

# Mdélisation mathématique - Apprentissage automatique

## Projet : Développer une IA dans un contexte spécifique

### 2024 - 2025

Issam Falih  
Chafik Samir

Dans le cadre des exigences du cours, vous devez réaliser un projet de classe. Le projet peut être réalisé en groupe de 2 étudiants.

Le projet doit contenir au minimum les parties suivantes:

- Collecte et prétraitement des données & transformation des données (c.-à-d. nettoyage des données, remplissage des valeurs manquantes,...).
- Analyse des données.
- Visualisation des données
- Appliquer des algorithmes d'apprentissage supervisé et/ou des approches d'apprentissage non supervisé. Les comparer en utilisant des indices internes et externes si une étiquette de classe est fournie dans les données sélectionnées.
- Rapports de projet (voir les détails sur ce qu'il faut soumettre ci-dessous).

Vous serez *évalués* sur les étapes de traitement, les phases d'apprentissage, votre code, le rapport ainsi que l'analyse et les commentaires.

Le calendrier et les livrables du projet sont les suivants. Tous les livrables doivent être remis avant minuit à la date d'échéance :

- **26 Janvier 2025:** Déposez sur Moodle une proposition de projet de 2 pages. La proposition doit contenir les informations suivantes:
  - Le titre du projet ainsi qu'une liste des membres de l'équipe.
  - La liste des sources de données qui seront utilisées (fournissez une URL pour les données). Indiquez si les données nécessitent un prétraitement supplémentaire.
  - Un court résumé décrivant les objectifs du projet.
  - Le calendrier du projet.
  - Le rôle de chaque membre de l'équipe (qui sera responsable de quoi).

- **2 Février 2025** : Déposez votre rapport final sur Moodle. Assurez-vous de soumettre un fichier **zip** qui inclut tout le code source ainsi que les données.

Le rapport doit contenir au moins :

- Un premier chapitre présentant les caractéristiques du jeu de données avec des statistiques de base et des graphiques.
- Chaque résultat doit être expliqué. Notez que de mauvais résultats peuvent aussi être intéressants à présenter s'ils sont bien argumentés.
- Un deuxième chapitre sur la sélection/extraction de variables. Si cela n'est pas nécessaire pour votre jeu de données, justifiez pourquoi.
- Un troisième chapitre sur les résultats D'apprentissage, avec des explications sur les meilleurs résultats/meilleurs paramètres.

**Remarque** : Le nombre de pages n'est pas essentiel, assurez-vous simplement de présenter tout ce qui est demandé. N'oubliez pas d'indiquer votre nom dans le rapport et de fournir votre code Python ainsi que la version de Python utilisée.

## Versionnement du code sur GitHub

Afin d'assurer un suivi continu de votre travail, vous êtes tenus de :

- Créer un dépôt GitHub (public ou privé) pour votre projet.
- Écrire et versionner votre code en effectuant des **commits** réguliers.
- Effectuer ces **commits** aux différentes échéances de cours (ou selon les jalons définis dans votre calendrier de projet).
- Fournir l'accès (si le dépôt est privé) à votre enseignant pour consultation.

## Données

Vous devez choisir vos propres données. Les données doivent être hétérogènes et comporter un nombre considérable de variables.

**Bon Courage**