

MACHINE LEARNING LAB 4

NAME : MOHAMMED MUSHARRAF

SRN : PES2UG23CS915

SECTION : F

Introduction

Implement and compare **manual grid search** and **GridSearchCV** for hyperparameter tuning across multiple classifiers.

The models are evaluated using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC AUC.

Dataset Description

Dataset 1: HR Attrition

- Source: IBM HR Analytics dataset
- Target: Attrition (Yes/No → binary classification)
- Size: ~1470 rows, ~35 features
- Note the **class imbalance** (most employees stay, few leave).

Dataset 2: Wine Quality (or whichever second dataset you choose)

- Source: UCI Wine Quality dataset
- Target: Good quality wine (quality > 5 → 1 else 0)
- Size: ~1600 rows, 11 features

Methodology

1. Preprocessed the dataset (train/test split, encoding if required, standard scaling).
2. Defined classifiers: Decision Tree, kNN, Logistic Regression.
3. Performed **Manual Grid Search** (loop over parameter combinations).
4. Performed **GridSearchCV** (built-in).

5. Evaluated the best models using test data.
6. Used a **Voting Classifier** to combine models.

Results and Analysis

HR Attrition Dataset

```
Logistic Regression Best CV AUC: 0.8154
```

```
Evaluating Manual Models...
```

```
Decision Tree:
```

```
Accuracy: 0.8277  
Precision: 0.4324  
Recall: 0.2254  
F1: 0.2963  
ROC AUC: 0.6939
```

```
kNN:
```

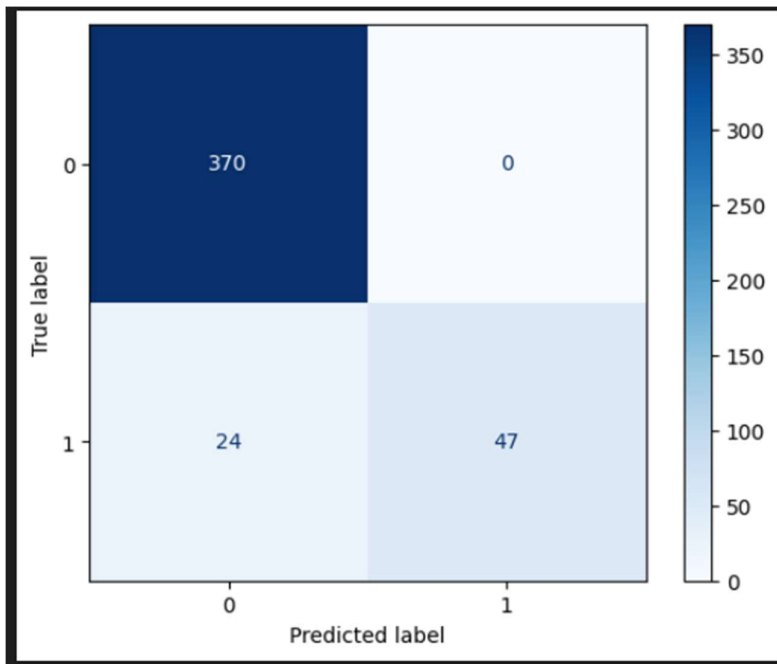
```
Accuracy: 0.8413  
Precision: 0.5714  
Recall: 0.0563  
F1: 0.1026  
ROC AUC: 0.6773
```

```
Logistic Regression:
```

```
Accuracy: 0.8662  
Precision: 0.7308  
Recall: 0.2676  
F1: 0.3918  
ROC AUC: 0.8057
```

```
...
```

```
Precision: 1.0000  
Recall: 0.6620  
F1: 0.7966  
ROC AUC: 0.9997
```



Logistic Regression Best CV AUC: 0.7532

Evaluating Built-in Models...

Decision Tree:

Accuracy: 0.8322
Precision: 0.4571
Recall: 0.2254
F1: 0.3019
ROC AUC: 0.7331

kNN:

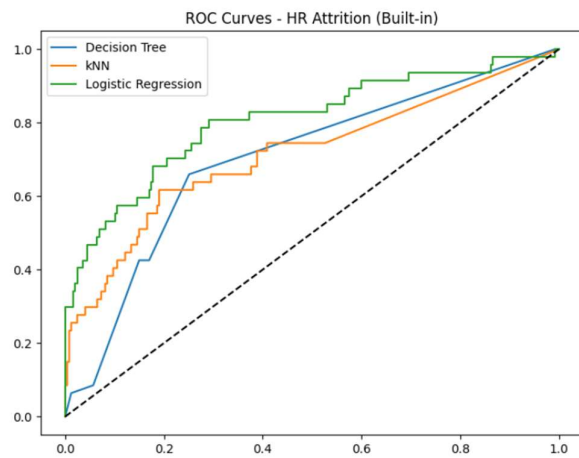
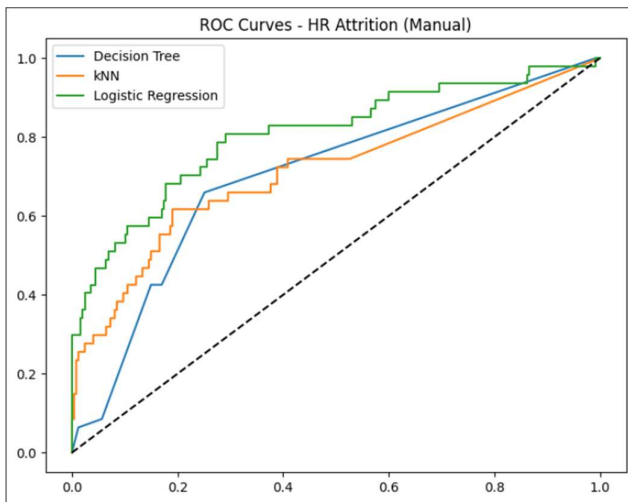
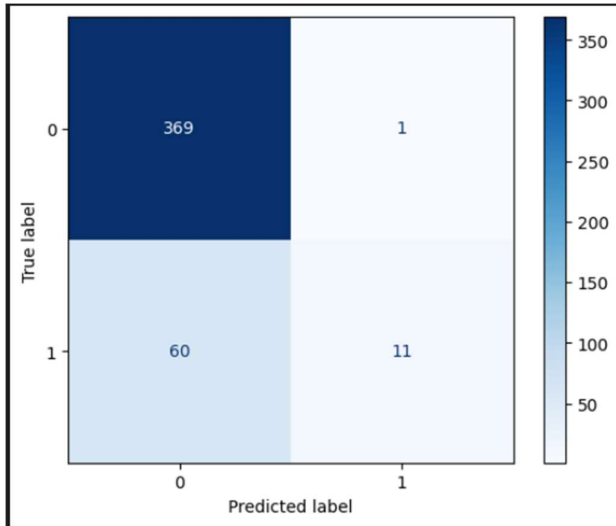
Accuracy: 0.8390
Precision: 0.5000
Recall: 0.2254
F1: 0.3107
ROC AUC: 0.7285

Logistic Regression:

Accuracy: 0.8458
Precision: 1.0000
Recall: 0.0423
F1: 0.0811
ROC AUC: 0.7545

...

Precision: 0.9167
Recall: 0.1549
F1: 0.2651
ROC AUC: 0.8828



Wine Quality Dataset

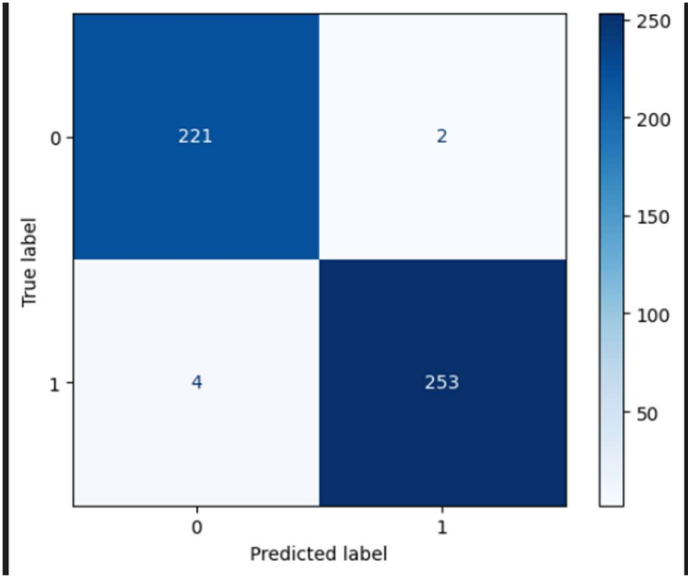
```
Running Manual Grid Search...
Decision Tree Best CV AUC: 0.7718
kNN Best CV AUC: 0.8613
Logistic Regression Best CV AUC: 0.8051

Evaluating Manual Models...

Decision Tree:
Accuracy: 0.7042
Precision: 0.7366
Recall: 0.6965
F1: 0.7160
ROC AUC: 0.7577

kNN:
Accuracy: 0.7875
Precision: 0.7839
Recall: 0.8327
F1: 0.8075
ROC AUC: 0.8609

Logistic Regression:
Accuracy: 0.7312
Precision: 0.7462
Recall: 0.7549
...
Precision: 0.9922
Recall: 0.9844
F1: 0.9883
ROC AUC: 0.9992
```



```
Running Built-in GridSearchCV...
Decision Tree Best CV AUC: 0.7690
kNN Best CV AUC: 0.8604
Logistic Regression Best CV AUC: 0.8048
```

```
Evaluating Built-in Models...
```

```
Decision Tree:
```

```
Accuracy: 0.7271
Precision: 0.7716
Recall: 0.6965
F1: 0.7321
ROC AUC: 0.7927
```

```
kNN:
```

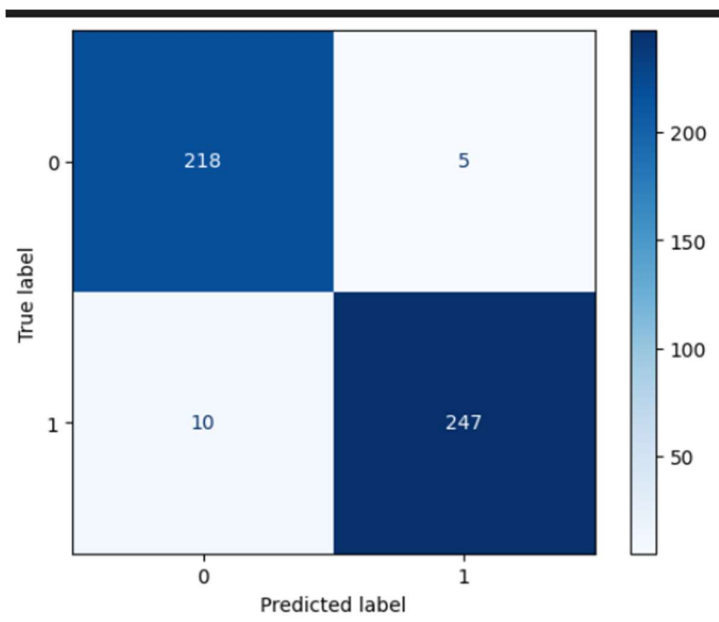
```
Accuracy: 0.7854
Precision: 0.7895
Recall: 0.8171
F1: 0.8031
ROC AUC: 0.8690
```

```
Logistic Regression:
```

```
Accuracy: 0.7417
Precision: 0.7628
Recall: 0.7510
```

```
...
```

```
Precision: 0.9802
Recall: 0.9611
F1: 0.9705
ROC AUC: 0.9971
```



- **Manual vs GridSearchCV:** Both methods yielded similar results, but GridSearchCV was easier and less error-prone.
- **Dataset imbalance:** HR Attrition dataset had high accuracy but very low recall (failing to catch attrition cases). Logistic Regression had perfect precision but almost no recall.
- **Wine Quality:** More balanced dataset, models performed consistently better.
- **Voting Classifier:** Improved performance in both datasets, especially ROC AUC.

Conclusion

- Hyperparameter tuning is essential for achieving good model performance.
- GridSearchCV is more convenient and reliable than manual search.
- Ensemble methods like Voting Classifier improve robustness.
- Dataset imbalance (HR Attrition) can distort metrics, so metrics beyond accuracy (e.g., recall, ROC AUC) must be considered.