PROJECT-6

Project Description:

In this project, we aim to build a classification model to predict the safety of cars based on various attributes. The dataset used contains information on car attributes like buying price, maintenance cost, number of doors, capacity, trunk size, and safety features, along with the safety classification as 'unacceptable', 'acceptable', 'good', or 'very good'.

Summary:

The project involves the following steps:

1. **Data Exploration and Preprocessing:**
   - Loaded a CSV dataset containing car evaluation data into a Pandas data frame.
   - Renamed columns for clarity and identified the data types (all categorical).
   - Explored the dataset's structure, dimensions, and checked for missing values (none found).
   - Analyzed the frequency distribution of categorical variables.

2. **Feature Engineering:**
   - Prepared the feature vector (X) and target variable (y) for model training.
   - Encoded categorical variables using ordinal encoding to prepare them for machine learning algorithms.

3. **Model Building:**
   - Split the dataset into training and testing sets using a 67:33 split ratio.
   - Utilized Decision Tree Classifier models with both Gini index and entropy criteria.

o   Trained the models on the training data and evaluated their performance on the test data.
o   Assessed model accuracy and checked for overfitting by comparing training and test set accuracies.

4. **<u>Model Evaluation:</u>**
    o   Visualized decision trees generated by the models to understand their decision-making processes.
    o   Employed confusion matrix and classification report to delve deeper into model performance, identifying types of errors made by the classifier.
    o   Evaluated precision, recall, and F1-score metrics for each class label to gauge overall model effectiveness.
5. **<u>Conclusion:</u>**
    o   Both Decision Tree models (using Gini index and entropy) achieved comparable accuracy scores of approximately 80.21% on the test set.
    o   The training and test set accuracies were consistent, suggesting no significant overfitting.
    o   The confusion matrix highlighted areas where the model excelled (e.g., predicting 'unacceptable' cars correctly) and areas for improvement (e.g., predicting 'good' and 'very good' cars).
    o   Overall, the models demonstrated strong predictive performance in classifying car safety based on provided attributes.

This project showcases the application of Decision Tree classifiers for multiclass classification tasks using Python's machine learning ecosystem, emphasizing data preprocessing, model training, evaluation, and interpretation of results.