Scientific Paper

"Estimating window opening in a room"

Students :
GARAT Alexandre
SAINT OMER Jeanne

Supervisor :
FERRARI Jérôme

January, 17th 2022

Contents

Table of figures

Table of tables

Abstract

Smart systems are now everywhere in our life. Our smartphones, our cars, our buildings, have a lot of sensors. With the Internet of Things (IoT), huge amounts of data is sent every second. To be useful these data have to be treated, to determine some habits or to predict some behaviors.

More and more people want to have automatic and well-adapted autonomous systems that can ideally predict their desires and also intelligently reduce their consumption without losing comfort. It is only the beginning of the data analysis field. Informatic tools are the key for analyzing and building models.

The purpose of this study is to analyze the impact of opening the window in a room on life and health parameters such as temperature, humidity or amount of CO_2. Then, building a model that will be able to predict when the room is ventilated or not. This work will be helpful to detect if the window of a room is open, in particular in the case where this was not desired due to an oversight of the closure of the window or a problem. This could also replace windows sensors and allow some money savings.

Key words : Data Analysis, Smarthouse, Sensors, Predictive model, Decision Tree, Window Openings

I. Introduction

For this project, the study of a smarthouse will be conducted thanks to the access to data in real time of about 340 measuring points of a 120 m² household where a 5 person family lives. The Expe-smarthouse was initiated in 2018. A wide variety of sensors are available (CO_2 concentration, humidity, temperature…) as well as recorded data measured for at least one year. The figure below shows all the possible data available in each of the rooms in the Expe-smarthouse.

Figure 1 : Sensors available in the smarthouse

The purpose of this project is to analyze data from a few sensors to find some correlations or to build predictive models by using the knowledge acquired with the Smart Systems module this year.

Our study will be focused on using machine learning in order to be able to predict window openings in a room using different parameters : the amount of CO_2 in the bedroom, the room's temperature and humidity and the outdoor's temperature and humidity. The goal of this work will be to build a predictive model which can determine the times when the bedroom is opened or not.

II. Data collection

In order to study the correlation between the aeration of the bedroom 1 of the smarthouse and the variation of humidity, temperature and the amount of CO_2 concentration, the access to this data as well as the state of the window according to time for this bedroom is needed. To have enough data to build a model, the decision has been taken to collect them between the 1st of July and the 31st of october, so four entirely months. These data were imported on a csv file from the website grafana.

The types of the different used data are :

- Temperature : float
- CO_2 : float
- Humidity : float
- Window : boolean

All the data were not collected at the same exact time. For example, for the state of the window, the sensor only detects the time when the window is opened or closed. For the future analysis, it was necessary to sample them to have all the data available for each datetime.
A first step was to put the time value in a datetime format. To construct the list of the window states at each of these times, some rules were defined. The states were initialized at 0 until a state of 1 was reached and after the states were put at 1 until the next already existing 0 (an illustration of this can be seen in Fig. 2). For the other parameters a linear upsampling has been applied to obtain the values for each 2-minute.

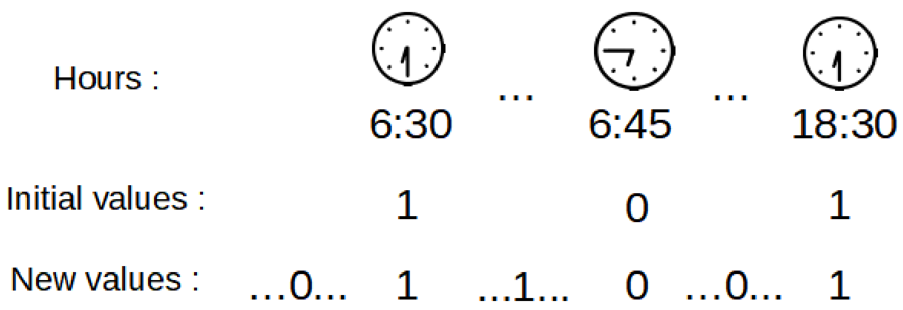

Figure 2 : Illustration of filling windows states

Sometimes, the opening of the windows lasts less than one minute because it corresponds to the opening for closing the shutters for example. These events were not considered interesting in the following study because the time of opening the window is not enough to observe some evolution on the different parameters of the bedroom.
This issue was treated by sampling data with a 2-minute resolution. This also allows to decrease the number of data and then the time needed to run the script. Finally, there are 88 543 timestep and therefore as much data for each feature.

An example of data used for the 2nd of July is show in the next figure with in purple the time where the window is open :

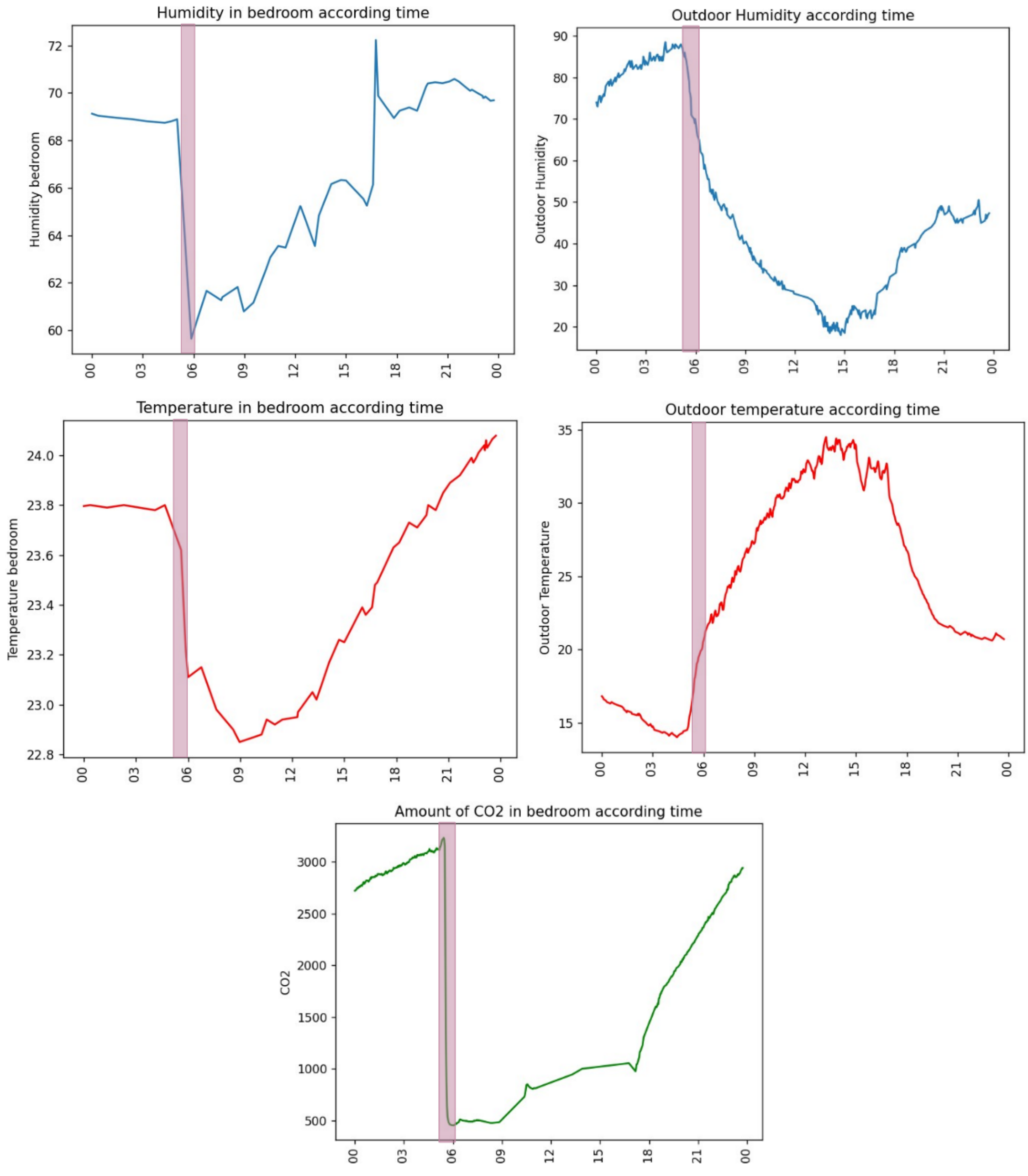

Figure 3 : Example of evolution of the parameters (2nd of July 2021)

In this example, the opening of the windows tends to significantly decrease the humidity, the temperature and the amount of CO_2 in the bedroom. It can be a first confirmation that these parameters play an important role in the predictive model we want to build.

III. Building predictive model

1. Decision Tree Classifier

The predictive model for estimating window opening in a room was built using supervised learning thanks to a set of training and testing data. The predictive model used the Decision Tree algorithm. Decision Tree algorithm is a Supervised learning technique. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

The data were separated into a training and a testing set. 70% of the data were randomly chosen to be part of the training data and the 30% remaining were constituting the testing set.

Decision Tree were built for different max depth in order to visualize the complexity of the algorithm.

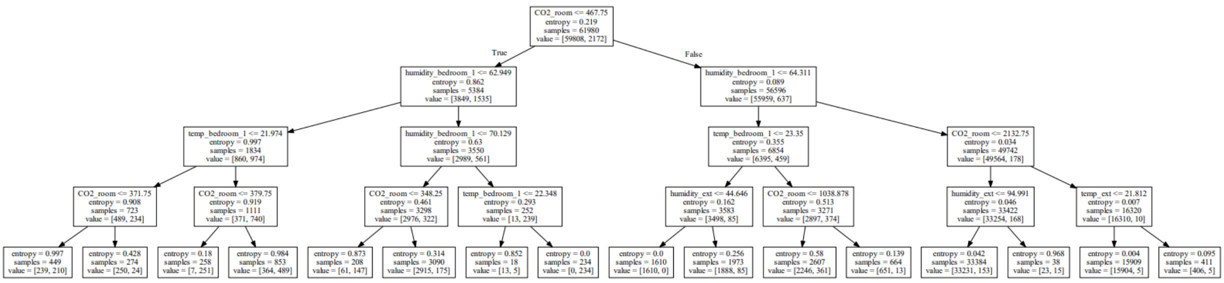

Figure 4 : Decision tree for a max depth of 4

With the max depth becoming high, the Decision trees quickly started to be very dense and complicated to read.

2. Choice of relevant sensors

The obtained Decision Trees were quite dense and complicated because of the numbers of different parameters and the complexity of the problem. Thus, a study of the evolution of the accuracy in function of the used parameters was conducted in order to determine the relevant sensors for the study of the window openings.
In order to minimize the number of simulations that needed to be run, it was estimated that the bedroom's temperature and humidity were needed to obtain a relevant model. Thus, only 8 scenarios needed to be compared (compared to 32 without this assumption) :

Sensors	Bedroom temperature	Bedroom humidity	Bedroom CO2 level	Outdoor humidity	Outdoor temperature
Scenario 1	✔	✔	✔	✔	✔
Scenario 2	✔	✔	✔	✔	✘
Scenario 3	✔	✔	✔	✘	✔
Scenario 4	✔	✔	✔	✘	✘
Scenario 5	✔	✔	✘	✔	✔
Scenario 6	✔	✔	✘	✔	✘
Scenario 7	✔	✔	✘	✘	✔
Scenario 8	✔	✔	✘	✘	✘

Table 1 : Chosen sensors for each scenarios

The choice of sensors was based on the accuracy obtained for the different scenarios.

One particularity of our data is the imbalance existing between the number of 1 (window is opened) and 0 (window is closed) since a window is often much more closed than opened in a day. Indeed, in our set of testing data, only 3,8% of window state values are equal to 1.
This means that even if the algorithm decides to affect 0 to all the window states, the accuracy would be 96,2%. Thus, in order to go above this, the scenarios were compared using a mean accuracy. This mean accuracy was calculated for each scenario as the mean value between the accuracy on the 0 values and the accuracy on the 1 values.

All the simulations were run with a max_depth of 7 for this comparison in order to limit the calculation time of the algorithm.

The following results were obtained :

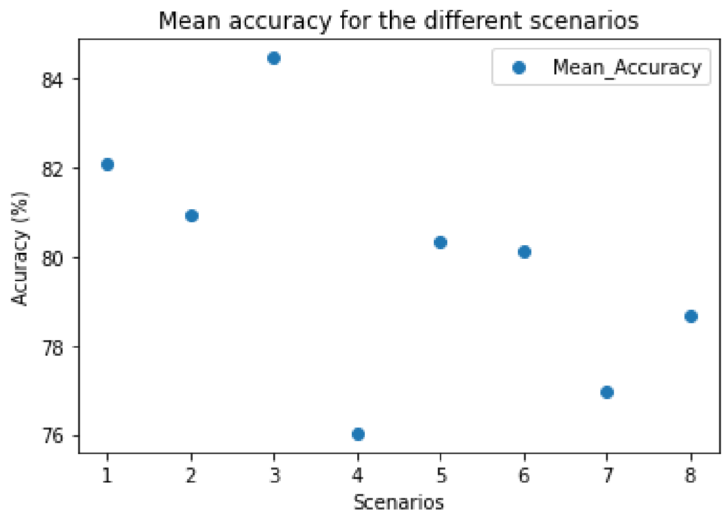

Figure 5 : Mean accuracy in function of the simulated scenarios

As shown by Figure 5, the best mean accuracy is obtained for the third scenario which is the one where the outdoor humidity data are not used.
For this scenario, the mean accuracy is 84,46% (the accuracy on 1 values is 69,7% and the accuracy on 0 values is 99,22%) whereas the global accuracy is 98,09%.

Thus, for the following parts, the simulations were run without using the outdoor humidity data.

IV. Obtained results and analysis

1. Study of the impact of the chosen max depth

A key parameter of the Decision Tree Classifier is the max_depth chosen. It indicates how deep the tree can be. Thus, the higher the max_depth, the more complex the Decision Tree will be and the longer the simulation will take to be run.

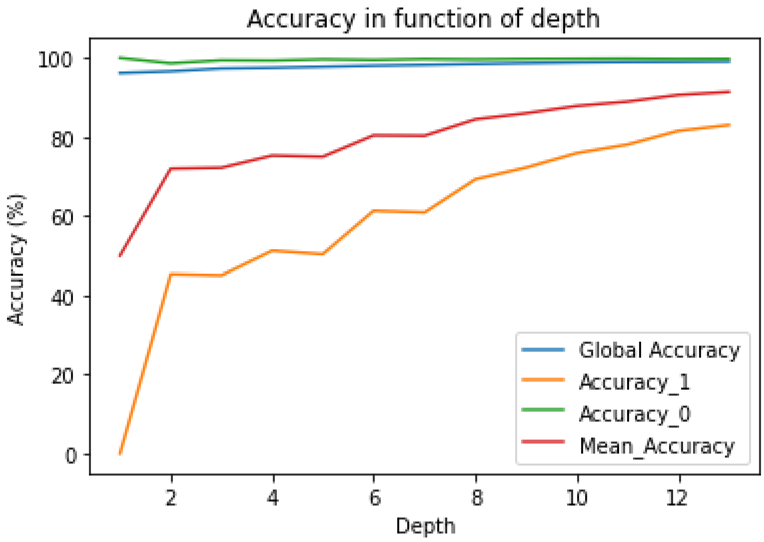

Figure 6 : Accuracy according to the depth value of the tree

Figure 6 shows that the higher the depth, the better the accuracy.
However, it also shows that even with a lower max_depth the obtained results can be considered relevant. Thus, by estimating that we need a mean accuracy of at least 85% and at least 75% for the accuracy of the 1 value, selecting a max_depth of 10 is convenient.
This allows to simplify the Decision Tree and to decrease the time needed for the algorithm to be run.

2. Analysis of the results

The different results (Figure 6) shows that the accuracy of the value 1 (window is opened) is much smaller than the one of the value 2 (window is closed). This can be explained by the fact that the value 1 has fewer occurrences in the training set so that it is more complicated for the algorithm to predict a 1 (opened window) than a 0 (closed window). Also, when opening the window, the different studied parameters can take some minutes to evolve. For example, the temperature might not evolve immediately when opening the window.

For a max_depth of 10, the following results are obtained :

Accuracy	Value
Accuracy_0	99,78%
Accuracy_1	75,91%
Mean_Accuracy	**87,85%**
Global_Accuracy	98,87%

Table 2 : Obtained results of accuracy with a max_depth of 10

The results show the importance to have in mind the imbalance in the repartition of data when the window is opened or closed.
However, the results are really satisfying since the algorithm is able to predict with a good accuracy if the window is opened or closed.

V. Conclusion

In this article, an application of a supervised method has been developed and the main results were described. The prediction model obtained has a satisfying accuracy which means the algorithm is able to predict if the window is opened or closed. This model could then replace the window opening sensors for this bedroom. An extra work could have been to adapt this model to other rooms of the house in order to be able to also replace the window opening sensors of these rooms and do some money savings.
Also, even if it was not treated in this study, it could be interesting to improve this algorithm by using a Random Forest classifier instead of a single Decision Tree in order to improve the accuracy.

References

[1] « EXPE-SMARTHOUSE The full connected living house ». Consulté le: 31 janvier 2022. url: https://expe-smarthouse.org/index.php/en/project/

[2] J. Ferrari, « Grafana ». Consulté le: 17 janvier 2022. url: https://jarvis-oneforall.duckdns.org:3000/login

[3] J. Ferrari, « Security risk on a smart home », 29 janvier 2021. Consulté le: 31 janvier 2022. url: https://miniprojets.net/index.php/2021/01/29/security-risk-on-a-smart-home/