

CS Game 2023

Apprentissage Automatique

Prédiction en milieu hydroélectrique

Hiver 2023

Auteur : Yoan Villeneuve

Instructions générales

- Vous avez **6 heures** pour terminer cette épreuve.
- Une seule remise à remettre par équipe.
- Ce travail compte pour **9% (1800 points)**.

Objectifs de l'épreuve

Développer à l'aide du langage de programmation Python et des bibliothèques connexes un modèle d'apprentissage automatique capable de **prédire les apports en eau** sur le territoire à l'aide d'un ensemble de données hydroélectrique.

Mise en Contexte

Dans cet exercice, le gouvernement norvégien est confronté à un problème concernant la production d'énergie hydroélectrique des centrales situées sur son territoire. Étant donné que le Québec est un leader dans le domaine de l'hydroélectricité, ceux-ci font appel à une PME de la région du Saguenay Lac-Saint-Jean, à savoir vous qui lisez ce document, pour résoudre son problème. Comme la majeure partie des centrales norvégiennes sont exploitées par des entreprises privées, le gouvernement a demandé à chaque centrale du pays de fournir un bilan de production au cours des dernières années. Il espère que l'expertise du Québec dans la gestion centralisée de ses centrales (grâce à la société paragonnementale *Hydro-Québec*) sera capable de modéliser la production électrique norvégienne.

European hydro-power plants

4182 plants, total capacity = 196.5 GW

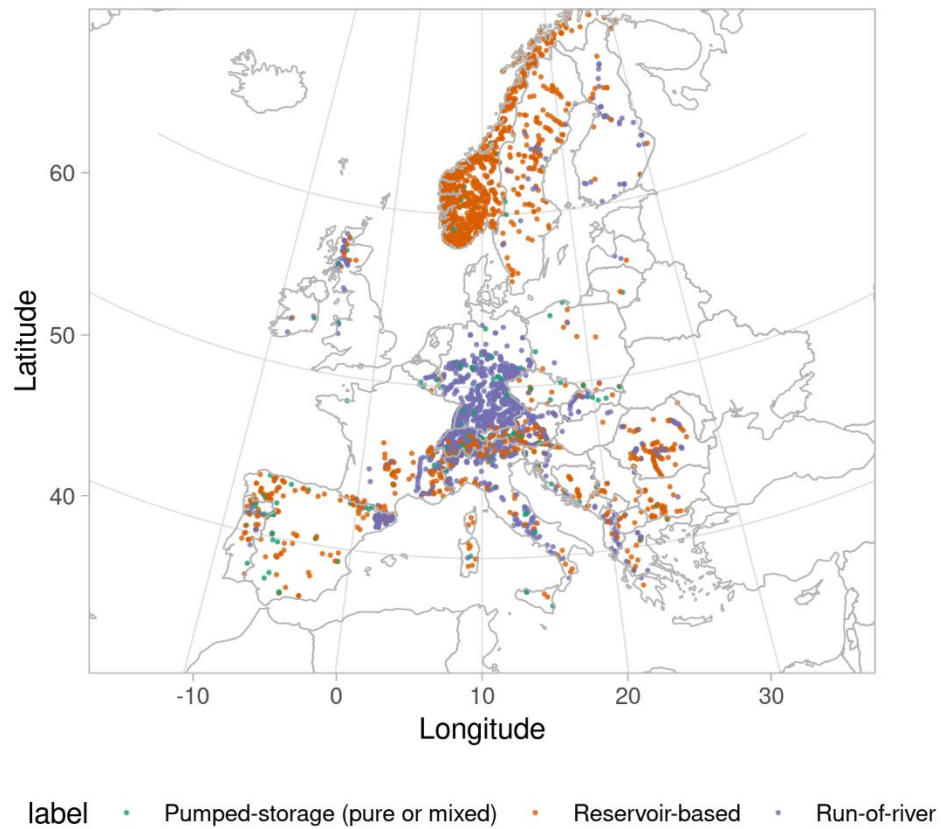


Figure 1 : Centrales hydroélectriques de l'Europe (<https://github.com/energy-modelling-toolkit/hydro-power-database/>)

Énoncé de l'épreuve

Vous avez reçu un ensemble de données nommé « *NORWAY_INFLOW_2018-2022.parquet* » contenant des informations **hebdomadaires** sur une période de **cinq ans** (début 2018 à fin 2022) relative à la production d'électricité des centrales hydroélectriques norvégiennes. Par mesure de sécurité, toutes les centrales ont été renommées « **H** » suivies d'un identifiant numérique (par exemple, H1006, H1016, H116, etc.). Chaque instance de l'ensemble de données représente une semaine. Ces instances sont constituées de 475 attributs (caractéristiques). Voici une liste de chaque attribut suivi d'une courte description.

- **Year** : L'année de l'enregistrement de l'instance.
- **Week** : La semaine de l'enregistrement de l'instance, par rapport à l'année.
- **Entsoe_inflow** : La quantité d'apports en eau observée pour la totalité des réservoirs des centrales « H ».
- **Time_id** : Un mix des caractéristiques **Year** et **Week**.
- **H + identifiant numérique** : Puissance (GW) produite par la centrale au courant de la semaine. Il y a 471 centrales hydroélectriques de fournies dans chaque instance.

Objectif de l'épreuve

Votre objectif est de concevoir un modèle capable de faire **la prédiction des apports en eau (Entsoe_inflow)** sur le territoire de la Norvège. Le gouvernement norvégien connaît déjà les prévisions hebdomadaires de production d'énergie pour chaque centrale sur leur territoire pour l'année. Il en revient à vous de fournir un modèle inspiré de l'apprentissage automatique capable de **prédire l'apport en eau** attendu pour l'année 2023. Vous avez accès au langage de programmation **Python 3.11** ainsi que les librairies suivantes :

- NumPy: une bibliothèque pour effectuer des calculs numériques en Python.
<https://numpy.org/doc/stable/reference/index.html>
- Pandas: une bibliothèque pour la manipulation et l'analyse de données en Python.
<https://pandas.pydata.org/docs/reference/>
- Matplotlib: Une bibliothèque pour la visualisation de données en Python.
<https://matplotlib.org/stable/api/index.html>

- Scikit-learn : une bibliothèque d'apprentissage automatique pour Python. <https://scikit-learn.org/stable/index.html>
- TensorFlow : une bibliothèque d'apprentissage automatique open source développée par Google. https://www.tensorflow.org/api_docs/python/tf/all_symbols
- PyTorch : une autre bibliothèque d'apprentissage automatique open source mettant l'accent sur l'apprentissage en profondeur. <https://pytorch.org/docs/stable/index.html>

L'objectif est de prédire la valeur de l'attribut **Entsoe_inflow** en se basant sur les données historiques.

- Il n'y a pas de restriction sur le type de modèle pouvant être utilisé pour cette épreuve.
- Il n'y a pas de restriction sur l'algorithme d'apprentissage de votre modèle.
- Il n'y a pas de restriction sur la manipulation et le prétraitement de l'ensemble de données fourni.
- Le programme final doit est fonctionnelle et sont fonctionnement doit être détaillé.

Méthode d'évaluation

La correction se fera à l'aide d'un fichier **d'une taille de 52 instances** (L'année 2023) nommé « *NORWAY_INFLOW_2023.parquet* » ayant les mêmes attributs que le fichier « *NORWAY_INFLOW_2018-2022.parquet* » vous ayant été fourni pour l'entraînement, y compris la valeur réelle de l'attribut à prédire. **Votre code doit être capable de prendre en entrée ce type de fichier et d'effectuer une prédiction pour chaque instance.** Votre code devrait aussi fournir un fichier texte en sortie permettant de visualiser les valeurs prédites.

Vous pouvez intégrer la méthode que vous désirez pour intégrer cet ensemble, mais **vous devez obligatoirement expliquer textuellement dans la remise comment utiliser votre programme** afin de faciliter le travail du correcteur. Vous devez donc indiquer comment lancer votre programme dans le but de faire la prédiction. Il est fortement conseillé de calculer des valeurs de performance (ex. MAE, MSE, RMSE, F1 Score, etc.). Vous pouvez remettre votre programme avec votre modèle déjà entraîné ou non. Néanmoins, votre programme ne doit pas prendre plus de 2 à 3 minutes pour faire la prédiction (et l'apprentissage si ce dernier n'a pas été fait à l'avance).

Remise et barème

Vous devez remettre votre code ainsi que tous documents associés avant la fin de la période de compétition allouée. Avec votre modèle, vous devez **remettre un document descriptif du fonctionnement de votre programme** afin de faciliter la correction et éviter les erreurs.

Vous devez **obligatoirement** remettre les fichiers suivants :

- Le dossier contenant le code Python permettant de lancer votre modèle de prédiction.
- Un rapport écrit détaillant :
 - Le fonctionnement de votre programme;
 - La phase de prétraitement;
 - Le type de modèle et la méthode d'entraînement;
 - Une justification du choix de votre modèle;
 - Un résumé des performances de votre modèle;
 - Autres commentaires, informations et graphes pertinents à la correction.

Pointage

Cette compétition vaut pour **1800 points** du total cumulable durant les CS games, soit **9%** du pointage total. Sur les 1800 points pouvant être amassé durant cette compétition :

- **1000 points** pour le développement d'un modèle fonctionnel et capable d'effectuer de bonne prédiction pendant la correction (MAE, RMSE et autres valeurs de performance).
- **600 points** pour la complexité ou la créativité du modèle développé.
- **200 points** pour la remise d'un rendu clair et détaillé pour la correction.