

# Consignes de projets AI28 – Machine Learning

1<sup>er</sup> juin 2023

## 1 Projet

Le projet est à effectuer en trinôme. Il consiste à traiter un problème de machine learning avec un jeu de données réelles, dont on vous les fournira. Il faut réaliser une démarche méthodologique de machine learning de bout en bout permettant de répondre au problème posé, à l'aide des algorithmes étudiés dans le cours. Justifier les choix de méthode, de paramètres, etc.; lier les résultats des éventuelles analyses préalables des algorithmes.

## 2 Rapport

Le rapport doit inclure une présentation claire et concise du problème et des données, une définition de l'objectif de l'étude, la démarche choisie et les résultats obtenus, ainsi qu'une conclusion, sans dépasser **10 pages de PDF** en utilisant le template LaTeX fourni. D'éventuelles illustrations complémentaires, non essentielles pour la compréhension du travail, pourront être mises dans des annexes (appendices).

Le rapport (obligatoirement en PDF) accompagné du code source (un jupyter notebook (et/ou un script Python) qui a servi à générer les résultats présentés dans le rapport devra être archivé dans un seul fichier .ZIP et déposé sur moodle dans l'activité "Dépôt Projets AI28-P23" avant **Samedi 17/06/2023 à 23h59**.

Chaque figure devra comporter une légende expliquant ce qui est illustré. Attention, une pénalité sera appliquée si une figure n'est pas reproductible.

## 3 Soutenance

Une séance de soutenance des projets aura lieu le **jeudi 22 juin de 10h15 à 12h45**. Les soutenances seront de durée 15 mm pour chaque trinôme et permettront de vérifier que chaque étudiant.e est bien auteur de son projet.

## 4 Groupes

(Groupe 1 - AAT) ##### Mostafa, ##### Ilian et ##### Thomas  
(Groupe 2 - ADG) ##### Yves, ##### César et ##### Clément  
(Groupe 3 - AMP) ##### Dorra, ##### Sophie et ##### Eden  
(Groupe 4 - BM) ##### Yassine et ##### Hugo  
(Groupe 5 - DGT) ##### Constant, ##### LY Anne-Soline et #####  
Alexandre

(Groupe 6 - DMW) #####É Damien, ##### Eloise et ##### Morgan  
(Groupe 7 - GNS) ##### Cécile, ##### Léo et ##### Adam  
(Groupe 8 - JLL) ##### Matthew, ##### Edouard et ##### Louis  
(Groupe 9 - KEZ) ##### Myriem, ##### Yassine et ##### Yasmine

## 5 Distribution de projets

Nous vous proposons 9 projets avec des jeux de données réelles répertoriés dans le site web UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>. La liste des titres de projets sont mis en ligne dans l'espace moodle. En cliquant sur un titre d'un projet, vous serez amenés à sa page de description ainsi que les données qui viennent avec.

Chaque projet peut être traité **au plus par deux groupes différentes**. Une fois les membres de groupes se mettent d'accord pour le choix d'un projet, vous devez me contacter par mail pour son attribution. Attention ! j'applique la règle "FIFO (First In First Out)", premier arrivé, premier servi.

Un fichier Excel **Groupes et distribution de données** dans la section de Projets est mis à jour régulièrement pour les projets affectés aux groupes. Avant m'envoyer votre mail, vous devez consulter ce fichier pour savoir si le sujet que vous iriez choisir est encore disponible.

## 6 Quantité de données de capacité des machines

Le nombres d'exemples dans certains projets dépasse les 20K. Si vos machines ne sont pas assez performantes pour l'entraînement, vous pouvez dans ce cas travailler avec un sous échantillon de taille entre 15K et 20K.

## 7 Points d'évaluation

### 7.1 Rapports

Le rapport doit contenir les principales étapes pour un projet machine learning de bout-en-bout :

- Description succincte de projet à mettre dans l'abstract (voir template).
- Analyse exploratoire de données.
- Pré-traitement de données.
- Modélisation : un large panel de modèles prédictifs doit figuré dans votre rapport.
- Hyper-optimisation des paramètres de modèles (validation croisée, GridSearCV, ...)
- Évaluation des modèles.
- Votre recommandation de modèle(s) pour le problème étudié.

La présentation des résultats et ses interprétations sont des points forts pour l'évaluation.

### 7.2 Code

Nous vous encourageons à écrire un code "pythonique", c'est à dire un code Python bien conçu et documenté pour faciliter sa compréhensible, en particulier vous pouvez coder

des petites fonctions de pre-processing, modélisation, évaluation, etc ... pour automatiser le plus possible l'exécution de vos programmes.