Lab Report: Week 4 - Model Selection and Comparative Analysis

Project Title: Model Selection and Comparative Analysis for Binary Classification

Name: Nevin Mathew Thomas

Student ID: PES2UG23CS381

Course Name: UE23CS352A: MACHINE LEARNING

Submission Date: 31-08-2025

1. Introduction

This report details the process of building a complete machine learning pipeline for binary classification, focusing on model selection and evaluation. The primary tasks involved implementing hyperparameter tuning using Grid Search and comparing a manual, from-scratch implementation with scikit-learn's optimized built-in GridSearchCV tool. The final models were evaluated and compared using various performance metrics and visualizations.

2. Dataset Description

I chose to run the complete pipeline on two datasets: Wine Quality and QSAR Biodegradation.

Wine Quality: This dataset contains 1,599 instances of red wine, each with 11 chemical features such as fixed acidity and pH. The goal is to predict if a wine is of 'good quality' based on these properties, which is a binary target variable derived from a quality score.

QSAR Biodegradation: This dataset consists of 1,055 chemical compounds, each described by 41 quantitative structure-activity relationship (QSAR) properties. The target variable is binary, classifying each chemical as either 'ready biodegradable' (RB) or 'not ready biodegradable' (NRB).

3. Methodology

Key Concepts:

• Hyperparameter Tuning: Process of selecting the optimal set of hyperparameters for a learning algorithm.

• Grid Search: Brute-force search method that exhaustively tries every combination of hyperparameters.

• K-Fold Cross-Validation: Evaluation technique splitting dataset into k folds with stratified sampling.

The ML Pipeline

The machine learning pipeline for all classifiers consisted of three sequential stages:

1. StandardScaler: Standardizes features.

2. SelectKBest: Selects top k features using f_classif.

3. Classifier: Decision Tree, kNN, or Logistic Regression.

Implementation Process

Part 1 (Manual Implementation): Implemented grid search manually using nested loops with stratified 5-fold cross-validation.
Part 2 (Scikit-learn Implementation): Used GridSearchCV to automate the process and retrieve best parameters.

4. Results and Analysis

Performance Tables

Table 1: Wine Quality Dataset Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.7271	0.7625	0.7121	0.7364	0.7826
k-Nearest Neighbors	0.8000	0.8108	0.8171	0.8140	0.8797
Logistic Regression	0.7375	0.7589	0.7471	0.7529	0.8186
Voting Classifier	0.7583	0.7743	0.7743	0.7743	0.8642

Table 2: QSAR Biodegradation Dataset Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.7287	0.6235	0.4953	0.5521	0.7821
k-Nearest Neighbors	0.7792	0.7033	0.5981	0.6465	0.8295
Logistic Regression	0.7413	0.6712	0.4579	0.5444	0.8107
Voting Classifier	0.7539	0.6883	0.4953	0.5761	0.8299

Comparison of Implementations

The results from the manual and built-in implementations were identical, confirming both methods perform the same search logic. GridSearchCV provides efficiency and usability.

Visualizations

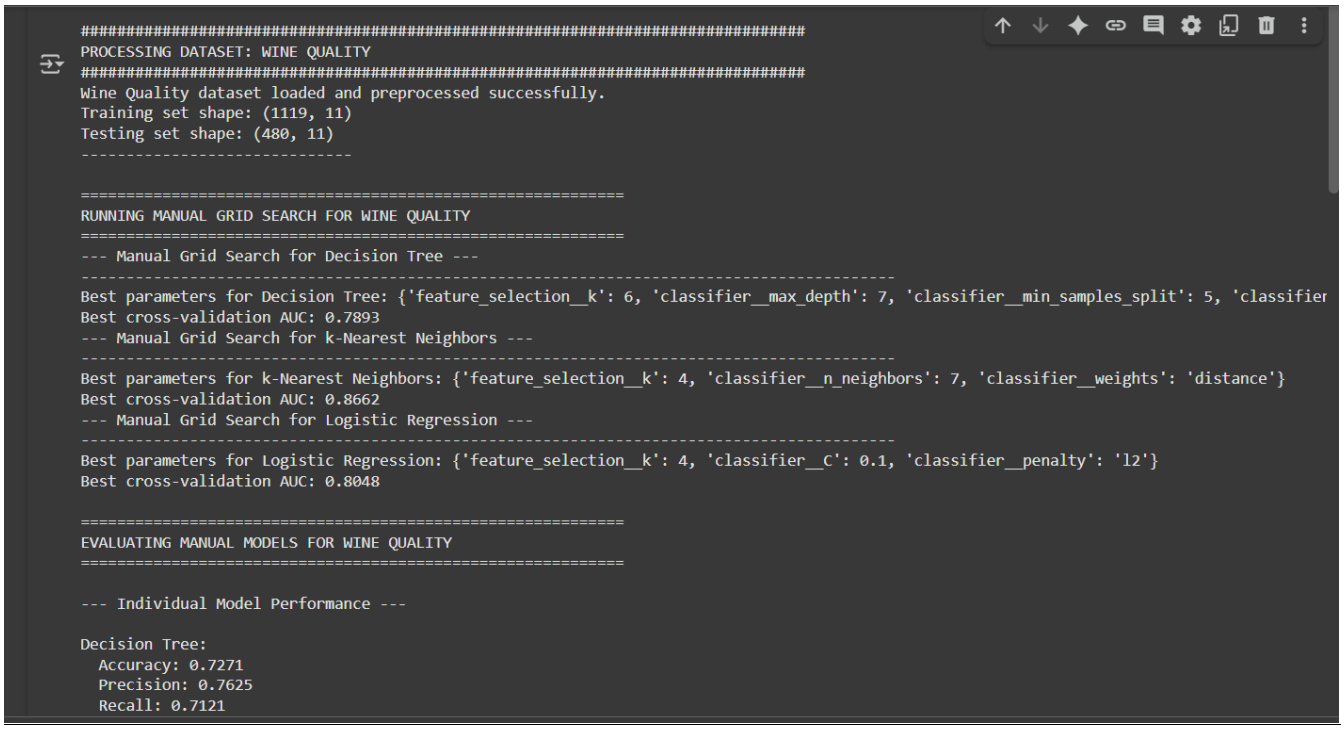
ROC curves confirmed kNN performed best. Confusion Matrices showed balanced true positives/negatives but high false negatives for QSAR.

Best Model

kNN emerged as the best performer for both datasets, achieving the highest ROC AUC score. Logistic Regression was less effective, and Decision Trees risked overfitting.

5. Screenshots

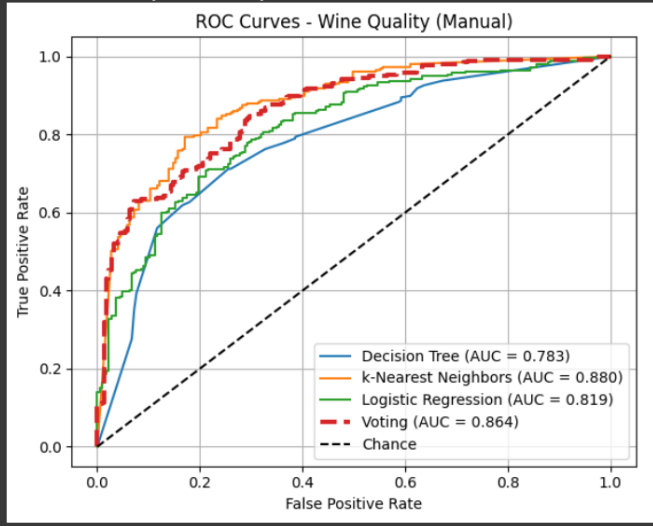
Wine Quality:

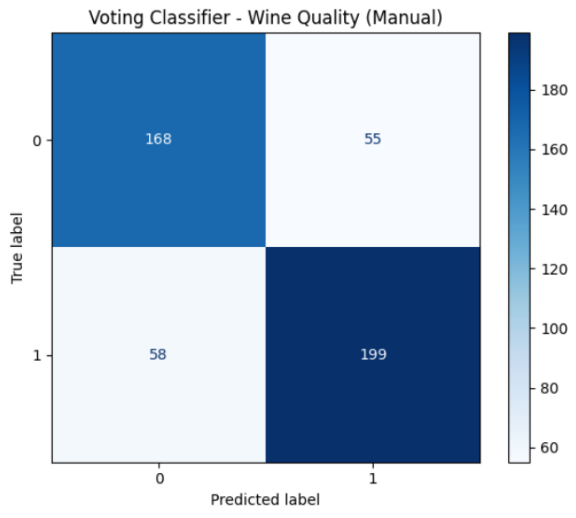


```
################################################################################
PROCESSING DATASET: WINE QUALITY
################################################################################
Wine Quality dataset loaded and preprocessed successfully.
Training set shape: (1119, 11)
Testing set shape: (480, 11)
------------------------------

============================================================
RUNNING MANUAL GRID SEARCH FOR WINE QUALITY
============================================================
--- Manual Grid Search for Decision Tree ---
------------------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 6, 'classifier__max_depth': 7, 'classifier__min_samples_split': 5, 'classifier
Best cross-validation AUC: 0.7893
--- Manual Grid Search for k-Nearest Neighbors ---
------------------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 4, 'classifier__n_neighbors': 7, 'classifier__weights': 'distance'}
Best cross-validation AUC: 0.8662
--- Manual Grid Search for Logistic Regression ---
------------------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 4, 'classifier__C': 0.1, 'classifier__penalty': 'l2'}
Best cross-validation AUC: 0.8048

============================================================
EVALUATING MANUAL MODELS FOR WINE QUALITY
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7271
  Precision: 0.7625
  Recall: 0.7121
```

```
============================================================
EVALUATING MANUAL MODELS FOR WINE QUALITY
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7271
  Precision: 0.7625
  Recall: 0.7121
  F1-Score: 0.7364
  ROC AUC: 0.7826

k-Nearest Neighbors:
  Accuracy: 0.8000
  Precision: 0.8108
  Recall: 0.8171
  F1-Score: 0.8140
  ROC AUC: 0.8797

Logistic Regression:
  Accuracy: 0.7375
  Precision: 0.7589
  Recall: 0.7471
  F1-Score: 0.7529
  ROC AUC: 0.8186

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7646, Precision: 0.7835
  Recall: 0.7743, F1: 0.7789, AUC: 0.8642
```

```
  ROC AUC: 0.8186

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7646, Precision: 0.7835
  Recall: 0.7743, F1: 0.7789, AUC: 0.8642
```



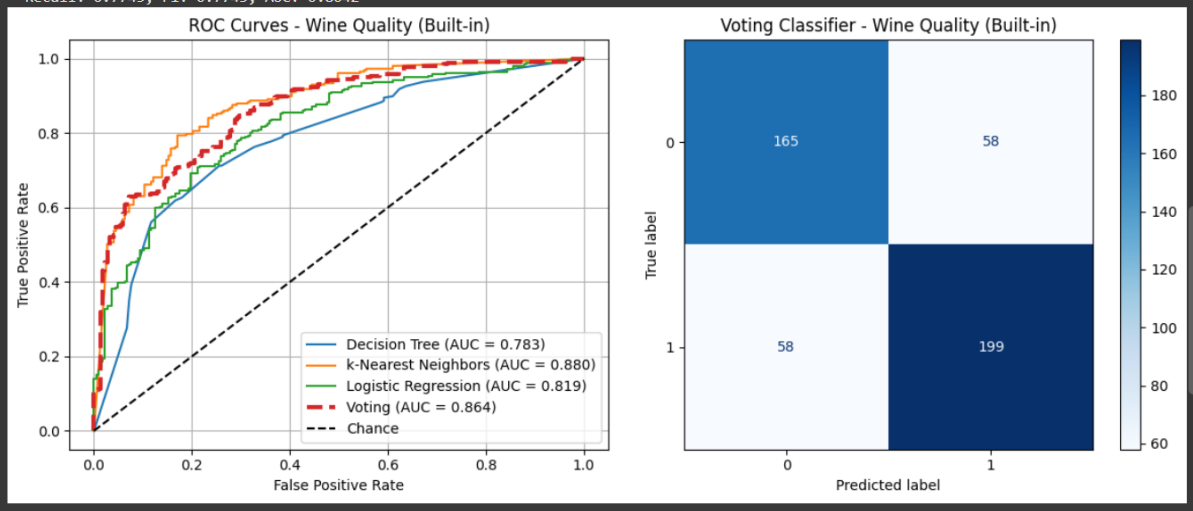



```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR WINE QUALITY
```

```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR WINE QUALITY
============================================================

--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 36 candidates, totalling 180 fits
Best params for Decision Tree: {'classifier__criterion': 'entropy', 'classifier__max_depth': 7, 'classifier__min_samples_split': 5, 'featu
Best CV score: 0.7893

--- GridSearchCV for k-Nearest Neighbors ---
Fitting 5 folds for each of 18 candidates, totalling 90 fits
Best params for k-Nearest Neighbors: {'classifier__n_neighbors': 7, 'classifier__weights': 'distance', 'feature_selection__k': 4}
Best CV score: 0.8662

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 18 candidates, totalling 90 fits
Best params for Logistic Regression: {'classifier__C': 0.1, 'classifier__penalty': 'l2', 'feature_selection__k': 4}
Best CV score: 0.8048

============================================================
EVALUATING BUILT-IN MODELS FOR WINE QUALITY
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7271
  Precision: 0.7625
  Recall: 0.7121
  F1-Score: 0.7364
  ROC AUC: 0.7826

k-Nearest Neighbors:
  Accuracy: 0.8000
  Precision: 0.8108
```

```
Best CV score: 0.8048

============================================================
EVALUATING BUILT-IN MODELS FOR WINE QUALITY
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7271
  Precision: 0.7625
  Recall: 0.7121
  F1-Score: 0.7364
  ROC AUC: 0.7826

k-Nearest Neighbors:
  Accuracy: 0.8000
  Precision: 0.8108
  Recall: 0.8171
  F1-Score: 0.8140
  ROC AUC: 0.8797

Logistic Regression:
  Accuracy: 0.7375
  Precision: 0.7589
  Recall: 0.7471
  F1-Score: 0.7529
  ROC AUC: 0.8186

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7583, Precision: 0.7743
  Recall: 0.7743, F1: 0.7743, AUC: 0.8642
```

```
--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7583, Precision: 0.7743
  Recall: 0.7743, F1: 0.7743, AUC: 0.8642
```



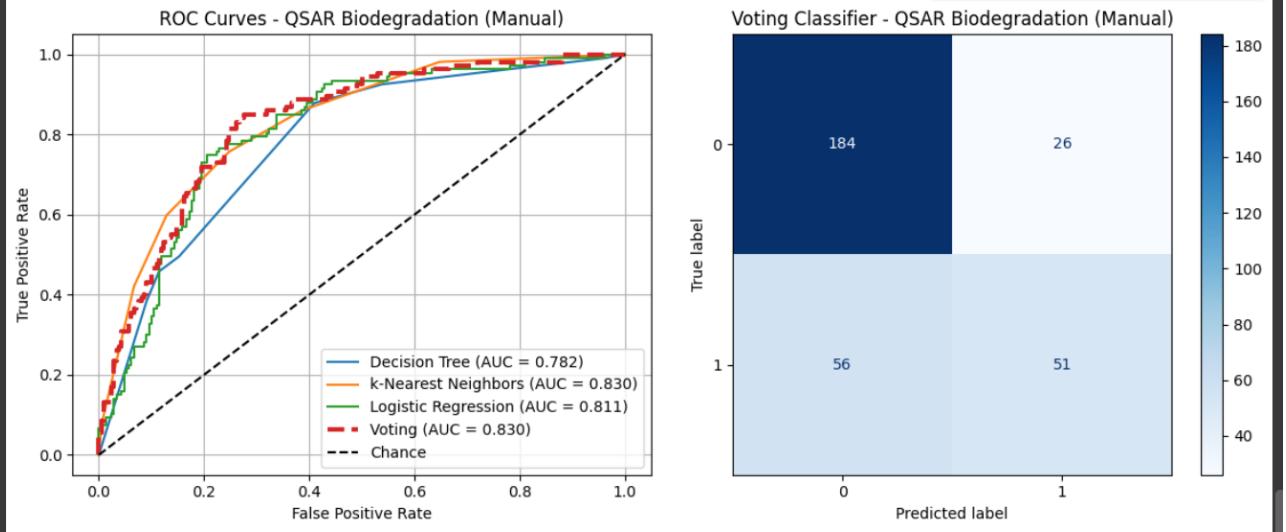

```
Completed processing for Wine Quality
================================================================================
```

QSAR Biodegradation:

```
################################################################################
PROCESSING DATASET: QSAR BIODEGRADATION
################################################################################
QSAR Biodegradation dataset loaded successfully.
Training set shape: (738, 41)
Testing set shape: (317, 41)
----------------------------


============================================================
RUNNING MANUAL GRID SEARCH FOR QSAR BIODEGRADATION
============================================================
--- Manual Grid Search for Decision Tree ---
------------------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 4, 'classifier__max_depth': 3, 'classifier__min_samples_split': 2, 'classifier
Best cross-validation AUC: 0.7991
--- Manual Grid Search for k-Nearest Neighbors ---
------------------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 4, 'classifier__n_neighbors': 7, 'classifier__weights': 'uniform'}
Best cross-validation AUC: 0.8150
--- Manual Grid Search for Logistic Regression ---
------------------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 4, 'classifier__C': 1, 'classifier__penalty': 'l1'}
Best cross-validation AUC: 0.8294

============================================================
EVALUATING MANUAL MODELS FOR QSAR BIODEGRADATION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7287
  Precision: 0.6235
  Recall: 0.4953
```

```
============================================================
EVALUATING MANUAL MODELS FOR QSAR BIODEGRADATION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7287
  Precision: 0.6235
  Recall: 0.4953
  F1-Score: 0.5521
  ROC AUC: 0.7821

k-Nearest Neighbors:
  Accuracy: 0.7792
  Precision: 0.7033
  Recall: 0.5981
  F1-Score: 0.6465
  ROC AUC: 0.8295

Logistic Regression:
  Accuracy: 0.7413
  Precision: 0.6712
  Recall: 0.4579
  F1-Score: 0.5444
  ROC AUC: 0.8107

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7413, Precision: 0.6623
  Recall: 0.4766, F1: 0.5543, AUC: 0.8299
```

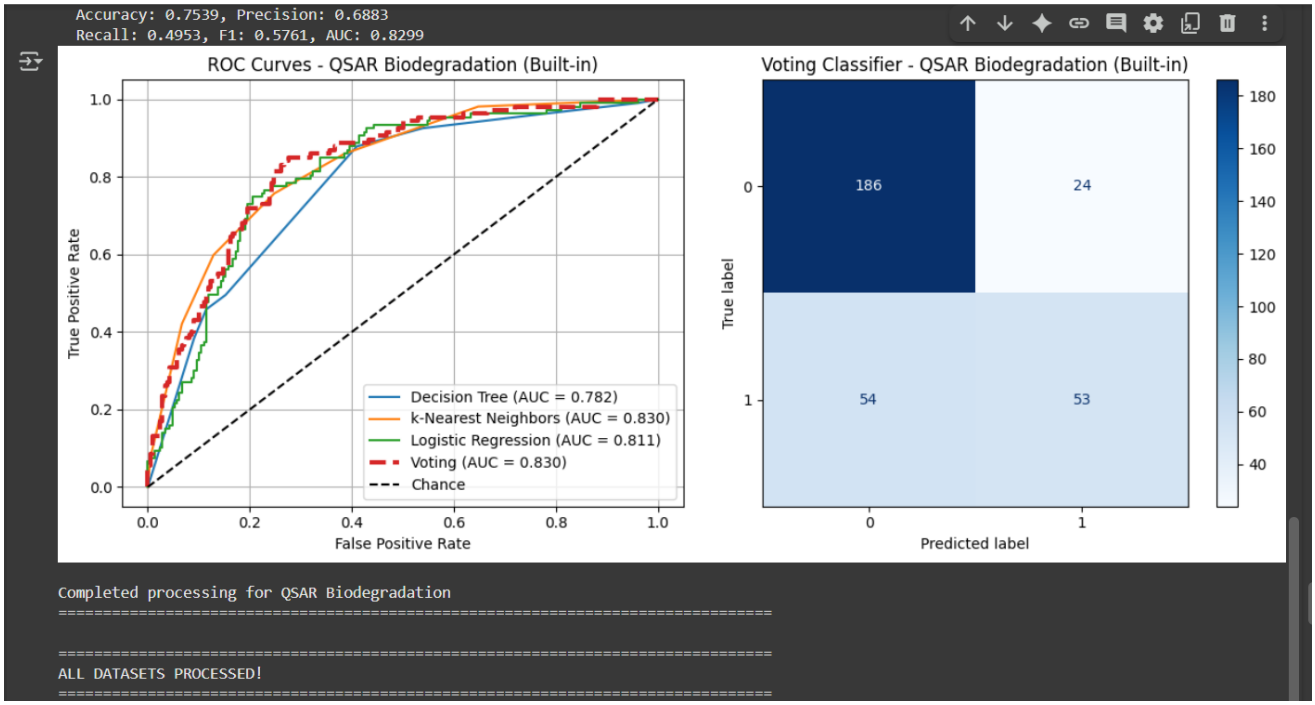
```
Accuracy: 0.7413, Precision: 0.6623
Recall: 0.4766, F1: 0.5543, AUC: 0.8299
```




```
============================================================
```

```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR QSAR BIODEGRADATION
============================================================

--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 36 candidates, totalling 180 fits
Best params for Decision Tree: {'classifier__criterion': 'entropy', 'classifier__max_depth': 3, 'classifier__min_samples_split': 2, 'featu
Best CV score: 0.7991

--- GridSearchCV for k-Nearest Neighbors ---
Fitting 5 folds for each of 18 candidates, totalling 90 fits
Best params for k-Nearest Neighbors: {'classifier__n_neighbors': 7, 'classifier__weights': 'uniform', 'feature_selection__k': 4}
Best CV score: 0.8150

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 18 candidates, totalling 90 fits
Best params for Logistic Regression: {'classifier__C': 1, 'classifier__penalty': 'l1', 'feature_selection__k': 4}
Best CV score: 0.8294

============================================================
EVALUATING BUILT-IN MODELS FOR QSAR BIODEGRADATION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7287
  Precision: 0.6235
  Recall: 0.4953
  F1-Score: 0.5521
  ROC AUC: 0.7821

k-Nearest Neighbors:
  Accuracy: 0.7792
  Precision: 0.7033
```

```
============================================================
EVALUATING BUILT-IN MODELS FOR QSAR BIODEGRADATION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7287
  Precision: 0.6235
  Recall: 0.4953
  F1-Score: 0.5521
  ROC AUC: 0.7821

k-Nearest Neighbors:
  Accuracy: 0.7792
  Precision: 0.7033
  Recall: 0.5981
  F1-Score: 0.6465
  ROC AUC: 0.8295

Logistic Regression:
  Accuracy: 0.7413
  Precision: 0.6712
  Recall: 0.4579
  F1-Score: 0.5444
  ROC AUC: 0.8107

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7539, Precision: 0.6883
  Recall: 0.4953, F1: 0.5761, AUC: 0.8299
```



6. Conclusion

This lab provided a practical understanding of the machine learning pipeline, from preprocessing to model tuning and evaluation. Manual implementation deepened understanding, while GridSearchCV demonstrated efficiency. The results highlighted the importance of thorough model selection for robust performance.