

MINI PROJECT

Comparing the Accuracy Score and Classification Report of Four Algorithms: Logistic Regression, KNN, SVM, and Random Forest (Iris Dataset)

Step 1 — Load and Prepare the Dataset

1. Import the **Iris dataset** from `sklearn.datasets`.
 2. Store features in **X** and class labels in **y**.
 3. Split the dataset into **80% training** and **20% testing** using `train_test_split()`.
 4. Standardize the data using `StandardScaler()` to improve model performance, especially for **KNN** and **SVM**.
-

Step 2 — Logistic Regression

Procedure:

- Train Logistic Regression using all scaled features.
- Predict the class labels for test data.
- Compute Accuracy Score and Classification Report.

Output (Typical):

- Accuracy: **1.00 (100%)**
 - High precision, recall and F1-score for all classes.
-

Step 3 — K-Nearest Neighbors (KNN)

Procedure:

- Train KNN with **k = 5** neighbors.
- Use standardized dataset for better performance.
- Predict on the test data.
- Compute evaluation metrics.

Output (Typical):

- Accuracy: **1.00 (100%)**
- All classes classified correctly due to clear separation in Iris dataset.

Step 4 — Support Vector Machine (SVM)

Procedure:

- Use SVM with **RBF kernel**.
- Fit on scaled training data.
- Predict test labels.
- Compute accuracy and classification report.

Output (Typical):

- Accuracy: **1.00 (100%)**
 - SVM handles the Iris dataset extremely well due to clean separability.
-

Step 5 — Random Forest Classifier

Procedure:

- Train a Random Forest model with 100 trees.
- Predict the classes for test data.
- Evaluate using accuracy and classification report.

Output (Typical):

- Accuracy: **1.00 (100%)**
 - Ensemble method gives stable results.
-

Step 6 — Summary of Results

Algorithm	Accuracy
Logistic Regression	100%
KNN	100%
SVM	100%
Random Forest	100%

All four algorithms achieved perfect accuracy on the Iris dataset due to:

- ✓ Clean data
- ✓ Balanced classes

- ✓ Simple features
 - ✓ Clear boundaries between classes
-

Step 7 — Comparison and Interpretation

1. Logistic Regression

- Works well because data is almost linearly separable.

2. KNN

- Very effective with scaling.
- Performs well due to clustered data.

3. SVM

- Robust and stable.
- Handles non-linear decision boundaries.

4. Random Forest

- Ensemble approach provides very strong performance.
 - Slightly more complex than necessary for Iris dataset.
-

Step 8 — Best Algorithm Justification

Even though all models score 100%,

SVM is the MOST SUITABLE algorithm because:

- Works best on small datasets
 - Handles non-linearity effectively
 - More stable and less likely to overfit than KNN
 - Performs consistently even without parameter tuning
-

Conclusion

This mini project successfully compares four machine learning classifiers on the Iris dataset. Each model achieved excellent performance, but **Support Vector Machine (SVM)** stands out as the best due to its robustness and capability to generalize well.

