Assignment: Decision Tree Classification using Gini Index and Information Gain

 Objective:

To understand and implement Decision Tree classifiers using **Gini Index** and **Information Gain (Entropy)** as splitting criteria, and evaluate their performance using a real-world dataset.

 Instructions:

1. **Dataset**: Use the **Car Evaluation** dataset from the UCI Repository
   URL: https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.data

2. **Tasks to Complete**:

Part A: Data Preprocessing

- Load the dataset using pandas.

- Assign appropriate column names: ['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class']

- Encode categorical features using LabelEncoder.

- Split the dataset into training and testing sets (80/20).

Part B: Decision Tree Classifier (Gini)

- Train a Decision Tree classifier using **Gini Index**.

- Predict and evaluate the model using **accuracy score**.

- Visualize the trained tree using plot_tree.

Part C: Decision Tree Classifier (Entropy)

- Train another Decision Tree classifier using **Information Gain (Entropy)**.

- Evaluate and compare its performance with the Gini-based tree.

- Visualize the second tree as well.

Part D: Analysis & Report

- Compare the accuracy of both trees.

- Discuss differences in tree structure and any overfitting.

- Provide interpretations for splits and decisions made by the tree.

Deliverables:

1. A **Jupyter Notebook** or Python script with:

   o   Clean and commented code

   o   Visualizations of both trees

   o   Accuracy scores and comparison

2. A brief **written report (PDF or Word)** (1–2 pages) covering:

   o   Dataset overview

   o   Your implementation steps

   o   Comparison between Gini and Entropy-based trees

   o   Conclusion on which performed better and why