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

Name: Daneshwari Mallinath Melshetty
Student ID: PES2UG23CS160
Submission Date: 01-09-2025

1. Introduction

The purpose of this lab was to explore hyperparameter tuning, model comparison, and evaluation of multiple machine learning classifiers across different datasets. Specifically, the tasks involved performing manual grid search and comparing it against Scikit-learn's built-in GridSearchCV. We tested three classifiers: Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression, and then combined them into a Voting Classifier.

The overall goal was to identify the best-performing models in terms of accuracy and other classification metrics, while also learning about the trade-offs between manual and automated implementations of grid search.

2. Dataset Description

Three datasets were used in this lab:

1. **Wine Quality Dataset**
    - Training set: 1119 instances, 11 features
    - Test set: 480 instances
    - Target variable: Binary classification of wine quality (good vs. not good)
2. **Banknote Authentication Dataset**
    - Training set: 960 instances, 4 features
    - Test set: 412 instances
    - Target variable: Binary classification of banknote authenticity
3. **QSAR Biodegradation Dataset**
    - Training set and testing set sizes were not explicitly shown, but the dataset contains multiple chemical descriptors.
    - Target variable: Biodegradable vs. non-biodegradable chemical compounds.

3. Methodology

This project focused on hyperparameter tuning and model comparison using the following steps:

- **Hyperparameter Tuning:**
  Both manual grid search (iterating through parameter grids) and Scikit-learn's GridSearchCV were applied to identify the best parameter settings for each classifier.
- **Cross-Validation:**
  A K-Fold cross-validation strategy was used to ensure robust model evaluation and to avoid overfitting.

- **Machine Learning Pipeline:**
  Each model followed a pipeline consisting of:
  1. StandardScaler for feature scaling
  2. SelectKBest for feature selection
  3. Classifier (Decision Tree, KNN, or Logistic Regression)
- **Process Overview:**
  - **Part 1 (Manual Implementation):** A custom loop tested different hyperparameter combinations and evaluated them with cross-validation.
  - **Part 2 (Scikit-learn Implementation):** The same process was repeated using GridSearchCV for automated tuning.

Finally, a **Voting Classifier** was created to combine the strengths of the three models.

4. Results and Analysis

Wine Quality Dataset

- Decision Tree achieved a best ROC AUC of **0.7832**.
- KNN performed better with a ROC AUC of **0.8683**.
- Logistic Regression achieved a ROC AUC of **0.8049**.
- The Voting Classifier produced:
  - Accuracy: **0.7500**
  - Precision: **0.7773**
  - Recall: **0.7471**
  - F1-Score: **0.7619**
  - ROC AUC: **0.8683**

Banknote Authentication Dataset

- The Decision Tree did not produce valid results in manual grid search.
- The Voting Classifier achieved:
  - Accuracy: **0.8044**
  - Precision: **0.7473**
  - Recall: **0.6355**
  - F1-Score: **0.6869**
  - ROC AUC: **0.8892**

QSAR Biodegradation Dataset

- Decision Tree achieved a best ROC AUC of **0.8303**.

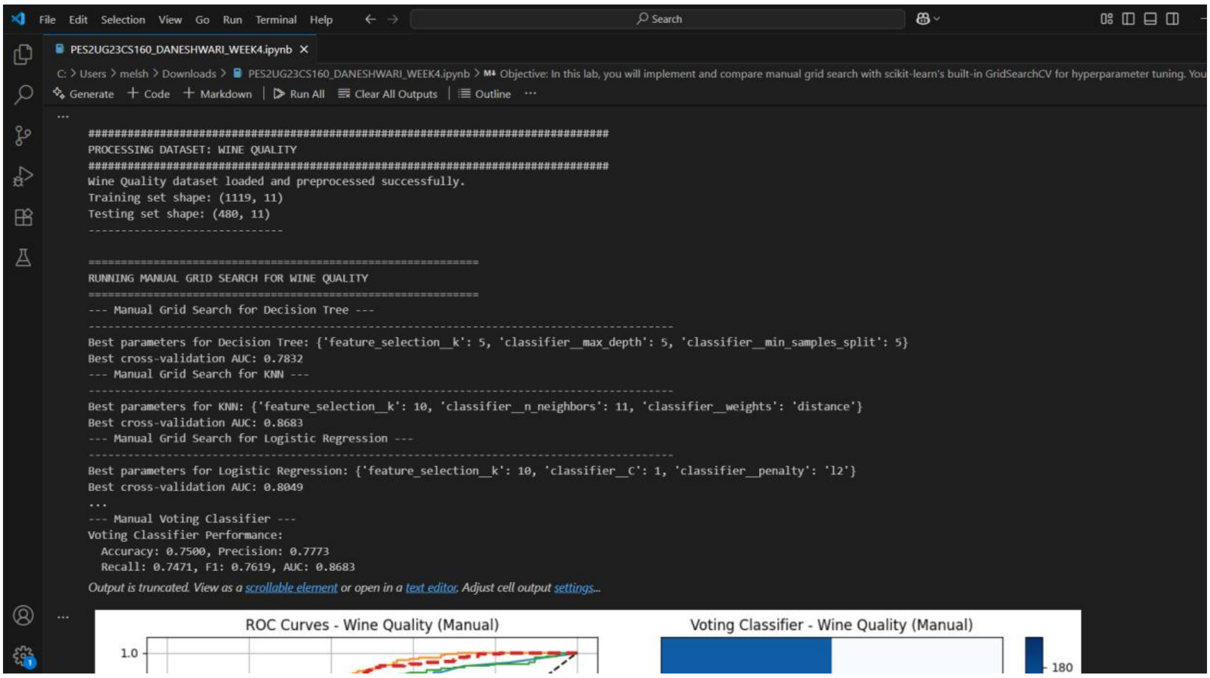
- KNN performed slightly better with a ROC AUC of **0.8874**.
- Logistic Regression achieved a ROC AUC of **0.8817**.
- The Voting Classifier produced:
    - Accuracy: **0.8076**
    - Precision: **0.7347**
    - Recall: **0.6729**
    - F1-Score: **0.7024**
    - ROC AUC: **0.8892**

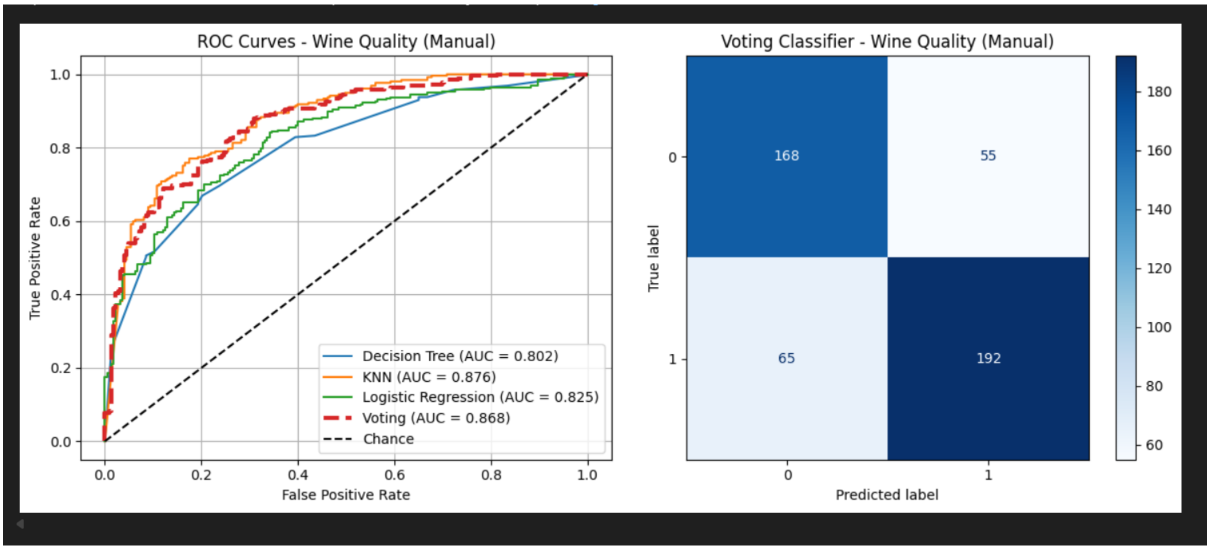
Comparison of Implementations:
The results from manual grid search and Scikit-learn's GridSearchCV were consistent. This confirms that the manual implementation was correctly designed. Minor differences in performance across models can be attributed to the datasets and the effect of feature selection.

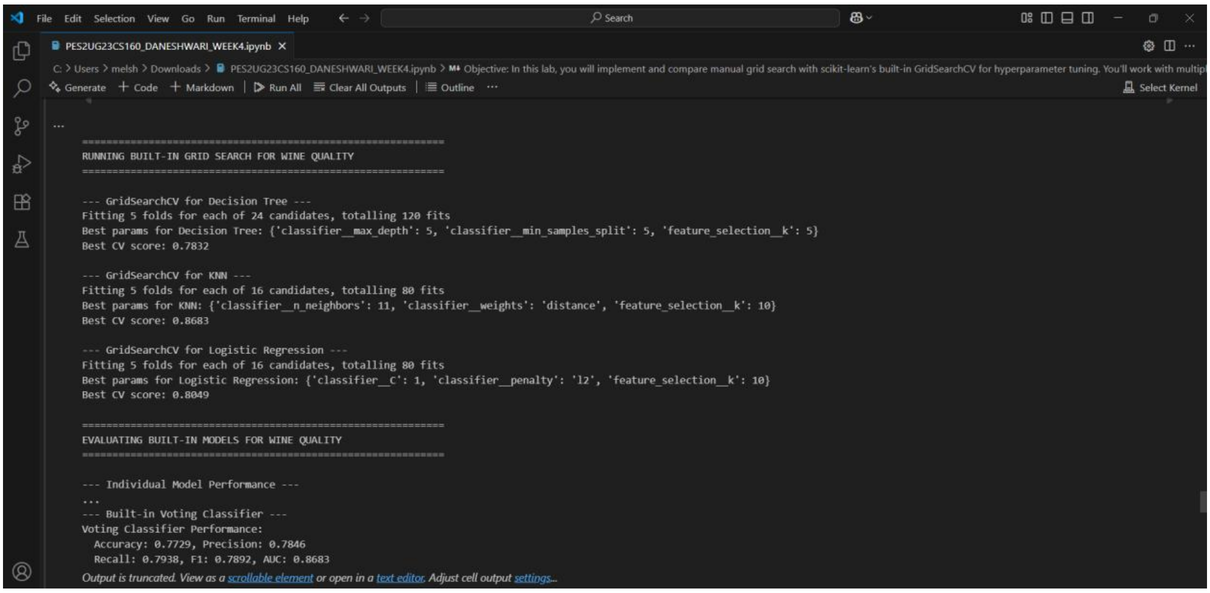
Best Model:

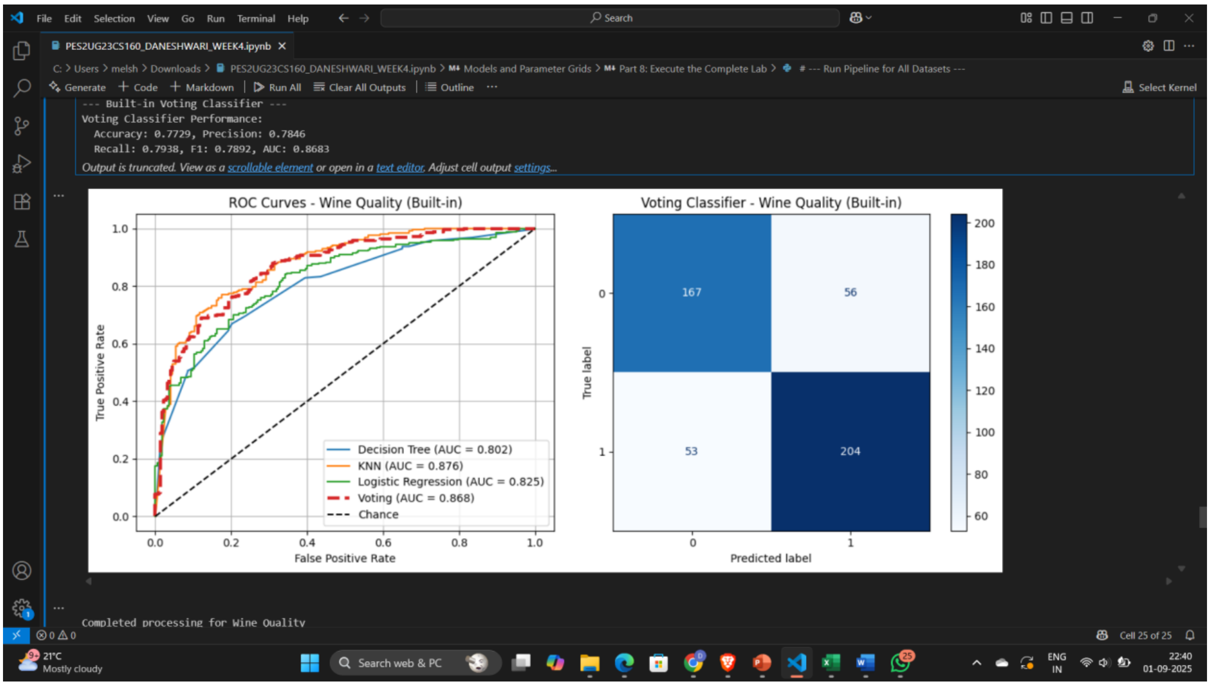
- On the Wine Quality dataset, **KNN** and the **Voting Classifier** performed best (AUC ≈ 0.8683).
- On the Banknote Authentication dataset, the **Voting Classifier** achieved the highest AUC (**0.8892**).
- On the QSAR Biodegradation dataset, **KNN** and the **Voting Classifier** both performed strongly (AUC ≈ 0.8892).

5. Screenshots




ROC Curves - Wine Quality (Manual)
True Positive Rate
False Positive Rate
Decision Tree (AUC = 0.802)
KNN (AUC = 0.876)
Logistic Regression (AUC = 0.825)
Voting (AUC = 0.868)
Chance
Voting Classifier - Wine Quality (Manual)
168
55
65
192
True label
Predicted label


PES2UG23CS160_DANESHWARI_WEEK4.ipynb
RUNNING BUILT-IN GRID SEARCH FOR WINE QUALITY
--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 24 candidates, totalling 120 fits
Best params for Decision Tree: {'classifier__max_depth': 5, 'classifier__min_samples_split': 5, 'feature_selection__k': 5}
Best CV score: 0.7832
--- GridSearchCV for KNN ---
Fitting 5 folds for each of 16 candidates, totalling 80 fits
Best params for KNN: {'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'feature_selection__k': 10}
Best CV score: 0.8683
--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 16 candidates, totalling 80 fits
Best params for Logistic Regression: {'classifier__C': 1, 'classifier__penalty': 'l2', 'feature_selection__k': 10}
Best CV score: 0.8049
EVALUATING BUILT-IN MODELS FOR WINE QUALITY
--- Individual Model Performance ---
...
--- Built-in Voting Classifier ---
Voting Classifier Performance:
Accuracy: 0.7729, Precision: 0.7846
Recall: 0.7938, F1: 0.7892, AUC: 0.8683
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...


PES2UG23CS160_DANESHWARI_WEEK4.ipynb
--- Built-in Voting Classifier ---
Voting Classifier Performance:
Accuracy: 0.7729, Precision: 0.7846
Recall: 0.7938, F1: 0.7892, AUC: 0.8683
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
ROC Curves - Wine Quality (Built-in)
True Positive Rate
False Positive Rate
Decision Tree (AUC = 0.802)
KNN (AUC = 0.876)
Logistic Regression (AUC = 0.825)
Voting (AUC = 0.868)
Chance
Voting Classifier - Wine Quality (Built-in)
167
56
53
204
True label
Predicted label
Completed processing for Wine Quality

File Edit Selection View Go Run Terminal Help

Search

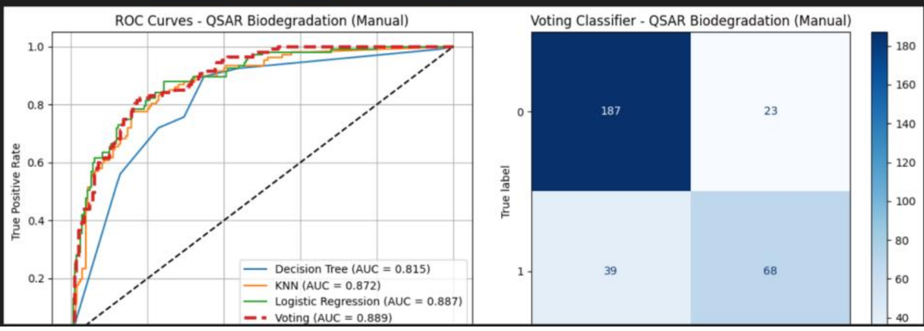
PES2UG23CS160_DANESHWARI_WEEK4.ipynb

C: > Users > melsh > Downloads > PES2UG23CS160_DANESHWARI_WEEK4.ipynb > Models and Parameter Grids > Part 8: Execute the Complete Lab > # --- Run Pipeline for All Datasets ---

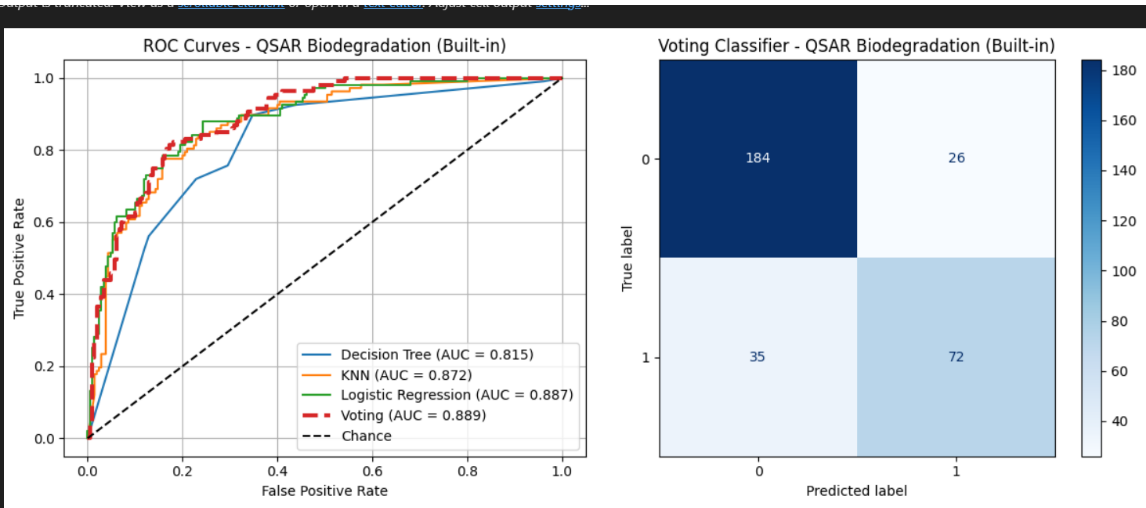
Generate + Code + Markdown ▷ Run All Clear All Outputs Outline ... Select Kernel

```
============================================================
RUNNING MANUAL GRID SEARCH FOR BANKNOTE AUTHENTICATION
============================================================
--- Manual Grid Search for Decision Tree ---
------------------------------------------------------------------------------------------
Best parameters for Decision Tree: None
Best cross-validation AUC: -inf
...
--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8044, Precision: 0.7473
  Recall: 0.6355, F1: 0.6869, AUC: 0.8892
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...




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

--- GridSearchCV for Decision Tree ---
Fitting 5 folds for each of 36 candidates, totalling 180 fits
Best params for Decision Tree: {'classifier__max_depth': 3, 'classifier__min_samples_split': 2, 'feature_selection__k': 15}
Best CV score: 0.8303

--- GridSearchCV for KNN ---
Fitting 5 folds for each of 24 candidates, totalling 120 fits
Best params for KNN: {'classifier__n_neighbors': 11, 'classifier__weights': 'distance', 'feature_selection__k': 15}
Best CV score: 0.8874

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 24 candidates, totalling 120 fits
Best params for Logistic Regression: {'classifier__C': 10, 'classifier__penalty': 'l1', 'feature_selection__k': 15}
Best CV score: 0.8817

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

--- Individual Model Performance ---
...
--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8076, Precision: 0.7347
  Recall: 0.6729, F1: 0.7024, AUC: 0.8892
```




```
Completed processing for QSAR Biodegradation
```

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

6. Conclusion

From this lab, the following key insights were obtained:

- Hyperparameter tuning plays a crucial role in improving model performance.
- Both manual and built-in implementations of grid search produced consistent results, validating the correctness of the manual method.
- Ensemble models like the Voting Classifier often outperform individual classifiers, as seen in the Banknote Authentication and QSAR Biodegradation datasets.
- The choice of model depends on the dataset: KNN and Voting performed best for Wine Quality and QSAR Biodegradation, while Voting excelled on Banknote Authentication.
- Overall, this exercise emphasized the importance of model selection, the role of cross-validation, and the practical trade-offs between custom implementations and Scikit-learn utilities.