Week 4: Model Selection and Comparative Analysis

Name: C Vishwa SRN: PES2UG23CS139 Course Code: UE23CS352A Date: 01-09-2025

1. Introduction

The purpose of this lab is to understand the working behind grid search which is used for hyper parameter tuning. This is to help with understanding model selection and comparing the respective models based on selected hyperparameters.

We are using the HR dataset and creating 2 Jupyter files – one contains a manually implemented grid search function and the other uses grid search CV.

2. Dataset Description

Name: HR–Employee–Attrition

Number of features: 35

Number of instances: 1470

3. Methodology

3.1. Conceptual Understanding

Hyperparameter Tuning: Is the process of cross-checking the various performance metrics for different sets of parameter values, such as learning rate, batch size, etc.

Grid Search: Is a hyperparameter tuning method that checks for the most optimal set of hyperparameters by checking every single combination in a specific order.

K-Fold Cross-Validation: For each combination of parameters chosen using grid search we divide the dataset into K equally sized folds and 1 of the folds is validated upon training the rest K-1 folds.

4. Results and Analysis

Performance

```
============================================================
EVALUATING MANUAL MODELS FOR HR ATTRITION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.8231
  Precision: 0.3333
  Recall: 0.0986
  F1-Score: 0.1522
  ROC AUC: 0.7107

kNN:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236

Logistic Regression:
...
  F1-Score: 0.3762
  ROC AUC: 0.7759
```

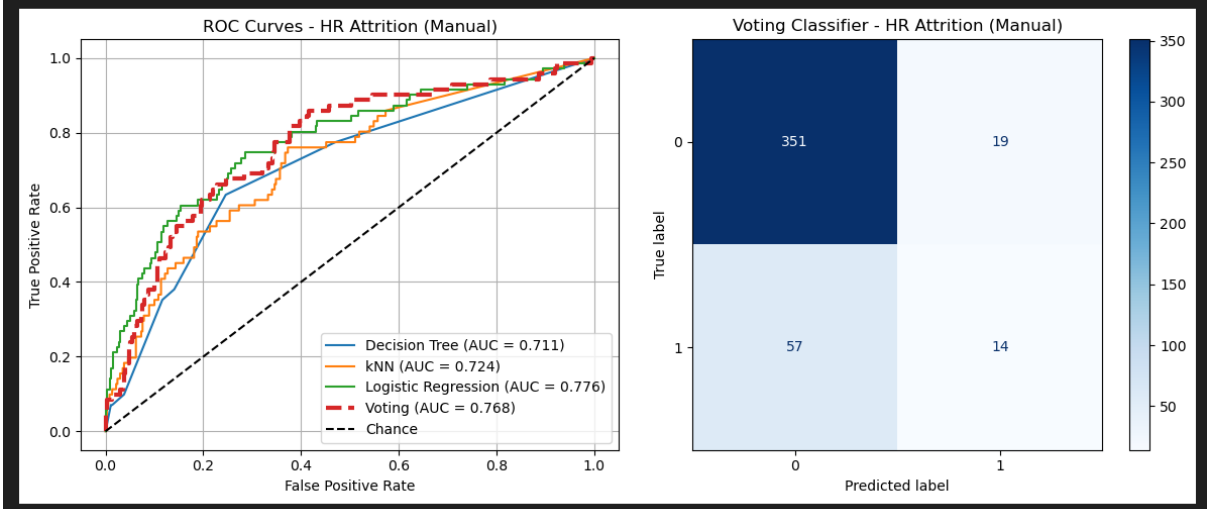
Manual

```
################################################################################
PROCESSING DATASET: HR ATTRITION
################################################################################
IBM HR Attrition dataset loaded and preprocessed successfully.
Training set shape: (1029, 46)
Testing set shape: (441, 46)
----------------------------


============================================================
RUNNING MANUAL GRID SEARCH FOR HR ATTRITION
============================================================
--- Manual Grid Search for Decision Tree ---
Testing 36 combinations...
```

```
Best parameters for Decision Tree: {'feature_selection__k': 5, 'classifier__max_depth': 3, 'classifier__min_samples_split': 2}
Best cross-validation AUC: 0.7152
--- Manual Grid Search for kNN ---
Testing 24 combinations...
New best score: 0.6578 with params: {'feature_selection__k': 5, 'classifier__n_neighbors': 3, 'classifier__weights': 'uniform'}
```

```
Voting Classifier Performance:
  Accuracy: 0.8277, Precision: 0.4242
  Recall: 0.1972, F1: 0.2692, AUC: 0.7679
```




```
============================================================
EVALUATING BUILT-IN MODELS FOR HR ATTRITION
============================================================

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.8231
  Precision: 0.3333
  Recall: 0.0986
  F1-Score: 0.1522
  ROC AUC: 0.7107

kNN:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236

Logistic Regression:
  Accuracy: 0.8571
...
  F1-Score: 0.3762
  ROC AUC: 0.7759

--- Built-in Voting Classifier ---
```