

소프트웨어학과 32204041 정다훈 4장 과제

결과 요약

Best Test Result: 0.7705627705627706

Decision Tree Result

하이퍼파라미터 튜닝 전: Train accuracy : 0.819366852886406 Test accuracy : 0.7359307359307359

하이퍼파라미터 튜닝 후: Train accuracy : 0.7821229050279329 Test accuracy : 0.7532467532467533

SVM Result

하이퍼파라미터 튜닝 전: Train accuracy : 0.7821229050279329 Test accuracy : 0.7229437229437229

하이퍼파라미터 튜닝 후: Train accuracy : 0.7821229050279329 Test accuracy : 0.7532467532467533

RF

하이퍼파라미터 튜닝 전: Train accuracy : 0.9981378026070763 Test accuracy : 0.7532467532467533

하이퍼파라미터 튜닝 후: Train accuracy : 0.8640595903165735 Test accuracy : 0.7705627705627706

XGBoost Result

하이퍼파라미터 튜닝 전: Test accuracy : 0.7272727272727273

하이퍼파라미터 튜닝 후: Train accuracy : 0.9162011173184358 Test accuracy : 0.7705627705627706

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
```

```
In [2]: df = pd.read_csv("C:/dankook/DeepLearning_Cloud/data/PimaIndiansDiabetes.csv")
```

```
In [3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   pregnant    768 non-null    int64
1   glucose     768 non-null    int64
2   pressure    768 non-null    int64
3   triceps     768 non-null    int64
4   insulin     768 non-null    int64
5   mass        768 non-null    float64
6   pedigree    768 non-null    float64
7   age         768 non-null    int64
8   diabetes    768 non-null    object
dtypes: float64(2), int64(6), object(1)
memory usage: 54.1+ KB
```

Decision tree

```
In [4]: df_X = df.loc[:, df.columns != 'diabetes']
df_y = df['diabetes']

# Split the data into training/testing sets
train_X, test_X, train_y, test_y = \
    train_test_split(df_X, df_y, test_size=0.3,\
                    random_state=1234)

# Define Learning model (basic) #####
model = DecisionTreeClassifier(random_state=1234)

# Train the model using the training sets
model.fit(train_X, train_y)

# performance evaluation
print('Train accuracy :', model.score(train_X, train_y))
print('Test accuracy :', model.score(test_X, test_y))

pred_y = model.predict(test_X)
confusion_matrix(test_y, pred_y)

# Define Learning model (tuening) #####
model = DecisionTreeClassifier(max_depth=4, random_state=1234)

# Train the model using the training sets
model.fit(train_X, train_y)

# performance evaluation
print('Train accuracy :', model.score(train_X, train_y))
print('Test accuracy :', model.score(test_X, test_y))

pred_y = model.predict(test_X)
confusion_matrix(test_y, pred_y)
```

```
Train accuracy : 1.0
Test accuracy : 0.7012987012987013
Train accuracy : 0.819366852886406
Test accuracy : 0.7359307359307359
```

```
Out[4]: array([[127, 20],
               [ 41, 43]], dtype=int64)
```

```
In [5]: from sklearn.tree import DecisionTreeClassifier
import optuna
from sklearn.metrics import confusion_matrix

def objective(trial):
    # 하이퍼파라미터 검색 공간 정의
    max_depth = trial.suggest_int('max_depth', 1, 150) # 최대 깊이
    min_samples_split = trial.suggest_int('min_samples_split', 2, 100) # 분할을
    min_samples_leaf = trial.suggest_int('min_samples_leaf', 1, 100) # 리프 노드
    criterion = trial.suggest_categorical('criterion', ['gini', 'entropy']) # 분

    # 모델 초기화
    model = DecisionTreeClassifier(max_depth=max_depth,
                                   min_samples_split=min_samples_split,
                                   min_samples_leaf=min_samples_leaf,
                                   criterion=criterion,
                                   random_state=1234)

    # 모델 학습
    model.fit(train_X, train_y)

    # 모델 평가 (여기서는 validation set이 없으므로 test set 사용)
    test_accuracy = model.score(test_X, test_y)

    return test_accuracy

# Optuna 스터디 생성
study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=100) # 100번의 시도로 최적화 수행

# 최적 하이퍼파라미터 출력
print('Best trial:')
trial = study.best_trial

print('  Value: {}'.format(trial.value))
print('  Parameters:')
for key, value in trial.params.items():
    print('    {}: {}'.format(key, value))

# 최적 하이퍼파라미터로 모델 학습
best_model = DecisionTreeClassifier(**trial.params, random_state=1234)
best_model.fit(train_X, train_y)

# 성능 평가
print('Train accuracy :', best_model.score(train_X, train_y))
print('Test accuracy :', best_model.score(test_X, test_y))

pred_y = best_model.predict(test_X)
print('Confusion Matrix:\n', confusion_matrix(test_y, pred_y))
```

```
C:\Users\jdh25\AppData\Local\Programs\Python\Python310\lib\site-packages\tqdm\auto.py:21: TqdmWarning: IProgress not found. Please update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html
  from .autonotebook import tqdm as notebook_tqdm
[I 2024-09-24 10:11:01,568] A new study created in memory with name: no-name-53224ce4-84d4-4661-9e07-8127ad1749c1
[I 2024-09-24 10:11:01,575] Trial 0 finished with value: 0.7316017316017316 and parameters: {'max_depth': 120, 'min_samples_split': 88, 'min_samples_leaf': 32, 'criterion': 'entropy'}. Best is trial 0 with value: 0.7316017316017316.
[I 2024-09-24 10:11:01,578] Trial 1 finished with value: 0.7402597402597403 and parameters: {'max_depth': 105, 'min_samples_split': 28, 'min_samples_leaf': 15, 'criterion': 'gini'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,591] Trial 2 finished with value: 0.7316017316017316 and parameters: {'max_depth': 101, 'min_samples_split': 59, 'min_samples_leaf': 18, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,598] Trial 3 finished with value: 0.7316017316017316 and parameters: {'max_depth': 99, 'min_samples_split': 100, 'min_samples_leaf': 49, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,607] Trial 4 finished with value: 0.7402597402597403 and parameters: {'max_depth': 52, 'min_samples_split': 26, 'min_samples_leaf': 39, 'criterion': 'gini'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,619] Trial 5 finished with value: 0.7142857142857143 and parameters: {'max_depth': 54, 'min_samples_split': 48, 'min_samples_leaf': 96, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,628] Trial 6 finished with value: 0.7142857142857143 and parameters: {'max_depth': 73, 'min_samples_split': 28, 'min_samples_leaf': 69, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,640] Trial 7 finished with value: 0.7142857142857143 and parameters: {'max_depth': 60, 'min_samples_split': 31, 'min_samples_leaf': 2, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,648] Trial 8 finished with value: 0.7142857142857143 and parameters: {'max_depth': 95, 'min_samples_split': 69, 'min_samples_leaf': 96, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,658] Trial 9 finished with value: 0.7142857142857143 and parameters: {'max_depth': 1, 'min_samples_split': 74, 'min_samples_leaf': 19, 'criterion': 'entropy'}. Best is trial 1 with value: 0.7402597402597403.
[I 2024-09-24 10:11:01,686] Trial 10 finished with value: 0.7489177489177489 and parameters: {'max_depth': 135, 'min_samples_split': 7, 'min_samples_leaf': 64, 'criterion': 'gini'}. Best is trial 10 with value: 0.7489177489177489.
[I 2024-09-24 10:11:01,708] Trial 11 finished with value: 0.7532467532467533 and parameters: {'max_depth': 149, 'min_samples_split': 4, 'min_samples_leaf': 67, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,736] Trial 12 finished with value: 0.7532467532467533 and parameters: {'max_depth': 150, 'min_samples_split': 2, 'min_samples_leaf': 69, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,758] Trial 13 finished with value: 0.7142857142857143 and parameters: {'max_depth': 150, 'min_samples_split': 4, 'min_samples_leaf': 75, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,783] Trial 14 finished with value: 0.7142857142857143 and parameters: {'max_depth': 149, 'min_samples_split': 14, 'min_samples_leaf': 77, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,798] Trial 15 finished with value: 0.7489177489177489 and parameters: {'max_depth': 120, 'min_samples_split': 45, 'min_samples_leaf': 58, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,829] Trial 16 finished with value: 0.7142857142857143 and parameters: {'max_depth': 16, 'min_samples_split': 18, 'min_samples_leaf': 83, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:01,849] Trial 17 finished with value: 0.7402597402597403 and parameters: {'max_depth': 129, 'min_samples_split': 2, 'min_samples_leaf': 50, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
```

[I 2024-09-24 10:11:01,868] Trial 18 finished with value: 0.7142857142857143 and parameters: {'max_depth': 140, 'min_samples_split': 39, 'min_samples_leaf': 81, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,888] Trial 19 finished with value: 0.7489177489177489 and parameters: {'max_depth': 83, 'min_samples_split': 15, 'min_samples_leaf': 59, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,903] Trial 20 finished with value: 0.7142857142857143 and parameters: {'max_depth': 121, 'min_samples_split': 19, 'min_samples_leaf': 86, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,927] Trial 21 finished with value: 0.7489177489177489 and parameters: {'max_depth': 136, 'min_samples_split': 7, 'min_samples_leaf': 64, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,948] Trial 22 finished with value: 0.7532467532467533 and parameters: {'max_depth': 150, 'min_samples_split': 10, 'min_samples_leaf': 68, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,968] Trial 23 finished with value: 0.7532467532467533 and parameters: {'max_depth': 150, 'min_samples_split': 12, 'min_samples_leaf': 69, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:01,988] Trial 24 finished with value: 0.7402597402597403 and parameters: {'max_depth': 115, 'min_samples_split': 38, 'min_samples_leaf': 42, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,008] Trial 25 finished with value: 0.7489177489177489 and parameters: {'max_depth': 35, 'min_samples_split': 2, 'min_samples_leaf': 59, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,028] Trial 26 finished with value: 0.7142857142857143 and parameters: {'max_depth': 131, 'min_samples_split': 22, 'min_samples_leaf': 91, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,049] Trial 27 finished with value: 0.7532467532467533 and parameters: {'max_depth': 140, 'min_samples_split': 10, 'min_samples_leaf': 72, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,068] Trial 28 finished with value: 0.7402597402597403 and parameters: {'max_depth': 110, 'min_samples_split': 34, 'min_samples_leaf': 45, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,088] Trial 29 finished with value: 0.7489177489177489 and parameters: {'max_depth': 122, 'min_samples_split': 58, 'min_samples_leaf': 55, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,108] Trial 30 finished with value: 0.7402597402597403 and parameters: {'max_depth': 142, 'min_samples_split': 85, 'min_samples_leaf': 30, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,131] Trial 31 finished with value: 0.7532467532467533 and parameters: {'max_depth': 149, 'min_samples_split': 12, 'min_samples_leaf': 68, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,154] Trial 32 finished with value: 0.7142857142857143 and parameters: {'max_depth': 132, 'min_samples_split': 23, 'min_samples_leaf': 76, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,174] Trial 33 finished with value: 0.7532467532467533 and parameters: {'max_depth': 148, 'min_samples_split': 10, 'min_samples_leaf': 66, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,198] Trial 34 finished with value: 0.7142857142857143 and parameters: {'max_depth': 127, 'min_samples_split': 16, 'min_samples_leaf': 89, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,225] Trial 35 finished with value: 0.7142857142857143 and parameters: {'max_depth': 109, 'min_samples_split': 8, 'min_samples_leaf': 81, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,249] Trial 36 finished with value: 0.7489177489177489 and parameters: {'max_depth': 143, 'min_samples_split': 23, 'min_samples_leaf': 55, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,269] Trial 37 finished with value: 0.7142857142857143 and parameters: {'max_depth': 85, 'min_samples_split': 4, 'min_samples_leaf': 72, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,295] Trial 38 finished with value: 0.7402597402597403 and parameters: {'max_depth': 125, 'min_samples_split': 13, 'min_samples_leaf': 33, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,339] Trial 39 finished with value: 0.7142857142857143 and parameters: {'max_depth': 101, 'min_samples_split': 28, 'min_samples_leaf': 100, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,389] Trial 40 finished with value: 0.7489177489177489 and parameters: {'max_depth': 142, 'min_samples_split': 56, 'min_samples_leaf': 64, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,428] Trial 41 finished with value: 0.7532467532467533 and parameters: {'max_depth': 139, 'min_samples_split': 10, 'min_samples_leaf': 72, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,479] Trial 42 finished with value: 0.7532467532467533 and parameters: {'max_depth': 135, 'min_samples_split': 18, 'min_samples_leaf': 71, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,573] Trial 43 finished with value: 0.7142857142857143 and parameters: {'max_depth': 150, 'min_samples_split': 7, 'min_samples_leaf': 78, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,662] Trial 44 finished with value: 0.7489177489177489 and parameters: {'max_depth': 145, 'min_samples_split': 2, 'min_samples_leaf': 63, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,738] Trial 45 finished with value: 0.7316017316017316 and parameters: {'max_depth': 137, 'min_samples_split': 11, 'min_samples_leaf': 54, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,788] Trial 46 finished with value: 0.7532467532467533 and parameters: {'max_depth': 67, 'min_samples_split': 100, 'min_samples_leaf': 69, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,830] Trial 47 finished with value: 0.7012987012987013 and parameters: {'max_depth': 115, 'min_samples_split': 21, 'min_samples_leaf': 2, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,874] Trial 48 finished with value: 0.7316017316017316 and parameters: {'max_depth': 93, 'min_samples_split': 27, 'min_samples_leaf': 61, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,898] Trial 49 finished with value: 0.7402597402597403 and parameters: {'max_depth': 128, 'min_samples_split': 71, 'min_samples_leaf': 47, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,938] Trial 50 finished with value: 0.7142857142857143 and parameters: {'max_depth': 48, 'min_samples_split': 64, 'min_samples_leaf': 75, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,969] Trial 51 finished with value: 0.7532467532467533 and parameters: {'max_depth': 146, 'min_samples_split': 13, 'min_samples_leaf': 68, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:02,999] Trial 52 finished with value: 0.7142857142857143 and parameters: {'max_depth': 148, 'min_samples_split': 7, 'min_samples_leaf': 84, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,031] Trial 53 finished with value: 0.7142857142857143 and parameters: {'max_depth': 150, 'min_samples_split': 16, 'min_samples_leaf': 79, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,068] Trial 54 finished with value: 0.7532467532467533 and parameters: {'max_depth': 139, 'min_samples_split': 32, 'min_samples_leaf': 73, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,138] Trial 55 finished with value: 0.7532467532467533 and parameters: {'max_depth': 132, 'min_samples_split': 5, 'min_samples_leaf': 66, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,196] Trial 56 finished with value: 0.7489177489177489 and parameters: {'max_depth': 142, 'min_samples_split': 12, 'min_samples_leaf': 53, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,246] Trial 57 finished with value: 0.7489177489177489 and parameters: {'max_depth': 134, 'min_samples_split': 80, 'min_samples_leaf': 59, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,302] Trial 58 finished with value: 0.7142857142857143 and parameters: {'max_depth': 116, 'min_samples_split': 45, 'min_samples_leaf': 88, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,344] Trial 59 finished with value: 0.7229437229437229 and parameters: {'max_depth': 145, 'min_samples_split': 92, 'min_samples_leaf': 68, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,386] Trial 60 finished with value: 0.7142857142857143 and parameters: {'max_depth': 124, 'min_samples_split': 19, 'min_samples_leaf': 7, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,428] Trial 61 finished with value: 0.7532467532467533 and parameters: {'max_depth': 150, 'min_samples_split': 10, 'min_samples_leaf': 66, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,472] Trial 62 finished with value: 0.7229437229437229 and parameters: {'max_depth': 144, 'min_samples_split': 9, 'min_samples_leaf': 74, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,511] Trial 63 finished with value: 0.7532467532467533 and parameters: {'max_depth': 137, 'min_samples_split': 5, 'min_samples_leaf': 70, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,559] Trial 64 finished with value: 0.7489177489177489 and parameters: {'max_depth': 150, 'min_samples_split': 15, 'min_samples_leaf': 62, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,608] Trial 65 finished with value: 0.7142857142857143 and parameters: {'max_depth': 138, 'min_samples_split': 3, 'min_samples_leaf': 81, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,681] Trial 66 finished with value: 0.7532467532467533 and parameters: {'max_depth': 131, 'min_samples_split': 11, 'min_samples_leaf': 66, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,731] Trial 67 finished with value: 0.7489177489177489 and parameters: {'max_depth': 20, 'min_samples_split': 25, 'min_samples_leaf': 57, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,768] Trial 68 finished with value: 0.7402597402597403 and parameters: {'max_depth': 145, 'min_samples_split': 6, 'min_samples_leaf': 50, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,811] Trial 69 finished with value: 0.7142857142857143 and parameters: {'max_depth': 139, 'min_samples_split': 18, 'min_samples_leaf': 77, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,841] Trial 70 finished with value: 0.7489177489177489 and parameters: {'max_depth': 128, 'min_samples_split': 2, 'min_samples_leaf': 61, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,866] Trial 71 finished with value: 0.7532467532467533 and parameters: {'max_depth': 141, 'min_samples_split': 10, 'min_samples_leaf': 71, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,898] Trial 72 finished with value: 0.7532467532467533 and parameters: {'max_depth': 146, 'min_samples_split': 14, 'min_samples_leaf': 73, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,946] Trial 73 finished with value: 0.7532467532467533 and parameters: {'max_depth': 136, 'min_samples_split': 9, 'min_samples_leaf': 66, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:03,979] Trial 74 finished with value: 0.7142857142857143 and parameters: {'max_depth': 140, 'min_samples_split': 21, 'min_samples_leaf': 84, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,027] Trial 75 finished with value: 0.7532467532467533 and parameters: {'max_depth': 146, 'min_samples_split': 6, 'min_samples_leaf': 69, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,060] Trial 76 finished with value: 0.7056277056277056 and parameters: {'max_depth': 132, 'min_samples_split': 16, 'min_samples_leaf': 80, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,099] Trial 77 finished with value: 0.7142857142857143 and parameters: {'max_depth': 150, 'min_samples_split': 12, 'min_samples_leaf': 75, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,149] Trial 78 finished with value: 0.7489177489177489 and parameters: {'max_depth': 142, 'min_samples_split': 8, 'min_samples_leaf': 63, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,189] Trial 79 finished with value: 0.7489177489177489 and parameters: {'max_depth': 120, 'min_samples_split': 24, 'min_samples_leaf': 58, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,242] Trial 80 finished with value: 0.7142857142857143 and parameters: {'max_depth': 135, 'min_samples_split': 4, 'min_samples_leaf': 93, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,283] Trial 81 finished with value: 0.7532467532467533 and parameters: {'max_depth': 146, 'min_samples_split': 17, 'min_samples_leaf': 71, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,310] Trial 82 finished with value: 0.7142857142857143 and parameters: {'max_depth': 135, 'min_samples_split': 20, 'min_samples_leaf': 77, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,339] Trial 83 finished with value: 0.7532467532467533 and parameters: {'max_depth': 141, 'min_samples_split': 13, 'min_samples_leaf': 72, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,369] Trial 84 finished with value: 0.7489177489177489 and parameters: {'max_depth': 147, 'min_samples_split': 9, 'min_samples_leaf': 64, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,402] Trial 85 finished with value: 0.7532467532467533 and parameters: {'max_depth': 130, 'min_samples_split': 5, 'min_samples_leaf': 68, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,429] Trial 86 finished with value: 0.7142857142857143 and parameters: {'max_depth': 125, 'min_samples_split': 2, 'min_samples_leaf': 75, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,453] Trial 87 finished with value: 0.7316017316017316 and parameters: {'max_depth': 143, 'min_samples_split': 14, 'min_samples_leaf': 61, 'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,481] Trial 88 finished with value: 0.7142857142857143 and parameters: {'max_depth': 138, 'min_samples_split': 38, 'min_samples_leaf': 82, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,510] Trial 89 finished with value: 0.7532467532467533 and parameters: {'max_depth': 134, 'min_samples_split': 7, 'min_samples_leaf': 67, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,540] Trial 90 finished with value: 0.7532467532467533 and parameters: {'max_depth': 148, 'min_samples_split': 18, 'min_samples_leaf': 71, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,589] Trial 91 finished with value: 0.7532467532467533 and parameters: {'max_depth': 68, 'min_samples_split': 100, 'min_samples_leaf': 69, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,653] Trial 92 finished with value: 0.7532467532467533 and parameters: {'max_depth': 64, 'min_samples_split': 12, 'min_samples_leaf': 73, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,717] Trial 93 finished with value: 0.7489177489177489 and parameters: {'max_depth': 78, 'min_samples_split': 52, 'min_samples_leaf': 65, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,848] Trial 94 finished with value: 0.7532467532467533 and parameters: {'max_depth': 58, 'min_samples_split': 94, 'min_samples_leaf': 70, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,916] Trial 95 finished with value: 0.7142857142857143 and parameters: {'max_depth': 75, 'min_samples_split': 66, 'min_samples_leaf': 77, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:04,959] Trial 96 finished with value: 0.7489177489177489 and parameters: {'max_depth': 54, 'min_samples_split': 10, 'min_samples_leaf': 60, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.

[I 2024-09-24 10:11:05,018] Trial 97 finished with value: 0.7489177489177489 and parameters: {'max_depth': 144, 'min_samples_split': 81, 'min_samples_leaf': 56, 'criterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.


```
[I 2024-09-24 10:11:05,061] Trial 98 finished with value: 0.7489177489177489 and
parameters: {'max_depth': 89, 'min_samples_split': 15, 'min_samples_leaf': 64, 'c
riterion': 'gini'}. Best is trial 11 with value: 0.7532467532467533.
[I 2024-09-24 10:11:05,115] Trial 99 finished with value: 0.7056277056277056 and
parameters: {'max_depth': 148, 'min_samples_split': 44, 'min_samples_leaf': 79,
'criterion': 'entropy'}. Best is trial 11 with value: 0.7532467532467533.
```

Best trial:

Value: 0.7532467532467533

Parameters:

max_depth: 149

min_samples_split: 4

min_samples_leaf: 67

criterion: gini

Train accuracy : 0.7821229050279329

Test accuracy : 0.7532467532467533

Confusion Matrix:

```
[[121  26]
```

```
[ 31  53]]
```

Decision Tree Result

하이퍼파라미터 튜닝 전: Train accuracy : 0.819366852886406 Test accuracy :
0.7359307359307359

하이퍼파라미터 튜닝 후: Train accuracy : 0.7821229050279329 Test accuracy :
0.7532467532467533

SVM

```
In [6]: from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
import pandas as pd

# Define Learning model (basic) #####
model = svm.SVC()
# Train the model using the training sets
model.fit(train_X, train_y)

# performance evaluation
print('Train accuracy :', model.score(train_X, train_y))
print('Test accuracy :', model.score(test_X, test_y))

pred_y = model.predict(test_X)
confusion_matrix(test_y, pred_y)

# Define Learning model (poly kernel) #####
model = svm.SVC(kernel='poly')

# Train the model using the training sets
model.fit(train_X, train_y)

# performance evaluation
print('Train accuracy :', model.score(train_X, train_y))
print('Test accuracy :', model.score(test_X, test_y))
```

```
pred_y = model.predict(test_X)
confusion_matrix(test_y, pred_y)
```

```
Train accuracy : 0.7746741154562383
Test accuracy : 0.7402597402597403
Train accuracy : 0.7821229050279329
Test accuracy : 0.7229437229437229
```

```
Out[6]: array([[138,  9],
               [ 55, 29]], dtype=int64)
```

SVM Result

하이퍼파라미터 튜닝 전: Train accuracy : 0.7821229050279329 Test accuracy : 0.7229437229437229

하이퍼파라미터 튜닝 후: Train accuracy : 0.7821229050279329 Test accuracy : 0.7532467532467533

RF

```
In [7]: from sklearn.ensemble import RandomForestClassifier
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import confusion_matrix
        import pandas as pd

        # Define Learning model (# of tree: 10) #####
        model = RandomForestClassifier(n_estimators=10, random_state=1234)
        # Train the model using the training sets
        model.fit(train_X, train_y)

        # performance evaluation
        print('Train accuracy :', model.score(train_X, train_y))
        print('Test accuracy :', model.score(test_X, test_y))

        pred_y = model.predict(test_X)
        confusion_matrix(test_y, pred_y)

        # Define Learning model ((# of tree: 50) #####
        model = RandomForestClassifier(n_estimators=50, random_state=1234)

        # Train the model using the training sets
        model.fit(train_X, train_y)

        # performance evaluation
        print('Train accuracy :', model.score(train_X, train_y))
        print('Test accuracy :', model.score(test_X, test_y))

        pred_y = model.predict(test_X)
        confusion_matrix(test_y, pred_y)
```

```
Train accuracy : 0.9851024208566108
Test accuracy : 0.7272727272727273
Train accuracy : 0.9981378026070763
Test accuracy : 0.7532467532467533
```

```
Out[7]: array([[129, 18],
               [ 39, 45]], dtype=int64)
```

```

In [8]: def objective(trial):
# 하이퍼파라미터 검색 공간 정의
n_estimators = trial.suggest_int('n_estimators', 10, 500)
max_depth = trial.suggest_int('max_depth', 1, 150)
min_samples_split = trial.suggest_int('min_samples_split', 2, 100)
min_samples_leaf = trial.suggest_int('min_samples_leaf', 1, 100)
criterion = trial.suggest_categorical('criterion', ['gini', 'entropy'])

# 모델 초기화
model = RandomForestClassifier(n_estimators=n_estimators,
                              max_depth=max_depth,
                              min_samples_split=min_samples_split,
                              min_samples_leaf=min_samples_leaf,
                              criterion=criterion,
                              random_state=1234)

# 모델 학습
model.fit(train_X, train_y)

# 모델 평가 (여기서는 validation set이 없으므로 test set 사용)
test_accuracy = model.score(test_X, test_y)

return test_accuracy

# Optuna 스터디 생성
study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=100) # 100번의 시도로 최적화 수행

# 최적 하이퍼파라미터 출력
print('Best trial:')
trial = study.best_trial

print('  Value: {}'.format(trial.value))
print('  Parameters:')
for key, value in trial.params.items():
    print('    {}: {}'.format(key, value))

# 최적 하이퍼파라미터로 모델 학습
best_model = RandomForestClassifier(**trial.params, random_state=1234)
best_model.fit(train_X, train_y)

# 성능 평가
print('Train accuracy :', best_model.score(train_X, train_y))
print('Test accuracy :', best_model.score(test_X, test_y))

pred_y = best_model.predict(test_X)
print('Confusion Matrix:\n', confusion_matrix(test_y, pred_y))

```

```
[I 2024-09-24 10:11:05,540] A new study created in memory with name: no-name-13ae
a73f-cad4-422f-bff1-f277527bd5d6
[I 2024-09-24 10:11:06,078] Trial 0 finished with value: 0.7186147186147186 and p
arameters: {'n_estimators': 359, 'max_depth': 98, 'min_samples_split': 86, 'min_s
amples_leaf': 49, 'criterion': 'entropy'}. Best is trial 0 with value: 0.71861471
86147186.
[I 2024-09-24 10:11:06,458] Trial 1 finished with value: 0.7142857142857143 and p
arameters: {'n_estimators': 213, 'max_depth': 87, 'min_samples_split': 42, 'min_s
amples_leaf': 63, 'criterion': 'gini'}. Best is trial 0 with value: 0.71861471861
47186.
[I 2024-09-24 10:11:06,600] Trial 2 finished with value: 0.7142857142857143 and p
arameters: {'n_estimators': 90, 'max_depth': 102, 'min_samples_split': 92, 'min_s
amples_leaf': 24, 'criterion': 'entropy'}. Best is trial 0 with value: 0.71861471
86147186.
[I 2024-09-24 10:11:07,399] Trial 3 finished with value: 0.7316017316017316 and p
arameters: {'n_estimators': 420, 'max_depth': 138, 'min_samples_split': 57, 'min_
samples_leaf': 33, 'criterion': 'gini'}. Best is trial 3 with value: 0.7316017316
017316.
[I 2024-09-24 10:11:07,599] Trial 4 finished with value: 0.7142857142857143 and p
arameters: {'n_estimators': 125, 'max_depth': 127, 'min_samples_split': 80, 'min_
samples_leaf': 73, 'criterion': 'gini'}. Best is trial 3 with value: 0.7316017316
017316.
[I 2024-09-24 10:11:08,283] Trial 5 finished with value: 0.7575757575757576 and p
arameters: {'n_estimators': 281, 'max_depth': 87, 'min_samples_split': 20, 'min_s
amples_leaf': 5, 'criterion': 'entropy'}. Best is trial 5 with value: 0.757575757
5757576.
[I 2024-09-24 10:11:08,890] Trial 6 finished with value: 0.7272727272727273 and p
arameters: {'n_estimators': 407, 'max_depth': 111, 'min_samples_split': 99, 'min_
samples_leaf': 49, 'criterion': 'gini'}. Best is trial 5 with value: 0.7575757575
757576.
[I 2024-09-24 10:11:09,327] Trial 7 finished with value: 0.7272727272727273 and p
arameters: {'n_estimators': 273, 'max_depth': 121, 'min_samples_split': 39, 'min_
samples_leaf': 34, 'criterion': 'entropy'}. Best is trial 5 with value: 0.7575757
5757576.
[I 2024-09-24 10:11:09,388] Trial 8 finished with value: 0.7142857142857143 and p
arameters: {'n_estimators': 30, 'max_depth': 56, 'min_samples_split': 61, 'min_sa
mples_leaf': 56, 'criterion': 'entropy'}. Best is trial 5 with value: 0.757575757
5757576.
[I 2024-09-24 10:11:09,988] Trial 9 finished with value: 0.696969696969697 and pa
rameters: {'n_estimators': 425, 'max_depth': 63, 'min_samples_split': 33, 'min_sa
mples_leaf': 98, 'criterion': 'gini'}. Best is trial 5 with value: 0.757575757575
7576.
[I 2024-09-24 10:11:10,379] Trial 10 finished with value: 0.7662337662337663 and
parameters: {'n_estimators': 265, 'max_depth': 15, 'min_samples_split': 2, 'min_s
amples_leaf': 7, 'criterion': 'entropy'}. Best is trial 10 with value: 0.76623376
62337663.
[I 2024-09-24 10:11:10,868] Trial 11 finished with value: 0.7575757575757576 and
parameters: {'n_estimators': 278, 'max_depth': 7, 'min_samples_split': 2, 'min_sa
mples_leaf': 2, 'criterion': 'entropy'}. Best is trial 10 with value: 0.766233766
2337663.
[I 2024-09-24 10:11:11,258] Trial 12 finished with value: 0.7619047619047619 and
parameters: {'n_estimators': 211, 'max_depth': 32, 'min_samples_split': 5, 'min_s
amples_leaf': 1, 'criterion': 'entropy'}. Best is trial 10 with value: 0.76623376
62337663.
[I 2024-09-24 10:11:11,538] Trial 13 finished with value: 0.7489177489177489 and
parameters: {'n_estimators': 191, 'max_depth': 15, 'min_samples_split': 2, 'min_s
amples_leaf': 16, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337
662337663.
[I 2024-09-24 10:11:12,049] Trial 14 finished with value: 0.7402597402597403 and
parameters: {'n_estimators': 343, 'max_depth': 36, 'min_samples_split': 17, 'min_
```

```

samples_leaf': 16, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:12,368] Trial 15 finished with value: 0.7662337662337663 and
parameters: {'n_estimators': 189, 'max_depth': 34, 'min_samples_split': 18, 'min_
samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:12,578] Trial 16 finished with value: 0.7402597402597403 and
parameters: {'n_estimators': 138, 'max_depth': 33, 'min_samples_split': 21, 'min_
samples_leaf': 16, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:13,240] Trial 17 finished with value: 0.658008658008658 and p
arameters: {'n_estimators': 499, 'max_depth': 1, 'min_samples_split': 28, 'min_sa
mples_leaf': 34, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:13,459] Trial 18 finished with value: 0.70995670995671 and pa
rameters: {'n_estimators': 162, 'max_depth': 53, 'min_samples_split': 12, 'min_sa
mples_leaf': 85, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:13,618] Trial 19 finished with value: 0.7402597402597403 and
parameters: {'n_estimators': 76, 'max_depth': 28, 'min_samples_split': 47, 'min_s
amples_leaf': 11, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:14,138] Trial 20 finished with value: 0.7272727272727273 and
parameters: {'n_estimators': 328, 'max_depth': 18, 'min_samples_split': 66, 'min_
samples_leaf': 25, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:14,582] Trial 21 finished with value: 0.7619047619047619 and
parameters: {'n_estimators': 217, 'max_depth': 42, 'min_samples_split': 10, 'min_
samples_leaf': 2, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:15,038] Trial 22 finished with value: 0.7619047619047619 and
parameters: {'n_estimators': 231, 'max_depth': 21, 'min_samples_split': 11, 'min_
samples_leaf': 9, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:15,319] Trial 23 finished with value: 0.7445887445887446 and
parameters: {'n_estimators': 173, 'max_depth': 71, 'min_samples_split': 28, 'min_
samples_leaf': 23, 'criterion': 'entropy'}. Best is trial 10 with value: 0.7662337662337663.
[I 2024-09-24 10:11:15,858] Trial 24 finished with value: 0.7705627705627706 and
parameters: {'n_estimators': 252, 'max_depth': 47, 'min_samples_split': 2, 'min_s
amples_leaf': 8, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.
[I 2024-09-24 10:11:16,443] Trial 25 finished with value: 0.7489177489177489 and
parameters: {'n_estimators': 302, 'max_depth': 47, 'min_samples_split': 17, 'min_
samples_leaf': 13, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.
[I 2024-09-24 10:11:16,833] Trial 26 finished with value: 0.7272727272727273 and
parameters: {'n_estimators': 246, 'max_depth': 69, 'min_samples_split': 27, 'min_
samples_leaf': 38, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.
[I 2024-09-24 10:11:17,529] Trial 27 finished with value: 0.7445887445887446 and
parameters: {'n_estimators': 381, 'max_depth': 45, 'min_samples_split': 9, 'min_s
amples_leaf': 22, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.
[I 2024-09-24 10:11:18,043] Trial 28 finished with value: 0.7272727272727273 and
parameters: {'n_estimators': 306, 'max_depth': 14, 'min_samples_split': 15, 'min_
samples_leaf': 43, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.
[I 2024-09-24 10:11:18,440] Trial 29 finished with value: 0.7402597402597403 and
parameters: {'n_estimators': 249, 'max_depth': 80, 'min_samples_split': 23, 'min_

```


samples_leaf': 9, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:19,058] Trial 30 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 364, 'max_depth': 25, 'min_samples_split': 34, 'min_samples_leaf': 19, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:19,407] Trial 31 finished with value: 0.7575757575757576 and parameters: {'n_estimators': 146, 'max_depth': 34, 'min_samples_split': 2, 'min_samples_leaf': 4, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:19,948] Trial 32 finished with value: 0.7662337662337663 and parameters: {'n_estimators': 197, 'max_depth': 56, 'min_samples_split': 6, 'min_samples_leaf': 2, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:20,188] Trial 33 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 101, 'max_depth': 62, 'min_samples_split': 8, 'min_samples_leaf': 30, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:20,608] Trial 34 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 192, 'max_depth': 42, 'min_samples_split': 8, 'min_samples_leaf': 8, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:20,989] Trial 35 finished with value: 0.70995670995671 and parameters: {'n_estimators': 197, 'max_depth': 53, 'min_samples_split': 75, 'min_samples_leaf': 62, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:21,358] Trial 36 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 232, 'max_depth': 79, 'min_samples_split': 15, 'min_samples_leaf': 26, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:21,599] Trial 37 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 112, 'max_depth': 94, 'min_samples_split': 52, 'min_samples_leaf': 9, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:21,978] Trial 38 finished with value: 0.7056277056277056 and parameters: {'n_estimators': 266, 'max_depth': 144, 'min_samples_split': 7, 'min_samples_leaf': 74, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:22,078] Trial 39 finished with value: 0.7186147186147186 and parameters: {'n_estimators': 62, 'max_depth': 9, 'min_samples_split': 23, 'min_samples_leaf': 42, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:22,878] Trial 40 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 309, 'max_depth': 25, 'min_samples_split': 14, 'min_samples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:23,298] Trial 41 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 214, 'max_depth': 36, 'min_samples_split': 2, 'min_samples_leaf': 4, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:23,591] Trial 42 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 173, 'max_depth': 60, 'min_samples_split': 6, 'min_samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:23,958] Trial 43 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 255, 'max_depth': 48, 'min_samples_split': 5, 'min_samples_leaf': 13, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:24,618] Trial 44 finished with value: 0.7575757575757576 and parameters: {'n_estimators': 281, 'max_depth': 29, 'min_samples_split': 17, 'min_

samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:24,979] Trial 45 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 203, 'max_depth': 40, 'min_samples_split': 38, 'min_samples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:25,316] Trial 46 finished with value: 0.7229437229437229 and parameters: {'n_estimators': 157, 'max_depth': 3, 'min_samples_split': 5, 'min_samples_leaf': 19, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:25,618] Trial 47 finished with value: 0.7359307359307359 and parameters: {'n_estimators': 128, 'max_depth': 13, 'min_samples_split': 21, 'min_samples_leaf': 16, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:26,168] Trial 48 finished with value: 0.7229437229437229 and parameters: {'n_estimators': 223, 'max_depth': 50, 'min_samples_split': 95, 'min_samples_leaf': 29, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:26,791] Trial 49 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 290, 'max_depth': 59, 'min_samples_split': 12, 'min_samples_leaf': 14, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:27,164] Trial 50 finished with value: 0.70995670995671 and parameters: {'n_estimators': 184, 'max_depth': 68, 'min_samples_split': 3, 'min_samples_leaf': 54, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:27,533] Trial 51 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 214, 'max_depth': 31, 'min_samples_split': 11, 'min_samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:28,022] Trial 52 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 231, 'max_depth': 38, 'min_samples_split': 8, 'min_samples_leaf': 5, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:28,450] Trial 53 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 265, 'max_depth': 44, 'min_samples_split': 19, 'min_samples_leaf': 11, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:29,004] Trial 54 finished with value: 0.7575757575757576 and parameters: {'n_estimators': 244, 'max_depth': 21, 'min_samples_split': 11, 'min_samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:29,545] Trial 55 finished with value: 0.7575757575757576 and parameters: {'n_estimators': 333, 'max_depth': 56, 'min_samples_split': 26, 'min_samples_leaf': 8, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:29,861] Trial 56 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 166, 'max_depth': 22, 'min_samples_split': 14, 'min_samples_leaf': 18, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:30,208] Trial 57 finished with value: 0.7012987012987013 and parameters: {'n_estimators': 204, 'max_depth': 128, 'min_samples_split': 5, 'min_samples_leaf': 94, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:30,443] Trial 58 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 144, 'max_depth': 31, 'min_samples_split': 46, 'min_samples_leaf': 5, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:30,698] Trial 59 finished with value: 0.7229437229437229 and parameters: {'n_estimators': 184, 'max_depth': 42, 'min_samples_split': 86, 'min_

samples_leaf': 12, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:31,019] Trial 60 finished with value: 0.7359307359307359 and parameters: {'n_estimators': 238, 'max_depth': 8, 'min_samples_split': 10, 'min_samples_leaf': 21, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:31,348] Trial 61 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 221, 'max_depth': 20, 'min_samples_split': 13, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:31,708] Trial 62 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 258, 'max_depth': 15, 'min_samples_split': 2, 'min_samples_leaf': 11, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:32,147] Trial 63 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 287, 'max_depth': 36, 'min_samples_split': 17, 'min_samples_leaf': 4, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:32,489] Trial 64 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 208, 'max_depth': 26, 'min_samples_split': 31, 'min_samples_leaf': 10, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:32,868] Trial 65 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 234, 'max_depth': 52, 'min_samples_split': 9, 'min_samples_leaf': 16, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:33,200] Trial 66 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 178, 'max_depth': 18, 'min_samples_split': 6, 'min_samples_leaf': 3, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:33,698] Trial 67 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 272, 'max_depth': 65, 'min_samples_split': 19, 'min_samples_leaf': 8, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:34,420] Trial 68 finished with value: 0.7359307359307359 and parameters: {'n_estimators': 477, 'max_depth': 48, 'min_samples_split': 65, 'min_samples_leaf': 14, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:34,689] Trial 69 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 156, 'max_depth': 32, 'min_samples_split': 23, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:35,268] Trial 70 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 304, 'max_depth': 74, 'min_samples_split': 4, 'min_samples_leaf': 3, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:35,798] Trial 71 finished with value: 0.7705627705627706 and parameters: {'n_estimators': 321, 'max_depth': 23, 'min_samples_split': 14, 'min_samples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:36,338] Trial 72 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 324, 'max_depth': 25, 'min_samples_split': 10, 'min_samples_leaf': 10, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:37,038] Trial 73 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 361, 'max_depth': 10, 'min_samples_split': 17, 'min_samples_leaf': 4, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:37,323] Trial 74 finished with value: 0.7012987012987013 and parameters: {'n_estimators': 194, 'max_depth': 39, 'min_samples_split': 7, 'min_s

amples_leaf': 76, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:37,668] Trial 75 finished with value: 0.7186147186147186 and parameters: {'n_estimators': 245, 'max_depth': 3, 'min_samples_split': 14, 'min_samples_leaf': 1, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:38,154] Trial 76 finished with value: 0.7489177489177489 and parameters: {'n_estimators': 317, 'max_depth': 35, 'min_samples_split': 25, 'min_samples_leaf': 13, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:38,768] Trial 77 finished with value: 0.7705627705627706 and parameters: {'n_estimators': 389, 'max_depth': 28, 'min_samples_split': 2, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:39,422] Trial 78 finished with value: 0.7662337662337663 and parameters: {'n_estimators': 391, 'max_depth': 44, 'min_samples_split': 4, 'min_samples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:40,030] Trial 79 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 398, 'max_depth': 46, 'min_samples_split': 2, 'min_samples_leaf': 20, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:40,720] Trial 80 finished with value: 0.7445887445887446 and parameters: {'n_estimators': 345, 'max_depth': 29, 'min_samples_split': 6, 'min_samples_leaf': 16, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:41,519] Trial 81 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 436, 'max_depth': 57, 'min_samples_split': 4, 'min_samples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:42,119] Trial 82 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 377, 'max_depth': 39, 'min_samples_split': 9, 'min_samples_leaf': 9, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:42,859] Trial 83 finished with value: 0.7619047619047619 and parameters: {'n_estimators': 436, 'max_depth': 43, 'min_samples_split': 12, 'min_samples_leaf': 3, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:43,518] Trial 84 finished with value: 0.7705627705627706 and parameters: {'n_estimators': 393, 'max_depth': 17, 'min_samples_split': 7, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:44,103] Trial 85 finished with value: 0.7186147186147186 and parameters: {'n_estimators': 404, 'max_depth': 16, 'min_samples_split': 4, 'min_samples_leaf': 65, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:44,707] Trial 86 finished with value: 0.7705627705627706 and parameters: {'n_estimators': 380, 'max_depth': 106, 'min_samples_split': 7, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:45,428] Trial 87 finished with value: 0.7532467532467533 and parameters: {'n_estimators': 391, 'max_depth': 110, 'min_samples_split': 7, 'min_samples_leaf': 13, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:45,988] Trial 88 finished with value: 0.7402597402597403 and parameters: {'n_estimators': 383, 'max_depth': 86, 'min_samples_split': 16, 'min_samples_leaf': 24, 'criterion': 'gini'}. Best is trial 24 with value: 0.7705627705627706.

[I 2024-09-24 10:11:46,846] Trial 89 finished with value: 0.7705627705627706 and parameters: {'n_estimators': 417, 'max_depth': 97, 'min_samples_split': 2, 'min_s

```

amples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.77056277
05627706.
[I 2024-09-24 10:11:47,428] Trial 90 finished with value: 0.7402597402597403 and
parameters: {'n_estimators': 429, 'max_depth': 111, 'min_samples_split': 2, 'min_
samples_leaf': 18, 'criterion': 'entropy'}. Best is trial 24 with value: 0.770562
7705627706.
[I 2024-09-24 10:11:48,221] Trial 91 finished with value: 0.7619047619047619 and
parameters: {'n_estimators': 460, 'max_depth': 99, 'min_samples_split': 8, 'min_s
amples_leaf': 6, 'criterion': 'entropy'}. Best is trial 24 with value: 0.77056277
05627706.
[I 2024-09-24 10:11:49,092] Trial 92 finished with value: 0.7532467532467533 and
parameters: {'n_estimators': 414, 'max_depth': 115, 'min_samples_split': 5, 'min_
samples_leaf': 10, 'criterion': 'entropy'}. Best is trial 24 with value: 0.770562
7705627706.
[I 2024-09-24 10:11:49,813] Trial 93 finished with value: 0.7662337662337663 and
parameters: {'n_estimators': 349, 'max_depth': 89, 'min_samples_split': 11, 'min_
samples_leaf': 8, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627
705627706.
[I 2024-09-24 10:11:50,398] Trial 94 finished with value: 0.7359307359307359 and
parameters: {'n_estimators': 372, 'max_depth': 125, 'min_samples_split': 57, 'min
_samples_leaf': 5, 'criterion': 'entropy'}. Best is trial 24 with value: 0.770562
7705627706.
[I 2024-09-24 10:11:51,001] Trial 95 finished with value: 0.7445887445887446 and
parameters: {'n_estimators': 419, 'max_depth': 102, 'min_samples_split': 4, 'min_
samples_leaf': 15, 'criterion': 'entropy'}. Best is trial 24 with value: 0.770562
7705627706.
[I 2024-09-24 10:11:51,688] Trial 96 finished with value: 0.7489177489177489 and
parameters: {'n_estimators': 394, 'max_depth': 11, 'min_samples_split': 13, 'min_
samples_leaf': 12, 'criterion': 'entropy'}. Best is trial 24 with value: 0.770562
7705627706.
[I 2024-09-24 10:11:52,568] Trial 97 finished with value: 0.7619047619047619 and
parameters: {'n_estimators': 445, 'max_depth': 103, 'min_samples_split': 8, 'min_
samples_leaf': 3, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627
705627706.
[I 2024-09-24 10:11:53,332] Trial 98 finished with value: 0.7272727272727273 and
parameters: {'n_estimators': 353, 'max_depth': 23, 'min_samples_split': 75, 'min_
samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627
705627706.
[I 2024-09-24 10:11:54,008] Trial 99 finished with value: 0.7489177489177489 and
parameters: {'n_estimators': 336, 'max_depth': 27, 'min_samples_split': 2, 'min_s
amples_leaf': 11, 'criterion': 'entropy'}. Best is trial 24 with value: 0.7705627
705627706.

```

Best trial:

Value: 0.7705627705627706

Parameters:

```

n_estimators: 252
max_depth: 47
min_samples_split: 2
min_samples_leaf: 8
criterion: entropy

```

Train accuracy : 0.8640595903165735

Test accuracy : 0.7705627705627706

Confusion Matrix:

```

[[134  13]
 [ 40  44]]

```

RF Result

하이퍼파라미터 튜닝 전: Train accuracy : 0.9981378026070763 Test accuracy : 0.7532467532467533

하이퍼파라미터 튜닝 후: Train accuracy : 0.8640595903165735 Test accuracy : 0.7705627705627706

XGBoost

```
In [9]: import xgboost as xgb
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np

# df_y를 0,1로 변환
df_y_exchanged = df_y.replace('pos', 1)
df_y_exchanged = df_y_exchanged.replace('neg', 0)

# Split the data into training/testing sets
train_X, test_X, train_y, test_y = \
    train_test_split(df_X, df_y_exchanged, test_size=0.3,\
                    random_state=1234)

D_train = xgb.DMatrix(train_X, label=train_y)
D_test = xgb.DMatrix(test_X, label=test_y)

# Define & train Learning model #####
param = {
    'eta': 0.2,
    'max_depth': 3,
    'objective': 'binary:logistic',
    'eval_metric': 'error'}

steps = 20 # The number of training iterations

model = xgb.train(param, D_train, steps)

pred = model.predict(D_test)
round_preds = np.round(pred) # real -> [0,1]

from sklearn.metrics import accuracy_score
print('Test accuracy :', accuracy_score(test_y, round_preds))
```

Test accuracy : 0.7272727272727273

C:\Users\jdh25\AppData\Local\Temp\ipykernel_7620\119717568.py:8: FutureWarning: D owncasting behavior in `replace` is deprecated and will be removed in a future ve rsion. To retain the old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to the future behavior, set `pd.set_option('future.no_silent_down casting', True)`
df_y_exchanged = df_y_exchanged.replace('neg', 0)

```
In [ ]: import xgboost as xgb
import optuna
from sklearn.metrics import confusion_matrix

def objective(trial):
    # 하이퍼파라미터 검색 공간 정의
    n_estimators = trial.suggest_int('n_estimators', 50, 500) # 부스팅할 트리의
```

```

max_depth = trial.suggest_int('max_depth', 1, 15) # 트리의 최대 깊이
learning_rate = trial.suggest_loguniform('learning_rate', 1e-3, 0.3) # 학습
min_child_weight = trial.suggest_int('min_child_weight', 1, 10) # 자식 노드
gamma = trial.suggest_loguniform('gamma', 1e-3, 10.0) # 트리를 가지치기할 초
subsample = trial.suggest_uniform('subsample', 0.5, 1.0) # 각 트리를 작성하
colsample_bytree = trial.suggest_uniform('colsample_bytree', 0.5, 1.0) # 각

# 모델 초기화
model = xgb.XGBClassifier(n_estimators=n_estimators,
                          max_depth=max_depth,
                          learning_rate=learning_rate,
                          min_child_weight=min_child_weight,
                          gamma=gamma,
                          subsample=subsample,
                          colsample_bytree=colsample_bytree,
                          random_state=1234,
                          use_label_encoder=False,
                          eval_metric='logloss')

# 모델 학습
model.fit(train_X, train_y)

# 모델 평가 (여기서는 validation set이 없으므로 test set 사용)
test_accuracy = model.score(test_X, test_y)

return test_accuracy

# Optuna 스터디 생성
study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=100) # 1000번의 시도로 최적화 수행

# 최적 하이퍼파라미터 출력
print('Best trial:')
trial = study.best_trial

print('  Value: {}'.format(trial.value))
print('  Parameters:')
for key, value in trial.params.items():
    print('    {}: {}'.format(key, value))

# 최적 하이퍼파라미터로 모델 학습
best_model = xgb.XGBClassifier(**trial.params, random_state=1234, use_label_encoder=False)
best_model.fit(train_X, train_y)

# 성능 평가
print('Train accuracy :', best_model.score(train_X, train_y))
print('Test accuracy :', best_model.score(test_X, test_y))

pred_y = best_model.predict(test_X)
print('Confusion Matrix:\n', confusion_matrix(test_y, pred_y))

```

XGBoost Result

하이퍼파라미터 튜닝 전: Test accuracy : 0.7272727272727273

하이퍼파라미터 튜닝 후: Train accuracy : 0.9162011173184358 Test accuracy : 0.7705627705627706