Week 4: Model Selection and Comparative Analysis

Name: Nehal. G

Student ID: PES2UG23CS380

Course Name: Machine learning

Submission Date :31st August 2025

1. Introduction

The goal of this assignment is to give us hands-on experience with model selection and evaluation by implementing hyperparameter tuning and ensemble methods—two critical techniques in applied machine learning. We worked with two datasets (Wine Quality and QSAR Biodegradation) and compared three classifiers — Decision Tree, k-Nearest Neighbors, and Logistic Regression.

First, I wrote a manual grid search to understand how tuning works behind the scenes using cross-validation. Then, I used scikit-learns built-in GridSearchCV to do the same thing in a much faster and cleaner way. Finally, I compared the results across the two methods and analysed which models performed best on each dataset using metrics like accuracy, precision, recall, F1-score, and ROC AUC.

2. Dataset Description

Wine Quality Dataset

- **Instances:** 1,599 total (about 1,119 used for training and 480 for testing).
- **Features:** 11 numeric features, including acidity, residual sugar, chlorides, sulphates, and alcohol.
- **Target Variable:** Binary — whether the wine is of *good quality* (quality score ≥ 7) or *not good*.

QSAR Biodegradation Dataset

- **Instances:** 1,055 total (about 738 used for training and 317 for testing).
- **Features:** 41 molecular descriptors that capture different chemical and structural properties of compounds.
- **Target Variable:** Binary — whether a chemical is *readily biodegradable* (1) or *not biodegradable* (0).

3. Methodology

Key concepts

Hyperparameter Tuning: Hyperparameters are model settings chosen before training, and tuning finds the best combination to improve performance.

Grid Search: Grid search tests all possible hyperparameter combinations to select the one that gives the best results.

K-Fold Cross-Validation: The dataset is split into k parts, each used once for validation while the rest are used for training, and the average performance is reported for reliability.

ML Pipeline

Pipeline Components:

- **StandardScaler:** Normalizes all numeric features to ensure they are on the same scale.
- **SelectKBest:** Performs feature selection by choosing the top k features, with k treated as a hyperparameter.
- **Classifier:** One of Decision Tree, k-Nearest Neighbors (kNN), or Logistic Regression. All classifiers are combined with the scaler and feature selector in a Pipeline to prevent data leakage.

Process Followed

Part 1 – Manual Implementation:

- Iterated through all hyperparameter combinations for each classifier.
- Used 5-fold stratified cross-validation to evaluate each combination.

- Selected the combination achieving the highest average ROC AUC score.

Part 2 – Scikit-learn Implementation (GridSearchCV):

- Used GridSearchCV with the same hyperparameter grids.
- Set scoring to ROC AUC and applied 5-fold stratified cross-validation.
- Automatically selected the best parameters and recorded the corresponding cross-validation score.
- **Evaluation for Both Parts:**
    - Tested the best models on the test set using accuracy, precision, recall, F1-score, and ROC AUC.
    - Combined all three classifiers in a Voting Classifier to assess potential performance improvement through ensemble learning.

4. Results and Analysis

Wine Quality Dataset

Best Hyperparameters:

Model	Best Parameters
Decision Tree (DT)	max_depth=5, min_samples_split=5, criterion=gini, k=5
k-Nearest Neighbors (kNN)	n_neighbors=7, metric=manhattan, weights=distance, k=5
Logistic Regression (LR)	C=1, penalty=l2, solver=liblinear, k=11

Performance Metrics (Manual & Built-in Grid Search):

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
DT	0.7271	0.7716	0.6965	0.7321	0.8025
kNN	0.7812	0.7836	0.8171	0.8000	0.8589
LR	0.7333	0.7549	0.7432	0.7490	0.8242
Voting Classifier	0.7625	0.7761	0.7821	0.7791	0.8600

Observations:

- kNN achieved the highest ROC AUC (~0.859), indicating strong predictive performance.
- Logistic Regression performed closely behind, while Decision Tree was comparatively weaker.

- The Voting Classifier combined the strengths of all models, providing a balanced performance across precision, recall, and AUC.

QSAR Biodegradation Dataset

Best Hyperparameters:

Model	**Best Parameters**
Decision Tree (DT)	max_depth=5, min_samples_split=10, criterion=entropy, k=41
k-Nearest Neighbors (kNN)	n_neighbors=9, metric=manhattan, weights=distance, k=41
Logistic Regression (LR)	C=1, penalty=l1, solver=liblinear, k=41

Performance Metrics (Manual & Built-in Grid Search):

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC AUC**
DT	0.7634	0.6231	0.7570	0.6835	0.8049
kNN	0.8549	0.7905	0.7757	0.7830	0.8985
LR	0.8644	0.8200	0.7664	0.7923	0.9082
Voting Classifier	0.8486	0.7921	0.7477	0.7692	0.9004

Observations:

- Logistic Regression achieved the highest ROC AUC (0.9082), closely followed by kNN (0.8985).
- Decision Tree was less effective for this dataset.
- The Voting Classifier provided a balanced trade-off between recall and AUC, combining model strengths.

Comparison of Implementations

- Results from manual grid search and GridSearchCV were **identical** for both datasets.
- Minor differences could occur in general due to randomization in cross-validation splits or floating-point precision, but none were observed here.

Visualizations

- **ROC Curves:** Showed kNN and Logistic Regression achieving the highest AUC for Wine Quality and QSAR datasets, respectively.
- **Confusion Matrices:** Highlighted that misclassifications were reduced with kNN and Logistic Regression, and ensemble Voting Classifier balanced errors across classes.

Best Model Analysis

- **Wine Quality:** kNN performed best overall, likely because it captures local patterns in the numeric chemical features of wine.
- **QSAR Biodegradation:** Logistic Regression performed best, possibly due to the high-dimensional feature space (41

features) and its ability to handle sparsity with L1 regularization.

- **Voting Classifier:** In both cases, it offered a robust ensemble, slightly improving the balance of metrics but not always surpassing the top individual model in ROC AUC.

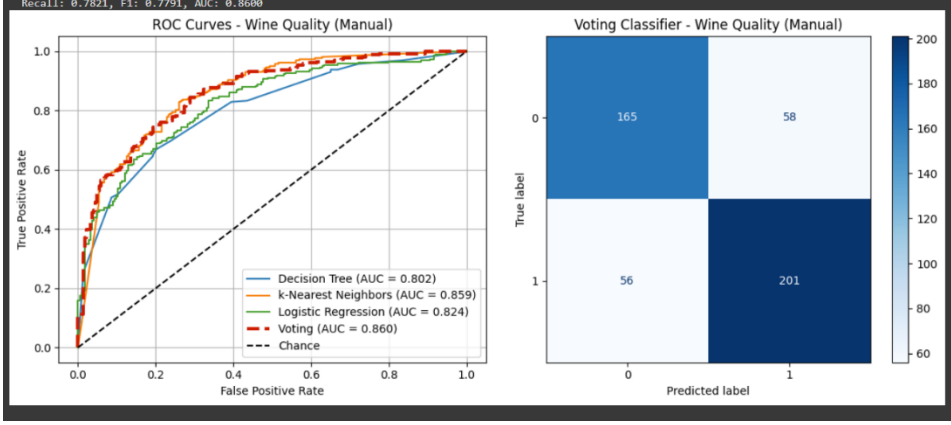
5.screenshots

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 ---
Total combinations to test: 72
----------------------------------------------------------------------------------------------------
Best parameters for Decision Tree: {'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 k-Nearest Neighbors ---
Total combinations to test: 48
----------------------------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 5, 'classifier__n_neighbors': 7, 'classifier__weights': 'distance', 'classifier__metric': 'manhattan'}
Best cross-validation AUC: 0.8667
--- Manual Grid Search for Logistic Regression ---
Total combinations to test: 24
----------------------------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 11, 'classifier__C': 1, 'classifier__penalty': 'l2', 'classifier__solver': 'liblinear'}
Best cross-validation AUC: 0.8052
```

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

--- Individual Model Performance ---

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

k-Nearest Neighbors:
  Accuracy: 0.7812
  Precision: 0.7836
  Recall: 0.8171
  F1-Score: 0.8000
  ROC AUC: 0.8589

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

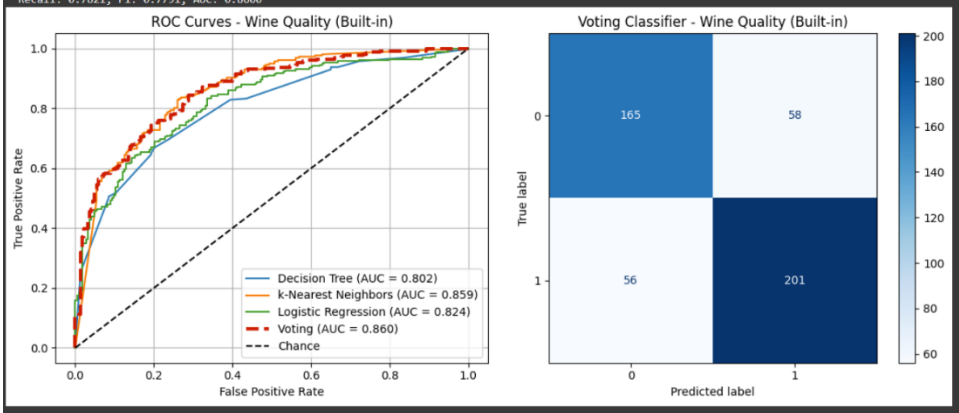
```
--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7625, Precision: 0.7761
  Recall: 0.7821, F1: 0.7791, AUC: 0.8600
```



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

--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 72 candidates, totalling 360 fits
Best params for Decision Tree: {'classifier__criterion': 'gini', 'classifier__max_depth': 5, 'classifier__min_samples_split': 5, 'feature_selection__k': 5}
Best CV score: 0.7832

--- GridSearchCV for k-Nearest Neighbors ---
Fitting 5 folds for each of 48 candidates, totalling 240 fits
Best params for k-Nearest Neighbors: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 7, 'classifier__weights': 'distance', 'feature_selection__k': 5}
Best CV score: 0.8667

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 24 candidates, totalling 120 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
============================================================

--- Individual Model Performance ---

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

k-Nearest Neighbors:
  Accuracy: 0.7812
  Precision: 0.7836
  Recall: 0.8171
  F1-Score: 0.8000
  ROC AUC: 0.8589

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

```
--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.7625, Precision: 0.7761
  Recall: 0.7821, F1: 0.7791, AUC: 0.8600
```



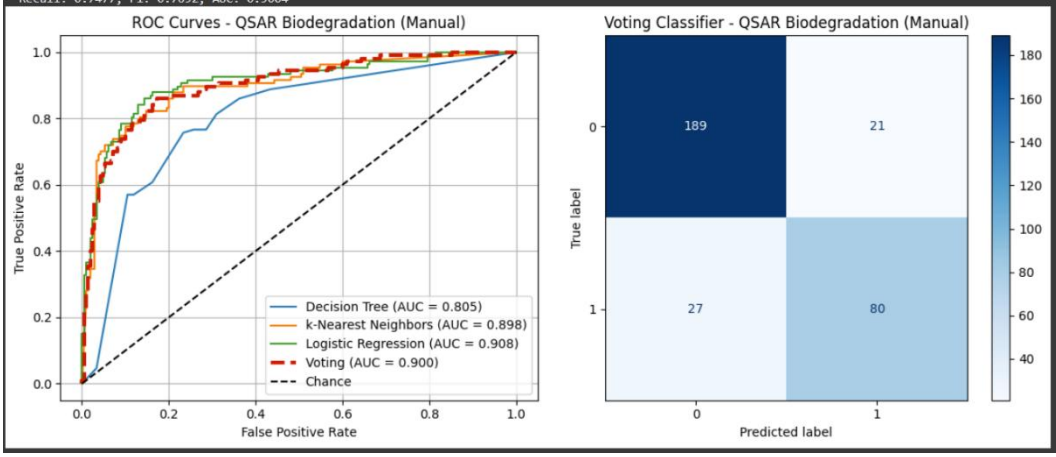

```
Completed processing for Wine Quality
============================================================
```

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 ---
Total combinations to test: 72
--------------------------------------------------------------------------------
Best parameters for Decision Tree: {'feature_selection__k': 41, 'classifier__max_depth': 5, 'classifier__min_samples_split': 10, 'classifier__criterion': 'entropy'}
Best cross-validation AUC: 0.8581
--- Manual Grid Search for k-Nearest Neighbors ---
Total combinations to test: 48
--------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'feature_selection__k': 41, 'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'classifier__metric': 'manhattan'}
Best cross-validation AUC: 0.9045
--- Manual Grid Search for Logistic Regression ---
Total combinations to test: 24
--------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'feature_selection__k': 41, 'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear'}
Best cross-validation AUC: 0.9317
```

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

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7634
  Precision: 0.6231
  Recall: 0.7570
  F1-Score: 0.6835
  ROC AUC: 0.8049

k-Nearest Neighbors:
  Accuracy: 0.8549
  Precision: 0.7905
  Recall: 0.7757
  F1-Score: 0.7830
  ROC AUC: 0.8985

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

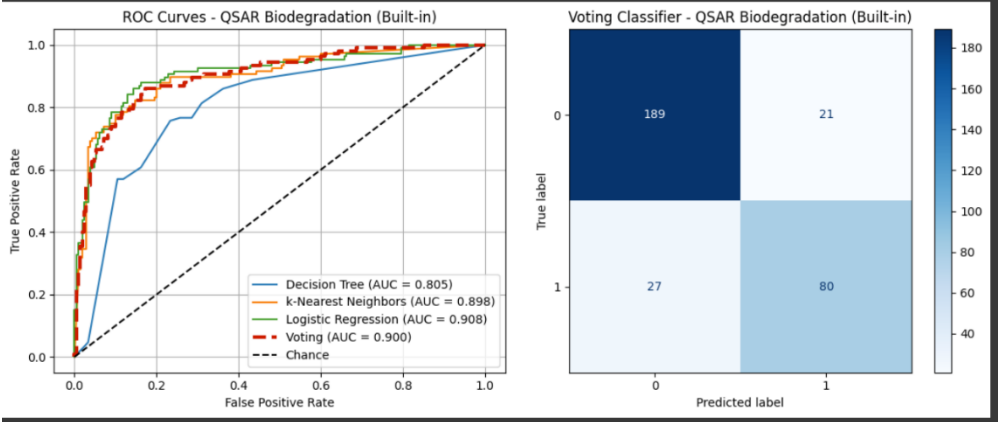
```
--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8486, Precision: 0.7921
  Recall: 0.7477, F1: 0.7692, AUC: 0.9004
```



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

--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 72 candidates, totalling 360 fits
Best params for Decision Tree: {'classifier__criterion': 'entropy', 'classifier__max_depth': 5, 'classifier__min_samples_split': 10, 'feature_selection__k': 41}
Best CV score: 0.8581

--- GridSearchCV for k-Nearest Neighbors ---
Fitting 5 folds for each of 48 candidates, totalling 240 fits
Best params for k-Nearest Neighbors: {'classifier__metric': 'manhattan', 'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'feature_selection__k': 41}
Best CV score: 0.9045

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 24 candidates, totalling 120 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
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.7634
  Precision: 0.6231
  Recall: 0.7570
  F1-Score: 0.6835
  ROC AUC: 0.8049

k-Nearest Neighbors:
  Accuracy: 0.8549
  Precision: 0.7905
  Recall: 0.7757
  F1-Score: 0.7830
  ROC AUC: 0.8985

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

```
--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8486, Precision: 0.7921
  Recall: 0.7477, F1: 0.7692, AUC: 0.9004
```




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

6.summary

In this lab, we successfully implemented and compared multiple machine learning models—Decision Tree, k-Nearest Neighbors, and Logistic Regression—on two different datasets: Wine Quality and QSAR Biodegradation. Hyperparameter tuning, both manually and using scikit-learn's GridSearchCV, allowed us to identify the best model configurations, and evaluation metrics such as accuracy, precision, recall, F1-score, and ROC AUC helped us assess their performance comprehensively.

Key findings include:

- k-Nearest Neighbors performed best for the Wine Quality dataset, while Logistic Regression excelled on QSAR Biodegradation.
- The Voting Classifier effectively combined individual model strengths, offering balanced performance across multiple metrics.
- Manual grid search and built-in library functions yielded identical results, demonstrating the reliability of scikit-learn while highlighting the extra effort required for manual implementation.

The main takeaway from this lab is that systematic model selection and hyperparameter tuning are critical for achieving optimal predictive performance. Using a library like scikit-learn not only saves time but also reduces the likelihood of errors, while manual implementation helps deepen understanding of the underlying process. Overall, this exercise reinforced the

importance of pipelines, cross-validation, and careful evaluation in building robust machine learning models.