Machine learning for detecting damages in wind turbines

Degree programme: BSc in Industrial Engineering and Management Science | Specialisation: Business Engineering
Thesis advisor: Prof. Dr. Angela Meyer
Expert: Prof. Florian Thürkow

Over the past years, renewable energy has gained more and more importance in the industry's efforts to decarbonize the energy system. Being able to manage the power production is crucial for the security of energy supply.

Samuel Trescoli Blasco

Initial situation and objectives

The company EDP offers its data from wind turbines on its website for free. The thesis used this data as main source of information from the wind turbines. The scope of the research is to explore the data available and try to predict future failures.

Research design

The starting point was to understand the datasets. Once the data was comprehensible, the next step was to create a hypothesis on how to predict the damages. A first dataset contained the measuraments of the sensors on the wind turbines. A second data set was used to record failures on each turbine for the years 2016 and 2017.
The main hypothesis was that the abnormal power production during a period could be an indicator of the failure of the wind turbine.
The last step was to apply the machine learning method Random Forest Classification to create the predictions. Model performance was defined with the confusion matrix and the parameters associated: Accuracy(0‘89), Precision(0‘89), Recall(0‘99) and the F1 score(0‘94). The confusion matrix counts the correct and the wrong classifications done by the model.

Results and recommendations

The result of the analysis did not show a significant correlation between the abnormal power production and failures in wind turbines. So, the first hypothesis was rejected. However, another possibility to predict the failures emerged during the analysis process related to the temperature and the season of the year. The temperatures of the ambient environment were correlated with the failures. The summer season, and especially during the month of August, the wind turbines tended to fail more often. (Figure 1)
After applying Random Forest as Machine learning classification algorithm, the feature importance showed that the temperature of the components was strongly correlated with the failures. The most important temperature variables were: Ambient(0‘062), Oil(0‘0531) and Spin(0‘052). In future research, it will be interesting to further explore this correlation and to investigate the predictability of wind turbine damage based on the recorded component temperatures.

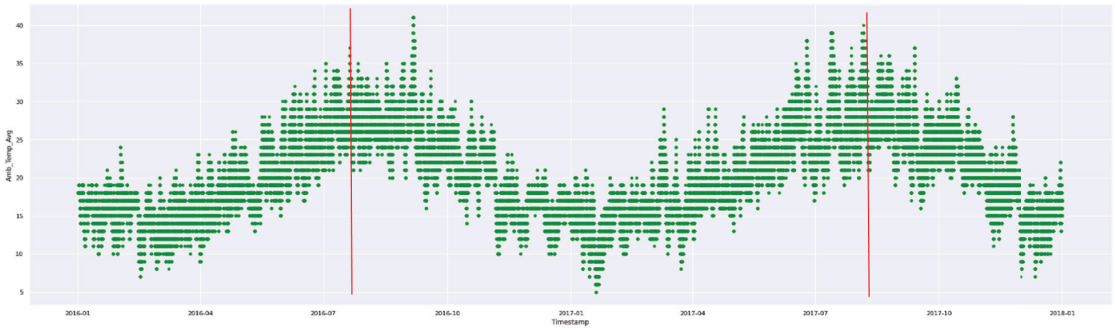

Figure 1: Distribution of failures and temperature during two years (Turbine 01)