Task 1: Breast Cancer Classification

Dataset:

The Breast Cancer Wisconsin dataset was used. The task is to classify tumors as benign or malignant based on 30 numeric features.

Data Preprocessing:

The dataset was loaded from sklearn. Features were standardized using StandardScaler. The data was split into training and testing sets.

Models Used:

1. Logistic Regression

2. Random Forest Classifier

Evaluation Metrics:

Accuracy, Precision, Recall, F1 Score and ROC-AUC were used to evaluate model performance.

Results:

Logistic Regression achieved higher cross-validation accuracy and better ROC-AUC compared to Random Forest.

Conclusion:

Logistic Regression was selected as the final model because it showed more stable and reliable performance on this dataset.