Electrical consumption depending on indoor temperature, humidity and CO2.

Julián Felipe Giraldo Cruz
julian-felipe.giraldo-cruz@grenoble-inp.org

Hugo Ronaldo Paipa Chaparro
hugo.paipa-chaparro@grenoble-inp.org

Presented to: Jerome Ferrari
jerome.ferrari@g2elab.grenoble-inp.fr

***Abstract*— Energy consumption in residential buildings depends for now in human behaviors, weather conditions, etc. During this report we want to predict the energy consumption in a residential building depending on the quantity of CO2, temperature and humidity using machine learning methods. Specifically in this study we used supervised learning with linear regression and Decision tree classifier.**

***Keywords*— machine learning, energy consumption, prediction and weather conditions.**

I. Introduction

The prediction of energy consumption in any type of building is important towards decreasing energy usage[1]. For prediction we can use libraries such as scikit learn, Pytorch, Tensor flow, that content different algorithms of classification, regression and clustering in machine learning. In this report we will analyze the energy consumption in a Smart Home, using variables of temperature, humidity and quantity of CO2 through lineal regression and Decission tree classifier in scikit learn.

II. Analysis and Results

For the analysis of our study, we started extracting the data of temperature, humidity and quantity of CO2 in one month from the raspberry for the timeline of 13th December 2022 to 10th January 2023. After cleaning wrong values in our data, we carried on plotting our data.

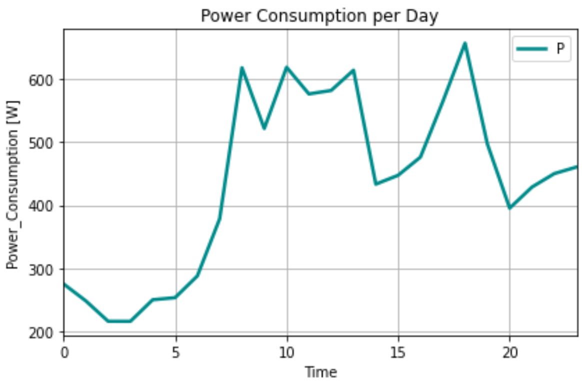

Figure 1. Average of power consumption during a day.

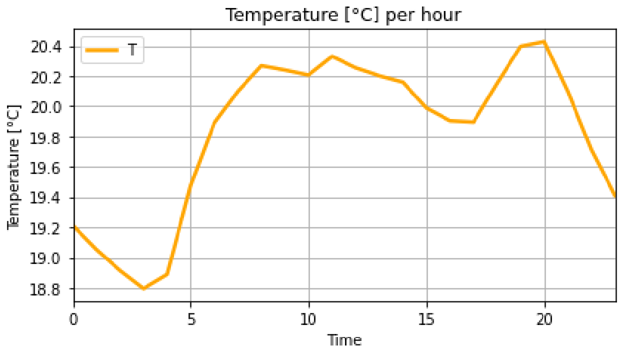

Figure 2. Average of temperature during a day.

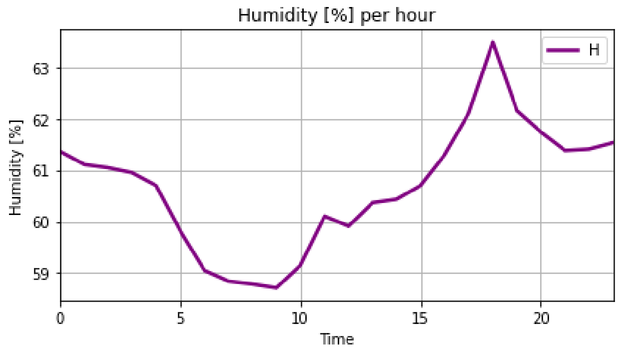

Figure 3. Average % of humidity during the whole month.

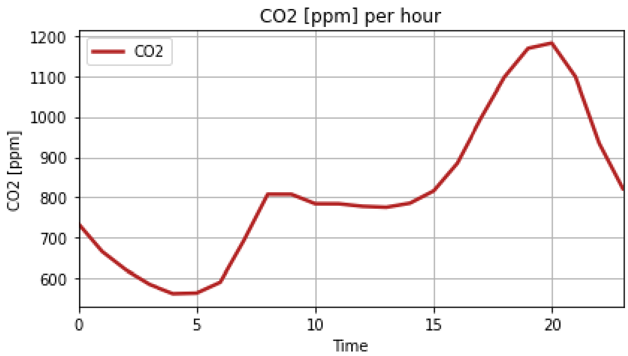

Figure 4. Average of quantity of CO2 during a day.

---

[1] Olu-Ajayi R, Alaka H, Sulaimon I, Sunmola F and Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. 2022

Linear Regression

For the prediction of energy consumption using linear regression, we follow the library of scikit learn preparing our data following the structure on the website. Thus, we wanted to understand whether the linear regression can be an accurate model to predict the power consumption with the three main features CO2, temperature and humidity.
As usual, we split our data into the training and testing ones, choosing the test size as 29%, because is a considerable portion of the data and most of it must be used to train the model. It must be stated that we directly thought into use the regression model, because the label is the power consumption, and it is continuous. On the other hand, if our label was discreet data, we could have used any classification model like decision three classifier.

The linear regression model is presented in the figure below:

```
# group_Day_E (y)
y = group_Day_E
# y = a_0 + a_1*x_1 + a_2 *x_2 + a_3*x_3
group_Day_T = Data[["Day", "Temperature [°C]"]].groupby('Day').mean()  # x_1
group_Day_CO2 = Data[["Day", "CO2 [ppm]"]].groupby('Day').mean()  # x_2
group_Day_H = Data[["Day", "Humidity [%]"]].groupby('Day').mean()  # x_3

X = np.concatenate((group_Day_T, group_Day_CO2, group_Day_H), axis=-1)
x_train, x_test, y_train, y_test = train_test_split(X, y , test_size= 0.29, random_state=0)
reg = LinearRegression()
# training the linear model
reg.fit(x_train, y_train)
y_pred= reg.predict(x_test)
```

Figure 5. Linear regressor code.

After training our regression model, we decided to check how accurate were the results that we were obtaining, by using the metrics of mean squared error , because it is the metric for a continuous label data , we have gotten as mean squared error of 1913,21 , which is super high, while the objective mean squared error could have been some number close to 0, but as our real error is a large value it means that the linear regression is not recommended to be used for the forecasting of power consumption based on CO2 levels, Temperature and Humidity as features.

We evaluate our model with the features data for the first date time as follows:

ds	Energy_Consumption [W]	Temperature [°C]	Humidity [%]	CO2 [ppm]	Year	Month	Day	Hour
2022-12-13 00:30:00	199.029851	19.420000	50.240000	621.250000	2022	12	13	0

Table 1. Real testing data example.

t = 19.13°C

h = 61.6 5%

co2 = 641.33 ppm

```
reg.predict(np.array([t, co2, h]).reshape(1, -1))
```

```
array([[338.68834721]])
```

As seen in Table 1, the real energy consumption at that specific time was 199[W], meanwhile our linear regression model gave us a power consumption of 338.68[W]. By comparing the real and the forecasted power consumption, it can be noted the big difference in the values.

Classification with Decision with Tree Classifier

We transform our problem in a problem of classification, classifying the energy consumption in tree levels: Low (smaller than 600W), Medium (Between 600W and 800W) and High Consumption (greater than 1000W). We shall find the best configuration of Tree Decision Classifier for our study, finding the best percentage of data for fit our data, best feature importance, and number of depths.

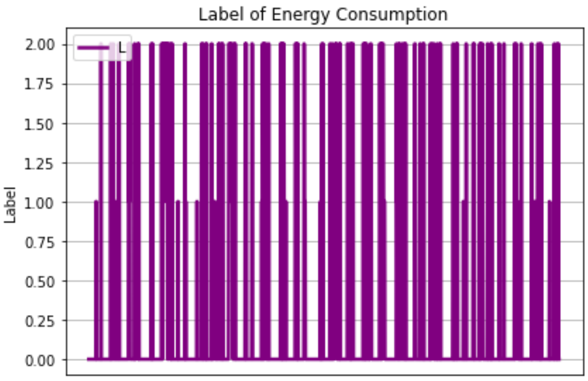

Figure 5. Labeled values for Energy Consumption.

Varying the size of test data and training we find the best portion for our study using metrics of scikit learn.

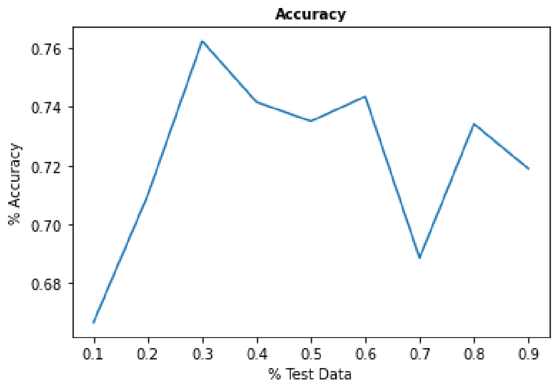

Figure 6. Labeled values for Energy Consumption.

According to Figure 6 with a 70% for training data and 30% for testing our data we find the best accuracy for our model.

Same in Figure 7 with F-score, showing the precision of our model.

According to clf.feature_importances_ the most important feature is the Humidity in order to classify our energy consumption.

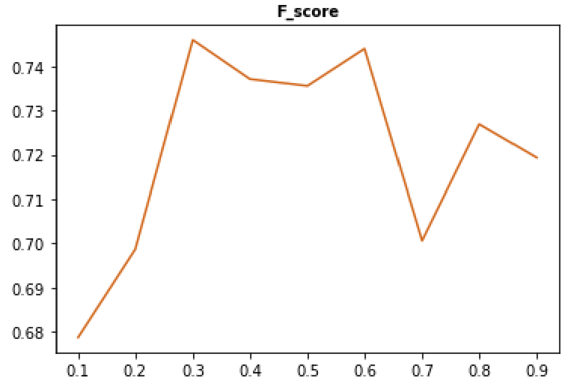

Figure 7. Labeled values for Energy Consumption.

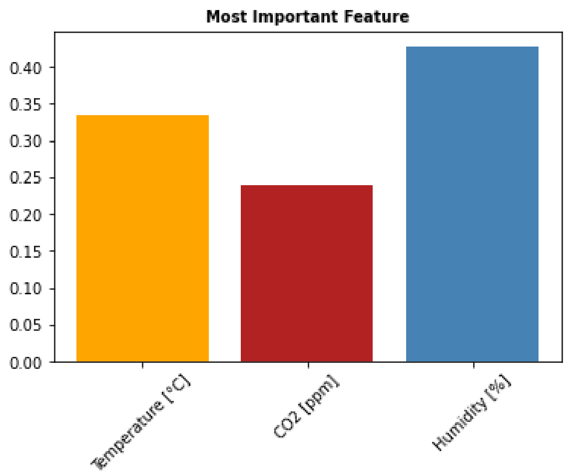

Figure 8. Labeled values for Energy Consumption.

Stablishing our size for training and testing we continue finding the best number of depths.

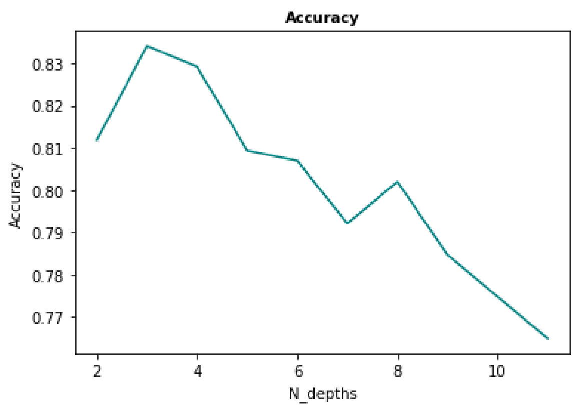

Figure 9. Labeled values for Energy Consumption.

The best number of depths is 3 according to figure 9.

Finally having the number of depths, we evaluate our function:

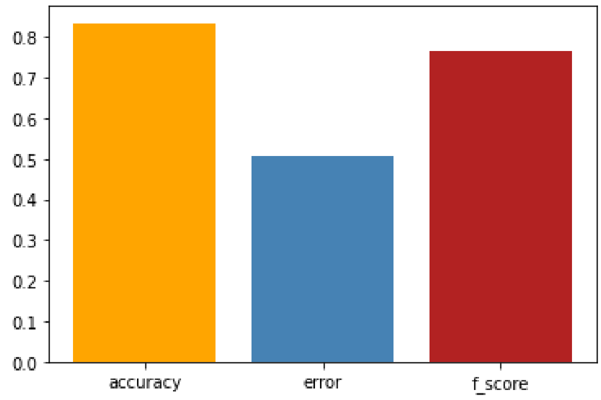

Figure 10. Accuracy, error, and f_score with Decision Tree Classifier.

Evaluating our model with the following values:

t = 19.13°C

h = 61.6 5%

co2 = 641.33 ppm

```
clf = tree.DecisionTreeClassifier(max_depth=3)
clf = clf.fit(x_train, y_train)
clf.predict([[t,h,co2]])
```

```
array([0.])
```

Our model give us low comsuption while in the real data is high consumption. Our model is not accurate and it can be seen in the error, accurancy and f-score.

The error 50% is super high.

Energy_Consumption [W]	Temperature [°C]	Humidity [%]	CO2 [ppm]
1111.534591	19.13	61.65	641.333333

Table 2. Testing Case.

Conclusion

The linear regression result, let us know that the relationship between the power consumption and the three different features is not linear so it could have been better to use another regression model like the random forest regressor, which might be more accurate for this forecasting. As well, after attacking this problem as a classification, dividing into three groups the label of power consumption, we also found a not accurate model, this is because the power consumption might not be really affected by the presence of CO2, high or low temperatures and humidity at home, or also because maybe our data was not enough to run a more detailed and accurate model.

References

[1] u-Ajayi R, Alaka H, Sulaimon I, Sunmola F and Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. 2022.