Master's in Electrical Engineering for Smart Grids & Buildings - (M2)

Year of study 2022-2023

Machine Learning & Optimization

Estimation of electrical consumption & analysis

Submitted to: *Prof. Jerome Ferrari & Hassan Issa*

Submitted by:

Albert Granollers

Mikayel Margaryan

January 26, 2023

Contents

List of Figures

1 Introduction

This report investigates the discussion of the electrical consumption with the proper policies to optimize them.

Currently, one of the hot topics in the engineering is the power supply in highly loaded periods like, especially after last geopolitical crisis. According to EU commission, Russia's unjustified military aggression against Ukraine and its weaponisation of gas supplies have provoked an unprecedented energy crisis for the EU. They have caused a sharp rise in energy prices and brought hardship for Europeans. Thus, the EU commission is taking strong action to address this.

In order to overcome unexpected challenges, one of the actions and policies published in September 2022 is about new measures that were implemented to reduce electricity demand and use energy surpluses for the benefit of citizens and industry. As a result, the focus of this study will be on how to use collected data to create opportunities for appropriate policies.

The methodology of our study is presented in the section 2. Then the results and data analysis are discussed in the section 3. The motivation and drivers of choosing current labels and features has been discussed in detail with different scenarios. In section 4, the designed model of Machine Learning(ML) is presented for the estimation. Finally, in section 5 some brief conclusive remarks are given with the future work prespective.

Figure 1: EXPE - Smart House

2 Methodology

For the reason of gathering data, the EXPE-SmartHome project is introduced as a use case. This project grant an access to about 340 measuring points for scientists, accessible in real time through a Grafana portal with Influxdb database. This use case smart home is developed based on Open Source Hardware and Software. So, the analysis was done according to the sensorsed installed in this home.

Figure 2: Different sensors in the EXPE - SMART home

There are measures of:

- Electricity, gas & water consumption of each device
- Temperature, humidity & brightness of each common room
- Opening position of each door & window
- Motion sensors
- Ligth state
- Air analysis of each room
- Outdoor weather conditions

3 Data Analysis

This section aims to deal with the data needed for the estimation of the electrical consumption of the EXPE smart home. The features that we have selected are temperature, humidity and luminosity. This data was acquired from a data base in a raspberry device and saved as csv files. After that, we have read those files and an analysis of the different features have been done. As well as that, we could also obtain the power consumption of the smart-home for 1 month which it is also presented in this section.

- Temperature data set:

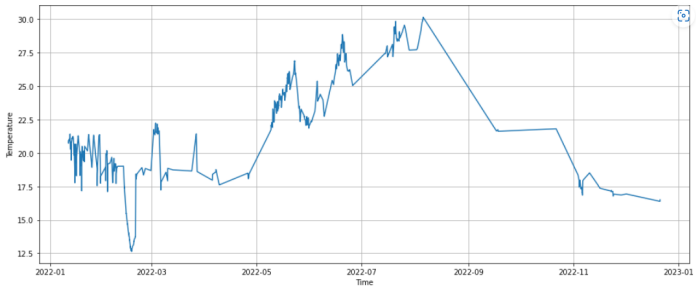

Figure 3: Temperature data set

- Humidity data set:

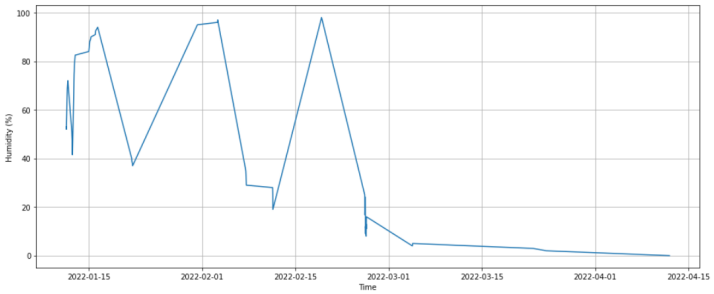

Figure 4: Humidity data set

- Luminosity data set:

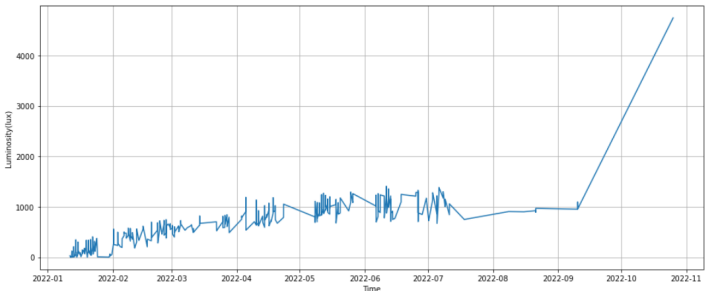

Figure 5: Luminosity data set

- Electrical consumption:

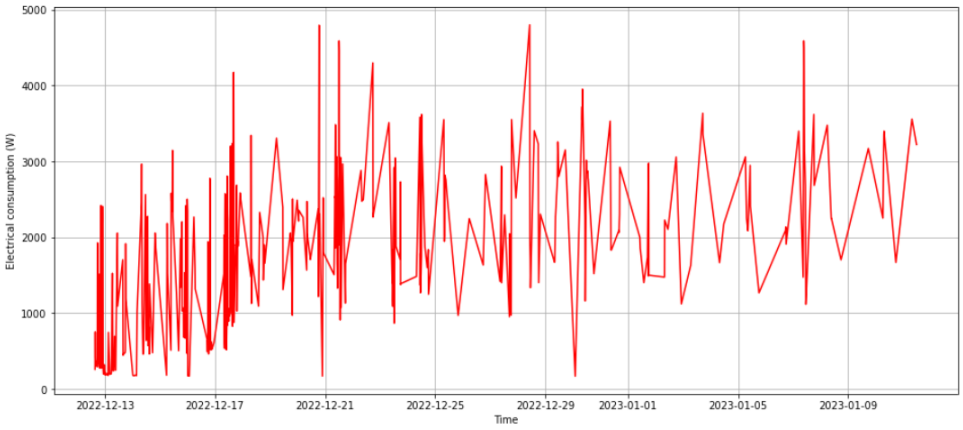

Figure 6: Electrical consumption during 1 month

First of all, regarding the features, it can be seen that the length of the data, is not the same for each feature. For instance, regarding the temperature inside the smart home it lasts for almost a year and the data seems quite coherent taking into account the different months of the year. Nevertheless, if we check the humidity it can be seen that the data has a different size (just from January to the end of April) and in some point the data gets values close to 0, therefore leading to erroneous data with a lot of noise and distortion. Although in the luminosity data, it seems to perform better, some noise and distortion is also presented at the end of the measurement.
Finally in figure 6 it can be seen the electrical consumption of the home during 1 month (the whole year data could not be achieved).

Therefore, the main problem that we are facing here, even though after the "re sampling", is the incoherence in terms of samples for each of the features and the label. So, if we want to implement a machine learning algorithm we should assume that we have the samples either for the features or the label. In order to so, an assumption is presented. Since the consumption is during the month of December, it has been decided to just maintain the data of the beginning of the year for each feature. The idea is that the feature values that correspond to the first months of the year, should be equivalent with the electrical consumption of the home at the end of the year (both cases in winter time).

4 Machine Learning model

Without losing our goal (Estimation of electrical consumption), in this section the machine learning method for predicting the electrical consumption will be presented. Nevertheless, first of all lets analyse the situation. In our case, since we assume that we know all the features values as well as the label (electric consumption) the type of machine learning algorithm would be a supervised method. Concretely we will focus in two kind of supervised methods:

- Decision tree model as a regression
- Random forest model as a regression

Since the output data (label) it is going to be continuous, the ML method for those kind of problems will be base on a regression method as can be seen above. Using the concatenation function of python it can be shown the table with all the features and the corresponding output:

temperature	humidity	luminosity	Power consumption
20.93	53.0	32.0	267.0
20.93	53.0	32.0	267.0
20.84	52.0	25.0	261.0
20.84	52.0	25.0	261.0
20.80	53.0	18.0	259.0
...	...	...	...
16.40	49.0	116.0	272.0
16.40	49.0	131.0	272.0
16.38	50.0	131.0	264.0
16.38	50.0	133.0	264.0
16.31	49.0	133.0	271.0

Figure 7: Features and the label of the smart home

In order to implement the ML algorithm, first of all it must be imported some packages from sklearn.tree. After that, since we need to teach our model, the data will be split in training data and testing data. Normally, the ratio between training data and testing data is about 20 % to 80 % or 30 % to 70%. In the image below, it can be seen the procedure followed in order to develop the decision tree model as a regression:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.metrics import mean_squared_error

X_train, X_test, y_train, y_test = train_test_split(data_features_label[[

regressor = DecisionTreeRegressor(random_state=44)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)

score = regressor.score(X_train, y_train)
print("R-squared:", score)
mse = mean_squared_error(y_test, y_pred)
print("MSE: ", mse)
print("RMSE: ", mse**(1/2.0))
```

Figure 8: Decision tree as regression

Finally, the results of predictions using decision tree or random forest are presented. Since we have observed that both results look quite similar, it will only be introduced the results concerning the decision tree.

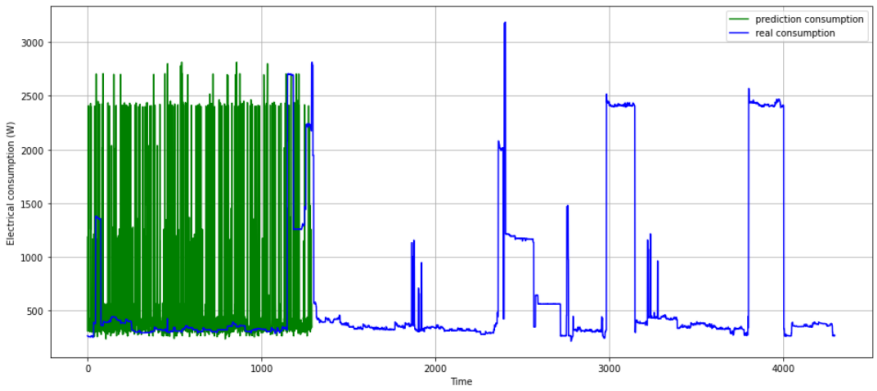

Figure 9: Predicted consumption (green) with 70 % of training data and real consumption (blue)

Analysing the results, it can be seen that the real consumption (in blue) and the prediction done with the DT-regressive model they do not have any similarity between them. It can also be seen such mismatch in the code of figure 9 where it is introduced the mean square error. Such value should be around 0, however in our case, it is a huge number which also shows the problem of our model. Some of the reasons that can explain such difference are introduced below:

- **Lack of enough data**: As it was discussed in section 3, we had to suppress a big amount of data, due to the incoherence of samples between the power consumption and the other features in order to be able to perform a prediction with the model. Therefore, we can conclude that suppressing data is not going to be good for our estimation, and we realise how important is it the management, quality and acquisition of data before starting implementing the machine learning algorithm.

- **Irrelevance of features**: Although, the features that we have selected can be directly or indirectly related with the electrical consumption of the house, it could be that they are not the main important ones. For instance, it is known, that for low and high temperatures, most probably the electrical consumption will increase due to HVAC systems, nevertheless when it comes about the humidity it is harder to see a direct relation between the increment or decrement of electrical consumption. As well as that, there are some relevant features which we got data and csv files, like motion detection in the house, which have not been analysed since we could not find a algorithm that links continuous and discrete data at the same time.

- **Non-representative training data**: Finally, another point that could lead to perform badly, is when the percentage of training data is not representative. For instance, one case could be that we have large training data but it does not represent what the model will see in the output.

5 Conclusion

For this project of 8 h lab, it can be said that the main goals have been developed and implemented. First of all, a work of data acquisition was performed using raspberry devices. From there, different data features were extracted and saved as csv files. After that, such data was imported and plotted in python environment and analysed. Finally, a machine learning model was built and a prediction of electrical consumption has been performed.
Regarding the results obtained, it can be said that they are not conclusive at all, which in a major part is due to several reasons. We have realised how important it is to work with huge amounts of data in order to get reasonable predictions. As well as that, the selection of features before gathering data and implementing ML models it is crucial. Besides that, it is also quite relevant that the sensors measuring different parameters are in good conditions and that the data obtained is coherent according to the field of measurement.