Capstone Project: Predictive Maintenance Classification

Data source: https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset

Acknowledgements/References:
Matzka, S. (2020). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Dataset information (as obtained from the website source above):

Since real predictive maintenance datasets are generally difficult to obtain, and in particular difficult to publish, we present and provide a synthetic dataset that reflects real predictive maintenance encountered in industry to the best of our knowledge.

Since real predictive maintenance datasets are generally difficult to obtain, and in particular difficult to publish, dataset author provides a synthetic dataset that reflects real predictive maintenance encountered in industry to the best of his knowledge.

Number of observations: 10,000
Number of features: 14

Target variable

- `Machine failure` – Machine failure status (0 = Not failed, 1 = Failed). This label is set to 1 if any of the Failure Modes described in the Features section below are true. Thus, in the case of more than 1 failure more, it is not apparent which of the failure modes has caused the process to fail.

Features/Attributes

Product information

- `UID` – Unique identifier ranging from 1 to 10000
- `Product ID` – Identifier consisting of the concatenation of the Type (see next feature) and a variant-specific serial number.
- `Type` – Classification of product quality variants. Labeled as low (L), medium (M), or high (H). Breakdown of types: L = 50% of all products, M = 30% of all products, and H = 20% of all products (is the breakdown 60/30/10?)

Process information

- `Air temperature [K]` – Air temperature in degrees Kelvin. Data generated using a random walk process later normalized to a standard deviation of 2 K around 300 K.
- `Process temperature [K]` – Process temperature in degrees Kelvin. Data generated using a random walk process normalized to a standard deviation of 1 K, then added to the air temperature plus 10 K.
- `Rotational speed [rpm]`: Revolutions per minute. Calculated from a power of 2860 W, overlaid with normally distributed noise.
- `Torque [Nm]`: Torque measured in newton-meters. Torque values are normally distributed around 40 Nm with a σ = 10 Nm and no negative values.

- `Tool wear [min]`: Tool wear in minutes. The quality variants H/M/L add 5/3/2 minutes of tool wear respectively to the used tool in the process.

Failure Mode Information

The machine failure consists of five independent failure modes:

- `TWF` – Tool wear failure (0/1). The tool will be replaced or fail at a randomly selected tool wear time between 200 - 240 mins (120 times in our dataset). At this point in time, the tool is replaced 69 times, and fails 51 times (randomly assigned).
- `HDF` – Heat dissipation failure. Heat dissipation causes a process failure, if the difference between air- and process temperature is below 8.6 K and the tool's rotational speed is below 1380 rpm. This is the case for 115 data points.
- `PWF` – Power failure. The product of torque and rotational speed (in rad/s) equals the power required for the process. If this power is below 3500 W or above 9000 W, the process fails, which is the case 95 times in our dataset.
- `OSF` – Overstrain failure. If the product of tool wear and torque exceeds 11,000 min-Nm for the L product variant (M: 12,000 min-Nm, H: 13,000 min-Nm), the process fails due to overstrain. This is true for 98 datapoints.
- `RNF` – Random failures. Each process has a chance of 0.1 % to fail regardless of its process parameters. This is the case for only 5 datapoints, less than could be expected for 10,000 datapoints in our dataset.