COUNTRY : Belgium
REGISTRATION NUMBER :
GROUP REF. :
PREF. SUBJECT :PS2
QUESTION N° : 1.1

Predicting the Belgian system imbalance better with AI

Monitoring and predicting the system imbalance is crucial when trying to maintain the balance in the grid. If we can predict the difference between generation and consumption better, we make the dispatchers' job easier and avoid the added costs that an imbalance gives.

To tackle this problem, Elia group is applying artificial intelligence to predict the imbalance for the current quarter hour, the next quarter hour and the quarter hour after that. We try to predict both an exact value for the imbalance and the most likely category it will fall into.

A lot of potential predictors and models were tested and we would like to share those results with you :

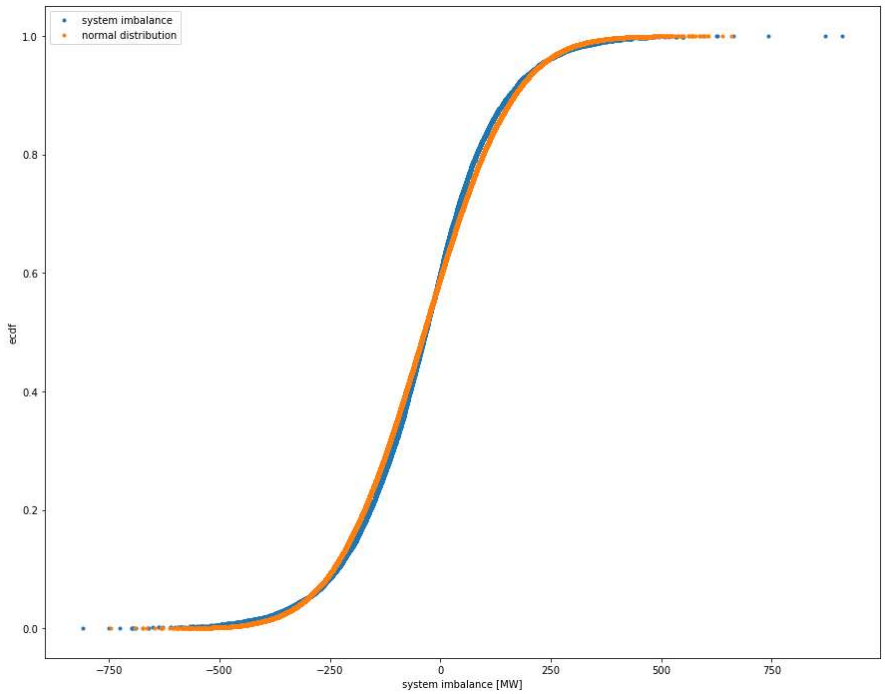
- To find the best predictors we applied techniques like recursive feature extraction, primary component analysis, genetic algorithms, … to hundreds of predictors
- Then there is the model itself. A whole bunch of models were evaluated : Linear regression, ridge regression, lasso regression, random forests, adaboost, gradient boost, … with the linear models coming out on top.
- To predict the categorical value of the system imbalance again the simplest models seem to perform very well. A one vs rest logistic regression classifier gave us the best result after testing a bunch of different models and techniques.

The conclusions are that the system imbalance is somewhat predictable, but there is still a lot of randomness in the data.

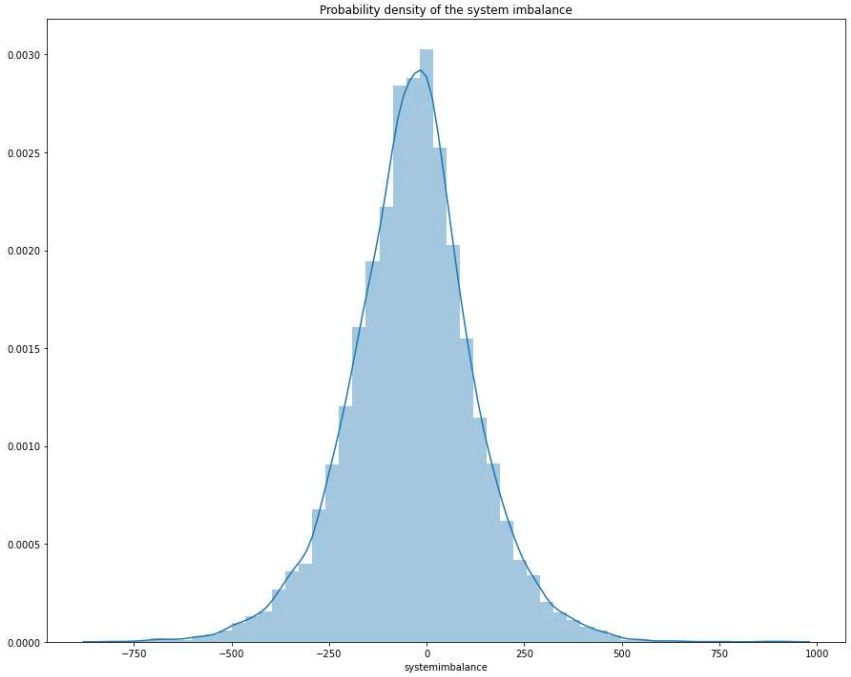
This brings us to question 1.1 :

Question 1.1: What experience on improving the performance of machine learning-based systems does exist in terms of addressing the anomalies (rare event with significant consequences), which may pose considerable impact on technological and/or economic performance of the power system? How should we distinct anomalies and data outliers between each other?

In our case we observed that the system imbalance distribution looks almost like a normal distribution but with too much outliers (fig 1 and 2). After investigation, we could conclude that the anomalies are there when we have a big outage in Belgium like a loss of a big powerplant. Those can be clearly distinguished from the outliers that naturally arise in a normal distribution. The mitigation here is to just exclude the datapoints from timestamps where there is a big outage from the train set. Excluding big outages from our data improves performance.



Figuur 1ECDF of system imbalance values



Figuur 2: Probability density of the system imbalance