```
In [ ]:  # #################################
         # Group ID: 222
         # Members: Adrián Gómez Repollés, Kresten Johansen, Misbah Unnisa, Mohammadali Zakeriharandi
         # Date: 22/03/2024
         # Lecture: Lecture 11 Decision Trees
         # Dependencies: Numpy, Scipy, Scikit-Learn, Pytorch and Matplotlib.
         # Python version: 3.12.2
         # Functionality: Jupyter notebook to examine the California housing data using decision trees.
         # #################################
```

Day 5 - Exercise

In this exercise, we examine the use of decision trees for both classification and regression. First import the California housing data set and split the data into a training set and a test set.

For classification:

Artificially create two classes based on the price. Use mean price as the threshold: class 1 if the price is larger than mean; otherwise, class 0. Split the data into training (80%) and test (20%). Afterwards, train a decision tree classifier on the training data and see how it performs on both the training data and the test data. Try to experiment with the depth of the tree and see what effect this has on the classification accuracy. Note all 13 dimensional features, as originally provided by the dataset, are used as input, and the labels/targets are 0 (low price) and 1 (high price).

```
In [ ]:  ## Imports
         import numpy as np
         import pandas as pd
         import seaborn as sns
         import matplotlib.pyplot as plt
         from sklearn.model_selection import GridSearchCV
         from sklearn.model_selection import train_test_split
         from sklearn.datasets import fetch_california_housing
         from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
         from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, accuracy_score, mean_squared_error
```

```
# from sklearn.ensemble import GradientBoostingClassifier
# from sklearn.ensemble import GradientBoostingRegressor
```

In [ ]:
```python
## Helper functions

# Load data
def load_data(mode):
    # Dataset
    X = fetch_california_housing(as_frame=True)['frame']
    if mode == "Classification":
        # Artificial class
        mean_price = X['MedHouseVal'].mean()
        y = [1 if i > mean_price else 0 for i in X['MedHouseVal']]
        # Remove original variable that produce the category
        X = X.drop(['MedHouseVal'], axis=1)
        # Split train-test
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y, shuffle=True)
    elif mode == "Regression":
        y = X['MedHouseVal']
        X = X.drop(['MedHouseVal'], axis=1)
        # Split train-test
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=True)
    # Return
    return(X_train, y_train, X_test, y_test)

# Plot train test error of grid search
def plot_train_test_error(gs):
    # Values
    iters = np.arange(1,len(gs.cv_results_['params'])+1,1)
    train_scores = gs.cv_results_['mean_train_score']
    test_scores = gs.cv_results_['mean_test_score']
    # Plot
    plt.figure(figsize=(6,6))
    plt.plot(iters, train_scores, c="orange", label="Train")
    plt.plot(iters, test_scores, c="blue", label="Test")
    plt.xlabel("Iterations")
    plt.ylabel("Score")
    plt.legend()
    plt.show()

# Plot confusion matrices
def plot_confusion_matrix(orig_train, pred_train, orig_test, pred_test, labels, normalize = None):
```

```python
    # Measure accuracies
    acc_train = accuracy_score(orig_train, pred_train)
    acc_test = accuracy_score(orig_test, pred_test)
    # Create confusion matrices
    cm_train = np.round(confusion_matrix(orig_train, pred_train, labels=labels, normalize=normalize),2)
    vcm_train = ConfusionMatrixDisplay(confusion_matrix=cm_train, display_labels=labels)
    cm_test = np.round(confusion_matrix(orig_test, pred_test, labels=labels, normalize=normalize),2)
    vcm_test = ConfusionMatrixDisplay(confusion_matrix=cm_test, display_labels=labels)
    # Plot confusion matrix
    fig, ax = plt.subplots(1,2, figsize=(8,8))
    plt.rcParams.update({'font.size': 12})
    # Confusion matrices
    vcm_train.plot(ax=ax[0], cmap="Blues", colorbar=None) # Train
    vcm_test.plot(ax=ax[1], cmap="Greens", colorbar=None) # Test
    ax[0].set_title(f"Train (Acc: {np.round(acc_train,2)})") # Train
    ax[1].set_title(f"Test (Acc: {np.round(acc_test,2)})") # Test
    plt.show()
```

In [ ]:
```python
# Data
X_train, y_train, X_test, y_test = load_data(mode="Classification")
```

In [ ]:
```python
## Grid Search: Random forest

# Model
rfcls = RandomForestClassifier()

# Parameters
params = {
    'n_estimators': [20,50,100,500],
    'max_depth': [10, 30, None],
    'max_samples': [35, None],
    'ccp_alpha': [0.0, 0.5],
}

# Search
rfcls_gs = GridSearchCV(rfcls, params, n_jobs=4, cv=5, verbose=3, scoring='f1', return_train_score=True).f:
```

Fitting 5 folds for each of 48 candidates, totalling 240 fits

```
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.707, test=0.713)
total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.686, test=0.685)
total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.728, test=0.708)
total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.691, test=0.676)
total time=   0.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.731, test=0.708)
total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.749, test=0.740)
total time=   0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.763, test=0.753)
total time=   0.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.725, test=0.729)
total time=   0.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.747, test=0.746)
total time=   0.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.736, test=0.726)
total time=   0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.740, test=0.73
9) total time=   0.4s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.742, test=0.73
3) total time=   0.3s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.749, test=0.75
5) total time=   0.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.761, test=0.77
2) total time=   0.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.745, test=0.71
8) total time=   0.4s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.763, test=0.75
4) total time=   1.6s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.749, test=0.75
7) total time=   1.6s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.759, test=0.75
9) total time=   1.9s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.900, test=0.83
3) total time=   0.5s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.753, test=0.76
2) total time=   1.9s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.899, test=0.83
3) total time=   0.7s
```

```
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.904, test=0.83
6) total time=   0.6s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.901, test=0.84
4) total time=   0.7s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.908, test=0.84
6) total time=   0.6s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.752, test=0.73
6) total time=   1.6s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.903, test=0.83
7) total time=   1.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.906, test=0.84
3) total time=   1.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.905, test=0.84
2) total time=   1.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.901, test=0.84
2) total time=   1.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.902, test=0.83
7) total time=   1.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.903, test=0.8
41) total time=   3.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.902, test=0.8
37) total time=   3.6s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.907, test=0.8
46) total time=   3.7s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.905, test=0.8
47) total time=   3.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.906, test=0.8
33) total time=   3.5s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.906, test=0.8
41) total time=  17.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.904, test=0.8
41) total time=  18.7s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.906, test=0.8
45) total time=  18.5s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.654, test=0.669)
total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.671, test=0.657)
total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.706, test=0.700)
total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.729, test=0.735)
total time=   0.1s
```

```
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.710, test=0.701)
total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.709, test=0.699)
total time=   0.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.733, test=0.723)
total time=   0.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.759, test=0.754)
total time=   0.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.747, test=0.753)
total time=   0.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.774, test=0.749)
total time=   0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.720, test=0.72
6) total time=   0.4s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.753, test=0.73
9) total time=   0.5s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.906, test=0.8
42) total time=  17.3s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.748, test=0.74
7) total time=   0.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.755, test=0.76
5) total time=   0.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.752, test=0.73
9) total time=   0.4s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.769, test=0.75
9) total time=   2.0s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.759, test=0.75
7) total time=   2.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.751, test=0.75
9) total time=   2.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.85
6) total time=   1.0s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.759, test=0.76
7) total time=   2.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.759, test=0.74
0) total time=   2.0s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.85
1) total time=   1.0s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.85
0) total time=   0.9s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.84
3) total time=   1.0s
```

```
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.997, test=0.84
6) total time=   1.0s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.85
7) total time=   2.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.85
6) total time=   2.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.85
8) total time=   2.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.85
8) total time=   2.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.84
9) total time=   2.4s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=1.000, test=0.8
63) total time=   5.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=1.000, test=0.8
57) total time=   4.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=1.000, test=0.8
60) total time=   4.3s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.904, test=0.8
36) total time=  18.5s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=1.000, test=0.8
64) total time=   5.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=1.000, test=0.8
50) total time=   4.4s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=1.000, test=0.8
63) total time=  23.6s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=1.000, test=0.8
60) total time=  24.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.712, test=0.71
1) total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.731, test=0.71
9) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.632, test=0.62
1) total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.720, test=0.72
9) total time=   0.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.711, test=0.68
5) total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.742, test=0.74
9) total time=   0.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.742, test=0.73
7) total time=   0.4s
```

```
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.763, test=0.76
4) total time=   0.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.703, test=0.71
4) total time=   0.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.739, test=0.72
6) total time=   0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.752, test=0.7
50) total time=   0.4s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.755, test=0.7
50) total time=   0.5s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=1.000, test=0.8
63) total time=   24.5s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.754, test=0.7
58) total time=   0.5s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.767, test=0.7
70) total time=   0.5s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.756, test=0.7
41) total time=   0.5s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=1.000, test=0.8
59) total time=   25.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.758, test=0.7
52) total time=   2.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.752, test=0.7
46) total time=   2.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.751, test=0.7
54) total time=   2.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.
850) total time=   0.9s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.766, test=0.7
73) total time=   2.4s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.
852) total time=   1.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.772, test=0.7
55) total time=   2.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.997, test=0.
854) total time=   1.0s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.998, test=0.
854) total time=   1.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.997, test=0.
843) total time=   1.0s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.
854) total time=   2.6s
```

```
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.
851) total time=   2.6s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.
857) total time=   2.6s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.
859) total time=   2.5s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=1.000, test=0.
849) total time=   2.5s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=1.000, test=
0.861) total time=   5.0s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=1.000, test=
0.858) total time=   5.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=1.000, test=
0.857) total time=   5.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=1.000, test=
0.856) total time=   4.9s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=1.000, test=0.8
54) total time=  26.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=1.000, test=
0.857) total time=   5.0s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=1.000, test=
0.864) total time=  24.8s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=1.000, test=
0.861) total time=  24.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
```

```
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=1.000, test=
0.859) total time=  24.9s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.4s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.5s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.4s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.5s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=1.000, test=
0.861) total time=  25.6s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.8s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.2s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.0s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.8s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.8s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.9s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.9s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   2.0s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   2.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   2.2s
```

```
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   1.8s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   1.7s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   4.0s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   3.6s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   3.7s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   3.9s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   4.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=1.000, test=
0.856) total time=  24.8s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  19.9s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  19.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.000)
total time=   0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  20.0s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.000)
total time=   0.3s
```

```
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.5s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.5s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.4s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.00
0) total time=   0.5s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  19.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   1.9s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.4s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.00
0) total time=   2.5s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   1.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   1.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   1.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   1.2s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   1.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   3.0s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   3.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   3.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   3.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   3.1s
```

```
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  20.6s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   6.4s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   6.5s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   6.4s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   6.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   6.4s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  32.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  32.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=0.000, test=0.00
0) total time=   0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   0.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   0.2s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   0.2s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=0.000, test=0.00
0) total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  32.5s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   0.5s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   0.4s
```

```
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   0.4s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   0.4s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=0.000, test=0.0
00) total time=   0.4s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=  32.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   2.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   2.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   2.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.
000) total time=   1.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   1.9s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   2.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.
000) total time=   1.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.
000) total time=   1.0s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.
000) total time=   1.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=0.000, test=0.
000) total time=   1.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.
000) total time=   2.7s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.
000) total time=   2.7s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.
000) total time=   3.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.
000) total time=   2.7s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=0.000, test=0.
000) total time=   2.7s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=0.000, test=
0.000) total time=   6.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=0.000, test=
0.000) total time=   5.3s
```

```
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=0.000, test=
0.000) total time=    5.4s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=0.000, test=
0.000) total time=    5.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=0.000, test=
0.000) total time=    6.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=0.000, test=0.0
00) total time=   30.9s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=0.000, test=
0.000) total time=   29.9s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=0.000, test=
0.000) total time=   27.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=0.000, test=
0.000) total time=   30.7s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=0.000, test=
0.000) total time=   26.7s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=0.000, test=
0.000) total time=   21.5s
```
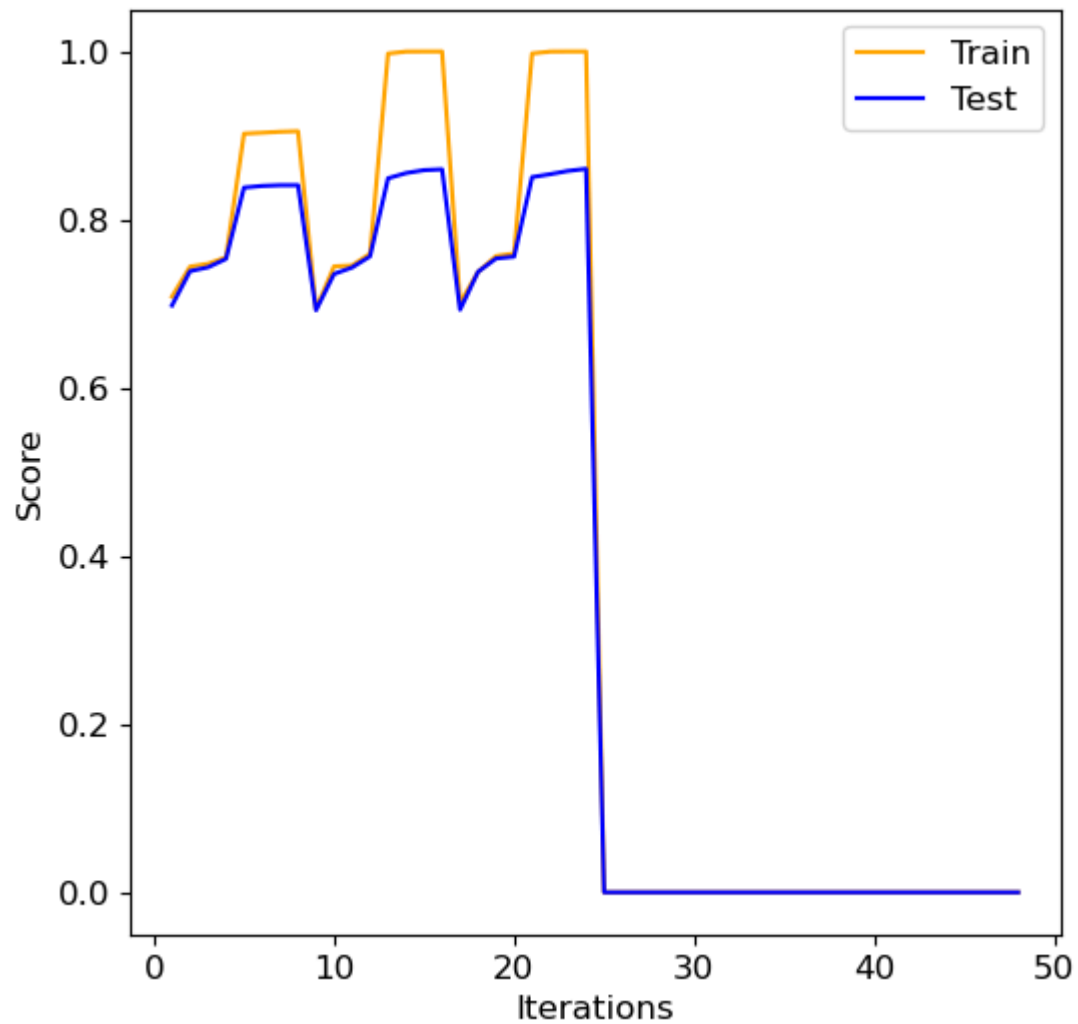
In [ ]:
```python
# Result
display(rfcls_gs.best_params_)
rfcls_gs.best_score_
```

```
{'ccp_alpha': 0.0, 'max_depth': 30, 'max_samples': None, 'n_estimators': 500}
```
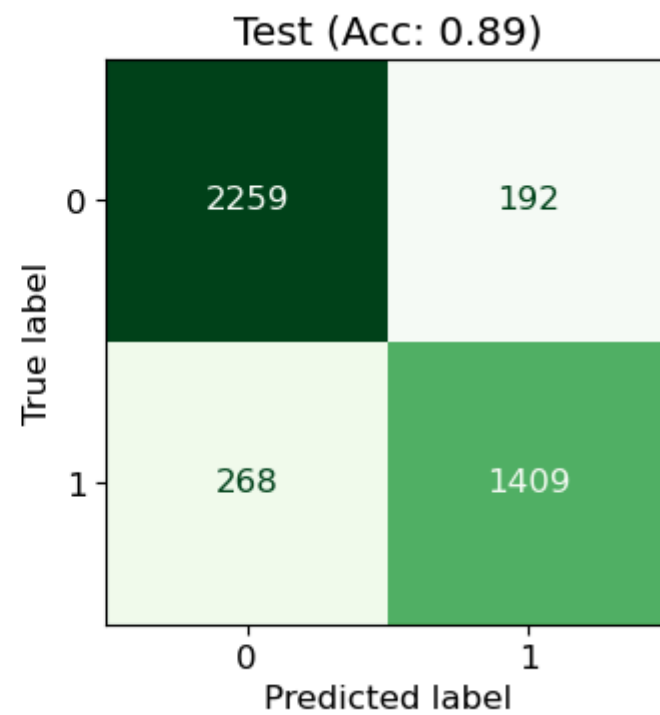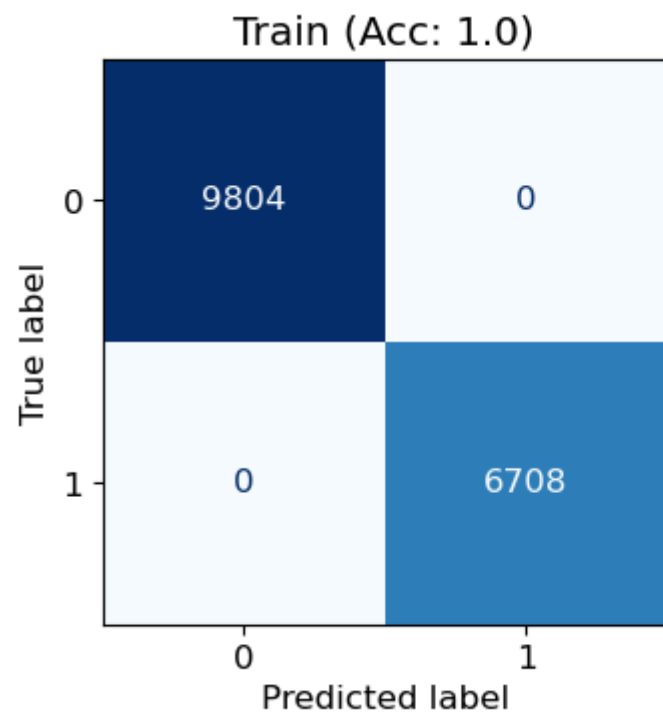
Out[ ]: 0.8615660171946644

In [ ]:
```python
# Train/Test scores
plot_train_test_error(rfcls_gs)
```
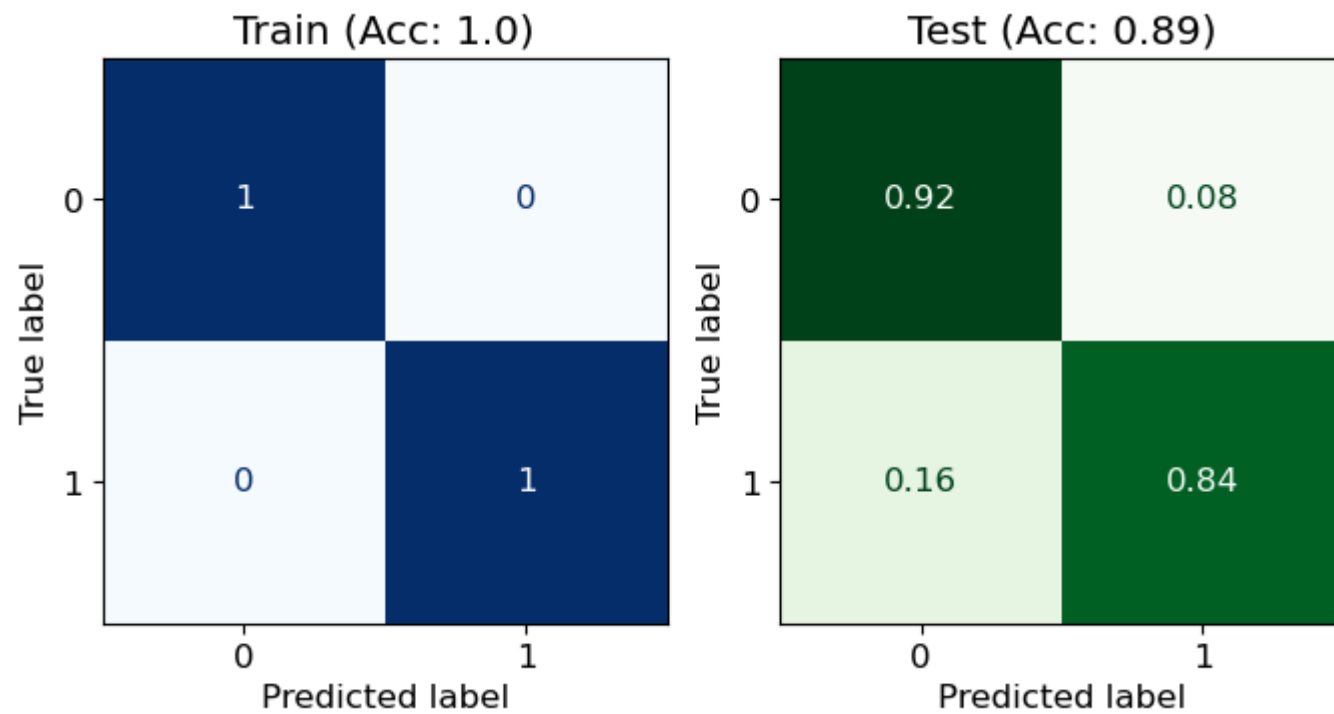
```python
In [ ]: # Prediction
pred_train = rfcls_gs.best_estimator_.predict(X_train)
pred_test = rfcls_gs.best_estimator_.predict(X_test)
```

```python
In [ ]: # Confusion matrix
display(plot_confusion_matrix(y_train, pred_train, y_test, pred_test, labels=rfcls_gs.classes_))
display(plot_confusion_matrix(y_train, pred_train, y_test, pred_test, labels=rfcls_gs.classes_, normalize=
```
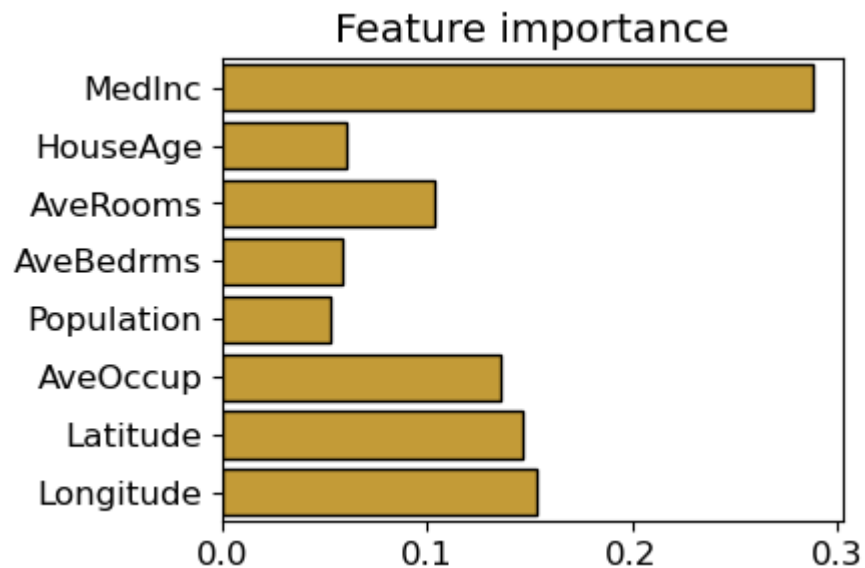
None

Train (Acc: 1.0)

	Predicted 0	Predicted 1
True 0	1	0
True 1	0	1

Test (Acc: 0.89)

	Predicted 0	Predicted 1
True 0	0.92	0.08
True 1	0.16	0.84

None

In [ ]:
```python
# Feature importance
plt.figure(figsize=(4,3))
sns.barplot(pd.DataFrame([i for i in zip(X_train.columns, rfcls_gs.best_estimator_.feature_importances_)])
plt.xlabel("")
plt.ylabel("")
plt.title("Feature importance")
plt.show()
```

Feature importance

For regression:

Train a decision tree regressor on the data with prices as the targets. As the classifier case, try to examine how the depth of your trees affect your results. Again, all 13 dimensional features are used as input, and the prices as the labels/targets.

In [ ]:
```python
# Data
X_train, y_train, X_test, y_test = load_data(mode = "Regression")
```

In [ ]:
```python
## Grid Search: Random forest

# Model
rfreg = RandomForestRegressor()

# Parameters
params = {
    'n_estimators': [20,50,100,500],
    'max_depth': [10, 30, None],
    'max_samples': [35, None],
    'ccp_alpha': [0.0, 0.5],
}
```

```python
# Search
rfreg_gs = GridSearchCV(rfreg, params, n_jobs=4, cv=5, verbose=3, scoring='neg_root_mean_squared_error', re
```

```
Fitting 5 folds for each of 48 candidates, totalling 240 fits
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.741, test=-0.74
7) total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.751, test=-0.75
5) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.779, test=-0.77
4) total time=   0.1s[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(tra
in=-0.763, test=-0.769) total time=   0.1s

[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.771, test=-0.77
3) total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.724, test=-0.72
8) total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.737, test=-0.74
2) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.726, test=-0.72
3) total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.734, test=-0.74
4) total time=   0.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.729, test=-0.72
3) total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.732, test=-0.7
34) total time=   0.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.730, test=-0.7
32) total time=   0.3s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.722, test=-0.7
17) total time=   0.3s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.730, test=-0.7
37) total time=   0.3s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.724, test=-0.7
20) total time=   0.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.725, test=-0.7
29) total time=   1.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.721, test=-0.7
24) total time=   1.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.729, test=-0.7
24) total time=   1.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.725, test=-0.7
34) total time=   1.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.724, test=-0.7
20) total time=   1.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-0.408, test=-0.
```

```
559) total time=   1.7s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-0.409, test=-0.
551) total time=   1.8s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-0.421, test=-0.
535) total time=   1.7s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-0.409, test=-0.
541) total time=   2.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-0.411, test=-0.
550) total time=   2.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-0.407, test=-0.
552) total time=   6.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-0.402, test=-0.
538) total time=   6.0s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-0.413, test=-0.
529) total time=   5.9s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-0.407, test=-0.
543) total time=   5.8s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-0.403, test=-0.
543) total time=   5.8s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-0.401, test=-
0.549) total time=  11.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-0.406, test=-
0.540) total time=  11.0s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-0.412, test=-
0.526) total time=  12.0s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-0.402, test=-
0.539) total time=  12.9s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-0.399, test=-
0.540) total time=  12.7s
[CV 2/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-0.401, test=-
0.538) total time=  54.5s
[CV 1/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-0.398, test=-
0.546) total time= 1.0min
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.774, test=-0.78
0) total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.732, test=-0.73
2) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.744, test=-0.73
7) total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.749, test=-0.75
6) total time=   0.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.754, test=-0.74
```

8) total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.736, test=-0.74
4) total time=   0.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.733, test=-0.73
5) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.725, test=-0.72
1) total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.726, test=-0.73
3) total time=   0.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.740, test=-0.73
6) total time=   0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.715, test=-0.7
17) total time=   0.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.722, test=-0.7
26) total time=   0.3s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.731, test=-0.7
29) total time=   0.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.729, test=-0.7
40) total time=   0.3s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.728, test=-0.7
29) total time=   0.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.720, test=-0.7
24) total time=   1.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-0.401, test=-
0.536) total time= 55.6s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.727, test=-0.7
30) total time=   1.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.724, test=-0.7
19) total time=   1.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-0.410, test=-
0.526) total time= 1.1min
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.724, test=-0.7
33) total time=   1.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.724, test=-0.7
23) total time=   1.4s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-0.211, test=-0.
525) total time=   3.9s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-0.207, test=-0.
519) total time=   4.0s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-0.211, test=-0.
507) total time=   3.9s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-0.207, test=-0.

529) total time=   3.8s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-0.206, test=-0.
530) total time=   3.9s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-0.194, test=-0.
514) total time=   9.5s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-0.197, test=-0.
511) total time=   9.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-0.199, test=-0.
500) total time=   9.7s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-0.194, test=-0.
508) total time=   9.6s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-0.193, test=-0.
517) total time=   9.9s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-0.192, test=-
0.516) total time= 20.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-0.400, test=-
0.542) total time= 55.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-0.192, test=-
0.508) total time= 19.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-0.194, test=-
0.496) total time= 19.8s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-0.190, test=-
0.510) total time= 20.0s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-0.189, test=-
0.515) total time= 20.0s
[CV 1/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-0.187, test=-
0.513) total time= 1.6min
[CV 2/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-0.188, test=-
0.505) total time= 1.6min
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.754, test=-0.
760) total time=   0.1s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.743, test=-0.
744) total time=   0.1s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.753, test=-0.
745) total time=   0.1s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.694, test=-0.
702) total time=   0.1s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.732, test=-0.
733) total time=   0.1s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.737, test=-0.
739) total time=   0.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.746, test=-0.

```
746) total time=    0.2s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.734, test=-0.
733) total time=    0.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.733, test=-0.
746) total time=    0.2s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.734, test=-0.
734) total time=    0.2s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.728, test=-
0.733) total time=    0.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.724, test=-
0.728) total time=    0.3s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.728, test=-
0.723) total time=    0.3s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.722, test=-
0.733) total time=    0.3s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.724, test=-
0.722) total time=    0.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.728, test=-
0.732) total time=    1.6s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.725, test=-
0.729) total time=    1.4s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.721, test=-
0.715) total time=    1.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.727, test=-
0.736) total time=    1.4s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.723, test=-
0.720) total time=    1.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-0.186, test=-
0.505) total time= 1.6min
[CV 3/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-0.190, test=-
0.493) total time= 1.7min
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-0.208, test=-
0.523) total time=    4.2s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-0.210, test=-
0.518) total time=    3.9s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-0.212, test=-
0.508) total time=    4.4s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-0.205, test=-
0.519) total time=    3.9s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-0.202, test=-
0.524) total time=    3.7s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-0.194, test=-
```

```
0.515) total time=   9.8s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-0.197, test=-
0.510) total time=   9.5s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-0.200, test=-
0.501) total time=  10.3s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-0.193, test=-
0.522) total time=   9.2s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-0.194, test=-
0.509) total time=  10.3s
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-0.191, test=
-0.514) total time=  19.6s
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-0.195, test=
-0.494) total time=  18.8s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-0.192, test=
-0.506) total time=  20.8s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-0.189, test=
-0.507) total time=  21.8s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-0.190, test=
-0.516) total time=  18.7s
[CV 5/5] END ccp_alpha=0.0, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-0.185, test=-
0.513) total time= 1.7min
[CV 1/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-0.188, test=
-0.511) total time= 1.8min
[CV 3/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-0.191, test=
-0.493) total time= 1.6min
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.949, test=-0.95
0) total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.961, test=-0.95
3) total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.980, test=-0.97
6) total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.907, test=-0.92
2) total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=20;, score=(train=-0.904, test=-0.89
5) total time=   0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.941, test=-0.94
2) total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.920, test=-0.91
3) total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.906, test=-0.90
6) total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.942, test=-0.95
```

```
8) total time=    0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=50;, score=(train=-0.961, test=-0.95
4) total time=    0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.921, test=-0.9
22) total time=    0.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.945, test=-0.9
38) total time=    0.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.945, test=-0.9
45) total time=    0.3s
[CV 2/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-0.188, test=
-0.506) total time= 1.7min
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.901, test=-0.9
17) total time=    0.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=100;, score=(train=-0.960, test=-0.9
52) total time=    0.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.940, test=-0.9
42) total time=    1.4s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.934, test=-0.9
27) total time=    1.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.934, test=-0.9
33) total time=    1.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.930, test=-0.9
46) total time=    1.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=35, n_estimators=500;, score=(train=-0.944, test=-0.9
36) total time=    1.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
148) total time=    2.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
146) total time=    2.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-1.154, test=-1.
150) total time=    2.5s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-1.148, test=-1.
174) total time=    2.5s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
150) total time=    2.3s
[CV 4/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-0.186, test=
-0.506) total time= 1.6min
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.
146) total time=    5.5s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.
148) total time=    6.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-1.154, test=-1.
```

150) total time=   5.8s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-1.148, test=-1.174) total time=   5.6s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.150) total time=   6.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-1.148) total time=   11.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-1.146) total time=   12.0s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-1.154, test=-1.150) total time=   11.5s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-1.148, test=-1.174) total time=   11.4s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-1.150) total time=   11.6s
[CV 5/5] END ccp_alpha=0.0, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-0.186, test=-0.514) total time= 1.6min
/home/au614901/Software/miniconda3/envs/MLAAU/lib/python3.12/site-packages/joblib/externals/loky/process_executor.py:752: UserWarning: A worker stopped while some jobs were given to the executor. This can be caused by a too short worker timeout or by a memory leak.
  warnings.warn(

```
[CV 1/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.148) total time=  57.4s
[CV 2/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.146) total time=  59.0s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.918, test=-0.91
9) total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.897, test=-0.89
1) total time=   0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.968, test=-0.96
8) total time=   0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.969, test=-0.98
8) total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=20;, score=(train=-0.937, test=-0.93
0) total time=   0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.928, test=-0.93
1) total time=   0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.967, test=-0.96
0) total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.913, test=-0.91
2) total time=   0.2s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.918, test=-0.93
3) total time=   0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=50;, score=(train=-0.962, test=-0.95
4) total time=   0.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.934, test=-0.9
38) total time=   0.2s
[CV 3/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-1.154, test=-
1.150) total time=  58.0s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.940, test=-0.9
32) total time=   0.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.949, test=-0.9
46) total time=   0.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.949, test=-0.9
67) total time=   0.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=100;, score=(train=-0.949, test=-0.9
42) total time=   0.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.932, test=-0.9
34) total time=   1.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.932, test=-0.9
25) total time=   1.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.938, test=-0.9
37) total time=   1.4s
```

```
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.948, test=-0.9
65) total time=   1.2s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=35, n_estimators=500;, score=(train=-0.932, test=-0.9
24) total time=   1.5s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
148) total time=  22.8s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
146) total time=  22.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-1.154, test=-1.
150) total time=  20.7s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-1.148, test=-1.
174) total time=  20.2s
[CV 4/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-1.148, test=-
1.174) total time=  59.6s
[CV 5/5] END ccp_alpha=0.5, max_depth=10, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.150) total time= 1.0min
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-1.
150) total time=  22.7s
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-1.154, test=-1.
150) total time=  52.5s
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.
148) total time=  53.7s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.
146) total time=  57.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-1.148, test=-1.
174) total time=  56.5s
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-1.
150) total time=  57.8s
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-
1.146) total time= 1.7min
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-
1.148) total time= 1.8min
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-1.154, test=-
1.150) total time= 1.7min
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-1.148, test=-
1.174) total time= 1.9min
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=100;, score=(train=-1.155, test=-
1.150) total time= 1.8min
[CV 1/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.148) total time= 8.3min
[CV 2/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.146) total time= 8.3min
```

```
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.913, test=-0.
914) total time=    0.1s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.931, test=-0.
923) total time=    0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.926, test=-0.
925) total time=    0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.935, test=-0.
951) total time=    0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=20;, score=(train=-0.962, test=-0.
954) total time=    0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.940, test=-0.
944) total time=    0.2s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.979, test=-0.
970) total time=    0.1s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.927, test=-0.
926) total time=    0.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.917, test=-0.
933) total time=    0.1s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=50;, score=(train=-0.921, test=-0.
915) total time=    0.1s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.924, test=-
0.925) total time=    0.3s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.953, test=-
0.945) total time=    0.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.932, test=-
0.931) total time=    0.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.945, test=-
0.961) total time=    0.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=100;, score=(train=-0.945, test=-
0.937) total time=    0.3s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.933, test=-
0.935) total time=    1.4s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.935, test=-
0.927) total time=    1.4s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.933, test=-
0.932) total time=    1.3s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.945, test=-
0.962) total time=    1.3s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=35, n_estimators=500;, score=(train=-0.931, test=-
0.923) total time=    1.6s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-
1.148) total time=   19.6s
```

```
[CV 3/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-1.154, test=-
1.150) total time= 8.4min
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-
1.146) total time=  19.7s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-1.154, test=-
1.150) total time=  19.4s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-1.148, test=-
1.174) total time=  19.5s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=20;, score=(train=-1.155, test=-
1.150) total time=  19.5s
[CV 4/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-1.148, test=-
1.174) total time= 8.3min
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-
1.148) total time=  49.9s
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-
1.146) total time=  48.3s
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-1.154, test=-
1.150) total time=  49.1s
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-1.148, test=-
1.174) total time=  49.0s
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=50;, score=(train=-1.155, test=-
1.150) total time=  49.2s
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-1.155, test=
-1.148) total time= 1.7min
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-1.155, test=
-1.146) total time= 1.8min
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-1.154, test=
-1.150) total time= 1.8min
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-1.148, test=
-1.174) total time= 1.8min
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=100;, score=(train=-1.155, test=
-1.150) total time= 1.8min
[CV 5/5] END ccp_alpha=0.5, max_depth=30, max_samples=None, n_estimators=500;, score=(train=-1.155, test=-
1.150) total time= 8.4min
[CV 1/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-1.155, test=
-1.148) total time= 8.5min
[CV 2/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-1.155, test=
-1.146) total time= 8.5min
[CV 3/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-1.154, test=
-1.150) total time= 8.6min
[CV 4/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-1.148, test=
-1.174) total time= 8.1min
```
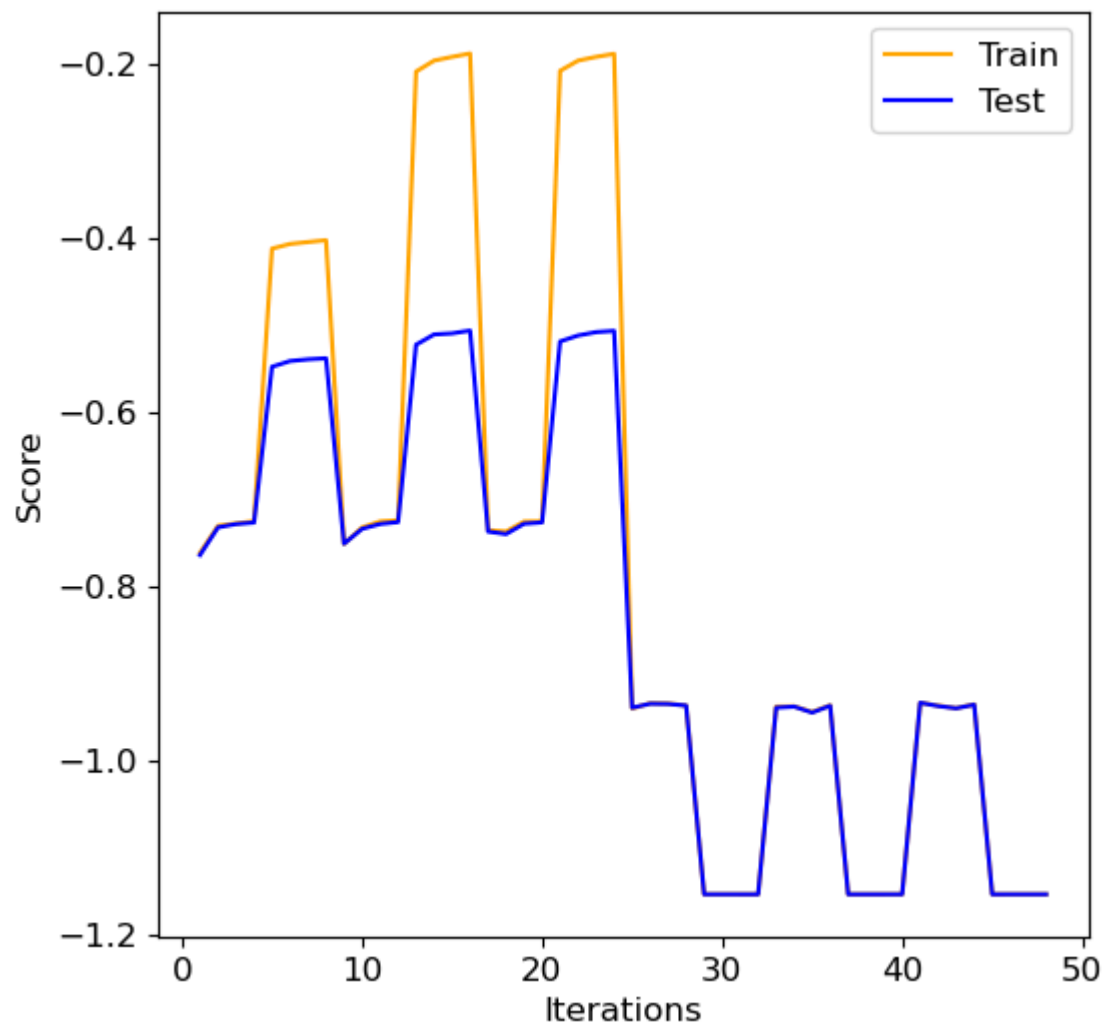
```
[CV 5/5] END ccp_alpha=0.5, max_depth=None, max_samples=None, n_estimators=500;, score=(train=-1.155, test=
-1.150) total time= 6.9min
```

In [ ]:
```
# Result
display(rfreg_gs.best_params_)
rfreg_gs.best_score_
```
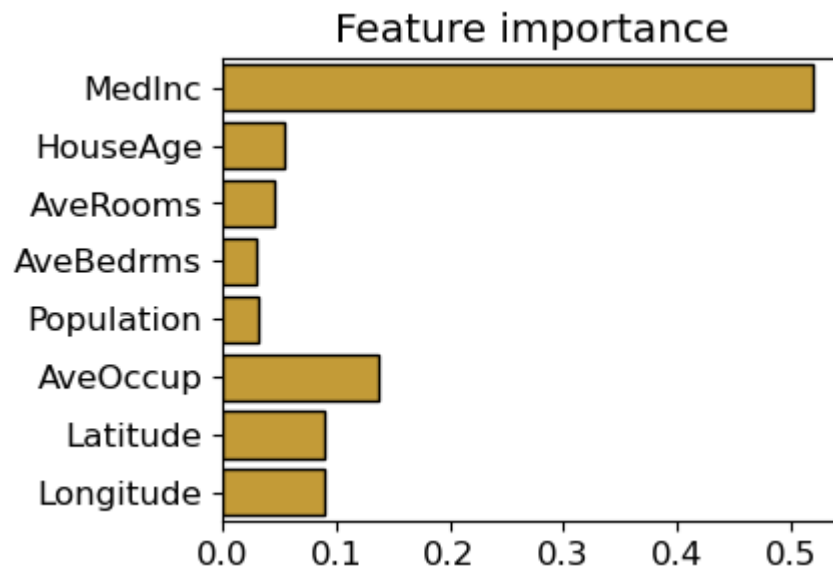
```
{'ccp_alpha': 0.0, 'max_depth': 30, 'max_samples': None, 'n_estimators': 500}
```

Out[ ]:   -0.5058012494222031

In [ ]:
```
# Train/Test scores
plot_train_test_error(rfreg_gs)
```

```
In [ ]:  # Feature importance
         plt.figure(figsize=(4,3))
         sns.barplot(pd.DataFrame([i for i in zip(X_train.columns, rfreg_gs.best_estimator_.feature_importances_)])
         plt.xlabel("")
         plt.ylabel("")
         plt.title("Feature importance")
         plt.show()
```
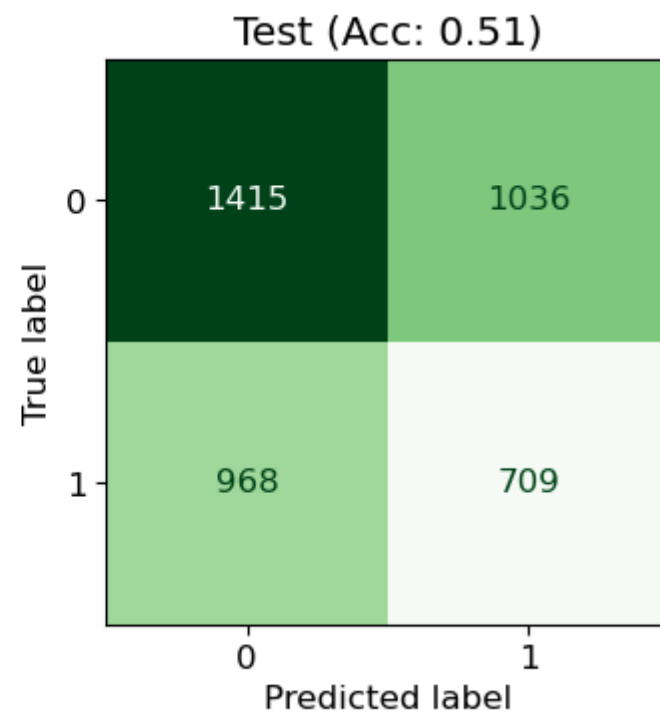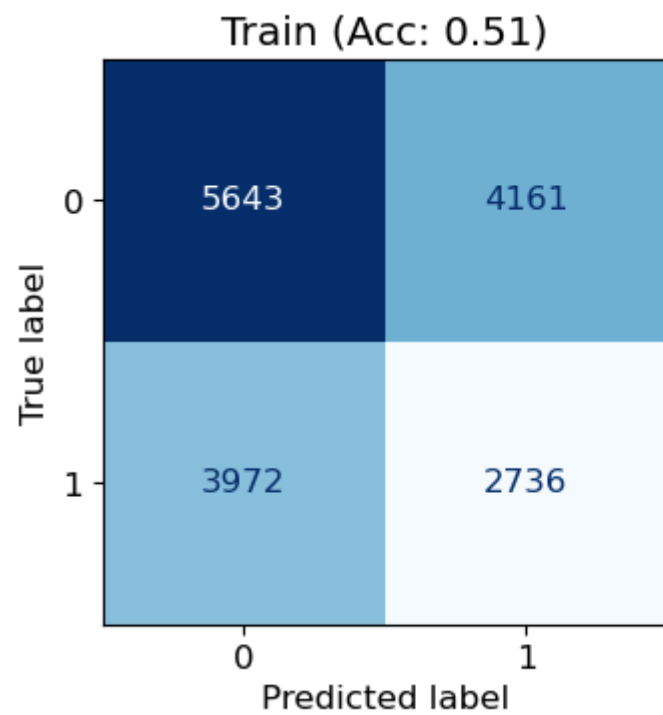
Feature importance


```
In [ ]:  # Predict
         pred_train = rfreg_gs.best_estimator_.predict(X_train)
         pred_test = rfreg_gs.best_estimator_.predict(X_test)
```

```
In [ ]:  # Measure score in test
         print(f"Mean squared error: {mean_squared_error(y_test, pred_test)}")
```
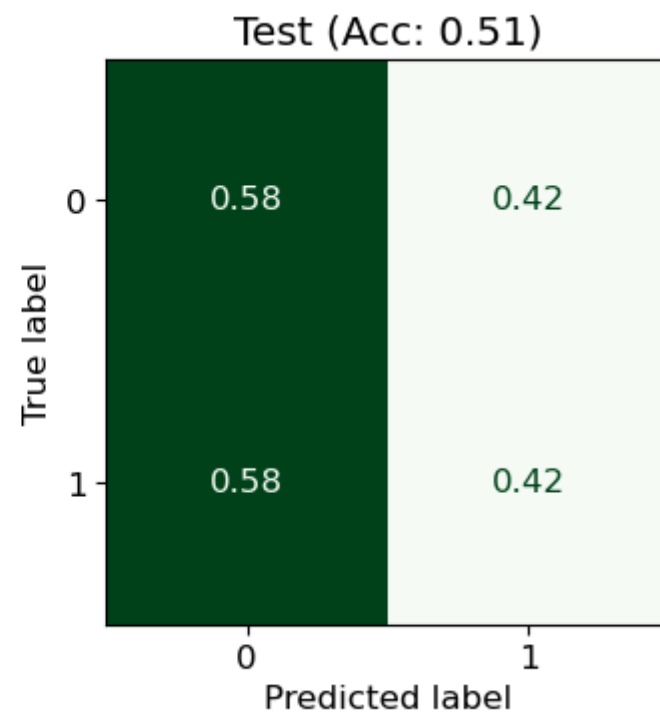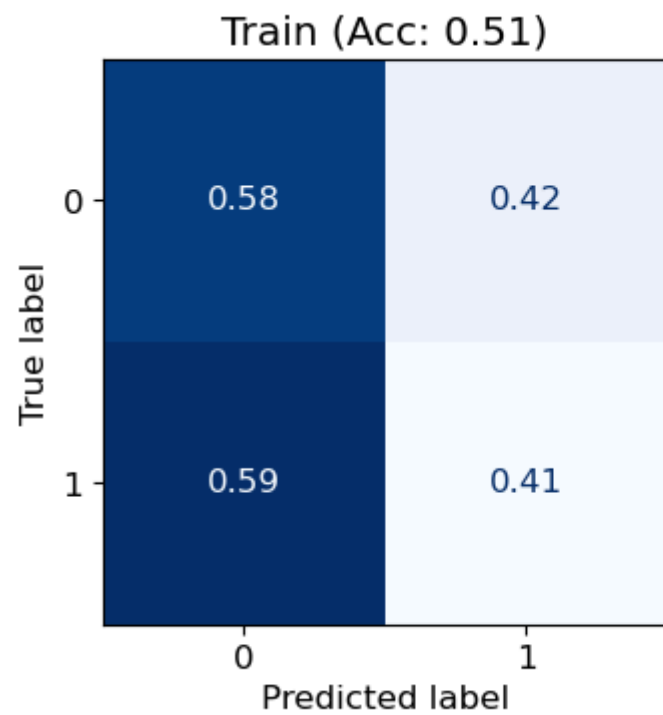
Mean squared error: 2.309608723397599

```
In [ ]:  # Transform prediction to class
         mean = pd.concat([y_train, y_test]).mean()
         pred_train_cat = [1 if i > mean else 0 for i in pred_train]
         pred_test_cat = [1 if i > mean else 0 for i in pred_test]
```

```
In [ ]:  # Confusion matrix
         X_train, y_train, X_test, y_test = load_data(mode="Classification")
         display(plot_confusion_matrix(y_train, pred_train_cat, y_test, pred_test_cat, labels=rfcls_gs.classes_))
         display(plot_confusion_matrix(y_train, pred_train_cat, y_test, pred_test_cat, labels=rfcls_gs.classes_, no
```

Train (Acc: 0.51)

True label	Predicted label 0	Predicted label 1
0	5643	4161
1	3972	2736

Test (Acc: 0.51)

True label	Predicted label 0	Predicted label 1
0	1415	1036
1	968	709

None

Train (Acc: 0.51)

	0	1
0	0.58	0.42
1	0.59	0.41

True label / Predicted label

Test (Acc: 0.51)

	0	1
0	0.58	0.42
1	0.58	0.42

True label / Predicted label

None