Exercice 3 : Classification avec un arbre de décision sur le jeu de donnée « Breast Cancer »

- L'objectif de cet exercice est d'utiliser un arbre de décision pour résoudre un problème de classification binaire. Nous allons utiliser le jeu de données **Breast Cancer** , disponible dans scikit-learn. Ce dataset contient des caractéristiques dérivées de biopsies de tumeurs mammaires, et l'objectif est de prédire si une tumeur est bénigne(1) ou maligne(0).
- **Étapes de l'Exercice:**

1. Charger les données : utiliser le jeu de données **load_breast_cancer** de scikit-learn.
2. Diviser les données : diviser les données en ensembles d'entraînement (train) et de test (test) avec une proportion de 70 % pour l'entraînement et 30 % pour le test.
3. Entraîner un modèle d'arbre de décision : utiliser le critère entropie pour construire l'arbre de décision.
4. Évaluer le modèle : calculer les métriques suivantes sur l'ensemble de test : accuracy, precision, recall et F1 Score.
5. Tester le modèle sur une nouvelle donnée (non vue lors de l'entraînement ou du test) et afficher les prédictions

Afin d'appliquer la validation croisée K-Fold, nous allons utiliser la méthode **cross_val_score** de scikit-learn. Cette méthode permet d'évaluer les performances du modèle en divisant les données en K plis et en entraînant/testant le modèle sur chaque pli. Cela donne une estimation plus robuste des performances du modèle, car toutes les données sont utilisées à la fois pour l'entraînement et la validation :

6. Ajouter la Validation Croisée K-Fold :
   - ✓ Diviser les données en K plis (par exemple, K=5 ou K=10).
   - ✓ Entraîner et évaluer le modèle sur chaque sous-ensemble.
   - ✓ Calculer la moyenne des métriques (accuracy, precision, recall, F1 score) sur tous les sous-ensembles.
7. Comparer les Résultats :
   - ✓ Comparer les résultats obtenus avec la validation croisée aux résultats obtenus avec la division simple (train-test split).