

Project: Task 14 - Model Comparison & Selection

Dataset: Breast Cancer Diagnostic Dataset

### 1. Models Evaluated

We tested four distinct algorithms to determine which generalized best for classifying tumors:

Logistic Regression: A linear baseline.

Decision Tree: A simple rule-based non-linear model.

Random Forest: An ensemble of trees to reduce overfitting.

Support Vector Machine (SVM): A distance-based margin classifier.

### 2. Performance Analysis

Accuracy: The Random Forest and SVM models typically tied for top performance (~96-98% accuracy).

Recall (Sensitivity): SVM achieved the highest recall. In a medical context, Recall is often the most critical metric because we must minimize "False Negatives" (failing to detect cancer in a sick patient).

Overfitting: The Decision Tree showed slightly lower test accuracy compared to its training score, indicating mild overfitting. The Random Forest corrected this by averaging multiple trees.

### 3. Final Verdict

Support Vector Machine (SVM) was selected as the best model.

Reason: It provided the best balance of Accuracy and Recall.

Robustness: With the linear kernel, it proved the data is well-separated, offering a stable and explainable boundary.