

## Lab Overview & Objectives

The goal of this assignment is to give you hands-on experience with model selection and evaluation by implementing hyperparameter tuning and ensemble methods—two critical techniques in applied machine learning.

### Learning Goals

- Understand and implement a pipeline for scaling, feature selection, and model training.
- Grasp the fundamentals of hyperparameter tuning using Grid Search.
- Implement k-fold cross-validation to get robust performance estimates.
- Effectively use scikit-learn's Pipeline and GridSearchCV tools.
- Evaluate and compare the performance of different classification models using various metrics.
- Analyze and interpret model results to draw meaningful conclusions.

## DataSets

### Wine Quality

1. Number of features: 11
2. Target Class: 'good' quality wine or the inverse

### HR Attrition

1. Number of features: 34
2. Target Class: Employee Attrition (Either 'yes' or 'no')

## Methodology

The experiments were conducted using a machine learning pipeline consisting of three main stages:

1. **StandardScaler** – Standardizes features to mean 0 and variance 1.
2. **SelectKBest** – Selects top k features using f\_classif statistical test.
3. **Classifier** – Decision Tree, kNN, or Logistic Regression.

Two approaches used:

4. Manual Grid Search – Implemented from scratch using nested loops and 5-fold Stratified Cross Validation.

5. Built-in `GridSearchCV` – Used `scikit-learn`'s optimized method with the same pipeline

## Results & Analysis

Model performance was compared in terms of

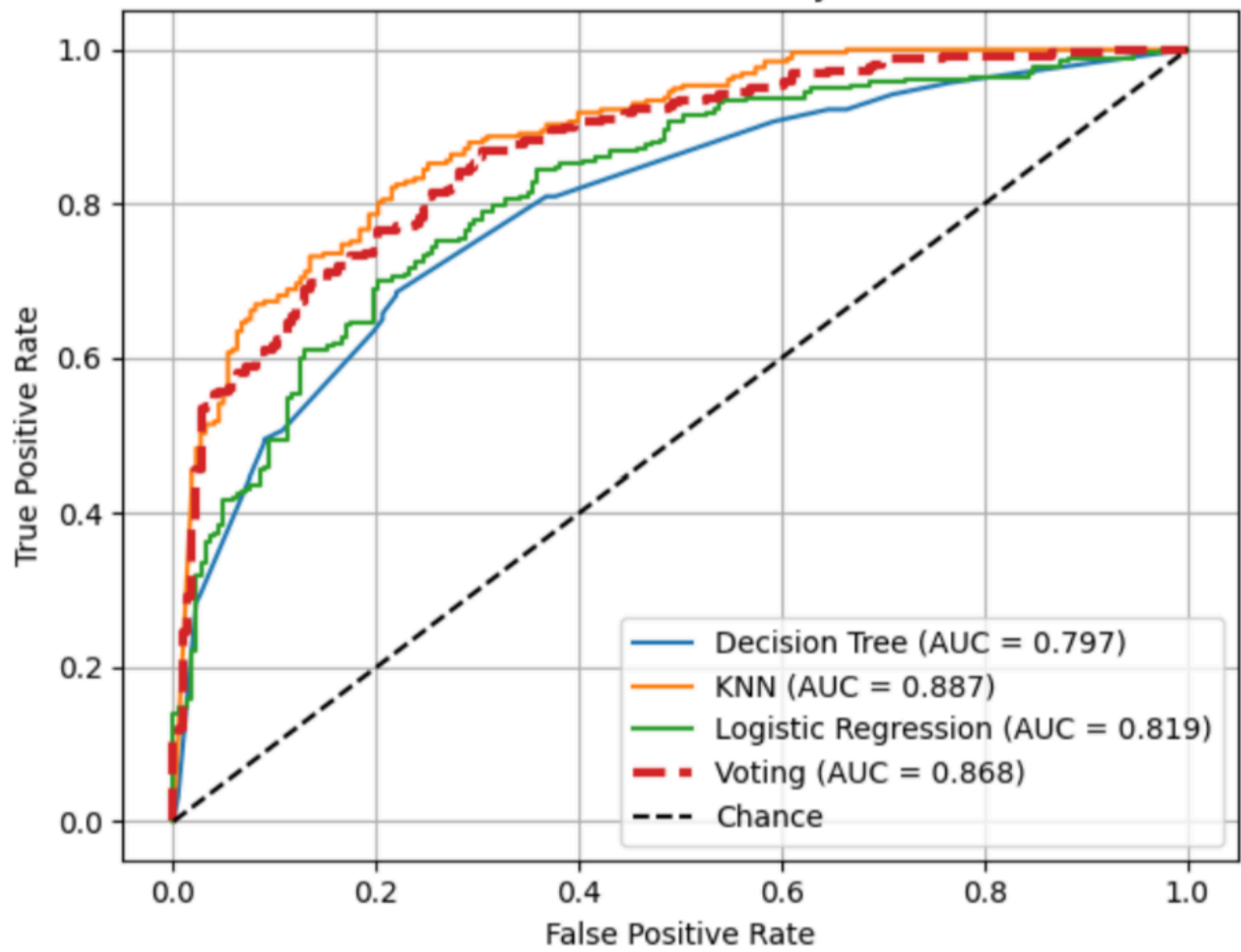
1. Accuracy
2. Precision
3. Recall
4. F1-score
5. ROC AUC.

## Wine Quality

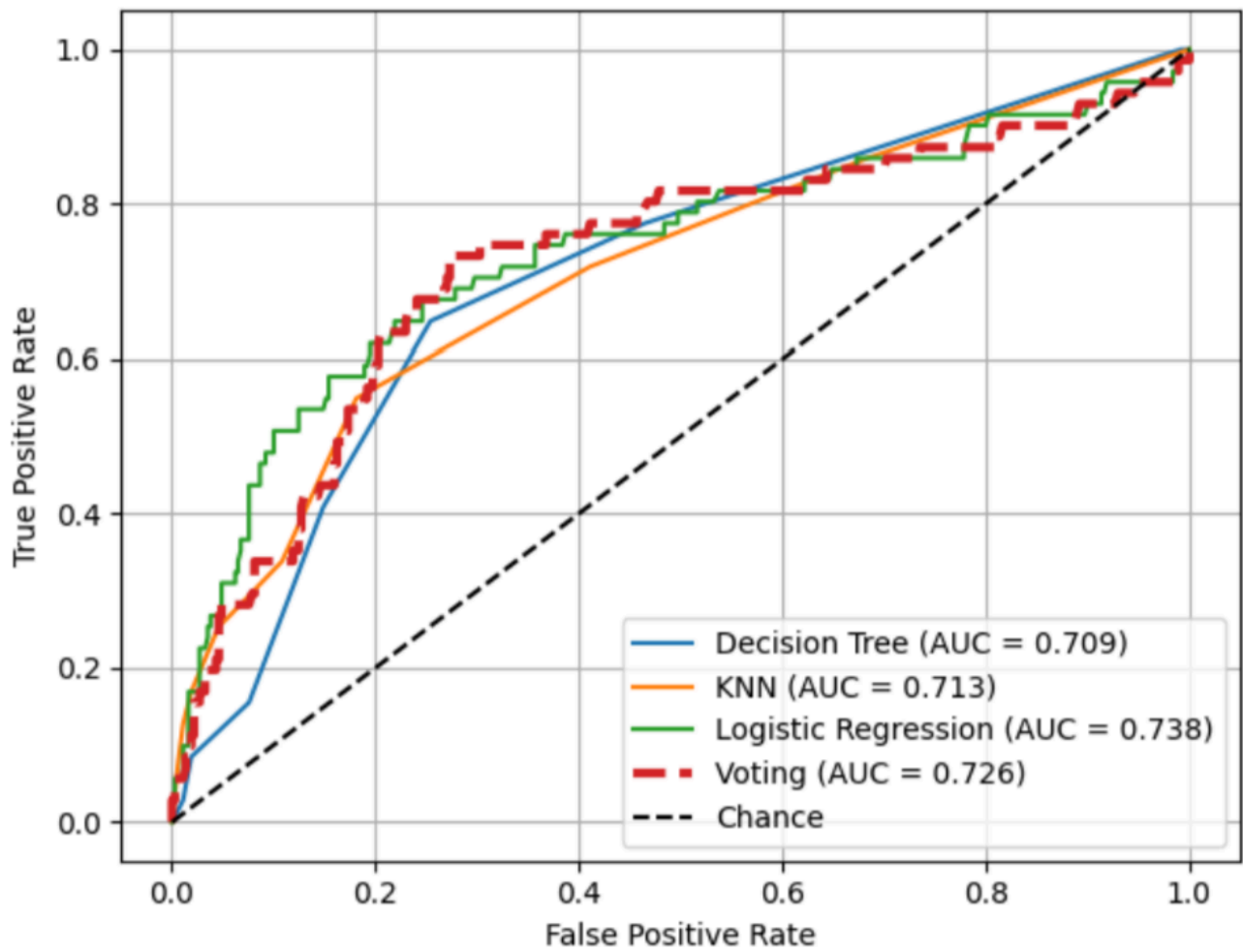
Metrics	Manual	GridSearchCV
Accuracy	0.7583	0.7771
Precision	0.7787	0.7799
Recall	0.7665	0.8132
F1	0.7725	0.7962
AUC	0.8678	0.8678

## ROC Curve

ROC Curves - Wine Quality (Built-in)

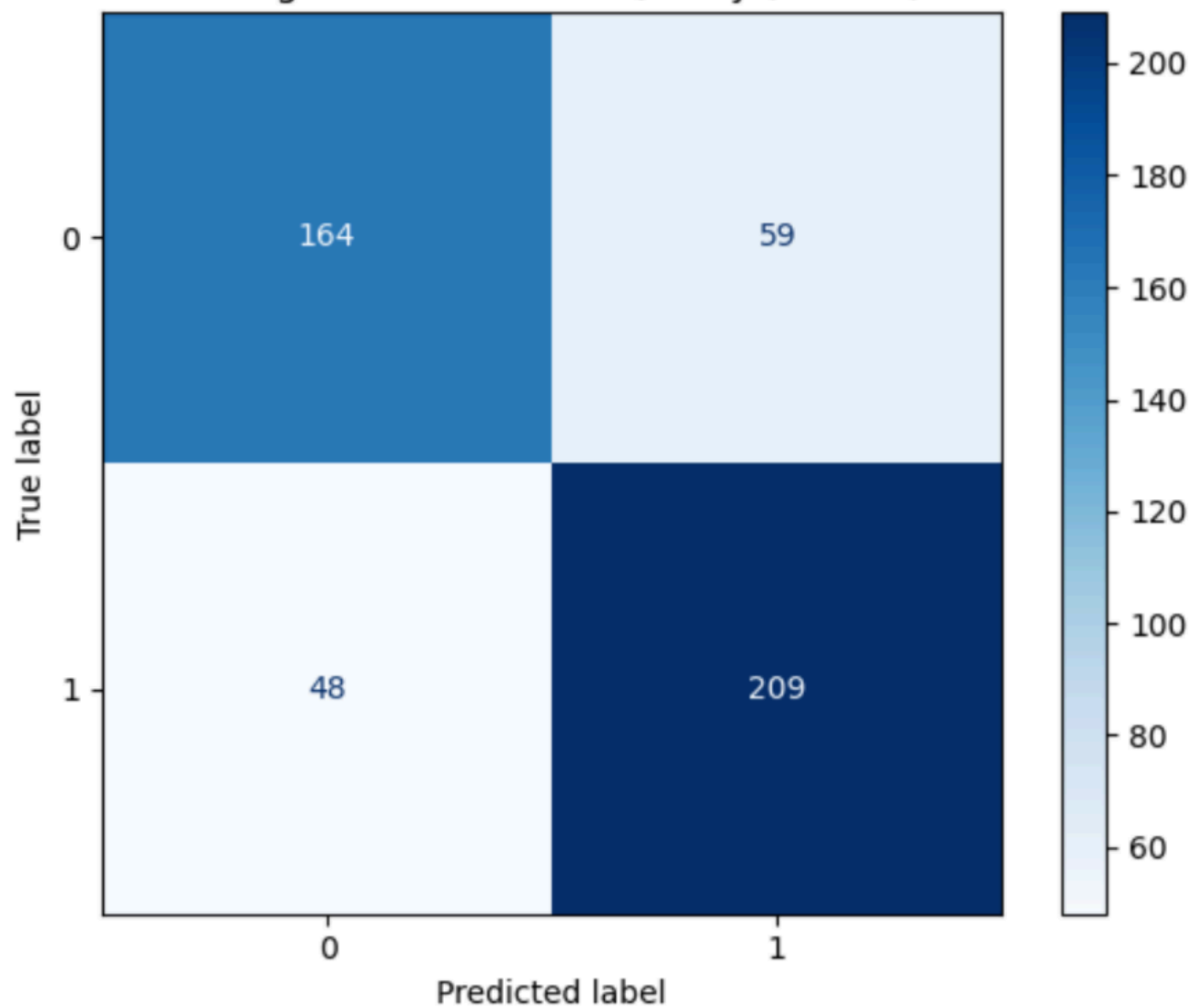


ROC Curves - HR Attrition (Manual)

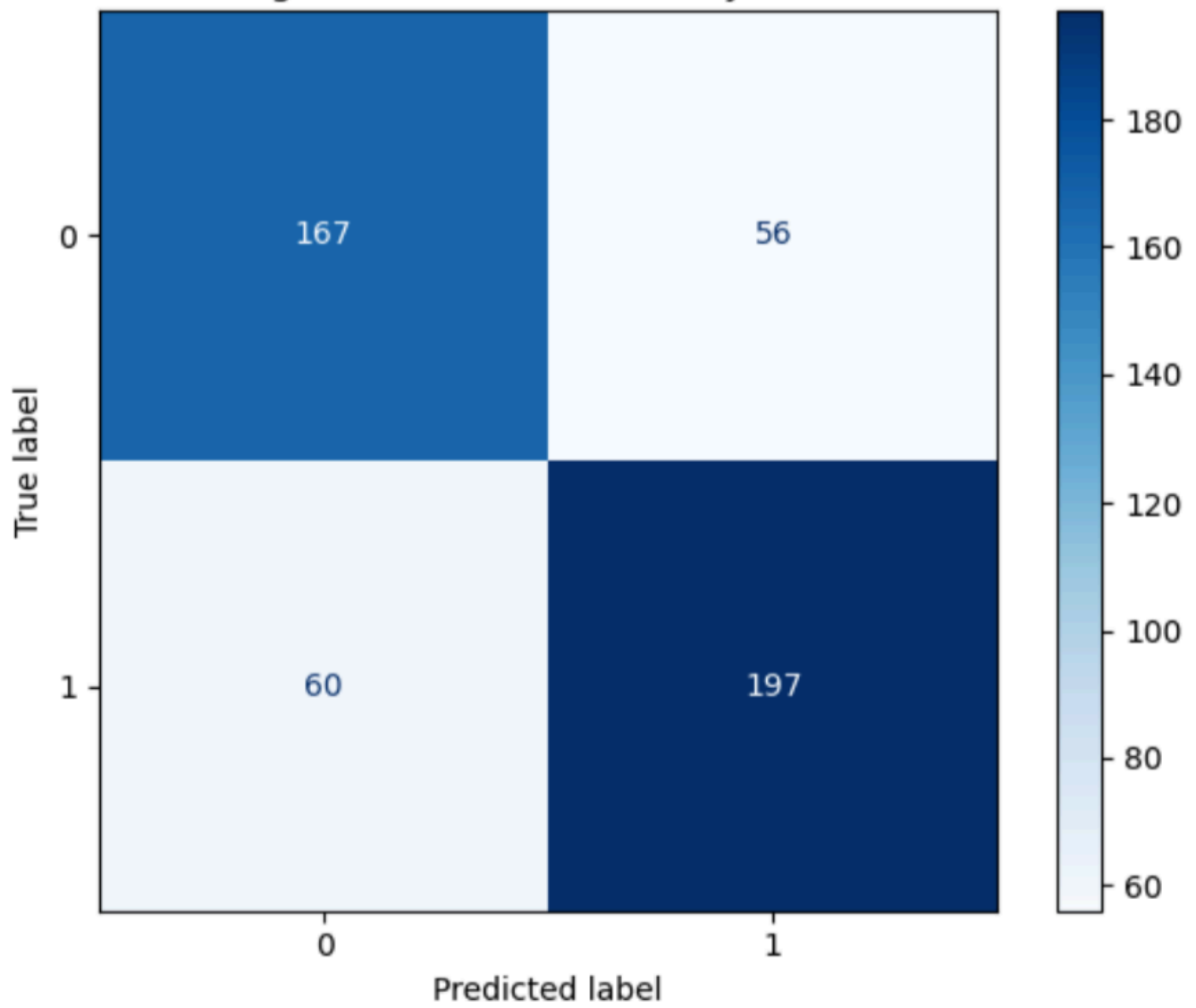


## Confusion Matrix

Voting Classifier - Wine Quality (Built-in)



Voting Classifier - Wine Quality (Manual)

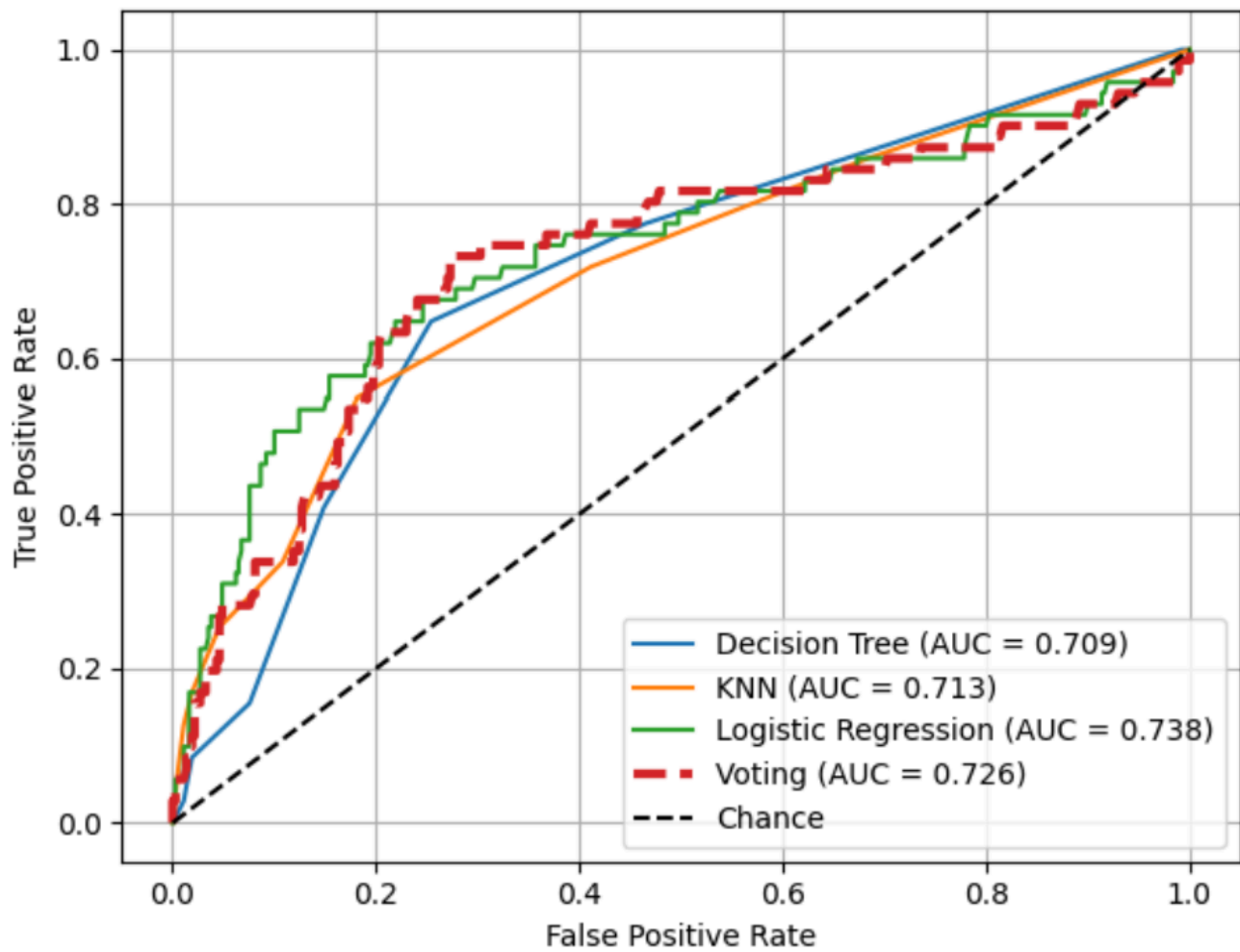


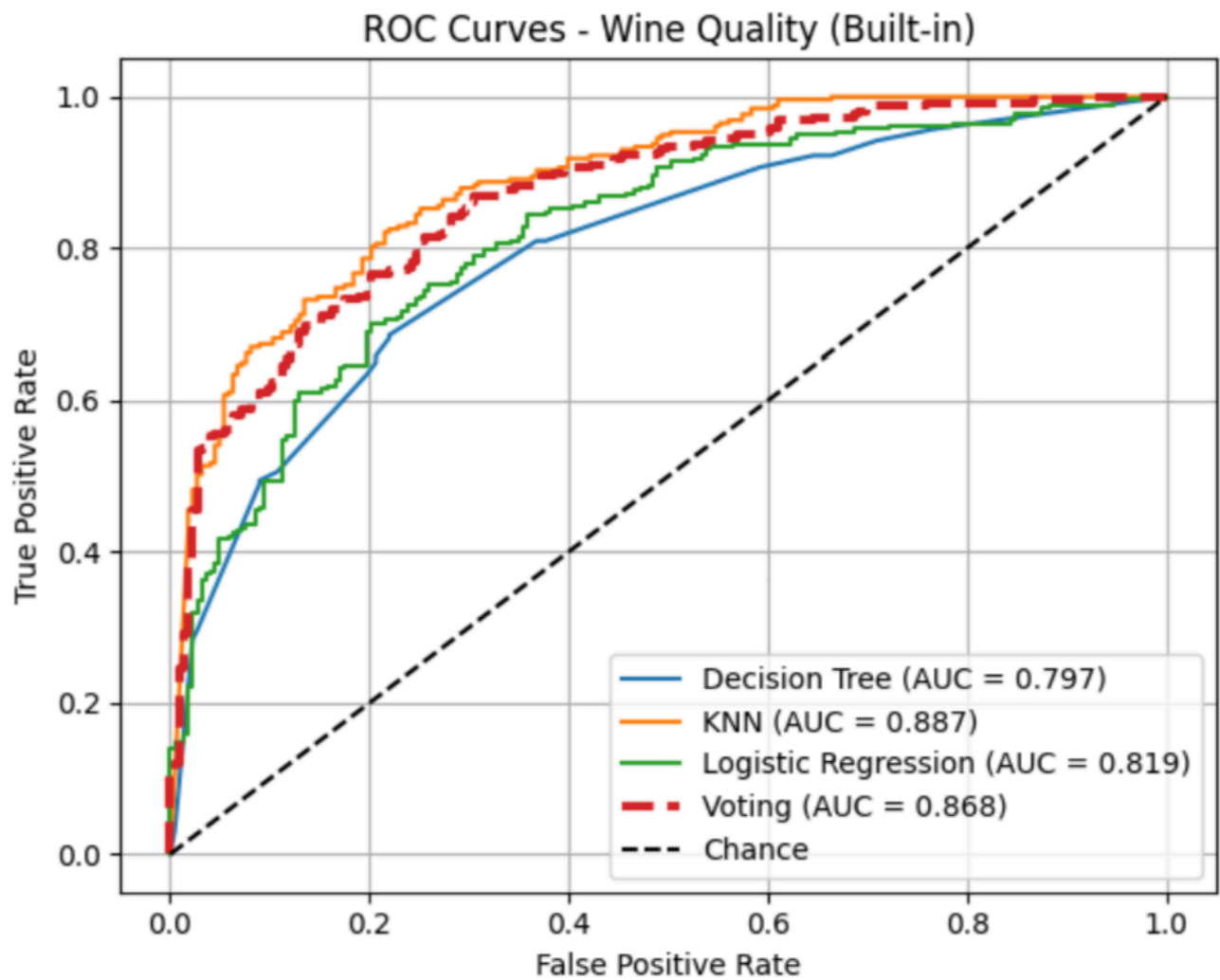
## HR Attrition

Metrics	Manual	GridSearchCV
Accuracy	0.8413	0.8413
Precision	0.5556	0.5294
Recall	0.0704	0.1268
F1	0.1250	0.2045
AUC	0.7256	0.7256

## ROC Curve

ROC Curves - HR Attrition (Manual)

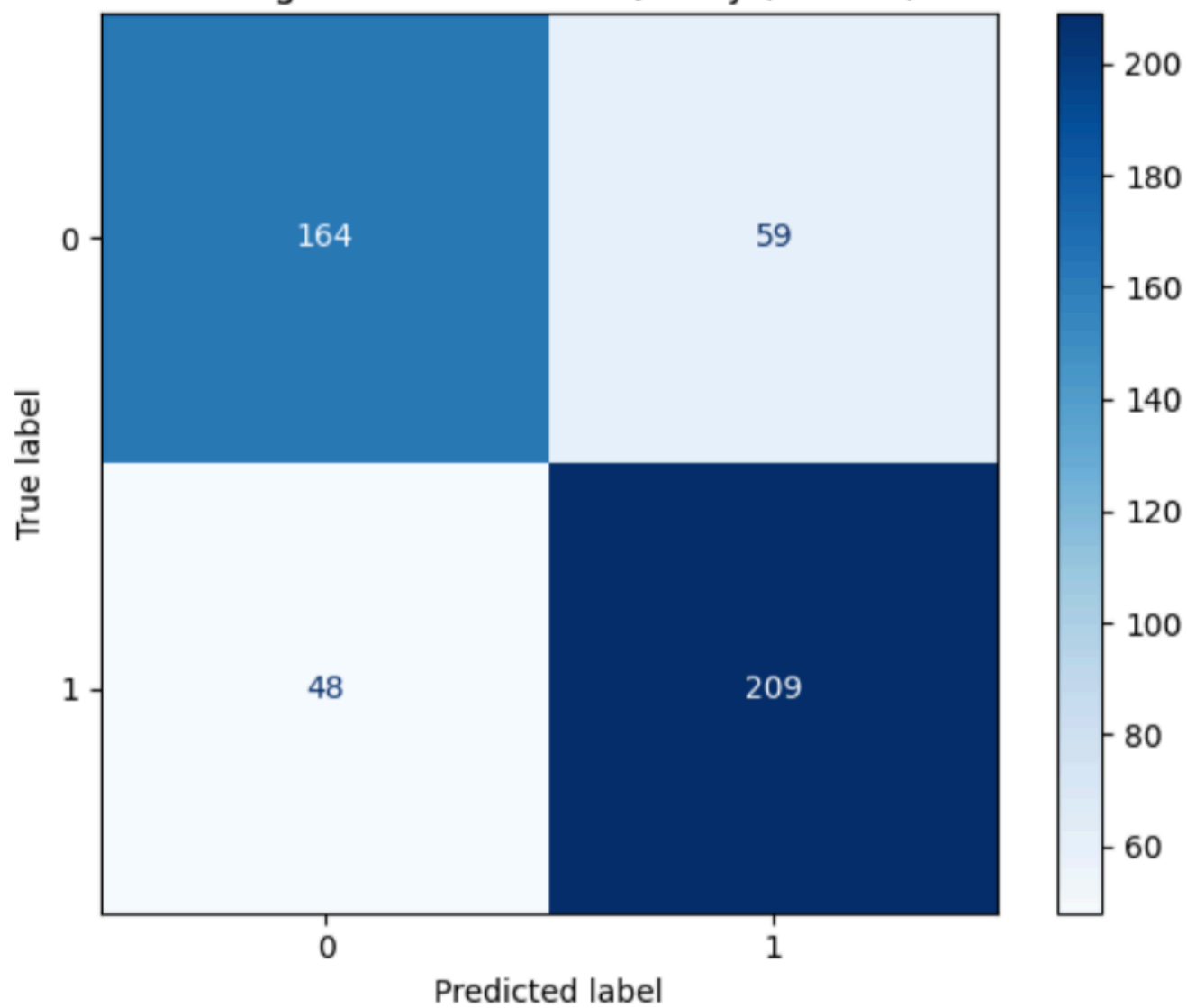




## Confusion Matrix



Voting Classifier - Wine Quality (Built-in)



Voting Classifier - HR Attrition (Manual)

