

# TP Machine Learning Construire un modèle de Machine Learning à partir de zéro avec le jeu de données Iris

## Objectif :

L'objectif de ce TP est d'appliquer plusieurs algorithmes de classification supervisée sur la célèbre base de données Iris. Vous devrez utiliser Decision Tree, Random Forest, KNN (K-Nearest Neighbors), Logistic Regression, et Naive Bayes pour prédire les espèces de fleurs d'Iris en fonction de leurs caractéristiques. Vous comparerez ensuite leurs performances.

## Prérequis :

- Connaissance de base en Python.
- Connaissance des bibliothèques : Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn.
- Connaissance de base des algorithmes de Machine Learning.

## Fichier de données :

Vous utiliserez un fichier CSV contenant les données d'Iris (fichier fourni). Ce fichier contient 150 lignes avec 4 caractéristiques et une colonne de classe pour les espèces d'Iris :

- Caractéristiques :
  1. Sepal\_length (longueur du sépale en cm)
  2. Sepal\_width (largeur du sépale en cm)
  3. Petal\_length (longueur du pétale en cm)
  4. Petal\_width (largeur du pétale en cm)
- Classe (cible) :

Espèce : Iris-setosa, Iris-versicolor, Iris-virginica

## Partie I : Préparation des données :

Question 1 : Importation et chargement de la dataset.

- Importer toutes les bibliothèques nécessaires
- Charger le jeu de données Iris avec la fonction `pd.read_csv()`

Question 2 : Analyser exploratoire des données.

- Affichez les 10 premières lignes du jeu de données.
- Vérifiez s'il y a des valeurs manquantes.
- Tracez la distribution de chaque caractéristique (longueur et largeur des sépales, longueur et largeur des pétales).

### Question 3 : Prétraitement des données

- Vérifier les valeurs manquantes et les outliers.
- Diviser les données en caractéristiques (X) et cible (y).
- Séparer les données en ensembles d'entraînement et de test (80%-20%).

## Partie 2 : Implémentation des algorithmes de classification :

### Question 4 : Decision Tree.

- Initialiser et entraîner le modèle.
- Evaluer le modèle

### Question 5 : Random Forest

- Initialiser et entraîner le modèle
- Evaluer le modèle
- Faire des prédictions
- Comparer les performances avec le Decision Tree

### Question 6 : KNN (K-Nearest Neighbors).

- Initialiser et entraîner le modèle
- Evaluer le modèle
- Ajuster le nombre de voisins et comparer les résultats

### Question 7 : Logistic Regression.

- Initialiser et entraîner le modèle
- Evaluer le modèle
- Calculer les performances du modèle

### Question 8 : Naive Bayes

- Initialiser et entraîner le modèle
- Evaluer le modèle
- Observer les performances du Naive Bayes par rapport aux autres modèles

### Question 9 : Support Vector Machine (SVM).

- Initialiser et entraîner le modèle
- Evaluer le modèle

## Partie 3 : Comparaison des modèles :

- Comparer les résultats des différents modèles en termes de précision, matrice de confusion et temps d'exécution
- Réaliser un tableau récapitulatif des performances pour chaque algorithme.

## Partie 4 : Discussion

- Quels modèles semblent les plus adaptés pour ce type de données
- Quels sont les avantages et inconvénients de chaque algorithme dans le contexte de ce jeu de données.

## Partie 5 : Conclusion.