# Assignment

### Goals

1. Demonstrate understanding of ML fundamentals

2. Show ability to handle real-world industrial data

3. Apply domain knowledge to feature engineering

4. Build and evaluate multiple ML approaches

5. Implement proper model validation

### Dataset

You will receive:

1. sensor\_data.csv: Historical sensor readings

This files contains 11 columns out of which ‘**Conditions**’ is the target value

It has 3 pressure sensors, 4 temperature sensors and 1 vibration sensor data and others are timestamp and pump id.

2.test\_data.csv

This is similar to sensor\_data, here it will not have conditions, which you have to predict.

Data Units

All pressure unit are in **bar**

All temperature are in **Deg C**

Vibrations are in **mm/s**

Description of conditions

Conditions describe the pump leakage with following values.

* Pump leakage:  
  1: no leakage  
  1: weak leakage  
  2: severe leakage

## Task Requirements

### Data Analysis & Preprocessing

* Handle missing values appropriately
* Identify and handle outliers
* Create time-based features
* Normalize/scale features

### Model Development

* Implement at least three different models.
* Generate the results with all the models and compare their results.

### Model Validation & Analysis

* Implement proper cross-validation
* Analyze feature importance and correlations.
* Evaluate predictions with multiple metrics like f2\_score, accuracy etc.

### Deliverables

* You must provide a Jupyter notebook with all the logic you built.
* The prediction on test\_data( add predictions in test\_data itself)
* Also the prediction files for all models you generated with comparison chart

Optional

If you can document the process and techniques you used to train the model , that will be great.