

머신러닝을 활용한 저평가 종목 탐색 및 최적 매도 타이밍 예측 시스템

8조 김민경 박지인 이민열 이현 홍수현



목차

데이터 수집 및 가공	01
모델 개요	02
저평가 종목 식별 모델	03
매도 타이밍 예측 모델	04
결론	05

01. 데이터 가공

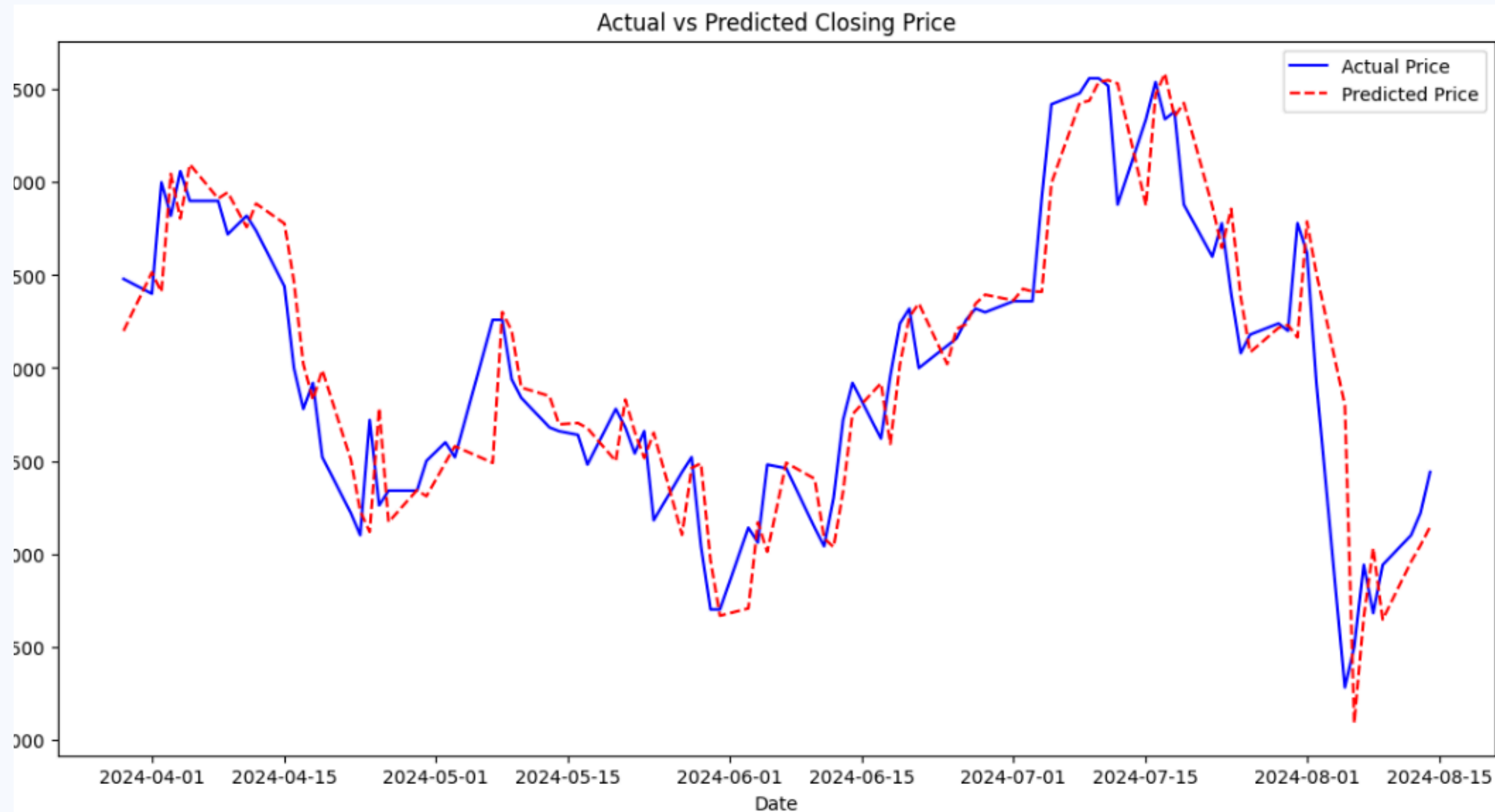
파생변수명	변수 설명	
	변수 생성 방식	변수 의미
PBR PER 배당수익률	최근 PBR, PER, 배당수익률을 찾아서 기입	기업의 주가가 순자산에 비해 얼마나 높은지, 기업의 주가가 수익에 비해 얼마나 높은지, 주가에 대한 배당금의 비율
PBR_점수	1미만 2점 / 1 or 2 1점 / 2초과 0점	2점 / 1점 / 0점으로 자산 대비 저평가된 정도를 반영하는 변수
PER_점수	1분위(7.26) 이하 2점 / 2분위(7.26 ~ 29.88) 1점 / 3분위(29.88) 0점 / 음수인 경우 -1점	2점 / 1점 / 0점 / -1점으로 주가가 기업의 수익 대비 얼마나 저평가되었는지를 반영 하는 변수
배당수익률_점수	4분위 이상 2점 / 2, 3 분위 1점 / 1분위 0점	2점 / 1점 / 0점으로 주가 대비 배당금이 얼마나 높은지를 반영하는 변수
총_점수	6점 이상을 저평가로 기준 잡음	PBR_점수, PER_점수, 배당수익률_점수와 코스피(+1) 조정을 합산하여 1점에서 7점으로 평가하는 함수

01. 데이터 가공

파생변수명	변수설명	
	변수 생성 방식	변수 의미
SMA 5 (5일 단순 이동 평균)	마지막 5일 동안의 증가(close)의 평균을 계산하여 생성	주가의 단기적 추세를 파악하기 위한 지표로 사용
Volatility 5 (5일 변동성)	마지막 5일 동안의 거래량(volume)의 평균을 계산하여 생성	가격의 변동 범위를 측정하고 주가의 일별 변화를 주식의 거래량 추세를 분석하기 위한 지표로 사용
EMA 5 (5일 지수 이동 평균)	종가(close)에 대해 5일 간의 지수 이동 평균을 계산하여 생성	주가의 최신 정보를 반영하기 위해 최근 데이터에 더 많은 가중치를 부여함
Volatility_Volume 5 (5일 거래량 변동성)	마지막 5일 동안의 거래량(volume)의 표준편차를 계산하여 생성	거래량이 얼마나 크게 변동하는지를 나타내는 지표
Volatility (변동성)	주식 가격의 변동 범위를 나타내는 지표.	주식의 하루 최고가(high)와 최저가(low)의 차이를 계산
turnover_ratio (회전율)	거래된 주식 수와 전체 발행 주식 수의 비율을 나타내는 지표로, 주식이 얼마나 활발하게 거래되고 있는지	거래량(volume)을 발행 주식 수(stocks)로 나누어 회전율을 계산
price_change_ratio (가격 변동 비율)	주식 가격의 변동 비율을 나타내는 지표로, 특정 기간 동안의 주식 가격 변화의 비율	주식 가격의 변화량(changes)을 그 날의 시가(open)로 나누어 가격변동비율을 계산

02.모델 개요

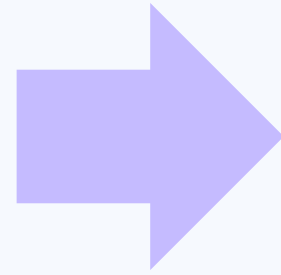
회귀 모델의 문제점



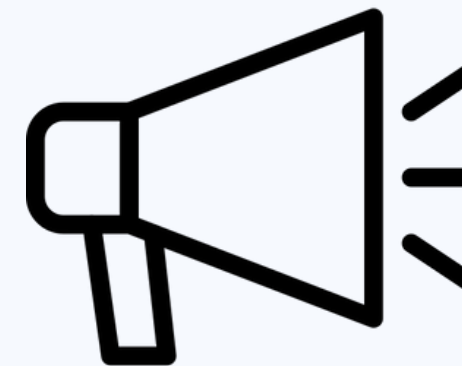
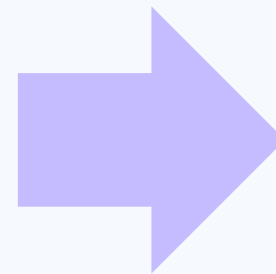
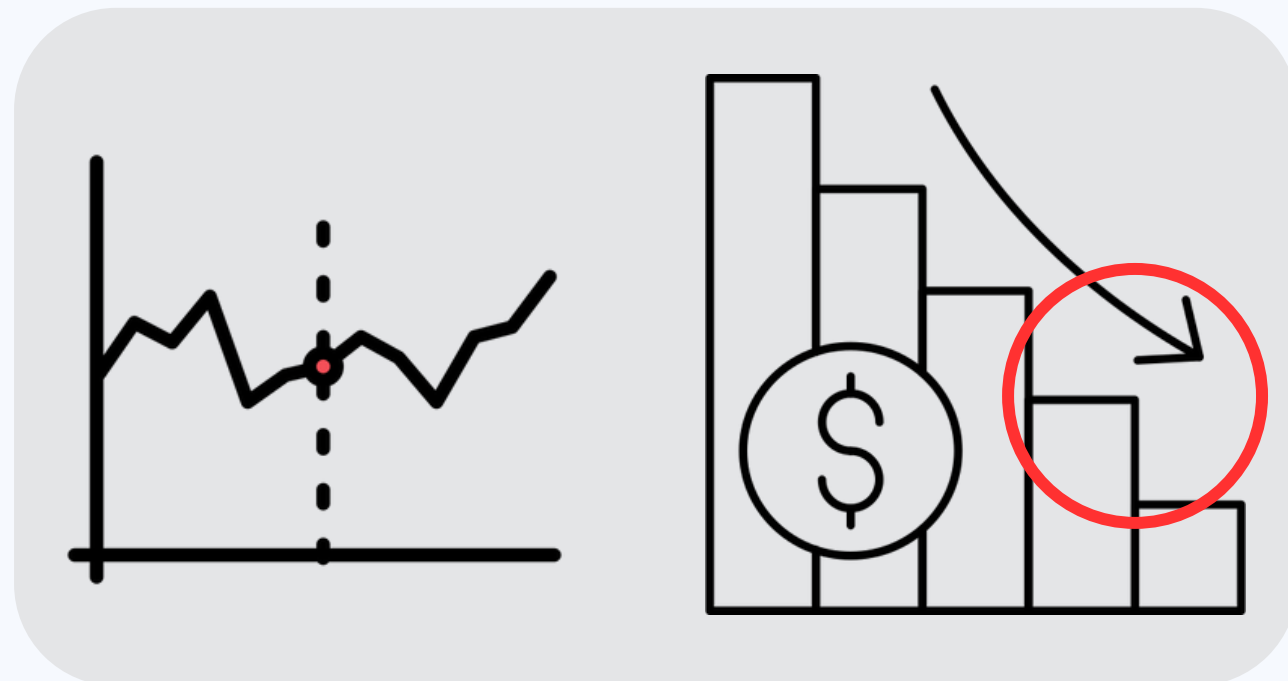
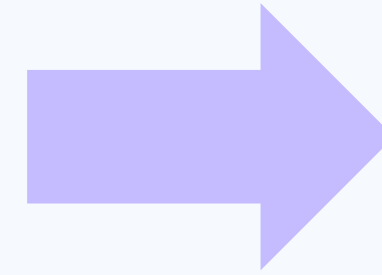
모델	RMSE
LSTM	1658

➤➤➤ 이전날의 값을 그냥 다음날에 그대로 예측하는 문제가 있음

02.모델 개요



저평가 기업입니다!
보시고 골라서 투자 하세요



하락 예정입니다
판매 추천드립니다

03.분류 모델링 과정 요약 및 결과

< Target >
0 : 1850 / 1 : 412

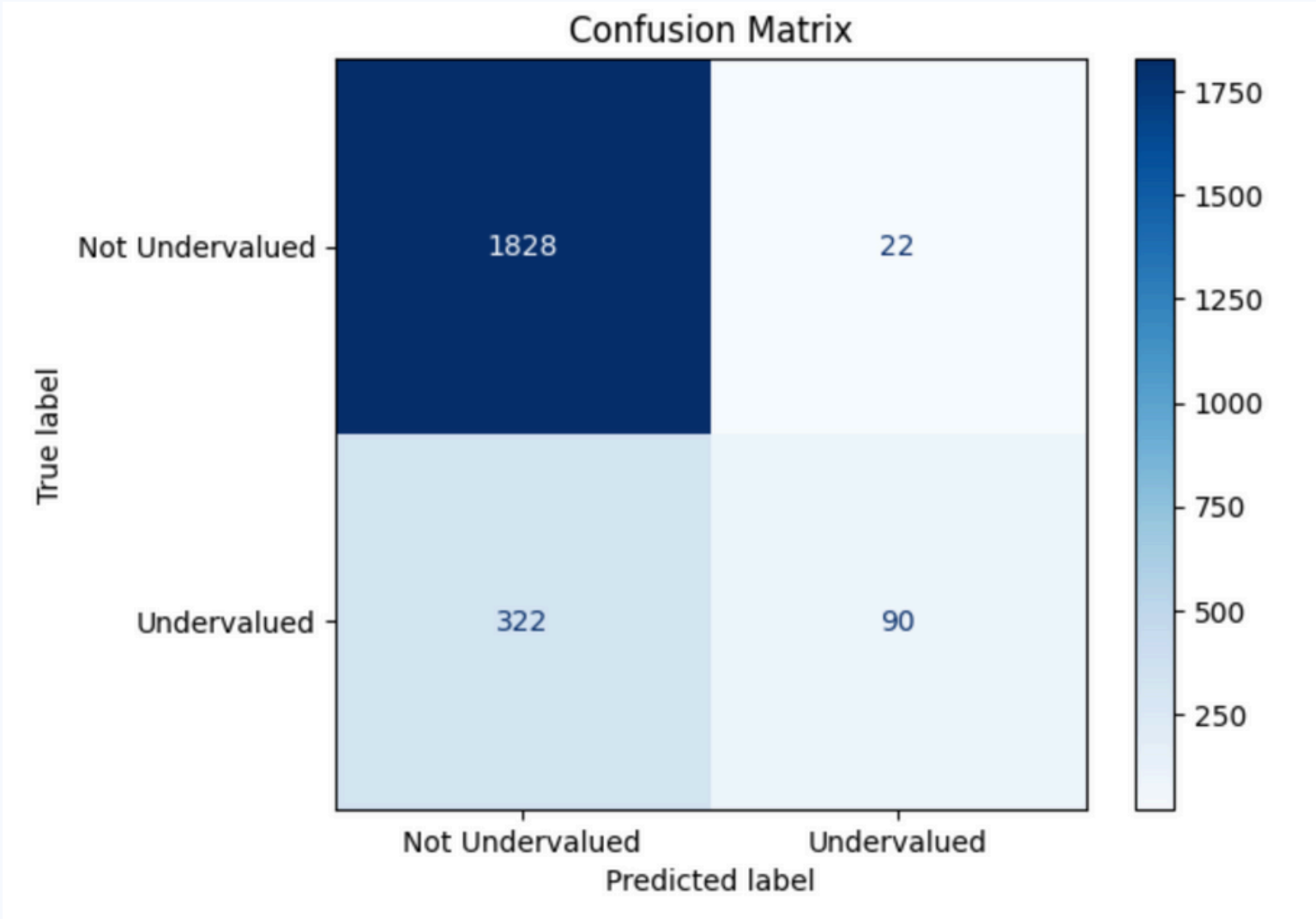
사용모델	AUC	F-1 score
랜덤포레스트	0.68	0.22
XGBoost	0.68	0.27
서포트벡터머신	0.43	0.005

사용 변수	AUC	F-1 score
후진변수법 n = 5	0.68	0.22
'High', 'Volume', 'Amount', 'Marcap', 'Stocks'		
후진변수법 n = 4	0.74	0.05
'Volume', 'Amount', 'Marcap', 'Stocks'		
직접 선택(그리드x)	0.73	0.3
직접 선택 (그리드o)	0.75	0.32
'Close', 'Volume', 'Marcap', 'Stocks', 'price_movement', 'volatility', 'turnover_ratio', 'price_change_ratio'		

>>> 최적의 분류 모델을 위해 여러 모델과 변수를 조합해봄

03.분류 모델링 과정 요약 및 결과

	내용
사용모델	XGBoost
사용 변수	'Close', 'Volume', 'Marcap', 'Stocks', 'price_movement', 'volatility', 'turnover_ratio', 'price_change_ratio'
하이퍼파라미터 튜닝	'learning_rate': 0.1 'max_depth': 3 'n_estimators': 50 'subsample': 0.8
AUC	0.75
F-1 score	0.32



장기투자를 목적으로 저평가된 모델을 찾아 추천

03.분류 모델링 과정 요약 및 결과



화승인더
1700 -> 11000

영풍제지
5000 -> 50000

화승인더
동일산업
대림B&Co
현대모비스
CJ

JB금융지주
효성티앤씨
롯데렌탈
동원f&b

영풍제지
사조대림
FnF 홀딩스
혜성 디에스
유나이티드 제약
삼천리
화신

대상홀딩스
대성에너지
대원강업
서울가스

케이씨
종근당홀딩스
송원산업
삼영무역
신대양제지
일진홀딩스
삼화페인트
일진다이아

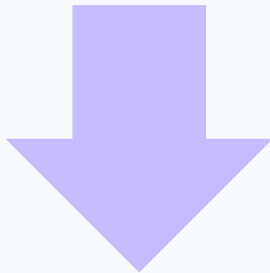
경농
무학
한미글로벌
벽산
삼양통상
금강공업
나이스정보통신
YG원
대한약품
대원산업
한국가구

➤➤➤ 평가기업의 약 10%가 급상하는 양상을 보임

04. 매도 타이밍 예측 모델

	Change	target
day1	+ 0.01	0
day2	- 0.01	1
day3	0.00	1
day4	+ 0.01	1
day5	-0.01	1
day6	-0.03	
...

day2에 target이 1이 나옴

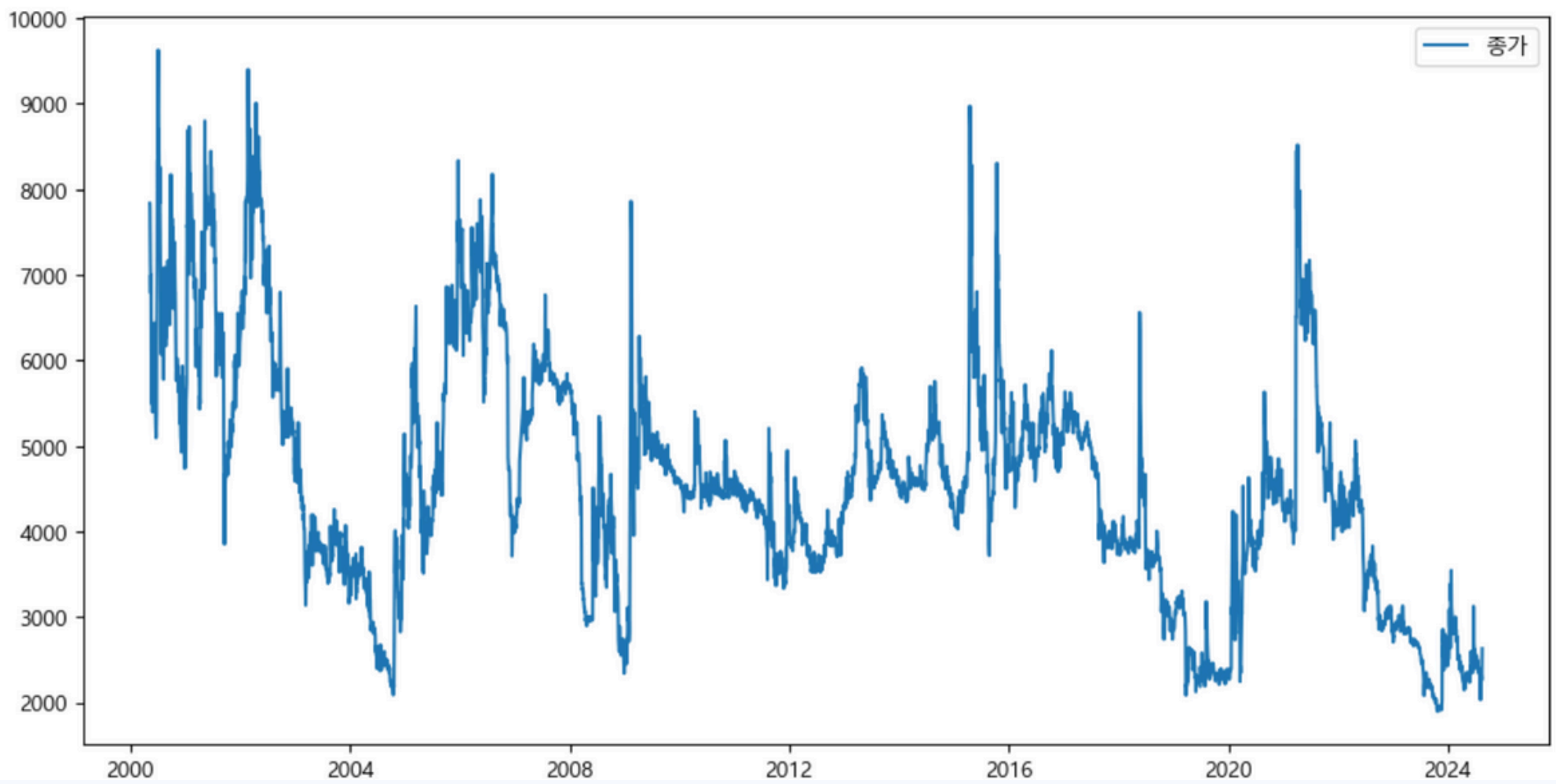


5일 이내에 종가 2% 이상 하락 예측

target이 1인 경우 5일 이내에 종가 2% 이상 하락 예측

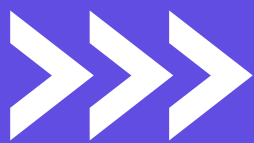
04. 매도 타이밍 예측 모델

깨끗한 나라



< Target >
0 : 3126 / 1 : 2832

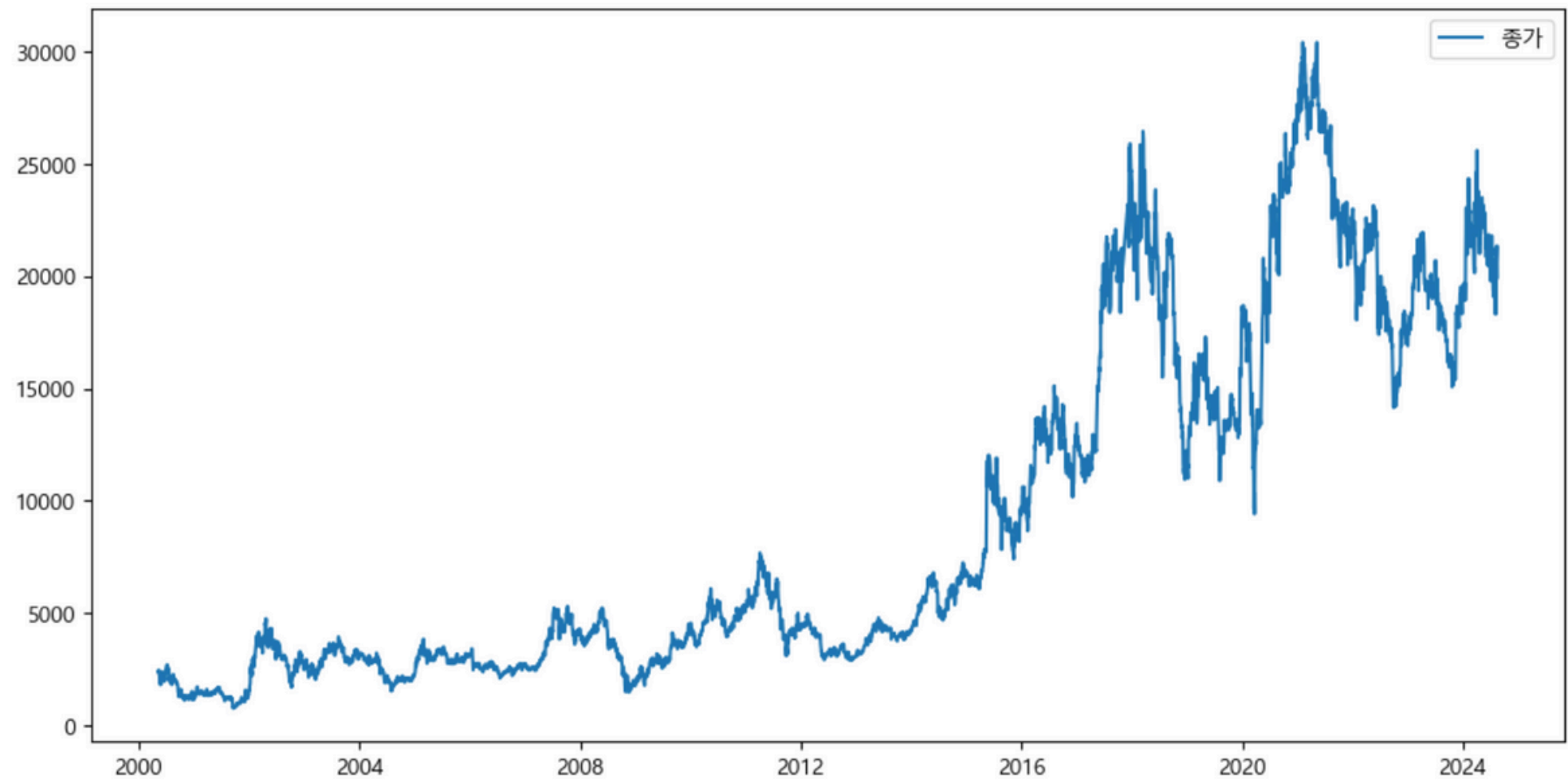
	내용
사용모델	Random Forest
사용 변수	'Close', 'SMA_5', 'Volatility_5', 'SMA_Volume_5'
하이퍼파라미터 튜닝	'n_estimators': 500 'min_samples_split': 2 'min_samples_leaf': 1 'max_depth': 40 'bootstrap': True
AUC	0.8191
Recall	0.75



75%의 재현율로 주가 하락을 예측함

04. 매도 타이밍 예측 모델

케이시



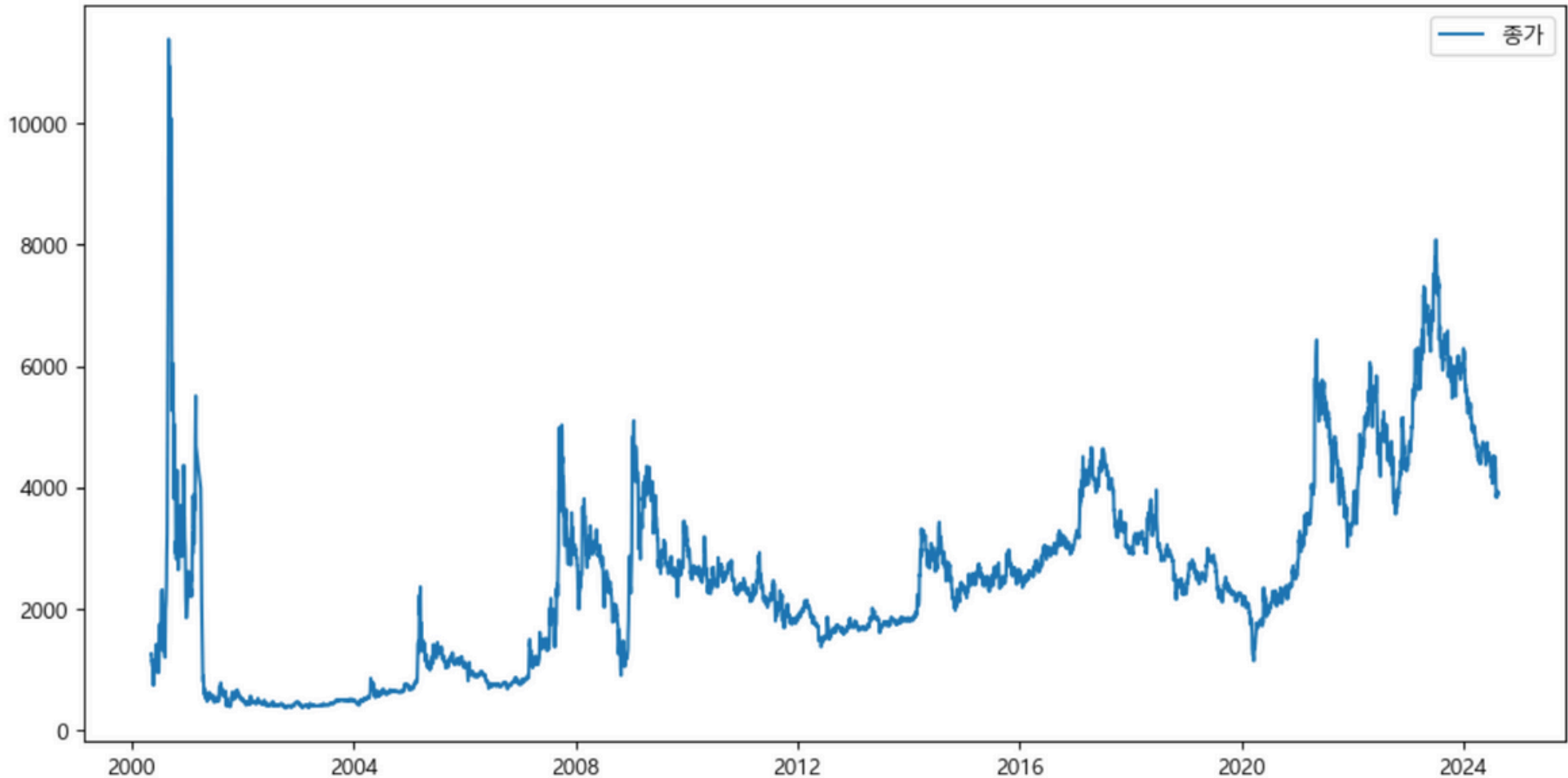
< Target >
0 : 2652 / 1 : 3298

	내용
사용모델	Random Forest
사용 변수	'Close', 'SMA_5', 'Volatility_5', 'SMA_Volume_5'
하이퍼파라미터 튜닝	'n_estimators': 500 'min_samples_split': 2 'min_samples_leaf': 1 'max_depth': 40 'bootstrap': True
AUC	0.7393
Recall	0.77

➤➤➤ 77%의 재현율로 주가 하락을 예측함

04. 매도 타이밍 예측 모델

NI 스틸



< Target >
0 : 2767 / 1 : 3198

	내용
사용모델	Random Forest
사용 변수	'Close', 'SMA_5', 'Volatility_5', 'SMA_Volume_5'
하이퍼파라미터 튜닝	'n_estimators': 500 'min_samples_split': 2 'min_samples_leaf': 1 'max_depth': 40 'bootstrap': True
AUC	0.7624
Recall	0.72

➤➤➤ 72%의 재현율로 주가 하락을 예측함

05. 결론 & 아쉬운 점

저평가 종목 추천

- 저평가 종목을 예측하여 장기 투자를 위한 기업 항목을 추천해준다.
- 모델이 F_1 Score가 낮아 아쉽다.

매도 타이밍 추천

- 기업별 특성을 반영하여 5일 내 하락장을 예측하여 매도 타이밍을 알려준다.
- 회귀모델을 통해 정확한 수치를 예측할 수 있었으면 좋았을 것 같다.

감사합니다