# **AI & Sustainability**

**Water shortage prediction** 🚱

## **Goal/Target**

The goal is to build an AI model that can predict the watertable/ground water levels of french piezometric stations, with a focus on the summer months. To build this model, you were given piezometric/watertable, weather, hydrology, water withdrawal and economic data.

But beyond producing an AI model, the competition will ask you to realistically project your solution in a market / real-world context.

## **Deliverables**

* Build an AI model and make predictions
* Create a scientific / business document
* Produce a presentation video
* Defend your project in front of a live audience (this deliverable only concerns 7-10 groups notified Sunday evening)

More information on the deliverables and the event in general can be found in the "Hickathon\_student\_guide.pdf" document.

## **Dataset**

The full dataset contains over 3 million rows with 136 columns. It was split into a train/test set.

* Train set (X\_train\_Hi5.csv): The dataset has around 2 800 000 rows. It contains data between 2020 and 2023, excluding the summer months (june, july, august, september) of 2022 and 2023.
* Test set (X\_test\_Hi5.csv): The contains has around 600 000 rows. It contains data for the 2022 and 2023 summer months (june, july, august, september).
* Test submission example (y\_test\_submission\_example\_Hi5.csv): Please follow this example to submit results to the leaderboard. The "row\_index" variable is a unique identifier of each row, to match the values

The target variable to predict is piezo\_groundwater\_level\_category.

The datasets are available in the Studio via this path ~/hfactory\_magic\_folders/water\_shortage\_prediction/ .

Make sure you work in the /shared\_storage folder since you don't have write permissions in the /water\_shortage\_prediction !

A google drive containing all of the sprint documents + the datasets is available [here](https://drive.google.com/drive/folders/1r630CoylbFw7DpnPIYfs1IymY_Nyv3uH?usp=sharing)

A dictionary with a description of each variable is available [here](https://docs.google.com/spreadsheets/d/1seb0FFxci_AnbdyJW-Kg2ZDcx-Y7qrgd/edit?usp=sharing&ouid=110879283880412136702&rtpof=true&sd=true)

Access the discord server of the event [here](https://discord.gg/DpF2Yg3A)

**Wifi 1:**

* Login: telecom411284-0628
* Password: DINGO/sheet

**Wifi 2:**

* Login: telecom411284-6031
* Password: TRACK IRON