

# Problem Set 5: Trees, Forests, and Fairness in ML

Your mission for this problem set is to use your knowledge of tree-based methods and supervised learning to -- among other things! -- explore issues of fairness in machine learning (ML). Unlike in previous psets, where we were implementing algorithms from scratch, this pset will rely more heavily on sklearn. Unless explicitly noted otherwise, you are allowed (encouraged!) to make good use of this wonderful library.

This problem set will also rely on the resources provided by the [folktables](#) project. Before getting started, take some time to understand what folktables is about and how it can be used to benchmark ML algorithms in social science. If you don't understand the basics, this pset will be very challenging!!

Some resources you might find useful:

[Folktables paper](#)

[Folktables video](#)

[PUMS\\_Data\\_Dictionary\\_2018](#)

[Fairness in ML](#)

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# !pip install folktables
import folktables
import seaborn as sns
```

## Part I: Setup of prediction task

### 1.1 Define the income prediction task

For this first question we are providing the code. Run it to download the data you'll be using throughout this problem set.

Begin by explaining what this prediction task is about and what is this code accomplishing:

- Where do the data come from
- What is the sample?
- What are we trying to predict?

- What are the features that we will be using?
- What is the "group" feature about?

```
In [ ]: from folktables import ACSDataSource, generate_categories

def adult_filter(data):
    """Mimic the filters in place for Adult data.

    Adult documentation notes: Extraction was done by Barry Becker from
    the 1994 Census database. A set of reasonably clean records was extracted
    using the following conditions:
    ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0))
    """
    df = data
    df = df[df['AGEP'] > 16]
    df = df[df['PINCP'] > 100]
    df = df[df['WKHP'] > 0]
    df = df[df['PWGTP'] >= 1]
    return df

ACSIIncome = folktables.BasicProblem(
    features=[
        'AGEP',
        'COW',
        'SCHL',
        'MAR',
        'POBP',
        'RELP',
        'WKHP',
        'SEX',
        'RAC1P',
    ],
    target='PINCP',
    target_transform=lambda x: x > 50000,
    group='RAC1P',
    preprocess=adult_filter,
    postprocess=lambda x: np.nan_to_num(x, -1),
)

data_source = ACSDataSource(survey_year='2018', horizon='1-Year', survey='pe
acs_data = data_source.get_data(states=["CA"], download=True)
definition_df = data_source.get_definitions(download=True)
categories = generate_categories(features=ACSIIncome.features, definition_df=
features, target, group = ACSIIncome.df_to_pandas(acs_data)

feature_names = ACSIIncome.features
target_name = ACSIIncome.target
group_name = ACSIIncome.group
```

*your answer here*

- Where do the data come from

- from the American Community Survey (part of the US Census), the 2018 1-year-summary version for the state of California
  - the data was accessed through the Folktables dataset
- What is the sample?
  - a set of 195k individual people in California
- What are we trying to predict?
  - We are trying to predict **PINCP** - Income
- What are the features that we will be using?
  - **'AGEP'** - Age (years) - Numeric
  - **'COW'** - Class of Worker (b, 1-9) - Categorical
  - **'SCHL'** - Educational Attainment (bb, 00-24) - Categorical
  - **'MAR'** - Marital Status - Categorical
  - **'POBP'** - Place of Birth (state fips/country code) - Categorical
  - **'RELP'** - Relationship (00-17) - Categorical
  - **'WKHP'** - hrs worked/week past 12 months (bb, 1-99) - Numeric
  - **'SEX'** - Sex (Male, Female) - Categorical
  - **'RAC1P'** - Race (1-9) - Categorical
- What is the "group" feature about?
  - the group **'RAC1P'** (Race), means we want to look at the values by race

## 1.2 Split into train and test

Split your data into 80%-20% train and test splits. Please use `sklearn.model_selection.train_test_split` and set the `random_state` parameter equal to 0.

```
In [ ]: from sklearn.model_selection import train_test_split

# your code here
seed = 0

X_train, X_test, y_train, y_test, group_train, group_test = train_test_split(
    features,
    target,
    group,
    test_size=0.2,
    random_state=0
)

# Asserts
assert X_train.shape[0] == y_train.shape[0] == group_train.shape[0]
assert X_test.shape[0] == y_test.shape[0] == group_test.shape[0]

# Replace 1d dataframes with series
y_train = y_train['PINCP']
y_test = y_test['PINCP']
```

```
group_train = group_train['RAC1P']
group_test = group_test['RAC1P']
```

### 1.3 Explore the data and set appropriate data types

Create 2-3 figures and tables that explore the training data, and tell us what insights you can draw from those figures. Which features are best represented as numerical data types? Which features are best represented as categorical data types? Set each feature to its desired data type in both train and test splits. Make sure that the target variable in both splits is numeric before moving ahead.

```
In [ ]: # your code here

# Data Types:
## Numeric: AGE, WKHP, SCHL
numeric_cols = ['AGE', 'WKHP', 'SCHL']
categorical_cols = list(categories.keys()) #['COW', 'MAR', 'POBP', 'RELP', 'SEX']

X_train.groupby('POBP').count()

# TODO
```

```
Out [ ]:
```

	AGE	COW	SCHL	MAR	RELP	WKHP	SEX	RAC1P
<b>POBP</b>								
<b>1.0</b>	275	275	275	275	275	275	275	275
<b>2.0</b>	144	144	144	144	144	144	144	144
<b>4.0</b>	846	846	846	846	846	846	846	846
<b>5.0</b>	227	227	227	227	227	227	227	227
<b>6.0</b>	75604	75604	75604	75604	75604	75604	75604	75604
...	...	...	...	...	...	...	...	...
<b>512.0</b>	6	6	6	6	6	6	6	6
<b>515.0</b>	63	63	63	63	63	63	63	63
<b>523.0</b>	36	36	36	36	36	36	36	36
<b>527.0</b>	14	14	14	14	14	14	14	14
<b>554.0</b>	8	8	8	8	8	8	8	8

219 rows × 8 columns

*your answer here*

## Part II: ML pipeline for a Classification Tree

Next, you'll build an ML pipeline using sklearn. Take some time to familiarize yourself with [sklearn.pipeline.Pipeline](#) before proceeding.

## 2.1 Pre-processing of features

The first step of the pipeline will standardize the numeric features and one-hot encode the categorical features. Write the code for this preprocessing step below.

**Hint:** Take a look at sklearn's ColumnTransformer, OneHotEncoder and StandardScaler for this task.

```
In [ ]: from sklearn.pipeline import Pipeline
        from sklearn.compose import ColumnTransformer
        from sklearn.preprocessing import OneHotEncoder, StandardScaler

        # your code here

        # Define Transformers
        numeric_TF = StandardScaler()
        categorical_TF = OneHotEncoder(handle_unknown='ignore') # Ignore because

        # Define Pre-processing step
        preprocessing_step = ColumnTransformer(
            transformers=[
                ('numeric', numeric_TF, numeric_cols),
                ('categorical', categorical_TF, categorical_cols)
            ]
        )
```

## 2.2 Use your pre-processed data to fit a basic DecisionTreeClassifier

Now, bring together the pre-processing step with a Decision Tree Classifier in a pipeline. For now, use the default values; we'll take care of hyperparameter optimization later. Fit the pipeline on the training data. How deep is the resulting tree? How many leafs are in the tree? Print the 10 variables with highest feature importance and tell us what those importances mean in plain English. Plot the top 5 levels of the fitted decision tree. Make sure that the nodes are appropriately labeled. Comment on anything that you find interesting.

```
In [ ]: from sklearn.tree import DecisionTreeClassifier
        from sklearn import tree

        # your code here
        # Define Pipeline
        pipe = Pipeline(
            steps = [
                # Preprocessing
                ('preprocessing', preprocessing_step),
```

```

        # Estimating
        ('decision_tree', DecisionTreeClassifier(max_features=10))
    ]
)

pipe.fit(X_train, y_train)

y_hat_train = pipe.predict(X_train)
y_hat_train.shape

#acc = pipe.score(X_test, y_test)

```

Out[ ]: (156532,)

```

In [ ]: # How deep is the resulting tree?
print(f'max_depth: {pipe['decision_tree'].tree_.max_depth}') # 90

# How many leafs are in the tree?
print(f'leafs: {pipe['decision_tree'].get_n_leaves()}') # 37,902

print(f'n_feats: {pipe.named_steps['decision_tree'].n_features_in_}') # 287

# Print the 10 variables with highest feature importance and tell us what th
key_features = pd.Series(pipe['decision_tree'].feature_importances_)
key_features.index = [s.split('_')[1] for s in pipe.named_steps['preprocess
kf_top10 = key_features.nlargest(10)
print(f'top 10 feats:')
print('\n'.join(f'\t{line}' for line in str(kf_top10).split('\n')))

# Plot the top 5 levels of the fitted decision tree. Make sure that the node
from sklearn.tree import plot_tree

plt.figure(figsize=(200,40))
plot_tree(
    pipe.named_steps['decision_tree'],
    filled=True,
    max_depth=4,
    feature_names=pipe.named_steps['preprocessing'].get_feature_names_out().
)
plt.show()

```

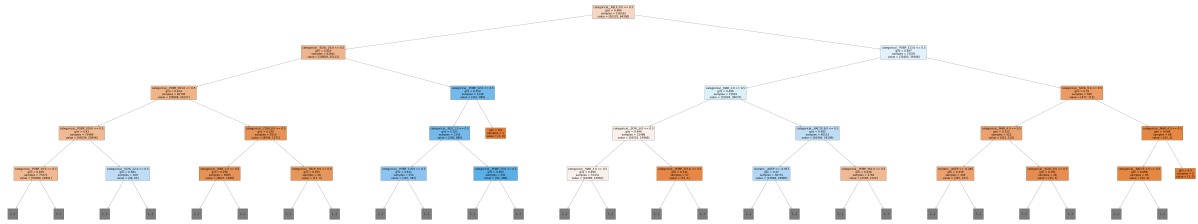
max\_depth: 127

leafs: 62349

n\_feats: 288

top 10 feats:

AGEP	0.267800
WKHP	0.168838
RELP_0.0	0.065333
MAR_1.0	0.056994
SCHL	0.055167
SCHL_21.0	0.020894
SCHL_22.0	0.019278
POBP_303.0	0.015529
RAC1P_8.0	0.013687
SEX_1.0	0.011456
dtype:	float64



*your answer here*

There are a lot of splits. This is a big tree and probably should be pruned a lot.

## 2.3 Estimate AUC using K-Fold CV

Calculate the 5-fold cross-validated AUC-ROC for this simple tree pipeline. Report the score for each fold as well as the average across all folds.

**Hint 1:** `sklearn.model_selection.cross_val_score` is your friend. Setting the verbose option to 3 is useful.

```
In [ ]: from sklearn.model_selection import cross_val_score, KFold
        from sklearn.metrics import make_scorer, roc_auc_score

        cv = KFold(n_splits=5, shuffle=True, random_state=0)
        AUCs_cv = cross_val_score(
            pipe,
            X_train,
            y_train,
            cv=cv,
            scoring='roc_auc',
            verbose=3
        )

        print(f'\nMean AUC: {np.mean(AUCs_cv):.3f}')

        # The error is that there are categories in one of the fold test sets that c
        # therefore its causing errors
```

```
[CV] END ..... score: (test=0.728) total time=
1.8s
[CV] END ..... score: (test=0.731) total time=
1.9s
[CV] END ..... score: (test=0.730) total time=
1.9s
[CV] END ..... score: (test=0.734) total time=
1.9s
[CV] END ..... score: (test=0.725) total time=
1.9s
```

Mean AUC: 0.730

## 2.4 Estimating AUC using nested CV

Let's try to improve on the results of the decision tree by tweaking its hyperparameters. Since you already are an expert using nested CV, this is going to be a walk in the park! Compute the nested cv AUC-ROC of a decision tree pipeline. Use 5 folds for the inner loop and 3 folds for the outer loop. Include different values for "max\_depth", "max\_features" and "max\_leave\_nodes" in your hyperparameter grid. Report the inner and outer fold scores for all combinations of hyperparams and interpret the results. Report the best hyperparameters used in each outer fold. How stable are the results?

What else do you notice in these results? To what extent do you think specific hyperparameters might lead to overfitting? How is run-time impacted by your choice of hyperparameters? Is hyperparameter optimization worth the trouble?

**Hint 1:** The pipeline remains the same. You can access the list of all the hyperparams in your pipeline with `pipeline.get_params_keys()`. Notice the naming convention.

**Hint 2:** `GridSearchCV` and `cross_val_score` are your friends-- especially if you run them with `verbose = 3`.

**Hint 3:** This gets computationally expensive quickly. Be intelligent about the hyperparameter values you include in the grid and the number of outer and inner folds you use (we use 5 inner folds and 3 outer folds).

```
In [ ]: from sklearn.model_selection import GridSearchCV
        from sklearn.model_selection import cross_validate

        # your code here
        outer_k, inner_k = 3, 5

        inner_cv = KFold(n_splits=inner_k, shuffle=True, random_state=0)
        outer_cv = KFold(n_splits=outer_k, shuffle=True, random_state=0)

        pipe.get_params()['preprocessing'].get_params()
        pipe.get_params()['decision_tree'].get_params()

        # Parameter space to search
        param_grid = {
            'decision_tree__max_depth': np.arange(3, 23, 10), # current is 89
            'decision_tree__max_features': np.arange(50, 150, 50), # Total Features:
            'decision_tree__max_leaf_nodes': np.arange(100, 500, 200) # Look at what
        }
```

```
In [ ]: # Assign the training set to a new value
        X_fold = X_train
        y_fold = y_train.values

        outer_scores_dt = []
        for outer_train_idx, outer_test_idx in outer_cv.split(X_fold):
            X_outer_train, X_outer_test = X_fold.iloc[outer_train_idx], X_fold.iloc[outer_test_idx]
            y_outer_train, y_outer_test = y_fold[outer_train_idx], y_fold[outer_test_idx]

            # Define Grid Search Object - uses all the params in `param_grid` to fit
```



```
grid_search_dt = GridSearchCV(  
    estimator=pipe,  
    param_grid=param_grid,  
    cv=inner_cv,  
    verbose=3,  
    n_jobs=-1  
)  
  
# Fit Model on all hyperparameters  
grid_search_dt.fit(X_outer_train, y_outer_train)  
  
inner_scores_dt = grid_search_dt.cv_results_  
  
#inner_scores = cross_val_score(grid_search, X_outer_train, y_outer_train)  
outer_scores_dt.append(inner_scores_dt)
```

Fitting 5 folds for each of 8 candidates, totalling 40 fits

```
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.629 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.699 total time= 0.3s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.681 total time= 0.3s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.730 total time= 0.3s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.648 total time= 0.3s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.684 total time= 0.3s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.690 total time= 0.3s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.721 total time= 0.3s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.733 total time= 0.3s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.713 total time= 0.3s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=100;; score=0.705 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=100;; score=0.722 total time= 0.1s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=100;; score=0.657 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=100;; score=0.749 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=100;; score=0.747 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=300;; score=0.729 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=300;; score=0.743 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=300;; score=0.732 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=300;; score=0.729 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, decision_tree__max_leaf_nodes=300;; score=0.683 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.780 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.774 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.772 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.775 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=100;; score=0.771 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.781 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, decision_tree__max_leaf_nodes=300;; score=0.779 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
```

```
cision_tree__max_leaf_nodes=300;; score=0.778 total time= 0.3s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.772 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.779 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.787 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.783 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.784 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.785 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.782 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.787 total time= 0.3s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.789 total time= 0.3s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.785 total time= 0.3s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.786 total time= 0.3s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.778 total time= 0.3s
Fitting 5 folds for each of 8 candidates, totalling 40 fits
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.728 total time= 0.1s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.742 total time= 0.1s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.732 total time= 0.1s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.655 total time= 0.1s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.657 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=300;; score=0.728 total time= 0.1s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=300;; score=0.640 total time= 0.1s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=300;; score=0.733 total time= 0.1s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=300;; score=0.712 total time= 0.1s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=300;; score=0.680 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.739 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.744 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.745 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.744 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.742 total time= 0.2s
```

```

[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.726 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.674 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.750 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.706 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.720 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.781 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.775 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.769 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.772 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.770 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.768 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.785 total time= 0.3s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.770 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.786 total time= 0.3s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.781 total time= 0.3s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.776 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.780 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.788 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.780 total time= 0.4s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.788 total time= 0.4s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.780 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.778 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.782 total time= 0.3s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.787 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.787 total time= 0.3s
Fitting 5 folds for each of 8 candidates, totalling 40 fits
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.660 total time= 0.1s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec
ision_tree__max_leaf_nodes=100;; score=0.714 total time= 0.1s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, dec

```

```
ision_tree__max_leaf_nodes=100;; score=0.731 total time= 0.1s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.661 total time= 0.1s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.733 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.668 total time= 0.1s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.744 total time= 0.1s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.746 total time= 0.1s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.678 total time= 0.1s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.698 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.713 total time= 0.1s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.748 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.751 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.741 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=100;; score=0.703 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.685 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.714 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.751 total time= 0.1s
[CV 4/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.740 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=3, decision_tree__max_features=100, de
cision_tree__max_leaf_nodes=300;; score=0.744 total time= 0.1s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.770 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.769 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.772 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.779 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.780 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.777 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.749 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=100;; score=0.781 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.789 total time= 0.3s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.784 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
```

```

cision_tree__max_leaf_nodes=300;; score=0.773 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.786 total time= 0.2s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.783 total time= 0.3s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=50, de
cision_tree__max_leaf_nodes=300;; score=0.777 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.786 total time= 0.3s
[CV 2/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.782 total time= 0.2s
[CV 4/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=100;; score=0.775 total time= 0.2s
[CV 1/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.780 total time= 0.2s
[CV 3/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.787 total time= 0.2s
[CV 5/5] END decision_tree__max_depth=13, decision_tree__max_features=100, d
ecision_tree__max_leaf_nodes=300;; score=0.781 total time= 0.2s
k=0 -- Mean CV Score: 0.689 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=0 -- Mean CV Score: 0.697 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=0 -- Mean CV Score: 0.716 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=0 -- Mean CV Score: 0.723 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}
k=0 -- Mean CV Score: 0.774 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=0 -- Mean CV Score: 0.778 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=0 -- Mean CV Score: 0.784 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=0 -- Mean CV Score: 0.785 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}
k=1 -- Mean CV Score: 0.703 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=1 -- Mean CV Score: 0.699 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=1 -- Mean CV Score: 0.743 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=1 -- Mean CV Score: 0.715 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}
k=1 -- Mean CV Score: 0.772 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=1 -- Mean CV Score: 0.777 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=1 -- Mean CV Score: 0.782 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=1 -- Mean CV Score: 0.785 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}
k=2 -- Mean CV Score: 0.700 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=2 -- Mean CV Score: 0.707 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=2 -- Mean CV Score: 0.731 with Parameters: {'decision_tree__max_depth': 3,

```

```

'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=2 -- Mean CV Score: 0.727 with Parameters: {'decision_tree__max_depth': 3,
'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}
k=2 -- Mean CV Score: 0.768 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 100}
k=2 -- Mean CV Score: 0.779 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_nodes': 300}
k=2 -- Mean CV Score: 0.781 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 100}
k=2 -- Mean CV Score: 0.783 with Parameters: {'decision_tree__max_depth': 1
3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_nodes': 300}

```

```

In [ ]: print('All scores:')
        for i in range(0, outer_k):
            for mean_score, params in zip(outer_scores_dt[i]['mean_test_score'], out
                print(f"outer_fold={i} -- Mean CV Score: {mean_score:.3f} with Param

```

All scores:

```
outer_fold=0 -- Mean CV Score: 0.689 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=0 -- Mean CV Score: 0.697 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=0 -- Mean CV Score: 0.716 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=0 -- Mean CV Score: 0.723 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}
outer_fold=0 -- Mean CV Score: 0.774 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=0 -- Mean CV Score: 0.778 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=0 -- Mean CV Score: 0.784 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=0 -- Mean CV Score: 0.785 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}
outer_fold=1 -- Mean CV Score: 0.703 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=1 -- Mean CV Score: 0.699 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=1 -- Mean CV Score: 0.743 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=1 -- Mean CV Score: 0.715 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}
outer_fold=1 -- Mean CV Score: 0.772 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=1 -- Mean CV Score: 0.777 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=1 -- Mean CV Score: 0.782 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=1 -- Mean CV Score: 0.785 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}
outer_fold=2 -- Mean CV Score: 0.700 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=2 -- Mean CV Score: 0.707 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=2 -- Mean CV Score: 0.731 with Parameters: {'decision_tree__max_d
```



```

epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=2 -- Mean CV Score: 0.727 with Parameters: {'decision_tree__max_d
epth': 3, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}
outer_fold=2 -- Mean CV Score: 0.768 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 100}
outer_fold=2 -- Mean CV Score: 0.779 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 50, 'decision_tree__max_leaf_node
s': 300}
outer_fold=2 -- Mean CV Score: 0.781 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 100}
outer_fold=2 -- Mean CV Score: 0.783 with Parameters: {'decision_tree__max_d
epth': 13, 'decision_tree__max_features': 100, 'decision_tree__max_leaf_node
s': 300}

```

```

In [ ]: grid_search_dt.get_params().keys()

print('All scores:')

outer_fold_print_dt = [np.mean(i['mean_test_score']) for i in outer_scores_c
print(outer_fold_print_dt)
print(f'total AUC: {np.mean(outer_fold_print_dt):.3f}')

```

All scores:

[0.7432512644068131, 0.7470353600689952, 0.7470964496190887]

total AUC: 0.746

*your answer here*

**What else do you notice in these results? To what extent do you think specific hyperparameters might lead to overfitting? How is run-time impacted by your choice of hyperparameters? Is hyperparameter optimization worth the trouble?**

There doesn't seem to be much difference between different folds in the AUC and the ideal paramaters don't seem to change much. This may be indiciatice that we haven't searched the parameter-space enough yet. Total AUC of 0.746 is a good baseline and slightly better than the original, which means there may be better options out there.

## Part III: ML pipeline for a random forest

### 3.1 Bare bones random forest

Just as you did for the classification tree, start by training a bare bones random forest using your training data. We will take care of the hyperparameter optimization later, but for the time being, use a maximum tree depth of 15. Hopefully is it clear by now why this is a good idea, since we are going to be training a bunch of trees. You should continue to use the same preprocessing step as before.

How many trees are in the forest? Which are the most important features? Compute the fraction of the trees that are splitting on age at their root node. Create a plot showing the 10 most important features and compare them to the results you obtained before. Explain the meaning of variable importance in the case of a random forest classifier.

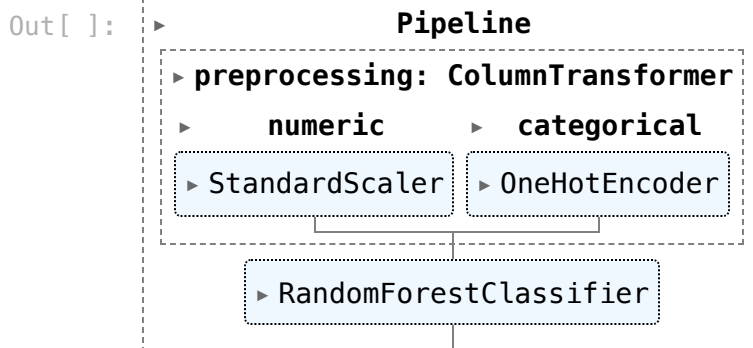
```
In [ ]: from sklearn.ensemble import RandomForestClassifier

# your code here

# Need to fit into an actual pipeline --

# Instantiate
pipe_rf = Pipeline(
    steps = [
        # Preprocessing
        ('preprocessing', preprocessing_step),
        # Estimating
        ('model', RandomForestClassifier(max_depth=15, random_state=0))
    ]
)
# Train
pipe_rf.fit(X_train, y_train)

# Predict
# y_hat_rf_test = pipe_rf.predict(X_test)
```



```
In [ ]: # How many trees?
print(f'Number of trees: \t{pipe_rf.named_steps['model'].n_estimators}')

# Key Features?
kf_rf = pipe_rf.named_steps['model'].feature_importances_
#print(kf_rf)

# Fraction of trees splitting on age at root
tree_list = pipe_rf.named_steps['model'].estimators_
root_node_list = [e.tree_.feature[0] for e in tree_list]
col_list = np.array([c.split('__')[1] for c in pipe_rf.named_steps['preproce
age_share = np.sum(col_list[root_node_list] == 'AGEP')/col_list.shape[0]
print(f'Share w/ root `AGEP`: \t{age_share:.2%}')
```

```

# Create a plot showing the 10 most important features and compare them to t
kf_rf_top10_idx = np.argsort(kf_rf)[::-1][0:10] # <-- reversed

```

```

kf_rf_top10 = kf_rf[kf_rf_top10_idx]

# Write a function to label the categorical feats correctly <---- Note: can p
def apply_label(feats_label):
    feat_split = feat_label.split('_')

    if(len(feat_split) == 1):
        return feat_label
    if(len(feat_split) == 2):
        cat_idx = int(float(feat_split[1]))
        cat_label = categories[feat_split[0]][cat_idx]
        return f'{feat_split[0]}\n({cat_idx} - {cat_label})'

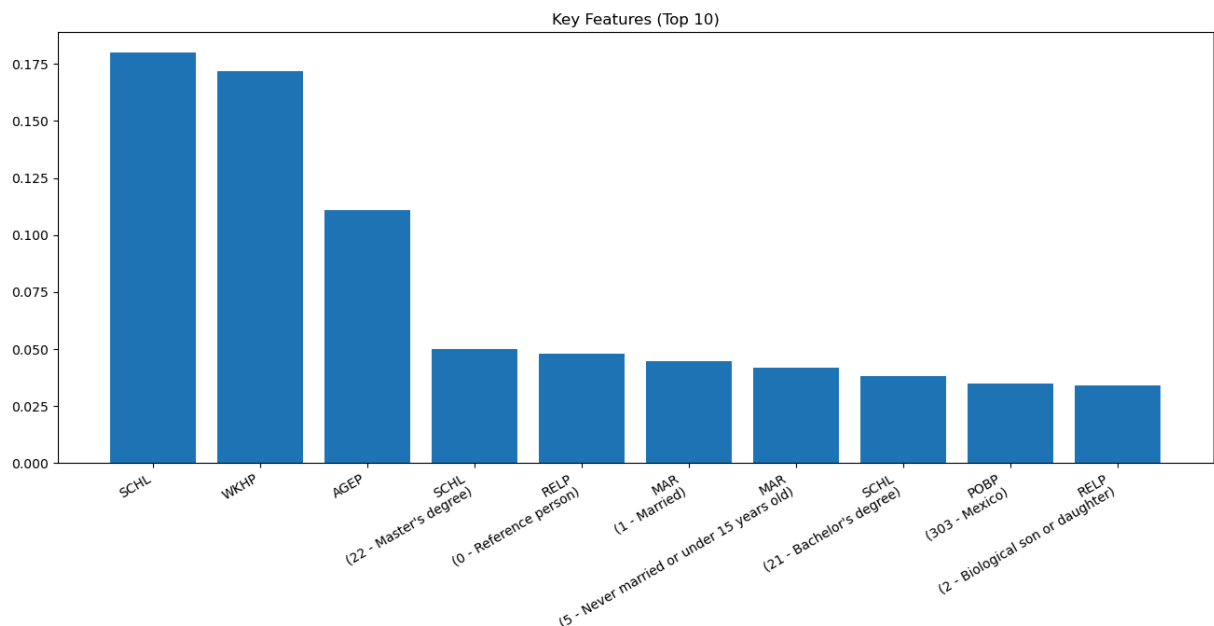
kf_rf_top10_labels = [apply_label(s) for s in col_list[kf_rf_top10_idx]]

plt.figure(figsize=(16,6))
plt.bar(np.arange(0,10), kf_rf_top10)
plt.xticks(np.arange(0,10), kf_rf_top10_labels, rotation = 30, ha='right')
plt.title('Key Features (Top 10)')
plt.plot()

```

Number of trees: 100  
 Share w/ root `AGEP`: 1.74%

Out[ ]: [ ]



*your answer here*

### Explain the meaning of variable importance in the case of a random forest

Variable importance in the context of random forest is a measure of how frequently and how powerful each feature is across all trees included in the forest.

Schooling and working hours are the most valuable differentiator as far as predicting income is concerned. Next up is age.

### 3.2 Estimating AUC using K-fold CV

Calculate the 5-fold cross-validated AUC-ROC for the random forests pipeline. Report the score for each fold as well as the average across all folds.

**Hint 1:** `sklearn.model_selection.cross_val_score` is your friend. Setting the verbose option to 3 is useful.

In [ ]: *# your code here*

```
# cv = KFold(n_splits=5, shuffle=True, random_state=0)
AUCs_cv_rf = cross_val_score(
    pipe_rf,
    X_train,
    y_train,
    cv=cv,
    scoring='roc_auc',
    verbose=2,
    n_jobs=-1
)

print('AUC for all folds:')
print(AUCs_cv_rf)
print(f'\nMean AUC: {np.mean(AUCs_cv_rf):.3f}\n')
```

[Parallel(n\_jobs=-1)]: Using backend LokyBackend with 10 concurrent workers.

```
[CV] END ..... total time=
9.9s
[CV] END ..... total time= 1
0.1s
[CV] END ..... total time= 1
0.2s
[CV] END ..... total time= 1
0.2s
[CV] END ..... total time= 1
0.3s
```

AUC for all folds:

```
[0.87243924 0.87112891 0.87471254 0.87031427 0.87507136]
```

Mean AUC: 0.873

[Parallel(n\_jobs=-1)]: Done 5 out of 5 | elapsed: 10.4s finished

### 3.3 Estimating AUC using nested CV

Let's try to improve on the results of the random forest by tweaking its hyperparameters. Build a pipeline that computes the nested cv AUC-ROC. Include different values for "n\_estimators", "max\_samples", "max\_features", "max\_depth" and "max\_leaf\_nodes" in your hyperparameter grid. Report the inner and outer fold scores for all combinations of hyperparams and interpret the results. Report the best hyperparams used in each outer fold.

What do you notice about these results? How are these results related to those from the decision tree pipeline? Is hyperparameter optimization worth the trouble?

**Hint 1:** The pipeline remains the same. You can access the list of all the hyperparams in your pipeline with `pipeline.get_params_keys()`. Notice the naming convention.

**Hint 2:** `GridSearchCV` and `cross_val_score` are your friends. Specially if you run them with `verbose = 3`.

**Hint 3:** This gets computationally expensive quickly. Be intelligent about the hyperparam values you include in the grid.

```
In [ ]: from sklearn.model_selection import GridSearchCV
```

```
# your code here
```

```
outer_k, inner_k = 3, 5
```

```
inner_cv_rf = KFold(n_splits=inner_k, shuffle=True, random_state=0)
```

```
outer_cv_rf = KFold(n_splits=outer_k, shuffle=True, random_state=0)
```

```
#pipe_rf.get_params()['preprocessing'].get_params()
```

```
#pipe_rf.get_params()['decision_tree'].get_params()
```

```
# Parameter space to search
```

```
param_grid_rf = {
```

```
    'model__n_estimators': [50, 100], # x2
```

```
    'model__max_samples': [None, 10], # x2
```

```
    'model__max_depth': [3, 15], # x2
```

```
    'model__max_features': [25, 100], # x2
```

```
    'model__max_leaf_nodes': [100, 500] # x2
```

```
}
```

```
In [ ]: #import logging
```

```
X_fold = X_train
```

```
y_fold = y_train.values
```

```
outer_scores_rf = []
```

```
for outer_train_idx, outer_test_idx in outer_cv_rf.split(X_fold):
```

```
    X_outer_train, X_outer_test = X_fold.iloc[outer_train_idx], X_fold.iloc[
```

```
    y_outer_train, y_outer_test = y_fold[outer_train_idx], y_fold[outer_test
```

```
# Define Grid Search Object - uses all the params in `param_grid` to fit
```

```
grid_search_rf = GridSearchCV(
```

```
    estimator=pipe_rf,
```

```
    param_grid=param_grid_rf,
```

```
    cv=inner_cv,
```

```
    verbose=3,
```

```
    scoring='roc_auc',
```

```
    n_jobs=-1
```

```
)
```

```
# Fit Model across all hyper-parameters
```

```
grid_search_rf.fit(X_outer_train, y_outer_train)
```

```
inner_scores_rf = grid_search_rf.cv_results_  
outer_scores_rf.append(inner_scores_rf)
```

Fitting 5 folds for each of 32 candidates, totalling 160 fits

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.8s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.8s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.836 total time= 0.8s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.844 total time= 0.9s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.9s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.822 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.823 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.811 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.819 total time= 0.6s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.818 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.849 total time= 1.6s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.845 total time= 1.6s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.842 total time= 1.6s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.848 total time= 1.6s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.845 total time= 2.0s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.822 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.9s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.822 total time= 0.8s

es=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.829 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.824 total time= 0.9s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.844 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.8s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.9s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.9s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.819 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.811 total time= 0.5s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.823 total time= 0.5s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.822 total time= 0.5s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.836 total time= 1.0s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.818 total time= 0.5s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.849 total time= 1.7s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.845 total time= 1.6s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.822 total time= 0.9s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.848 total time= 1.6s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.842 total time= 1.6s



[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.845 total time= 1.7s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.829 total time= 1.2s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.824 total time= 0.9s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.846 total time= 1.4s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.850 total time= 1.4s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.849 total time= 1.4s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.844 total time= 1.4s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.852 total time= 1.4s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.843 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.835 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.826 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.834 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.834 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.849 total time= 2.7s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.839 total time= 0.8s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.840 total time= 0.9s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.846 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.845 total time= 0.8s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.845 total

```
time= 0.9s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.847 tota
l time= 2.9s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.851 tota
l time= 2.9s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.843 tota
l time= 2.8s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.846 total
time= 1.4s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.844 total
time= 1.4s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.850 total
time= 1.4s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.849 total
time= 1.5s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.852 total
time= 1.5s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.847 tota
l time= 4.3s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.826 total t
ime= 0.5s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.843 total t
ime= 0.5s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.834 total t
ime= 0.5s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.835 total t
ime= 0.5s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.834 total t
ime= 0.6s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.840 total
time= 0.8s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.847 tota
l time= 2.7s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.839 total
time= 0.8s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.849 tota
l time= 2.7s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
```

des=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.845 total time= 0.8s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.846 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.851 total time= 2.8s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.847 total time= 2.8s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.845 total time= 1.0s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.843 total time= 3.0s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 1.9s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 2.1s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.864 total time= 2.0s  
[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.865 total time= 2.1s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.867 total time= 2.2s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.810 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.825 total time= 0.5s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.815 total time= 0.5s  
[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.829 total time= 0.5s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.801 total time= 0.5s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.827 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.817 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.832 total time= 0.9s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.838 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.872 total time= 3.9s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.868 total time= 4.0s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.870 total time= 4.1s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.865 total time= 4.0s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.867 total time= 4.9s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.876 total time= 3.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.871 total time= 3.7s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.876 total time= 3.8s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.874 total time= 3.9s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 4.0s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.810 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.815 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.829 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.825 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.801 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.827 total

```
time= 0.9s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.832 total
time= 0.9s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.838 total
time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.817 total
time= 0.9s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.877 tota
l time= 7.5s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.875 tota
l time= 7.5s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.876 tota
l time= 7.6s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.872 tota
l time= 7.5s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=50;; score=0.874 tota
l time= 5.3s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=50;; score=0.874 tota
l time= 5.4s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=50;; score=0.876 tota
l time= 5.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=50;; score=0.870 tota
l time= 5.4s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=50;; score=0.871 tota
l time= 5.7s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=50;; score=0.841 total
time= 0.5s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=50;; score=0.827 total
time= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=50;; score=0.837 total
time= 0.5s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=50;; score=0.830 total
time= 0.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=50;; score=0.827 total
time= 0.5s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=100;; score=0.839 total
time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
```

```

odes=100, model__max_samples=10, model__n_estimators=100;; score=0.843 total
time= 0.9s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=100;; score=0.847 total
time= 0.9s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=100;; score=0.845 total
time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.871 tota
l time= 10.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=10, model__n_estimators=100;; score=0.842 total
time= 0.9s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=100;; score=0.877 tot
al time= 11.2s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=100;; score=0.874 tot
al time= 11.2s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=100;; score=0.875 tot
al time= 11.5s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=100;; score=0.871 tot
al time= 11.2s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=100, model__max_samples=None, model__n_estimators=100;; score=0.870 tot
al time= 11.1s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=50;; score=0.881 tota
l time= 13.0s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=50;; score=0.879 tota
l time= 13.1s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=50;; score=0.880 tota
l time= 12.9s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=50;; score=0.875 tota
l time= 12.8s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.841 total
time= 0.5s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.827 total
time= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.837 total
time= 0.5s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.830 total
time= 0.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.827 total
time= 0.5s

```

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.839 total  
time= 0.9s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.843 total  
time= 0.9s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.847 total  
time= 0.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.845 total  
time= 0.9s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.842 total  
time= 0.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.876 tota  
l time= 16.7s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.880 tot  
al time= 20.1s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.881 tot  
al time= 20.4s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.880 tot  
al time= 19.7s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.877 tot  
al time= 19.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.875 tot  
al time= 19.1s

Fitting 5 folds for each of 32 candidates, totalling 160 fits

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.838 total  
time= 0.8s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.844 total  
time= 0.8s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total  
time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.846 total  
time= 0.9s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.845 total  
time= 0.9s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.818 total ti  
me= 0.4s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod  
es=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.792 total ti  
me= 0.5s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nod

es=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.782 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.827 total time= 0.5s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.807 total time= 0.5s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.847 total time= 1.4s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.840 total time= 1.5s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.849 total time= 1.6s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.853 total time= 1.6s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.840 total time= 2.0s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.825 total time= 0.9s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.810 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.9s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.840 total time= 0.8s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.9s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.846 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.844 total time= 0.9s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.838 total time= 0.8s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.792 total time= 0.4s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.818 total time= 0.4s



[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.827 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.782 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.845 total time= 1.0s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.807 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.840 total time= 1.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.825 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.840 total time= 1.5s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.847 total time= 1.6s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.849 total time= 1.6s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.810 total time= 0.9s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.853 total time= 1.6s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.9s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.840 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.849 total time= 1.5s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.851 total time= 1.4s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.846 total time= 1.5s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.837 total time= 1.5s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.855 total

```
time= 1.6s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.844 total t
ime= 0.5s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.806 total t
ime= 0.5s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.810 total t
ime= 0.5s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.808 total t
ime= 0.5s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.830 total t
ime= 0.5s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.848 tota
l time= 2.8s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.833 total
time= 1.0s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.832 total
time= 0.9s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.833 total
time= 0.9s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.851 tota
l time= 2.9s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.835 total
time= 0.9s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.851 tota
l time= 2.9s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.839 total
time= 0.9s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.853 tota
l time= 2.8s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.842 tota
l time= 3.4s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.849 total
time= 1.3s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.855 total
time= 1.3s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.846 total
time= 1.4s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
```

des=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;;, score=0.837 total time= 1.4s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;;, score=0.851 total time= 1.5s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;;, score=0.844 total time= 0.5s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;;, score=0.806 total time= 0.5s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;;, score=0.808 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;;, score=0.810 total time= 0.5s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;;, score=0.830 total time= 0.5s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;;, score=0.833 total time= 0.8s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;;, score=0.848 total time= 2.6s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;;, score=0.832 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;;, score=0.839 total time= 0.8s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;;, score=0.835 total time= 0.8s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;;, score=0.833 total time= 0.9s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;;, score=0.851 total time= 2.8s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;;, score=0.851 total time= 2.8s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;;, score=0.842 total time= 2.8s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;;, score=0.853 total time= 3.2s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;;, score=0.867 total time= 2.1s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;;, score=0.870 total time= 2.1s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.869 total time= 2.0s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 2.1s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.865 total time= 2.1s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.807 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.810 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.823 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.794 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.814 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.837 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.811 total time= 0.9s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.832 total time= 0.9s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total time= 0.8s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.839 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.869 total time= 4.1s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.868 total time= 4.1s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.869 total time= 4.0s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.866 total time= 4.2s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.872 total time= 4.4s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_degree=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.875 total

```
time= 3.9s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.872 total
time= 3.9s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.875 total
time= 3.9s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.873 total
time= 3.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.876 total
time= 3.9s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.807 total t
ime= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.810 total t
ime= 0.5s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.823 total t
ime= 0.5s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.814 total t
ime= 0.5s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.794 total t
ime= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.832 total
time= 0.8s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.837 total
time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.811 total
time= 0.9s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.828 total
time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=100;; score=0.839 total
time= 0.9s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.875 tota
l time= 7.7s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.875 tota
l time= 7.7s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.874 tota
l time= 7.8s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=100;; score=0.872 tota
l time= 7.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
```

```

des=500, model__max_samples=None, model__n_estimators=100;; score=0.877 total time= 7.9s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=50;; score=0.874 total time= 5.7s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=50;; score=0.874 total time= 5.7s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=50;; score=0.874 total time= 5.6s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=50;; score=0.871 total time= 5.6s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=50;; score=0.876 total time= 5.7s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=50;; score=0.837 total time= 0.5s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=50;; score=0.809 total time= 0.5s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=50;; score=0.807 total time= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=50;; score=0.798 total time= 0.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=50;; score=0.835 total time= 0.5s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=100;; score=0.829 total time= 0.9s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=100;; score=0.826 total time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=100;; score=0.834 total time= 0.9s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=100;; score=0.835 total time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=10, model__n_estimators=100;; score=0.842 total time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=100;; score=0.874 total time= 10.8s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=100;; score=0.874 total time= 10.9s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_nodes=100, model__max_samples=None, model__n_estimators=100;; score=0.874 total time= 10.9s

```

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.871 tot  
al time= 10.8s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.876 tot  
al time= 10.6s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.880 tota  
l time= 12.4s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.878 tota  
l time= 12.4s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.877 tota  
l time= 12.4s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.880 tota  
l time= 12.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.882 tota  
l time= 12.4s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.837 total  
time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.809 total  
time= 0.5s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.798 total  
time= 0.5s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.807 total  
time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.835 total  
time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.834 total  
time= 0.9s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.829 total  
time= 0.9s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.826 total  
time= 0.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total  
time= 0.8s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.842 total  
time= 0.8s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.879 tot  
al time= 19.2s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.877 tot

```
al time= 19.0s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.880 tot
al time= 19.4s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.880 tot
al time= 19.6s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.882 tot
al time= 19.1s
Fitting 5 folds for each of 32 candidates, totalling 160 fits
[CV 3/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=50;; score=0.847 total
time= 0.8s
[CV 5/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=50;; score=0.850 total
time= 0.8s
[CV 2/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=50;; score=0.840 total
time= 0.8s
[CV 1/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=50;; score=0.841 total
time= 0.9s
[CV 4/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=50;; score=0.842 total
time= 0.9s
[CV 5/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=10, model__n_estimators=50;; score=0.811 total ti
me= 0.4s
[CV 3/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=10, model__n_estimators=50;; score=0.809 total ti
me= 0.4s
[CV 1/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=10, model__n_estimators=50;; score=0.796 total ti
me= 0.5s
[CV 2/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=10, model__n_estimators=50;; score=0.820 total ti
me= 0.5s
[CV 4/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=10, model__n_estimators=50;; score=0.798 total ti
me= 0.5s
[CV 3/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=100;; score=0.849 total
time= 1.5s
[CV 2/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=100;; score=0.845 total
time= 1.5s
[CV 5/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=100;; score=0.853 total
time= 1.5s
[CV 4/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=100;; score=0.841 total
time= 1.7s
[CV 1/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=100, model__max_samples=None, model__n_estimators=100;; score=0.841 total
time= 1.7s
```



[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.8s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.821 total time= 0.8s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.821 total time= 0.9s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.817 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.830 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.841 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.840 total time= 0.8s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.847 total time= 0.9s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.850 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.842 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.796 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.820 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.809 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.811 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.798 total time= 0.5s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.821 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.835 total time= 0.8s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.845 total time= 1.6s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=25, model\_\_max\_leaf\_nodes=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.841 total

```
time= 1.6s
[CV 4/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=None, model__n_estimators=100;; score=0.841 total
time= 1.5s
[CV 3/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=10, model__n_estimators=100;; score=0.821 total t
ime= 0.8s
[CV 3/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=None, model__n_estimators=100;; score=0.849 total
time= 1.6s
[CV 5/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=None, model__n_estimators=100;; score=0.853 total
time= 1.6s
[CV 5/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=10, model__n_estimators=100;; score=0.817 total t
ime= 0.8s
[CV 4/5] END model__max_depth=3, model__max_features=25, model__max_leaf_nod
es=500, model__max_samples=10, model__n_estimators=100;; score=0.830 total t
ime= 0.9s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=50;; score=0.848 total
time= 1.4s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=50;; score=0.857 total
time= 1.5s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=50;; score=0.842 total
time= 1.4s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=50;; score=0.851 total
time= 1.5s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=50;; score=0.853 total
time= 1.5s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.827 total t
ime= 0.5s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.772 total t
ime= 0.5s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.822 total t
ime= 0.5s
[CV 4/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.803 total t
ime= 0.5s
[CV 5/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=50;; score=0.818 total t
ime= 0.5s
[CV 3/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.850 tota
l time= 2.6s
[CV 1/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.834 total
time= 0.9s
[CV 2/5] END model__max_depth=3, model__max_features=100, model__max_leaf_no
```

des=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.856 total time= 2.8s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.820 total time= 0.8s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.829 total time= 0.9s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.847 total time= 2.8s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.848 total time= 2.7s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.819 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.853 total time= 2.8s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.827 total time= 0.9s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.848 total time= 1.4s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.857 total time= 1.4s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.842 total time= 1.4s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.853 total time= 1.4s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.851 total time= 1.5s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.827 total time= 0.5s  
[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.772 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.822 total time= 0.5s  
[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.803 total time= 0.5s  
[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.818 total time= 0.5s  
[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.834 total time= 0.9s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.829 total time= 0.9s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.820 total time= 0.8s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.848 total time= 2.7s

[CV 4/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.819 total time= 0.9s

[CV 1/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.847 total time= 2.8s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.853 total time= 2.8s

[CV 5/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.827 total time= 0.9s

[CV 2/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.856 total time= 2.9s

[CV 3/5] END model\_\_max\_depth=3, model\_\_max\_features=100, model\_\_max\_leaf\_node\_size=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.850 total time= 3.0s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.865 total time= 2.1s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.871 total time= 2.1s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.867 total time= 2.1s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.863 total time= 2.1s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 2.2s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.799 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.817 total time= 0.5s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.803 total time= 0.5s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.804 total time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node\_size=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.816 total time= 0.5s

```
ime= 0.5s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.823 total
time= 0.8s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.836 total
time= 0.9s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.821 total
time= 0.9s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.837 total
time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=10, model__n_estimators=100;; score=0.826 total
time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.869 tota
l time= 4.1s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.866 tota
l time= 4.1s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.872 tota
l time= 4.1s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.863 tota
l time= 4.1s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=100, model__max_samples=None, model__n_estimators=100;; score=0.871 tota
l time= 4.1s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.871 total
time= 3.8s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.874 total
time= 3.7s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.877 total
time= 3.7s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.870 total
time= 3.8s
[CV 5/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=None, model__n_estimators=50;; score=0.875 total
time= 3.8s
[CV 1/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.799 total t
ime= 0.4s
[CV 2/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.817 total t
ime= 0.4s
[CV 3/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
des=500, model__max_samples=10, model__n_estimators=50;; score=0.803 total t
ime= 0.4s
[CV 4/5] END model__max_depth=15, model__max_features=25, model__max_leaf_no
```

des=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.804 total time= 0.4s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.816 total time= 0.5s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.823 total time= 0.9s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.836 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.821 total time= 0.9s  
[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.837 total time= 0.9s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.826 total time= 0.9s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.877 total time= 7.4s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.872 total time= 7.5s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.875 total time= 7.5s  
[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.870 total time= 7.6s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=25, model\_\_max\_leaf\_node=500, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.876 total time= 7.6s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.872 total time= 5.5s  
[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.875 total time= 5.5s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.877 total time= 5.5s  
[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.870 total time= 5.4s  
[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.874 total time= 5.4s  
[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.827 total time= 0.5s  
[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_node=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.821 total time= 0.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.762 total  
time= 0.5s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.806 total  
time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.815 total  
time= 0.5s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.836 total  
time= 0.8s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.819 total  
time= 0.9s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.824 total  
time= 0.9s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.823 total  
time= 0.9s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=10, model\_\_n\_estimators=100;; score=0.828 total  
time= 0.9s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.872 tot  
al time= 10.5s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.875 tot  
al time= 10.7s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.877 tot  
al time= 10.8s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.870 tot  
al time= 10.8s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=100, model\_\_max\_samples=None, model\_\_n\_estimators=100;; score=0.875 tot  
al time= 10.8s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.877 tota  
l time= 12.3s

[CV 3/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.882 tota  
l time= 12.3s

[CV 2/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.880 tota  
l time= 12.4s

[CV 4/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.875 tota  
l time= 12.4s

[CV 1/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=10, model\_\_n\_estimators=50;; score=0.827 total  
time= 0.5s

[CV 5/5] END model\_\_max\_depth=15, model\_\_max\_features=100, model\_\_max\_leaf\_n  
odes=500, model\_\_max\_samples=None, model\_\_n\_estimators=50;; score=0.879 tota

```

l time= 12.5s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.762 total
time= 0.5s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.821 total
time= 0.5s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.806 total
time= 0.5s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=50;; score=0.815 total
time= 0.5s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.836 total
time= 0.8s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.819 total
time= 0.8s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.824 total
time= 0.9s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.828 total
time= 0.8s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=10, model__n_estimators=100;; score=0.823 total
time= 0.9s
[CV 2/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.880 tot
al time= 19.2s
[CV 1/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.878 tot
al time= 19.5s
[CV 3/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.882 tot
al time= 19.3s
[CV 5/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.880 tot
al time= 19.2s
[CV 4/5] END model__max_depth=15, model__max_features=100, model__max_leaf_n
odes=500, model__max_samples=None, model__n_estimators=100;; score=0.876 tot
al time= 19.7s

```

```

In [ ]: # Print all
        for i in range(0, outer_k):
            for mean_score, params in zip(outer_scores_rf[i]['mean_test_score'], out
                print(f"outer_fold: {i} -- Mean CV Score: {mean_score:.3f} with Para

```



outer\_fold: 0 -- Mean CV Score: 0.840 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.846 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.819 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.828 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.840 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.846 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.819 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.828 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.848 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.847 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.834 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.843 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.848 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.847 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.834 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.843 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.868 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 0 -- Mean CV Score: 0.868 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 0 -- Mean CV Score: 0.816 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

```

ples': 10, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.873 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.874 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.816 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.873 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': None, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.873 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': None, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.833 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': 10, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.843 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': 10, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.878 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sa
mples': None, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.879 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sa
mples': None, 'model__n_estimators': 100}
outer_fold: 0 -- Mean CV Score: 0.833 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sa
mples': 10, 'model__n_estimators': 50}
outer_fold: 0 -- Mean CV Score: 0.843 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sa
mples': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.842 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_samp
les': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.846 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_samp
les': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.805 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_samp
les': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_samp
les': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.842 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_samp
les': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.846 with Parameters: {'model__max_depth':

```

```
3, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_samp
les': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.805 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_samp
les': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
3, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_samp
les': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.847 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.849 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.820 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.834 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.847 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.849 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.820 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.834 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.868 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.869 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.810 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.829 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.874 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.875 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 1 -- Mean CV Score: 0.810 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 1 -- Mean CV Score: 0.829 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
```

outer\_fold: 1 -- Mean CV Score: 0.874 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 1 -- Mean CV Score: 0.874 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 1 -- Mean CV Score: 0.817 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 1 -- Mean CV Score: 0.833 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 1 -- Mean CV Score: 0.879 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 1 -- Mean CV Score: 0.879 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 1 -- Mean CV Score: 0.817 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 1 -- Mean CV Score: 0.833 with Parameters: {'model\_\_max\_depth': 15, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.844 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 2 -- Mean CV Score: 0.846 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.807 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 2 -- Mean CV Score: 0.825 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.844 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 2 -- Mean CV Score: 0.846 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.807 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 50}

outer\_fold: 2 -- Mean CV Score: 0.825 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 25, 'model\_\_max\_leaf\_nodes': 500, 'model\_\_max\_samples': 10, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.850 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

outer\_fold: 2 -- Mean CV Score: 0.851 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 100}

outer\_fold: 2 -- Mean CV Score: 0.809 with Parameters: {'model\_\_max\_depth': 3, 'model\_\_max\_features': 100, 'model\_\_max\_leaf\_nodes': 100, 'model\_\_max\_samples': None, 'model\_\_n\_estimators': 50}

```
ples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.826 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.850 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.851 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.809 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.826 with Parameters: {'model__max_depth':
3, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.867 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.868 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.808 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 100, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.873 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.874 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': None, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.808 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.828 with Parameters: {'model__max_depth':
15, 'model__max_features': 25, 'model__max_leaf_nodes': 500, 'model__max_sam
ples': 10, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.874 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': None, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.874 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': None, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.806 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.826 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 100, 'model__max_sa
mples': 10, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.879 with Parameters: {'model__max_depth':
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_sa
mples': None, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.879 with Parameters: {'model__max_depth':
```

```
15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_samples': None, 'model__n_estimators': 100}
outer_fold: 2 -- Mean CV Score: 0.806 with Parameters: {'model__max_depth': 15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_samples': 10, 'model__n_estimators': 50}
outer_fold: 2 -- Mean CV Score: 0.826 with Parameters: {'model__max_depth': 15, 'model__max_features': 100, 'model__max_leaf_nodes': 500, 'model__max_samples': 10, 'model__n_estimators': 100}
```

```
In [ ]: grid_search_rf.get_params().keys()

print('All scores:')
#print(np.array(outer_scores_rf))
#print(f'\nMean AUC: {np.mean(outer_scores_rf):.3f}')

outer_fold_print = [np.mean(i['mean_test_score']) for i in outer_scores_rf]
print(outer_fold_print)
print(f'total AUC: {np.mean(outer_fold_print):.3f}')
```

```
All scores:
[0.8448987252076575, 0.841040250950414, 0.838686462617982]
total AUC: 0.842
```

*your answer here*

AUC of .842 is worse than the previous iteration, which implies that these grid search parameters would probably not be the most fruitful place to continue searching to optimize the model

### 3.4 Take stock of parts II and III

What do you conclude from parts II and III? If you were to choose between a decision tree and a random forest for this classification task, what would be your choice and why?

*your answer here*

It appears the Random Forest algorithm slightly better (measuring on AUC) than the Basic Decision Tree. They are not too close (.842 to .746). As such I probably choose the Random Forest algorithm in this instance, however, if performance were a huge factor there could be an argument to go with the DecisionTree because it is much less computationally expensive.

This result is unsurprising but the true difference could be actually pretty different than what we see here

- The data is significantly underfit or overfit. We did not search a large segment of the feature-space to ensure there were no better options available for hyperparameters. As such, the trees we are building may have significant gaps in learning.
- Finally, AUC is not a perfect method of scoring a classifier. In fact its probably not an ideal way to measure one in most real-life use-cases. As such, while this one model

may perform better on AUC, it might not be the best on other metrics that could align better with real-world use-cases

## Part IV: Test data, ROC curves and fairness thresholds

Note that until now, we have not used the test data for any purpose. Let's now incorporate it into our analysis and use it to find a "fair" threshold value.

### 4.1 Find the optimal hyperparameters and fit a decision tree pipeline

Use all of your training data to find the optimal hyperparameters and fit a decision tree pipeline. Report the optimal hyperparameters.

**Note:** Recall that nested cross-validation is only providing you with an estimate of the out-of-sample performance of the model finding procedure that involves hyperparameter optimization. At training time you should carry out the entire model finding procedure, including hyperparameter optimization.

```
In [ ]: # your code here
        grid_search_rf.fit(X_train, y_train)

        y_hat_test_dt = grid_search_rf.predict(X_test)

        y_hat_test_dt
```

```
Out [ ]: array([ True,  True,  True, ...,  True, False, False])
```

### 4.2 Scores and the ROC curve

Use the fitted pipeline to compute score values for all instances in the test set. Show the distribution of the scores in a histogram. Additionally, compute the (FPR,TPR) combinations for all relevant threshold values and use them to plot the ROC curve in a different figure. Following the convention, include a dashed line along the diagonal. Remember to label the axes and to make the figures as polished as possible.

**Note 1:** You should NOT be using `sklearn.metrics.auc` or `sklearn.metrics.RocCurveDisplay` to calculate or display the ROC curve. Please code this part yourself.

```
In [ ]: # your code here
```

### 4.3 ROC curves by racial groups

Compute and plot in the same figure the ROC curves for the following racial groups identified by the ACS:

- White (RAC1P==1) (plot in blue)
- African American (RAC1P==2) (plot in orange)
- Asian American (RAC1P==6) (plot in green)

Interpret the results.

```
In [ ]: # your code here
```

*your observations here*

## 4.4 Achieving error parity

Implement a thresholding strategy that satisfies error parity for all racial groups with FPR = 0.3, TPR = 0.6 and  $\epsilon = 0.025$ . In plain english, find a way of setting thresholds for the members of each group in the test data that, when evaluated on the test data, delivers FPR and TPR values that differ at most from the objective values by  $\epsilon$ . Plot the estimated TPRs and FPRs of the racial groups in the ROC plot. Use star markers, colored accordingly.

**Hint:** Consider using group-specific stochastic thresholds.

```
In [ ]: # your code here
```

## 4.5 Improving the results [extra-credit]

Can you improve the results from 4.4? That is, can you tweak your algorithm to deliver a higher TPR and lower FPR while still satisfying error parity with epsilon = 0.025? What is the best result that you are able to achieve? Is there a hard limit on how much you can improve?

```
In [ ]: # your code here
```

*your observations here*

## 4.6 Conclusion

What can you say about welfare and error parity? Some things to consider:

- In this income prediction task, is enforcing error parity costly?
- Would you expect this results to generalize to other predictions problems?
- Is the group definition relevant?
- Who benefits from enforcing error parity? Who doesn't?



*your observations here*