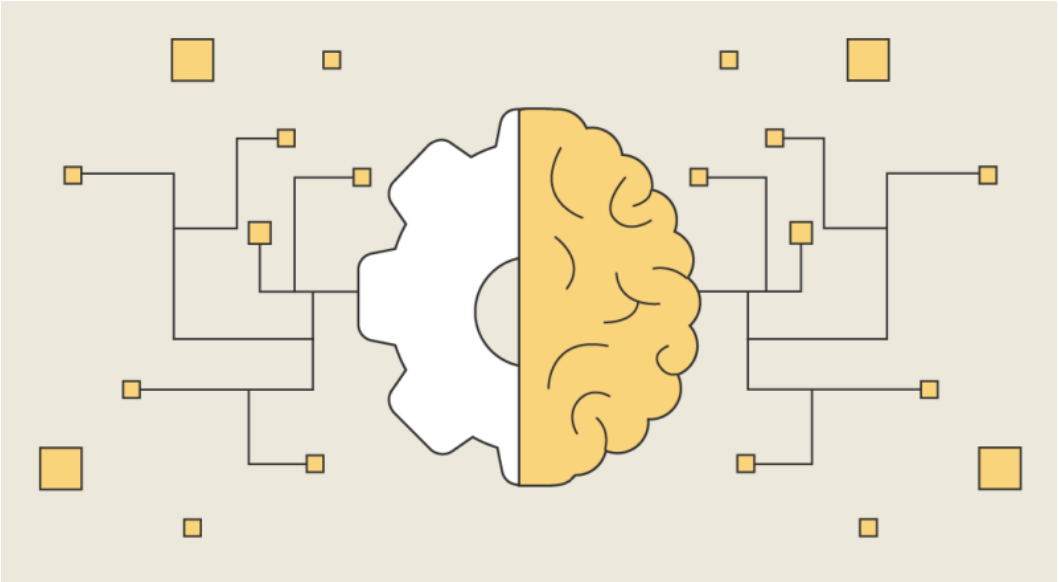
GRENOBLE INP Ense3 UGA

Machine Learning & Optimization



TP - Electrical Consumption Estimation in the Smart Home

M. Jerome FERRARI & M. Hassan ISSA

Presented By:

Shariq NADEEM

Solomon Berihu ARAYA

26th Jan, 2023

Table of Contents

1. Introduction

Demand prediction and demand response are becoming more and more crucial because of the rising proportion of renewable energies in the electrical mix and the accompanying unpredictability of electricity generation. The need for electricity frequently varies with the weather; for example, when it is cloudier, there is a greater demand for inside illumination. Solar panels are also currently producing less electricity than usual. In order to ensure grid stability and timely activation of backup resources, this requires the forecasting of the electrical demand. The necessary generation can then be coordinated by grid operators to meet the demand.

This report discusses the estimation of the electrical consumption inside the Smart Home based upon the previous consumption data and also to see the relationship with certain physical quantities like temperature, humidity, luminosity etc. This smart house project grants access to about 340 measuring points for scientists, accessible in real time through a Grafana portal with Influxdb database. It has been developed based on Open Source Hardware and Software. It is also flexible in adding new technologies or sensors with time.

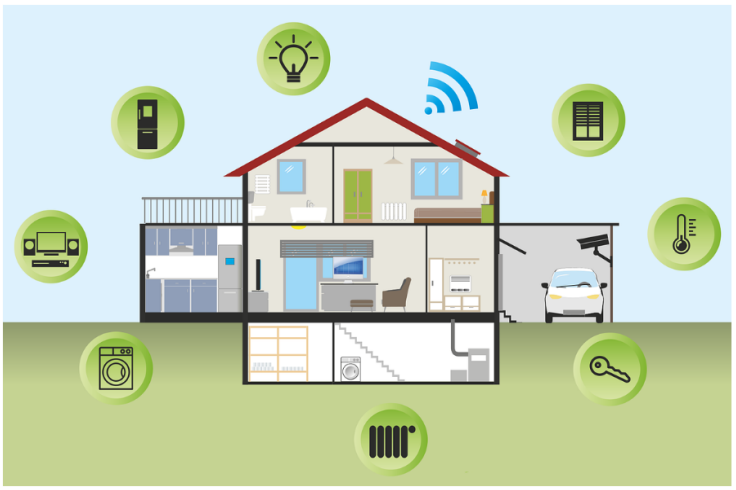


Figure 1 : Smart Home

There are measurements with the help of sensors of different quantities like:

- ➔ electricity, gas and water consumption of each device
- ➔ temperature, humidity and brightness of each common room
- ➔ opening position of each door and window
- ➔ motion sensors
- ➔ air analysis of each room
- ➔ outdoor weather conditions etc.

Figure 2 : Map of different sensors on the platform

2. Methodology

First of all, the data acquisition was performed with the help of *Putty.exe* software which was used to retrieve the necessary sensor data of one year values from the database. The first session of the TP was about acquiring the data and making our own local database in *raspberry pi.*

The following instruction was used to name and make an empty dataset:
sudo nano database

The below instruction was used to access the required data file and download it to the raspberry pi.
sudo python3 database

Afterwards, the data was accessed by a flash drive from the raspberry and used for the modeling and estimation purposes.

Selected sensors:

Below are the sensors selected for which data was obtained:

'id': '5224'	'name': 'Full_Home_PZEM004/Puissance'

'id': '7029'	'name': 'Luminosité Salon V2/Luminosité (Lux)'

'id': '3917'	'name': 'Capteur TH Ch1/Température'
'id': '3918'	'name': 'Capteur TH Ch1/Humidité'

id': '734'	'name': 'Détecteur Ouverture Porte Entrée/Ouverture'

'id': '4296'	'name': 'Détecteur Ouverture Fenêtre Chambre 1/Ouverture'

Sensor Data Plotting:

The data obtained from the sensors has been plotted as below to see the trending and nature of the data.

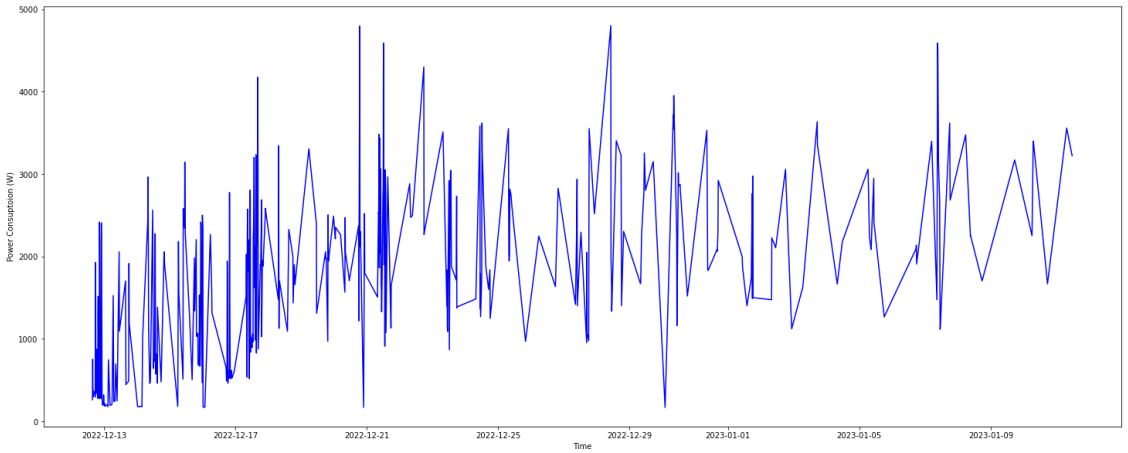


Figure 3 : Power Consumption

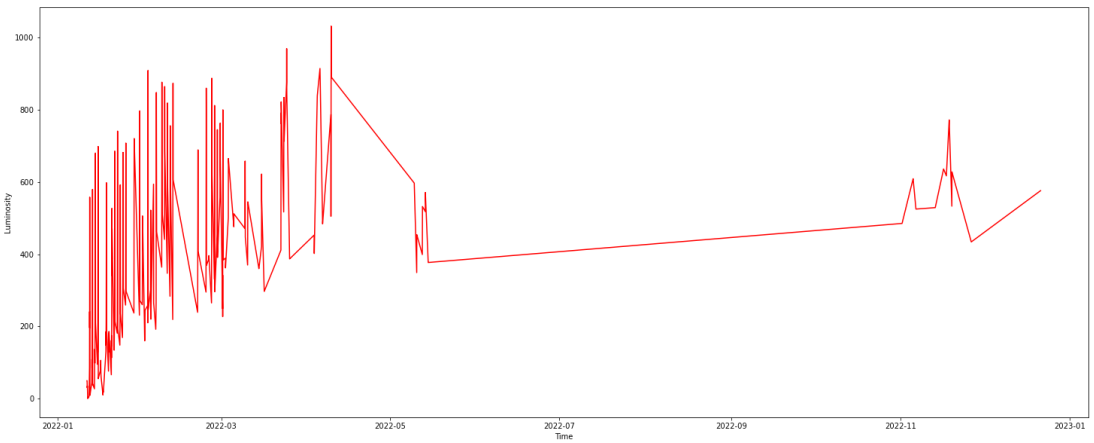


Figure 4 : Luminosity (Room)

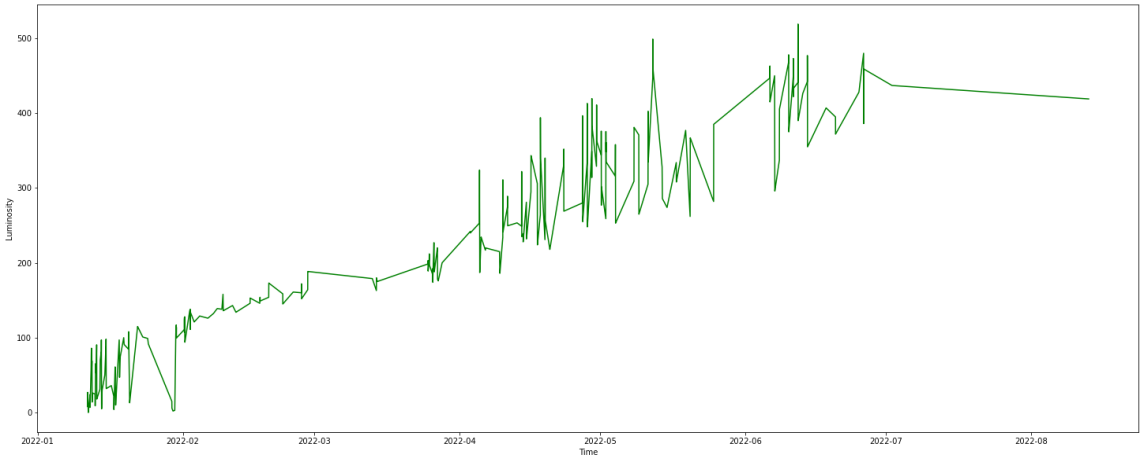


Figure 5 : Luminosity (Kitchen)

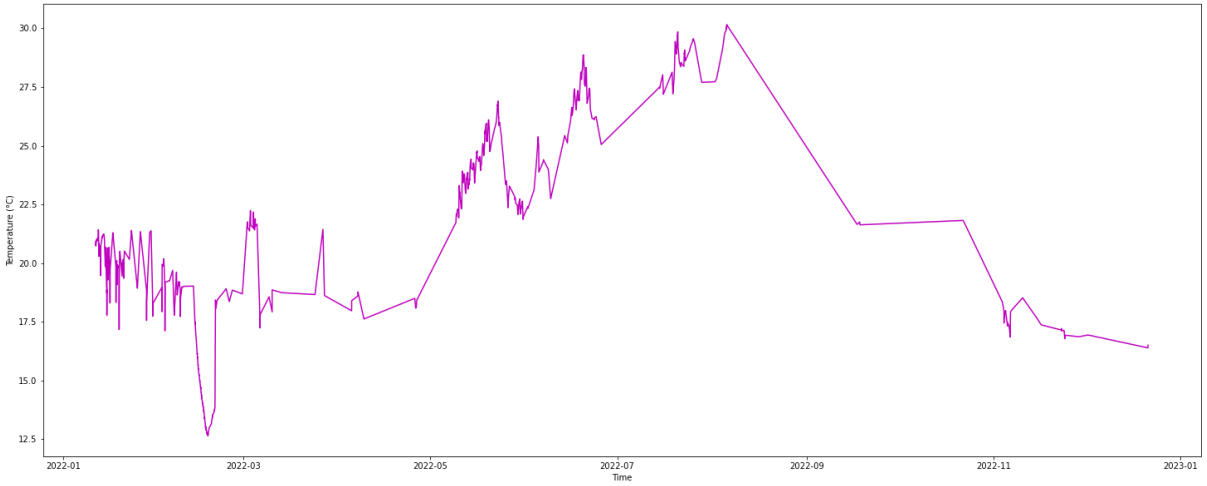


Figure 6 : Temperature

Figure 7 : Humidity

3. Machine Learning Model & Results:

The data available in the form of .csv files is continuous therefore the model developed will be regression based. Machine Learning Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. It's used as a method for predictive modeling in machine learning, in which an algorithm is used to predict continuous outcomes.

Solving regression problems is one of the most common applications for machine learning models, especially in supervised machine learning. Algorithms are trained to understand the relationship between independent variables and an outcome or dependent variable. The model can then be leveraged to predict the outcome of new and unseen input data, or to fill a gap in missing data.

The predicted power consumption along with the actual power consumption has been obtained by the data tree regressor and the results are as below:

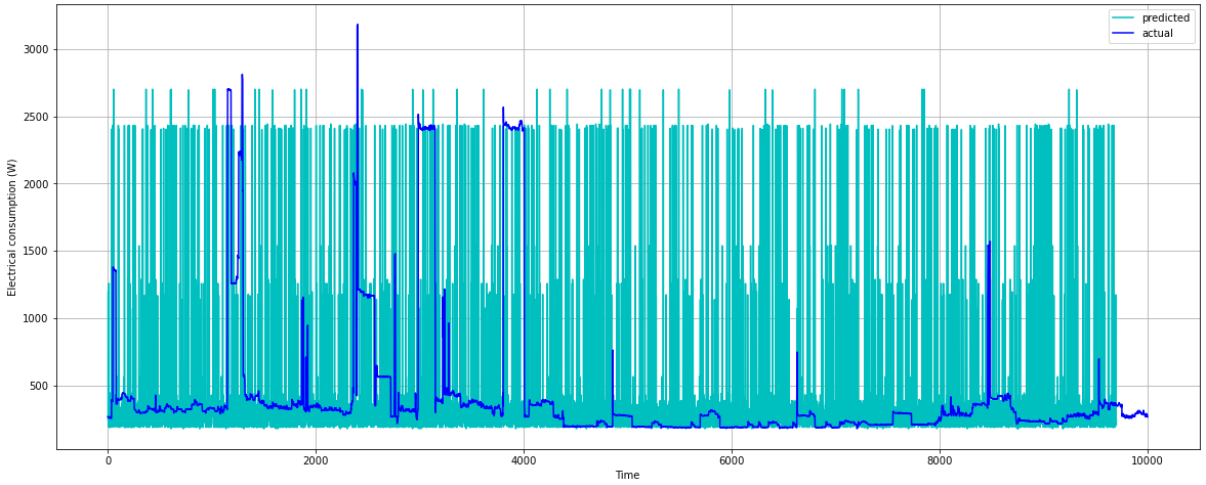


Figure 8 : Estimated and Actual Power Consumption

The predicted power consumption is definitely not accurate. The model has been able to predict accurately and precisely only at a few points of the year. As with all supervised machine learning, special care should be taken to ensure that the labeled training data is representative of the overall population. If the training data is not representative, the predictive model will be overfit to data that doesn't represent new and unseen data. This will result in inaccurate predictions once the model is deployed. Because regression analysis involves the relationships of features and outcomes, care should be taken to include the right selection of features too.

The relation between the estimated consumption and the features is plotted as below:

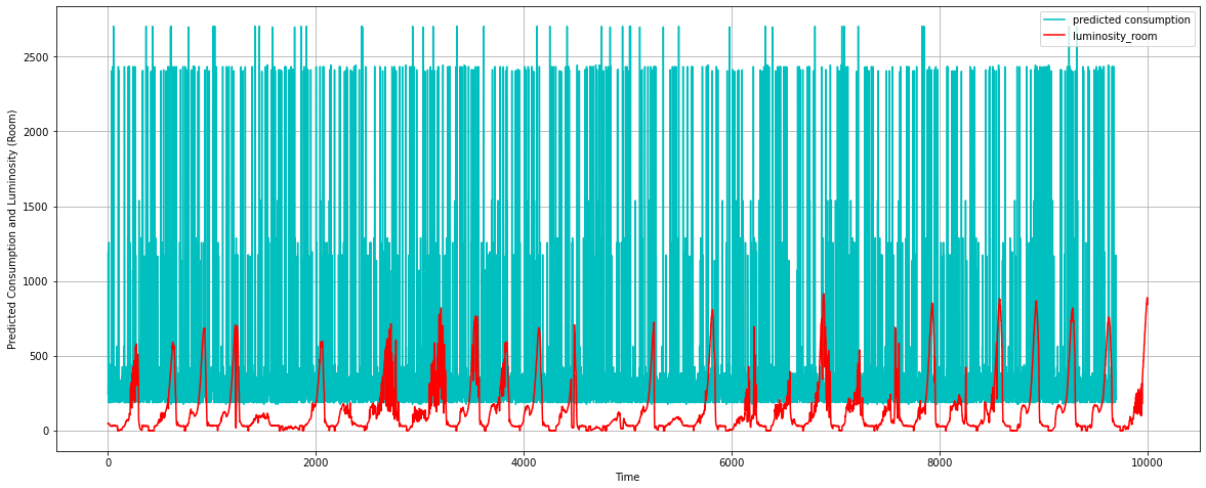


Figure 9 : Estimated Power Consumption & Luminosity (Room)

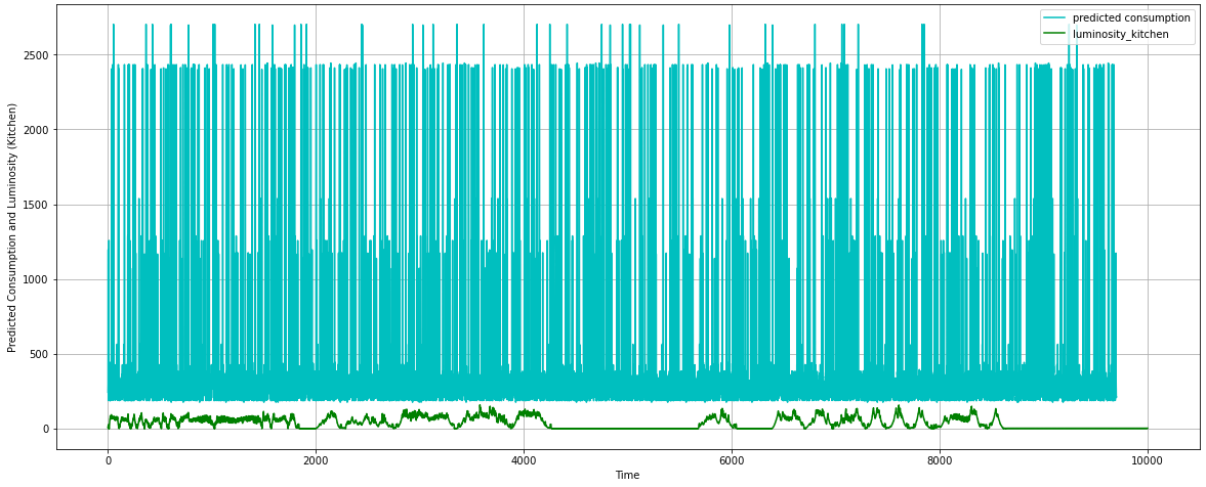


Figure 10 : Estimated Power Consumption & Luminosity (Kitchen)

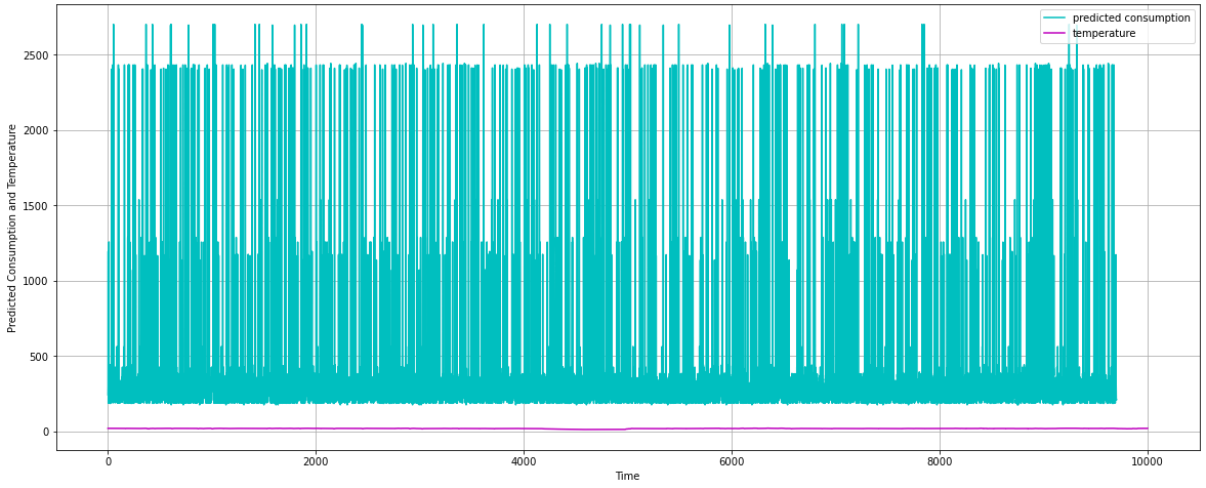


Figure 11 : Estimated Power Consumption & Temperature

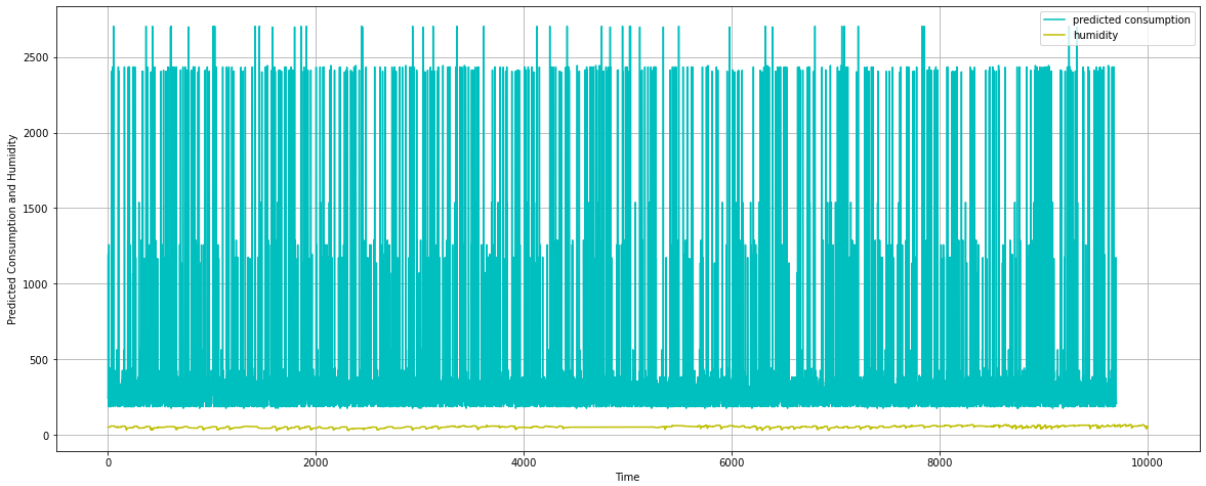


Figure 12 : Estimated Power Consumption & Humidity

The power consumption is low when the luminosity is high like in the day times, the electrical consumption reduces significantly. The power consumption is also low when temperature is high like in summers and it is high when temperature is low like in winters.

4. Conclusion

The relation between all these features and power consumption could be more clear and relatable if the data was resampled in a good way. The data provided has different time steps therefore even after concatenation, the difference could not be reduced enough to have accurate results.

As the relationship between the accuracy and data health is direct, that has led to not impressive results. The labeled training data is supposed to be representative of the overall population and as the training data is not representative of the whole population, the predictive model is overfit to data that doesn't represent new and unseen data. This has resulted in inaccurate predictions along with the correct predictions.

Moreover, regression analysis involves the relationships of features and outcomes, therefore the right selection of features is also very important. If some other features would have been selected for this estimation, results could have been improved in that case. In short, better data sampling, better choice of features and an improved machine learning model would be a better choice to obtain acceptable and useful results for this smart home.

Thank you!