04_model_development

August 21, 2025

0.1 Hyperparameter Tuning with XGBoost

In this section, we will focus on improving the performance of our predictive model by tuning the hyperparameters of **XGBoost**.

Hyperparameter tuning is a crucial step in model development, as it allows us to control the learning process, optimize bias–variance tradeoff, and achieve better generalization on unseen data. We will explore different configurations using systematic search strategies (grid search, randomized search, or Bayesian optimization) and evaluate the results based on metrics such as accuracy, ROC-AUC, and precision-recall AUC.

The goal is to identify the best combination of parameters that enhances model performance and ensures robust predictions.

```python
[128]: # Standard Library
       import logging

       # Third-Party Libraries
       import pandas as pd
       import numpy as np

       # Hyperparameter Tuning
       import optuna
       # For visualizations
       from optuna.visualization import plot_optimization_history,␣
        ↪plot_parallel_coordinate, plot_slice, plot_contour, plot_param_importances

       # Model Development & Evaluation
       from xgboost import XGBClassifier
       from sklearn.model_selection import cross_val_score
       from sklearn.metrics import accuracy_score
       from sklearn.model_selection import GridSearchCV

       # --- Get curve data + plots for ROC and Precision-Recall ---
       from sklearn.metrics import roc_curve, roc_auc_score, precision_recall_curve,␣
        ↪auc
       import matplotlib.pyplot as plt
```

```
[2]: # Configure logging
     logging.basicConfig(
         level=logging.INFO,
         format="%(asctime)s | %(levelname)s | %(message)s"
     )
```

```
[3]: df = pd.read_csv('final_feature_eng.csv')
     df
```

```
[3]:       federal_action_obligation   total_dollars_obligated  \
     0                     23116.94                  23116.94
     1                     84000.00                 284319.29
     2                        66.00                     66.00
     3                      1216.40                   1216.40
     4                       259.70                    259.70
     ...                        ...                       ...
     2627                  22973.12                 119973.12
     2628                   1008.64                   1008.64
     2629                    374.70                    374.70
     2630                    289.34                    289.34
     2631                     26.62                     26.62

           current_total_value_of_award   potential_total_value_of_award  \
     0                         23116.94                         23116.94
     1                        284319.29                        599885.84
     2                            66.00                            66.00
     3                          1216.40                          1216.40
     4                           259.70                           259.70
     ...                            ...                              ...
     2627                     119973.12                        119973.12
     2628                       1008.64                          1008.64
     2629                        374.70                           374.70
     2630                        289.34                           289.34
     2631                         26.62                            26.62

           action_date_fiscal_year   funding_agency_code   award_type  \
     0                        2022                    13            1
     1                        2022                    36            1
     2                        2022                    15            1
     3                        2022                    97            0
     4                        2022                    97            0
     ...                       ...                   ...          ...
     2627                     2023                    97            1
     2628                     2022                    15            0
     2629                     2022                    15            0
     2630                     2022                    97            0
     2631                     2022                    15            0
```

```
     type_of_contract_pricing  extent_competed  \
0                           0                1
1                           0                1
2                           0                3
3                           1                0
4                           1                0
...                       ...              ...
2627                        0                4
2628                        0                0
2629                        0                0
2630                        1                0
2631                        0                0

      government_furnished_property  ...        revt        ib         lt  \
0                                 0  ...  39211.000  7725.000  54146.000
1                                 0  ...     40.697   -54.454    125.427
2                                 0  ...    583.187     7.729    400.966
3                                 0  ...  12401.021   631.232   3804.587
4                                 0  ...  12401.021   631.232   3804.587
...                             ...  ...        ...       ...        ...
2627                              0  ...     26.074     0.481      9.036
2628                              0  ...    583.187     7.729    400.966
2629                              0  ...    583.187     7.729    400.966
2630                              0  ...  12401.021   631.232   3804.587
2631                              0  ...    583.187     7.729    400.966

            ceq      oancf       xrd       cogs  psc_3digit_freq  lapse_flag  \
0     40793.000   9312.000  1406.000  18171.000         120352.0           1
1        52.413    -48.746    85.641    115.855         635345.0           1
2       667.099    -66.537    29.307    118.470         635345.0           0
3      3425.126    709.580     0.000   8534.570         120352.0           0
4      3425.126    709.580     0.000   8534.570         120352.0           0
...         ...        ...       ...        ...              ...         ...
2627     20.020      0.250     1.828     16.821            459.0           1
2628    667.099    -66.537    29.307    118.470         635345.0           0
2629    667.099    -66.537    29.307    118.470         635345.0           0
2630   3425.126    709.580     0.000   8534.570         463837.0           0
2631    667.099    -66.537    29.307    118.470         635345.0           0

      ThemeCodeAlpha
0                  1
1                  6
2                  6
3                  2
4                  2
...              ...
```

```
2627              4
2628              3
2629              6
2630              1
2631              6

[2632 rows x 34 columns]
```

Feature Scaling of Financial Variables   The financial features in our dataset come from two different sources:
- **Government contract data** (e.g., federal obligations, award values)
- **Compustat firm data** (e.g., assets, sales, income, expenses)

These values can be recorded in **different magnitudes**.

However, since our modeling approach is based on **XGBoost**, feature scaling is **not required**.

Tree-based algorithms are insensitive to differences in units, as they rely on threshold-based splits rather than the absolute magnitude of values.

0.1.1   Train/Test data distribution

```python
[9]:   # Train = 2022 + 2023
       train_df = df[df["action_date_fiscal_year"].isin([2022, 2023])].copy()

       # Test = 2024
       test_df = df[df["action_date_fiscal_year"] == 2024].copy()

       # Drop date column
       train_df = train_df.drop(columns="action_date_fiscal_year", errors="ignore")
       test_df = test_df.drop(columns="action_date_fiscal_year", errors="ignore")
```

```python
[10]:  # Separate features and target
       X_train = train_df.drop(columns=["lapse_flag"])
       y_train = train_df["lapse_flag"]

       X_test  = test_df.drop(columns=["lapse_flag"])
       y_test  = test_df["lapse_flag"]
```

0.1.2   Hyperparameter Optimization with Optuna

We will use **Optuna**, a modern hyperparameter optimization framework.

We define an **objective function** that Optuna will repeatedly evaluate with different hyperparameter configurations. The goal is to maximize the model's validation performance (accuracy in this case).

Hyperparameters to Tune   Below are the XGBoost hyperparameters we will optimize:

- **n_estimators**: Number of boosting rounds (trees to build).

- **max_depth**: Maximum depth of a tree (higher depth = more complex model).

- **learning_rate (eta)**: Shrinks the contribution of each tree to prevent overfitting.

- **subsample**: Fraction of training data sampled for each tree (controls overfitting).

- **colsample_bytree**: Fraction of features sampled for each tree (adds randomness).

- **gamma (min_split_loss)**: Minimum loss reduction required to make a split.

- **min_child_weight**: Minimum sum of instance weights needed in a leaf node.

- **reg_alpha (L1 regularization)**: Adds penalty to weights (encourages sparsity).

- **reg_lambda (L2 regularization)**: Adds penalty to weights (controls overfitting).

0.1.3 Randomized Search

```python
# Objective function
def objective(trial):
    # Hyperparameters to tune
    n_estimators = trial.suggest_int('n_estimators', 100, 200)
    max_depth = trial.suggest_int('max_depth', 10, 15)
    learning_rate = trial.suggest_float('learning_rate', 0.05, 0.15, log=True)
    subsample = trial.suggest_float('subsample', 0.7, 0.9)
    colsample_bytree = trial.suggest_float('colsample_bytree', 0.5, 0.7)
    gamma = trial.suggest_float('gamma', 0, 1)
    min_child_weight = trial.suggest_int('min_child_weight', 1, 4)
    reg_alpha = trial.suggest_float('reg_alpha', 1, 3)
    reg_lambda = trial.suggest_float('reg_lambda', 2, 5)

    # Build model
    model = XGBClassifier(
        n_estimators=n_estimators,
        max_depth=max_depth,
        learning_rate=learning_rate,
        subsample=subsample,
        colsample_bytree=colsample_bytree,
        gamma=gamma,
        min_child_weight=min_child_weight,
        reg_alpha=reg_alpha,
        reg_lambda=reg_lambda,
        eval_metric="logloss",
        random_state=42
    )
```

```python
    # Perform 5-fold cross-validation
    score = cross_val_score(
        model, X_train, y_train, cv=5, scoring='accuracy'
    ).mean()

    return score
```

[65]:
```python
study = optuna.create_study(direction='maximize', sampler=optuna.samplers.
  ↪RandomSampler())  # We aim to maximize accuracy
study.optimize(objective, n_trials=100)  # Run 100 trials to find the best␣
  ↪hyperparameters
```

```
[I 2025-08-21 20:33:46,086] A new study created in memory with name: no-
name-46162c1b-48c3-4846-93c4-5be2abf399ed
[I 2025-08-21 20:33:46,717] Trial 0 finished with value: 0.9494407734512137 and
parameters: {'n_estimators': 128, 'max_depth': 12, 'learning_rate':
0.06832906450468945, 'subsample': 0.7061206524339927, 'colsample_bytree':
0.5999429909469579, 'gamma': 0.9985823996130839, 'min_child_weight': 1,
'reg_alpha': 1.6608641818655723, 'reg_lambda': 2.190545813363128}. Best is trial
0 with value: 0.9494407734512137.
[I 2025-08-21 20:33:47,636] Trial 1 finished with value: 0.9498690389758389 and
parameters: {'n_estimators': 138, 'max_depth': 13, 'learning_rate':
0.06336033448756266, 'subsample': 0.8860032253902534, 'colsample_bytree':
0.6626171875554742, 'gamma': 0.29516753198931645, 'min_child_weight': 1,
'reg_alpha': 1.336584272800695, 'reg_lambda': 4.155173743999046}. Best is trial
1 with value: 0.9498690389758389.
[I 2025-08-21 20:33:48,742] Trial 2 finished with value: 0.9520122046484272 and
parameters: {'n_estimators': 143, 'max_depth': 13, 'learning_rate':
0.122735540147556, 'subsample': 0.8835212290328968, 'colsample_bytree':
0.550234562558559, 'gamma': 0.0018382926311870662, 'min_child_weight': 1,
'reg_alpha': 1.6143030673686416, 'reg_lambda': 2.7033008980487963}. Best is
trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:49,549] Trial 3 finished with value: 0.9485833233772322 and
parameters: {'n_estimators': 148, 'max_depth': 13, 'learning_rate':
0.05056449610684099, 'subsample': 0.8738099079456719, 'colsample_bytree':
0.6830251292918592, 'gamma': 0.5191010661345722, 'min_child_weight': 1,
'reg_alpha': 2.875207681111255, 'reg_lambda': 4.145305919404561}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:50,181] Trial 4 finished with value: 0.9481541388278758 and
parameters: {'n_estimators': 180, 'max_depth': 14, 'learning_rate':
0.08767028701161299, 'subsample': 0.7794944897328948, 'colsample_bytree':
0.5958958569472382, 'gamma': 0.7332479070456567, 'min_child_weight': 2,
'reg_alpha': 2.489049460956166, 'reg_lambda': 4.574138076604402}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:50,659] Trial 5 finished with value: 0.9481541388278758 and
parameters: {'n_estimators': 107, 'max_depth': 12, 'learning_rate':
0.08012052892387854, 'subsample': 0.7360626709610478, 'colsample_bytree':
```

0.6682847246768854, 'gamma': 0.7123552169331484, 'min_child_weight': 3, 'reg_alpha': 2.736509456905183, 'reg_lambda': 2.352716591160345}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:51,389] Trial 6 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 161, 'max_depth': 10, 'learning_rate': 0.14780438176653946, 'subsample': 0.8220903909280908, 'colsample_bytree': 0.5957210140470248, 'gamma': 0.14278770577821343, 'min_child_weight': 4, 'reg_alpha': 2.2035714667050774, 'reg_lambda': 3.1075546812127834}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:51,832] Trial 7 finished with value: 0.9477249542785197 and parameters: {'n_estimators': 129, 'max_depth': 12, 'learning_rate': 0.11571880793652316, 'subsample': 0.8012314670339244, 'colsample_bytree': 0.6261751795590353, 'gamma': 0.7965044258368278, 'min_child_weight': 1, 'reg_alpha': 2.897237136268241, 'reg_lambda': 3.3433351197539616}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:52,292] Trial 8 finished with value: 0.9485824043525012 and parameters: {'n_estimators': 138, 'max_depth': 11, 'learning_rate': 0.11802403637438313, 'subsample': 0.8892239612735231, 'colsample_bytree': 0.518460507393397, 'gamma': 0.9485116609457169, 'min_child_weight': 4, 'reg_alpha': 1.718108115540682, 'reg_lambda': 2.843675152088348}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:53,336] Trial 9 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 193, 'max_depth': 14, 'learning_rate': 0.12055171827612668, 'subsample': 0.8064464189772091, 'colsample_bytree': 0.5227213119222114, 'gamma': 0.05110427746882884, 'min_child_weight': 2, 'reg_alpha': 2.3788999744263606, 'reg_lambda': 2.0368961619405166}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:53,932] Trial 10 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 159, 'max_depth': 15, 'learning_rate': 0.10761042255013732, 'subsample': 0.7373016733476079, 'colsample_bytree': 0.5059773529383489, 'gamma': 0.5501196974836168, 'min_child_weight': 2, 'reg_alpha': 2.485743130657636, 'reg_lambda': 3.375427067846054}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:54,383] Trial 11 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 153, 'max_depth': 13, 'learning_rate': 0.1282167871692459, 'subsample': 0.8065202268274576, 'colsample_bytree': 0.5940832408349463, 'gamma': 0.8067350600011226, 'min_child_weight': 3, 'reg_alpha': 2.68646735099513, 'reg_lambda': 2.9919085561676404}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:55,185] Trial 12 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 166, 'max_depth': 12, 'learning_rate': 0.05371151280907974, 'subsample': 0.7275890799778244, 'colsample_bytree': 0.5844514849670484, 'gamma': 0.3866432575559482, 'min_child_weight': 2, 'reg_alpha': 2.9954516666194166, 'reg_lambda': 3.5144614073426403}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:56,120] Trial 13 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 156, 'max_depth': 12, 'learning_rate': 0.05712955301664093, 'subsample': 0.7398290283533616, 'colsample_bytree':

0.6892221791272362, 'gamma': 0.3806105394424142, 'min_child_weight': 2, 'reg_alpha': 1.1648184244522328, 'reg_lambda': 4.18634337381225}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:56,746] Trial 14 finished with value: 0.9472966887538945 and parameters: {'n_estimators': 104, 'max_depth': 11, 'learning_rate': 0.05044329672925113, 'subsample': 0.7268834085739163, 'colsample_bytree': 0.5532566357760328, 'gamma': 0.04574081296560284, 'min_child_weight': 2, 'reg_alpha': 2.783153525547653, 'reg_lambda': 4.499625004284014}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:57,388] Trial 15 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 178, 'max_depth': 13, 'learning_rate': 0.07129777389314314, 'subsample': 0.7309409438719019, 'colsample_bytree': 0.5876384653985302, 'gamma': 0.9330898993344806, 'min_child_weight': 3, 'reg_alpha': 2.477041816596858, 'reg_lambda': 3.005931809995615}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:58,103] Trial 16 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 150, 'max_depth': 15, 'learning_rate': 0.10861283879175747, 'subsample': 0.7424976221295699, 'colsample_bytree': 0.5400741585462115, 'gamma': 0.4427895258321336, 'min_child_weight': 3, 'reg_alpha': 1.2482570846765202, 'reg_lambda': 4.114463808548271}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:58,687] Trial 17 finished with value: 0.9494398544264827 and parameters: {'n_estimators': 185, 'max_depth': 15, 'learning_rate': 0.14406341710296353, 'subsample': 0.8306946756532456, 'colsample_bytree': 0.5213342217752562, 'gamma': 0.6898485315337916, 'min_child_weight': 1, 'reg_alpha': 1.5978173902303359, 'reg_lambda': 2.179500834284385}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:59,186] Trial 18 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 103, 'max_depth': 15, 'learning_rate': 0.0905576536726716, 'subsample': 0.7788835491399162, 'colsample_bytree': 0.5569953337018285, 'gamma': 0.6536825611728443, 'min_child_weight': 3, 'reg_alpha': 1.811619547991131, 'reg_lambda': 2.249029574145199}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:33:59,801] Trial 19 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 160, 'max_depth': 12, 'learning_rate': 0.14579441638864926, 'subsample': 0.832449490913919, 'colsample_bytree': 0.60425445424366, 'gamma': 0.5420477702452771, 'min_child_weight': 1, 'reg_alpha': 1.2248431363540657, 'reg_lambda': 2.937173961023092}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:00,474] Trial 20 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 132, 'max_depth': 15, 'learning_rate': 0.05373357643152068, 'subsample': 0.7167715154997207, 'colsample_bytree': 0.5147619948843208, 'gamma': 0.7825561598930285, 'min_child_weight': 3, 'reg_alpha': 1.2704634257884007, 'reg_lambda': 3.9459373066243546}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:01,048] Trial 21 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 103, 'max_depth': 14, 'learning_rate': 0.07016960003716208, 'subsample': 0.7426951029071387, 'colsample_bytree':

0.57081012279659255, 'gamma': 0.12336951927708051, 'min_child_weight': 3, 'reg_alpha': 2.6307951820905644, 'reg_lambda': 2.5099457814079003}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:01,646] Trial 22 finished with value: 0.9494398544264827 and parameters: {'n_estimators': 104, 'max_depth': 10, 'learning_rate': 0.10616014002235166, 'subsample': 0.7111591971519716, 'colsample_bytree': 0.5159779592757994, 'gamma': 0.19351622768812815, 'min_child_weight': 2, 'reg_alpha': 1.7638189096750867, 'reg_lambda': 3.81974457372355}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:02,299] Trial 23 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 136, 'max_depth': 12, 'learning_rate': 0.05488968639116548, 'subsample': 0.8177217041110073, 'colsample_bytree': 0.6081743839056412, 'gamma': 0.9452082207686887, 'min_child_weight': 1, 'reg_alpha': 2.330144695350154, 'reg_lambda': 4.596083868314272}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:02,939] Trial 24 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 132, 'max_depth': 13, 'learning_rate': 0.05853227071015496, 'subsample': 0.8099704333362923, 'colsample_bytree': 0.6454277033758098, 'gamma': 0.6643372828590296, 'min_child_weight': 2, 'reg_alpha': 2.700874678453764, 'reg_lambda': 2.34553310357943}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:03,667] Trial 25 finished with value: 0.9472985268033562 and parameters: {'n_estimators': 194, 'max_depth': 11, 'learning_rate': 0.07881098039976742, 'subsample': 0.8455046405322325, 'colsample_bytree': 0.5206065249675522, 'gamma': 0.8490242847504804, 'min_child_weight': 4, 'reg_alpha': 1.072901192705319, 'reg_lambda': 4.887393747412958}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:04,430] Trial 26 finished with value: 0.9502973045004641 and parameters: {'n_estimators': 146, 'max_depth': 11, 'learning_rate': 0.11358355472357472, 'subsample': 0.800649185114184, 'colsample_bytree': 0.6562820895998851, 'gamma': 0.29677881232263736, 'min_child_weight': 2, 'reg_alpha': 1.5859852865333528, 'reg_lambda': 3.8592693067062567}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:05,100] Trial 27 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 172, 'max_depth': 14, 'learning_rate': 0.10284372189272858, 'subsample': 0.7289647756190197, 'colsample_bytree': 0.5609546953164803, 'gamma': 0.762488605224028, 'min_child_weight': 2, 'reg_alpha': 1.3781712562587, 'reg_lambda': 3.7201381901698696}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:05,751] Trial 28 finished with value: 0.9485824043525012 and parameters: {'n_estimators': 158, 'max_depth': 15, 'learning_rate': 0.06366659887458932, 'subsample': 0.8342181017774297, 'colsample_bytree': 0.5844794244018023, 'gamma': 0.8103992225739837, 'min_child_weight': 1, 'reg_alpha': 2.8712058289253997, 'reg_lambda': 3.3944175859198755}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:06,292] Trial 29 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 131, 'max_depth': 10, 'learning_rate': 0.09559616357558047, 'subsample': 0.821816650598372, 'colsample_bytree':

0.691542582616287, 'gamma': 0.4754228352384996, 'min_child_weight': 4, 'reg_alpha': 2.801945453923362, 'reg_lambda': 2.5999888898887153}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:06,700] Trial 30 finished with value: 0.9468693422540001 and parameters: {'n_estimators': 116, 'max_depth': 12, 'learning_rate': 0.10212316356823158, 'subsample': 0.7266402411676778, 'colsample_bytree': 0.697477662843466, 'gamma': 0.9306835858928314, 'min_child_weight': 4, 'reg_alpha': 2.7351750713232468, 'reg_lambda': 3.7462147416007365}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:07,598] Trial 31 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 178, 'max_depth': 10, 'learning_rate': 0.08013205113447822, 'subsample': 0.8507208331981568, 'colsample_bytree': 0.5377641416993975, 'gamma': 0.48793293416544636, 'min_child_weight': 2, 'reg_alpha': 1.0241182946486764, 'reg_lambda': 4.9419740260267195}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:08,050] Trial 32 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 109, 'max_depth': 13, 'learning_rate': 0.078740093920670831, 'subsample': 0.7054136468969521, 'colsample_bytree': 0.5252477692578931, 'gamma': 0.9969579619896364, 'min_child_weight': 3, 'reg_alpha': 2.361212325035581, 'reg_lambda': 4.620766097703648}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:08,794] Trial 33 finished with value: 0.9502973045004641 and parameters: {'n_estimators': 119, 'max_depth': 11, 'learning_rate': 0.12222049578006333, 'subsample': 0.8063830549120995, 'colsample_bytree': 0.5519430056457937, 'gamma': 0.011341013399899125, 'min_child_weight': 2, 'reg_alpha': 2.524687607306168, 'reg_lambda': 4.758434005033132}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:09,401] Trial 34 finished with value: 0.9485824043525012 and parameters: {'n_estimators': 121, 'max_depth': 10, 'learning_rate': 0.06311208137420518, 'subsample': 0.7062221854902607, 'colsample_bytree': 0.6004256198196799, 'gamma': 0.7614799030251872, 'min_child_weight': 1, 'reg_alpha': 2.624093657986749, 'reg_lambda': 2.3015514482985764}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:10,037] Trial 35 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 185, 'max_depth': 10, 'learning_rate': 0.09860933534701831, 'subsample': 0.7274025656844232, 'colsample_bytree': 0.6585277172711097, 'gamma': 0.9282161929254932, 'min_child_weight': 2, 'reg_alpha': 1.6591658078985136, 'reg_lambda': 2.262056621679057}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:10,873] Trial 36 finished with value: 0.9494407734512137 and parameters: {'n_estimators': 128, 'max_depth': 15, 'learning_rate': 0.0514480962906697, 'subsample': 0.8138948118742523, 'colsample_bytree': 0.6425383555213222, 'gamma': 0.08680044766640227, 'min_child_weight': 3, 'reg_alpha': 1.0925975629550575, 'reg_lambda': 4.832898583284797}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:11,539] Trial 37 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 171, 'max_depth': 14, 'learning_rate': 0.07531008401812536, 'subsample': 0.8992509195867069, 'colsample_bytree':

0.6062966153315427, 'gamma': 0.6949421822699129, 'min_child_weight': 3, 'reg_alpha': 2.0881462359765637, 'reg_lambda': 4.55102875965798}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:11,994] Trial 38 finished with value: 0.9472966887538945 and parameters: {'n_estimators': 103, 'max_depth': 13, 'learning_rate': 0.08662977482667984, 'subsample': 0.8844839229185624, 'colsample_bytree': 0.592149420013988, 'gamma': 0.7243263047335633, 'min_child_weight': 3, 'reg_alpha': 2.637135896855224, 'reg_lambda': 2.528836723246512}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:12,581] Trial 39 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 147, 'max_depth': 12, 'learning_rate': 0.09280913635060321, 'subsample': 0.7227287167749054, 'colsample_bytree': 0.5402755304446037, 'gamma': 0.6594323738724834, 'min_child_weight': 2, 'reg_alpha': 2.562158828266748, 'reg_lambda': 4.440798648354454}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:13,102] Trial 40 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 171, 'max_depth': 14, 'learning_rate': 0.14559789965112302, 'subsample': 0.8458767819321079, 'colsample_bytree': 0.6315852852505257, 'gamma': 0.642890824568698, 'min_child_weight': 2, 'reg_alpha': 2.015864318024419, 'reg_lambda': 3.057407680618312}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:13,617] Trial 41 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 120, 'max_depth': 12, 'learning_rate': 0.06594232839867774, 'subsample': 0.824221182003489, 'colsample_bytree': 0.6088156061175672, 'gamma': 0.9134781249611784, 'min_child_weight': 4, 'reg_alpha': 2.833960302898258, 'reg_lambda': 2.6624561225590706}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:14,150] Trial 42 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 198, 'max_depth': 12, 'learning_rate': 0.14424002956829338, 'subsample': 0.8039076566200654, 'colsample_bytree': 0.6255346511673052, 'gamma': 0.9918826903966608, 'min_child_weight': 4, 'reg_alpha': 1.5819936810150388, 'reg_lambda': 3.534558758327245}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:14,921] Trial 43 finished with value: 0.9477249542785197 and parameters: {'n_estimators': 149, 'max_depth': 12, 'learning_rate': 0.1039107019670128, 'subsample': 0.8927773975660606, 'colsample_bytree': 0.6426577137364192, 'gamma': 0.19920997208050706, 'min_child_weight': 2, 'reg_alpha': 2.016587798826923, 'reg_lambda': 2.908803024880231}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:15,657] Trial 44 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 198, 'max_depth': 12, 'learning_rate': 0.13211569012022226, 'subsample': 0.8664189200484744, 'colsample_bytree': 0.5423411867187106, 'gamma': 0.4189129407561647, 'min_child_weight': 1, 'reg_alpha': 1.5495815252331357, 'reg_lambda': 4.281787640301291}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:16,231] Trial 45 finished with value: 0.9494407734512137 and parameters: {'n_estimators': 109, 'max_depth': 15, 'learning_rate': 0.10941475572570405, 'subsample': 0.7949836227272263, 'colsample_bytree':

0.6610057105972189, 'gamma': 0.20787613568691188, 'min_child_weight': 2, 'reg_alpha': 2.780989195522966, 'reg_lambda': 3.4799640076815397}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:16,648] Trial 46 finished with value: 0.9472976077786253 and parameters: {'n_estimators': 123, 'max_depth': 14, 'learning_rate': 0.14733895407321543, 'subsample': 0.890163851951413, 'colsample_bytree': 0.6372478447300816, 'gamma': 0.7260292168423824, 'min_child_weight': 1, 'reg_alpha': 1.8406165366934129, 'reg_lambda': 4.672723080790161}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:17,307] Trial 47 finished with value: 0.9507264890498204 and parameters: {'n_estimators': 175, 'max_depth': 10, 'learning_rate': 0.1444448978696616, 'subsample': 0.7523646473179986, 'colsample_bytree': 0.5046116274172302, 'gamma': 0.42998412132243335, 'min_child_weight': 1, 'reg_alpha': 1.987622517613373, 'reg_lambda': 2.916038224444164}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:18,241] Trial 48 finished with value: 0.9511538355497147 and parameters: {'n_estimators': 149, 'max_depth': 12, 'learning_rate': 0.13079532492455867, 'subsample': 0.7023133859149285, 'colsample_bytree': 0.65588465634044, 'gamma': 0.21321538744879476, 'min_child_weight': 2, 'reg_alpha': 1.0475398560661708, 'reg_lambda': 4.799210955071511}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:19,071] Trial 49 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 200, 'max_depth': 11, 'learning_rate': 0.08878284574990788, 'subsample': 0.7467239677452236, 'colsample_bytree': 0.5382198748733934, 'gamma': 0.23261037942662788, 'min_child_weight': 3, 'reg_alpha': 2.3811526704521238, 'reg_lambda': 4.214077323663826}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:19,846] Trial 50 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 192, 'max_depth': 14, 'learning_rate': 0.07812754809049481, 'subsample': 0.7882092395786967, 'colsample_bytree': 0.5178455168127862, 'gamma': 0.6123496160045772, 'min_child_weight': 3, 'reg_alpha': 1.5833834632052406, 'reg_lambda': 4.768506760661705}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:20,613] Trial 51 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 177, 'max_depth': 11, 'learning_rate': 0.055132053816450526, 'subsample': 0.8181396491659181, 'colsample_bytree': 0.5727872358165648, 'gamma': 0.50429242431102, 'min_child_weight': 4, 'reg_alpha': 2.769914155330893, 'reg_lambda': 4.162179149807781}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:21,491] Trial 52 finished with value: 0.9507255700250894 and parameters: {'n_estimators': 181, 'max_depth': 10, 'learning_rate': 0.076711203976696, 'subsample': 0.8709117615698094, 'colsample_bytree': 0.5019636228944524, 'gamma': 0.30236266475045315, 'min_child_weight': 1, 'reg_alpha': 2.038973322816653, 'reg_lambda': 3.4450275183654293}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:21,921] Trial 53 finished with value: 0.9468693422540001 and parameters: {'n_estimators': 111, 'max_depth': 13, 'learning_rate': 0.10349399457370759, 'subsample': 0.7800947441419327, 'colsample_bytree':

0.5986428375344792, 'gamma': 0.8122970470767471, 'min_child_weight': 3, 'reg_alpha': 2.3267542431030384, 'reg_lambda': 4.201858055221688}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:22,452] Trial 54 finished with value: 0.948584242401963 and parameters: {'n_estimators': 190, 'max_depth': 13, 'learning_rate': 0.12984818530317718, 'subsample': 0.854641534007358, 'colsample_bytree': 0.6556407630726512, 'gamma': 0.7800492392315074, 'min_child_weight': 4, 'reg_alpha': 2.0813695084857367, 'reg_lambda': 4.027914101133323}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:23,138] Trial 55 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 162, 'max_depth': 12, 'learning_rate': 0.06652975111650286, 'subsample': 0.7431662935697627, 'colsample_bytree': 0.5113601403350788, 'gamma': 0.7982999580960123, 'min_child_weight': 2, 'reg_alpha': 1.933259836672481, 'reg_lambda': 4.128840839508116}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:23,964] Trial 56 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 200, 'max_depth': 11, 'learning_rate': 0.10334778916024208, 'subsample': 0.7608829415744112, 'colsample_bytree': 0.6978566898903544, 'gamma': 0.3424715401611057, 'min_child_weight': 2, 'reg_alpha': 2.2613444552873965, 'reg_lambda': 3.421398235060336}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:24,374] Trial 57 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 126, 'max_depth': 13, 'learning_rate': 0.12447242074567254, 'subsample': 0.8118532306356078, 'colsample_bytree': 0.5005963143108539, 'gamma': 0.8670559306217968, 'min_child_weight': 1, 'reg_alpha': 2.7042710487070076, 'reg_lambda': 4.699227663986383}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:25,315] Trial 58 finished with value: 0.9502963854757333 and parameters: {'n_estimators': 174, 'max_depth': 13, 'learning_rate': 0.12117426481419706, 'subsample': 0.7043016593620306, 'colsample_bytree': 0.5637412463246502, 'gamma': 0.21532454244250643, 'min_child_weight': 1, 'reg_alpha': 1.8222614006074478, 'reg_lambda': 2.45393533551338}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:26,247] Trial 59 finished with value: 0.9485824043525012 and parameters: {'n_estimators': 196, 'max_depth': 11, 'learning_rate': 0.08776573804331349, 'subsample': 0.7952834443188853, 'colsample_bytree': 0.6327444722554992, 'gamma': 0.4850190064802, 'min_child_weight': 1, 'reg_alpha': 1.3551171258889347, 'reg_lambda': 2.707960528013282}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:27,094] Trial 60 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 166, 'max_depth': 11, 'learning_rate': 0.052662249525047555, 'subsample': 0.7493242271581334, 'colsample_bytree': 0.5534227854044855, 'gamma': 0.582919121700372, 'min_child_weight': 1, 'reg_alpha': 1.8868402820853762, 'reg_lambda': 2.233380953701156}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:27,675] Trial 61 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 118, 'max_depth': 13, 'learning_rate': 0.06318490723638906, 'subsample': 0.8865843223740745, 'colsample_bytree':

0.648576233176584, 'gamma': 0.647554356738976, 'min_child_weight': 4,
'reg_alpha': 2.273738993611893, 'reg_lambda': 3.892081880340489}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:28,254] Trial 62 finished with value: 0.9485833233772322 and
parameters: {'n_estimators': 161, 'max_depth': 14, 'learning_rate':
0.1237478971558728, 'subsample': 0.752281268175999, 'colsample_bytree':
0.6518715275827757, 'gamma': 0.6240172277544602, 'min_child_weight': 2,
'reg_alpha': 1.9696459833879152, 'reg_lambda': 4.761603070367904}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:29,391] Trial 63 finished with value: 0.9515821010743398 and
parameters: {'n_estimators': 178, 'max_depth': 13, 'learning_rate':
0.05654376414090278, 'subsample': 0.8045099031013505, 'colsample_bytree':
0.5039131626532406, 'gamma': 0.28845231262969706, 'min_child_weight': 1,
'reg_alpha': 1.0080003197469682, 'reg_lambda': 4.304371236077509}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:30,041] Trial 64 finished with value: 0.9481541388278758 and
parameters: {'n_estimators': 137, 'max_depth': 10, 'learning_rate':
0.06410212299169266, 'subsample': 0.848318924959335, 'colsample_bytree':
0.5336158387193178, 'gamma': 0.6081387899607208, 'min_child_weight': 3,
'reg_alpha': 2.1302251537703087, 'reg_lambda': 3.892786808039461}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:30,710] Trial 65 finished with value: 0.9481550578526068 and
parameters: {'n_estimators': 148, 'max_depth': 11, 'learning_rate':
0.12616471281345482, 'subsample': 0.7824140324278249, 'colsample_bytree':
0.6198886842060142, 'gamma': 0.18156588023194853, 'min_child_weight': 3,
'reg_alpha': 2.8326815196752575, 'reg_lambda': 2.574595877974513}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:31,185] Trial 66 finished with value: 0.9468684232292691 and
parameters: {'n_estimators': 124, 'max_depth': 12, 'learning_rate':
0.05922193158849854, 'subsample': 0.7159034703063484, 'colsample_bytree':
0.5279471703836859, 'gamma': 0.7874339793012785, 'min_child_weight': 4,
'reg_alpha': 2.5417848525717908, 'reg_lambda': 4.4605019211900965}. Best is
trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:31,632] Trial 67 finished with value: 0.948584242401963 and
parameters: {'n_estimators': 122, 'max_depth': 12, 'learning_rate':
0.10641992666681732, 'subsample': 0.8119665145433405, 'colsample_bytree':
0.6076395757282135, 'gamma': 0.6356870036762355, 'min_child_weight': 3,
'reg_alpha': 2.297273688449872, 'reg_lambda': 2.16842846781199}. Best is trial 2
with value: 0.9520122046484272.
[I 2025-08-21 20:34:32,103] Trial 68 finished with value: 0.9502973045004641 and
parameters: {'n_estimators': 161, 'max_depth': 13, 'learning_rate':
0.138166741598715, 'subsample': 0.7046900626844119, 'colsample_bytree':
0.530877957157552, 'gamma': 0.9127155429798095, 'min_child_weight': 2,
'reg_alpha': 1.825164058613596, 'reg_lambda': 2.551096193502845}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:32,727] Trial 69 finished with value: 0.9490115889018573 and
parameters: {'n_estimators': 177, 'max_depth': 10, 'learning_rate':
0.07831441217283568, 'subsample': 0.8707555678759834, 'colsample_bytree':

0.5452877281623818, 'gamma': 0.9959117755297433, 'min_child_weight': 1, 'reg_alpha': 1.5922688579434898, 'reg_lambda': 3.6582192199692836}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:33,266] Trial 70 finished with value: 0.9481559768773378 and parameters: {'n_estimators': 173, 'max_depth': 10, 'learning_rate': 0.14942786439557132, 'subsample': 0.8733260008713272, 'colsample_bytree': 0.5744106537170746, 'gamma': 0.3880213783009995, 'min_child_weight': 4, 'reg_alpha': 2.6509295969049393, 'reg_lambda': 4.86584889016505}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:33,881] Trial 71 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 177, 'max_depth': 14, 'learning_rate': 0.09918789792286443, 'subsample': 0.8209069341760192, 'colsample_bytree': 0.5972201101919611, 'gamma': 0.8067128963266318, 'min_child_weight': 1, 'reg_alpha': 1.701802151959366, 'reg_lambda': 3.7369658651393434}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:34,727] Trial 72 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 183, 'max_depth': 11, 'learning_rate': 0.07029879886554893, 'subsample': 0.8225610713633671, 'colsample_bytree': 0.6990345205202046, 'gamma': 0.3961765682893055, 'min_child_weight': 1, 'reg_alpha': 2.7579543456696163, 'reg_lambda': 3.987692328661043}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:35,297] Trial 73 finished with value: 0.9481559768773378 and parameters: {'n_estimators': 195, 'max_depth': 12, 'learning_rate': 0.11798355699787658, 'subsample': 0.7437643275071579, 'colsample_bytree': 0.5084128182139659, 'gamma': 0.821331845525663, 'min_child_weight': 2, 'reg_alpha': 2.3357598473254715, 'reg_lambda': 2.2111056329696974}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:36,031] Trial 74 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 122, 'max_depth': 15, 'learning_rate': 0.10605402239475205, 'subsample': 0.8682346158397213, 'colsample_bytree': 0.6349852800024446, 'gamma': 0.03643126498263005, 'min_child_weight': 4, 'reg_alpha': 1.754104185099064, 'reg_lambda': 3.557271779739872}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:36,624] Trial 75 finished with value: 0.9485833233772322 and parameters: {'n_estimators': 141, 'max_depth': 12, 'learning_rate': 0.08601818163261421, 'subsample': 0.7431871659771125, 'colsample_bytree': 0.6032550863940671, 'gamma': 0.37995306017658237, 'min_child_weight': 2, 'reg_alpha': 2.9985668448856186, 'reg_lambda': 4.3061103132795875}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:37,452] Trial 76 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 175, 'max_depth': 12, 'learning_rate': 0.14201741197220596, 'subsample': 0.8213346602054226, 'colsample_bytree': 0.6768781222857924, 'gamma': 0.09690956442922938, 'min_child_weight': 3, 'reg_alpha': 2.216870473435872, 'reg_lambda': 2.600293621830087}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:37,989] Trial 77 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 184, 'max_depth': 12, 'learning_rate': 0.13908902208396864, 'subsample': 0.8220608245259743, 'colsample_bytree':

0.6530915239627391, 'gamma': 0.7585012366722129, 'min_child_weight': 4, 'reg_alpha': 1.3041238758876375, 'reg_lambda': 2.2083115372390805}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:38,411] Trial 78 finished with value: 0.9477267923279816 and parameters: {'n_estimators': 110, 'max_depth': 11, 'learning_rate': 0.0916728815818674, 'subsample': 0.8827990317536053, 'colsample_bytree': 0.6303615314665666, 'gamma': 0.6233333100026895, 'min_child_weight': 3, 'reg_alpha': 2.770589310202318, 'reg_lambda': 4.087170886721737}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:39,102] Trial 79 finished with value: 0.9498690389758389 and parameters: {'n_estimators': 117, 'max_depth': 12, 'learning_rate': 0.08512606978484491, 'subsample': 0.720116448351916, 'colsample_bytree': 0.6653520973561371, 'gamma': 0.3322228058132747, 'min_child_weight': 1, 'reg_alpha': 1.5723040517513633, 'reg_lambda': 2.2898675583694494}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:40,020] Trial 80 finished with value: 0.9494407734512137 and parameters: {'n_estimators': 188, 'max_depth': 10, 'learning_rate': 0.055792294661880654, 'subsample': 0.840611008854633, 'colsample_bytree': 0.653253519774826, 'gamma': 0.4401636042512067, 'min_child_weight': 3, 'reg_alpha': 2.6416625190128817, 'reg_lambda': 2.1144345955416277}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:40,636] Trial 81 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 189, 'max_depth': 12, 'learning_rate': 0.09450753309979391, 'subsample': 0.805205474830961, 'colsample_bytree': 0.5269715596984302, 'gamma': 0.9993691502004199, 'min_child_weight': 3, 'reg_alpha': 1.424567738078394, 'reg_lambda': 4.527837567044466}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:41,492] Trial 82 finished with value: 0.9494407734512137 and parameters: {'n_estimators': 151, 'max_depth': 12, 'learning_rate': 0.05749434935071007, 'subsample': 0.8069320929184418, 'colsample_bytree': 0.6138889822535653, 'gamma': 0.3960225123804473, 'min_child_weight': 2, 'reg_alpha': 1.687423256842185, 'reg_lambda': 3.138562997181223}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:41,992] Trial 83 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 107, 'max_depth': 15, 'learning_rate': 0.05454114429349953, 'subsample': 0.7127898712997481, 'colsample_bytree': 0.6056340893057307, 'gamma': 0.5794141864513631, 'min_child_weight': 4, 'reg_alpha': 2.616632978335323, 'reg_lambda': 2.3820178824018217}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:42,567] Trial 84 finished with value: 0.9485833233772322 and parameters: {'n_estimators': 137, 'max_depth': 14, 'learning_rate': 0.0700490762902502, 'subsample': 0.7695405301975684, 'colsample_bytree': 0.6121142710049705, 'gamma': 0.8176605620408494, 'min_child_weight': 3, 'reg_alpha': 1.7324005053706564, 'reg_lambda': 2.32366643134249}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:43,199] Trial 85 finished with value: 0.9490125079265883 and parameters: {'n_estimators': 133, 'max_depth': 15, 'learning_rate': 0.05863884354822237, 'subsample': 0.8277719436471879, 'colsample_bytree':

0.5781113577149322, 'gamma': 0.5172204600083908, 'min_child_weight': 2, 'reg_alpha': 2.375457972671981, 'reg_lambda': 2.3940700706495583}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:43,647] Trial 86 finished with value: 0.948584242401963 and parameters: {'n_estimators': 119, 'max_depth': 11, 'learning_rate': 0.11818372692408591, 'subsample': 0.8427171010201091, 'colsample_bytree': 0.6298570513533965, 'gamma': 0.41275498611896677, 'min_child_weight': 2, 'reg_alpha': 2.9284707414528057, 'reg_lambda': 2.309782659271092}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:44,220] Trial 87 finished with value: 0.948584242401963 and parameters: {'n_estimators': 109, 'max_depth': 10, 'learning_rate': 0.06344139420317951, 'subsample': 0.8011332960351466, 'colsample_bytree': 0.6246693376413123, 'gamma': 0.222690401167855, 'min_child_weight': 4, 'reg_alpha': 1.8528468082605565, 'reg_lambda': 2.249542994190138}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:45,064] Trial 88 finished with value: 0.9485824043525012 and parameters: {'n_estimators': 152, 'max_depth': 15, 'learning_rate': 0.06058772360409419, 'subsample': 0.8192235878507986, 'colsample_bytree': 0.6831832790597562, 'gamma': 0.2824232375617788, 'min_child_weight': 3, 'reg_alpha': 1.6940823189416472, 'reg_lambda': 2.4093632417483812}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:45,830] Trial 89 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 183, 'max_depth': 10, 'learning_rate': 0.12774013945089488, 'subsample': 0.8304695973272279, 'colsample_bytree': 0.6200503447214787, 'gamma': 0.14135186303730507, 'min_child_weight': 4, 'reg_alpha': 2.3645894916773447, 'reg_lambda': 4.239236206321599}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:46,478] Trial 90 finished with value: 0.9502963854757333 and parameters: {'n_estimators': 144, 'max_depth': 13, 'learning_rate': 0.10367684320146266, 'subsample': 0.8998307437856197, 'colsample_bytree': 0.6901793193112474, 'gamma': 0.14684585393452343, 'min_child_weight': 2, 'reg_alpha': 2.8501108021317125, 'reg_lambda': 2.349621728776931}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:47,067] Trial 91 finished with value: 0.9481550578526068 and parameters: {'n_estimators': 186, 'max_depth': 10, 'learning_rate': 0.09925815004698826, 'subsample': 0.7335824859700015, 'colsample_bytree': 0.6244648174782046, 'gamma': 0.7236733726535681, 'min_child_weight': 4, 'reg_alpha': 2.176523987740641, 'reg_lambda': 4.78443408455571}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:47,499] Trial 92 finished with value: 0.9494407734512137 and parameters: {'n_estimators': 100, 'max_depth': 13, 'learning_rate': 0.1029700652873935, 'subsample': 0.7238703522405109, 'colsample_bytree': 0.523016001130456, 'gamma': 0.7250580588692915, 'min_child_weight': 1, 'reg_alpha': 2.2880980987570427, 'reg_lambda': 2.3944621131613104}. Best is trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:48,184] Trial 93 finished with value: 0.9481541388278758 and parameters: {'n_estimators': 191, 'max_depth': 11, 'learning_rate': 0.058035526342476086, 'subsample': 0.7025058382244633, 'colsample_bytree':

0.507071054643387, 'gamma': 0.8947444061610484, 'min_child_weight': 4,
'reg_alpha': 1.8483133777642924, 'reg_lambda': 2.4826389043774073}. Best is
trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:48,714] Trial 94 finished with value: 0.9490125079265883 and
parameters: {'n_estimators': 150, 'max_depth': 12, 'learning_rate':
0.10162973278769492, 'subsample': 0.7173964454859203, 'colsample_bytree':
0.5999751903606022, 'gamma': 0.6564565269172564, 'min_child_weight': 1,
'reg_alpha': 2.913158455706325, 'reg_lambda': 4.946388192128716}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:49,056] Trial 95 finished with value: 0.9485833233772322 and
parameters: {'n_estimators': 104, 'max_depth': 10, 'learning_rate':
0.13705636713332034, 'subsample': 0.7103204903513395, 'colsample_bytree':
0.5887361894129605, 'gamma': 0.6930050390888259, 'min_child_weight': 2,
'reg_alpha': 2.8274337597294252, 'reg_lambda': 2.9636443683670093}. Best is
trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:49,853] Trial 96 finished with value: 0.9490115889018573 and
parameters: {'n_estimators': 157, 'max_depth': 15, 'learning_rate':
0.10002536904825243, 'subsample': 0.8973148054758765, 'colsample_bytree':
0.6539771239629499, 'gamma': 0.31054508283007975, 'min_child_weight': 3,
'reg_alpha': 1.0844433052098914, 'reg_lambda': 4.972667341590343}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:50,527] Trial 97 finished with value: 0.9490125079265883 and
parameters: {'n_estimators': 145, 'max_depth': 15, 'learning_rate':
0.05340998128896371, 'subsample': 0.8271004182440389, 'colsample_bytree':
0.5232086211520687, 'gamma': 0.717943733543281, 'min_child_weight': 3,
'reg_alpha': 1.3978523960284115, 'reg_lambda': 3.6589119487487776}. Best is
trial 2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:51,027] Trial 98 finished with value: 0.9490115889018573 and
parameters: {'n_estimators': 160, 'max_depth': 14, 'learning_rate':
0.11791639681584642, 'subsample': 0.7639916997049724, 'colsample_bytree':
0.5858758237467845, 'gamma': 0.5852850208487468, 'min_child_weight': 1,
'reg_alpha': 2.981514595272574, 'reg_lambda': 4.901329922465549}. Best is trial
2 with value: 0.9520122046484272.
[I 2025-08-21 20:34:51,549] Trial 99 finished with value: 0.9490115889018573 and
parameters: {'n_estimators': 139, 'max_depth': 12, 'learning_rate':
0.08131636121834576, 'subsample': 0.7057407501906526, 'colsample_bytree':
0.6414735049956724, 'gamma': 0.7439703335922098, 'min_child_weight': 1,
'reg_alpha': 2.985509503263726, 'reg_lambda': 2.456202904818124}. Best is trial
2 with value: 0.9520122046484272.

```python
[66]:  # Print the best result
       print(f'Best trial accuracy: {study.best_trial.value}')
       print(f'Best hyperparameters: {study.best_trial.params}')
```

Best trial accuracy: 0.9520122046484272
Best hyperparameters: {'n_estimators': 143, 'max_depth': 13, 'learning_rate':
0.122735540147556, 'subsample': 0.8835212290328968, 'colsample_bytree':
0.550234562558559, 'gamma': 0.0018382926311870662, 'min_child_weight': 1,

```
      'reg_alpha': 1.6143030673686416, 'reg_lambda': 2.7033008980487963}
```

```
[69]: # Train an XGBClassifier using the best hyperparameters from Optuna
      best_model = XGBClassifier(
          **study.best_trial.params,
          random_state=42,
          eval_metric="logloss"    # avoids warnings
      )

      # Fit the model to the training data
      best_model.fit(X_train, y_train)

      # Make predictions on the test set
      y_pred = best_model.predict(X_test)

      # Calculate the accuracy on the test set
      test_accuracy = accuracy_score(y_test, y_pred)

      # Print the test accuracy
      print(f"Test Accuracy with best hyperparameters: {test_accuracy:.2f}")
```

```
      Test Accuracy with best hyperparameters: 0.83
```

```
[68]: # Optimization History
      plot_optimization_history(study).show()
```

0.1.4 Bayesian Optimization

```
[91]: def objective(trial):
          # Hyperparameters to tune (refined from latest Random Search best trial)

          # Best so far: n_estimators = 143
          n_estimators = trial.suggest_int('n_estimators', 110, 180)

          # Best so far: max_depth = 13
          max_depth = trial.suggest_int('max_depth', 10, 15)

          # Best so far: learning_rate = 0.1227
          learning_rate = trial.suggest_float('learning_rate', 0.09, 0.16, log=True)

          # Best so far: subsample = 0.884
          subsample = trial.suggest_float('subsample', 0.80, 0.95)

          # Best so far: colsample_bytree = 0.550
          colsample_bytree = trial.suggest_float('colsample_bytree', 0.4, 0.7)

          # Best so far: gamma = 0.0018
          gamma = trial.suggest_float('gamma', 0.0, 0.1)
```

```python
    # Best so far: min_child_weight = 1
    min_child_weight = trial.suggest_int('min_child_weight', 1, 4)

    # Regularization
    # Best so far: reg_alpha = 1.614
    reg_alpha = trial.suggest_float('reg_alpha', 1.0, 2.5)

    # Best so far: reg_lambda = 2.703
    reg_lambda = trial.suggest_float('reg_lambda', 2.0, 4)

    # Create model with suggested hyperparameters
    model = XGBClassifier(
        n_estimators=n_estimators,
        max_depth=max_depth,
        learning_rate=learning_rate,
        subsample=subsample,
        colsample_bytree=colsample_bytree,
        gamma=gamma,
        min_child_weight=min_child_weight,
        reg_alpha=reg_alpha,
        reg_lambda=reg_lambda,
        eval_metric="logloss",
        random_state=42
    )

    # Cross-validation
    score = cross_val_score(
        model, X_train, y_train, cv=5, scoring="accuracy"
    ).mean()

    return score
```

```python
[92]: # Create a study object and optimize the objective function
      study = optuna.create_study(direction='maximize', sampler=optuna.samplers.
      ↪TPESampler())  # Aim to maximize accuracy
      study.optimize(objective, n_trials=50)  # Run 50 trials to find the best␣
      ↪hyperparameters
```

```
[I 2025-08-21 20:44:09,518] A new study created in memory with name: no-
name-371696bd-2e5a-4ac4-885c-039c7ffacfe4
[I 2025-08-21 20:44:10,586] Trial 0 finished with value: 0.9490125079265883 and
parameters: {'n_estimators': 162, 'max_depth': 10, 'learning_rate':
0.10261532611425381, 'subsample': 0.8268238907190912, 'colsample_bytree':
0.5699870614413879, 'gamma': 0.05105882820506183, 'min_child_weight': 3,
'reg_alpha': 1.649206084663557, 'reg_lambda': 3.0647658421883346}. Best is trial
0 with value: 0.9490125079265883.
[I 2025-08-21 20:44:11,248] Trial 1 finished with value: 0.9490125079265883 and
```

parameters: {'n_estimators': 110, 'max_depth': 15, 'learning_rate': 0.09163475413196215, 'subsample': 0.9226364117822137, 'colsample_bytree': 0.5201458670028765, 'gamma': 0.06729720490547579, 'min_child_weight': 4, 'reg_alpha': 2.086025332856344, 'reg_lambda': 2.324470348810527}. Best is trial 0 with value: 0.9490125079265883.
[I 2025-08-21 20:44:12,058] Trial 2 finished with value: 0.9498662819016459 and parameters: {'n_estimators': 143, 'max_depth': 12, 'learning_rate': 0.1252757506371609, 'subsample': 0.9129194139748001, 'colsample_bytree': 0.6785162576970036, 'gamma': 0.09771168583455003, 'min_child_weight': 1, 'reg_alpha': 2.3874797300397725, 'reg_lambda': 3.67209820946087}. Best is trial 2 with value: 0.9498662819016459.
[I 2025-08-21 20:44:13,155] Trial 3 finished with value: 0.9502973045004641 and parameters: {'n_estimators': 172, 'max_depth': 12, 'learning_rate': 0.09118310130820434, 'subsample': 0.8320602012138888, 'colsample_bytree': 0.4226030989559514, 'gamma': 0.023575698486938124, 'min_child_weight': 2, 'reg_alpha': 2.463696598731252, 'reg_lambda': 2.963429162290989}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:14,328] Trial 4 finished with value: 0.9477258733032506 and parameters: {'n_estimators': 173, 'max_depth': 13, 'learning_rate': 0.10718465095065093, 'subsample': 0.8217121525203733, 'colsample_bytree': 0.5763989121718703, 'gamma': 0.0629344461838256, 'min_child_weight': 3, 'reg_alpha': 2.147625840727687, 'reg_lambda': 2.268145181966326}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:15,472] Trial 5 finished with value: 0.9502973045004641 and parameters: {'n_estimators': 172, 'max_depth': 10, 'learning_rate': 0.14675886889633802, 'subsample': 0.82757592132731, 'colsample_bytree': 0.690813585457375, 'gamma': 0.0272081515154777, 'min_child_weight': 1, 'reg_alpha': 2.309821617497361, 'reg_lambda': 2.715160195479423}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:16,484] Trial 6 finished with value: 0.9490115889018573 and parameters: {'n_estimators': 148, 'max_depth': 10, 'learning_rate': 0.10210831383146371, 'subsample': 0.8700837596899766, 'colsample_bytree': 0.6035029842485016, 'gamma': 0.060881048248237535, 'min_child_weight': 2, 'reg_alpha': 1.7006957783468273, 'reg_lambda': 3.5805405943013637}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:17,830] Trial 7 finished with value: 0.9498690389758387 and parameters: {'n_estimators': 179, 'max_depth': 11, 'learning_rate': 0.09422079181599115, 'subsample': 0.9412961781358006, 'colsample_bytree': 0.6870756493531555, 'gamma': 0.06431971546009589, 'min_child_weight': 2, 'reg_alpha': 1.0038278551694473, 'reg_lambda': 3.8295807611198525}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:18,738] Trial 8 finished with value: 0.9494398544264827 and parameters: {'n_estimators': 134, 'max_depth': 10, 'learning_rate': 0.14181554959124074, 'subsample': 0.8476758276929958, 'colsample_bytree': 0.6377968955690965, 'gamma': 0.024804925595136897, 'min_child_weight': 1, 'reg_alpha': 2.4561618559164433, 'reg_lambda': 3.269030093698859}. Best is trial 3 with value: 0.9502973045004641.
[I 2025-08-21 20:44:19,712] Trial 9 finished with value: 0.9507264890498203 and

parameters: {'n_estimators': 149, 'max_depth': 12, 'learning_rate': 0.13587494351364068, 'subsample': 0.8069328612199825, 'colsample_bytree': 0.671931096012376, 'gamma': 0.08765338733313652, 'min_child_weight': 1, 'reg_alpha': 1.9074979810615351, 'reg_lambda': 3.2487853750166833}. Best is trial 9 with value: 0.9507264890498203.
[I 2025-08-21 20:44:20,393] Trial 10 finished with value: 0.94901250792665883 and parameters: {'n_estimators': 125, 'max_depth': 14, 'learning_rate': 0.15993351673068728, 'subsample': 0.804188604551241, 'colsample_bytree': 0.5084207542987, 'gamma': 0.09600697069462887, 'min_child_weight': 4, 'reg_alpha': 1.352777284046569, 'reg_lambda': 3.391952552229725}. Best is trial 9 with value: 0.9507264890498203.
[I 2025-08-21 20:44:21,364] Trial 11 finished with value: 0.9502963854757333 and parameters: {'n_estimators': 158, 'max_depth': 12, 'learning_rate': 0.1227495851187829, 'subsample': 0.8006535736412087, 'colsample_bytree': 0.4003942777025164, 'gamma': 0.0030028639881811148, 'min_child_weight': 2, 'reg_alpha': 1.9655890856157403, 'reg_lambda': 2.771915704760829}. Best is trial 9 with value: 0.9507264890498203.
[I 2025-08-21 20:44:22,414] Trial 12 finished with value: 0.9524395511483214 and parameters: {'n_estimators': 159, 'max_depth': 13, 'learning_rate': 0.13151561408514884, 'subsample': 0.8633855115402435, 'colsample_bytree': 0.42753851549672517, 'gamma': 0.02760779545876145, 'min_child_weight': 1, 'reg_alpha': 1.8541908402933511, 'reg_lambda': 2.8929008932234606}. Best is trial 12 with value: 0.9524395511483214.
[I 2025-08-21 20:44:23,305] Trial 13 finished with value: 0.9511529165249837 and parameters: {'n_estimators': 156, 'max_depth': 13, 'learning_rate': 0.1340487246117481, 'subsample': 0.8885901043729023, 'colsample_bytree': 0.46266690247023123, 'gamma': 0.08049980756380588, 'min_child_weight': 1, 'reg_alpha': 1.8664547216361407, 'reg_lambda': 2.6007159504924915}. Best is trial 12 with value: 0.9524395511483214.
[I 2025-08-21 20:44:24,353] Trial 14 finished with value: 0.9511547545744456 and parameters: {'n_estimators': 158, 'max_depth': 14, 'learning_rate': 0.13123611750691375, 'subsample': 0.8871740727130596, 'colsample_bytree': 0.46087025981428764, 'gamma': 0.036731377722010246, 'min_child_weight': 1, 'reg_alpha': 1.509200055247094, 'reg_lambda': 2.493025890442427}. Best is trial 12 with value: 0.9524395511483214.
[I 2025-08-21 20:44:25,462] Trial 15 finished with value: 0.9541553703210154 and parameters: {'n_estimators': 164, 'max_depth': 15, 'learning_rate': 0.11326035104234311, 'subsample': 0.875248567224301, 'colsample_bytree': 0.46071767557754806, 'gamma': 0.04162674744248357, 'min_child_weight': 1, 'reg_alpha': 1.4797852566106324, 'reg_lambda': 2.466774006625462}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:26,653] Trial 16 finished with value: 0.9498681199511079 and parameters: {'n_estimators': 164, 'max_depth': 15, 'learning_rate': 0.11406251065991518, 'subsample': 0.8601839573581251, 'colsample_bytree': 0.4598961922741633, 'gamma': 0.006318887029675764, 'min_child_weight': 3, 'reg_alpha': 1.289421955184708, 'reg_lambda': 2.906030803429983}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:27,624] Trial 17 finished with value: 0.9528668976482157 and

parameters: {'n_estimators': 136, 'max_depth': 14, 'learning_rate': 0.11632603019632681, 'subsample': 0.8982029477338772, 'colsample_bytree': 0.5044472089480128, 'gamma': 0.04050958226644458, 'min_child_weight': 2, 'reg_alpha': 1.460068556995651, 'reg_lambda': 2.204812636527895}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:28,613] Trial 18 finished with value: 0.9528678166729467 and parameters: {'n_estimators': 133, 'max_depth': 14, 'learning_rate': 0.113724900001048532, 'subsample': 0.9012082993409474, 'colsample_bytree': 0.5110130822541127, 'gamma': 0.04594985363283286, 'min_child_weight': 2, 'reg_alpha': 1.0831935198536202, 'reg_lambda': 2.0297414053587826}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:29,547] Trial 19 finished with value: 0.9528668976482157 and parameters: {'n_estimators': 121, 'max_depth': 15, 'learning_rate': 0.11026650350478007, 'subsample': 0.9105502262554316, 'colsample_bytree': 0.5360414449620481, 'gamma': 0.04676629036793674, 'min_child_weight': 2, 'reg_alpha': 1.0096928956431126, 'reg_lambda': 2.06661870368219044}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:30,517] Trial 20 finished with value: 0.9528678166729467 and parameters: {'n_estimators': 138, 'max_depth': 14, 'learning_rate': 0.1035980012871621, 'subsample': 0.8820999306844356, 'colsample_bytree': 0.48505242268403254, 'gamma': 0.01277856394742348, 'min_child_weight': 3, 'reg_alpha': 1.2059444584824197, 'reg_lambda': 2.0125795749422433}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:31,451] Trial 21 finished with value: 0.9511547545744456 and parameters: {'n_estimators': 136, 'max_depth': 14, 'learning_rate': 0.09971810315924944, 'subsample': 0.8819759572326246, 'colsample_bytree': 0.48936977684872396, 'gamma': 0.013699087658394837, 'min_child_weight': 3, 'reg_alpha': 1.2050464367713027, 'reg_lambda': 2.065810202318257}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:32,362] Trial 22 finished with value: 0.9520122046484272 and parameters: {'n_estimators': 129, 'max_depth': 15, 'learning_rate': 0.10784448904239599, 'subsample': 0.9030166739699681, 'colsample_bytree': 0.48176987673537747, 'gamma': 0.0508129822907096, 'min_child_weight': 3, 'reg_alpha': 1.1717903797805376, 'reg_lambda': 2.4004730594163135}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:33,335] Trial 23 finished with value: 0.9498681199511079 and parameters: {'n_estimators': 143, 'max_depth': 14, 'learning_rate': 0.11816771417575105, 'subsample': 0.9281227676086231, 'colsample_bytree': 0.5487052453221208, 'gamma': 0.017501878449186558, 'min_child_weight': 4, 'reg_alpha': 1.1277061453043, 'reg_lambda': 2.0444304552150983}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:34,114] Trial 24 finished with value: 0.9515821010743398 and parameters: {'n_estimators': 116, 'max_depth': 15, 'learning_rate': 0.09749925799148364, 'subsample': 0.8557868206105168, 'colsample_bytree': 0.4376110213309901, 'gamma': 0.03832484795400139, 'min_child_weight': 3, 'reg_alpha': 1.5246514183541526, 'reg_lambda': 2.1968110267028473}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:35,086] Trial 25 finished with value: 0.9520112856236962 and

parameters: {'n_estimators': 129, 'max_depth': 14, 'learning_rate': 0.11122321758961302, 'subsample': 0.8785162355488689, 'colsample_bytree': 0.48646222919440973, 'gamma': 0.013074951390334147, 'min_child_weight': 2, 'reg_alpha': 1.2784703038951286, 'reg_lambda': 2.4751823131404933}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:36,045] Trial 26 finished with value: 0.9507246510003584 and parameters: {'n_estimators': 151, 'max_depth': 13, 'learning_rate': 0.10515808592638622, 'subsample': 0.8444075619157689, 'colsample_bytree': 0.5304356092889226, 'gamma': 0.07352227493985705, 'min_child_weight': 3, 'reg_alpha': 1.4206024703223359, 'reg_lambda': 2.0347114768742545}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:36,971] Trial 27 finished with value: 0.9524386321235905 and parameters: {'n_estimators': 140, 'max_depth': 15, 'learning_rate': 0.12032161395161968, 'subsample': 0.896591089682392, 'colsample_bytree': 0.44746594948956137, 'gamma': 0.055829773551791466, 'min_child_weight': 2, 'reg_alpha': 1.1121714967749063, 'reg_lambda': 2.6131217809896166}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:37,765] Trial 28 finished with value: 0.9498699580005698 and parameters: {'n_estimators': 132, 'max_depth': 14, 'learning_rate': 0.12635194894966695, 'subsample': 0.868870311954463, 'colsample_bytree': 0.563554752943519, 'gamma': 0.03658108944624059, 'min_child_weight': 4, 'reg_alpha': 1.3507241741831277, 'reg_lambda': 2.174133851522378}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:38,767] Trial 29 finished with value: 0.9498672009263769 and parameters: {'n_estimators': 165, 'max_depth': 13, 'learning_rate': 0.10380769024484346, 'subsample': 0.9404214469314017, 'colsample_bytree': 0.5861230068381648, 'gamma': 0.04594110581708007, 'min_child_weight': 3, 'reg_alpha': 1.6039194120296079, 'reg_lambda': 2.3415881942494052}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:39,710] Trial 30 finished with value: 0.9507255700250894 and parameters: {'n_estimators': 153, 'max_depth': 15, 'learning_rate': 0.11274522995293954, 'subsample': 0.8906805302366333, 'colsample_bytree': 0.4741720944609904, 'gamma': 0.05341047355520023, 'min_child_weight': 2, 'reg_alpha': 1.6316870117537365, 'reg_lambda': 2.433278170600439}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:40,617] Trial 31 finished with value: 0.9520103665989652 and parameters: {'n_estimators': 138, 'max_depth': 14, 'learning_rate': 0.11604297090650839, 'subsample': 0.90181933252316, 'colsample_bytree': 0.5021179385752037, 'gamma': 0.041213664471040676, 'min_child_weight': 2, 'reg_alpha': 1.4756786818094703, 'reg_lambda': 2.1935277982944843}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:41,471] Trial 32 finished with value: 0.9524404701730523 and parameters: {'n_estimators': 122, 'max_depth': 14, 'learning_rate': 0.11684607800503671, 'subsample': 0.9213991733516527, 'colsample_bytree': 0.5113407031431626, 'gamma': 0.03243516197625984, 'min_child_weight': 2, 'reg_alpha': 1.2483205422298067, 'reg_lambda': 2.225584254627254}. Best is trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:42,519] Trial 33 finished with value: 0.9502982235251951 and

24

parameters: {'n_estimators': 144, 'max_depth': 15, 'learning_rate':
0.09850404562880204, 'subsample': 0.9118913120646905, 'colsample_bytree':
0.5236690269638855, 'gamma': 0.04630983184742699, 'min_child_weight': 1,
'reg_alpha': 1.4022227058302632, 'reg_lambda': 2.0039946071753265}. Best is
trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:43,464] Trial 34 finished with value: 0.9507264890498204 and
parameters: {'n_estimators': 140, 'max_depth': 14, 'learning_rate':
0.10890261012773751, 'subsample': 0.879113176357133, 'colsample_bytree':
0.4999707455809371, 'gamma': 0.07074455801195509, 'min_child_weight': 3,
'reg_alpha': 1.1175277938043917, 'reg_lambda': 2.144835395103741}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:44,333] Trial 35 finished with value: 0.9498672009263769 and
parameters: {'n_estimators': 132, 'max_depth': 13, 'learning_rate':
0.12518665826164865, 'subsample': 0.8946101914284673, 'colsample_bytree':
0.5416638679349134, 'gamma': 0.05782584865220722, 'min_child_weight': 2,
'reg_alpha': 1.7837298342386343, 'reg_lambda': 2.312851673309867}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:45,470] Trial 36 finished with value: 0.9520122046484272 and
parameters: {'n_estimators': 146, 'max_depth': 15, 'learning_rate':
0.10515613393841303, 'subsample': 0.9043910849392329, 'colsample_bytree':
0.5603421334951462, 'gamma': 0.03129306192640869, 'min_child_weight': 1,
'reg_alpha': 1.3241931504490676, 'reg_lambda': 2.580490465219187}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:46,363] Trial 37 finished with value: 0.9485824043525012 and
parameters: {'n_estimators': 127, 'max_depth': 14, 'learning_rate':
0.11336265862834309, 'subsample': 0.9198717513721477, 'colsample_bytree':
0.412576765881020586, 'gamma': 0.021184334967526072, 'min_child_weight': 2,
'reg_alpha': 1.5515701080753, 'reg_lambda': 2.3250305168852767}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:47,251] Trial 38 finished with value: 0.9502973045004641 and
parameters: {'n_estimators': 113, 'max_depth': 15, 'learning_rate':
0.09548490614789745, 'subsample': 0.8730097524617161, 'colsample_bytree':
0.4505665658275649, 'gamma': 0.04073845993379301, 'min_child_weight': 3,
'reg_alpha': 1.0518007805687986, 'reg_lambda': 3.075271612613018}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:48,409] Trial 39 finished with value: 0.9507255700250894 and
parameters: {'n_estimators': 169, 'max_depth': 11, 'learning_rate':
0.10134841641652972, 'subsample': 0.931829943760373, 'colsample_bytree':
0.4713384453037794, 'gamma': 0.010682419564784128, 'min_child_weight': 2,
'reg_alpha': 1.7399720549502933, 'reg_lambda': 2.7480119473815856}. Best is
trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:49,517] Trial 40 finished with value: 0.9515830200990709 and
parameters: {'n_estimators': 136, 'max_depth': 13, 'learning_rate':
0.11950484533354624, 'subsample': 0.88337105344268, 'colsample_bytree':
0.5181185787582083, 'gamma': 0.000673631270040484, 'min_child_weight': 1,
'reg_alpha': 1.4187840885142662, 'reg_lambda': 2.112216490882899}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:50,417] Trial 41 finished with value: 0.9511538355497147 and

parameters: {'n_estimators': 114, 'max_depth': 15, 'learning_rate':
0.11033123147875573, 'subsample': 0.91031577837668, 'colsample_bytree':
0.5411722597728904, 'gamma': 0.04680348933507258, 'min_child_weight': 2,
'reg_alpha': 1.000032219801387, 'reg_lambda': 2.2671381093192515}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:51,318] Trial 42 finished with value: 0.9520112856236962 and
parameters: {'n_estimators': 120, 'max_depth': 15, 'learning_rate':
0.10766244988975028, 'subsample': 0.8972829016905488, 'colsample_bytree':
0.5305199770155696, 'gamma': 0.04550999138672532, 'min_child_weight': 2,
'reg_alpha': 1.074752595420894, 'reg_lambda': 2.004274246733531}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:52,203] Trial 43 finished with value: 0.9524404701730523 and
parameters: {'n_estimators': 123, 'max_depth': 14, 'learning_rate':
0.11517158319850344, 'subsample': 0.9170473548731481, 'colsample_bytree':
0.49549421164381924, 'gamma': 0.03288647610021463, 'min_child_weight': 2,
'reg_alpha': 1.1781647299209637, 'reg_lambda': 2.1257646297396833}. Best is
trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:53,105] Trial 44 finished with value: 0.9485814853277702 and
parameters: {'n_estimators': 178, 'max_depth': 15, 'learning_rate':
0.12096665452467573, 'subsample': 0.9088314237085611, 'colsample_bytree':
0.5933928805568814, 'gamma': 0.058971048138942224, 'min_child_weight': 3,
'reg_alpha': 2.1804434275926607, 'reg_lambda': 2.2559604053512667}. Best is
trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:53,986] Trial 45 finished with value: 0.9485824043525012 and
parameters: {'n_estimators': 132, 'max_depth': 14, 'learning_rate':
0.11092123682240596, 'subsample': 0.8481405475063741, 'colsample_bytree':
0.6328320163256993, 'gamma': 0.06676697973726804, 'min_child_weight': 2,
'reg_alpha': 1.207732193946815, 'reg_lambda': 3.99220606158872}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:54,911] Trial 46 finished with value: 0.9511556735991766 and
parameters: {'n_estimators': 117, 'max_depth': 15, 'learning_rate':
0.0923315726455507, 'subsample': 0.8672030388503582, 'colsample_bytree':
0.5119475372888822, 'gamma': 0.05071558797188447, 'min_child_weight': 1,
'reg_alpha': 1.057562046919737, 'reg_lambda': 2.1220819964512527}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:55,953] Trial 47 finished with value: 0.9498699580005698 and
parameters: {'n_estimators': 141, 'max_depth': 11, 'learning_rate':
0.10565137163792157, 'subsample': 0.9281934631803254, 'colsample_bytree':
0.5702632339024046, 'gamma': 0.02152702092191655, 'min_child_weight': 2,
'reg_alpha': 1.0039946708673093, 'reg_lambda': 2.3850854638928447}. Best is
trial 15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:56,952] Trial 48 finished with value: 0.9502973045004641 and
parameters: {'n_estimators': 147, 'max_depth': 12, 'learning_rate':
0.12403096844753747, 'subsample': 0.9486432919493897, 'colsample_bytree':
0.4763419670222533, 'gamma': 0.04278610689840389, 'min_child_weight': 1,
'reg_alpha': 1.2081794343849785, 'reg_lambda': 2.530498780038285}. Best is trial
15 with value: 0.9541553703210154.
[I 2025-08-21 20:44:57,693] Trial 49 finished with value: 0.9507264890498204 and

```
parameters: {'n_estimators': 127, 'max_depth': 14, 'learning_rate':
0.10264711379728857, 'subsample': 0.8878784451262696, 'colsample_bytree':
0.43416395423602644, 'gamma': 0.027832508912942602, 'min_child_weight': 4,
'reg_alpha': 1.6802342555288454, 'reg_lambda': 2.6707458621818514}. Best is
trial 15 with value: 0.9541553703210154.
```

[93]:
```python
# Print the best result
print(f'Best trial accuracy: {study.best_trial.value}')
print(f'Best hyperparameters: {study.best_trial.params}')
```

```
Best trial accuracy: 0.9541553703210154
Best hyperparameters: {'n_estimators': 164, 'max_depth': 15, 'learning_rate':
0.11326035104234311, 'subsample': 0.875248567224301, 'colsample_bytree':
0.46071767557754806, 'gamma': 0.04162674744248357, 'min_child_weight': 1,
'reg_alpha': 1.4797852566106324, 'reg_lambda': 2.466774006625462}
```

[97]:
```python
# Train an XGBClassifier using the best hyperparameters from Optuna
best_model = XGBClassifier(
    **study.best_trial.params,
    random_state=42,
    eval_metric="logloss"   # avoids warnings
)

# Fit the model to the training data
best_model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = best_model.predict(X_test)

# Calculate the accuracy on the test set
test_accuracy = accuracy_score(y_test, y_pred)

# Print the test accuracy
print(f"Test Accuracy with best hyperparameters: {test_accuracy:.2f}")
```

```
Test Accuracy with best hyperparameters: 0.83
```

[98]:
```python
# Optimization History
plot_optimization_history(study).show()
```

[99]:
```python
# Contour Plot
plot_contour(study).show()
```

0.1.5 Grid Search

[122]:
```python
# Define search space (refined grid from Random + Bayesian results)
search_space = {
    "n_estimators": [150, 200, 300, 400],        # Random ~143, Bayesian ~164
    "max_depth": [7, 8, 10, 15],                 # Random ~13, Bayesian ~15
```

```python
        "learning_rate": [0.1, 0.15, 0.2],      # Random ~0.123, Bayesian ~0.113
        "subsample": [0.85, 0.9],            # Random ~0.884, Bayesian ~0.875
        "colsample_bytree": [0.45, 0.7, 0.9],   # Random ~0.550, Bayesian ~0.461
        "gamma": [0.025, 0.05, 0.1],               # Random ~0.002, Bayesian ~0.042
        "min_child_weight": [1, 2],             # Both suggest ~1
        "reg_alpha": [0, 1, 1.5, 1.8],             # Random ~1.614, Bayesian ~1.480
        "reg_lambda": [0, 1, 2.0, 2.5, 2.75]          # Random ~2.703, Bayesian
    ↪~2.467
}
```

[136]:
```python
# Objective function
def objective(trial):
    # Pick hyperparameters from the fixed grid
    params = {
        "n_estimators": trial.suggest_categorical("n_estimators",
    ↪search_space["n_estimators"]),
        "max_depth": trial.suggest_categorical("max_depth",
    ↪search_space["max_depth"]),
        "learning_rate": trial.suggest_categorical("learning_rate",
    ↪search_space["learning_rate"]),
        "subsample": trial.suggest_categorical("subsample",
    ↪search_space["subsample"]),
        "colsample_bytree": trial.suggest_categorical("colsample_bytree",
    ↪search_space["colsample_bytree"]),
        "gamma": trial.suggest_categorical("gamma", search_space["gamma"]),
        "min_child_weight": trial.suggest_categorical("min_child_weight",
    ↪search_space["min_child_weight"]),
        "reg_alpha": trial.suggest_categorical("reg_alpha",
    ↪search_space["reg_alpha"]),
        "reg_lambda": trial.suggest_categorical("reg_lambda",
    ↪search_space["reg_lambda"]),
        "eval_metric": "logloss",
        "random_state": 42
    }

    # Build model
    model = XGBClassifier(**params)

    # Cross-validation (5-fold for speed)
    score = cross_val_score(
        model, X_train, y_train, cv=5, scoring="accuracy"
    ).mean()

    return score
```

```
# Run Grid Search with Optuna
study = optuna.create_study(
    direction="maximize",
    sampler=optuna.samplers.GridSampler(search_space)
)

study.optimize(objective)
```

```
# Results
print("Best Accuracy:", study.best_value)
print("Best Parameters:", study.best_params)
```

```
Best Accuracy: 0.9537252667469283
Best Parameters: {'n_estimators': 150, 'max_depth': 15, 'learning_rate': 0.2,
'subsample': 0.9, 'colsample_bytree': 0.45, 'gamma': 0.1, 'min_child_weight': 1,
'reg_alpha': 1, 'reg_lambda': 2.0}
```

```
# Train an XGBClassifier using the best hyperparameters from Optuna
best_model = XGBClassifier(
    **study.best_trial.params,
    random_state=42,
    eval_metric="logloss"    # avoids warnings
)

# Fit the model to the training data
best_model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = best_model.predict(X_test)

# Calculate the accuracy on the test set
test_accuracy = accuracy_score(y_test, y_pred)

# Print the test accuracy
print(f"Test Accuracy with best hyperparameters: {test_accuracy:.2f}")
```

```
Test Accuracy with best hyperparameters: 0.84
```

```
# Probabilities for the positive class
y_proba = best_model.predict_proba(X_test)[:, 1]

# ROC data
fpr, tpr, roc_thresholds = roc_curve(y_test, y_proba)
roc_auc = roc_auc_score(y_test, y_proba)

# Precision-Recall data
precision, recall, pr_thresholds = precision_recall_curve(y_test, y_proba)
pr_auc = auc(recall, precision)
```

```python
print(f"ROC AUC: {roc_auc:.3f}")
print(f"PR AUC : {pr_auc:.3f}")

# Plot ROC
plt.figure()
plt.plot(fpr, tpr, label=f"ROC (AUC={roc_auc:.3f})")
plt.plot([0, 1], [0, 1], linestyle="--")
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curve")
plt.legend()
plt.tight_layout()
plt.show()

# Plot Precision-Recall
plt.figure()
plt.plot(recall, precision, label=f"PR (AUC={pr_auc:.3f})")
plt.xlabel("Recall")
plt.ylabel("Precision")
plt.title("Precision-Recall Curve")
plt.legend()
plt.tight_layout()
plt.show()

# If you want the raw curve arrays later:
roc_data = {"fpr": fpr, "tpr": tpr, "thresholds": roc_thresholds, "auc":
  ↪roc_auc}
pr_data  = {"precision": precision, "recall": recall, "thresholds":
  ↪pr_thresholds, "auc": pr_auc}
```
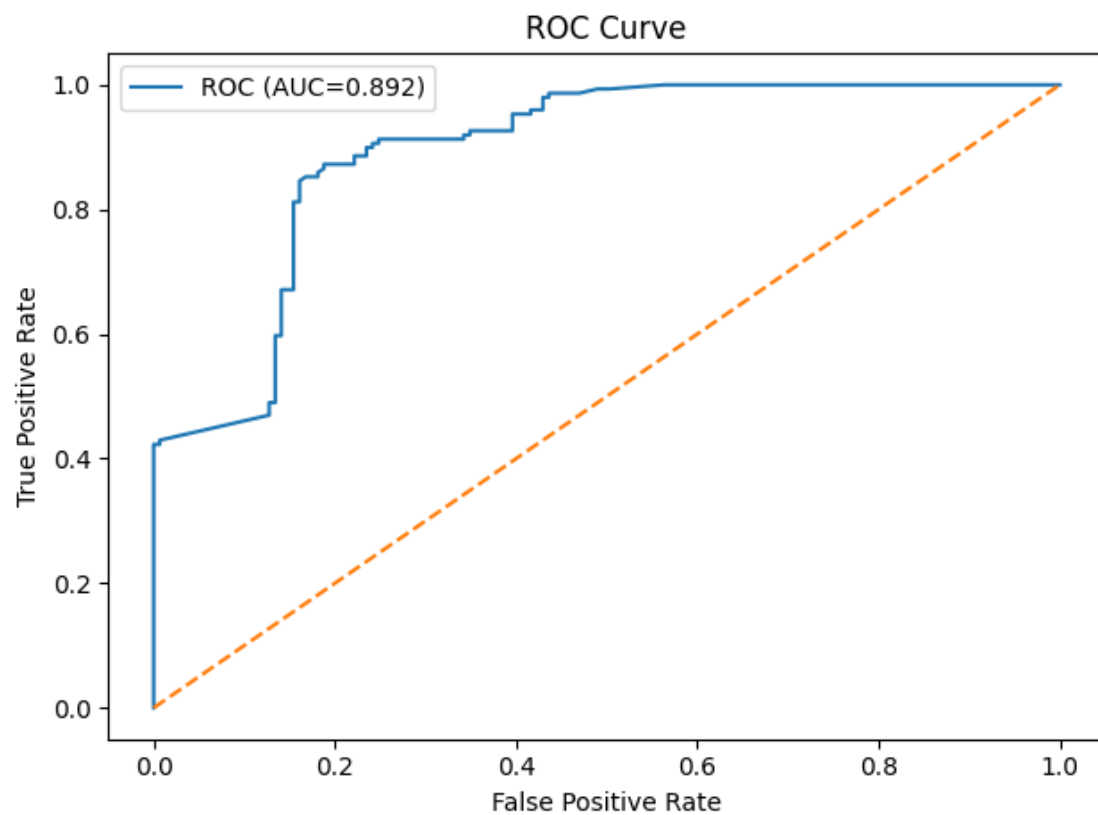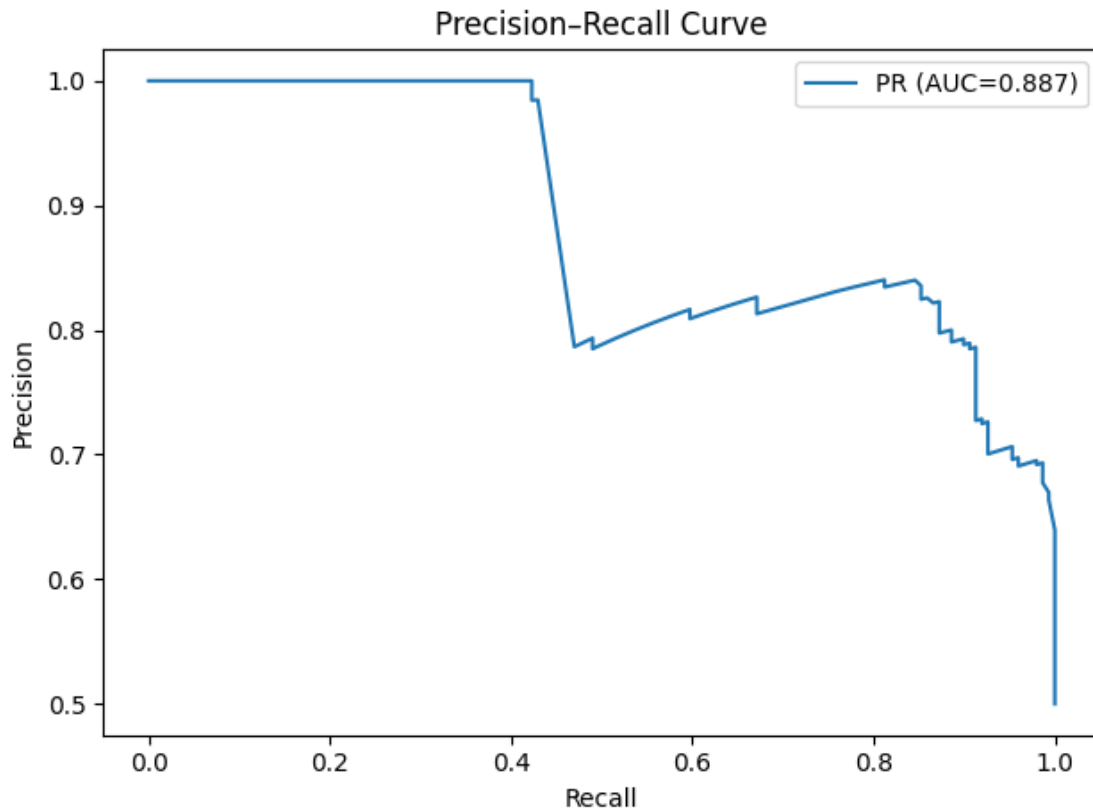
```
ROC AUC: 0.892
PR AUC : 0.887
```

ROC Curve

Precision–Recall Curve

```
[142]: # Optimization History
       plot_optimization_history(study).show()
```

```
[143]: # Parallel Coordinates Plot
       plot_parallel_coordinate(study).show()
```

```
[144]: # Slice Plot
       plot_slice(study).show()
```

```
[145]: # Contour Plot
       plot_contour(study).show()
```

```
[146]: # Hyperparameter Importance
       plot_param_importances(study).show()
```

0.2 Conclusion

Our tuned **XGBoost model** demonstrates that it is possible to **predict whether a government contract will lapse with more than 80% accuracy**.
Specifically, the model achieved:

- **Accuracy**: 84%

- **ROC-AUC**: 0.892

- **PR-AUC**: 0.887

with the following best hyperparameters:
"'python {'n_estimators': 150, 'max_depth': 15, 'learning_rate': 0.2, 'subsample': 0.9, 'colsample_bytree': 0.45, 'gamma': 0.1, 'min_child_weight': 1, 'reg_alpha': 1, 'reg_lambda': 2.0}

0.2.1 Next Goals

1. **Firm-Level Impact**
   Investigate the relationship between **government contracts** and firm outcomes such as **revenue growth** and **R&D investment**.

2. **Macro-Level Impact**
   Explore how **government spending on contracts** influences the **labor market**, including **job creation** and **sectoral employment trends**.

3. **Extended Modeling**
   Integrate additional **economic and financial variables** to evaluate whether contract lapse risk correlates with **broader industry dynamics**.