Gruppe 10 SWMAL Bjarke G, Torsten von Wildenradt Nielsen, Yevhen Parolia **08.05.2023**

Introduction

The objective of the project is to be able to make predictions about the validity of the TSO forecast and to forecast energy generation and energy prices based on weather and energy datasets. More specifically, the two datasets will be used to answer the question of whether the weather forecast affects the size of discrepancy between the price day ahead and the actual price of the electricity. This will help TSOs to determine the validity of their own forecasts.

Materials and methods

Polynomial regression assumes a nonlinear relationship between the variables, compared to the linear relationship in linear regression.

We want to find the polynomial equation that best fits the data, without underfitting or overfitting. We have to be careful not to overfit, as this would cause poor performance on newly presented data.

The goal is to hit the optimal capacity, where the training error as well as the generalization error is the smallest.

A method to not overfit would be to implement early stopping.

Results

Main goal question:
Can we determine which features cause the discrepancy between the price day ahead and the price actual, and can we train a model that can help the TSO to predict the discrepancy based on the aforementioned features?

	price day ahead	price actual
mean	49.87	57.88

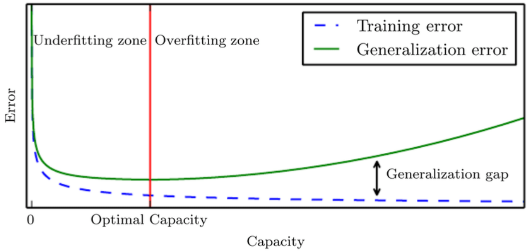
=> Big discrepancy between price day ahead and price actual (4 years of hourly data !)

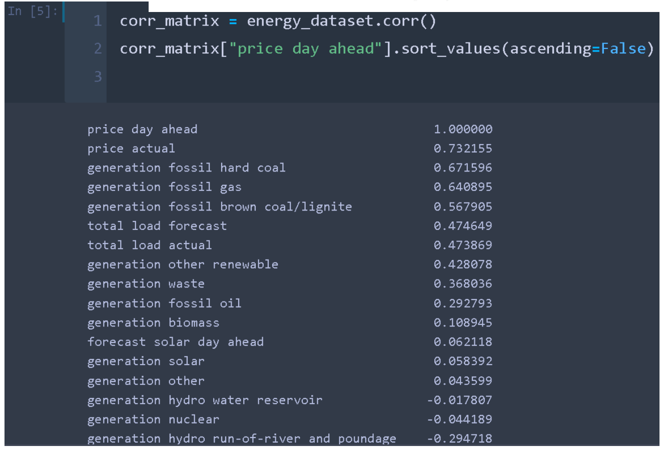
Dataset:

The dataset contains 35040 samples (1 for each hour in 4 years)

Sources of error in the dataset:

- Accuracy of weather data
- Merging of the two datasets
  - Datetime index error
- Take an average of weather data
  - For the 5 cities

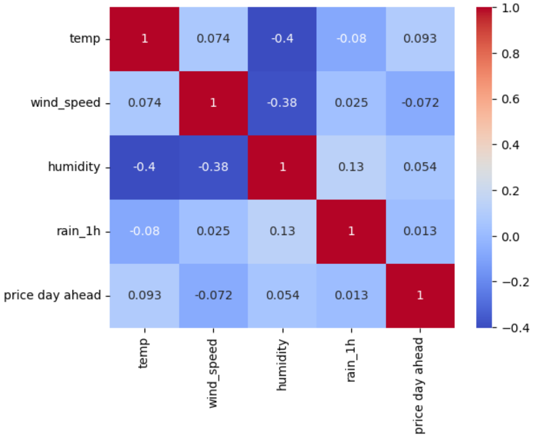




```
In [5]:
corr_matrix = energy_dataset.corr()
corr_matrix["price day ahead"].sort_values(ascending=False)
```

```
price day ahead                               1.000000
price actual                                  0.732155
generation fossil hard coal                   0.671596
generation fossil gas                         0.640895
generation fossil brown coal/lignite          0.567905
total load forecast                           0.474649
total load actual                             0.473869
generation other renewable                    0.428078
generation waste                              0.368036
generation fossil oil                         0.292793
generation biomass                            0.108945
forecast solar day ahead                      0.062118
generation solar                              0.058392
generation other                              0.043599
generation hydro water reservoir             -0.017807
generation nuclear                           -0.044189
generation hydro run-of-river and poundage   -0.294718
```

Conclusions: **We find that some features correlate strongly with the actual price and the price day ahead. This can potentially simplify the number of features we have to include in the model and shorten the search space for training.**



	temp	wind_speed	humidity	rain_1h	price day ahead
temp	1	0.074	-0.4	-0.08	0.093
wind_speed	0.074	1	-0.38	0.025	-0.072
humidity	-0.4	-0.38	1	0.13	0.054
rain_1h	-0.08	0.025	0.13	1	0.013
price day ahead	0.093	-0.072	0.054	0.013	1

Literature cited

https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather?select=energy_dataset.csv