Activity: #2

Lab Title: Applying Decision Tree Classifier on a Real-Life Scenario Using Python

Learning Objectives:

By the end of this lab, students will be able to:
- Understand the working principles of Decision Trees.
- Implement the Decision Tree algorithm using Python and libraries like scikit-learn.
- Evaluate the performance of Decision Tree models on real-world datasets.

Instructions to Students:

- **Application Title Registration**:
    - Choose a real-world classification problem (e.g., fraud detection in banking, disease classification in healthcare).
    - Register the application title with the course tutor. The same application area cannot be repeated among students.
- **Dataset Selection**:
    - Select a relevant dataset with enough features for a meaningful decision tree.
    - Ensure the dataset is suitable for classification problems.
- **Decision Tree Implementation**:
    - Implement the Decision Tree algorithm in Python using scikit-learn.
    - Apply appropriate preprocessing techniques like encoding categorical variables and handling missing values.
    - Visualize the Decision Tree structure.
- **Model Fitting and Testing**:
    - Fit the model on the training data and test its performance on a test set.
    - Analyze the effect of tree depth and other parameters.
- **Performance Evaluation**:
    - Evaluate the model's performance using metrics like accuracy and interpret confusion matrix.

Rubrics:

- **Dataset (0.25 Mark)**:
    - Relevant dataset used, with sufficient features for classification.
- **Algorithm (0.5 Mark)**:
    - Correct implementation of Decision Tree with parameters like tree depth, criterion, and feature splitting.
- **Fitting and Visualization (0.5 Mark)**:
    - Model trained and tested, with analysis of tree depth effects on performance.
- **Performance (0.25 Mark)**