Developing and Evaluating Classification Models

1. Dataset

Dataset: Breast Cancer Wisconsin Dataset (binary classification).

Features: 30 numeric features (mean radius, mean texture, etc.).

Target: Malignant (0) or Benign (1).

2. Models Compared

- Logistic Regression
- k-Nearest Neighbors (k=5)
- Decision Tree Classifier

3. Metrics Summary

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.96	0.97	0.97	0.97	0.99
k-NN (k=5)	0.95	0.95	0.97	0.96	0.98
Decision Tree	0.93	0.93	0.95	0.94	0.94

4. Observations

- Logistic Regression provided the best ROC-AUC and balanced performance.
- k-NN performed well but depends on feature scaling and k value.
- Decision Tree offers interpretability but may overfit.

5. Recommendation

Based on the evaluation, Logistic Regression is recommended for its stable, interpretable, and high-performance results across metrics, making it suitable for binary classification tasks like fraud detection or medical diagnosis.