

NOMS:  
CLASSE:  
EQUIPE:

## TP1 : CLASSIFICATION

### KNN

L'objectif du TP est d'appliquer la méthode KNN pour prédire si une plante est comestible (0) ou toxique (1). Vous suivrez les étapes vues dans le cours : chargement des données, séparation, entraînement, test de plusieurs valeurs de K, calcul des métriques et analyse scientifique.

#### **Partie 1 — Téléchargement et chargement du dataset**

1. Charger le fichier plants.csv et afficher ses 5 premières lignes.

Question : Combien d'observations et de variables contient le dataset ?

Question : Que représente la variable 'class' ?

2. Séparer X et y.

Question : Pourquoi doit-on isoler la variable cible du reste

#### **Partie 2 — Préparation des données (train/test)**

1. Créer un ensemble d'entraînement et un ensemble de test contenant environ 80%/20% des données.

2. Vérifier les shapes de x\_train, y\_train, x\_test, y\_test.

Question : Pourquoi K ne doit pas dépasser la taille de x\_train

#### **Partie 3 — Entraînement du modèle**

I. Tester les valeurs de K suivantes : [1,3,5,7,9,11,13,15].

Pour chaque K : entraîner le modèle, prédire, calculer accuracy, precision, recall.

Question : Quelle métrique est prioritaire dans un problème de toxicité ?

Pourquoi

## II. Découverte de GridSearchCV:

On vous donne ci-dessous un extrait de code utilisant GridSearchCV.

L'objectif est de voir comment scikit-learn cherche automatiquement le meilleur K.

Comparer le meilleur K automatique avec celui trouvé manuellement

```
from sklearn.model_selection import GridSearchCV

params = {"n_neighbors": [1,3,5,7,9,11,13,15]}

grid = GridSearchCV(KNeighborsClassifier(), params, cv=5)

grid.fit(x_train, y_train)

print("Meilleur K trouvé par GridSearch :", grid.best_params_)
```

## **Partie 4 — Analyse scientifique**

Afficher accuracy, precision, recall pour chaque K.

Question : Pour quelles valeurs de K observe-t-on le plus de faux négatifs

## **Partie 5 — Courbes de performance**

Tracer sur les graphiques : Accuracy(K), Precision(K), Recall(K).

Question : Pourquoi la courbe du Recall est-elle essentielle dans un problème de toxicité ?

Question : Quel est le meilleur K selon Recall et selon l'accur