

MINI PROJECT

COMPARING FOUR CLASSIFICATION ALGORITHMS (IRIS DATASET)

Step 1 — Load and Prepare the Dataset

1. Import the built-in Iris dataset from sklearn.datasets.
2. Extract features (X) and labels (y).
3. Split dataset into 80% training and 20% testing using train_test_split().
4. Standardize data using StandardScaler() for KNN and SVM performance.

Step 2 — Logistic Regression

- Train Logistic Regression using all scaled features.
- Predict classes for the test set.
- Compute Accuracy and Classification Report.

Expected Accuracy: ~100%

Step 3 — K-Nearest Neighbors (KNN)

- KNN Classifier with k=5 neighbors.
- Sensitive to scaling — scaling improves accuracy.
- Predict test labels and compute metrics.

Expected Accuracy: ~100%

Step 4 — Support Vector Machine (SVM)

- Use SVM with RBF kernel.
- Performs well on small and clean datasets.
- Predict test labels and compute accuracy.

Expected Accuracy: ~100%

Step 5 — Random Forest Classifier

- Ensemble model using multiple decision trees.
- Robust and handles non-linear patterns well.
- Predict test labels and compute metrics.

Expected Accuracy: ~100%

Step 6 — Results Comparison

All four algorithms show excellent performance on the Iris dataset.

Summary (Typical Values)

- Logistic Regression: 100%
- KNN: 100%
- SVM: 100%
- Random Forest: 100%

Step 7 — Best Algorithm Justification

Although all models achieve similar accuracy, SVM is the best-suited due to:

- High stability
- Handles complex boundaries
- Low error margin

Conclusion: SVM is the most suitable algorithm for the Iris classification task.