Assignment 1

Assignment Overview

This assignment will guide you through implementing and comparing the performance of three machine learning models—**K-Nearest Neighbors (KNN)**, **Decision Trees**, and **Random Forest**—using the **Breast Cancer dataset** from sklearn.

You will:

1. Load and preprocess the dataset, including feature scaling for KNN.
2. Train and evaluate each model using metrics such as accuracy, precision, recall, and F1-score.
3. Explore the impact of hyperparameter tuning on model performance.
4. Analyze and compare the models in a written report.
5. Submit your code via GitHub and a report summarizing your work.

Instructions

1. **Dataset**:
   Use the **Breast Cancer dataset** provided by sklearn. It includes 30 features and a binary classification task (malignant vs. benign).
2. **Tasks**:
   - **Data Preprocessing**:
     - Load the Breast Cancer dataset using `load_breast_cancer` from sklearn.
     - Partition the data into an 80% training set and a 20% test set.
     - Scale the features using `StandardScaler` for KNN.
   - **Model Training**:
     - Train three classifiers:
       1. **K-Nearest Neighbors (KNN)**: Start with `n_neighbors=5`.
       2. **Decision Tree**: Use the default settings initially, then experiment with `max_depth`.
       3. **Random Forest**: Start with 100 trees (`n_estimators=100`) and explore the effect of different `max_depth` or `min_samples_split`.
   - **Evaluation**:
     - Use the following metrics to evaluate performance:
       - Accuracy
       - Precision
       - Recall
       - F1-score
     - Include a confusion matrix for each model.

        - Compare the results across the models in a tabular or graphical format.
    - **Ablation Study**:
        - Modify key hyperparameters (e.g., `n_neighbors` for KNN, `max_depth` for Decision Trees and Random Forest) and observe the impact on performance.
3. **Deliverables**:
    1. **Code Submission**:
        - Upload all your code to a GitHub repository. Provide the repository link in your report. Ensure your code is well-documented with comments.
    2. **Report Submission**:
        - Write a maximum 4-page report. Submit the report as a PDF file through the course platform.