. In some cases, the other methods had better results. However, as the system at every iteration recognizes the most trustable method during the last 20 iterations and uses the results of the selected method, it is guaranteed that always the rulesets use the most exact forecasted values of the system. To evaluate the rulesets, the generated recommendations for 24 hours of an official day have been analyzed and has been proved that the recommended actions by the system in most cases are rational and are able to meet the comforts of the users in the most affordable way.

6 Conclusions

This chapter presents the main achievements of this work and discusses how they contribute to accomplishing the defined objectives. Additionally, the limitations of this project and the future work perspectives are also presented.

6.1 Achieved objectives

During the past decades, due to the amount of the energy consumption in the buildings, the energy management systems for different types of building have become one of the most addressed objectives of projects related to power and energy systems. However, only a few studies consider taking advantage of AI techniques to control the consumption of the building in order to meet the comfort and demand of the residents of the building. This project proposed an intelligent energy management system for the buildings which uses several AI techniques to adapt the functionality of the electrical appliances of the building with the comforts and habits of the residents while considering the consumption cost of every action. Furthermore, all of the decision processes are context aware. This system has been successfully implemented for the building N of GECAD facilities and has achieved all expected functionalities. The detailed achievements of this project are:

- Implementation of data preparation and pre-processing methodologies, including data cleaning and aggregation.
- Implementation of four forecasting, one classification and one clustering methods in R, as well as the creation of three web services to execute these methods.
- Creation and implementation of three rulesets to recommend the state of the AC system and the lights for the upcoming period.
- Design and implementation of clustering strategies and input data structure to create the required data contexts.
- Design and implementation of forecasting and classification strategies and input data structure to generate the needed predictions for the created rulesets.
- Creation of two databases to store the forecasting results and errors as well as generated recommendations.
- Experimentation and evaluation of the developed clustering, forecasting, and classification processes through various case studies based on the data from building N.

- Experimentation and evaluation of the created recommender rulesets by analyzing the recommended actions for the AC system and Lights of room 103 in building N.

All these objectives have been successfully reached, and the system has been implemented, including all the considered features. The experimentation and evaluation of the results prove that the designed system is able to forecast acceptable values for the necessary variables using a sequence of data mining techniques. The context creation of the system recognizes the considered contexts and the forecasting methods are able to predict reliable values by using the created context. In most of the times, the SVM method presents the best results, however, in some cases HyFIS and WM predicted the most trustable values. According to the analyzed results, it can be concluded that it is not possible to choose one of the forecasting methods as the most trustable one, but the system ensures that at every iteration the most exact results by the four forecasting methods of the system will be used to generate the recommendations. The generated recommendations by the rulesets also have been verified in different situations and it has been concluded that the system recommends rational actions and the recommendations are trustable to be activated by an automatic system.

The contributions of this work have resulted 6 scientific papers, namely the publication of 3 papers in international peer-reviewed journals, and 3 papers in international conference proceedings.

Journals:

- A.Jozi, T.Pinto, I.Praça, Z.Vale, "Decision Support Application for Energy Consumption Forecasting". Appl. Sci. 2019, 9, 699, doi: 10.3390/app9040699
- L.Gomes, C.Ramos, A.Jozi, B.Serra, L.Paiva, Z.Vale, "IoH: A Platform for the Intelligence of Home with a Context Awareness and Ambient Intelligence Approach". Future Internet 2019, 11, 58. doi:10.3390/fi11030058
- A. Jozi, T. Pinto, I. Praça, F. Silva, B. Teixeira, Z. Vale, "Genetic fuzzy rule-based system using MOGUL learning methodology for energy consumption forecasting", ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 2019, (Accepted)

Conference:

- A. Jozi, T. Pinto, I. Praça and Z. Vale, "Day-ahead forecasting approach for energy consumption of an office building using support vector machines," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1620-1625. doi: 10.1109/SSCI.2018.8628734

- A. Jozi, T. Pinto, G. Marreiros, Z. Vale, "Electricity consumption forecasting in office buildings: an artificial intelligence approach", IEEE PowerTech 2019, Milano, Italy, 2019 (Accepted)
- A. Jozi, D. Ramos, L. Gomes, P. Faria, T. Pinto, Z. Vale, "Demonstration of an Energy Consumption Forecasting System for Energy Management in Buildings", 2019 EPIA Conference on Artificial Intelligence, Vila Real, Portugal (Accepted)

The work developed in the scope of this dissertation was supported by several projects under the scope of the GECAD research center. The regarded projects are:

- **COLORS -** Contextual Load Flexibility Remuneration Strategies.
- **DOMINOES -** Smart Distribution Grid: a Market Driven Approach for the Next Generation of Advanced Operation Models and Services.
- **SIMOCE -** Sistema Inteligente e Segura para a monitorização e otimização do consumo energético.

6.2 Limitations and future work

All the objectives of the project have been achieved, but some limitations can still be identified, which can relate directly to the introduction of more features to the application. These limitations are:

- Limited data: the available data is one of the most important facts for the system to have a good performance. The design of all created data structures for the data mining techniques and created rulesets are limited by the available data from the database of the building. Having more measured variables from the building such as energy consumption of every room or the number of the persons in every room can improve the results of the forecasting process as well as the rationality of the recommendations. Having more measured variable requires more sensors and energy meters in the building.
- Accuracy of the data: More exact measured data results as a better performance of the system. The used sensors in the building sometimes have a considerable percentage of mistakes, especially in the case of movement sensors.
- Data failure: It is possible to clean the historical data, but the system is highly dependent on the expected data from the last periods. This way, the

recommendations of the system can easily be distracted if the data collection system of the building stops recording the data.

- The necessity of installing R and Java: As the implementation is based on Java and R programming languages, the system only can work on a server that supports these two languages.

As future work, the system will be improved by the following objectives:

- Including more forecasting, classification and clustering methods to achieve more exact predictions. Such as Artificial Neural Network (ANN) for forecasting and classification or G-Means method for clustering.
- Implementing the system for a larger building with more controllable appliances.
- Creating a system which receives the recommendations and applying them to the appliances. Having the recommendations activated in the building can help the evaluation of the system to be able to verify the automatic changes and satisfaction of the residents.

Also, more scientific papers based on this project, have been designed and will be submitted for publication, as detailed in the introductory chapter.

References

Adomavicius, G., & Tuzhilin, A. (2015). Context-aware recommender systems. *Recommender Systems Handbook, Second Edition*, 191–226. https://doi.org/10.1007/978-1-4899-7637-6_6

Ballenger, A. P., Herath, P., Caceres, M. N., Venayagamoorthy, G. K., & Corchado, J. M. (2017). Influencing behavior of electricity consumers to enhance participation in demand response. *2017 North American Power Symposium, NAPS 2017*. https://doi.org/10.1109/NAPS.2017.8107300

Brereton, R. G., & Lloyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

Cao, X., Dai, X., & Liu, J. (2016). Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy and Buildings*, *128*, 198–213. https://doi.org/10.1016/j.enbuild.2016.06.089

Casillas, J.; Córdon, O.; Herrera, F. (2000). Improving the Wang and Mendel ’ s Fuzzy Rule Learning Method by Inducing Cooperation Among Rules. *8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference*, 1682–1688. https://doi.org/10.1016/j.tiv.2010.06.017

Chen, W. J., Shao, Y. H., Deng, N. Y., & Feng, Z. L. (2014). Laplacian least squares twin support vector machine for semi-supervised classification. *Neurocomputing*, *145*, 465–476. https://doi.org/10.1016/j.neucom.2014.05.007

Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons Among Clustering Techniques for Electricity Customer Classification. *IEEE Transactions on Power Systems*, *21*(2), 933–940. https://doi.org/10.1109/TPWRS.2006.873122

Corchado, J. M., Abrishambaf, O., Vale, Z., Fernandes, F., Pinto, T., & Faia, R. (2017). Case based reasoning with expert system and swarm intelligence to determine energy reduction in buildings energy management. *Energy and Buildings*, *155*, 269–281.

https://doi.org/10.1016/j.enbuild.2017.09.020

Denoeux, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics*, *25*(5), 804–813. https://doi.org/10.1109/21.376493

Dimiter Driankov, Hans Hellendoorn , Michael Reinfrank, L. Ljung, R. Palm , B. Graham, A. O. (1996). An Introduction to Fuzzy Control.

Doukas, H., Patlitzianas, K. D., Iatropoulos, K., & Psarras, J. (2007). Intelligent building energy management system using rule sets. *Building and Environment*, *42*(10), 3562–3569. https://doi.org/10.1016/j.buildenv.2006.10.024

Dounis, A. I., & Caraiscos, C. (2009). Advanced control systems engineering for energy and comfort management in a building environment-A review. *Renewable and Sustainable Energy Reviews*, *13*(6–7), 1246–1261. https://doi.org/10.1016/j.rser.2008.09.015

Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, *35*(3), 1701–1708. https://doi.org/10.1016/j.enpol.2006.05.009

Elkan, G. and, & Hamerly, C. (2003). Learning the K in K-means. In *International Conference on Neural Information Processing Systems,* (pp. 281–288).

Faria, P., Spinola, J., & Vale, Z. (2016). Aggregation and Remuneration of Electricity Consumers and Producers for the Definition of Demand Response Programs. *IEEE Transactions on Industrial Informatics*, *PP*(99), 1–1. https://doi.org/10.1109/TII.2016.2541542

Fernandes, F., Morais, H., Vale, Z., & Ramos, C. (2014). Dynamic load management in a smart home to participate in demand response events. *Energy and Buildings*, *82*, 592–606. https://doi.org/10.1016/j.enbuild.2014.07.067

Georgios Drakos. (2018). How to select the Right Evaluation Metric for Machine Learning Models: Part 3 Classification Metrics. Retrieved June 15, 2019, from https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-3-classification-3eac420ec991

Herrera, F., Lozano, M., & Verdegay, J. L. (1995). A Learning Process for Fuzzy Control Rules using Genetic Algorithms, (DECSAI-95108).

Hosseinzadeh Aghdam, M. (2019). Context-aware recommender systems using hierarchical hidden Markov model. *Physica A: Statistical Mechanics and Its Applications*, *518*, 89–98. https://doi.org/10.1016/j.physa.2018.11.037

Hruschka, E. R., Campello, R. J. G. B., Freitas, A. A., & de Carvalho, A. C. P. L. F. (2009). A Survey of Evolutionary Algorithms for Clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *39*(2), 133–155. https://doi.org/10.1109/TSMCC.2008.2007252

Interpret all statistics and graphs for Cluster K-Means. (2018). Retrieved June 11, 2019, from https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/cluster-k-means/interpret-the-results/all-statistics-and-graphs/

ISEP meteo. (2018). Retrieved July 3, 2019, from http://meteo.isep.ipp.pt

Jamil, M., & Mittal, S. (2017). Building Energy Management System: A Review. *2017 14th IEEE India Council International Conference, INDICON 2017*. https://doi.org/10.1109/INDICON.2017.8488004

Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., & Mishra, S. (2016). Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid. *IEEE Transactions on Industrial Informatics*, *12*(3), 1005–1016. https://doi.org/10.1109/TII.2016.2543145

Jonathan, L. H., Joseph, A. K., Loren, G. T., & John, T. R. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, *22*(1), 5–53.

Jozi, A., Pinto, T., Praça, I., Ramos, S., Vale, Z., Goujon, B., & Petrisor, T. (2017). Energy Consumption Forecasting using Neuro-Fuzzy Inference Systems : Thales TRT building case study. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*.

Jozi, A., Pinto, T., Praca, I., Silva, F., Teixeira, B., & Vale, Z. (2016). Energy consumption forecasting based on Hybrid Neural Fuzzy Inference System. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–5). https://doi.org/10.1109/SSCI.2016.7849859

Jozi, A., Pinto, T., Praca, I., Silva, F., Teixeira, B., & Vale, Z. (2017). Energy consumption forecasting using genetic fuzzy rule-based systems based on MOGUL learning methodology. *ASTES*, *703689*(703689), 1–5. https://doi.org/10.1109/PTC.2017.7981219

Jozi, A., Pinto, T., Praça, I., Silva, F., Teixeira, B., & Vale, Z. (2016). Wang and Mendel's Fuzzy Rule Learning Method for Energy Consumption Forecasting considering the Influence of Environmental Temperature. In *Giis 2016*.

Khosravani, H., Castilla, M., Berenguel, M., Ruano, A., & Ferreira, P. (2016). A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building. *Energies*, *9*(1), 57. https://doi.org/10.3390/en9010057

Kim, J., & Kasabov, N. (1999). HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. *Neural Networks*, *12*(9), 1301–1319.

Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, *39*(4), 261–283. https://doi.org/10.1007/s10462-011-9272-4

Kwac, J., Flora, J., & Rajagopal, R. (2014). Household Energy Consumption Segmentation Using Hourly Data. *IEEE Transactions on Smart Grid*, *5*(1), 420–430. https://doi.org/10.1109/TSG.2013.2278477

Labeeuw, W., Stragier, J., & Deconinck, G. (2015). Potential of Active Demand Reduction With Residential Wet Appliances: A Case Study for Belgium. *IEEE Transactions on Smart Grid*, *6*(1), 315–323. https://doi.org/10.1109/TSG.2014.2357343

Lewis, H. G., & Brown, M. (2001). A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, *22*(16), 3223–3235. https://doi.org/10.1080/01431160152558332

Lezama, F., Rodriguez-Gonzalez, A. Y., & de Cote, E. M. (2016). Load pattern clustering using differential evolution with Pareto Tournament. In *2016 IEEE Congress on Evolutionary Computation (CEC)* (pp. 241–248). IEEE. https://doi.org/10.1109/CEC.2016.7743801

Macedo, M. N. Q., Galo, J. J. M., de Almeida, L. A. L., & de C. Lima, A. C. (2015). Demand side management using artificial neural networks in a smart grid environment. *Renewable and Sustainable Energy Reviews*, *41*, 128–133. https://doi.org/10.1016/j.rser.2014.08.035

Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. Hornik, K. (2014). cluster: Cluster Analysis Basics and Extensions. R package version 1.15.2., 2014.

Mahmood, T., & Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies, 73. https://doi.org/10.1145/1557914.1557930

Manic, M., Wijayasekara, D., Amarasinghe, K., & Rodriguez-Andina, J. J. (2016). Building Energy Management Systems: The Age of Intelligent and Adaptive Buildings. *IEEE Industrial Electronics Magazine*, *10*(1), 25–39. https://doi.org/10.1109/MIE.2015.2513749

Marreiros, G., Santos, R., Ramos, C., & Neves, J. (2010). Context-aware emotion-based model for group decision making. *IEEE Intelligent Systems*, *25*(2), 31–39. https://doi.org/10.1109/MIS.2010.46

Meyer, D., Evgenia Dimitriadou, Kurt Hornik, A. W., Leisch, F., Chang, C.-C., & Lin, C.-C. (2019). Package 'e1071.'

OMIE. (2019). Retrieved May 31, 2019, from http://www.omie.es/inicio

Pedrycz, W., Ekel, P., & Parreiras, R. (2010). Notions and Concepts of Fuzzy Sets: An Introduction. *Fuzzy Multicriteria Decision-Making*, 21–62. https://doi.org/10.1002/9780470974032.ch2

Pinto, T., Sousa, T. M., Praça, I., Vale, Z., & Morais, H. (2016). Support Vector Machines for decision support in electricity markets' strategic bidding. *Neurocomputing*, *172*, 438–445. https://doi.org/10.1016/j.neucom.2015.03.102

Platon, R., Dehkordi, V. R., & Martel, J. (2015). Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings*, *92*, 10–18. https://doi.org/10.1016/j.enbuild.2015.01.047

POPEANGA, J. (2015). Data Mining Smart Energy Time Series. *Database Systems Journal, 6*(1), 14–22.

Ramos, S., Duarte, J. M. M., Soares, J., Vale, Z., & Duarte, F. J. (2012). Typical load profiles in the smart grid context - A clustering methods comparison. In *2012 IEEE Power and Energy Society General Meeting* (pp. 1–8). https://doi.org/10.1109/PESGM.2012.6345565

Ribeiro, C., Pinto, T., Silva, M., Ramos, S., & Vale, Z. (2015). Data Mining Approach for Decision Support in Real Data Based Smart Grid Scenario. In *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 73–77). IEEE. https://doi.org/10.1109/DEXA.2015.33

Riza, L. S., Bergmeir, C., Herrera, F., & Benítez, J. M. (2015). {frbs}: Fuzzy Rule-Based Systems for Classification and Regression in {R}. *R Package Version 3.1-0*.

Robin Burke. (2007). Hybrid web recommender systems. *The Adaptive Web*, 377–408. Retrieved from http://www.ifm.com/ifmar/web/apps-by-industry/cat_060_010_040.html

Rodriguez-Fernandez, J., Pinto, T., Silva, F., Praça, I., Vale, Z., & Corchado, J. M. (2019). Context aware Q-Learning-based model for decision support in the negotiation of energy contracts. *International Journal of Electrical Power and Energy Systems*, *104*, 489–501. https://doi.org/10.1016/j.ijepes.2018.06.050

Rudd, K., Muro, G. Di, & Ferrari, S. (2014). A constrained backpropagation approach for the adaptive solution of partial differential equations. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(3), 571–584. https://doi.org/10.1109/TNNLS.2013.2277601

Santos, G., Femandes, F., Pinto, T., Silva, M., Abrishambaf, O., Morais, H., & Vale, Z. (2017). House management system with real and virtual resources: Energy efficiency in residential microgrid. *2016 Global Information Infrastructure and Networking Symposium, GIIS 2016*. https://doi.org/10.1109/GIIS.2016.7814943

Schaefer, A. M., Udluft, S., & Zimmermann, H.-G. (2007). A Recurrent Control Neural Network for Data Efficient Reinforcement Learning. In *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning* (pp. 151–157). IEEE. https://doi.org/10.1109/ADPRL.2007.368182

Soares, N., Lopes, M. A. R., Monteiro, H., Rodrigues, E., Bastos, J., Pereira, L. D., … Amaral, A. R. (2017). A review on current advances in the energy and environmental performance of buildings towards a more sustainable built environment. *Renewable and Sustainable Energy Reviews*, *77*, 845–860. https://doi.org/10.1016/j.rser.2017.04.027

Spinola, J., Faria, P., & Vale, Z. (2017). Aggregation validation approach for the management of resources in a smart grid. In *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016* (pp. 1–7). IEEE. https://doi.org/10.1109/SSCI.2016.7849861

Talib, A. M., & Elshaiekh, N. E. M. (2014). Multi Agent System-Based on Case Based Reasoning for Cloud Computing System. *Academic Platform Journal of Engineering and Science*, *2*(2), 34–38. https://doi.org/10.5505/apjes.2014.42714

Thinsungnoen, T., Kaoungku, N., Durongdumronchai, P., Kerdprasop, K., & Kerdprasop, N. (2015). The Clustering Validity with Silhouette and Sum of Squared Errors, 44–51.

https://doi.org/10.12792/iciae2015.012

V. Vapnik, A. L. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, *24*, 774–780.

Vale, Z., Morais, H., Ramos, S., Soares, J., & Faria, P. (2011). Using data mining techniques to support DR programs definition in smart grids. In *2011 IEEE Power and Energy Society General Meeting* (pp. 1–8). IEEE. https://doi.org/10.1109/PES.2011.6039081

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. (Springer, Ed.). New york.

Vinagre, E., Gomes, L., & Vale, Z. (2015). Electrical Energy Consumption Forecast Using External Facility Data. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 659–664). IEEE. https://doi.org/10.1109/SSCI.2015.101

Vinagre, E., Pinto, T., Ramos, S., Vale, Z., & Corchado, J. M. (2016). Electrical Energy Consumption Forecast Using Support Vector Machines. In *2016 27th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 171–175). IEEE. https://doi.org/10.1109/DEXA.2016.046

Wang, L. X., & Mendel, J. M. (1992). Generating Fuzzy Rules by Learning from Examples. *IEEE Transactions on Systems, Man and Cybernetics*, *22*(6), 1414–1427. https://doi.org/10.1109/21.199466

Wang, S., Dong, Z. Y., Luo, F., Zhao, J., Kong, W., & Ranzi, G. (2017). Non-intrusive energy saving appliance recommender system for smart grid residential users. *IET Generation, Transmission & Distribution*, *11*(7), 1786–1793. https://doi.org/10.1049/iet-gtd.2016.1615

Wang, Z., & Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, *75*, 796–808. https://doi.org/10.1016/j.rser.2016.10.079

Wilamowski, B M, & Hao Yu. (2010). Neural Network Learning Without Backpropagation. *IEEE Transactions on Neural Networks*, *21*(11), 1793–1803. https://doi.org/10.1109/TNN.2010.2073482

Wilamowski, Bogdan M., & Yu, H. (2010). Neural network learning without backpropagation. *IEEE Transactions on Neural Networks*, *21*(11), 1793–1803. https://doi.org/10.1109/TNN.2010.2073482

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, *2*(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1

Yang, X. H., Guo, S. X., Zhang, Y. H., & Yang, H. T. (2008). A sort of new energy consumption forecast method. *Chinese Control and Decision Conference, 2008, CCDC 2008*, 3611–3615. https://doi.org/10.1109/CCDC.2008.4598003

Yang, X., Tan, L., & He, L. (2014). A robust least squares support vector machine for regression and classification with noise. *Neurocomputing*, *140*, 41–52. https://doi.org/10.1016/j.neucom.2014.03.037

Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, *42*(10), 1637–1646. https://doi.org/10.1016/j.enbuild.2010.04.006

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049

Zhao, J., Lasternas, B., Lam, K. P., Yun, R., & Loftness, V. (2014). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings*, *82*, 341–355. https://doi.org/10.1016/j.enbuild.2014.07.033

Zheng, Y., & Jose, A. A. (2019). Context-aware recommendations via sequential predictions, 2525–2528. https://doi.org/10.1145/3297280.3297639

isep Instituto Superior de Engenharia do Porto

Instituto Superior de Engenharia do Porto ⊙ Rua Dr. António Bernardino de Almeida, 431 ⊙ 4249-015 Porto
Tel. +351 228 340 500 ⊙ Fax +351 228 321 159 ⊙ mail@isep.ipp.pt ⊙ www.isep.ipp.pt