TP 4 : Heart Disease Prediction with Decision Trees

Problem:

Predict whether a patient has heart disease based on clinical and demographic features.

Data:

Use the "Heart Disease" dataset from the UCI Machine Learning Repository: Heart Disease Dataset.

Tasks:

1. Data Exploration and Pre-processing:
   - Load the "Heart Disease" dataset.
   - Explore the dataset to understand its structure and features.
   - Pre-process the data by handling missing values, encoding categorical features, and scaling numerical features.
2. Train a Decision Tree Model:
   - Split the dataset into training and testing sets.
   - Implement a decision tree model using a suitable library.
3. Model Evaluation:
   - Evaluate the decision tree model's performance using metrics such as accuracy, precision, recall, and F1 score.
   - Visualize the confusion matrix to understand the distribution of correct and incorrect classifications.
4. Visualize the Decision Tree:
   - Visualize the decision tree to understand the decision rules learned by the model.
   - Discuss the key decision points and leaf nodes
5. Feature Importance Analysis:
   - Analyze the importance of different features in the decision-making process.
   - Discuss how certain features influence the prediction of heart disease.
6. Bonus: Prediction for New Patients:
   - Acquire or generate information for new patients.
   - Use the trained decision tree model to predict whether these patients have heart disease.

- Discuss potential challenges and considerations for deploying the model to new data.

Deliverables:

- Collab Notebook containing code for data exploration, pre-processing, model training, and evaluation.
- Visualizations illustrating the dataset features, confusion matrix, and any relevant insights.
- A report summarizing key findings, insights from the decision tree analysis, and recommendations for deploying the model.

This TP provides a practical experience in healthcare analytics using decision trees with a dataset that is relevant to a critical healthcare condition. It also offers insights into feature importance in predicting heart disease.