DECISION TREE ASSIGNMENT

Problem Statement:
This assignment aims to utilize Decision Tree algorithms to analyze given datasets, extracting meaningful insights and predicting outcomes based on the decision rules learned from the data. By exploring the datasets and employing Decision Tree techniques, students are expected to build predictive models and interpret the decision rules to derive valuable insights.

Guidelines:
1. Foundational Knowledge:
  - Understand the principles of decision trees and how they partition data based on attribute values.
  - Familiarize yourself with Decision Tree algorithms.
  - Recognize different splitting criteria and their impact on decision tree construction.

2. Data Exploration:
  - Analyze the dataset's structure and characteristics using various exploratory techniques such as histograms, scatter plots, and correlation matrices.
  - Gain insights into the dataset's attributes to guide the decision tree modeling process.

3. Preprocessing and Feature Engineering:
  - Handle missing values and categorical variables appropriately.
  - Encode categorical variables if necessary.
  - Split the dataset into training and testing sets.

4. Decision Tree Construction:
  - Choose appropriate hyperparameters such as maximum depth, minimum samples per leaf, and splitting criteria based on data exploration.
  - Implement Decision Tree algorithms using chosen parameters.
  - Train the Decision Tree model on the training data.

5. Model Evaluation:
  - Evaluate the trained model using appropriate metrics such as accuracy, precision, recall, and F1-score.
  - Visualize the decision tree to interpret the learned decision rules and feature importance.

6. Hyperparameter Tuning and Model Optimization:
  - Perform hyperparameter tuning using techniques like grid search or random search to optimize model performance.
  - Validate the optimized model using cross-validation techniques.

Step-by-Step Approach to Decision Tree Modeling:

1. Setup and Data Preparation:
  - Import necessary libraries: pandas, matplotlib, scikit-learn.
  - Load the dataset for decision tree modeling.
  - Preprocess the data, handle missing values, and encode categorical variables.

2. Decision Tree Parameters:
  - Choose appropriate hyperparameters such as maximum depth, minimum samples per leaf, and splitting criteria based on data exploration.

3. Building the Decision Tree:
  - Initialize the Decision Tree model with selected parameters.
  - Train the Decision Tree model on the prepared training data.

4. Model Evaluation:
  - Evaluate the trained model using appropriate metrics such as accuracy, precision, recall, and F1-score.
  - Visualize the decision tree to interpret the learned decision rules and feature importance.

5. Hyperparameter Tuning and Optimization:
  - Perform hyperparameter tuning using techniques like grid search or random search to optimize model performance.
  - Validate the optimized model using cross-validation techniques.

Links to Datasets for the Assignment:
- **Obesity Classification Dataset**
[https://www.kaggle.com/datasets/sujithmandala/obesity-classification-dataset]

- **Email Spam Classification Dataset**
[https://www.kaggle.com/datasets/balaka18/email-spam-classification-dataset-csv]

- **Heart Attack Analysis & Prediction Dataset**
[https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset]