**Final Project**

Important information:

* This is a project that needs to be solved in groups of two. You will need to fill out a Google Forms specifying your partner (one answer per group is enough) **before February 17th, 23:59**. Everyone that has not answered this form by that date will be assigned a random classmate.
* **The deadline for the project is March 21, 23:59. No extensions will be granted.**
* The submission will be through BruinLearn. Each group will have to submit **one** project package consisting of:
  + Jupyter Notebook with the code and discussion detailing your solution.
* We highly encourage you to start working **early**. No extra discussion sessions, or office hours will be granted for the project, **so do not wait until the last week to start asking questions.** This is not a project you can start and finish in a couple of days.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Consider a mining processing line dedicated to copper production. From the collected mineral stockpile, the material is transported through a series of conveyor belts to different components, aiming to make the extraction process of the mineral from the rock easier. Here, two feeders in a parallel conﬁguration transport the raw mineral from a stockpile to a belt conveyor, a sifter, and ﬁnally the crusher. The copper mining crusher (CMC) is the principal component in this processing line and is the focus of this project.

A simpliﬁed diagram of the monitored components is presented in Figure 1.

A diagram of a belt

AI-generated content may be incorrect.

Figure 1 - Simpliﬁed diagram of copper mining crusher processing line.

A condition-monitoring sensor network collects operational data from the CMC system’s components. Sensor measurements, including the time variable, are available from 22 sensors with a sampling time of 2 min. Data are acquired from July 1st/2017 to October 1st/2019.

Data logs containing information on all the interruptions and stoppages that the system experienced are also provided. The system’s logs are available from January 2nd/2017 to August 31st/2019. These data document the interruptions of the system’s operation classiﬁed into two categories: interruption of normal operation (including programmed maintenance and inspections) and failures detected in the system. There are 2,595 operational interruptions reported during this period, where only 382 explicitly refer to failures in the system.

Table 1 presents the internal classiﬁcation of these failures. As it can be seen, several causes do not necessarily correspond to failures of the CMC system, such as sensor communication errors, inspection, programmed maintenance, and other external events. Thus, the system downtime events logs are ﬁltered only considering events related to failures.

A list of instructions on a white background

AI-generated content may be incorrect.

Table 1 – List of most significant identified interruptions.

All the monitoring variables are strictly positive. The nomenclature used for the available sensor monitoring variables is presented in Table 2.

A close-up of a sensor name

AI-generated content may be incorrect.

Table 2 – Sensor nomenclature in the crusher system.

The raw sensor data for the entire copper mining processing line is stored in the CSV file: “Copper\_Mining\_Raw\_Sensor\_Data.csv”. In this file, there are the sensor data not only relevant to the crusher, but also to the other equipment in the processing line.

The copper mining processing line downtime events logs are available in the CSV file: “Downtime\_Events.csv”.

Based on the above information and data sources, you must develop a predictive solution that provides health state diagnostics for the copper mining crusher. To this end, the following goals should be completed:

1. Preprocess the raw sensor data
2. Preprocess the raw downtime events log data
3. Generate the labels (health states) for the crusher from the synchronization between the sensor data and the downtime events (relevant to the crusher)
4. Perform PCA to the cleaned data
5. Develop various machine learning models based on the sensor variables as well as on the principal components. You should at least try the following models (other options are also possible and welcome!):
   1. Logistic Regression
   2. Support Vector Machine
   3. Decision Tree
   4. Random Forest
   5. Gradient Boosting
   6. Dense (Feedforward) Neural Network
6. For each of the implemented models, you should report and discuss the following:
   1. Model architecture and hyper-parameters
   2. Performance metrics:
      1. Confusion matrix
      2. Accuracy (balanced and unbalanced)
      3. Precision
      4. Recall
      5. F1-Socre
      6. Plot of Train x Validation losses for the neural network
7. Select one of the solutions above and discuss your choice
8. For the selected model, perform Grid Search:
   1. Discuss the chosen hyper-parameters and their ranges
9. Based on the grid search, perform Cross-Validation on the model with the best set of hyper-parameters:
   1. Report and discuss the results