

TP Final à rendre

Analyse de Clustering et de Classification sur le Jeu de Données des Maladies Cardiaques

I. Jeu de Données :

✓ **Jeu de Données :** [Heart Attack UCI](#)

✓ **Description :** Ce jeu de données comprend divers attributs liés aux maladies cardiaques. L'objectif est de prédire la présence ou l'absence de maladies cardiaques.

II. Tâches de Classification :

1. Préparation des Données :

- Utiliser le jeu de données des maladies cardiaques avec la variable cible indiquant la présence ou l'absence de maladies cardiaques.
- Effectuer le prétraitement nécessaire des données, traiter les valeurs manquantes, encoder les variables catégorielles, etc.

2. Algorithmes de Classification :

a. Forêt Aléatoire :

- Implémenter un classifieur de forêt aléatoire.
- Visualiser l'importance des caractéristiques.
- Ajuster les hyperparamètres si nécessaire.
- Évaluer le classifieur.

b. AdaBoost :

- Implémenter un classifieur AdaBoost.
- Visualiser l'importance des classificateurs faibles.
- Évaluer le classifieur.

c. Machine à Vecteurs de Support (SVM) :

- Implémenter un classifieur SVM.
- Visualiser la séparation des classes dans l'espace des caractéristiques.
- Ajuster les hyperparamètres si nécessaire.
- Évaluer le classifieur.

III. Tâches de Clustering :

1. Chargement des Données et Visualisation :

- Charger le jeu de données sur les maladies cardiaques.
- Visualiser le jeu de données dans une grille 2D avec des points en utilisant des caractéristiques pertinentes.

2. Algorithmes de Clustering :

a. k-Means Clustering :

- Appliquer k-Means avec un nombre prédéfini de clusters
- Visualiser les résultats du clustering.
- Explorer les métriques (silhouette score, etc.) pour évaluer la qualité du clustering.
- Itérer pour trouver le nombre optimal de clusters en utilisant les métriques d'évaluation. Mesurer le temps de calcul.
- Expliquer les limites de K-means avec un exemple de jeu de données

VI. Analyse Comparative :

1. Analyse de la classification

- Comparer les performances des différents classifieurs.
- Analyser les forces et les faiblesses de chaque classifieur.
- Fournir des insights sur les classifieurs qui se comportent mieux sur le jeu de données des maladies cardiaques.

2. Analyse du clustering

- Comparer les performances des différents classifieurs.
- Analyser les forces et les faiblesses de chaque classifieur.
- Fournir des insights sur les classifieurs qui se comportent mieux sur le jeu de données des maladies cardiaques.