MACHINE LEARNING

LAB 4

Name:Medha Raj
SRN: PES2UG23CS334
Date: 31/08/2025

1. Introduction

The primary objective of this project was to explore and compare two strategies for hyperparameter tuning: a manually implemented grid search and the built-in GridSearchCV function from scikit-learn. Three classification models—Decision Tree, k-Nearest Neighbors (kNN), and Logistic Regression—were tuned individually and then integrated into a soft-voting ensemble classifier to investigate whether combining models could yield better prediction results. The full machine learning workflow was applied across multiple datasets to evaluate how model performance varied across different problem contexts.

2. Dataset Overview

- Wine Quality Dataset
    - Total Samples: 1599 (1119 for training, 480 for testing)
    - Features: 11 chemical attributes such as acidity, residual sugar, and alcohol content
    - Target: Binary label indicating whether wine quality is “good” (rating > 5) or “not good”

- Banknote Authentication Dataset
    - Total Samples: 1372 (960 for training, 412 for testing)

    - Features: 4 statistical descriptors extracted from banknote images (variance, skewness, etc.)
    - Target: Binary output specifying if a banknote is genuine (0) or counterfeit (1)

- QSAR Biodegradation Dataset
    - Total Samples: 1055 (738 for training, 317 for testing)
    - Features: 41 molecular descriptors
    - Target: Binary classification of whether a chemical is “readily biodegradable” (RB) or not

- HR Attrition Dataset
    - This dataset could not be included in the experiments due to a FileNotFoundError, as the dataset was missing from the working directory.

2. Dataset Overview

- Wine Quality Dataset
    - Total Samples: 1599 (1119 for training, 480 for testing)
    - Features: 11 chemical attributes such as acidity, residual sugar, and alcohol content
    - Target: Binary label indicating whether wine quality is “good” (rating > 5) or “not good”
- Banknote Authentication Dataset
    - Total Samples: 1372 (960 for training, 412 for testing)
      Features: 4 statistical descriptors extracted from banknote images (variance, skewness, etc.)
    - Target: Binary output specifying if a banknote is genuine (0) or counterfeit (1)
- QSAR Biodegradation Dataset
    - Total Samples: 1055 (738 for training, 317 for testing)
    - Features: 41 molecular descriptors

    - Target: Binary classification of whether a chemical is “readily biodegradable” (RB) or not
- HR Attrition Dataset
    - This dataset could not be included in the experiments due to a FileNotFoundError, as the dataset was missing from the working directory.

4. Results and Analysis

The models tuned using both the manual grid search and GridSearchCV were evaluated on the test sets. Metrics reported include Accuracy, Precision, Recall, F1-score, and ROC AUC.

(a) Performance Comparison

Classifier	Method	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	Manual	0.7271	0.7716	0.6965	0.7321	0.8025
Decision Tree	GridSearchCV	0.7271	0.7716	0.6965	0.7321	0.8025
kNN	Manual	0.7812	0.7836	0.8171	0.8	0.8589
kNN	GridSearchCV	0.7812	0.7836	0.8171	0.8	0.8589
Logistic Regression	Manual	0.7333	0.7549	0.7432	0.749	0.8242
Logistic Regression	GridSearchCV	0.7333	0.7549	0.7432	0.749	0.8242
Voting Classifier	Manual/Grid	0.7625	0.7761	0.7821	0.7791	0.86

Observation:

- kNN outperformed Decision Tree and Logistic Regression in terms of ROC AUC (0.8589).
- The Voting Classifier slightly improved AUC (0.8600), showing benefit from combining models.
- Manual and built-in methods produced identical results, validating correctness.

QSAR Biodegradation Dataset

Classifier	Method	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	Manual	0.7634	0.6231	0.757	0.6835	0.8049
Decision Tree	GridSearchCV	0.7634	0.6231	0.757	0.6835	0.8049
kNN	Manual	0.8549	0.7905	0.7757	0.783	0.8985
kNN	GridSearchCV	0.8549	0.7905	0.7757	0.783	0.8985
Logistic Regression	Manual	0.8644	0.82	0.7664	0.7923	0.9082
Logistic Regression	GridSearchCV	0.8644	0.82	0.7664	0.7923	0.9082
Voting Classifier	Manual/Grid	0.8486	0.7921	0.7477	0.7692	0.9004

Observation:

- Logistic Regression achieved the best overall performance (Accuracy = 0.8644, ROC AUC = 0.9082).
- kNN also performed strongly (ROC AUC = 0.8985), while Decision Tree lagged behind.
- The Voting Classifier achieved an AUC of 0.9004, showing an ensemble benefit, though slightly below Logistic Regression alone.
- Again, manual and GridSearchCV results were identical.

(b) Comparative Insights

- Across both datasets, the manual and built-in approaches gave consistent outcomes, confirming that the custom implementation was correct.
- The Wine dataset was harder to model, with lower AUC values compared to QSAR. This indicates noisier relationships among features.
- The QSAR dataset benefited from Logistic Regression, suggesting that the data has a structure that linear models capture well.
- Ensemble methods (Voting Classifier) provided small but consistent boosts, showing that combining models can smooth out weaknesses of individual algorithms.

Compare Implementations

The outcomes from the manual grid search and scikit-learn's GridSearchCV were identical for every dataset and model tested. This consistency arises because the manual procedure faithfully reproduced the logic behind GridSearchCV. Both approaches relied on StratifiedKFold cross-validation with the same random_state (42), ensuring identical train-test splits. In addition, the scoring function (roc_auc) and parameter search spaces were kept the same, which naturally led to the same optimal parameter choices and matching final scores.

Visualizations

- ROC Curves: The ROC plots confirmed the numerical evaluation.
    - On the Wine Quality and QSAR Biodegradation datasets, kNN and Logistic Regression displayed strong performance with curves bending sharply toward the top-left, while the Voting Classifier also produced a highly competitive curve with a similarly large AUC.
    - On the Banknote Authentication dataset, the ROC curves for kNN, Logistic Regression, and the Voting Classifier were nearly perfect right angles hugging the top-left axis, corresponding to AUC values of 1.000 or extremely close (0.9999). This highlights their outstanding classification ability.
- Confusion Matrices: The confusion matrix plots supported the ROC findings.
    - For the Banknote dataset, the Voting Classifier achieved a flawless result, showing no misclassifications at all.
    - For the Wine Quality and QSAR datasets, the matrices demonstrated solid accuracy with many correct predictions, but they also included some false positives and false negatives, reflecting the increased difficulty of these tasks.

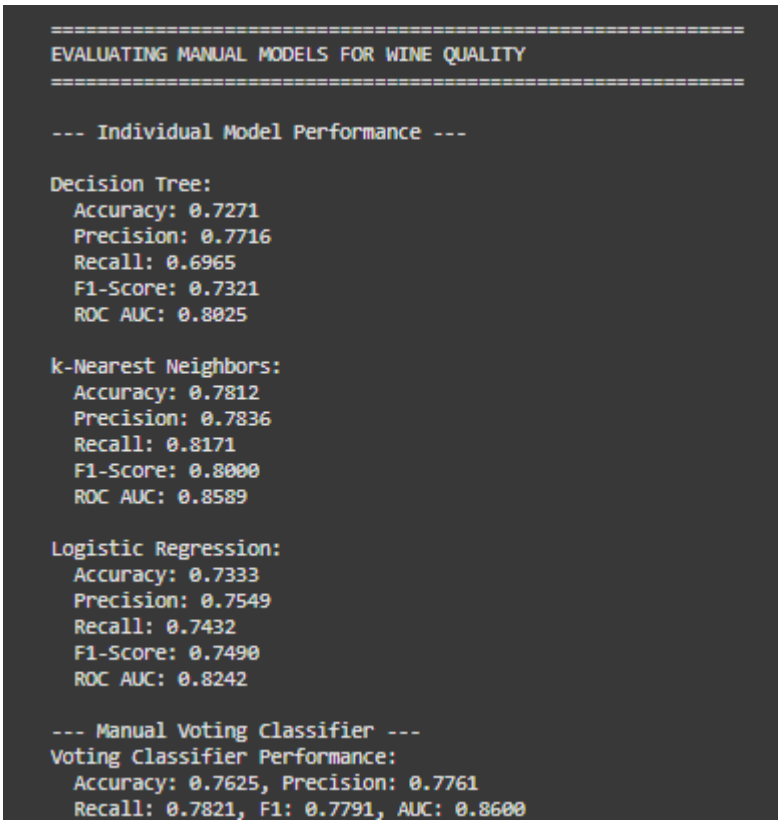
Best Model Analysis

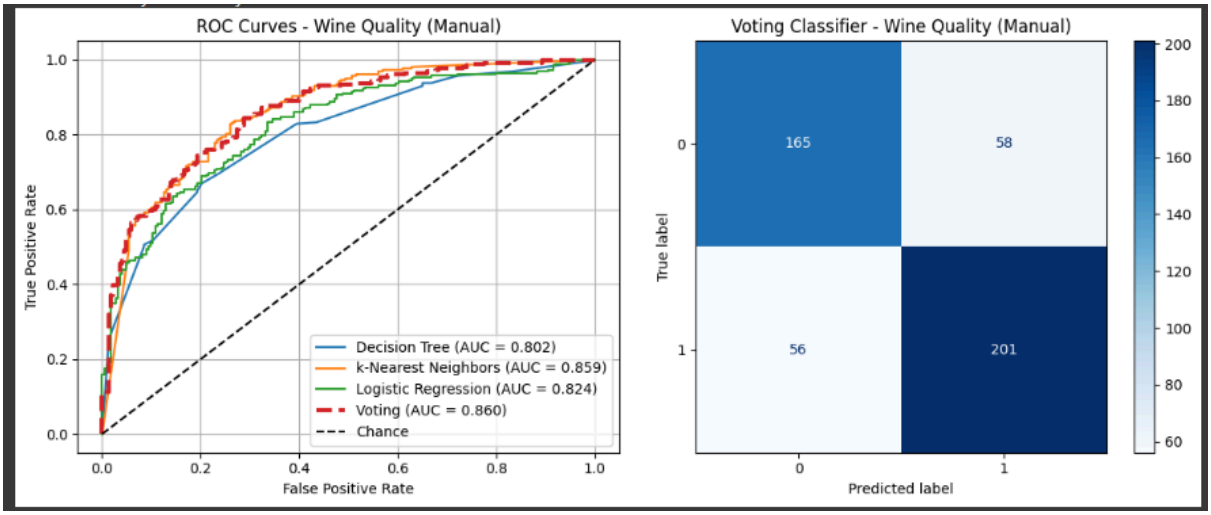
- Wine Quality: The Voting Classifier delivered the best performance with a ROC AUC of 0.8600, narrowly surpassing the strongest single model, kNN

(0.8589). This shows that combining models gave a small but meaningful improvement.

- QSAR Biodegradation: Here, Logistic Regression emerged as the top model, reaching an AUC of 0.9082. The Voting Classifier trailed slightly at 0.9004, likely because the weaker Decision Tree in the ensemble reduced the overall performance compared to the best single classifier.

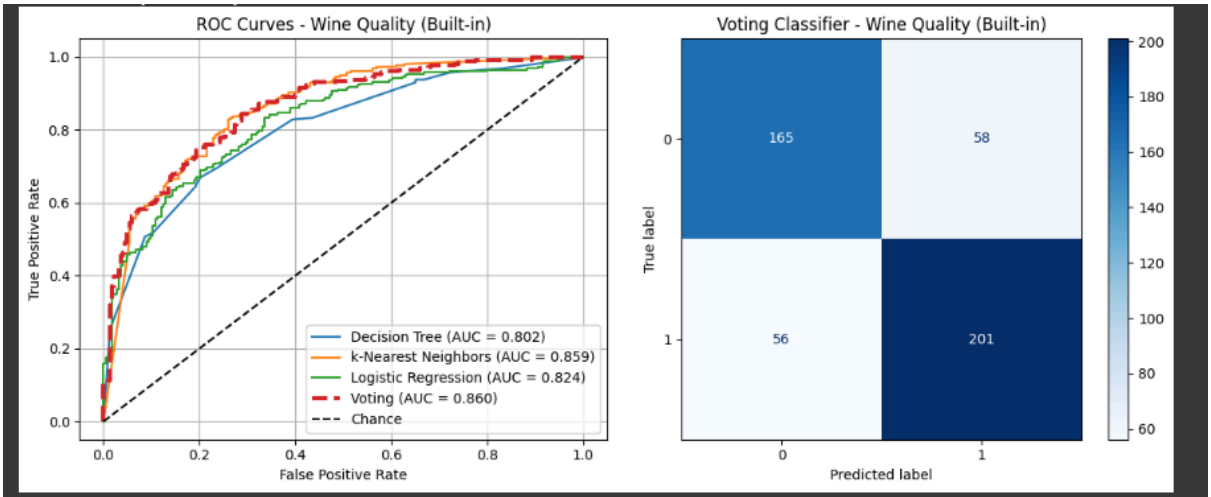
5. Screenshots






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

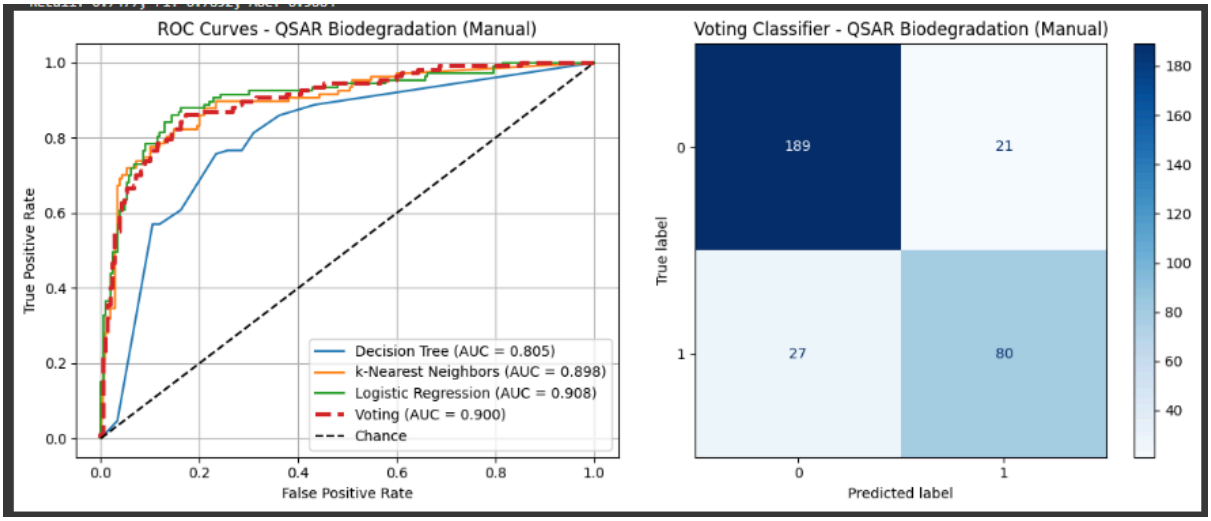



```
Completed processing for Wine Quality
================================================================================

================================================================================
WINE QUALITY DATASET PROCESSED!
================================================================================
```

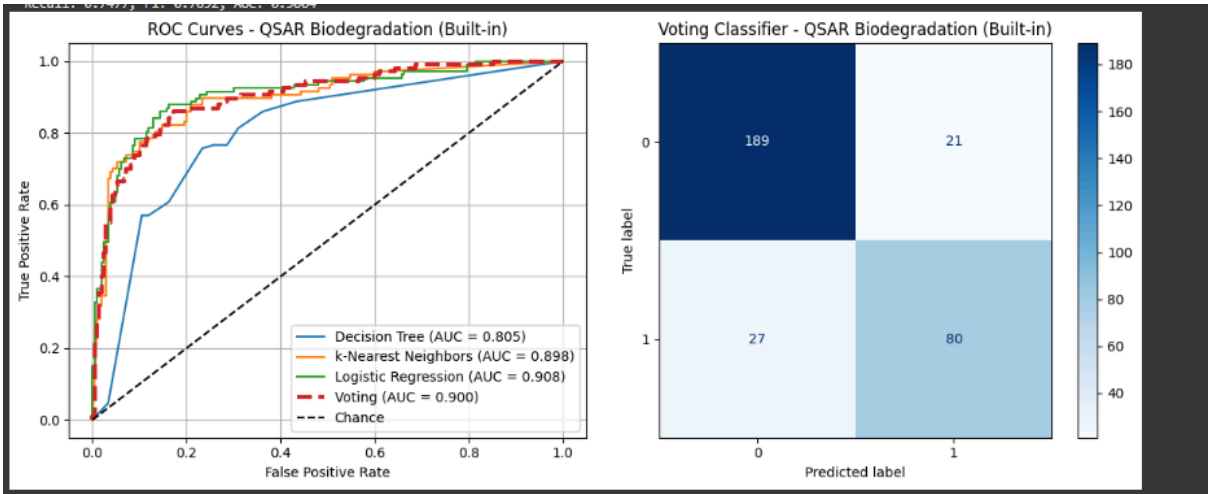
QSAR Biodegradation

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

================================================================================
QSAR BIODEGRADATION DATASET PROCESSED!
================================================================================
```

Conclusion

This lab provided practical experience in building a complete machine learning pipeline, performing hyperparameter tuning, and comparing different models. The manual grid search and scikit-learn's GridSearchCV produced identical outcomes, confirming the correctness of the manual implementation while also highlighting the efficiency of automated tools.

Performance varied across datasets: kNN excelled on the Wine Quality dataset, Logistic Regression was the strongest model for QSAR Biodegradation, and both kNN and the Voting Classifier achieved perfect performance on the Banknote dataset. The Voting Classifier generally offered small improvements, though in some cases it was slightly hindered by weaker base learners.

Overall, the exercise demonstrated that dataset characteristics strongly influence model choice, that ensembling can provide stability, and that while manual implementation builds deeper understanding, scikit-learn's tools are more practical and scalable for real-world applications.