

Uptake Data Science Case Study

Mami Saito

Uptake Data Science Case Study Code

- csv load – load revenue.csv, testing.csv, training.csv, and weather.csv
- Preprocessing training data
 - NaN, empty cell, and zero are removed from data
 - Outliers are removed from data
 - sensor_8, sensor_21, sensor_87, sensor_97, sensor_140, and sensor_167 are selected
- MATLAB TreeBagger is selected to do classification
 - 'failure' is set to 1 and 'no_failure' is set to 0
- Predict with testing data
 - This part has an issue

What I Would Do If I Had More Time (1)

- Predict with test data needs to be examined since it does not seem to produce the result
- After predicting test.csv, the model needs to be evaluated
- Look for additional variables that would help the model to be improved (e.g., other sensors, weather data)
- Profit needs to be calculated

What I Would Do If I Had More Time (2)

- Code needs to be optimized and converted into the one such that it could be reused for other energy companies that own wind turbines
- Compare with other machine learning methods such as SVM, logistic regression
- Create presentation summarizing results for non-technical audience