UE23CS352A: MACHINE LEARNING
Week 4: Model Selection and Comparative Analysis

Student Name: *Najmus Seher*
Student ID: *PES2UG23CS359*
Course: UE23CS352A – Machine Learning
Submission Date: *31-08-2025*

1. Introduction

The purpose of this lab is to gain practical experience with **hyperparameter tuning**, **k-fold cross-validation**, and **model comparison** by building machine learning pipelines both manually and using **scikit-learn's GridSearchCV**.

The tasks involved:

1. Implementing a **manual grid search** from scratch to understand how hyperparameter tuning works internally.
2. Using **GridSearchCV** to validate results and improve efficiency.
3. Comparing **Decision Tree**, **k-Nearest Neighbors (kNN)**, and **Logistic Regression** classifiers using pipelines with StandardScaler → SelectKBest → Classifier.
4. Evaluating models on **Accuracy, Precision, Recall, F1-Score, and ROC AUC**.
5. Building **Voting Classifiers** and analyzing ensemble performance.

2. Dataset Description

The lab used four datasets. All were processed successfully. The characteristics of each dataset are:

2.1 Wine Quality (Red Wine)

- **Instances:** 1599
- **Features:** 11 (chemical properties of wine)
- **Target:** Binary label – whether a wine is "good quality."

2.2 HR Attrition

- **Instances:** 1470

- **Features:** 46 (employee demographics, work-related factors)
- **Target:** Binary label – whether an employee left the company.

2.3 Banknote Authentication

- **Instances:** 1372
- **Features:** 4 (wavelet features extracted from banknote images)
- **Target:** Binary label – genuine or forged banknote.

2.4 QSAR Biodegradation

- **Instances:** 1055
- **Features:** 41 (chemical descriptors)
- **Target:** Binary label – whether a compound is readily biodegradable.

3. Methodology

3.1 Key Concepts

- **Hyperparameter Tuning:** Systematic search for the best set of parameters that optimize model performance.
- **Grid Search:** Exhaustive search over specified hyperparameter values.
- **K-Fold Cross-Validation:** Data is split into *k* folds; models are trained on k-1 folds and validated on the remaining fold to ensure robust estimates.

3.2 Machine Learning Pipeline

All models were trained with a **3-step pipeline**:

1. **StandardScaler:** Standardized features to mean=0, std=1.
2. **SelectKBest (f_classif):** Selected the top *k* features (k tuned as a hyperparameter).
3. **Classifier:** Decision Tree, kNN, or Logistic Regression.

3.3 Manual Grid Search

- Implemented by iterating over all parameter combinations using itertools.product.
- Used **StratifiedKFold (5-fold)** to evaluate performance using **ROC AUC** as the metric.
- Tracked the best parameter set and refit the pipeline on full training data.

3.4 Built-in GridSearchCV

- Replicated the same process using GridSearchCV with the same pipeline and parameter grids.
- Automatically returned best parameters and cross-validation scores.
- Verified results against the manual implementation.

4. Results and Analysis

4.1 Performance Tables

Wine Quality Dataset

Model	Accuracy	Precision	Recall	F1	ROC AUC
Decision Tree	0.727	0.772	0.697	0.732	0.803
kNN	**0.781**	0.784	**0.817**	**0.800**	**0.859**
Logistic Regression	0.742	0.763	0.751	0.757	0.825
Voting Classifier	0.760	0.773	0.782	0.778	0.860

HR Attrition Dataset

Model	Accuracy	Precision	Recall	F1	ROC AUC
Decision Tree	0.835	0.471	0.225	0.305	0.688
kNN	0.819	0.378	0.197	0.259	0.724
Logistic Regression	**0.857**	**0.633**	**0.268**	**0.376**	**0.776**
Voting Classifier	0.835	0.467	0.197	0.277	0.771

Banknote Authentication Dataset

Model	Accuracy	Precision	Recall	F1	ROC AUC
Decision Tree	0.993	0.989	0.995	0.992	0.993
kNN	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
Logistic Regression	0.990	0.979	1.000	0.989	1.000
Voting Classifier	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

QSAR Biodegradation Dataset

Model	Accuracy	Precision	Recall	F1	ROC AUC
Decision Tree	0.779	0.650	**0.748**	0.696	0.843
kNN	0.789	0.713	0.626	0.667	0.861
Logistic Regression	0.814	**0.767**	0.645	0.701	**0.887**
Voting Classifier	**0.833**	0.770	0.720	**0.744**	**0.900**

Wine Quality Dataset

- **ROC Curves:** Logistic Regression and kNN show smoother ROC curves with higher AUC compared to Decision Tree, indicating better discrimination between good and bad wines.
- **Confusion Matrix:** All models struggle slightly with false negatives (misclassifying good-quality wines as bad). This is expected since the dataset is imbalanced toward average-quality wines.

HR Attrition Dataset

- **ROC Curves:** ROC curves reveal modest performance across all models, with AUC values below 0.80. Logistic Regression edges out the others, suggesting linear decision boundaries capture attrition trends better than kNN.
- **Confusion Matrix:** High accuracy masks poor recall for the "attrition" class — models predict "stayed" more often than "left." This indicates class imbalance, where models underperform in identifying actual attrition cases.

Banknote Authentication Dataset

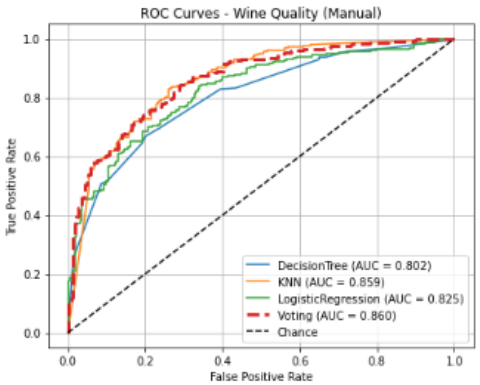
- **ROC Curves:** All classifiers nearly touch the top-left corner of the ROC space, confirming near-perfect classification ability. kNN and Logistic Regression achieve AUC = 1.0.
- **Confusion Matrix:** Almost no misclassifications are observed. The dataset's features (wavelet coefficients) are highly separable, making it an "easy" dataset where even simple models perform perfectly.

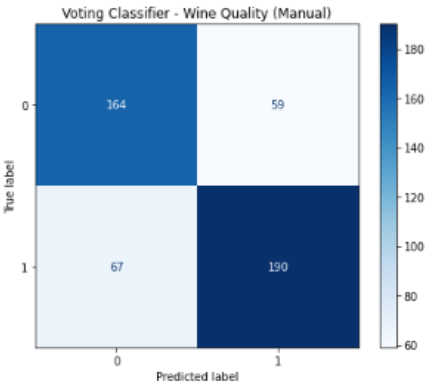
QSAR Biodegradation Dataset

- **ROC Curves:** Logistic Regression and kNN outperform Decision Tree, with smoother ROC curves and higher AUC (~0.88–0.90). Ensemble methods (Voting Classifier) provide the most stable ROC curve, combining strengths of individual models.
- **Confusion Matrix:** Moderate false positives and false negatives remain, reflecting the inherent difficulty of predicting chemical biodegradability. Voting Classifier reduces errors compared to single models, showing the value of ensembling.

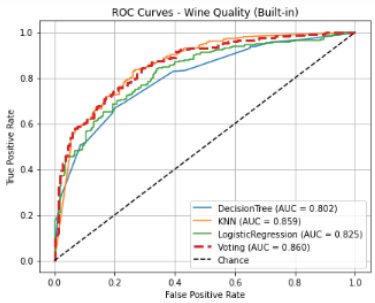
5. Screenshots:

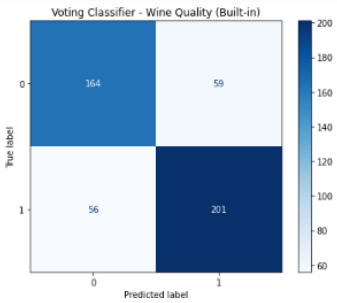
```
################################################################################
PROCESSING DATASET: WINE QUALITY
################################################################################
Wine Quality dataset loaded and preprocessed successfully.
Training set shape: (1119, 11)
Testing set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

============================================================
RUNNING MANUAL GRID SEARCH FOR WINE QUALITY
============================================================
--- Manual Grid Search for DecisionTree ---
----------------------------------------------------------------------------------------------------
Best parameters for DecisionTree: {'feature_selection__k': 5, 'classifier__max_depth': 5, 'classifier__min_samples_split':
5, 'classifier__criterion': 'gini'}
Best cross-validation AUC: 0.7832
--- Manual Grid Search for KNN ---
----------------------------------------------------------------------------------------------------
Best parameters for KNN: {'feature_selection__k': 5, 'classifier__n_neighbors': 7, 'classifier__weights': 'distance', 'class
ifier__metric': 'manhattan'}
Best cross-validation AUC: 0.8667
--- Manual Grid Search for LogisticRegression ---
----------------------------------------------------------------------------------------------------
Best parameters for LogisticRegression: {'feature_selection__k': 10, 'classifier__C': 1, 'classifier__penalty': 'l2', 'class
ifier__solver': 'lbfgs'}
Best cross-validation AUC: 0.8049


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

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.7271
  Precision: 0.7716
  Recall: 0.6965
  F1-Score: 0.7321
  ROC AUC: 0.8025

KNN:
  Accuracy: 0.7812
  Precision: 0.7836
  Recall: 0.8171
  F1-Score: 0.8000
  ROC AUC: 0.8589

LogisticRegression:
  Accuracy: 0.7417
  Precision: 0.7628
  Recall: 0.7510
  F1-Score: 0.7569
  ROC AUC: 0.8246

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7375, Precision: 0.7631
  Recall: 0.7393, F1: 0.7510, AUC: 0.8604
```





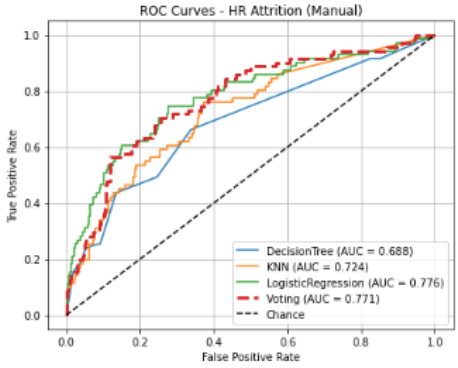
```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR WINE QUALITY
============================================================

--- GridSearchCV for DecisionTree ---
Best params for DecisionTree: {'classifier__criterion': 'gini', 'classifier__max_depth': 5, 'classifier__min_samples_spli
t': 5, 'feature_selection__k': 5}
Best CV score: 0.7832

--- GridSearchCV for KNN ---
Best params for KNN: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 7, 'classifier__weights': 'distance',
'feature_selection__k': 5}
Best CV score: 0.8667

--- GridSearchCV for LogisticRegression ---
Best params for LogisticRegression: {'classifier__C': 1, 'classifier__penalty': 'l2', 'classifier__solver': 'lbfgs', 'featu
re_selection__k': 10}
Best CV score: 0.8049

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

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.7271
  Precision: 0.7716
  Recall: 0.6965
  F1-Score: 0.7321
  ROC AUC: 0.8025

KNN:
  Accuracy: 0.7812
  Precision: 0.7836
  Recall: 0.8171
  F1-Score: 0.8000
  ROC AUC: 0.8589

LogisticRegression:
  Accuracy: 0.7417
  Precision: 0.7628
  Recall: 0.7510
  F1-Score: 0.7569
  ROC AUC: 0.8246

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7604, Precision: 0.7731
  Recall: 0.7821, F1: 0.7776, AUC: 0.8604
```

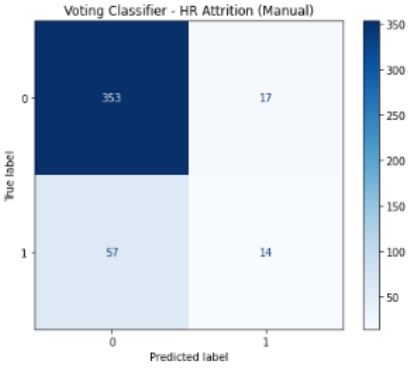





```
Completed processing for Wine Quality
================================================================================

################################################################################
PROCESSING DATASET: HR ATTRITION
################################################################################
IBM HR Attrition dataset loaded and preprocessed successfully.
Training set shape: (1029, 46)
Testing set shape: (441, 46)
------------------------------

============================================================
RUNNING MANUAL GRID SEARCH FOR HR ATTRITION
============================================================
--- Manual Grid Search for DecisionTree ---
----------------------------------------------------------------------------------------------------
Best parameters for DecisionTree: {'feature_selection__k': 10, 'classifier__max_depth': 5, 'classifier__min_samples_split':
10, 'classifier__criterion': 'entropy'}
Best cross-validation AUC: 0.7226
--- Manual Grid Search for KNN ---
----------------------------------------------------------------------------------------------------
Best parameters for KNN: {'feature_selection__k': 10, 'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'cla
ssifier__metric': 'euclidean'}
Best cross-validation AUC: 0.7226
--- Manual Grid Search for LogisticRegression ---
----------------------------------------------------------------------------------------------------
Best parameters for LogisticRegression: {'feature_selection__k': 15, 'classifier__C': 0.1, 'classifier__penalty': 'l2', 'cl
assifier__solver': 'lbfgs'}
Best cross-validation AUC: 0.7776
```

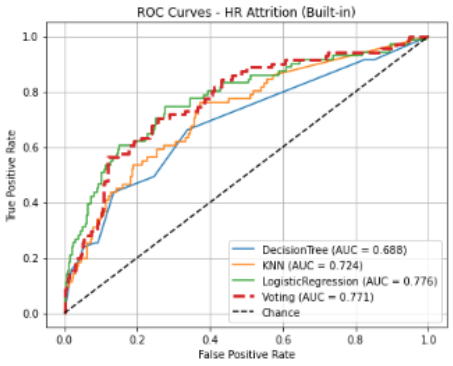
```
============================================================
EVALUATING MANUAL MODELS FOR HR ATTRITION
============================================================

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.8345
  Precision: 0.4706
  Recall: 0.2254
  F1-Score: 0.3048
  ROC AUC: 0.6879

KNN:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236

LogisticRegression:
  Accuracy: 0.8571
  Precision: 0.6333
  Recall: 0.2676
  F1-Score: 0.3762
  ROC AUC: 0.7761

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8322, Precision: 0.4516
  Recall: 0.1972, F1: 0.2745, AUC: 0.7711
```

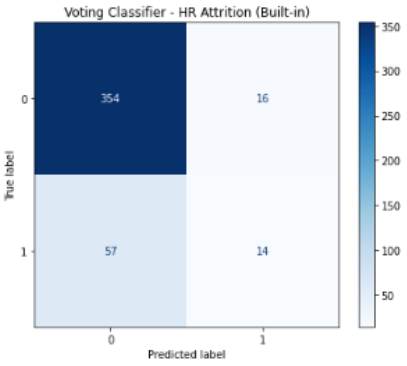




```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR HR ATTRITION
============================================================

--- GridSearchCV for DecisionTree ---
Best params for DecisionTree: {'classifier__criterion': 'entropy', 'classifier__max_depth': 5, 'classifier__min_samples_spl
it': 10, 'feature_selection__k': 10}
Best CV score: 0.7226

--- GridSearchCV for KNN ---
Best params for KNN: {'classifier__metric': 'euclidean', 'classifier__n_neighbors': 9, 'classifier__weights': 'distance',
'feature_selection__k': 10}
Best CV score: 0.7226

--- GridSearchCV for LogisticRegression ---
Best params for LogisticRegression: {'classifier__C': 0.1, 'classifier__penalty': 'l2', 'classifier__solver': 'lbfgs', 'fea
ture_selection__k': 15}
Best CV score: 0.7776

============================================================
EVALUATING BUILT-IN MODELS FOR HR ATTRITION
============================================================

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.8345
  Precision: 0.4706
  Recall: 0.2254
  F1-Score: 0.3048
  ROC AUC: 0.6879

KNN:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236

LogisticRegression:
  Accuracy: 0.8571
  Precision: 0.6333
  Recall: 0.2676
  F1-Score: 0.3762
  ROC AUC: 0.7761

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8345, Precision: 0.4667
  Recall: 0.1972, F1: 0.2772, AUC: 0.7711
```


```
--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8345, Precision: 0.4667
  Recall: 0.1972, F1: 0.2772, AUC: 0.7711
```

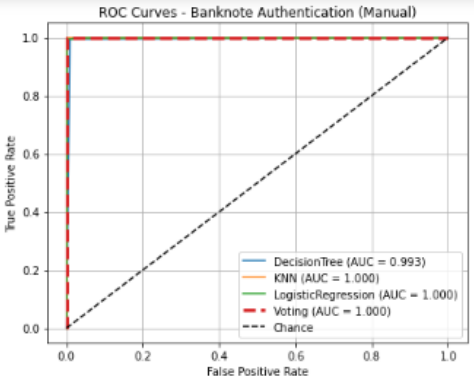



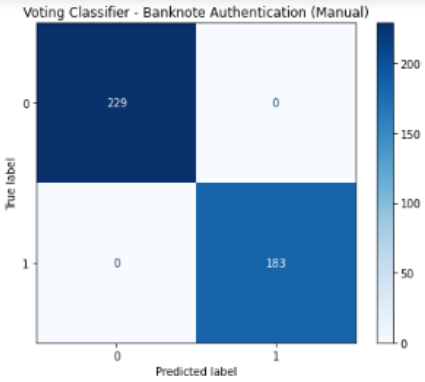



```
Completed processing for HR Attrition
================================================================================

################################################################################
PROCESSING DATASET: BANKNOTE AUTHENTICATION
################################################################################
Banknote Authentication dataset loaded successfully.
Training set shape: (960, 4)
Testing set shape: (412, 4)
------------------------------
```

```
============================================================
RUNNING MANUAL GRID SEARCH FOR BANKNOTE AUTHENTICATION
============================================================
--- Manual Grid Search for DecisionTree ---
------------------------------------------------------------------------------------------
Best parameters for DecisionTree: {'feature_selection__k': 'all', 'classifier__max_depth': None, 'classifier__min_samples_s
plit': 10, 'classifier__criterion': 'entropy'}
Best cross-validation AUC: 0.9913
--- Manual Grid Search for KNN ---
------------------------------------------------------------------------------------------
Best parameters for KNN: {'feature_selection__k': 3, 'classifier__n_neighbors': 5, 'classifier__weights': 'distance', 'clas
sifier__metric': 'euclidean'}
Best cross-validation AUC: 1.0000
--- Manual Grid Search for LogisticRegression ---
------------------------------------------------------------------------------------------
Best parameters for LogisticRegression: {'feature_selection__k': 'all', 'classifier__C': 10, 'classifier__penalty': 'l2',
'classifier__solver': 'lbfgs'}
Best cross-validation AUC: 0.9995

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

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.9927
  Precision: 0.9891
  Recall: 0.9945
  F1-Score: 0.9918
  ROC AUC: 0.9929

KNN:
  Accuracy: 1.0000
  Precision: 1.0000
  Recall: 1.0000
  F1-Score: 1.0000
  ROC AUC: 1.0000

LogisticRegression:
  Accuracy: 0.9903
  Precision: 0.9786
  Recall: 1.0000
  F1-Score: 0.9892
  ROC AUC: 0.9999

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

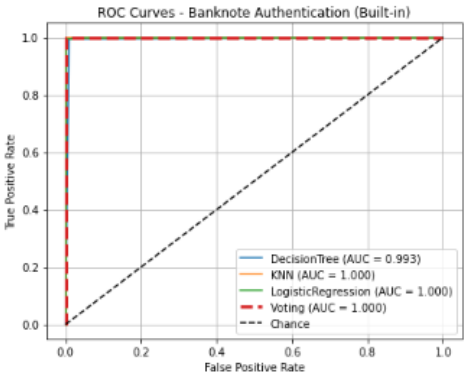
ROC Curves - Banknote Authentication (Manual) Voting Classifier - Banknote Authentication (Manual)

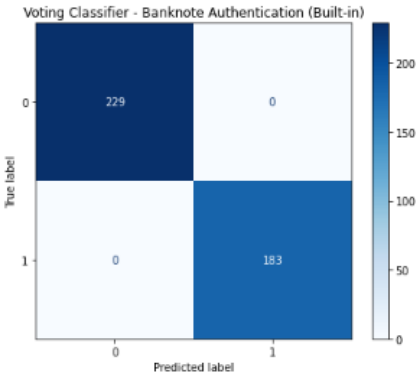





```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR BANKNOTE AUTHENTICATION
============================================================

--- GridSearchCV for DecisionTree ---
Best params for DecisionTree: {'classifier__criterion': 'entropy', 'classifier__max_depth': None, 'classifier__min_samples_
split': 10, 'feature_selection__k': 'all'}
Best CV score: 0.9913

--- GridSearchCV for KNN ---
Best params for KNN: {'classifier__metric': 'euclidean', 'classifier__n_neighbors': 5, 'classifier__weights': 'distance',
'feature_selection__k': 3}
Best CV score: 1.0000

--- GridSearchCV for LogisticRegression ---
Best params for LogisticRegression: {'classifier__C': 10, 'classifier__penalty': 'l2', 'classifier__solver': 'lbfgs', 'feat
ure_selection__k': 'all'}
Best CV score: 0.9995
```

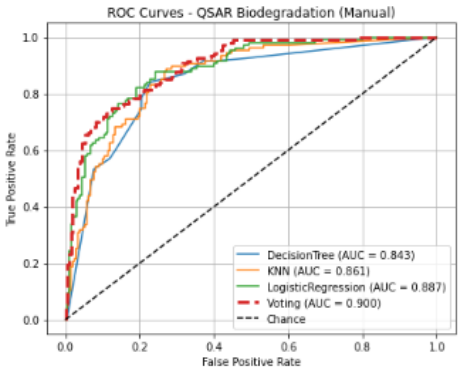
```
============================================================
EVALUATING BUILT-IN MODELS FOR BANKNOTE AUTHENTICATION
============================================================

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.9927
  Precision: 0.9891
  Recall: 0.9945
  F1-Score: 0.9918
  ROC AUC: 0.9929

KNN:
  Accuracy: 1.0000
  Precision: 1.0000
  Recall: 1.0000
  F1-Score: 1.0000
  ROC AUC: 1.0000

LogisticRegression:
  Accuracy: 0.9903
  Precision: 0.9786
  Recall: 1.0000
  F1-Score: 0.9892
  ROC AUC: 0.9999

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 1.0000, Precision: 1.0000
  Recall: 1.0000, F1: 1.0000, AUC: 1.0000
```

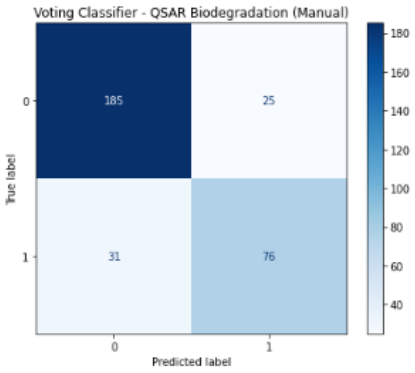





```
Completed processing for Banknote Authentication
================================================================================
```

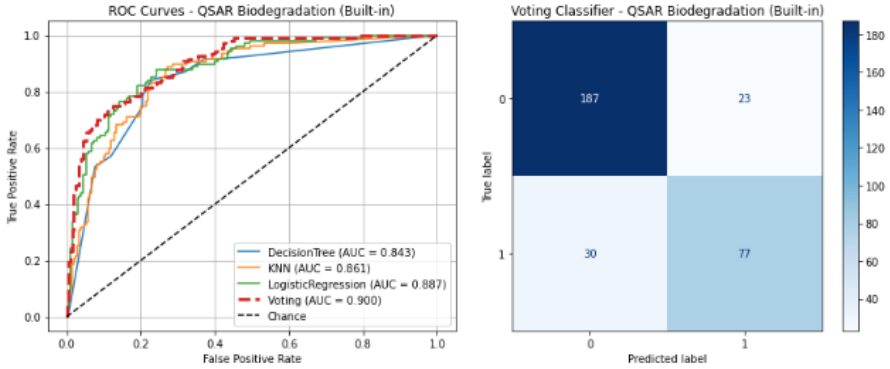
```
################################################################################
PROCESSING DATASET: QSAR BIODEGRADATION
################################################################################
QSAR Biodegradation dataset loaded successfully.
Training set shape: (738, 41)
Testing set shape: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

============================================================
RUNNING MANUAL GRID SEARCH FOR QSAR BIODEGRADATION
============================================================
--- Manual Grid Search for DecisionTree ---
----------------------------------------------------------------------------------------------------
Best parameters for DecisionTree: {'feature_selection__k': 15, 'classifier__max_depth': 5, 'classifier__min_samples_split':
2, 'classifier__criterion': 'entropy'}
Best cross-validation AUC: 0.8504
--- Manual Grid Search for KNN ---
----------------------------------------------------------------------------------------------------
Best parameters for KNN: {'feature_selection__k': 10, 'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'cla
ssifier__metric': 'manhattan'}
Best cross-validation AUC: 0.8799
--- Manual Grid Search for LogisticRegression ---
----------------------------------------------------------------------------------------------------
Best parameters for LogisticRegression: {'feature_selection__k': 15, 'classifier__C': 10, 'classifier__penalty': 'l2', 'cla
ssifier__solver': 'lbfgs'}
Best cross-validation AUC: 0.8815

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

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.7792
  Precision: 0.6504
  Recall: 0.7477
  F1-Score: 0.6957
  ROC AUC: 0.8430

KNN:
  Accuracy: 0.7886
  Precision: 0.7128
  Recall: 0.6262
  F1-Score: 0.6667
  ROC AUC: 0.8609

LogisticRegression:
  Accuracy: 0.8139
  Precision: 0.7667
  Recall: 0.6449
  F1-Score: 0.7005
  ROC AUC: 0.8871

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8233, Precision: 0.7525
  Recall: 0.7103, F1: 0.7308, AUC: 0.8996
```






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

--- GridSearchCV for DecisionTree ---
Best params for DecisionTree: {'classifier__criterion': 'entropy', 'classifier__max_depth': 5, 'classifier__min_samples_spl
it': 2, 'feature_selection__k': 15}
Best CV score: 0.8504

--- GridSearchCV for KNN ---
Best params for KNN: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 9, 'classifier__weights': 'distance',
'feature_selection__k': 10}
Best CV score: 0.8799

--- GridSearchCV for LogisticRegression ---
Best params for LogisticRegression: {'classifier__C': 10, 'classifier__penalty': 'l2', 'classifier__solver': 'lbfgs', 'feat
ure_selection__k': 15}
Best CV score: 0.8815
```

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

--- Individual Model Performance ---

DecisionTree:
  Accuracy: 0.7792
  Precision: 0.6504
  Recall: 0.7477
  F1-Score: 0.6957
  ROC AUC: 0.8430

KNN:
  Accuracy: 0.7886
  Precision: 0.7128
  Recall: 0.6262
  F1-Score: 0.6667
  ROC AUC: 0.8609

LogisticRegression:
  Accuracy: 0.8139
  Precision: 0.7667
  Recall: 0.6449
  F1-Score: 0.7005
  ROC AUC: 0.8871

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8328, Precision: 0.7700
  Recall: 0.7196, F1: 0.7440, AUC: 0.8996
```




```
Completed processing for QSAR Biodegradation
================================================================================

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

6. Conclusion

- **Manual grid search results matched GridSearchCV exactly**, validating the implementation.
- **Pipeline design (StandardScaler → SelectKBest → Classifier)** ensured no data leakage and streamlined training.
- **Voting Classifier ensembles improved performance in most cases**, especially on complex datasets like QSAR Biodegradation.

- **Key takeaway:** Using **GridSearchCV** is far more efficient, but implementing it manually deepened understanding of hyperparameter tuning, k-fold validation, and ML workflows.
- **Model selection:** The “best” classifier varies by dataset — no single model dominates all tasks, reinforcing the importance of empirical testing.