# Smart Home-Measuring the correlation between Energy Consumption and Weather

## Machine Learning & Optimization

## (29-01-2023)

**Submitted to Professor:**

**Jérôme Ferrari**

**Done by:**

**Manu Rajesh**
**Omar Khalil**
**Hussein Khalife**

# Contents

**Introduction**......................................................................................................................2

**Methodology**......................................................................................................................2

**Results**......................................................................................................................4

**Conclusion** ......................................................................................................................6

# 1. Introduction

Forecasting power consumption is a critical aspect of energy management, as it helps in ensuring a stable and reliable electricity supply. Weather conditions play a significant role in determining power consumption, with factors such as temperature, humidity, and luminosity affecting the demand for electricity. For example, cold weather entices people to utilize more electricity for heating. In this study, we aim to explore the relationship between weather data and power consumption by developing a predictive model based on multivariate linear Regression. By analyzing the correlations between weather variables and power usage, we aim to identify the key factors that drive power consumption and gain insights into the behavior of the system. The model will be developed using a dataset of weather data and power consumption readings, and will employ machine learning techniques in order to establish the relationship between the two features and labels. The goal of this study is to provide a more accurate and robust method for predicting power consumption, which can help smart homes and energy providers to better manage energy supply and demand.

# 2. Methodology

In this study, the Regressor will be used to predict power consumption based on weather data as features (Rain, Wind, Temperature and light) and label (electrical consumption), by establishing a relationship between the input weather variables and the target power consumption variable. First, we needed to clean the data and assign reference time frame that they all share to get sensible results. Then, the algorithm will analyze the input variables, such as temperature, humidity, and luminosity, and identify the key relationships between these variables and power consumption. The data will be split into training and testing data (Optimally 80% for training and 20% for testing). The results of the analysis will be represented as a tree-based model, which can be used to make predictions about future power consumption based on the current weather conditions. The performance of the Regressor will be evaluated using statistical metrics, such as accuracy and mean squared error, to determine its suitability for this application. We have fitted the features into the same time frame after concatenation.

First after cleaning the preparing the data, we described the data as follows:
The below table of Figure 1 displays all our four features and their values, However for the simplicity of calculation, study and plotting we need to put them in the same range using “MinMaxScaler”. The scaled data is displayed in figure 2.

| time | rain | wind | light | temp |
|---|---|---|---|---|
| **2022-01-11 16:00:00** | 0.000603 | 2.800000 | 90.870000 | 6.532500 |
| **2022-01-11 16:10:00** | 0.000723 | 3.000000 | 58.218889 | 6.486667 |
| **2022-01-11 16:20:00** | 0.000844 | 3.071429 | 27.341000 | 6.440833 |
| **2022-01-11 16:30:00** | 0.000964 | 3.142857 | 7.587778 | 6.395000 |
| **2022-01-11 16:40:00** | 0.001085 | 3.214286 | 4.080000 | 6.237500 |
| **...** | ... | ... | ... | ... |
| **2022-05-27 14:40:00** | 1.566996 | 18.667523 | 33264.431111 | 26.441250 |
| **2022-05-27 14:50:00** | 1.567655 | 18.667476 | 31142.308000 | 26.542500 |
| **2022-05-27 15:00:00** | 1.568315 | 18.667429 | 29344.590000 | 26.643750 |
| **2022-05-27 15:10:00** | 1.568975 | 18.667383 | 26844.411000 | 26.745000 |
| **2022-05-27 15:20:00** | 1.569635 | 18.667336 | 24312.642000 | 26.846250 |

**Figure 1: Features and labels description**

| time | rain | wind | light | temp |
|---|---|---|---|---|
| **2022-01-11 16:00:00** | 0.000000 | 0.147368 | 0.001383 | 0.195364 |
| **2022-01-11 16:10:00** | 0.000014 | 0.157895 | 0.000862 | 0.193862 |
| **2022-01-11 16:20:00** | 0.000028 | 0.161654 | 0.000371 | 0.192360 |
| **2022-01-11 16:30:00** | 0.000043 | 0.165414 | 0.000056 | 0.190858 |
| **2022-01-11 16:40:00** | 0.000057 | 0.169173 | 0.000000 | 0.185698 |
| **...** | ... | ... | ... | ... |
| **2022-05-27 14:40:00** | 0.184642 | 0.982501 | 0.529846 | 0.847682 |
| **2022-05-27 14:50:00** | 0.184720 | 0.982499 | 0.496040 | 0.850999 |
| **2022-05-27 15:00:00** | 0.184798 | 0.982496 | 0.467402 | 0.854317 |
| **2022-05-27 15:10:00** | 0.184876 | 0.982494 | 0.427573 | 0.857634 |
| **2022-05-27 15:20:00** | 0.184953 | 0.982491 | 0.387241 | 0.860952 |

**Figure 2: Scaled feature data**

We polished the data and plotted the features on the same graph, and we obtained this plot:

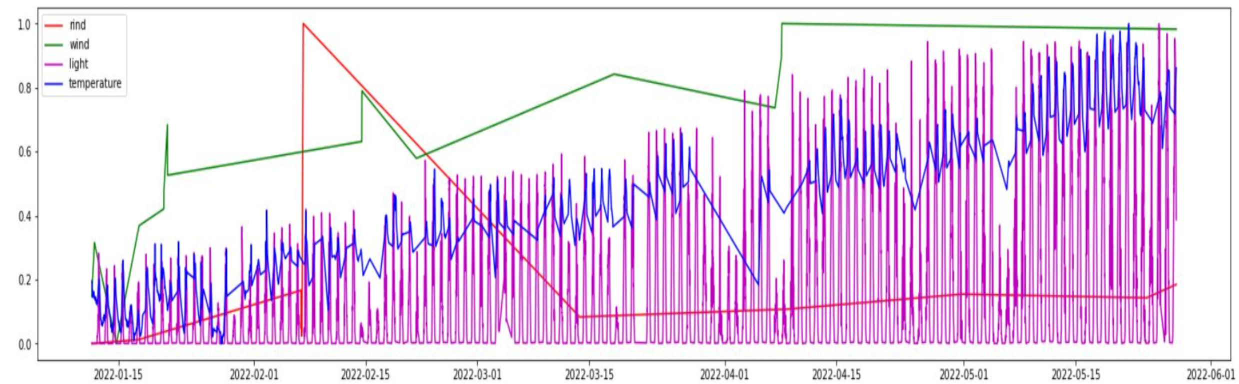


**Figure 3: Graphical Representation of Features (Rain, Wind, Temp, light)**

Then we Split the data, implemented the Regression model.

# 3. Results

After implementing our model, we've achieved the following results:
First the following graph visualizes the high compatibility of the tested results with actual results.

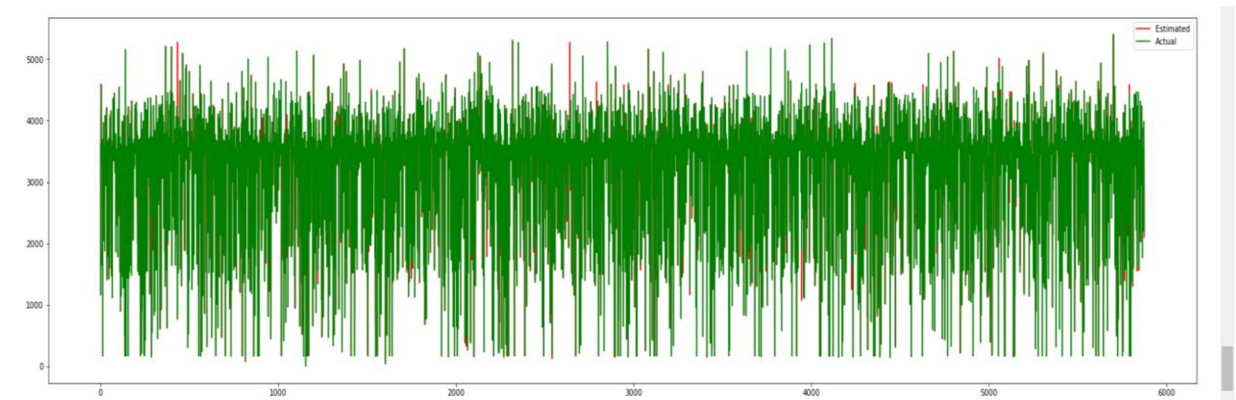


**Figure 4: Predicted vs Actual Results**

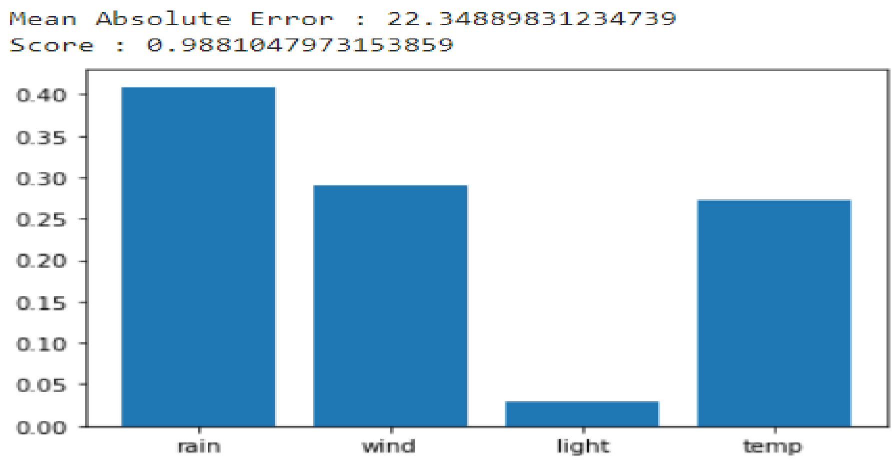


**Figure 5: Most important Features and Evaluation of model through F-score, MAE**

The Decision Tree is displayed in figure 5 below:

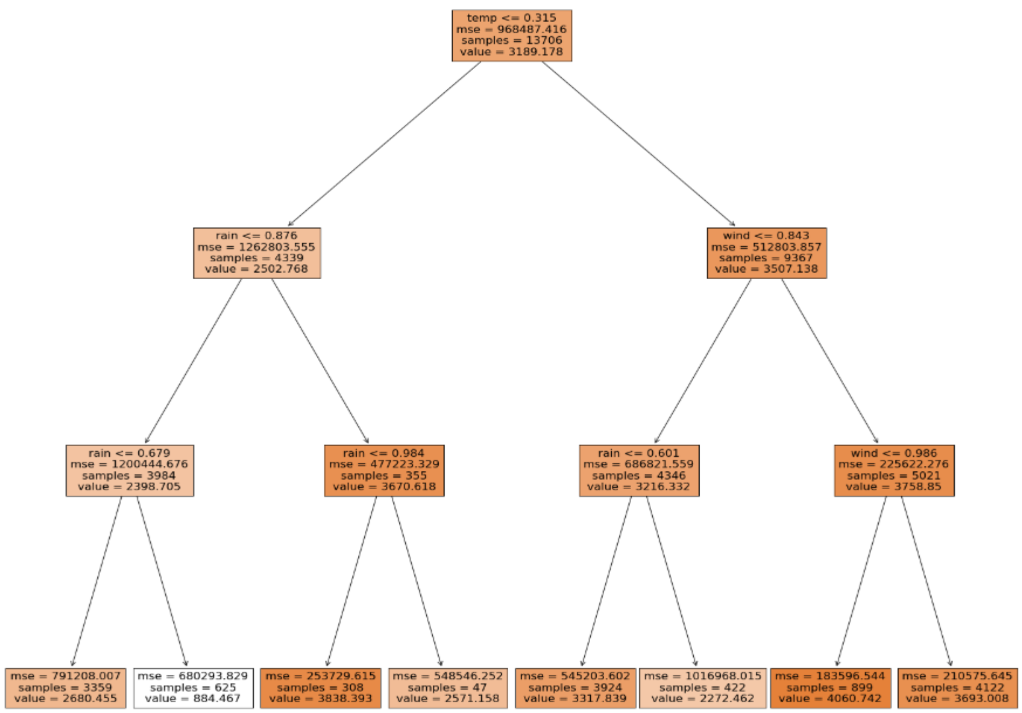


**Figure 6: Decision Tree based on the predicted model**

The results of the regression showed that the most important feature with the lowest impurity was the Rain, followed by the wind and temperature (which were very close) and the light comes last with the lowest in terms of importance.

We also implemented the Random Forest Regressor to enhance more our accuracy and the below figure shows our actual vs predicted consumption as well as the F-score and mean absolute error.

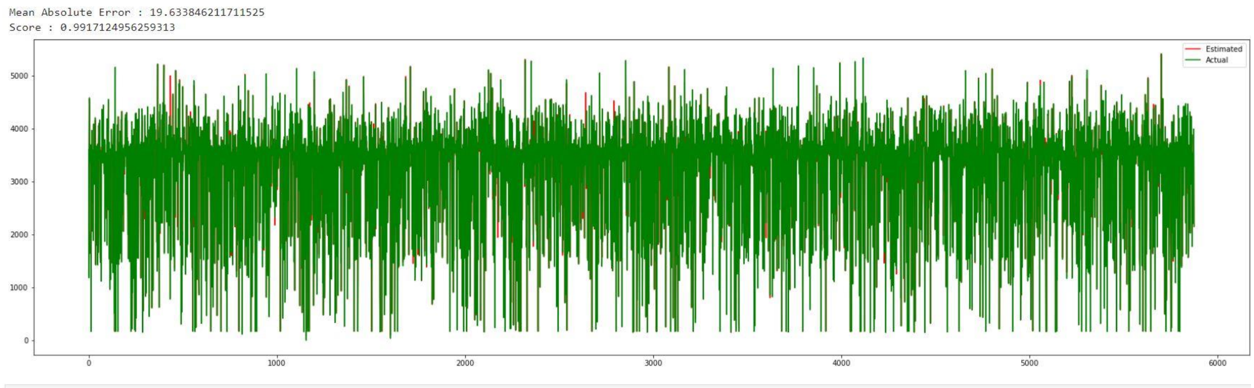


**Figure 7: Predicted vs Actual Results using Random Forest Regressor**

We can test the model by giving random data to get the predicted Electrical consumption.
For example:

```
# Testing the model with random sample
# rain = 1.56, wind = 18.65, light = 26000, temperature = 26.8
test_sample = scaler.transform(np.array([1.56, 18.65, 26000, 26.8]).reshape(1, -1))
rfr_model.predict(test_sample)
# Output is the predicted Electrical Consumption
```

```
array([3790.15936255])
```

**Figure 8: Testing our model with random sample**

In the above figure we can see random data has been put to the model to predict consumption.

# 4. Conclusion

Our study has demonstrated that Rain is the most significant factor in predicting power consumption based on weather data. This result highlights the importance of rain in determining energy demand and highlights the need for accurate temperature forecasts to improve energy management. People are common to have cold adapting behavior and witness an increase in load demand when it rains. The use of Regression technique (Random Forest) was found to be an effective approach for predicting power consumption, demonstrating good accuracy and mean squared error results.