# Project 1: Predictive Maintenance

Data Science and Machine Learning for Engineering Applications

Politecnico di Torino

## Introduction

In this dataset, various characteristics of the machines and their status (working or failure) were measured. The data are synthetic but represent a real case.

## Objective

Classify when a certain state of machinery leads to failure (i.e., "failure" and "working" binary classification given the other input features). Optional: for the "failure" examples, predict the failure type (i.e., multi-class classification of the failure type for only the failure rows). The two classifiers can be used together to predict failure and, in case of failure, the type of failure.

## Dataset and Task

### Task

The task is a binary classification problem. The dataset consists of two files in CSV format: a *predictive_maIntenance_train.csv* file that contains the data to be used for training (and possibly validation) and a file *predictive_maIntenance_test.csv* that should be used for testing. Remember that the test data should not be used in the training phase under any circumstances. Once the model has been trained, the final evaluation should be done on the test data.

### Dataset Attributes

The following attributes are present in the dataset:

- UID: unique identifier ranging from 1 to 10000
- productID: consisting of a letter L, M, or H for low (50% of all products), medium (30%), and high (20%) as product quality variants and a variant-specific serial number
- air temperature [K]: generated using a random walk process later normalized to a standard deviation of 2 K around 300 K
- process temperature [K]: generated using a random walk process normalized to a standard deviation of 1 K, added to the air temperature plus 10 K.
- rotational speed [rpm]: calculated from powepower of 2860 W, overlaid with a normally distributed noise
- torque [Nm]: torque values are normally distributed around 40 Nm with an $\ddot{\text{I}}\text{f} = 10$ Nm and no negative values.
- tool wear [min]: The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool in the process. and a 'machine failure' label that indicates, whether the machine has failed in this particular data point for any of the following failure modes are true.

### Reference

The original dataset can be found in the following URL: https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification

## Note

The dataset used in the project is a re-sampling of the original dataset (i.e., it is a modified version of the original dataset). The data collection is the result of manual collection, so it is necessary to pay attention to possible errors, missing values, etc.

For binary classification, we are interested in predicting the **Failure Label** (i.e., the **Failure** column in the dataset). For this task, the **Failure Type** column must be ignored during training(i.e., not included in the input features) and also for prediction. It can be used only as a label for the additional multi-class classification task.

**Important**: The dataset is rather **imbalanced**, having few failure cases. It is important to highly identify instances in which a failure could occur.