Week 4: Model Selection & Comparative Analysis
Tasks :

1. Evaluate & Compare efficiency of own built grid search vs Scikit builtin gridsearchCV
2. Dataset : IBM HR Attrition
3. Classification Algorithms -> Decision Tree , K Nearest Neighbor & Logistic Regression
4. Data Learn Pipeline :

StandardScaler: standardizes every feature to have a mean 0 and a standard deviation 1 (models like knn and logistic regression) are sensitive to scale

SelectKbest: Selects the best features from the dataset , it uses statistical tests (f-test) to score each features relationship with target
then it keeps only the top k features
Where k is a hyper parameter where we experiment with different values .

Classifier: actual machine learning model at the end of the pipeline .

What to submit :

1. **Manual Grid Search** (run_manual_grid_search)
   - Specify the hyperparameter grids for each model.
   - Write a search loop that tests every parameter combination using 5-fold stratified cross-validation and computes the average ROC AUC.
   - Re-train the pipeline with the best parameter set on the full training data.
2. **Built in Grid Search** (run_builtin_grid_search)
   - Use scikit-learn's gridsearchCV to perform automated hyperparameter tuning with the same pipeline and evaluation settings.
   - Extract and report the best estimator, chosen parameters, and corresponding cross-validation performance.

Introduction

This report covers the work completed for Week 4 of the Machine Learning lab. The focus was on creating a complete pipeline for model selection and evaluation. The main goals were to practice hyperparameter tuning and model comparison using two methods: a manually written grid search and the built-in GridSearchCV function from scikit-learn. The lab was designed to provide practical experience with these important techniques and to highlight the differences in performance and ease of use between a custom approach and a library-based approach.

Dataset Description

In this lab, experiments were carried out on the HR Attrition dataset. The dataset includes both personal details and job-related attributes, and the task is to predict whether an employee stays or leaves the company. After preprocessing and splitting, the training set contained 1029 records with 46 attributes, and the test set contained 441 records with the same number of features.

Methodology

The lab involved a series of steps to build, tune, and evaluate machine learning models.

- Hyperparameter tuning was used to identify the best settings for each classifier. Since these parameters are not learned directly by the model, they had to be chosen through experimentation.
- Grid search was the chosen tuning method, where different combinations of parameter values were systematically tested to see which produced the strongest results.
- To ensure fair evaluation, 5-fold stratified cross-validation was applied. This technique divides the dataset into five parts, trains the model on four parts, and validates it on the remaining part, repeating the process so that every fold is used for testing once.

The machine learning pipeline had three stages: s
caling the input features with StandardScaler,
selecting the most relevant features using SelectKBest,
and finally applying a classifier (Decision Tree, k-Nearest Neighbors, or Logistic Regression).
Two approaches were followed. In the first part, a manual grid search was coded using loops to explore parameter combinations and measure performance. In the second part, the same task was repeated using scikit-learn's GridSearchCV, which automates the search and evaluation process.

ML Pipeline

The machine learning pipeline used in this lab consisted of three main stages:

1. **StandardScaler** – All numerical features were standardized so that they have a mean of 0 and a standard deviation of 1. This prevents features with larger scales from dominating those with smaller scales.
2. **SelectKBest** – A feature selection step that chooses the top *k* features based on statistical tests (ANOVA F-test in this case). The number of features *k* was treated as a hyperparameter and tuned during the grid search.
3. **Classifier** – The final step of the pipeline where the actual model is trained. Three classifiers were evaluated in this lab: Decision Tree, k-Nearest Neighbors (kNN), and Logistic Regression.

Implementation Process

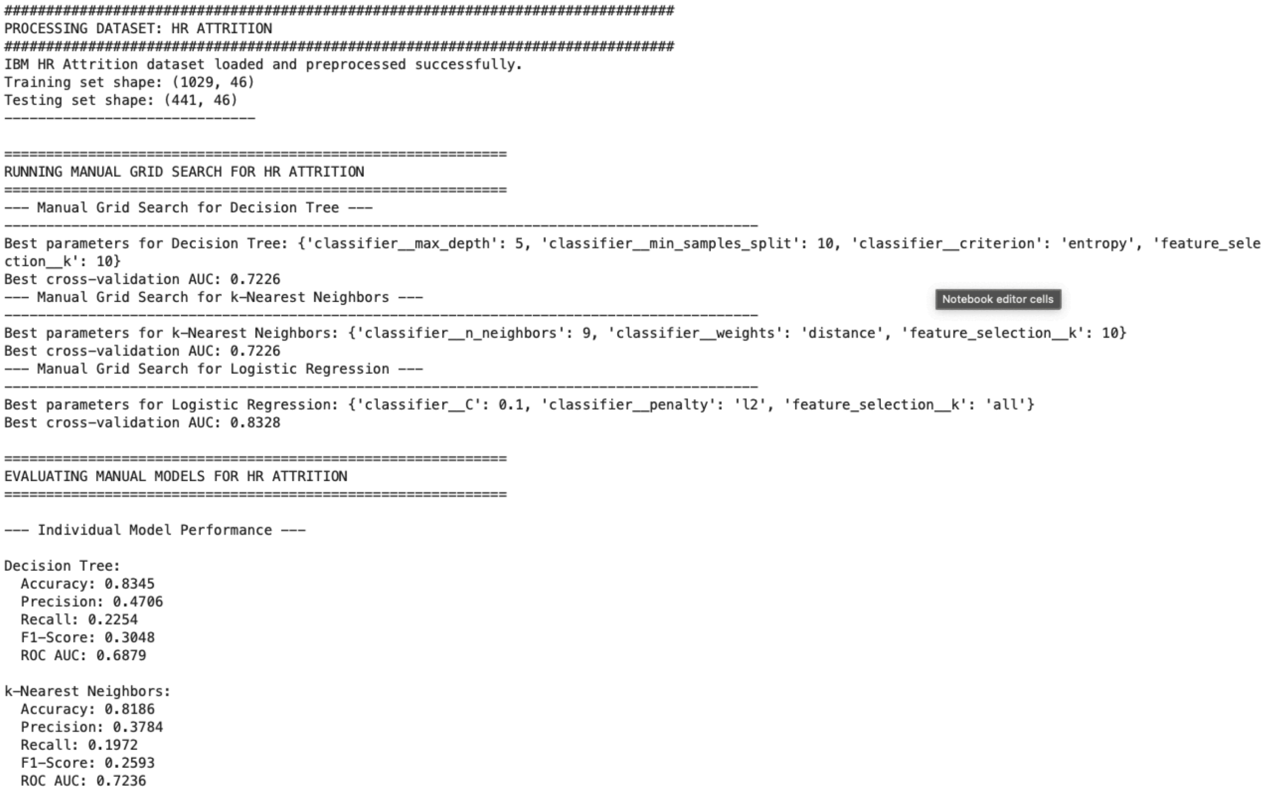
The experiment followed two approaches to perform hyperparameter tuning and model evaluation.

- **Manual Grid Search**
  - Parameter grids were defined for Decision Tree, k-Nearest Neighbors, and Logistic Regression.
  - A nested loop was used to generate all parameter combinations. Each combination was evaluated using **5-fold stratified cross-validation**, with the mean ROC AUC recorded as the score.
  - The parameter set achieving the highest average AUC was chosen as the best configuration.
  - A final model was retrained on the training set using these optimal parameters.
- **Built-in Grid Search with GridSearchCV**
  - The same parameter grids were passed to scikit-learn's GridSearchCV, which automatically handled cross-validation and scoring.
  - The pipeline was defined as:
    StandardScaler → SelectKBest → Classifier.
  - Best parameters, cross-validation AUC, and final models were obtained directly from the fitted GridSearchCV objects.
- **Evaluation**
  - Each tuned model (from both manual and built-in approaches) was tested on the HR Attrition dataset's hold-out test set.
  - Performance was measured with **Accuracy, Precision, Recall, F1-Score, and ROC AUC**.
  - Additionally, a **Voting Classifier** (soft voting) was built to combine predictions from all three classifiers.

Model	Best Parametrs	Avg CV Score (ROC AUC)
Decision Tree	{'criterion': 'entropy', 'max_depth': 5, 'min_samples_split': 10, 'k': 10}	0.7226
k-Nearest Neighbors	{'n_neighbors': 9, 'weights': 'distance', 'k': 10}	0.7226
Logistic Regression	{'C': 0.1, 'penalty': 'l2', 'k': 'all'}	0.8328

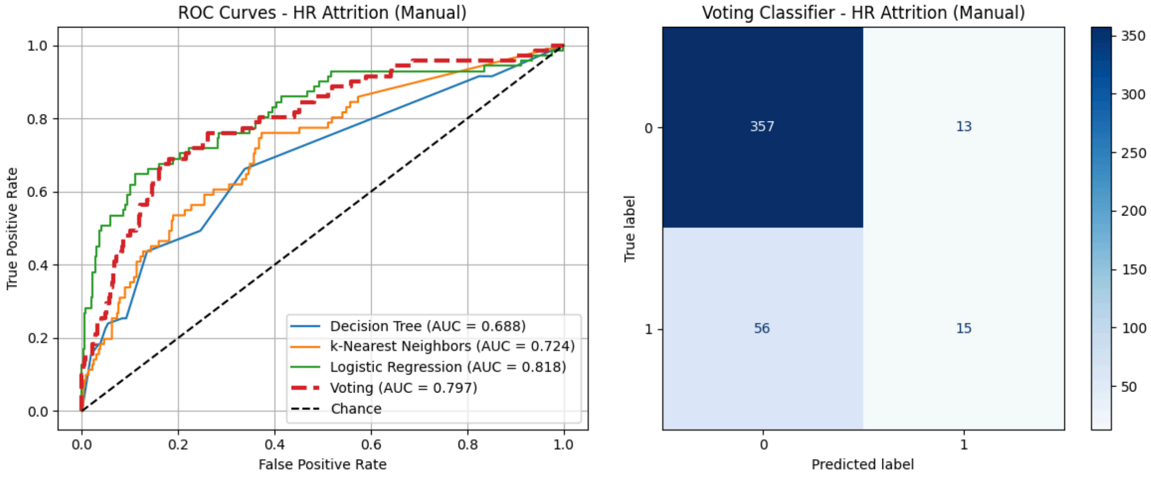
Model (manual)	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.8345	0.4706	0.2254	0.3048	0.6879
k-Nearest Neighbors	0.8186	0.3784	0.1972	0.2593	0.7236
Logistic Regression	0.8798	0.7368	0.3944	0.5138	0.8177
Voting Classifier	0.8435	0.5357	0.2113	0.3030	0.7971

Model (gridsearch)	Accuracy	Precision	Recall	F1-Score	ROC AUC
Decision Tree	0.8345	0.4706	0.2254	0.3048	0.6879
k-Nearest Neighbors	0.8186	0.3784	0.1972	0.2593	0.7236
Logistic Regression	0.8798	0.7368	0.3944	0.5138	0.8177
Voting Classifier	0.8481	0.5769	0.2113	0.3093	0.7971



```
################################################################################
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 ---
------------------------------------------------------------------------------------------
Best parameters for Decision Tree: {'classifier__max_depth': 5, 'classifier__min_samples_split': 10, 'classifier__criterion': 'entropy', 'feature_sele
ction__k': 10}
Best cross-validation AUC: 0.7226
--- Manual Grid Search for k-Nearest Neighbors ---
------------------------------------------------------------------------------------------
Best parameters for k-Nearest Neighbors: {'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'feature_selection__k': 10}
Best cross-validation AUC: 0.7226
--- Manual Grid Search for Logistic Regression ---
------------------------------------------------------------------------------------------
Best parameters for Logistic Regression: {'classifier__C': 0.1, 'classifier__penalty': 'l2', 'feature_selection__k': 'all'}
Best cross-validation AUC: 0.8328

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

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.8345
  Precision: 0.4706
  Recall: 0.2254
  F1-Score: 0.3048
  ROC AUC: 0.6879

k-Nearest Neighbors:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236
```

```
Logistic Regression:
  Accuracy: 0.8798
  Precision: 0.7368
  Recall: 0.3944
  F1-Score: 0.5138
  ROC AUC: 0.8177

--- Manual Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8435, Precision: 0.5357
  Recall: 0.2113, F1: 0.3030, AUC: 0.7971
```

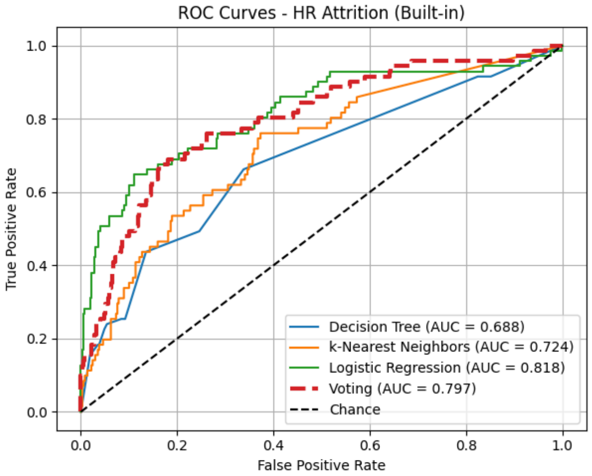


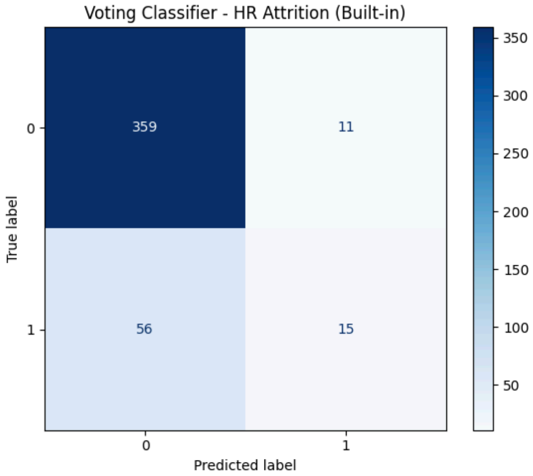
```
============================================================
RUNNING BUILT-IN GRID SEARCH FOR HR ATTRITION
============================================================

--- 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_selectio
n__k': 10}
Best CV score: 0.7226

--- GridSearchCV for k-Nearest Neighbors ---
Fitting 5 folds for each of 24 candidates, totalling 120 fits
```

```
Best params for k-Nearest Neighbors: {'classifier__n_neighbors': 9, 'classifier__weights': 'distance', 'feature_selection__k': 10}
Best CV score: 0.7226

--- GridSearchCV for Logistic Regression ---
Fitting 5 folds for each of 9 candidates, totalling 45 fits
```

```
Best params for Logistic Regression: {'classifier__C': 0.1, 'classifier__penalty': 'l2', 'feature_selection__k': 'all'}
Best CV score: 0.8328

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

--- Individual Model Performance ---

Decision Tree:
  Accuracy: 0.8345
  Precision: 0.4706
  Recall: 0.2254
  F1-Score: 0.3048
  ROC AUC: 0.6879

k-Nearest Neighbors:
  Accuracy: 0.8186
  Precision: 0.3784
  Recall: 0.1972
  F1-Score: 0.2593
  ROC AUC: 0.7236

Logistic Regression:
  Accuracy: 0.8798
  Precision: 0.7368
  Recall: 0.3944
  F1-Score: 0.5138
  ROC AUC: 0.8177

--- Built-in Voting Classifier ---
Voting Classifier Performance:
  Accuracy: 0.8481, Precision: 0.5769
  Recall: 0.2113, F1: 0.3093, AUC: 0.7971
```





```
Completed processing for HR Attrition
================================================================================

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