Machine learning and Optimization

Mini-project

Estimation of Energy Consumption in a Smart-House via meteorological variables

Authors:
Germán Alejandro Niño Pallares
José Ricardo Cabral Vianna

Grenoble
2023

1. Introduction

Among the perks of disposing of several sensors in a smart house lies the possibility of analyzing how the different sources of data can be used to estimate different parameters of interest. In the past this study would be carried out with different outlooks but the most predominant would remain being an empirical approach which would need a good amount of experience to make it work. Nowadays we possess larger amounts of data and tools that may allow us to find correlations between data that are not to follow a deterministic nature but are easily modifiable in order to find the best results. In this scope, the present report verses on how to use meteorological data in order to estimate energy consumption for a smart house.

2. Data sources

The Smart House in question disposes several sensors throughout the house and thus the first step in this project is to determine which variables that we have measures of we can use to estimate another variable [labels] that we can also measure [feature]. In this case we are interested in predicting the energy consumption using meteorological data, so we begin by selecting our data sources.
First we inspected the available sensors imprinted in figure here below:

Figure 1. Sensors available for data extraction

From there we determined that we were interested in the sensors located outside and we hypothesized that under colder weather the energy use would be higher, that on gray days the electricity demand is higher and so on. The sensors selected were:

- Total Energy Consumption [feature]
- Humidity
- Lux
- Rainfall
- Temperature
- UV
- Wind

Here we found one of the more critical parts of our project. The data is not even and a strict regularization process must be carried over. For once, the periodicity of our data is very different so we underwent a regularization process. In this instance we observed that energy consumption data was much more frequent than the other parameters and furthermore we had shortcomings in acquiring the data as we were only capable to retrieve 2 months of energy consumption at a time whereas we could retrieve up to 1 year of data for the labels. This incompatibility made us use a resampling process using interpolation of the mean value for the labels and using the total energy consumption for the feature (integration over a time lapse). Furthermore, as data were very diverse in their values and in some cases, constricted to the natural phenomena they represent it was necessary to do a normalization of the datasets using their mean and standard deviation.

3. Exploratory Analysis

We began by doing a scatterplot of the Energy Consumption as a function of Temperature. The results yielded that for a large range of temperatures several energy consumptions were probable, this is a first indication that our model might not be well founded.

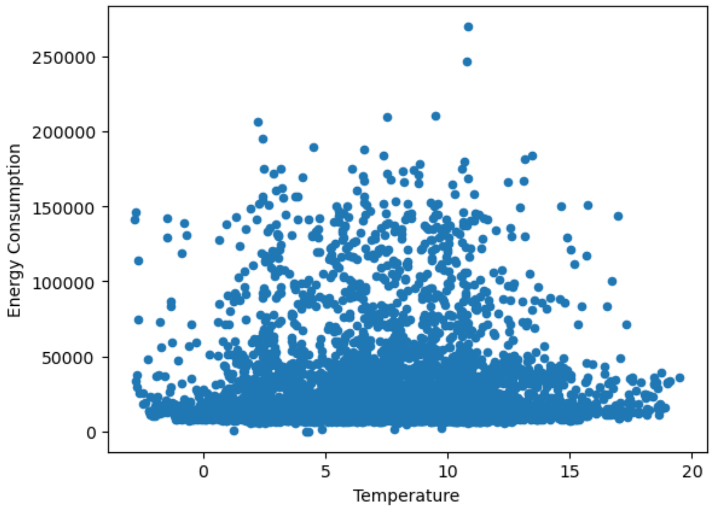

Figure 2. Scatter plot energy consumption vs Temperature

We proceeded to perform a pairplot in order to visualize how our labels and features are interconnected. Here we realized that Lux and UV sensors were complementary as they follow the same pattern and thus for a linear regression model we would find they are redundant as they carry similar information.

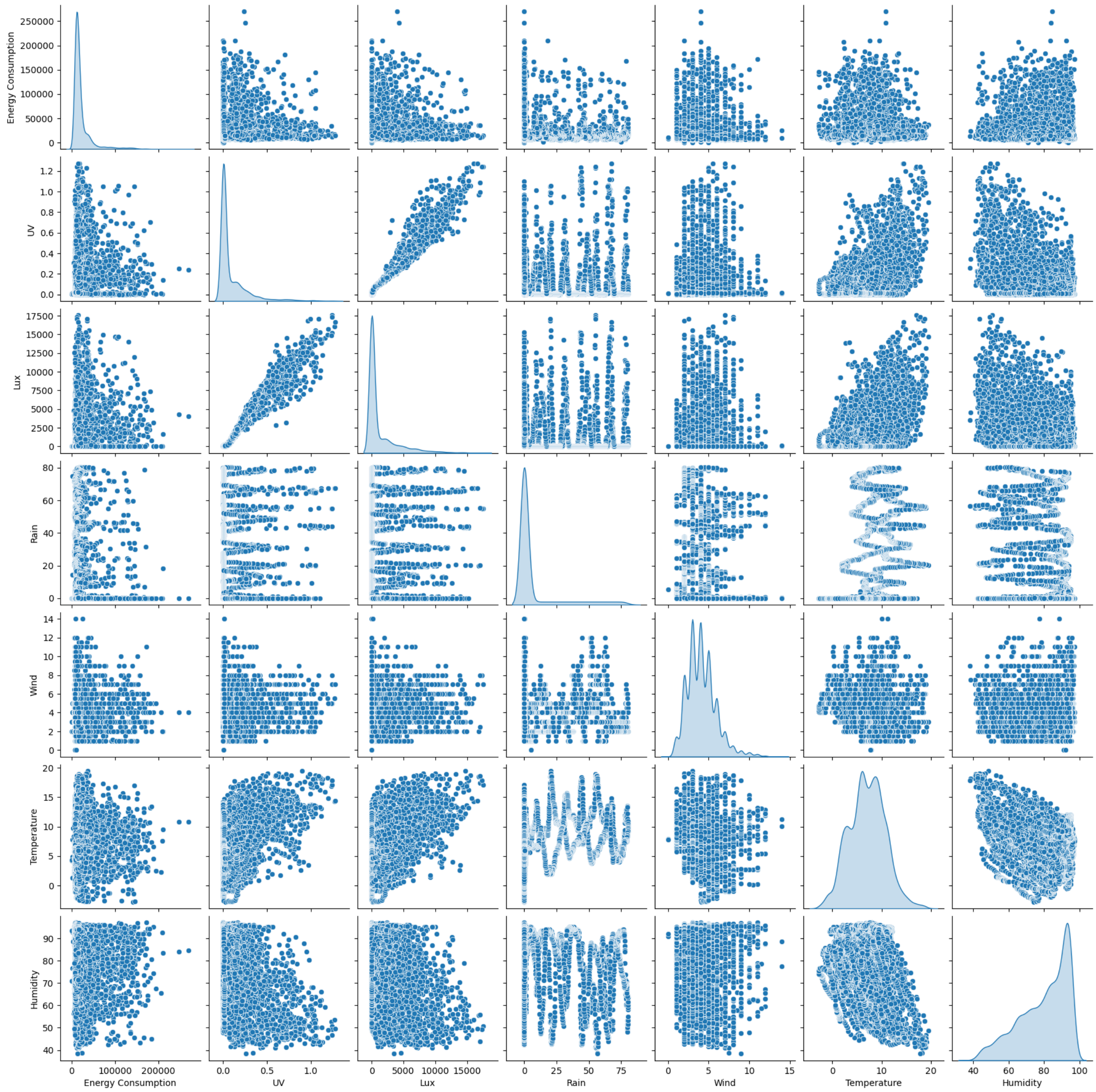

Figure 3. Pairplot between extracted data using seaborn package

From left to right and from up to down(Energy Consumption, UV, Lux, Rain, Wind, Temperature, Humidity)

4. Models Used

After the analysis, we decided to firstly implement a neural network in order to try to predict the energy consumption of the household. The data was splitted in such a way that 80% was reserved for training, 10% for testing and 10% for validation. We used the *Keras* library to generate our regressor, as the problem's expected output is a continuous value, not a category. The model had 4 layers: the first one for normalization of features; the second one has ReLu as activation function with 64 neurons; the third one also has ReLu as activation function, but with 32 neurons; the output layer has only one neuron and has the linear activation function. In the compilation step, we decided to use the Adam optimizer and two loss functions were tested, the mean absolute error and the mean squared error. We could not achieve satisfactory results, as the mean error was always too far apart from the expected.

As we could not achieve good results with the neural network, we decided to try other methods, such as Decision Tree Regressor and a Linear Regressor.

The Decision Tree is able to minimize the training error to a very low value, but when considering the test and validation data, the error is similar to the one found in the neural network. The same happens with the Linear Regressor, as its accuracy could not go higher than 2%.

5. Discussion of Results

The results of the modeling phase yielded that for our case, there was not good enough correlation between the features in order to predict the energy consumption. This might be due to the bad tuning of the models or the data might not have been enough.

Literature review englobes successful models that have tried to answer the same question under different conditions. For example Quiao et al. (2022) used the data from a meteorological station over the course of 3 months in 30 minute intervals using some of the data we used as samples. In this case the research yielded that Decision Tree, Support Vector Machine, Random Forest and Voting Regressor methods were capable of performing this model although SVM was not accurate. On the other hand, the case studied Prabakar et al. (2018) used open data form the netherlands climate service reflected that energy consumption patterns and meteorological data vary depending on geographical location which points to the fact that models are subjected to a specific domain, moreover one practical advice from this paper is to not shuffle all data into one but to rather separate correlated data, for example taking as a training set one specific months measurements and then testing it with a completely different set from another month. This approach is recommended by the two previous papers and its objective is not to mix different tendencies in the modeling phase as it is true that meteorology follows patterns in time.

6. Conclusion

The results obtained compared to the different models available showed that the data was flawed. It is advised to get a hand on the data before proceeding with model implementation. This project showed that even though the premise at first made sense, that the weather might influence the energy consumption of a household, the reality may not match these expectations. Also, the project also showed the importance of working with a good dataset and that correlation is really important in for modeling process.

Bibliography

Qiao, Qingyao; Yunusa-Kaltungo, Akilu; Edwards, Rodger (2020). *[IEEE 2020 IEEE PES/IAS PowerAfrica - Nairobi, Kenya (2020.8.25-2020.8.28)] 2020 IEEE PES/IAS PowerAfrica - Predicting building energy consumption based on meteorological data. , (), 1–5.* doi:10.1109/powerafrica49420.2020.9219909

A. Prabakar, L. Wu, L. Zwanepol, N. Van Velzen and D. Djairam, "Applying Machine Learning to Study the Relationship Between Electricity Consumption and Weather Variables Using Open Data," 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sarajevo, Bosnia and Herzegovina, 2018, pp. 1-6, doi: 10.1109/ISGTEurope.2018.8571430.