Hackathon 2025

IE University

Machine Learning for electricity market bidding

3 February 2025

Context

The Spanish electricity market is a benchmark for its high renewable penetration (50%+ in 2024), its integration with the European market and its efficient matching model, which prioritizes energies with lower production costs. In this context, Wind and Solar renewable energy generation are protagonists in this market.

For this challenge, data on energy auctions will be provided, specifically on producers, who are those who submit offers in terms of energy quantity and price. Producers interact with buyers and the regulator so that at the end there is a matching of supply and demand, and the final price is set by the last accepted offer, the one with the highest marginal cost to cover the demand.

The objective of this challenge is to develop Machine Learning models to better understand and forecast the electricity market's capacity. This type of tool helps to provide greater flexibility and efficiency to companies when participating in the market and managing their assets.

We're thrilled to have you participating in this challenge! Good luck!

Data

In order to simplify and narrow down the challenge, the data will contain data from Wind and Solar power generation plants, with a historical supply record of approximately 3 months (March to May 2024). In addition, in order to prioritize and simplify the modeling, the data set contains information on a subset of the total number of plants. Of the total wind and solar power plants, those that contribute the most (80%) to power generation have been filtered out.

1. df_omie_labelled.csv

Register of energy offers by power plants from 29/02/2024 to 01/06/2024. This data file contains the total energy offer per plant and the maximum price at which it is sold at different time intervals.

Variables included: *Codigo, Descripcion, fechaHora, PrecEuro y Energia*.

2. df_omie_blind.csv

Record of total energy bids by power plants 01/06/2024 through 29/06/2024. This file does not include the variable *Codigo*, and will be used in challenge #2 presented below.

Variables included: *fechaHora, PrecEuro y Energia.*

3. filtered_categories.csv

Description of the generation technology for each plant code. Relates the power plant code to the type of technology (Wind, Solar).

Variables included: *Categoria, Codigo.*

4. unit_list.csv

Complementary information on generation units. The use of this file is optional, in case it can help in providing some specific analysis at the time of developing the models.

Variables included: *Codigo, Descripción, Agente, Porcentaje_Propiedad, Tipo_Unidad, Zona/Frontera, Tecnología.*

Challenge

Based on the data, the objective of this challenge is to develop models that allow predicting the behavior of offers, through the forecasting of offers and the identification of plants that participate in the market. In this context, two challenges arise:

Challenge #1: Multiforecasting model for forecasting energy supply by power plant.

Given that the data has information from about 300 plants, using methods that allow forecasting several plants simultaneously will be privileged instead of using manual methods for forecasting individual curves.

The output of the forecasting models must consider:

- Forecast from 1/06/2024 23:00:00:00 to 29/06/2024 23:00:00:00
- Forecast displayed at hourly level

Challenge #2: Associating a plant code to anonymized offers

From the df_omie_blind.csv file, add the field *Codigo* and assign with an analytical code for each row or group of variables that represents a bid for the same forecast period as challenge #1.

The challenge is structured but not limited so that the models identified in Challenge #1 can be used as a reference point for Challenge #2.

Evaluation

The evaluation team will have a hidden file with the actual data to compare the results of each challenge.

Challenge #1:

- Each series will be evaluated, using error metrics between the results obtained (MAPE and MAE). Then an average of both metrics will be calculated for all series.
- The weight of this challenge will account for 70% of the final evaluation grade

Challenge # 2:

The correct assignment of *Code* of the plants will be evaluated from the df_omie_blind data. Using as metric the % hits on centrals.

- The weight of this challenge will account for 30% of the final evaluation grade.

Delivery and Submission

Delivery will consist of two Jupyter notebooks (ipynb files) containing the solutions for each challenge, with one file designated for each challenge.

Additionally, two CSV files with model outputs will be provided for both challenges.

- Challenge #1: file with forecast values (*Codigo, fechaHora y Energia*).
- Challenge #2: file that includes plant codes per record (*Codigo, fechaHora y Energia*)

The format of the name of the files (ipynb and csv) must include the Challenge Number followed by Group Identifier, for example:

- Challenge1_GroupA.ipynb, Challenge2_GroupA.csv.

Files in the appropriate format must be submitted to the following shared folder:
https://www.dropbox.com/request/3RluhTGgFO8Ym0tRlW9S

The submission deadline is Tuesday 18th, at 12:00 PM. Teams that submit after this deadline will not be considered for evaluation.

Other Considerations

Although it is not a strict part of the evaluation, the structure and order provided in the documents might be considered.

The code delivered is expected to run smoothly in local environments, so please be careful not to change the names of input files and column names, among other considerations.