# Using Machine Learning to Develop a Supernatural Martian Substance

### **TASK DESCRIPTION**

A visionary Company X is on a mission to build a self-sustaining city on the red planet. As part of this mission, scientists at Company X are developing a supernatural substance that can withstand Mars' extraterrestrial conditions. This substance will be the building block of this futuristic Martian city. Can Company X turn this idea into a reality? In this assignment, you will use machine learning to predict the quality of this substance based on historical data.

The task is formulated as a regression problem where you will predict a numerical metric that measures the quality of the Martian substance. You will be evaluated with the Mean Absolute Error (MAE) metric. Be aware that MAE is a negatively-oriented score, which means lower values are better!

## **DATASET DESCRIPTION**

**Number of instances:** 4000 **Number of attributes:** 15

**Target Variable:** A numerical quality metric of the Martian substance that Company X is developing.

### Attribute Information:

Name	Description	Туре	Values
1	Chemical property A of the substance	Numerical	Positive and negative float
2	Chemical property B of the substance	Numerical	Positive and negative float
3	Chemical property C of the substance	Numerical	Positive and negative float
4	Chemical property D of the substance	Numerical	Positive and negative float
5	Chemical property E of the substance	Numerical	Positive and negative float
6	Chemical property F of the substance	Numerical	Positive and negative float
7	Chemical property G of the substance	Numerical	Positive and negative float
8	Whether or not a refraction test was	Categorical	0: Yes
	performed on the substance		1: No
9	Physical property A of the substance	Numerical	Positive and negative float
10	Whether or not a radioactivity test	Categorical	0: Yes
	was performed on the substance		1: No
11	Physical property B of the substance	Numerical	Positive and negative float
12	Physical property C of the substance	Numerical	Positive and negative float
13	Physical property D of the substance	Numerical	Positive and negative float
14	Physical property E of the substance	Numerical	Positive and negative float
15	Physical property F of the substance	Numerical	Positive and negative float

#### PREDICTION FILE SUBMISSION

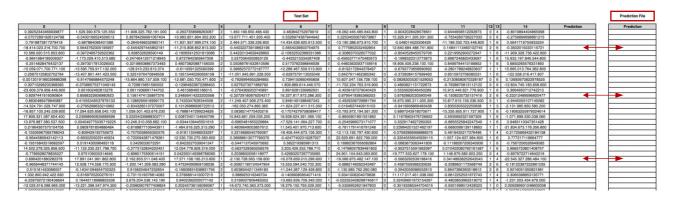
You are kindly requested to strictly follow the described submission guidelines:

File Format: .csv

**Filename:** Student code (e.g. 123456.csv) **Column Format:** 1 Column named *Prediction* 

Row Format: Your predictions with the same number of rows and in the same order as the test set

provided to you (see below)



### PDF FILE SUBMISSION

Along with your predictions, you are asked to kindly submit the following supporting information:

- A brief description of the step by step methodology (i.e. data cleaning, pre-processing, training, evaluation, etc.) that you have followed to do the assignment, with the aim of illustrating the motivation behind your selected approach.
- 2. The python code that you used to do the assignment, with a fair amount of comments within the code to ensure that they can be clearly understood.

File Format: .pdf

Filename: Student code (e.g. 123456.pdf)

### **ASSIGNMENT DEADLINES**

A. November 27th - The labeled dataset will be released on the BeeP platform;

- B. December 4th The unlabeled dataset will be released on the BeeP platform;
- C. December 5th to December 7th ASSIGNMENT DUE DATE: you are requested to upload both your PREDICTION FILE and the PDF FILE on the BeeP platform. The Assignment folder will be open from December 5<sup>th</sup> to December 7<sup>th</sup> at 19pm.