Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 417–422
www.elsevier.com/locate/egyr

The 8th International Conference on Energy and Environment Research ICEER 2021, 13–17 September

Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building

D. Ramos, P. Faria*, A. Morais, Z. Vale

Polytechnic of Porto (P. Porto), R. Dr. Antonio Bernardino de Almeida 431, 4249-015 Porto, Portugal

Received 31 December 2021; accepted 10 January 2022
Available online 3 February 2022

Abstract

The flexibility and management in the storage and control of building expertise in the energy optimization can be enhanced with the support of algorithms involved in forecasting tasks. These play an important role on obtaining anticipated and accurate consumption predictions associated to different contexts through extensive consumption patterns analysis. This paper evaluates the most viable forecasting algorithm for consumption predictions of a building in different contexts according to two alternatives: artificial neural networks and k-nearest neighbors. These algorithms use patterns of data from consumptions integrated in different contexts while retaining additional information from sensors data. The different contexts are classified on a sequence of periods that take place from five-to-five minutes. The decision criterion to evaluate which of the two forecasting algorithms is the most suitable in each five minutes periods is supported with decision trees that select the forecasting algorithms that looks to be more suitable followed by a logical answer that clarifies if the selection was the most viable option. Parameterization updates concerning the depth are studied to understand the forecasting accuracy impact. The decision trees approach has the potential to improve the accuracy of prediction as it plays a promising role in decision making.

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

Keywords: Decision tree; Load forecast; Neural networks

1. Introduction

The forecast of electricity consumption is very relevant for the and energy system planning, but also for the operation of smart grids in real-time, namely regarding the flexible control of electricity consumption, the so-called Demand Response (DR) [1]. A decision tree method is proposed in [2] to predict the energy consumption of a certain smart city. The decision tree uses three methods, the Fine Tree, the Medium Tree, and the Coarse Tree. The dataset is composed of 7 parameters which are, “date time”, “temperature”, “var1”, “pressure”, “wind speed”, “var2”, and “electricity consumption”. These parameters are used by the decision tree to predict the electricity consumptions. The decision tree method with the highest percentage of success is the Fine Tree method with 76,77%, followed

* Corresponding author.
E-mail address: pnf@isep.ipp.pt (P. Faria).

https://doi.org/10.1016/j.egyr.2022.01.046

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

by the Medium Tree method with 76,42% and the Coarse Tree method with 73,92%. According to [3] a decision tree method optimizes the energy management system and reduces the cost of energy consumption. An energy management system has been implemented in three floors of an office building and smart meters used to collect data. Three methods were used, Neural Network, Decision Trees and Gradient Boosting Decision Tree, and after the comparison of the three methods the conclusion is that the Gradient Boosting Decision Tree is the more capable method to predict the energy consumption. An alternative optimization application concerning a Decision Tree Method is proposed in [4]. Several steps of building a decision tree are explained, such as building a training set for it that is going to be use by the decision tree to predict the outcome of a certain input. Another important step of building a decision tree is forming the rules on which the decision tree, taking in account these rules and the training set, will choose the correct path in its prediction to generate the outcome. Having this into account, an optimized decision tree is an ideal tool to serve customers as well as the enterprises. According to [5] a decision tree method optimizes the use of energy resources. An agriculture system composed of Photovoltaic Panels, a synchronous generator, and the electricity network to drive the electric motors is considered. A decision tree, with a training set of 60 000 possibilities, was formed to control the amount of electricity produced by the generator and therefore preventing the excessive use of the electrical network with the objective of prioritize the use of a renewable power source.

The present work makes use of the forecast made in [6]. Although decision trees are used in a lot of applications, decision rules integrated in hierarchical methodologies, is has not been used to consider the benefit of analyzing the best forecasting algorithm in different contexts as evidenced in this paper. Section 2 presents the methodology followed by Section 3 featuring the case study and results, Section 4 presents all the conclusions.

2. Methodology

This section explores the sequence of steps of a proposed methodology consisting in the confirmation that the forecasting algorithm selection was the best choice in each period. Consumption forecasts are provided according to different prediction algorithms to support the decision rules obtained from the learning process. These decisions verify in each period if the selected forecasting algorithm was the most appropriate choice. The methodology process illustrated in Fig. 1 starts with the obtaining of energy forecasting consumptions according to different algorithms. This proceeds to the decision rule-based learning consisting in the decision rules influence to confirm the forecasting algorithms selections as the best choices depending on adjusted decision tree depth updates. The decision tree depth influences how many splits the decision tree performs to make the predictions. The selection of the best algorithm for the target context is the ending step for the proposed methodology.

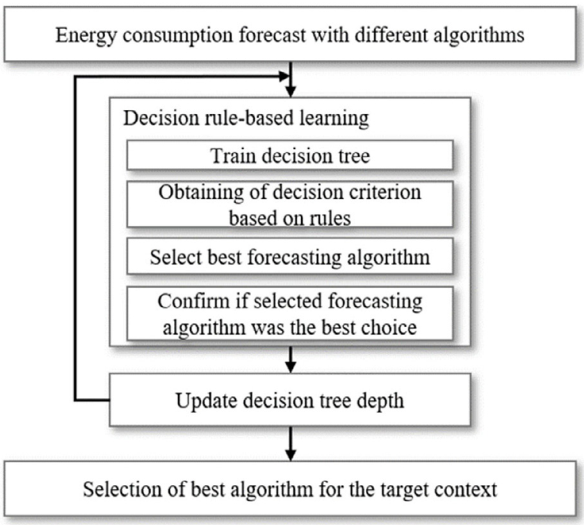

Fig. 1. Proposed methodology.

The methodology's first step consists in obtaining energy consumption forecasts according to two different prediction algorithms: Artificial Neural Networks (ANN) and K-nearest Neighbors (KNN). These forecasts are feed to a decision-rule based learning with the final goal of checking if the forecasting algorithm selection in each period was the most appropriate. This learning process starts to train a decision tree according to the observations obtained from the forecasts in the previous step and additionally with recording devices information from the actual and previous period. This information consists in the consumption and two sensor devices measuring CO2 and light intensity in the previous period and additionally time features which includes the day of the week and the actual period. Afterwards the decision tree automatically constructs decision rules involved in a decision criterion that selects the best forecasting algorithm and afterwards it confirms if the selection is the most appropriate choice. Verifications concerning that the selected forecasting algorithm was the most appropriate choice will result in the selection being consistent with final checking either if this selection was artificial neural networks or k-nearest neighbors. On the other hand, disagreements in the best choice application will result in the opposed algorithm being the most appropriated choice for each period. This learning process is applied to each period followed each time by a decision tree depth update to change for the best how these decision rules are created thus enhancing the decision criterion. Once the learning process is applied to all forecasting periods, the target period has selections of the best algorithm for the different contexts.

3. Case study and results

This section is structured in two sub-sections, namely 3.1 Case study, and 3.2 Results.

3.1. Case study

A sample of data obtained from electronic devices measuring different factors is studied for a whole week from 18 to 24 November 2019 considering 5 min contexts as illustrated in Fig. 2. The electronic devices measure unique factors including light intensity, CO2, the consumption and the period allocated. The profile concerning the day of the week is not studied as this has a linear behavior and it would make the remaining factors analysis more difficult to perform observations.

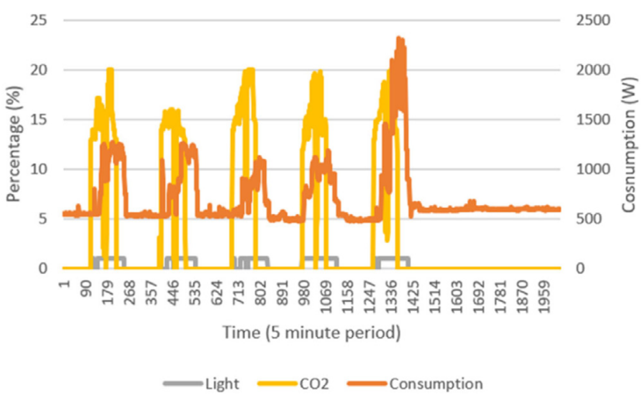

Fig. 2. Input parameters for train data.

3.2. Results

The results are conducted with the test data considering all five minutes periods that take place first in each hour beginning at 8 AM in the week from 18 to 24 November 2019 as illustrated in Fig. 3. The first five minutes periods in each hour between 12 AM and 7 AM are discarded as this integrates in a schedule out of the activity time considering the daily profile. The weekly profile of the different factors is studied considering five minutes contexts similarly to what was done to the train data in the case study section. The test considers the same features

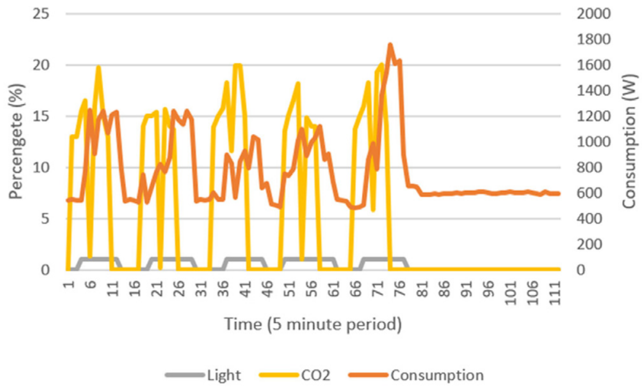

Fig. 3. Input parameters for test data.

of the training version including the consumption, the CO2 and the light intensity integrated in the first five minutes periods of each hour.

The results research the weekly profile of the different factors monitoring the changes that happen in the first five minutes of each hour. The test presents similar aspects with the training dataset evidencing for the different factors a sequence of 5 similar patterns each one with daily activity followed by 2 similar patterns presenting low activity. The first five similar patterns correspond each one to the activity of each day of the week from Monday to Friday. The following two similar patterns evidence the low activity respectively of Saturday and Sunday. The consumption weekly profile shows low consumption right in 8 AM and keeping the same behavior at the very least until 11 AM describing behaviors between 500 and 600 W. The activity periods start at some point between 11 AM and 12 PM describing consumptions above 700 and below 1300 W. This reaches higher activity between 4 PM and 7 PM. The low consumption is resumed at some point after 7 PM which stays until the 8 AM of the next day. The light intensity detects activity between 8 AM and 1 PM switching to no activity between 7 PM and 9 PM. The CO2 detects daily activity between 8 AM and 9 AM and ends it between 5 PM and 6 PM. The CO2 activity describes behaviors between 12.5 and nearly 20% from Monday to Friday. Between 12 PM and 2 PM there is a break for CO2 activity reaching the 0% before resuming the usual behavior. While light intensity and CO2 sensors present the daily profile with activity from Monday to Friday, the weekend presents daily profiles with no activity more specifically on Saturdays and Sundays as the sensors are inactive during this time. The consumption present from Monday to Friday is the usual with activity production, however weekends present low consumptions describing behaviors between 500 and 600 W.

A decision tree is used to predict a logical answer whether the select forecasting algorithm to predict consumption in each five minutes context was the most viable option. The data used by the decision rules considers the same one present in the case study involving the allocated period, the consumption and the CO2 and light intensity sensors allocated in the previous period. This scenario is target for all five minutes periods that start in each hour present in the week from 18 to 24 November 2019 as featured in the results section. Several scenarios are considered with this approach changing to different depth parameterizations which change the decision rules split complexity. The depth considers discrete values between 3 and 6 leading to four possible scenarios. These present the accuracy resulted from the depth parameterization used in each scenario as evidenced in Table 1.

Table 1. Accuracy of each depth scenario.

Depth	3	4	5	6
Accuracy	66.96%	66.96%	67.86%	71.43%

The accuracy results present in Table 1 show that increasing the decision tree depth results in higher accuracy. Thus, the decision tree depth increase makes the selection of the forecasting algorithm more robust meaning that the number of observations that the selected forecasting algorithm was the most convenient option increases. This is an understandable observation as increasing the complexity of rules will make these more accurate on the decision criterion involved in the forecasting algorithm selection. While the lower depth scenario featuring the parameterization assigned to 3 shows still a higher accuracy of 66.96%, it is observed some different accuracy variations while toggling this parameter. Increases from 3 to 4 show no accuracy improvement leading to the same results of depth assigned 3 of about 66.96%. Depth increases to 5 results in low accuracy improvements increasing it just to 67.86%. However, depth increases to 6 results on higher accurate results of about 71.43%. This means that the decision rules of depth scenarios concerning parameterizations assigned to 6 are complex enough to result on more accurate results being the forecasting algorithm selection more robust in this scenario.

The decision tree is presented for the first scenario featuring depth parameterization assigned to 3 as illustrated in Fig. 4. The decision tree constructs rules that are dependent on several parameters starting with the day of the week, following with the consumption with the possibility of using the sensors data known as CO2 and light intensity.

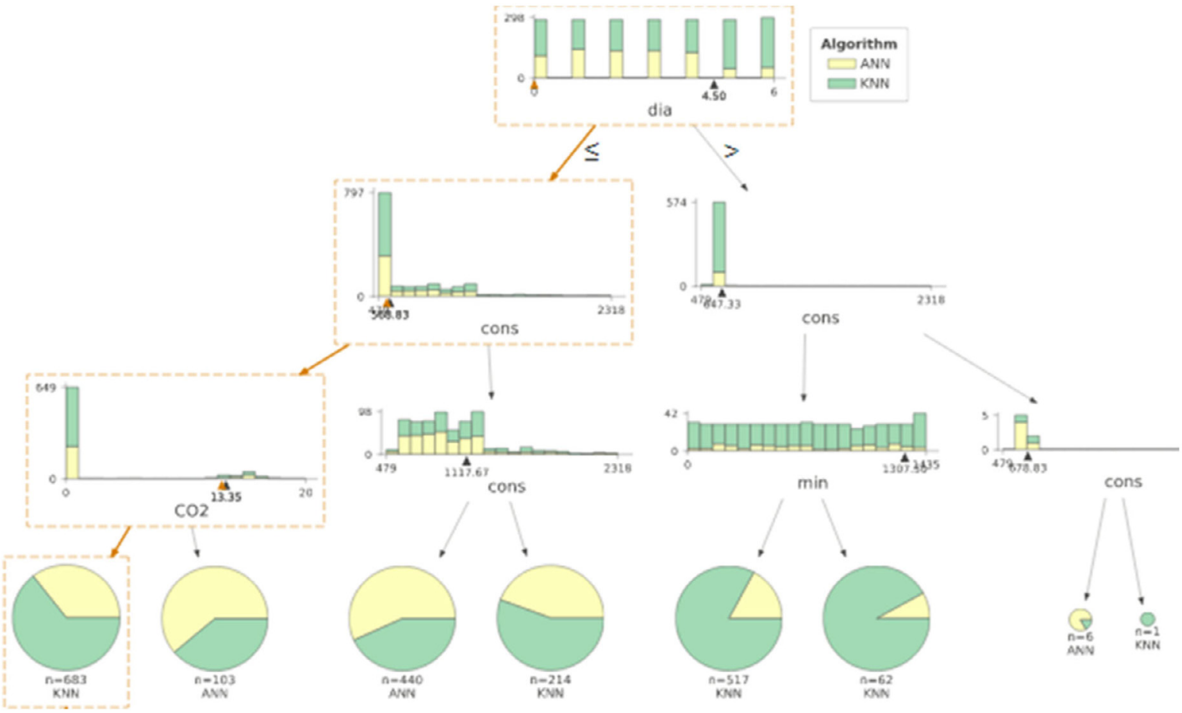

Fig. 4. Input parameters for train data.

The decision tree starts to split the data in two possible sets of data on left and right belonging respectively from Monday to Friday and the two days of the weekend. The rules applied from Monday to Friday option may use only the consumption data for values equal or above 479 W or this factor enhanced with CO2 data for observations with consumptions below 479 W which is expected to lead to more informative data. The two alternatives favor more KNN than ANN as the forecasting algorithm selection on five minutes contexts. However, using CO2 data has enough data to make KNN more trustable than ANN on more five minutes observations. The rules applied to weekend data do not use CO2 as the sensor is inactive during this period. The consumption is seen as expected to have low activity in the weekends which is not much reliable for the forecasting tasks selection as these present only variation during the whole weekend in very low ranges. Observations presenting consumptions above or equal to 479 W lead to low targets classified as ANN or KNN while on the other hand consumption below 479 W are reliable on the allocated period for forecasting tasks. Additionally, the light intensity sensor was not used as the rules created by the decision tree are not complex enough to add this information relying this on the depth parameter that in this scenario it is assigned to 3.

4. Conclusion

This paper presents pros and cons concerning the forecasting algorithm that looks more appropriate in each five minutes context among the two alternatives: artificial neural networks and k-nearest neighbors. These are supported by trial and test studies that verify if the forecasting algorithm selection was the most appropriate through a rule integrated in decision tree process. The depth parameterization shows that low values will result in high accuracy concerning the forecasting algorithm selection, however increasing enough may result in accuracy nearer to 100%. The forecasting tasks also show that these only benefit if these are applied to the week days from Monday to Friday as the weekends does not benefit from the sensors inactivity and the consumption presents low activity concerning a low range variation to present any insight if any of the two algorithm should be more trustable in each five minutes context. The sensors show to enhance the rules created in the decision tree supporting the expected insights that KNN is the most appropriated algorithm.

CRediT authorship contribution statement

D. Ramos: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **P. Faria:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **A. Morais:** Data curation, Formal analysis, Investigation, Software, Validation, Visualization. **Z. Vale:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The present work has been developed under the EUREKA - ITEA3 Project TIoCPS, Portugal (ITEA-18008), Project TIoCPS, Portugal (ANI|P2020 POCI-01-0247-FEDER-046182), and has received funding from European Regional Development Fund through COMPETE 2020 - Operational Programme for Competitiveness and Internationalization. The work has been done also in the scope of projects UIDB/00760/2020, and CEECIND/02887/2017, financed by FEDER Funds through COMPETE program and from National Funds through (FCT, Portugal).

References

[1] Faria P, Vale Z. A demand response approach to scheduling constrained load shifting. Energies 2019;12(9):1752.
[2] Yaman O, Yetis H, Karakose M. Decision tree based customer analysis method for energy planning in smart cities. In: International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain; 2020, p. 1–4.
[3] Yu X, et al. Load forecasting based on smart meter data and gradient boosting decision tree. In: Chinese Automation Congress (CAC), Hangzhou, China; 2019, p. 4438–42.
[4] Kang H, Zhao H, others, Ai T. The Description of Optimal Decision Tree Algorithm and Its Application in Customer Consumption Behavior. In: IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China; 2020, p. 655–59.
[5] Abrishambaf O, Faria P, Vale Z. Energy resource scheduling in an agriculture system using a decision tree approach. In: 20th International Conference on Intelligent System Application to Power Systems (ISAP), New Delhi, India; 2019, p. 1–5.
[6] Ramos D, Khorram M, Faria P, Vale Z. Load forecasting in an office building with different data structure and learning parameters. Forecasting 2021;3:242–55.