

# Modelling Week

Problem proposal

|                                   |
|-----------------------------------|
| Modelling Week – Problem proposal |
|-----------------------------------|

**Title: Design of a product recommendation model****Description:**

Maintenance operations have always been a concern for industrial companies. Minimizing unexpected downtimes translates to minimizing costs and maximizing production output.

With the new era of digitalization, more and more companies are starting to monitor their machines using sensors that measure temperature, pressure, and other physical quantities for key components of the production chain. These data are represented as a large time series, and the objective is clear: to predict future breakdowns.

In this project, we propose analyzing a public dataset recently released from the metro of Portugal (<https://arxiv.org/abs/2207.05466>). The goal is to predict future anomalies, whether they result in failures or not.

**Development details:**

You can download the data from <https://zenodo.org/record/6854240#.ZBMQ1XZw2UJ>.

Then:

1. Split the time series data into two halves while respecting the order of the data points.
2. Use the first half to train unsupervised or supervised models to predict anomalies in the data. For example, you could use clustering algorithms or regression models with anomaly detection techniques. Explore different methods to find the best approach for this data.
3. Use the trained model to make predictions on the second half of the data and identify the times when anomalies occur.

The final output should be a list of times when anomalies were detected, and the code used to train your model and make predictions.

You will be scored based on (in order of importance):

1. Your code cleanup, comments, code reproducibility, and using good programming practices.
2. Your explanation of the model you chose and why you believe it is appropriate for this dataset.
3. The algorithms you use to detect anomalies.

Please note that accuracy or performance will not be the primary scoring criteria. Instead, we are looking for clear and well-documented code, an understanding of the problem and the chosen approach, and the ability to explain your reasoning behind the model selection.