

3 유형 실전문제 - 주가 상승/하락 예측 (Full Analysis)

문제

어떤 종목의 과거 일별 주가 데이터를 바탕으로 다음날 종가가 상승할지 하락할지를 예측하는 분류 모델을 만드시오.

- 주어진 데이터는 Open, High, Low, Close, Volume 컬럼으로 구성되어 있음
- Target 컬럼은 다음날 종가가 오늘보다 상승했는지 여부를 1(상승), 0(하락)으로 표현
- 분류 모델을 훈련하여 Target 을 예측하고, 실제값과 비교하여 정확도(accuracy), 혼동행렬, AUC 를 평가하시오.

데이터 예시 컬럼

Date, Open, High, Low, Close, Volume, Target

정답 코드 예시

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

# 데이터 불러오기
df = pd.read_csv("type3_stock_prediction.csv")

# 특성과 타겟 분리
X = df[['Open', 'High', 'Low', 'Close', 'Volume']]
y = df['Target']

# 학습/테스트 분할
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```

# 모델 훈련
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# 예측
y_pred = model.predict(X_test)
y_prob = model.predict_proba(X_test)[:, 1]

# 정확도
print("Accuracy:", accuracy_score(y_test, y_pred))

# 혼동행렬
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

# AUC
print("AUC:", roc_auc_score(y_test, y_prob))

# ROC 곡선
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
plt.plot(fpr, tpr)
plt.xlabel("FPR")
plt.ylabel("TPR")
plt.title("ROC Curve")
plt.grid()
plt.show()

```

해설

- RandomForestClassifier 는 비선형 분류 문제에 강한 모델로서 주가 데이터처럼 복잡한 패턴에도 잘 작동한다.
- Target 이 다음날 상승(1)/하락(0)을 나타내므로 이진 분류 문제로 접근하며, accuracy 외에도 AUC 로 모델 성능을 종합적으로 평가해야 한다.
- Volume 은 종종 시장의 강도를 나타내는 지표로 작용할 수 있다.

- ROC Curve 는 TPR 과 FPR 간의 trade-off 를 시각화해주며, AUC 는 이 곡선 아래 면적으로써 1 에 가까울수록 성능이 좋음을 의미한다.