

Faculté des Sciences et Techniques

Université Moulay Ismail

ERRACHIDIA



DATA MINIG

Mini projet

Prédiction des maladies cardiaques :

Réaliser par : **BOUMAAZAH Omar**

Encadre Par:

Prof. Mohamed Sabiri

Introduction

Le data mining, ou fouille de données, est une discipline qui explore et analyse des grandes quantités de données pour en extraire des connaissances utiles. Ses principales étapes comprennent la définition des objectifs, la collecte et le prétraitement des données, l'application d'algorithmes, et l'interprétation des résultats. Les applications du data mining sont vastes, allant du marketing à la santé, en passant par la finance. Les principaux concepts incluent l'extraction de connaissances, le prétraitement des données, les algorithmes spécifiques à chaque tâche, et la considération éthique de la confidentialité des données. En somme, le data mining offre des outils puissants pour transformer les données en informations exploitables, mais nécessite une approche éthique et responsable. notre projet de prédiction des maladies cardiaques à l'aide du data mining est réalisé en utilisant Python dans Jupyter Notebook. voici une approche générale sur les étapes de notre code

1. Collecte des Données :

- Oua un ensemble de données disponibles

2. Importation des Bibliothèques :

- Importez des bibliothèques Python telles que
- `import numpy as np`
- `import pandas as pd`
- `import datetime`
- `from collections import Counter`
- `import os`
- `import pickle`
- `import io`
- `import time`

- `import seaborn as sns`
- `import matplotlib.pyplot as plt`

3. Exploration des Données :

- Effectuez une exploration des données pour comprendre la structure et les caractéristiques des données.
- Visualisez les distributions, les corrélations et d'autres aspects importants.

4. Prétraitement des Données :

- Gérez les valeurs manquantes.
- Effectuez la normalisation ou la standardisation des caractéristiques si nécessaire.
- Encodez les variables catégorielles.

5. Division des Données :

- Divisez les données en ensembles d'entraînement et de test.

6. Choix du Modèle :

- Sélectionnez un algorithme de prédiction adapté aux problèmes de classification, comme les forêts aléatoires,
- les machines à vecteurs de support (SVM), ou les réseaux de neurones.

7. Entraînement du Modèle :

- Entraînez le modèle sur l'ensemble d'entraînement.

8. Évaluation du Modèle :

- Évaluez la performance du modèle sur l'ensemble de test en utilisant des métriques telles que la précision, le rappel, la F1-score, et la matrice de confusion.

9. Optimisation du Modèle :

- Optimisez les hyperparamètres du modèle pour améliorer la performance.

10. Interprétation des Résultats :

- Analysez les résultats pour comprendre les caractéristiques importantes dans la prédiction des maladies cardiaques.

11. Visualisation :

- Utilisez des graphiques et des visualisations pour présenter les résultats de manière compréhensible