Power Consumption Prediction

1. Exploration

The purpose of this analysis is to predict the power consumption based on the weather condition.

The dataset includes the date and time of the measurement, power consumption of appliances, power consumption of lights, the values of 9 indoor temperature sensors, 9 indoor humidity sensors, outdoor temperature, temperature of dew point, humidity, wind speed, and relative velocity.

Analysing the range of measured temperatures show that the values of sensor 6 (T6) is quite like the outdoor temperature, which means that sensor 6 has been located at the outdoor area (e.g. alfresco). Therefore, when analysing the indoor temperature or humility, sensor 6's data should be excluded. The rest of indoor sensors have almost similar values (Figure 1).

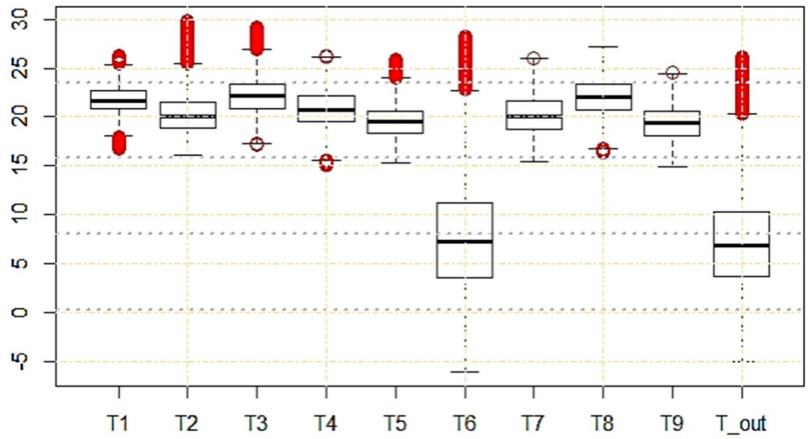

Figure 1. Indoor and outdoor temperature range

Considering the correlation between the measured parameters, their distribution pattern, and range of their values, I selected the key features as the temperature difference between indoor and outdoor, humidity difference, outdoor wind speed, relative velocity, air pressure, and visibility (Figure 2).

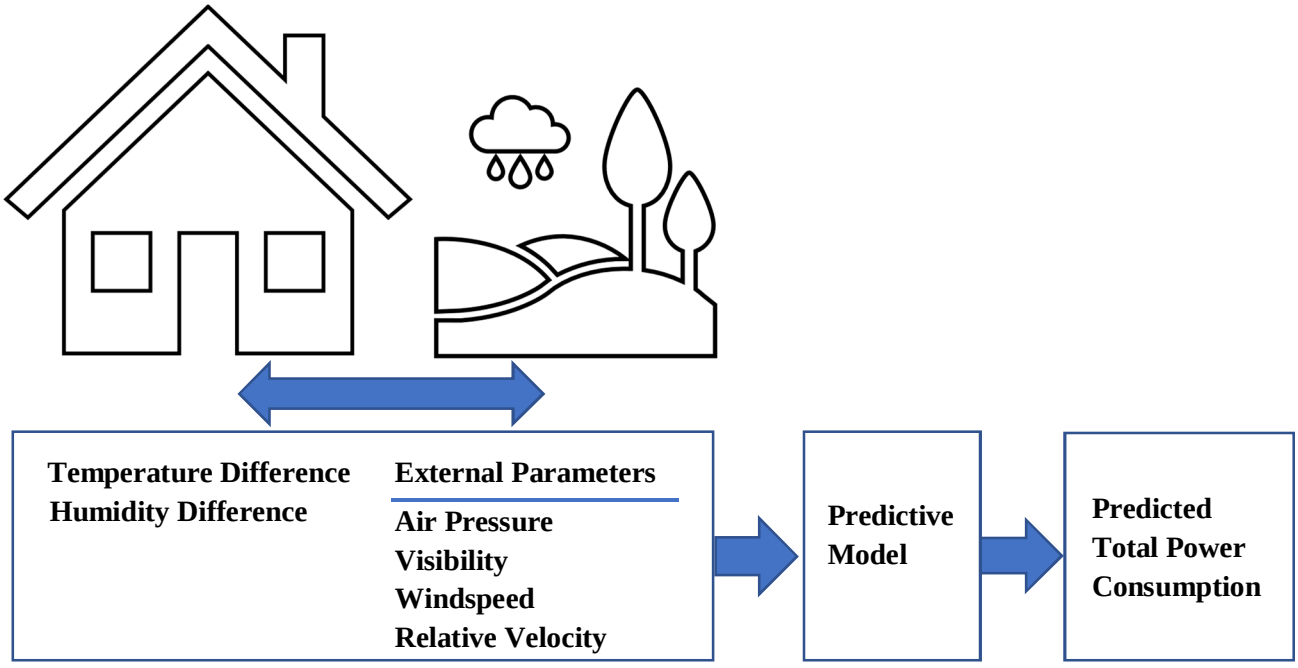

Figure 2. Overview of the project and key parameters

2. Modelling

I have used R for data analysis and presented my coding in R-Markdown.

For data pre-processing, at first, I managed the missing values by predicting their values. Next, I went through the outliers to confirm if there is any data as the outlier available. In order to make the model less complex and more practical, instead of indoor and outdoor temperatures and humidity, I have considered the temperature and humidity difference. The last step before modelling was normalising the values of inputs.

2.1. The selected modelling options for the predictions

To choose the best predictor model, I developed 11 regression models using various machine learning techniques. The techniques are Arima, random forest, multiple regression modelling, neural network, gradient boosting machine, generalized linear net, generalised additive model, rpart, party, and earth. After comparing their prediction errors, I found generalized linear net, random forest, and Arima as the best. I have also presented the predicted output using Arima and compared the predicted and actual values. The statistics formula to predict power consumption using Arima technique is as follows:

Power consumption (KW) = 13.4 + 0.3 temperature difference – 0.4 humidity difference – 0.8 pressure + 0.2 wind speed – 0.2 visibility – 0.2 relative velocity

2.2. Risks of applying machine learning methods to time series data

In the time series data, at some certain period, the data can vary a lot. For example, during the day time, the power consumption at some hours is higher than the others. To predict the power consumption accurately, the dataset should be subdivided, and different algorithm should present the prediction at different time point.

3. Benefits

1- Identify the infrastructure requirement: Predicting the power consumption helps to understand the electricity demand in one area at different time-frames. Therefore, it can help to recognize if the current electricity transmission method would be enough for the future and when an upgrade is needed.

2- Reduce the operational cost: Predicting the electricity usage helps to predict the demand for the other sources of energy in the future and investigate how energy can be supplied with the least cost.

3- Better customer service: This prediction helps the power companies better allocate enough power in each area. So, customer will not face any situation where the electricity is overloaded or cut off due to lack of resources. It makes the customers happy with the provided service.