Course Project: Electricity Price Explanation

QRT Challenge context

Raphael Monges, Romain Mondelice, Antoine Bossan

10/01/2024

Abstract

Every day, a multitude of factors impact the price of electricity. This challenge aims to model the electricity price from weather, energy, and commercial data for two European countries - France and Germany. It is not a prediction problem, but an explanatory one, to understand the daily price variation of electricity futures contracts.

1 Challenge Goals

The goal is to learn a model that outputs from these explanatory variables a good estimation for the daily price variation of electricity futures contracts in France and Germany.

2 Data Description

Three CSV file datasets are provided: training inputs X_train, training outputs Y_train, and test inputs X_test. The input data represents the same explanatory variables over two different time periods.

2.1 Input Data

Input datasets comprise 35 columns, including:

- **ID:** Unique row identifier, associated with a day and a country.
- **DAY_ID:** Day identifier.
- **COUNTRY:** Country identifier - DE = Germany, FR = France.
- **Weather measures:** Temperature, Rainfall, Wind, etc.
- **Energy production measures:** Natural gas, Hard coal, Hydro reservoir, etc.

- **Electricity use metrics:** Total electricity consumption, Residual load, Net import/export, etc.

2.2 Output Data

Output datasets are composed of two columns:

- **ID:** Unique row identifier - corresponding to the input identifiers.
- **TARGET:** Daily price variation for futures of 24H electricity baseload.

3 Benchmark Description

The benchmark for this challenge consists of a simple linear regression, with minimal data cleaning. The missing values are filled with zeros, and the country column is dropped.

4 Tasks for the Project

1. Apply all approaches taught in the course and practiced in lab sessions on this dataset. This includes Decision Trees, Bagging, Random Forests, Boosting, Gradient Boosted Trees, and AdaBoost. The goal is to explain the variability in the daily price variation of electricity futures contracts for France and Germany, as represented by the target variable.
2. Compare the performances of these models using various metrics learned in class, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), to assess the explanatory power of each model.
3. Write a concise report format that addresses all the guidelines provided, with a maximum page limit of five.

5 Access to data

On the website of ENS data Challenge, you can find the challenge. Here is the precise link of the challenge where you can find the data:
https://challengedata.ens.fr/challenges/97