Week 2: Supervised Learning

Task 2.2: Decision Trees and Random Forests

Objective: Learn to implement tree-based methods for classification and compare their performance.

Dataset: **Loan Prediction Dataset from Kaggle**. This dataset involves predicting loan approval based on attributes like gender, marital status, income, and loan amount.

- **Link to dataset**: https://www.kaggle.com/code/bhavikbb/loan-prediction-dataset/output

Activities:

1. **Model Implementation**:
    - Build a decision tree model and visualize the tree.
    - Implement a random forest to improve model performance and reduce overfitting.
2. **Performance Comparison**:
    - Compare the performance of the decision tree and random forest using accuracy, F1-score, and other relevant metrics.

Expected Output:

- A Jupyter notebook with implementations and performance comparisons.
- Visuals for tree diagrams and performance metrics.

Documentation:

- Detail the decisions made during the model building, including why certain parameters were chosen and the reasons for choosing tree-based methods over others.

General Guidelines for Tasks:

- **Code Quality**: Write clean, modular code.
- **Documentation**: Use the provided template to maintain consistency in documentation.