

# projet\_maladies\_cardiaques

March 19, 2024

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

### 0.0.2 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.

### 0.0.3 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. Arbre de Décision :

- Implémenter un classifieur d'arbre de décision.
- Visualiser l'arbre de décision généré.
- Évaluer le classifieur en utilisant des métriques (précision, rappel, etc.).

#### b. Forêt Aléatoire :

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

#### c. AdaBoost :

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

#### d. 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.

**e. K-Nearest Neighbors (KNN) :**

- Implémenter un classifieur KNN.
- Visualiser les régions de décision.
- Ajuster les hyperparamètres si nécessaire.
- Évaluer le classifieur.

**f. Naive Bayes :**

- Implémenter un classifieur Naive Bayes.
- Visualiser les probabilités a priori et postérieures.
- Évaluer le classifieur.

**0.0.4 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

**b. Clustering Hiérarchique Agglomératif :** - Appliquer le clustering hiérarchique agglomératif avec un nombre prédéfini de clusters. - Visualiser les dendrogrammes et les résultats du clustering. - Utiliser les métriques d'évaluation pour analyser la qualité du clustering. - Itérer pour trouver le nombre optimal de clusters. Mesurer le temps de calcul. - Expliquer les limites du clustering agglomératif avec un exemple de jeu de données

**3. Deep learning** - Utiliser les réseaux de neurones

**0.0.5 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.

### 3- Analyse de tous les algorithmes

- Choisir le modèle le plus performant

[ ]: