

Introduction

Context

- Planete OUI offers green electricity supply with prices adapted to the consumption profiles of its clients.
- The electricity prices are highly variable during time (depending on the market, the global consumption and its client needs).
- The consumption profile of an installation has to be appraised to compute the best estimation of supply tariffs
 - Most sites have a consumption varying strongly with temperature because of electrical heating systems.
 - Except industrial installations because their consumption might be highly related their uses

The 1st objectives are to **analyze thermosensitivity uses and / or other factors (times series...)** affecting consumption

Based on:

- the potential client historical consumption data (i.e its profile)
- electricity prices
- a given percentile used to cover supply costs for different scenarios the 2nd objective of Planète OUI is to compute a distribution of supply costs in €/MWh.

Planète OUI needs

Extrapolation of one or several years consumption data rebuilt from a single year of measured data supplied by the client (in order to be Tcombined with electricity prices) and to get a larger data set of analysis.

Data

- The client's data is often incomplete and spread over a relatively short period.
- 3 datasets:
 - x_train : input data of the training set
 - y_train : output data of the training set
 - x_test : input data of the testing set
- Features:
 - "ID": Data point ID;
 - "timestamp": Complete timestamps with year, month, day and hour, in local time (CET and CEST);
 - "temp1", "temp2", "meannationaltemp": Local and mean national temperatures (°C);
 - "humidity1", "humidity2": Local relative humidities (%);
 - "loc1", "loc2" "locsecondary1", "locsecondary2", "locsecondary3": the coordinates of the studied and secondary sites, in decimal degrees and of the form (latitude, longitude).
 - "consumptionsecondary1", "consumptionsecondary2", "consumptionsecondary3": the consumption data of three secondary sites, whose correlations with studied sites may be of use (kWh). Indeed, the two studied sites and the three secondary sites are used for the same purposes; The output data of the model to be developed takes the following form:
 - "ID": Data point ID;
 - "consumption1", "consumption2": the consumption data of the two studied sites (kWh).

Relative humidities are provided with temperature data because they represent variables of importance for electricity consumption: humidity indeed strongly available for all. It has influences thermal comfort. To replicate operational conditions, some temperature integrated BCM Energy's and humidity data points will be missing. The imputation method must be carefully perimeter in 2017. considered. The "consumptionsecondaryi" variables are the consumption data of several sites with metering power higher than 250 kVA of the Planète OUIs portfolio.

This correlation of the various sites consumptions shall be studied to precise data completion or interpolation. Timestamps may be expressed as month or day of year, day of week and hours, to study the impact of annual, weekly and daily seasonalities. Particular attention should be paid to national holidays processing.

Challenge goal

Prediction of the consumption of two given sites during a year, based on measured data of other profiles.