Project Title:

Manual vs. Scikit-learn GridSearchCV for Hyperparameter Tuning and Model Comparison

Name:

Maya Nithyanand Bhagath

Student ID:

PES2UG23CS331

Course Name:

Machine Learning Lab

Submission Date:

30th August, 2025

1. Introduction

This project is used to build and optimize classification models for two datasets: Wine Quality and QSAR Biodegradation. The goal is to predict a binary outcome for each dataset.

Hyperparameter tuning was used for different algorithms: Decision Tree, k-NN and Logistic Regression.

Two methods were used: manual implementation of grid search with cross validation and scikit-learn's built-in GridSearchCV. The performance of these models was compared using various metrics.

---

2. Dataset Description

- **Wine Quality Dataset**
    - Features: 11
    - Instances: 1599
    - Target variable: A binary variable representing good quality (rating>5).
- **QSAR Biodegradation Dataset**
    - Features: 41
    - Instances: 1055
    - Target variable: A binary variable indicating if a chemical is "ready biodegradable" (RB) or not.

---

3. Methodology

- **Key Concepts**
    - ***Hyperparameter Tuning:*** The process of finding the optimal combination of hyperparameters for a learning algorithm. Hyperparameters are those parameters which can be set before the learning process begins.
    - ***Grid Search:*** This is a hyperparameter tuning technique where the subsets of the hyperparameter space of an algorithm is searched. The models are trained and evaluated for each combination of parameters to find the best-performing model.
    - ***K-fold Cross Validation:*** The data is split into 'k' subsets or folds. The model is trained on k-1 folds and validated on the remaining fold. This process is repeated k times. The average of the results is used to produce a better estimate of the values.

- **ML Pipeline**
    - ***StandardScalar:*** standardizes features by removing the mean and scaling to unit variance.
    - ***SelectKBest:*** Selects the top 'k' features based on their statistical scores from the ANOVA F-test. The number of features 'k' was one of the hyperparameters tuned during the grid search.
    - ***Classifier:*** The algorithm being trained.

- **Implementation Process**
  - ***Manual Implementation (Part 1):***
    - For each classifier, a grid of hyperparameters was defined.
    - All possible combinations of the hyperparameters were generated using itertools.product.
    - A 5-fold StratifiedKFold splitter was created.
    - The code iterated through every parameter combination.
    - In each fold, a new pipeline was built, trained on the training portion, and used to predict probabilities on the validation portion.
    - The ROC AUC score was calculated for each fold, and the average score across the 5 folds was computed.
    - The parameter combination with the highest average ROC AUC was stored as the best.
    - A new parameter was initialized with the best parameters and trained on the entire training dataset.
  - ***Scikit-learn Implementation (Part 2):***
    - For each classifier, the same pipeline and parameter grid from the manual implementation were used.
    - A GridSearchCV object was created, specifying the pipeline, parameter grid, 5-fold stratified cross-validation, and roc_auc as the scoring metric.
    - Scikit-learn automatically handled the cross-validation, model training, and evaluation for all parameter combinations.

---

4. Results and Analysis

Performance Tables:

• Wine Quality Dataset

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.7271	0.7716	0.6965	0.7321	0.8025
k-Nearest Neighbors	0.7812	0.7836	0.8171	0.8000	0.8589
Logistic Regression	0.7333	0.7549	0.7432	0.7490	0.8242
Voting Classifier	0.7625	0.7761	0.7821	0.7791	0.8600

• QSAR Biodegradation Dataset

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.7634	0.6231	0.7570	0.6835	0.8049
k-Nearest Neighbors	0.8549	0.7905	0.7757	0.7830	0.8985
Logistic Regression	0.8644	0.8200	0.7664	0.7923	0.9082
Voting Classifier	0.8486	0.7921	0.7477	0.7692	0.9004

Comparison of Implementations:

- For both datasets, the results from the manual and Scikit-learn implementations were identical.
- This is because the base methods used in both the implementations are the same: the data splits, the cross-validation strategy, the same models, and the same parameter grids.

Visualizations and Analysis:

- ***Wine Quality Dataset***
  - *ROC Curve Analysis:* The ROC curves for the Wine Quality dataset show that the k-Nearest Neighbors and Voting Classifier models perform the best, with the highest Area Under the Curve (AUC) scores of 0.859 and 0.860, respectively.
  - *Confusion Matrix Analysis:* There are plenty of true positives and true negatives, but also some false positives and false negatives.

- ***QSAR Biodegradation Dataset***
  - *ROC Curve Analysis:* Logistic Regression clearly has the highest AUC of 0.908. The k-NN and Voting Classifier models also perform very strongly with AUCs around 0.900.
  - *Confusion Matrix Analysis:* This had a higher number of false negatives than false positives, making it more likely to wrongly classify biodegradable as non-biodegradable.

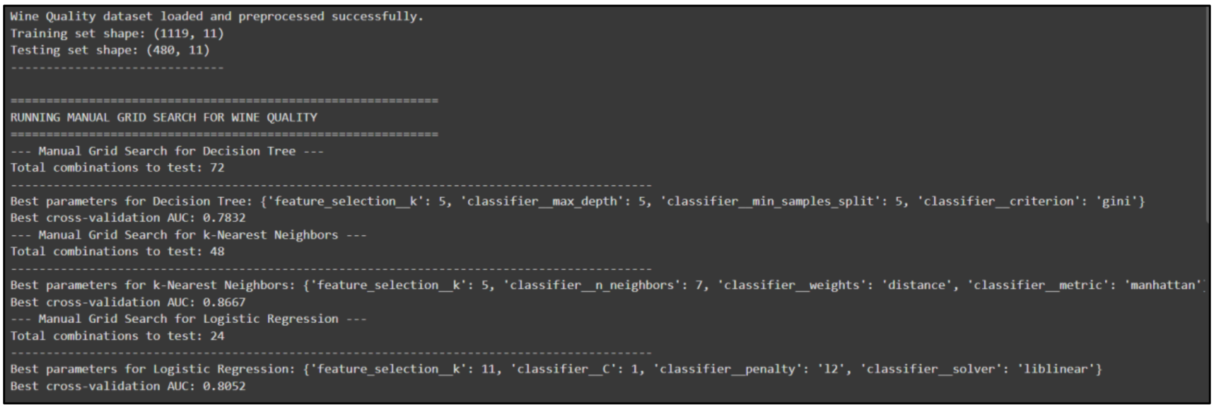
Best Model Analysis:

- ***Wine Quality Dataset***
  - The Voting Classifier model achieved the highest ROC AUC value.
- ***QSAR Biodegradation Dataset:***
  - The Logistic Regression model performed best overall, with the highest accuracy, precision, F1-score, and ROC AUC.

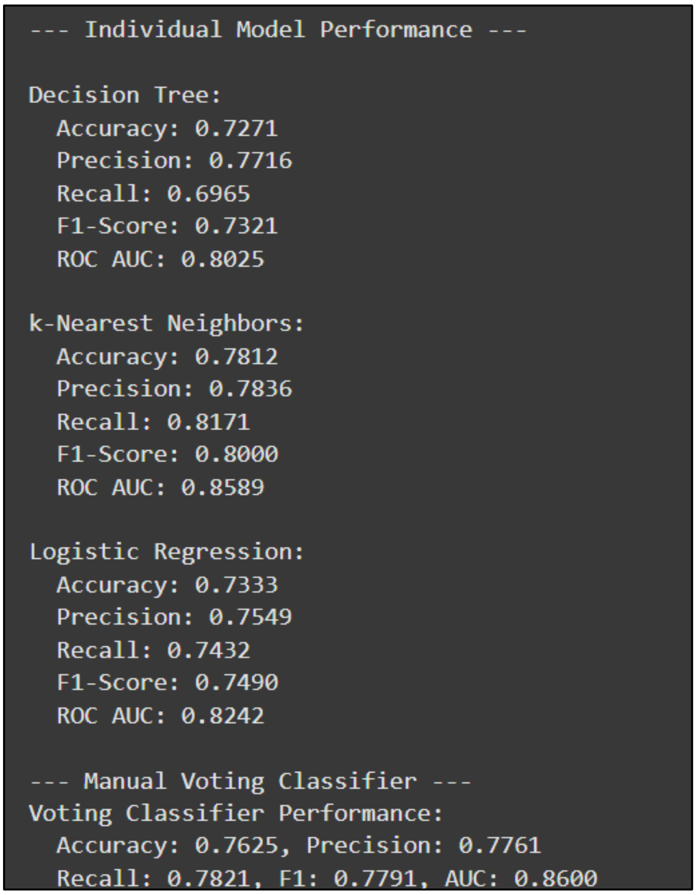
---

5. Screenshots

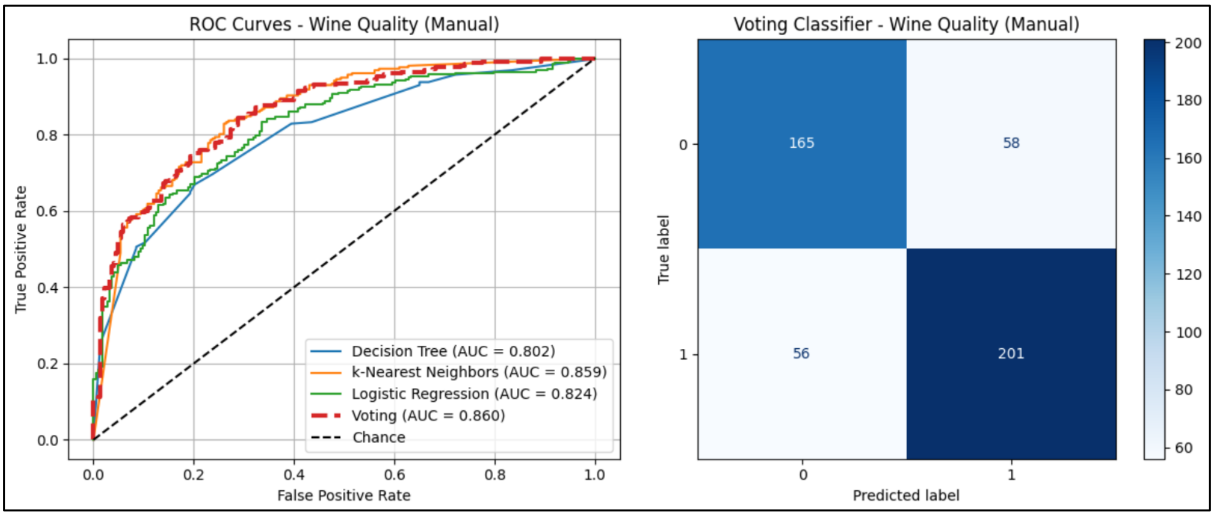
• Wine Quality Dataset



```
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
```

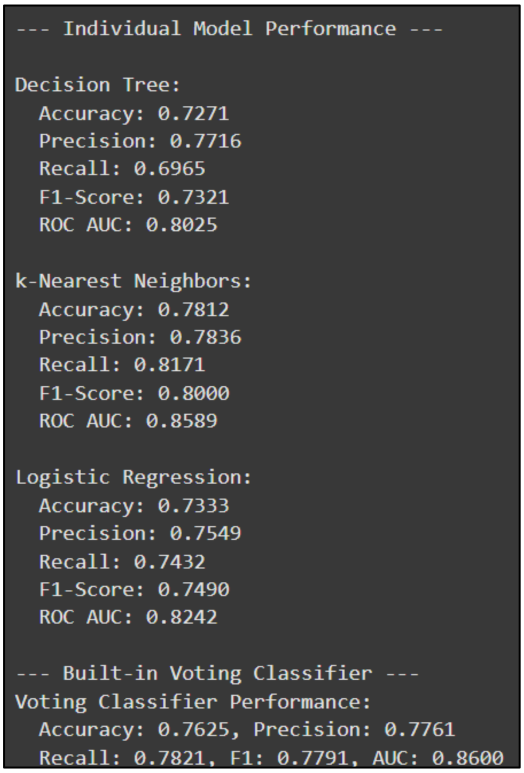


```
--- Individual Model Performance ---

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
```

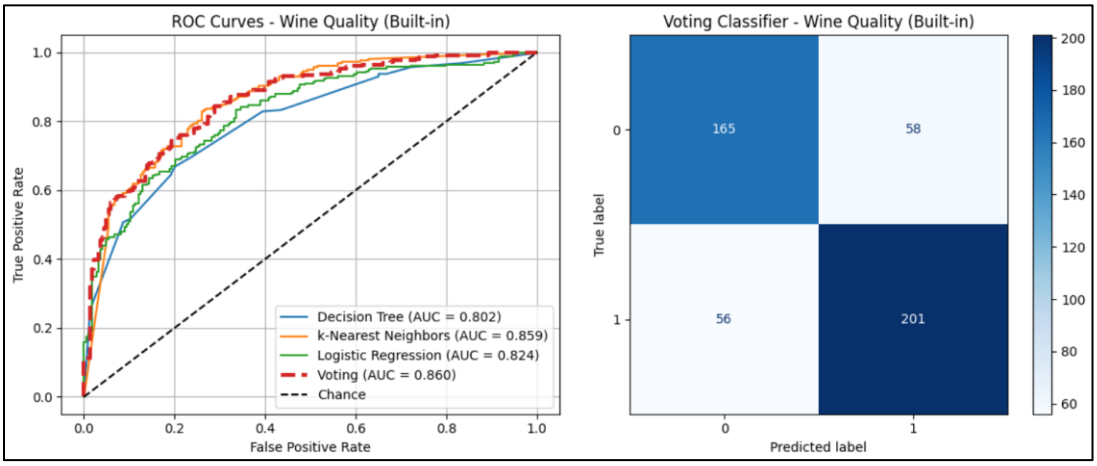


ROC Curves - Wine Quality (Manual)

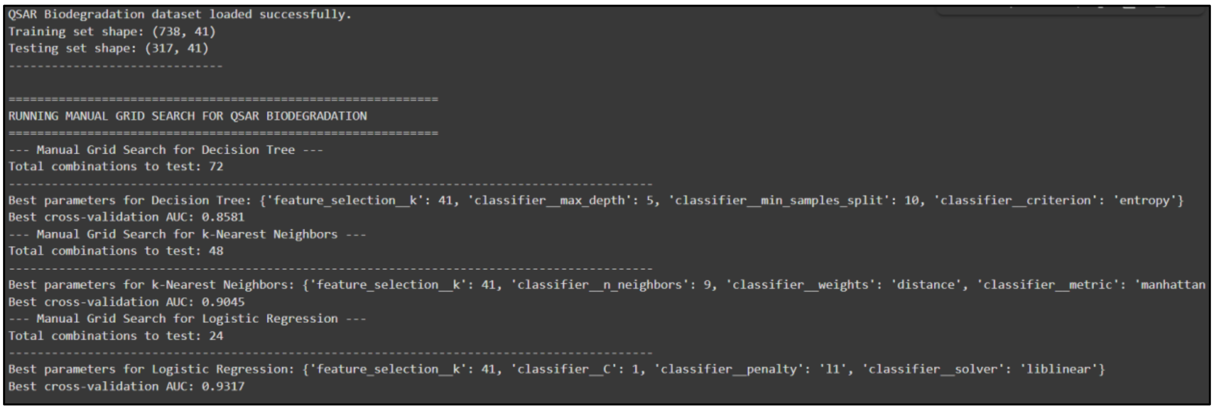
Voting Classifier - Wine Quality (Manual)



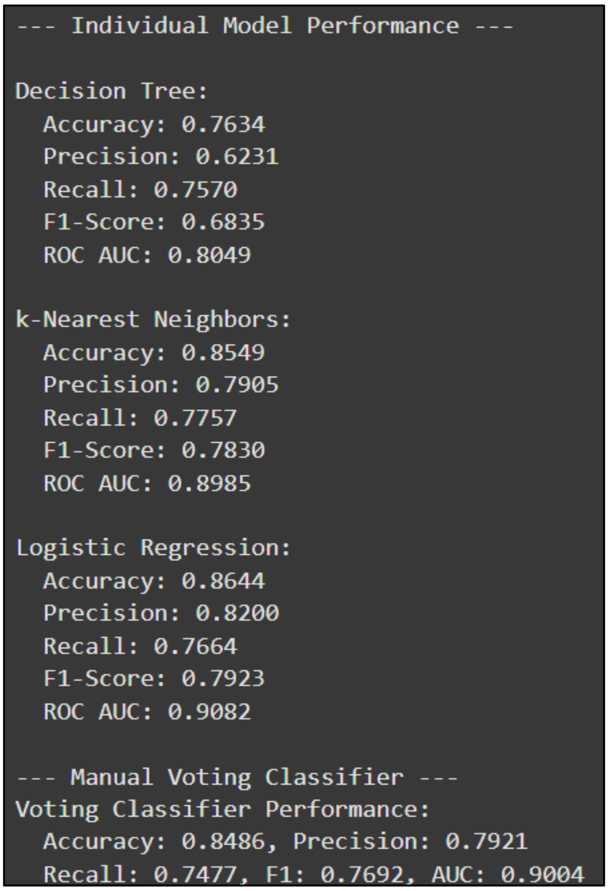
```
--- Individual Model Performance ---

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
```



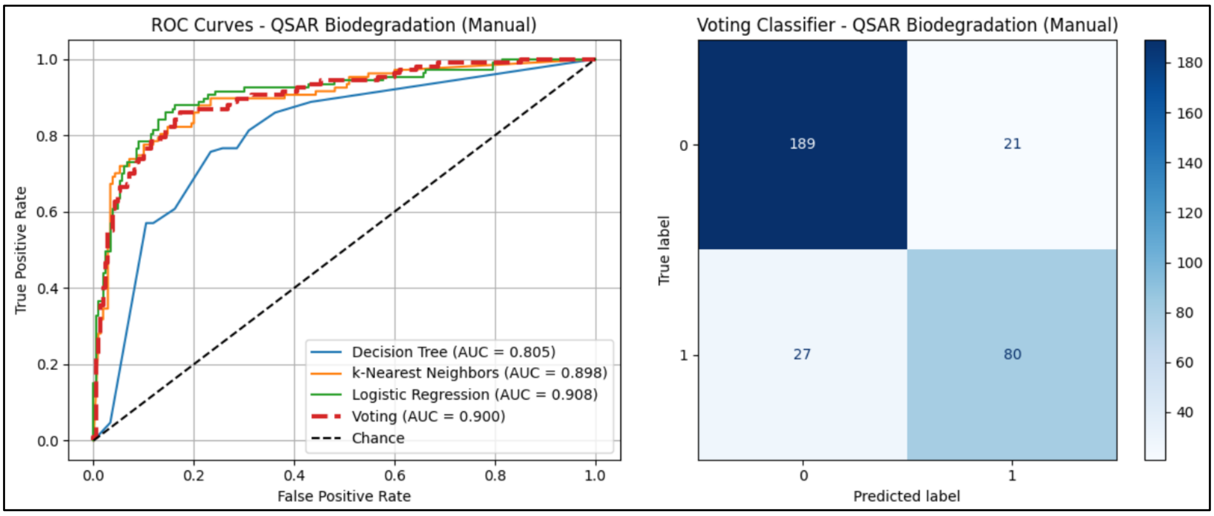
- ***QSAR Biodegradation Dataset***

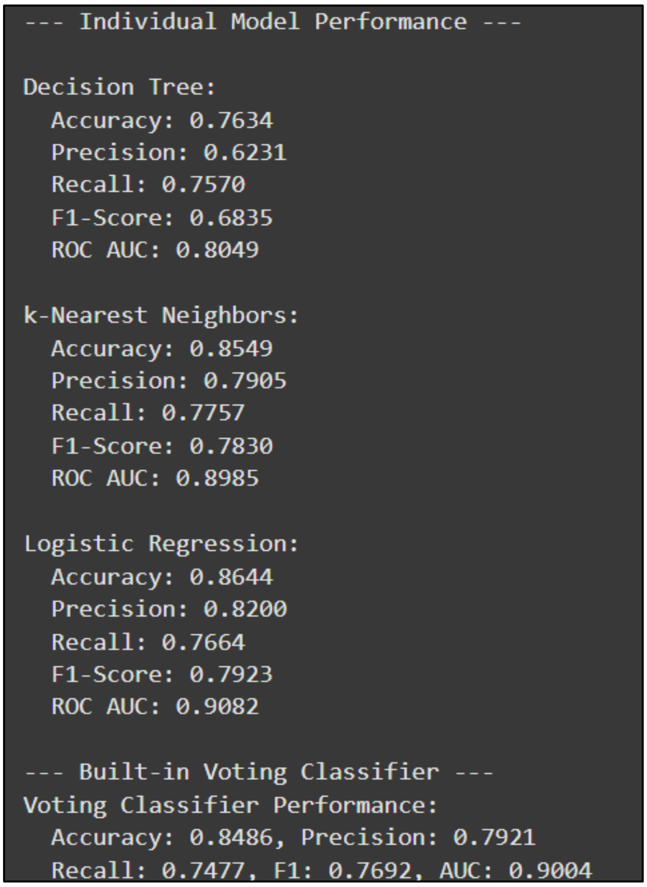


```
QSAR Biodegradation dataset loaded successfully.
Training set shape: (738, 41)
Testing set shape: (317, 41)
------------------------------

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

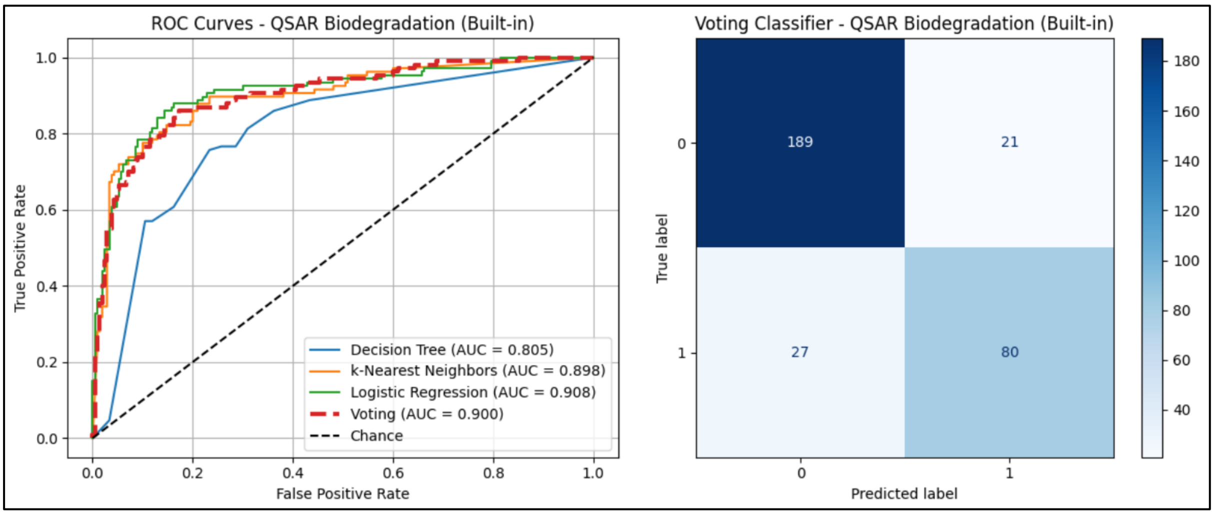


```
--- Individual Model Performance ---

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
```





```
--- 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
```



6. Conclusion

Key Findings:

- The manual and Scikit-learn implementation both produces the same results.
- In the Wine Quality Dataset, k-NN was the optimal model.
- In the QSAR Biodegradation dataset, the Logistic Regression model was the optimal model.

Main Takeaways:

- This lab highlights the understanding about manual implementation and Scikit-learn implementation.
- The manual implementation gives us a deep understanding of the cross validation and parameter iteration. However, it is prone to error and takes more time.
- The Scikit-learn implementation uses the built-in function GridSearchCV, and therefore is much more concise and less error-prone.