M2 Parcours SID - ECUE: Apprentissage automatique pour la science des données
Projet d'évaluation
Objectif :
Implémentation, évaluation et benchmarking d'un système de classification multiclasses en Machine Learning.

Mise en œuvre du projet
Le projet sera réalisé en groupes de maximum 4 étudiants. Il devra être implémenté en Python en utilisant les bibliothèques Keras, Scikit-Learn, et TensorFlow.

Étapes du projet :
1. **Sélection des données :**
    - Trouver un jeu de données contenant au moins **5000 instances** et **20 caractéristiques** (catégorielles, dates, numériques).
    - Le jeu de données doit être validé par l'enseignante avant son utilisation.
2. **Nettoyage des données :**
    - Identification et traitement des **valeurs manquantes**.
    - Sélection des variables pertinentes.
    - Traitement des données déséquilibrées si nécessaire.
3. **Sélection des variables :**
    - Fournir une liste des variables les plus utiles pour effectuer les analyses.
4. **Sélection du meilleur modèle :**
    - Entraînement avec une **grid search** et une **validation croisée (3-fold)** d'au moins **trois modèles différents**, dont au moins un modèle de réseaux de neurones.
5. **Évaluation de la performance du modèle :**
    - Choisir des métriques d'évaluation adaptées à l'objectif du système.
    - Présenter des **tableaux comparatifs de performance**.
    - Afficher l'**AUC** (Area Under Curve) pour tous les modèles.
6. **Justification des choix :**
    - Expliquer les traitements effectués.
    - Sélectionner le meilleur modèle et justifier ce choix.

Rendu du projet :
Soumission sur AMETICE :
1. **Présentation PowerPoint :**
    - Expliquer les objectifs du projet et le pipeline mis en place pour l'entraînement des modèles ainsi que les résultats obtenus.
2. **Jupyter Notebook :**
    - Fournir un notebook structuré et documenté contenant le code du projet.

Soutenance :
- Présentation orale de **10 minutes par équipe**, suivie de **5 minutes de questions**.

Évaluation :
Les critères suivants seront pris en compte dans l'évaluation :
1. **Pertinence des modèles ML implémentés :**

    - Par rapport au jeu de données et à l'objectif du projet.
2. **Qualité du prétraitement des données.**
3. **Pertinence des métriques d'évaluation utilisées.**
4. **Qualité de l'entraînement des modèles :**
    - Usage approprié de la grid search et de la validation croisée.
5. **Qualité de la présentation.**
6. **Qualité du code fourni.**