

Projet du cours d'Intelligence artificielle

Master 1 Informatique 2022-2023

Agitel Formation

Soit les 4 jeux de données suivants :

1. <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
2. [kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention](https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention)
3. <https://www.kaggle.com/datasets/anushonkar/network-anomaly-detection>
4. [https://www.kaggle.com/datasets/khusheekapoor/vehicle-insurance-fraud-detection\\$](https://www.kaggle.com/datasets/khusheekapoor/vehicle-insurance-fraud-detection)

Travail à faire

1. Faire un pré-traitement des données (résoudre les problèmes de doublons, de données manquantes, de jeu de données déséquilibrés, etc.)
2. Entraîner les données en utilisant 3 classifieurs différents et choisir le classifieur le plus efficace en termes de taux de justesse (*accuracy*)
3. Proposer une interface graphique pour tester le classifieur sélectionné à l'étape 2 (sur de nouvelles données) et la réaliser avec *Streamlit*
4. Déployer l'application avec *Streamlit Cloud*¹

N.B : Chaque étudiant devra utiliser qu'un seul jeu de données. Les travaux jugés similaires seront pénalisés.

Date limite de soumission des travaux : 18/03/2021 à 23h59

¹ <https://streamlit.io/cloud>