PES2UG23CS346

MOHAMMED MIR FAZLAI ALI

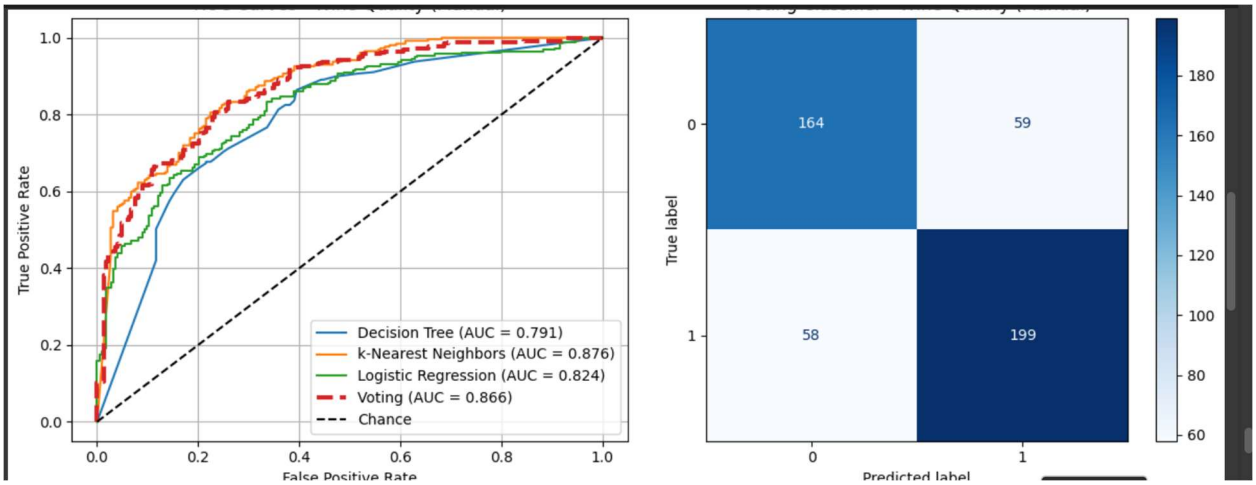
OUTPUT SCREENSHOTS

```
################################################################################
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 ---
Testing 135 parameter combinations...
  Tested 10/135 combinations. Current best AUC: 0.7796
  Tested 20/135 combinations. Current best AUC: 0.7846
  Tested 30/135 combinations. Current best AUC: 0.7846
  Tested 40/135 combinations. Current best AUC: 0.7850
  Tested 50/135 combinations. Current best AUC: 0.7850
  Tested 60/135 combinations. Current best AUC: 0.7850
  Tested 70/135 combinations. Current best AUC: 0.7850
  Tested 80/135 combinations. Current best AUC: 0.7850
  Tested 90/135 combinations. Current best AUC: 0.7850
  Tested 100/135 combinations. Current best AUC: 0.7850
  Tested 110/135 combinations. Current best AUC: 0.7850
  Tested 120/135 combinations. Current best AUC: 0.7850
  Tested 130/135 combinations. Current best AUC: 0.7850
  Tested 135/135 combinations. Current best AUC: 0.7850
------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 5, 'classifier__max_depth': 10, 'classifier__min_samples_split': 10, 'classifier__min_sam
Best cross-validation AUC: 0.7850
--- Manual Grid Search for k-Nearest Neighbors ---
Testing 60 parameter combinations...
  Tested 10/60 combinations. Current best AUC: 0.8579
  Tested 20/60 combinations. Current best AUC: 0.8696
  Tested 30/60 combinations. Current best AUC: 0.8696
  Tested 40/60 combinations. Current best AUC: 0.8696
  Tested 50/60 combinations. Current best AUC: 0.8696
  Tested 60/60 combinations. Current best AUC: 0.8696
--------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 5, 'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'classifier__m
Best cross-validation AUC: 0.8696
--- Manual Grid Search for Logistic Regression ---
Testing 30 parameter combinations...
  Tested 10/30 combinations. Current best AUC: 0.8035
  Tested 20/30 combinations. Current best AUC: 0.8049
  Tested 30/30 combinations. Current best AUC: 0.8052
--------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 11, 'classifier__C': 1, 'classifier__penalty': 'l2', 'classifier__solver': 'libline
Best cross-validation AUC: 0.8052

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

```
Decision Tree:
  Accuracy: 0.7250
  Precision: 0.7593
  Recall: 0.7121
  F1-Score: 0.7349
  ROC AUC: 0.7908

k-Nearest Neighbors:
  Accuracy: 0.7917
  Precision: 0.7940
  Recall: 0.8249
  F1-Score: 0.8092
  ROC AUC: 0.8765

Logistic Regression:
  Accuracy: 0.7333
  Precision: 0.7549
  Recall: 0.7432
  F1-Score: 0.7490
  ROC AUC: 0.8242

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7562, Precision: 0.7713
  Recall: 0.7743, F1: 0.7728, AUC: 0.8664
```

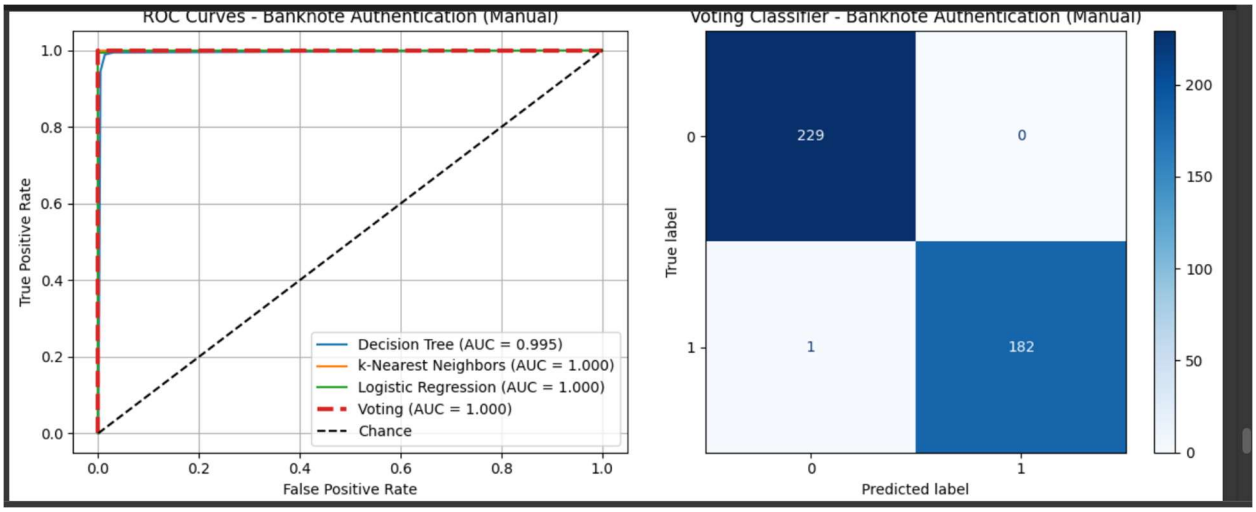



```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR WINE QUALITY
============================================================

--- GridSearchCV for Decision Tree ---
Fitting GridSearchCV for Decision Tree...
Fitting 5 folds for each of 135 candidates, totalling 675 fits
Best params for Decision Tree: {'classifier__max_depth': 10, 'classifier__min_samples_leaf': 4, 'classifier__min_samples_split': 10, 'feature_selecti
Best CV score: 0.7850

--- GridSearchCV for k-Nearest Neighbors ---
Fitting GridSearchCV for k-Nearest Neighbors...
Fitting 5 folds for each of 60 candidates, totalling 300 fits
Best params for k-Nearest Neighbors: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'feature_s
Best CV score: 0.8696

--- GridSearchCV for Logistic Regression ---
Fitting GridSearchCV for Logistic Regression...
Fitting 5 folds for each of 30 candidates, totalling 150 fits
Best params for Logistic Regression: {'classifier__C': 1, 'classifier__penalty': 'l2', 'classifier__solver': 'liblinear', 'feature_selection__k': 11}
Best CV score: 0.8052

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

```
--------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 4, 'classifier__max_depth': 7, 'classifier__min_samples_split': 2, 'classifier__min_sampl
Best cross-validation AUC: 0.9879
--- Manual Grid Search for k-Nearest Neighbors ---
Testing 20 parameter combinations...
  Tested 10/20 combinations. Current best AUC: 0.9990
  Tested 20/20 combinations. Current best AUC: 0.9990
--------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 4, 'classifier__n_neighbors': 7, 'classifier__weights': 'uniform', 'classifier__met
Best cross-validation AUC: 0.9990
--- Manual Grid Search for Logistic Regression ---
Testing 10 parameter combinations...
  Tested 10/10 combinations. Current best AUC: 0.9996
--------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 4, 'classifier__C': 100, 'classifier__penalty': 'l1', 'classifier__solver': 'liblir
Best cross-validation AUC: 0.9996

============================================================
EVALUATING MANUAL MODELS FOR BANKNOTE AUTHENTICATION
============================================================
```

```
--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.9879
  Precision: 0.9837
  Recall: 0.9891
  F1-Score: 0.9864
  ROC AUC: 0.9946

k-Nearest Neighbors:
  Accuracy: 1.0000
  Precision: 1.0000
  Recall: 1.0000
  F1-Score: 1.0000
  ROC AUC: 1.0000

Logistic Regression:
  Accuracy: 0.9879
  Precision: 0.9785
  Recall: 0.9945
  F1-Score: 0.9864
  ROC AUC: 0.9999

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.9976, Precision: 1.0000
  Recall: 0.9945, F1: 0.9973, AUC: 1.0000
```


```
--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.9879
  Precision: 0.9837
  Recall: 0.9891
  F1-Score: 0.9864
  ROC AUC: 0.9946

k-Nearest Neighbors:
  Accuracy: 1.0000
  Precision: 1.0000
  Recall: 1.0000
  F1-Score: 1.0000
  ROC AUC: 1.0000

Logistic Regression:
  Accuracy: 0.9879
  Precision: 0.9785
  Recall: 0.9945
  F1-Score: 0.9864
  ROC AUC: 0.9999
```

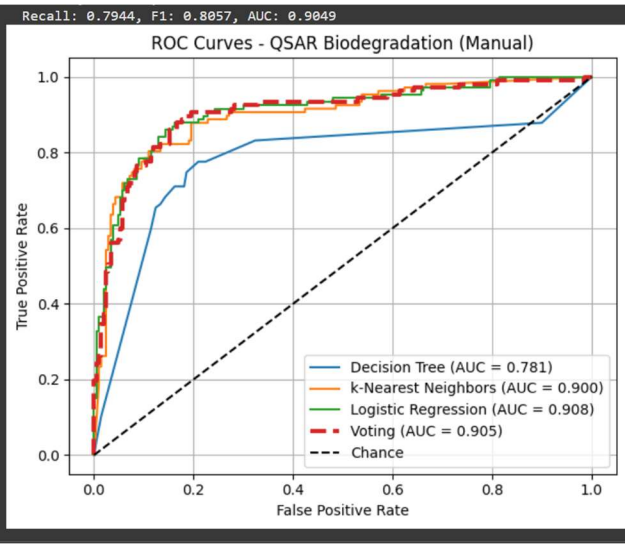
```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR BANKNOTE AUTHENTICATION
============================================================

--- GridSearchCV for Decision Tree ---
Fitting GridSearchCV for Decision Tree...
Fitting 5 folds for each of 45 candidates, totalling 225 fits
Best params for Decision Tree: {'classifier__max_depth': 7, 'classifier__min_samples_leaf': 4, 'classifier__min_samples_split': 2, 'feature_selectior
Best CV score: 0.9879

--- GridSearchCV for k-Nearest Neighbors ---
Fitting GridSearchCV for k-Nearest Neighbors...
Fitting 5 folds for each of 20 candidates, totalling 100 fits
Best params for k-Nearest Neighbors: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 7, 'classifier__weights': 'uniform', 'feature_sel
Best CV score: 0.9990

--- GridSearchCV for Logistic Regression ---
Fitting GridSearchCV for Logistic Regression...
Fitting 5 folds for each of 10 candidates, totalling 50 fits
Best params for Logistic Regression: {'classifier__C': 100, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear', 'feature_selection__k': 4
Best CV score: 0.9996
```

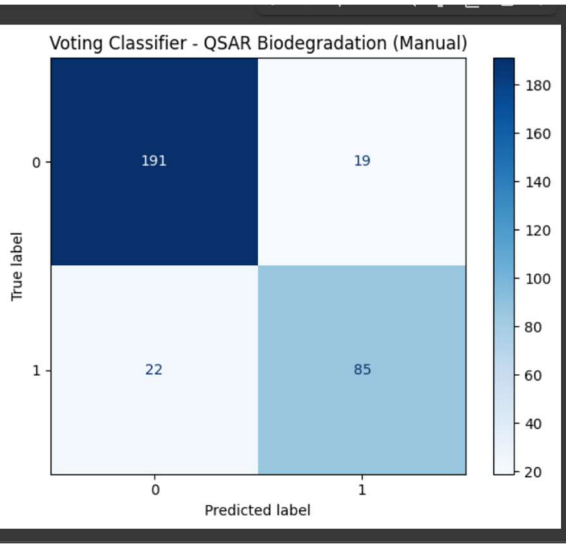
```
EVALUATING BUILT-IN MODELS FOR BANKNOTE AUTHENTICATION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.9879
  Precision: 0.9837
  Recall: 0.9891
  F1-Score: 0.9864
  ROC AUC: 0.9946

k-Nearest Neighbors:
  Accuracy: 1.0000
  Precision: 1.0000
  Recall: 1.0000
  F1-Score: 1.0000
  ROC AUC: 1.0000

Logistic Regression:
  Accuracy: 0.9879
  Precision: 0.9785
  Recall: 0.9945
  F1-Score: 0.9864
  ROC AUC: 0.9999
```

```
  Tested 180/180 combinations. Current best AUC: 0.8465
--------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 41, 'classifier__max_depth': 5, 'classifier__min_samples_split': 10, 'classifier__min_sam
Best cross-validation AUC: 0.8465
--- Manual Grid Search for k-Nearest Neighbors ---
Testing 80 parameter combinations...
  Tested 10/80 combinations. Current best AUC: 0.8294
  Tested 20/80 combinations. Current best AUC: 0.8294
  Tested 30/80 combinations. Current best AUC: 0.8687
  Tested 40/80 combinations. Current best AUC: 0.8822
  Tested 50/80 combinations. Current best AUC: 0.8822
  Tested 60/80 combinations. Current best AUC: 0.8874
  Tested 70/80 combinations. Current best AUC: 0.9020
  Tested 80/80 combinations. Current best AUC: 0.9047
--------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 41, 'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'classifier__
Best cross-validation AUC: 0.9047
--- Manual Grid Search for Logistic Regression ---
Testing 40 parameter combinations...
  Tested 10/40 combinations. Current best AUC: 0.8227
  Tested 20/40 combinations. Current best AUC: 0.8585
  Tested 30/40 combinations. Current best AUC: 0.8818
  Tested 40/40 combinations. Current best AUC: 0.9317
--------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 41, 'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'libline
Best cross-validation AUC: 0.9317
```

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

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7950
  Precision: 0.6909
  Recall: 0.7103
  F1-Score: 0.7005
  ROC AUC: 0.7810

k-Nearest Neighbors:
  Accuracy: 0.8580
  Precision: 0.7818
  Recall: 0.8037
  F1-Score: 0.7926
  ROC AUC: 0.8996

Logistic Regression:
  Accuracy: 0.8644
  Precision: 0.8200
  Recall: 0.7664
  F1-Score: 0.7923
  ROC AUC: 0.9082

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8707, Precision: 0.8173
  Recall: 0.7944, F1: 0.8057, AUC: 0.9049
```

```
Recall: 0.7944, F1: 0.8057, AUC: 0.9049
```






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

--- GridSearchCV for Decision Tree ---
Fitting GridSearchCV for Decision Tree...
Fitting 5 folds for each of 180 candidates, totalling 900 fits
Best params for Decision Tree: {'classifier__max_depth': 5, 'classifier__min_samples_leaf': 4, 'classifier__min_samples_split': 10, 'feature_selectic
Best CV score: 0.8465

--- GridSearchCV for k-Nearest Neighbors ---
Fitting GridSearchCV for k-Nearest Neighbors...
Fitting 5 folds for each of 80 candidates, totalling 400 fits
Best params for k-Nearest Neighbors: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'feature_s
Best CV score: 0.9047

--- GridSearchCV for Logistic Regression ---
Fitting GridSearchCV for Logistic Regression...
Fitting 5 folds for each of 40 candidates, totalling 200 fits
Best params for Logistic Regression: {'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear', 'feature_selection__k': 41}
Best CV score: 0.9317

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

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

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7950
  Precision: 0.6909
  Recall: 0.7103
  F1-Score: 0.7005
  ROC AUC: 0.7810

k-Nearest Neighbors:
  Accuracy: 0.8580
  Precision: 0.7818
  Recall: 0.8037
  F1-Score: 0.7926
  ROC AUC: 0.8996

Logistic Regression:
  Accuracy: 0.8644
  Precision: 0.8200
  Recall: 0.7664
  F1-Score: 0.7923
  ROC AUC: 0.9082
```

```
  --- Built-in Voting Classifier ---
  Error processing QSAR Biodegradation: name 'X_train' is not defined

  ================================================================================
  ALL DATASETS PROCESSED!
  ================================================================================
```