CS Game 2023
Machine Learning
Prediction in a hydroelectric environment

Winter 2023 **Author : Yoan Villeneuve**

General Instructions:

- You have **6 hours** to complete this exam.
- Only one submission per team is allowed.
- This challenge counts for **9% (1800 points)**.

Objective of the competition

Develop a machine learning model using the Python programming language and related libraries that can **predict water inflows** in the territory using a hydroelectric dataset.

Context

In this exercise, the Norwegian government is facing a problem with the hydroelectric power production of the plants located on its territory. Since Quebec is a leader in the field of hydroelectricity, they are calling on a small or medium-sized enterprise from the Saguenay Lac-Saint-Jean region, namely you who are reading this document, to solve their problem. As most of the Norwegian power plants are operated by private companies, the government has asked each plant in the country to provide a production report for the past few years. They hope that Quebec's expertise in centralized management of its plants (thanks to the government-owned company Hydro-Quebec) will be able to model Norwegian power production.

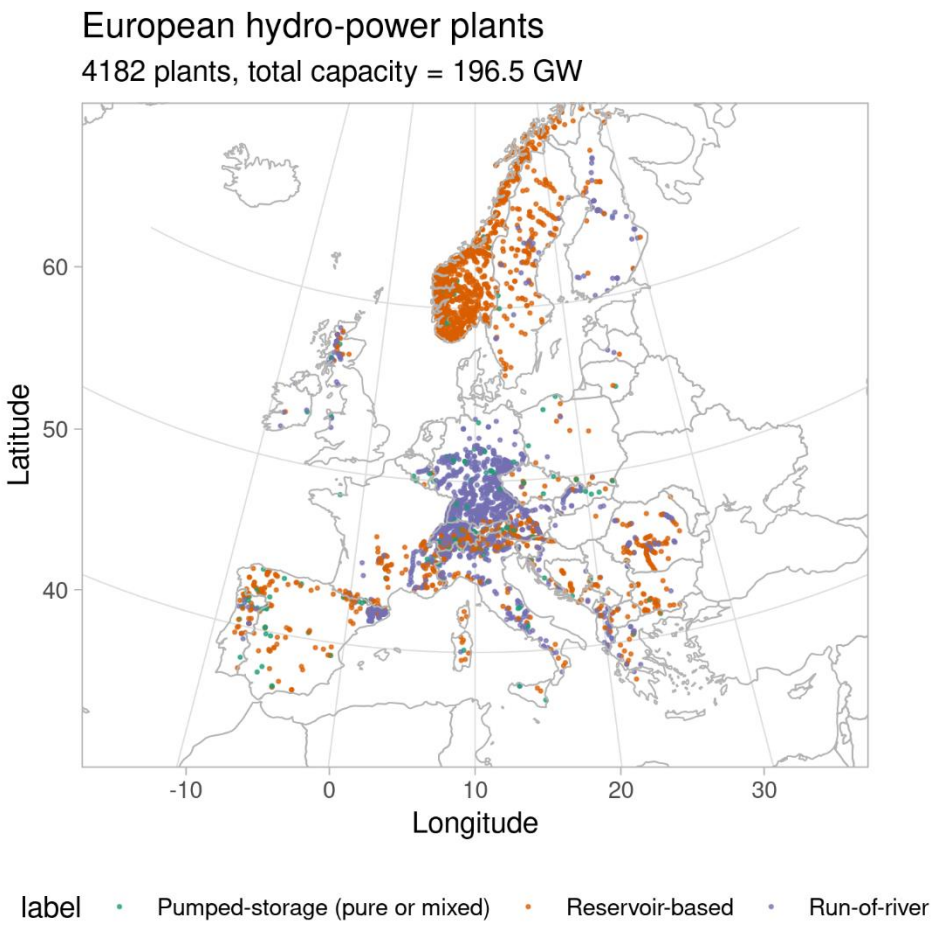

Figure 1 : Hydropower plants of Europe (https://github.com/energy-modelling-toolkit/hydro-power-database/)

Description of the competition

You have received a dataset named "NORWAY_INFLOW_2018-2022.parquet" containing **weekly information over a period of five years** (early 2018 to late 2022) related to the electricity production of Norwegian hydroelectric power plants. For security reasons, all power plants have been renamed "**H**" followed by a numeric identifier (for example, H1006, H1016, H116, etc.). Each instance in the dataset represents a week. These instances consist of 475 attributes (features). Here is a list of each attribute followed by a brief description.

- **Year**: The year of the instance recording.
- **Week**: The week of the instance recording (relative to the year).
- **Entsoe_inflow**: The amount of water inflow observed for all "H" power plants' reservoirs.
- **Time_id**: A mix of the Year and Week features.
- **H + numeric identifier**: Power (GW) produced by the power plant during the week. There are 471 hydroelectric power plants provided in each instance.

Objective of the Challenge

Your goal is to design a model capable of predicting the water inflow (**Entsoe_inflow**) on the territory of Norway. **The Norwegian government already knows the weekly energy production forecasts for each power plant on their territory for the year**. It is up to you to provide a machine learning-inspired model capable of predicting the expected water inflow for the year 2023. You have access to the **Python 3.11** programming language and the following libraries:

- NumPy : a library for performing numerical computations in Python. https://numpy.org/doc/stable/reference/index.html
- Pandas : a library for data manipulation and analysis in Python. https://pandas.pydata.org/docs/reference/
- Matplotlib : a library for data visualization in Python. https://matplotlib.org/stable/api/index.html
- Scikit-learn : a machine learning library for Python. https://scikit-learn.org/stable/index.html
- TensorFlow : an open-source machine learning library developed by Google. https://www.tensorflow.org/api_docs/python/tf/all_symbols
- PyTorch : another open-source machine learning library that focuses on deep learning. https://pytorch.org/docs/stable/index.html

The goal is to predict the value of the **Entsoe_inflow** attribute based on historical data.

- There are no restrictions on the type of model that can be used for this challenge.
- There are no restrictions on the learning algorithm used by your model.
- There are no restrictions on the manipulation and pre-processing of the provided dataset.
- The final program must be functional and its operation must be detailed.

Correction method

After the end of the competition, **the correction will be made using a file with a size of 52 instances (The year 2023)** named "*NORWAY_INFLOW_2023.parquet*" having the same attributes as the "*NORWAY_INFLOW_2018-2022.parquet*" file that was provided to you for training. This file also includes the real "**Entsoe_inflow**" attribute to be predicted. **Your code must be able to take this file as input and make a prediction for each instance**. Your code should also output a text file to visualize the predicted values.

You can use the method you want to integrate this set, but you must explain textually in the delivery how to use your program in order to facilitate the work of the corrector. You must therefore indicate how to run your program in order to make the predictions. It is strongly advised to calculate performance values (ex. MAE, MSE, RMSE, F1 Score, etc.). You can submit your program with your model already trained or not. Nevertheless, **your program should not take more than 2 to 3 minutes to make the prediction** (and the learning if the latter has not been done in advance).

Submission and Marking Scheme

You must submit your code and all associated documents before the end of the authorized competition period. Along with your model, you must submit a document describing how to operate your program in order to facilitate correction and avoid errors.

You must submit the following files:

- The folder containing the Python code used to run your prediction model.
- A written report including:
    - How to operate your program;
    - The pre-processing phase;
    - Model type and training method;
    - A rationale for choosing your model;
    - A summary of your model's performance;
    - Other comments, information and graphics relevant to the correction.

Score

This competition is worth **1800 points** of the cumulative total during the CS games, or **9%** of the total score. Of the **1800 points** that can be earned during this competition:

- **1000 points** for the development of a functional model capable of performing good prediction during the correction (MAE, RMSE and other performance values).
- **600 points** for the complexity or creativity of the model developed.
- **200 points** for submitting a clear and detailed rendering for the correction.