

8IAR403 : Apprentissage Automatique (AA) pour la science des données**Devoir #4****ANALYSE COMPARATIVE D'ALGORITHMES****Le travail peut être réalisé en équipe de 2****Date de remise : le 22 avril 2024****1. But**

- Mettre en place une expérimentation avec des données d'apprentissage dans un contexte réel en utilisant un algorithme de votre choix;
- Faire une analyse comparative des techniques d'apprentissage.

2. Étude de cas

Il s'agit d'explorer une problématique de votre choix, non seulement au niveau algorithmique d'apprentissage automatique mais aussi relativement au contexte d'étude et des données d'apprentissage. Cela peut être dans le domaine de la santé, comme l'analyse des données sur le Covid-19, l'analyse des sentiments en lien avec les réseaux sociaux, la vision artificielle, le jeu vidéo ou d'autres champs applications. Le but visé est de faire une analyse comparative entre une technique vue en cours et celle que vous avez choisie, en s'appuyant sur les mesures de performance des modèles appris. Vous devez choisir un des algorithmes qui n'a pas été vu en cours, déjà implémenté et fourni par une des bibliothèques d'apprentissage connues, comme scikit-learn ou Keras ou d'autres. Vous pouvez récupérer du code sur internet et le réutiliser. Voici un exemple d'une liste non exhaustive d'algorithmes que je vous propose mais vous pouvez recourir à d'autres :

1. **Variantes de CNN**, comme par exemple ResNet, Xception, ou d'autres et qui sont supportés par Keras (<https://keras.io/api/applications/>) en se basant sur **le transfert d'apprentissage**;
2. **Boosting Extrême du Gradient (XGBoost)** dont la documentation est disponible en utilisant le lien suivant: (<https://xgboost.readthedocs.io/en/stable/>)
3. ...

La démarche d'expérimentation, doit suivre celle que j'ai recommandée et qui s'inscrit dans la même lignée de conduite de projet de Machine Learning.

3. Livrable

Le livrable de l'étude menée doit être remis le lundi **22 avril 2024 et aucun livrable ne sera accepté au-delà de cette date.**

- 3.1 Vous devez fournir un fichier ipynb décrivant les étapes de la démarche d'entraînement du modèle. Vous pouvez ajouter tout au début de ce fichier une description en quelques lignes du contexte de l'étude et ainsi que et le dataset utilisé.
- 3.2 Vous devez inclure une « courte » vidéo de présentation d'une durée de dix minutes maximale où chaque membre de l'équipe doit présenter sa contribution. Le but visé est de donner une vue d'ensemble de l'algorithme utilisé, l'analyse comparative des résultats obtenus et une conclusion.

NB : voici quelques exemples d'entrepôt de données en source libre bien connus :

- <https://archive.ics.uci.edu/ml/datasets.php> : entrepôt de Machine Learning de l'Université d'Irvine, Californie
- <https://www.kaggle.com/datasets> : jeux de données Kaggle
- <https://registry.opendata.aws/> : jeux de données AWS d'amazon

Bonne continuation !