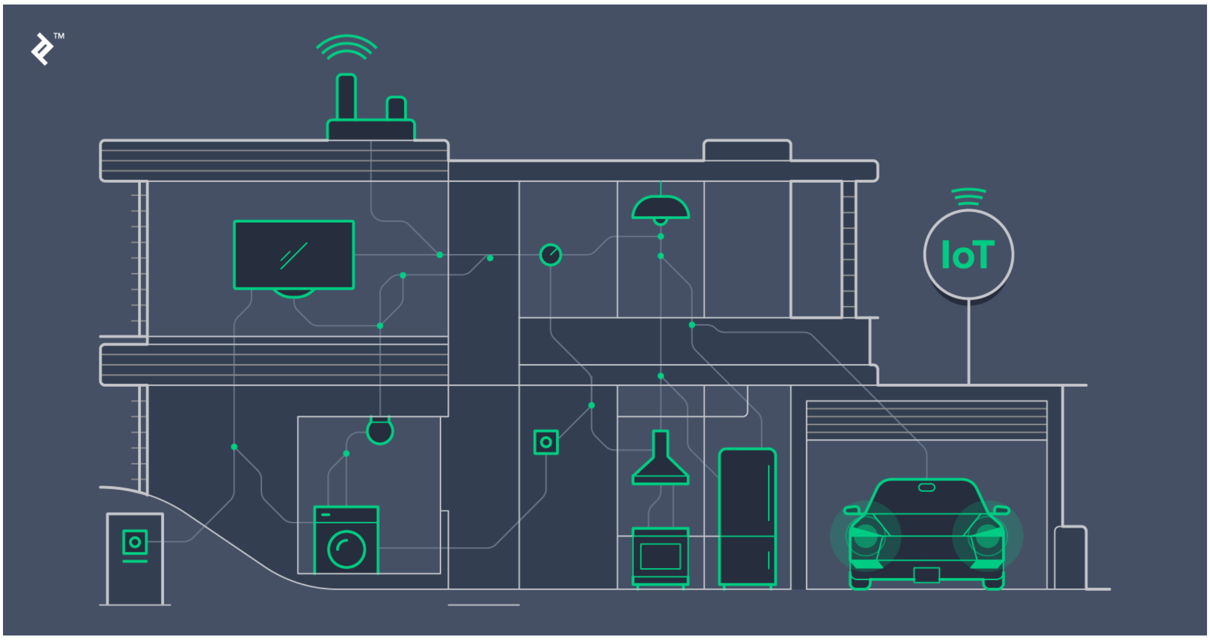
Grenoble INP

École Nationale Supérieur de l’Énergie, l’Eau et l’Environnement

Lab Work BE - M2 SGB



Machine Learning & Optimization (5EU9MLO0)

Should I open the window?

How to detect the optimal temperature and air flow for a room

Jérôme FERRARI

Hassan ISSA

Constanza MUÑOZ FILIPPIG

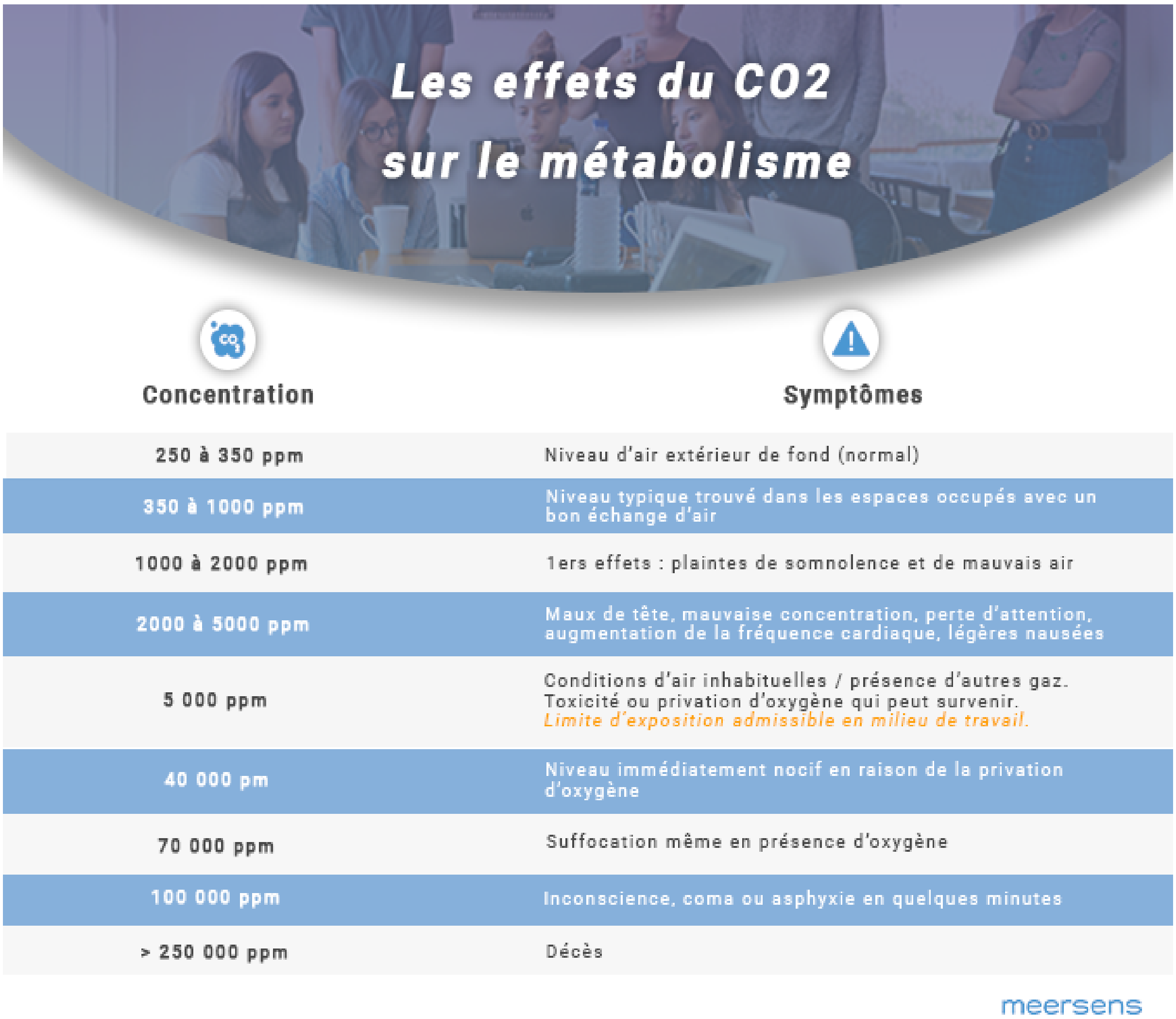
Francisco Javier CUESTA ELIAS

Due date: January 29th, 2022

=INTRODUCTION=

Buildings are a vital stakeholder toward energy savings. Advanced research has been done on this domain to adapt and change behaviors for an efficient use of the energy inside them. From using better insulation materials to sensors acquiring data to make decisions on the optimal temperature or amount of air flowing; these controls will allow to make cuts on the energy bill. Smart buildings are currently developed not only on new construction projects, which are a must, but also on actual existent buildings which are adapted to these new technologies to really generate an impact on reducing energy consumption.

In this project, an analysis is made on the living room which includes multiple sensors. The data obtained by them will undergo a treatment and a machine learning model will be executed to understand the relation between these factors and answer to the following question: When is the right moment to open the window considering to avoid bad indoor air quality. It's relevant to know what can happen if air quality is not good and the effects it might have on the human body, taking into account the recent development of diseases that can be transmitted through air such as COVID-19. Temperature impact will also be taken into account for window opening.



Concentration	Symptômes
250 à 350 ppm	Niveau d'air extérieur de fond (normal)
350 à 1000 ppm	Niveau typique trouvé dans les espaces occupés avec un bon échange d'air
1000 à 2000 ppm	1ers effets : plaintes de somnolence et de mauvais air
2000 à 5000 ppm	Maux de tête, mauvaise concentration, perte d'attention, augmentation de la fréquence cardiaque, légères nausées
5 000 ppm	Conditions d'air inhabituelles / présence d'autres gaz. Toxicité ou privation d'oxygène qui peut survenir. *Limite d'exposition admissible en milieu de travail.*
40 000 pm	Niveau immédiatement nocif en raison de la privation d'oxygène
70 000 ppm	Suffocation même en présence d'oxygène
100 000 ppm	Inconscience, coma ou asphyxie en quelques minutes
> 250 000 ppm	Décès

=DATA ACQUISITION=

For this experience a database with several sensors in the living room of a smart home was used. The data of the sensor was acquired through a raspberry pi connection to acquire csv files of: temperature, humidity, CO2 concentration, Sound intensity, electrical consumptions of appliances and window opening.

=DATA TREATMENT=

Once this data was acquired, we found out that each of the measurements had a different time stamp; therefore, as the first step of the data treatment, these measurements were standardized into timescales. Then the data set was prepared on a list according to these timescales, NaN values appeared and they had to be clean, so we just played with data that we had in similar times to compare. After cleaning the data, a model will be executed according to the information we have to answer to the previous posed question.

=MODEL=

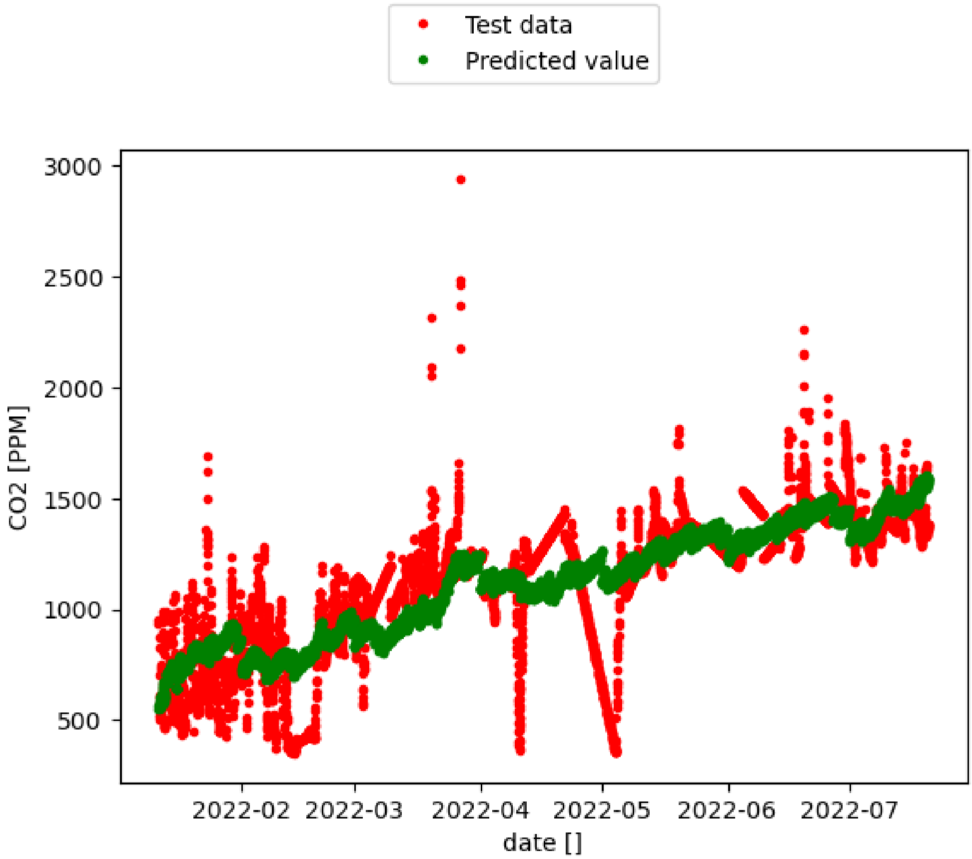
The first part of the model will determine the temperatures in the room and with these results determine if it’s needed to open the window or not. Two statistical methods were used to analyze the relationship between the variables which are: linear regression and ridge regression. The first one allows us to estimate the relationship between the quantitative variables and the latter uses a regularization coefficient which helps to avoid overfitting. The difference between the two R2 scores was not significant; hence, regression was chosen.

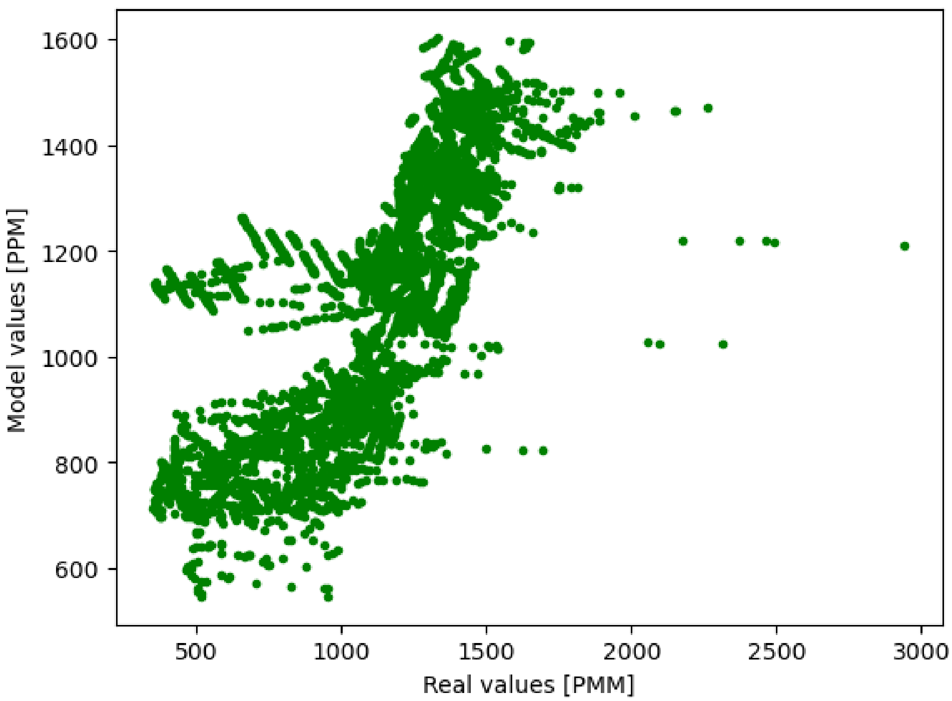
=RESULTS=

According to multiple sources, the level of CO_2 concentration should be below 1000 ppm. The following picture depicts the classification of air quality according to the concentration.

CO_2 [ppm]	Air Quality
2100	BAD Heavily contaminated indoor air Ventilation required
2000	
1900	
1800	
1700	
1600	
1500	MEDIOCRE Contaminated indoor air Ventilation recommended
1400	
1300	
1200	
1100	
1000	FAIR
900	
800	GOOD
700	
600	EXCELLENT
500	
400	

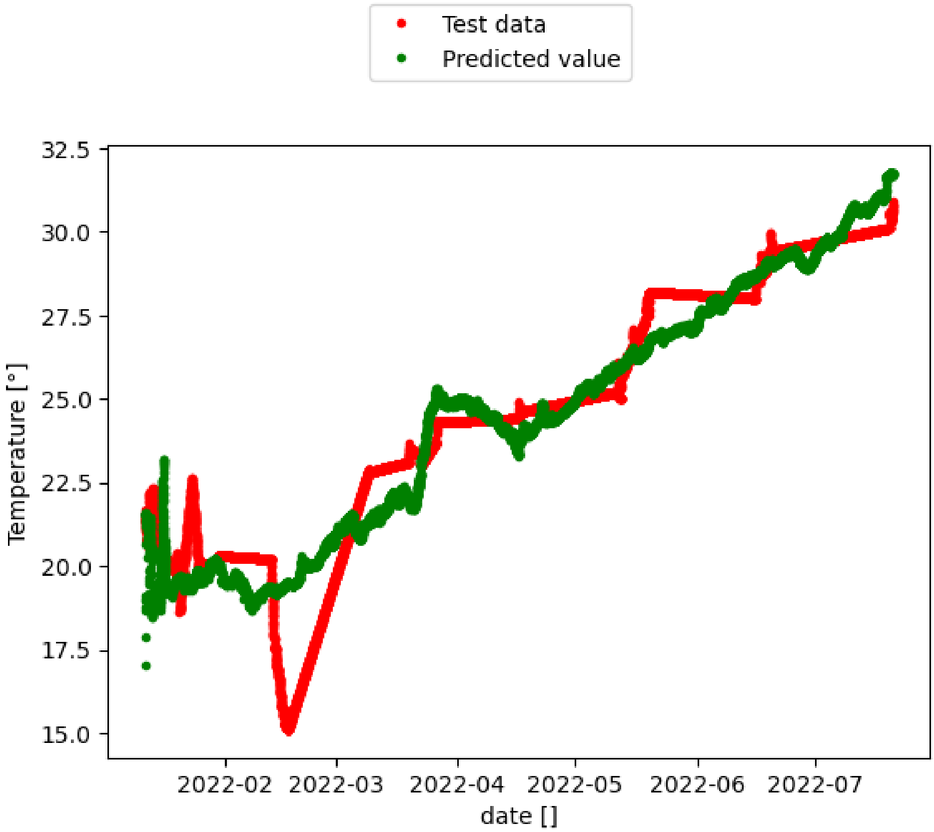
The following graphs show the relationship between the test data and the predicted values of our model considering the CO_2 concentration. The first one is a comparison within the same timeframe and the latter between them directly which gives an R2 score value of almost 60%.

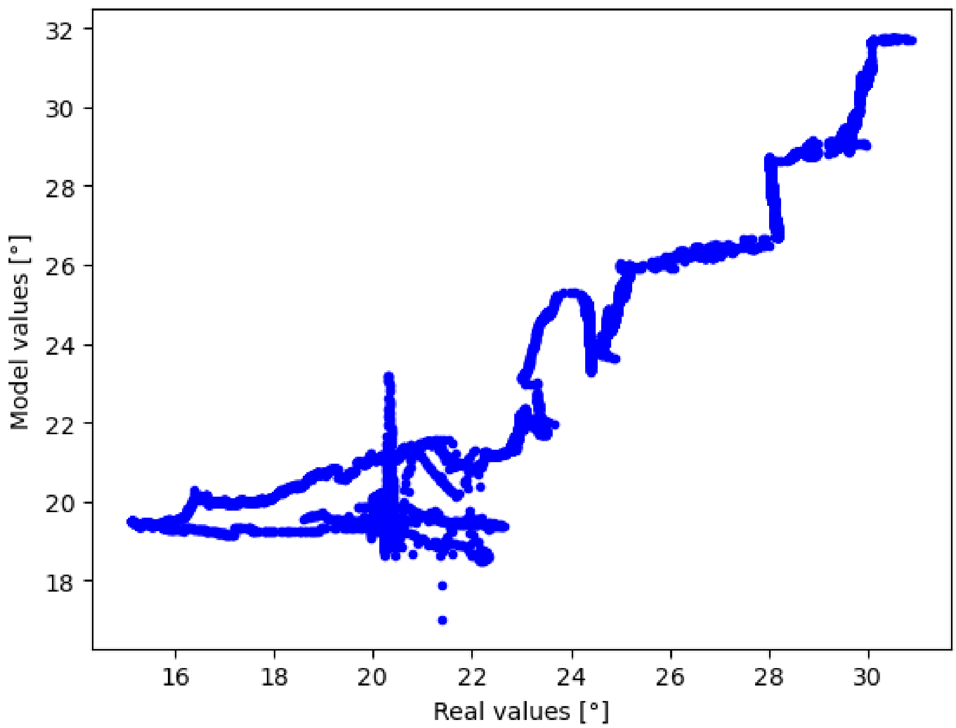




R2 Score = 0.5913479445878065

The following graphs show the relationship between the test data and the predicted values of our model considering temperatures. The first one is a comparison within the same timeframe and the latter between them directly which gives an R2 score value of 91%.

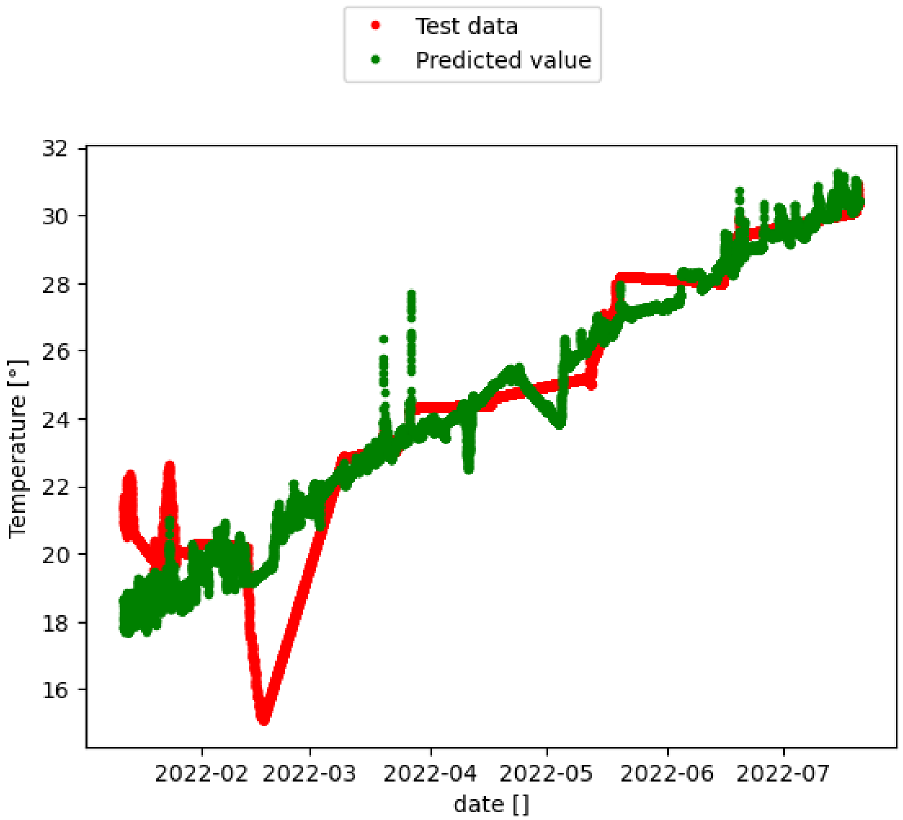


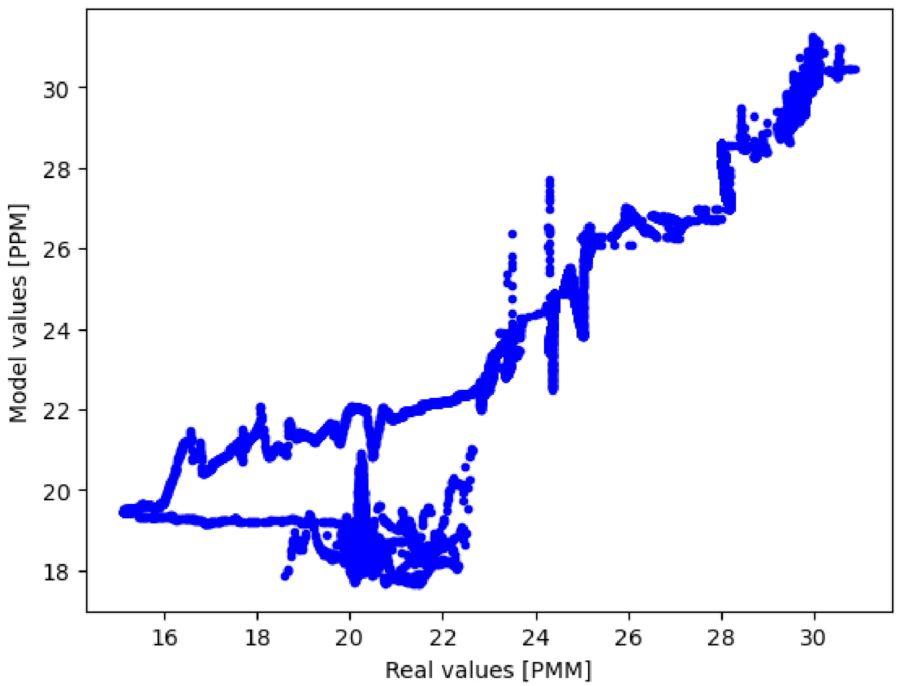


R2 Score = 0.9126802751489851

This final model considers both temperature and CO_2 by having the sensor contrary to the two before separated cases; this scenario is more complete and whilst losing only 2% of R2 score to 89% from the lone temperature model, drastically shows an improvement for CO_2.

Final model + a sensor of CO2

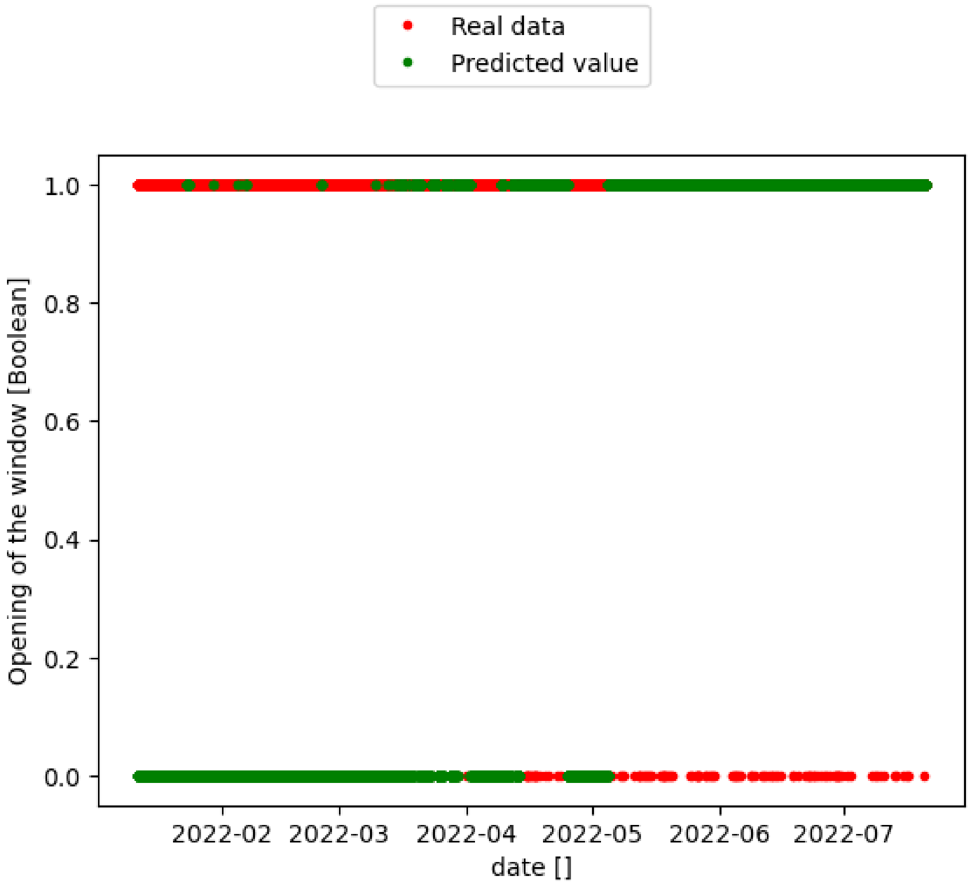




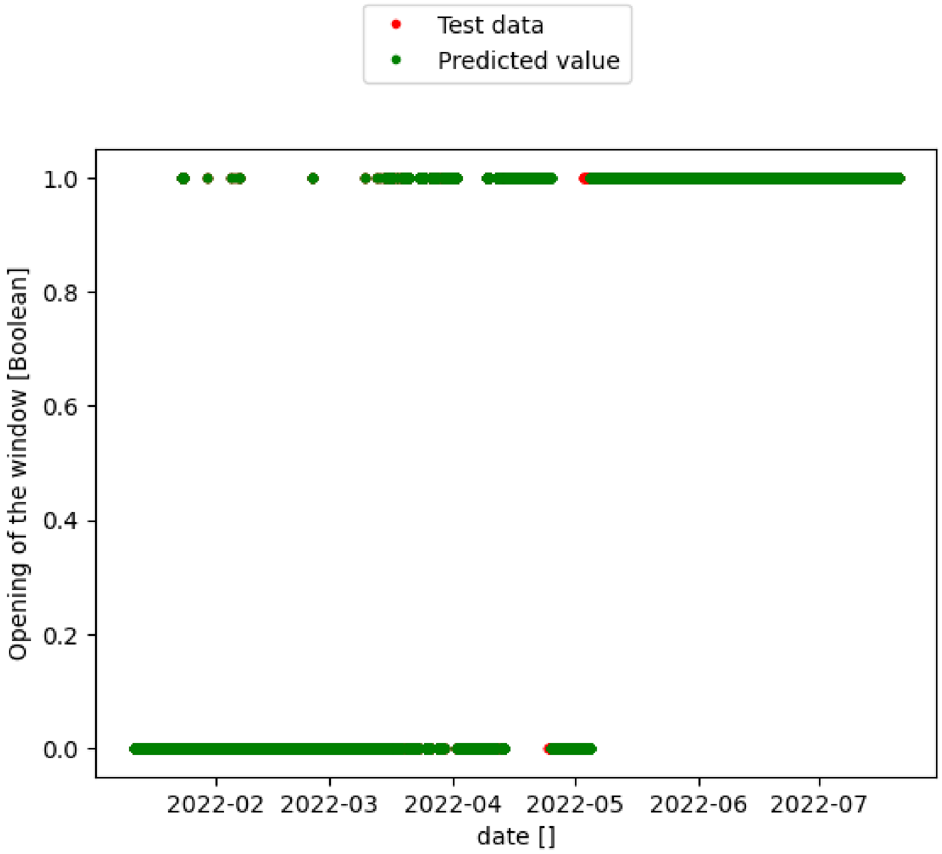
R2 Score = 0.8911663847052653

Afterwards a simple model for controlling the opening of the window was developed with the data acquired from a sensor and the data obtained from our machine learning method. The commands where to open if the measured concentration is higher than 1200 PPM or if the estimated temperature is higher than 25°C.

The trained model was used with test data to know when it's needed to open the window. Then we compare it with the given known data of the opening of the window in the room, it was concluded that there’s no concrete evidence or relation that people open the window when the CO_2 concentration or the temperature are high; therefore, realizing that this window is opened according to their own will.



Finally to assure that the model for opening the window would work properly we applied the same control strategy with the measured temperature and measured concentration of CO2 (the real data) in that period and it was compared with the results of the model with the data from a sensor of CO2 and the estimation of the temperature. It obtained a solid result of R2 score of 95% which means that this method and the models used work well to represent those actions.



R2 Score = 0.9518706535886697

=CONCLUSION=

One of the most important steps in machine learning is having useful data, cleaning it to compare within similar parameters and characteristics (standardization) and skills to identify and interpret results.

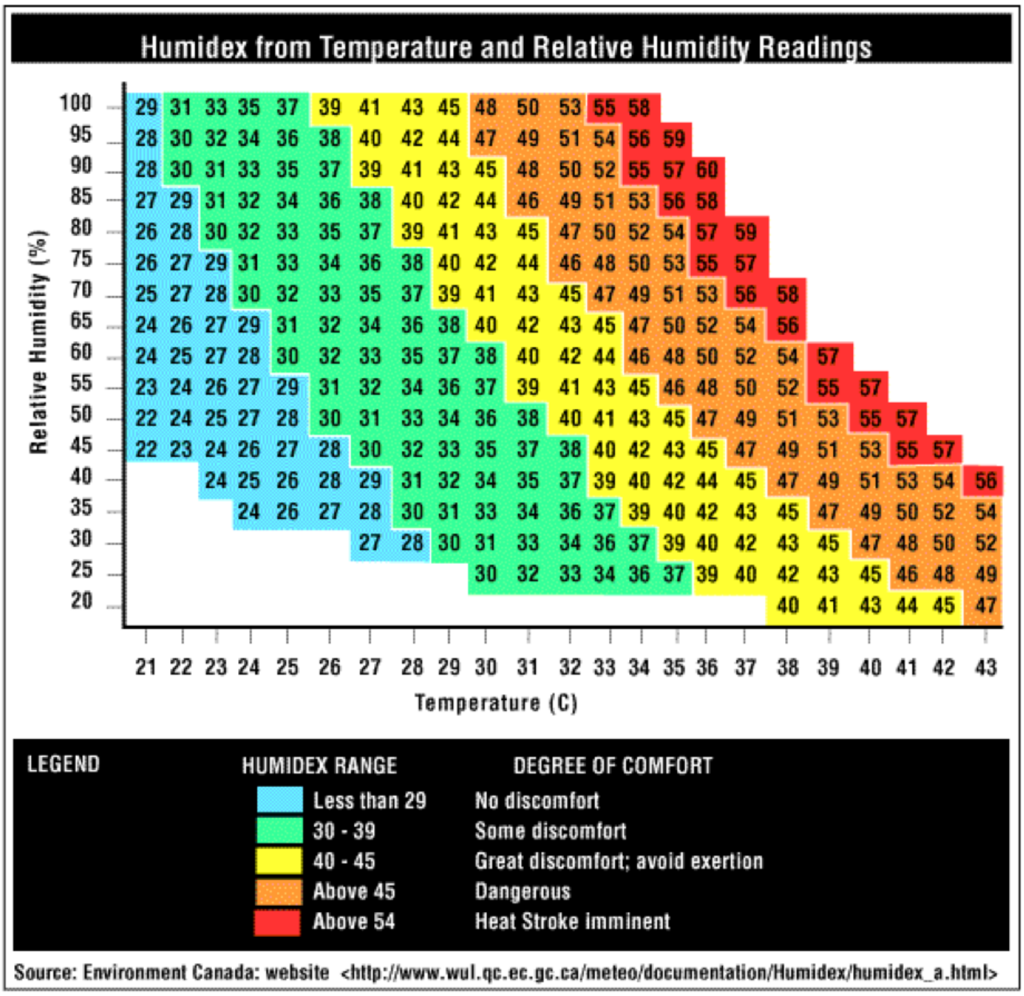
The model obtained through this experience is a good starting point. It is possible to have a rough idea of the moments during the day when it could be useful to open the window to allow air flow and a better general comfort in the room, however this model has several flaws such as the necessity to install a CO_2 sensor as well as not knowing exactly how much time should it be open and how this would impact into the model for estimating the temperature of the room.

A possible improvement is to better study the control strategy to assure a proper airflow as well as temperature in the room, in other words how long should we maintain open the window and how this will impact the temperature of the room. It would be necessary to have a better reference of the temperature outside of the building during the year; this would allow us to study the implementation of a model to predict the temperatures or implement temperature sensors outside. Even some estimation of the dimensions of the window would be necessary to obtain an estimated heat transfer model.

For further analysis, since there's a boiler sensor in another room. For an additional analysis, the model could be replicated to turn on the boiler for certain outside temperatures and inside values considering hours of movement.

Another possible study considering the data we got is to determine the degree of comfort by the relationship between the temperature and the air humidity. The following graph gives an idea of how the relationship between these two variables is and; therefore, search a relationship, if any, that people in the house open the window after getting to an uncomfortable range.

Machine Learning methods are a good method when you have an idea of what you're looking for, to identify patterns and interpret results. This basis allows in-depth analysis to determine decision making.



=REFERENCES=

https://meersens.com/indoor-air-quality-the-impact-of-co2-on-health-and-wellbeing-at-work/?lang=en

https://www.iqhome.org/index.php?route=extension/d_blog_module/post&post_id=17

https://meersens.com/indoor-air-quality-the-impact-of-co2-on-health-and-wellbeing-at-work/?lang=en

https://medium.com/mlearning-ai/linear-regression-vs-ridge-regression-vs-lasso-regression-467d0df39cc9