ML Tutorial-02: Dimensionality Reduction and Model Evaluation with the Iris Dataset

Objective: Gain hands-on experience with PCA and LDA dimensionality reduction techniques and evaluate their effects on machine learning model performance using the Iris dataset.

1. Data Preparation:

- Load the Iris dataset (sklearn.datasets.load_iris).
- Split into 70% training and 30% testing sets (train_test_split).
- Standardize features using StandardScaler.

2. Model Training & Evaluation (No Dimensionality Reduction): Train and evaluate the following models on the standardized dataset:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)

Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

3. Dimensionality Reduction with PCA:

- Apply PCA(n_components=2) on training data.
- Transform both training and test data.
- Train the same models listed in section 2.
- Evaluate using the same metrics.

4. Dimensionality Reduction with LDA:

- Apply LinearDiscriminantAnalysis(n_components=2) on training data.
- Transform both training and test data.
- Train and evaluate the same models using the same metrics.

5. Comparison and Analysis:

- Compare model performance across three scenarios:
    - Without dimensionality reduction
    - With PCA
    - With LDA
- Analyze the effect of dimensionality reduction on each metric.
- Provide observations and insights on model behavior and dimensionality impact.