



TEAM 2

목차

01



**프로젝트
배경**

02



**프로젝트
기능**

03



데이터 분석

04



시연 영상



**주식 투자 입문자를 위한
모의 투자 기반 종목 추천 플랫폼**

프로젝트 배경

"새 정부 랠리 올라타자"...주식 계좌·예탁금 급증

조아라 기자 ☆

입력 2025.06.12 18:23 수정 2025.06.13 00:33 지면 A24





가



국내 주식거래 계좌 수가 사상 최대를 기록했다. 이재명 대통령 취임 이후 국내 증시가 랠리를 이어가자 수익을 낼 수 기회라고 판단한 투자자들이 물린 영향이다.

12일 금융투자협회에 따르면 국내 주식거래 활동계좌는 지난 11일 기준 9059만622개로 집계돼 신기록을 세웠다. 전년 동일(2024년 6월 11일 ·8239만3370개) 대비 약 10% 증가했다. 주식거래 활동계좌는 10만원 이상의 금액이 들어 있으면서 최근 6개월 동안 한 번 이상 거래가 이뤄진 계좌를 의미한다. 거래가 없는

많이 본 뉴스

- 1 "올해 119번이나 밥 먹으러 갔어요"...가성비에 '복적...' 
- 2 '다이소보다 싸네' 인기 폭 발...4만개 팔린 이마트 야... 
- 3 "기모노 입은 K유아복이래" MZ부모 열광...다서 '100... 
- 4 "이건 안 들면 손해라는 데"...연 12% 금리' 적금에... 

오피니언 | 기자수첩

'커뮤니티냐, 도박장이나' 토스증권 종토팡이 불편한 이유

등록 2025.01.22 10:22

유선희 기자



'진정한 타짜만 들어오세요. 인생 뭐 있어! 상(상한가) 가면 대박이고 하따(하한가 따라잡기) 치면 한강가면 되지!'
'삼십(만원) 먹었는데 이걸로 마무리합니다. 감사합니다.'

주요뉴스



- 1 HDC현산 임금 문제 '화해 권고'...노사 절충안 촉각
- 2 "우려가 현실로"...삼성전자, 2분기 어닝쇼크(종합)
- 3 '관세-삼성전자 실적' 겹악 데 지며...코스피 벅토서...

주식 시장에 대한 대중의 관심이 급증

BUT !

개인 투자자들은 **전문적인 분석 도구나 정보 접근성 부족**으로 인해

비체계적인 투자 결정을 내리는 경우가 많음

주가 등락을 사전에 예측할 수 있는 데이터 기반 분석 모델의 필요성
실시간 뉴스의 감정 정보를 분석해 시장 심리를 반영한 투자 판단을 도울 수 있는 시스템의 부재
안전하게 투자 전략을 실험해볼 수 있는 모의투자 환경의 부재



Stockr

프로젝트 기능

주식 관련 기능 제공

전체 종목 실시간 시세 조회

종목 상세 정보 확인

시장가 매수·매도 주문

유사 종목 추천

모델 기반 주가 상승/하락 예측

사용자 맞춤 종목 추천

종목 즐겨찾기

즐겨찾기 관리 및 삭제 가능

유사 투자자 기반 추천

투자 성향 및 포트폴리오 유사도 분석

나와 비슷한 사용자들이 투자한 종목 추천

뉴스 기반 정보 제공

금융 뉴스 조회

관심 종목 관련 뉴스 자동 추천

모델을 활용한 뉴스 영향도 시각화

긍정: 🌞 맑음 / 중립: 🌤 안개 / 부정: 🌧 폭우

부가 기능

커뮤니티 기능

내 정보 & 투자 성향 관리

회원가입 후 가상 투자 계좌 개설

대시보드에서 자산 및 거래 내역 확인

설문 기반 투자 성향 분석

데이터 분석



사용한 데이터 목록



한국거래소 정보데이터 시스템



네이버페이 증권

종목별 재무지표, 개장일(한국 증시가 열린 날)

종목코드	종목명	종가	대비	등락률	EPS	PER	BPS	PBR	주당배당금	배당수익률
060310	3S	2,155	▼ 30	-1.37	54	39.91	998	2.16	0	0.00
095570	AJ네트웍스	4,220	▼ 45	-1.06	485	8.70	9,902	0.43	270	6.40
006840	AK홀딩스	11,910	▼ 250	-2.06	-	-	40,295	0.30	400	3.36
054620	APS	4,985	▼ 65	-1.29	-	-	11,326	0.44	0	0.00
265520	AP시스템	17,080	▼ 60	-0.35	3,448	4.95	22,405	0.76	530	3.10
211270	AP위성	12,530	▼ 20	-0.16	733	17.09	7,393	1.69	0	0.00
109960	AP헬스케어	320	▼ 9	-2.74	-	-	1,503	0.21	0	0.00
139050	BF랩스	2,805	0	0.00	-	-	2,348	1.19	0	0.00

종목별 시가, 고가, 저가, 종가, 거래량 데이터 KOSPI/KOSDAQ 지수 기관 순매매량, 외국인 순매매량, 외국인 보유율

날짜	종가	전일비	시가	고가	저가	거래량
2025.07.07	21,750	▲ 900	21,000	22,200	21,000	889,756
2025.07.04	20,850	▼ 250	21,100	21,150	20,700	196,116
2025.07.03	21,100	▲ 200	21,200	21,250	20,950	198,400
2025.07.02	20,900	▲ 200	20,850	20,950	20,700	179,656
2025.07.01	20,700	▲ 100	20,650	20,900	20,600	219,028

사용한 데이터 목록



yfinance API

[미국 지수]

S&P500, NASDAQ,

DOW JONES, VIX

[환율/금리/원유]

USD/KRW,

미국 10년물 금리,

WTI 유가



FinanceDataReader

FinanceDataReader

**종목별 거래량,
이동 평균선**



네이버 뉴스

종목별 관련 기사

Datasets:  takala / financial_phrasebank

Finance Phrase Bank

금융/경제 기사

데이터 전처리 - 결측치 처리 I

[유사 종목 추천]

재무지표의 '-'값은 결측치로 간주 후, dropna를 통해 제거

-> 총 460개의 종목에 대해서 데이터 분석 진행

```
# '-'를 결측값으로 바꾸기
df[target_columns] = df[target_columns].replace('-', pd.NA)
# 결측값 제거
filtered_df = df.dropna(subset=target_columns)
```

종목코드	종목명	증가	대비	등락률	EPS	PER	BPS	PBR	주당배당금	배당수익률
060310	3S	2,155	▼ 30	-1.37	54	39.91	998	2.16	0	0.00
095570	AJ네트웍스	4,220	▼ 45	-1.06	485	8.70	9,902	0.43	270	6.40
006840	AK홀딩스	11,910	▼ 250	-2.06	-	-	40,295	0.30	400	3.36
054620	APS	4,985	▼ 65	-1.29	-	-	11,326	0.44	0	0.00
265520	AP시스템	17,080	▼ 60	-0.35	3,448	4.95	22,405	0.76	530	3.10
211270	AP위성	12,530	▼ 20	-0.16	733	17.09	7,393	1.69	0	0.00
109960	AP헬스케어	320	▼ 9	-2.74	-	-	1,503	0.21	0	0.00
139050	BF랩스	2,805	0	0.00	-	-	2,348	1.19	0	0.00
027410	BGF	4,170	▲ 40	+0.97	963	4.33	18,202	0.23	130	3.12
282330	BGF리테일	124,600	▲ 3,400	+2.81	11,301	11.03	68,546	1.82	4,100	3.29
126600	BGF에코머티리얼즈	3,030	▼ 30	-0.98	259	11.70	6,475	0.47	50	1.65



종목코드	종목명	증가	대비	등락률	EPS	PER	BPS	PBR	주당배당금	배당수익률
008500	일정실업	12,730	▲ 350	+2.83	2,888	4.41	5,141	2.48	0	0.00
019540	일지테크	4,515	▲ 190	+4.39	1,703	2.65	11,378	0.40	100	2.21
081000	일진다이아	13,320	▼ 40	-0.30	908	14.67	29,059	0.46	300	2.25
020760	일진디스플레이	1,058	▲ 19	+1.83	110	9.62	685	1.54	0	0.00
103590	일진전기	36,950	▲ 750	+2.07	983	37.59	10,494	3.52	300	0.81
094820	일진파워	11,820	0	0.00	692	17.08	8,861	1.33	330	2.79
271940	일진하이솔루스	16,460	▼ 140	-0.84	50	329.20	8,743	1.88	0	0.00
015860	일진홀딩스	5,800	▲ 170	+3.02	491	11.81	10,068	0.58	150	2.59
226320	잇츠한불	13,100	▲ 90	+0.69	937	13.98	21,525	0.61	250	1.91
049550	잉크테크	3,320	▼ 30	-0.90	157	21.15	3,872	0.86	0	0.00
389020	자람테크놀로지	39,250	▼ 650	-1.63	323	121.52	6,860	5.72	0	0.00

데이터 전처리 - 결측치 처리 II

[주가 등락 예측 모델]

관련 뉴스가 없어 감성 분석 결과가 결측인 경우
-> 중립값으로 대체

```
def analyze_sentiment(self, text):
    try:
        text = text[:512]
        result = nlp_pipeline(text)

        label = result[0]['label']
        score = result[0]['score']

        if label == 'positive':
            return round(score, 3)
        elif label == 'negative':
            return round(-score, 3)
        else:
            return 0.0
    except Exception as e:
        logger.warning(f"감성 분석 오류 - 기사 없음: {e}")
        return 0.0
```

날짜	종목명	종목코드	뉴스감성수치	Target
2024.10.2	하이트진로	80		1
2025.1.3	하이트진로	80		0
2025.2.11	하이트진로	80		0
2025.2.19	하이트진로	80		1
2024.7.23	CJ대한통운	120		1
2024.8.7	CJ대한통운	120		1
2025.3.28	CJ대한통운	120		1
2025.6.27	CJ대한통운	120		0
2024.7.4	한국엔컴퍼니	240		1
2024.7.9	한국엔컴퍼니	240		0
2024.7.26	한국엔컴퍼니	240		1
2024.7.1	노루홀딩스	320		0
2024.7.3	노루홀딩스	320		1



날짜	종목명	종목코드	뉴스감성수치	Target
2024.10.2	하이트진로	80	0	1
2025.1.3	하이트진로	80	0	0
2025.2.11	하이트진로	80	0	0
2025.2.19	하이트진로	80	0	1
2024.7.23	CJ대한통운	120	0	1
2024.8.7	CJ대한통운	120	0	1
2025.3.28	CJ대한통운	120	0	1
2025.6.27	CJ대한통운	120	0	0
2024.7.4	한국엔컴퍼니	240	0	1
2024.7.9	한국엔컴퍼니	240	0	0
2024.7.26	한국엔컴퍼니	240	0	1
2024.7.1	노루홀딩스	320	0	0
2024.7.3	노루홀딩스	320	0	1

데이터 전처리 - 결측치 처리 III

[주가 등락 예측 모델]

한국 증시는 열렸으나 미국 증시가 열리지 않은 날 결측값

-> 가장 최근 값으로 대체

```
# 날짜 기준으로 머지
df = pd.merge(krx_calendar, us_index_df, on='date', how='left')

# 미국 지수가 결측인데 한국 증시는 열렸을 경우 → ffill 적용
us_cols = ['sp500', 'nasdaq', 'dow']

for col in us_cols:
    df[col] = df[col].where(~(df['kr_open'] & df[col].isna()), None)
    df[col] = df[col].ffill()
```

S&P500_종가	NASDAQ_종가	DOWJONES_종가
5473.22998	17857.01953	38778.10156
5487.029785	17862.23047	38834.85938
0	0	0
5473.169922	17721.58984	39134.76172

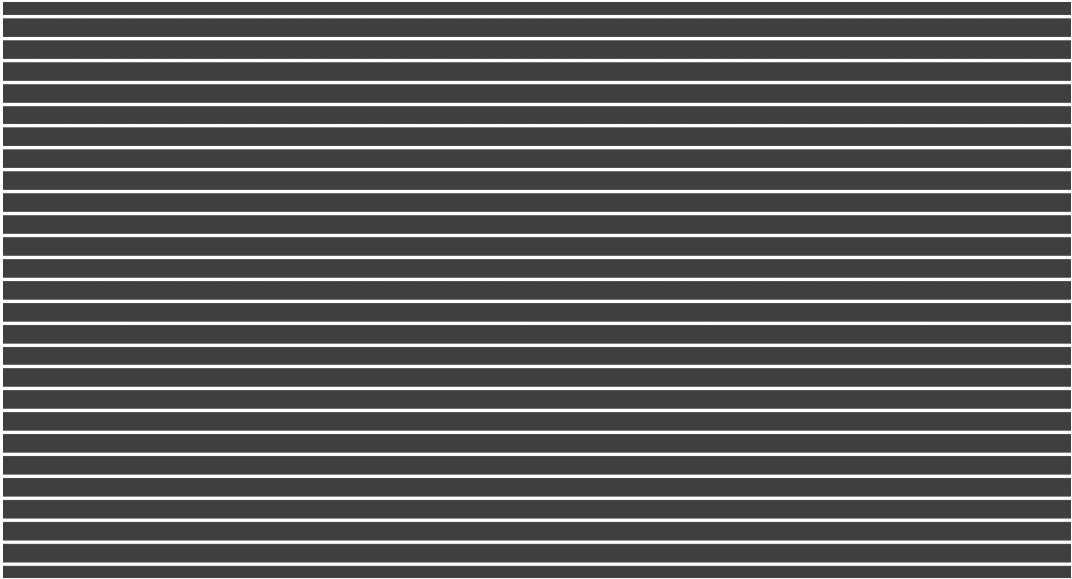


S&P500_종가	NASDAQ_종가	DOWJONES_종가
5473.22998	17857.01953	38778.10156
5487.029785	17862.23047	38834.85938
5487.029785	17862.23047	38834.85938
5473.169922	17721.58984	39134.76172

데이터 전처리 - 결측치 처리 결과



▲ 결측치 처리 전



▲ 결측치 처리 후

[뉴스 감정 분석 모델]

-> Undersampling & Class Weight 조정

중복된 기사 19개

→ 제거

labels	sentence	kor_sentence
neutral	According	Gran에 따르면, 그 회사는 회사가 성장하고 있는 곳이지만, 모든 생산을 러시아로 옮길 계획이 없다고 한다.
neutral	Technopo	테크노폴리스는 컴퓨터 기술과 통신 분야에서 일하는 회사들을 유치하기 위해 10만 평방미터 이상의 면적을 단계적으로 개발할 계획이라고 성명은 밝혔다.
negative	The intern	국제 전자산업 회사인 엘코텍은 탈린 공장에서 수십 명의 직원을 해고했으며, 이전의 해고와는 달리 회사는 사무직 직원 수를 줄였다고 일간 포스티메스가 보도했다.
positive	With the r	새로운 생산공장으로 인해 회사는 예상되는 수요 증가를 충족시킬 수 있는 능력을 증가시키고 원자재 사용을 개선하여 생산 수익성을 높일 것이다.
positive	According	2009-2012년 회사의 업데이트된 전략에 따르면, Basware는 20% - 40% 범위의 장기적인 순매출 성장을 목표로 하고 있으며, 영업이익률은 순매출액의 10% - 20%를 목표로 하고 있습니다.
positive	FINANCIN	ASPOCOMP의 성장기에 대한 자금 조달은 기술적으로 더 까다로운 HDI 인쇄 회로 기판 PCB에 점점 더 초점을 맞추면서 성장 전략을 공격적으로 추진하고 있다.
positive	For the las	2010년 4분기 Componenta의 순매출은 전년 동기의 7600만 유로에서 2배인 13100만 유로로 증가했으며, 세전 손실 7백만 유로에서 세전 수익 0으로 전환되었다.
positive	In the thir	2010년 3분기 순매출은 5.2%, 영업이익은 34.9% 증가한 23.5MN을 기록했다.
positive	Operating	영업이익은 2007년 해당 기간의 8.7 mn에서 13.1 mn으로 증가하여 순매출액의 7.7%를 차지했다.
positive	Operating	영업이익은 총 21.1 유로로 2007년 18.6 mn에서 증가하여 순매출의 9.7%를 차지했다.
positive	TeliaSoner	텔리아소네라 TLSN은 이번 제안이 핵심 사업 보유에 대한 지분을 늘리기 위한 전략과 부합하며 에스티텔레콤의 고객 제안을 강화할 것이라고 밝혔다.
positive	STORA EN	STORA ENSO, NORSKE SKOG, M-REAL, UPM-KYMMENE Credit Suisse First Boston(CFSB)은 북유럽 삼림 그룹 중 4곳의 주식에 대한 공정가치를 상향 조정했다.
positive	A purchas	핀란드 하미나 터미널에서 인도하는 휘발유 7,200t에 대한 구매협약이 올 9월 플랫폼 평균지수에서 네스테오일 OYJ와 월 8달러(약 8만 원)로 체결됐다.
positive	Finnish Ta	핀란드 탈렌툼은 영업이익이 2004년 9.3 mn에서 2005년 20.5 mn으로 증가했으며, 순매출액은 96.4 mn에서 103.3 mn으로 증가했다고 발표했다.
positive	Clothing r	의류 소매 체인 셉+엘+엘+엘은 매출이 8% 증가한 155.2만 유로, 영업이익은 2004년 17.1만 유로에서 31.1만 유로로 증가했다.
positive	Consolida	연결 순매출액은 16% 증가하여 74.8백만 유로에 달했고, 영업이익은 전년도의 0.7백만 유로에 비해 0.9백만 유로에 달했다.

파생 변수 생성 I - 유사 종목 추천

CAPM(자본자산가격결정모형)의 핵심 요소인 Beta(β)를 파생변수로 사용

$$R_i = \alpha + \beta R_m + \varepsilon$$

Beta(β): 해당 종목이 시장 수익률 변화에 얼마나 민감하게 반응하는지 측정
-> 시장의 변화에 따라 주가가 얼마나 출렁이는지를 수치화한 것

CAPM: 기업의 가치를 계산하거나 자산에 대한 투자 결정을 보조할 때에 가장 많이 사용되는 재무 모델

종목코드	시장구분	Beta	Alpha	R2
80	KOSPI	0.4123	-0.0001	0.1958
100	KOSPI	1.0626	-0.0019	0.1778
120	KOSPI	0.5555	0	0.0988
210	KOSPI	1.2927	-0.0002	0.3051
240	KOSPI	0.6231	-0.0001	0.093
270	KOSPI	0.9781	-0.0009	0.2846
320	KOSPI	0.6238	0.0025	0.1422
370	KOSPI	1.0531	-0.0005	0.2689
500	KOSPI	1.3315	0.0022	0.1174
640	KOSPI	0.2365	-0.0003	0.0202
660	KOSPI	1.741	0.0013	0.4226

종목코드	종목명	Beta	EPS	PER	선형 EPS
80	하이트진로	0.4123	1378	14.66	1870
100	유한양행	1.0626	943	109.86	1800
120	CJ대한통운	0.5555	12088	7.24	11869
210	DL	1.2927	3948	11.42	5464
240	한국엔컴퍼니	0.6231	3703	5.07	4039
270	기아	0.9781	24893	3.92	23100
320	노루홀딩스	0.6238	4836	4.88	3214
370	한화손해보험	1.0531	2121	2.69	2503
500	가온전선	1.3315	3270	19.54	3897
640	동아쏘시오홀딩스	0.2365	9017	12.37	14616
660	SK하이닉스	1.741	28732	9.88	42212

```
# Beta 계산
results = []

for code in stock_df['종목코드'].unique():
    sub = stock_df[stock_df['종목코드'] == code].copy()
    market_type = sub['시장구분'].iloc[0] # KOSPI or KOSDAQ

    stock = sub[['날짜', '수익률']].dropna()
    market = kospi_df if market_type == 'KOSPI' else kosdaq_df
    merged = pd.merge(stock, market[['날짜', '시장수익률']], on='날짜').dropna()

    if len(merged) < 30:
        print(f"[SKIP] {code} ({market_type}) - insufficient data")
        continue

    X = merged['시장수익률'].values.reshape(-1, 1)
    y = merged['수익률'].values
    model = LinearRegression().fit(X, y)

    beta = model.coef_[0]
    alpha = model.intercept_
    r2 = model.score(X, y)

    results.append({
        '종목코드': code,
        '시장구분': market_type,
        'Beta': round(beta, 4),
        'Alpha': round(alpha, 4),
        'R2': round(r2, 4),
        '데이터수': len(merged)
    })

print(f"[BETA] {code} ({market_type}) -> Beta: {beta:.4f}, R²: {r2:.4f}")
```


파생 변수 생성 II - 주가 등락 예측 모델

$$\text{gap_ratio} = \frac{\text{Close} - \text{Open}}{\text{Open}}$$

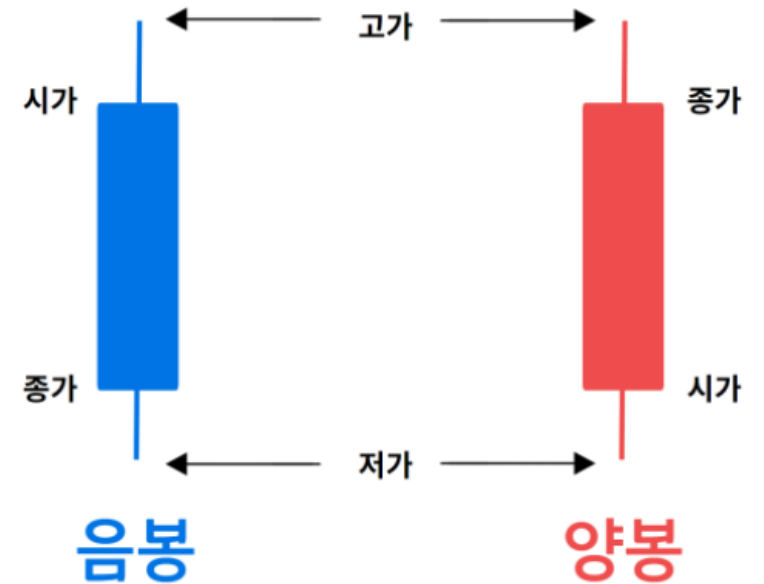
종가와 시가의 차이 비율
→ 등락폭 (모멘텀)을 나타냄

$$\text{high_low_ratio} = \frac{\text{High} - \text{Low}}{\text{Open}}$$

고가와 저가의 차이 비율
→ 변동성 수준

$$\text{is_bullish} = \begin{cases} 1 & \text{if Close} > \text{Open} \\ 0 & \text{otherwise} \end{cases}$$

종가 > 시가 여부 (1 또는 0)
→ 양봉/음봉 여부



```
merged_df["gap_ratio"] = (merged_df["Close"] - merged_df["Open"]) / merged_df["Open"]
merged_df["high_low_ratio"] = (merged_df["High"] - merged_df["Low"]) / merged_df["Open"]
merged_df["is_bullish"] = (merged_df["Close"] > merged_df["Open"]).astype(int)
```

최종 데이터셋

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 460 entries, 0 to 459
Data columns (total 29 columns):
#   Column                Non-Null Count  Dtype
---  -
0   종목 코드             460 non-null    int64
1   종목명                 460 non-null    object
2   종가                   460 non-null    int64
3   대비                   460 non-null    int64
4   등락률                 460 non-null    float64
5   EPS                   460 non-null    int64
6   PER                   460 non-null    float64
7   선행 EPS               460 non-null    int64
8   선행 PER                460 non-null    float64
9   BPS                   460 non-null    int64
10  PBR                    460 non-null    float64
11  주당배당금             460 non-null    int64
12  배당수익률             460 non-null    float64
13  시장구분               460 non-null    object
14  업종명                 460 non-null    object
15  시가총액               460 non-null    int64
16  EPS_표준화             460 non-null    float64
17  PER_표준화             460 non-null    float64
18  선행 EPS_표준화         460 non-null    float64
19  선행 PER_표준화         460 non-null    float64
20  BPS_표준화             460 non-null    float64
21  PBR_표준화             460 non-null    float64
22  주당배당금_표준화       460 non-null    float64
23  배당수익률_표준화       460 non-null    float64
24  시가총액_표준화         460 non-null    float64
25  Beta                   460 non-null    float64
26  1년_수익률(%)           460 non-null    float64
27  수익률_변동성           460 non-null    float64
28  1년_수익률(%)_표준화    460 non-null    float64
dtypes: float64(18), int64(8), object(3)
```

호가

종목 상세

추천 종목

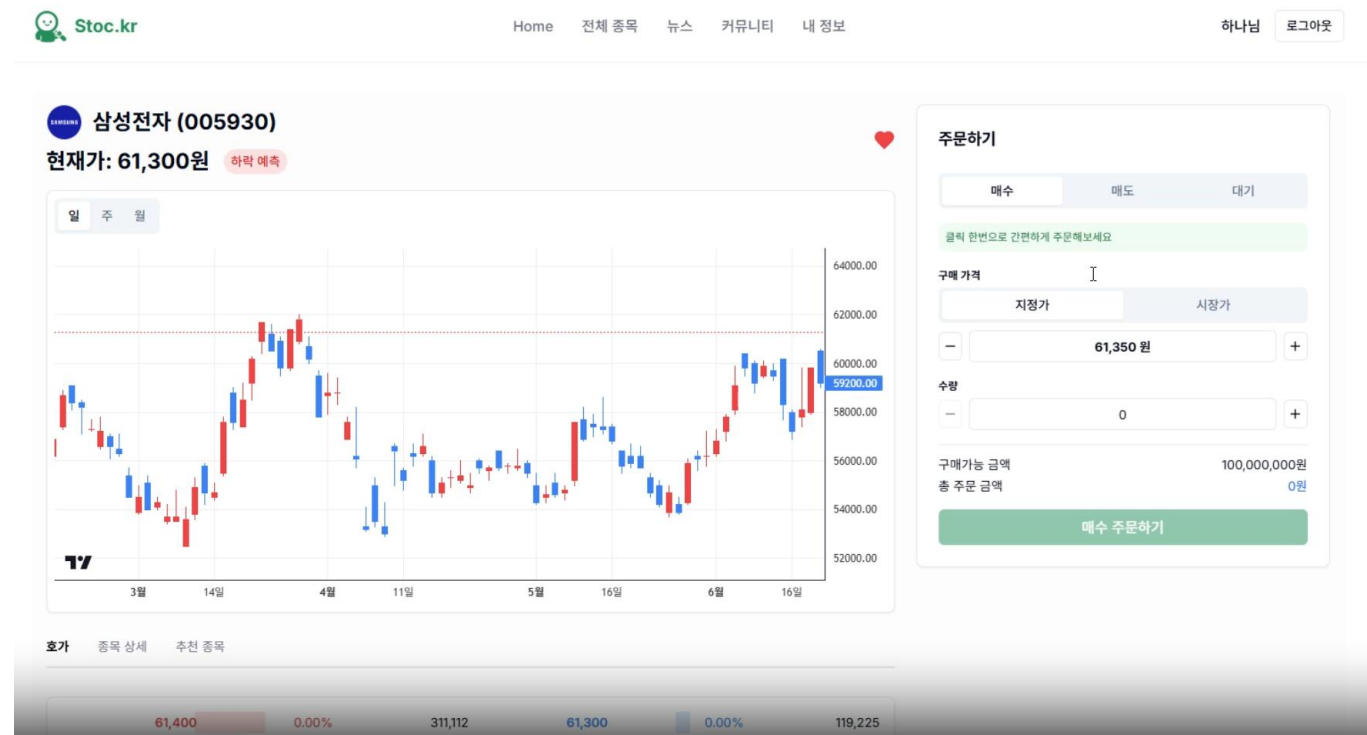
순위	종목명	업종	유사도
1	 LG전자	전기·전자	77.91%
2	 삼성SDI	전기·전자	77.47%
3	 삼성전기	전기·전자	77.06%
4	 리노공업	전기·전자	76.92%
5	 LG이노텍	전기·전자	76.74%

종목 추천 시 활용되는 유사도 분석 데이터셋

최종 데이터셋 II

```
RangeIndex: 275300 entries, 0 to 275299
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  -
0   날 짜                  275300 non-null object
1   Open                  275300 non-null int64
2   High                  275300 non-null int64
3   Low                   275300 non-null int64
4   Close                 275300 non-null int64
5   Volume                275300 non-null int64
6   Change                275281 non-null float64
7   S&P500_종 가          275300 non-null int64
8   NASDAQ_종 가          275300 non-null int64
9   DOWJONES_종 가        275300 non-null int64
10  S&P500_등 락 른        275300 non-null int64
11  NASDAQ_등 락 른        275300 non-null int64
12  DOWJONES_등 락 른      275300 non-null int64
13  gap_ratio              275300 non-null float64
14  high_low_ratio         275131 non-null float64
15  is_bullish             275300 non-null int64
16  Target                 275300 non-null int64
17  종 목 코드             275300 non-null int64
dtypes: float64(3), int64(14), object(1)
```

주가 등락 예측 모델 학습용 데이터셋



최종 데이터 셋 III

```
RangeIndex: 4846 entries, 0 to 4845
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  -
0   labels           4846 non-null   object
1   sentence         4846 non-null   object
2   kor_sentence      4846 non-null   object
dtypes: object(3)
memory usage: 113.7+ KB
```

뉴스 감성 분석 모델 학습용 데이터셋

네이버 치지직, e스포츠 대회 'EWC' 3년간 한국어 독점 중계

2025-07-08 10:01:17 · 조선비즈

NAVER

▶ 읽음

IT

EWC X 치지직

EWC 2025 한국어 독점 중계

총 상금 1천억원!

뜨거운 리야드의 열기 오직 치지직에서!



유사 종목 추천 - 유사도 분석

코사인 유사도(Cosine Similarity)를 사용하여 계산
각 종목을 n차원 벡터로 보고, 두 벡터 간 코사인 각도 유사도를 계산

$$\text{cos_sim}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

유사도가 1에 가까울수록 종목 간 특성이 유사한것으로 간주.
같은 업종이면 가중치 1.0, 다른 업종이면 가중치 0.8로 유사도 조정

$$\text{adjusted_sim} = \text{cos_sim} \times \text{업종_가중치}$$

추천 종목 선정:유사도 상위 5개 종목 추천 (자기 자신 제외)

```
# 코사인 유사도 계산
cos_sim = cosine_similarity(X)

# 업종 가중치 적용
adjusted_scores = []
for i, score in enumerate(raw_scores):
    if i == target_index:
        continue
    same_industry = df.loc[i, '업종명'] == target_industry
    weight = 1.0 if same_industry else 0.8
    adjusted_scores.append((i, score * weight))
```

주식 예측 모델 I - 이미지 데이터 기반 모델

CNN



학습 및 예측에 사용된 이미지 데이터

전문가들이 사용하는 14가지 대표 차트 패턴
(예: 이브닝 스타, 모닝 스타, 흑삼병 등)



사람이 시각적으로 판단하는 패턴을 CNN으로 대체하고자 함

결과> 테스트 정확도 56.1%, 검증 정확도 51.7%

-> 매우 낮은 성능

차트를 이미지로 변환하는 과정에서 정확한 가격·시간 값이 소실되어
모델이 수치 기반의 패턴을 정확히 학습하기 어려움

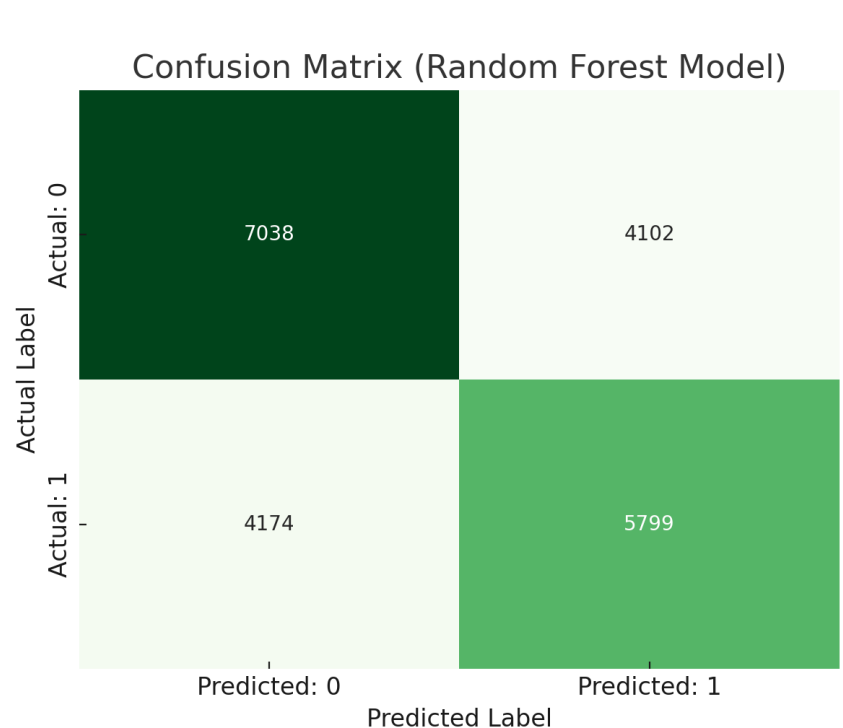
동일한 주가 흐름이라도 차트 툴 설정 (축 비율, 색상, 배경 등)에 따라
이미지 상에서 다르게 표현됨 이로 인해 같은 패턴이 서로 다른 패턴처럼
인식되는 문제 발생

-> 결과적으로 일관성 부족으로 모델 성능 저하

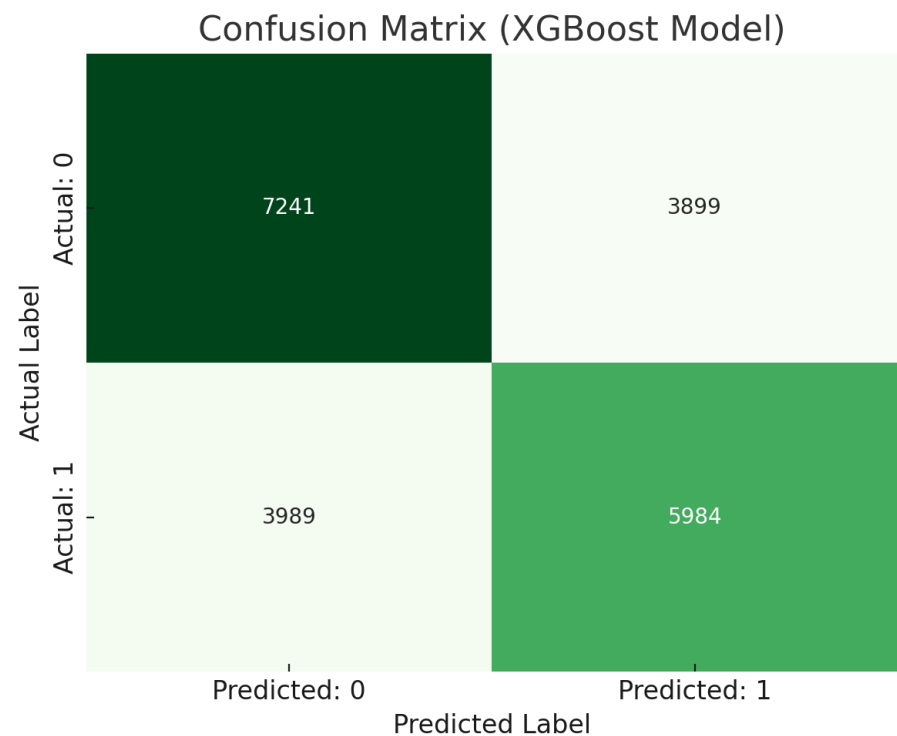
수치 기반으로 전환하여 차트의 외형보다 내부의 정확한 수치 변화에 주목할 필요가 있음

주식 예측 모델 II - 수치 데이터 기반 모델

Random Forest / XGBoost



정확도 60%

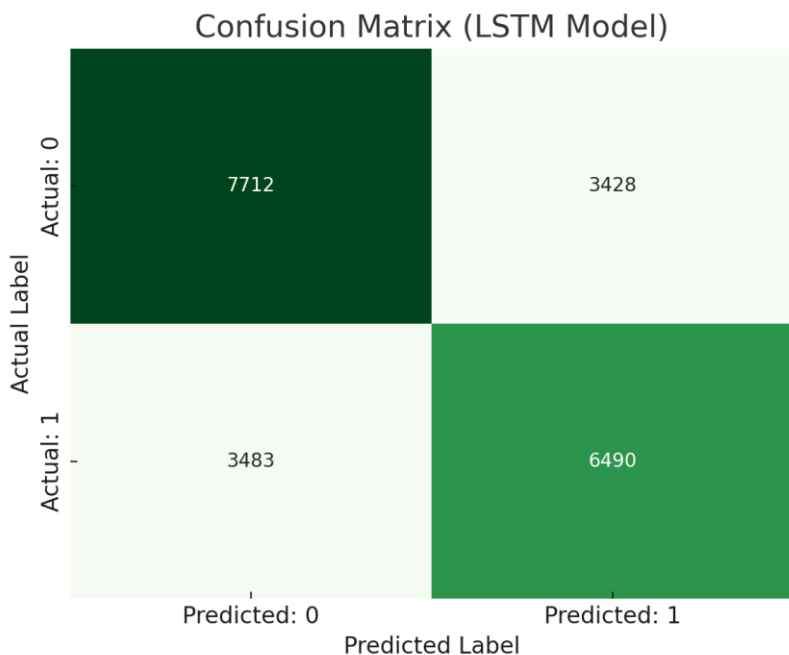
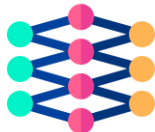


정확도 61%

기존 머신러닝(랜덤포레스트, XGBoost 등)은 시계열의 순서를 반영하기 어려움

주식 예측 모델 II - 수치 데이터 기반 모델

Bidirectional LSTM

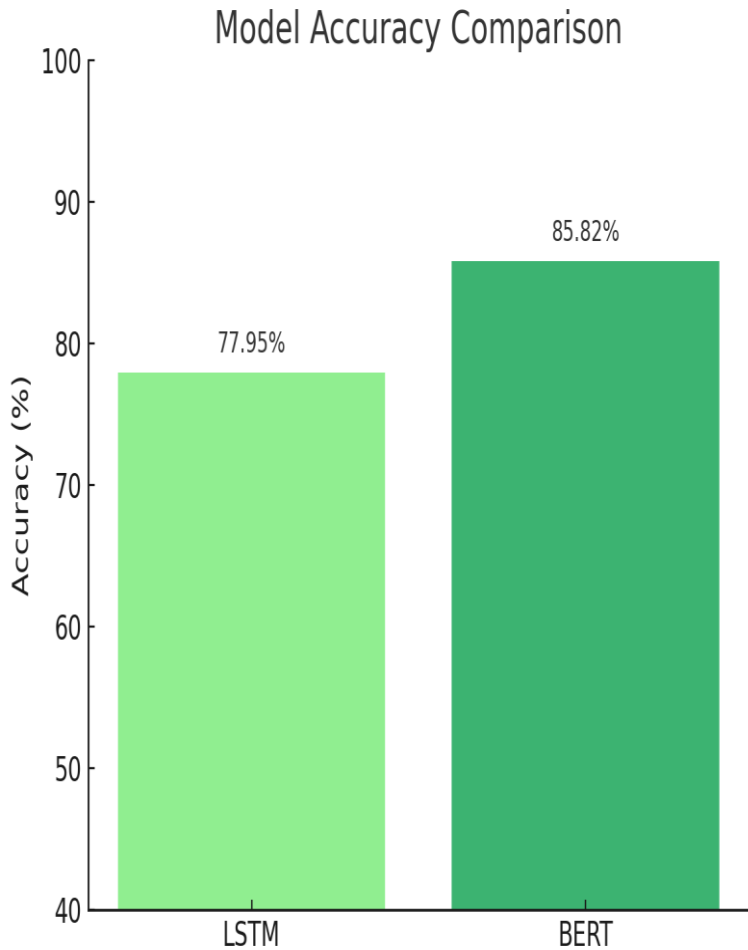
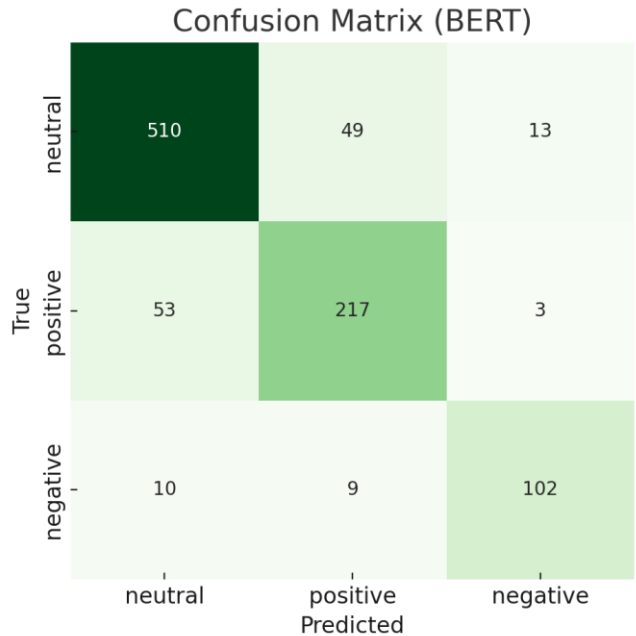
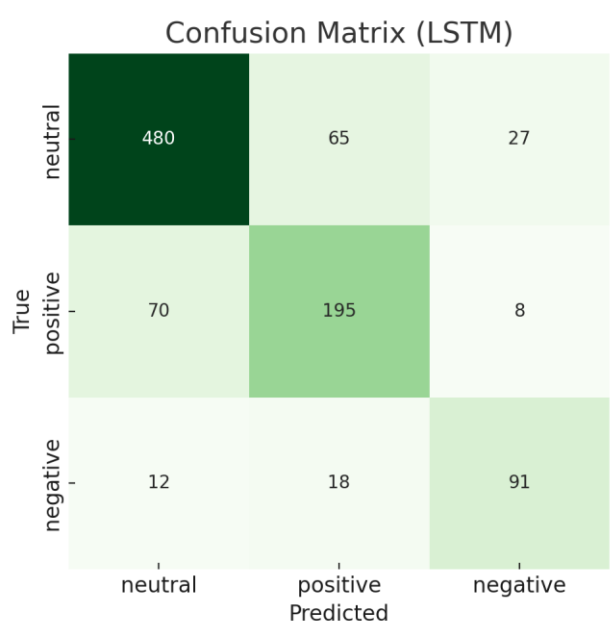


주가 데이터는 과거의 흐름이 미래에 영향을 주는 전형적인 시계열
따라서, LSTM을 선택, 시계열 순서를 모델 내부에 반영
일반 **LSTM**은 과거 → 현재 방향만 흐름 학습
Bidirectional LSTM은 정방향 + 역방향 LSTM 두 개를 학습
양쪽 정보를 모두 고려하여 시계열 특성 반영

★ ★ ★ 정확도 67%

뉴스 감성 분석 모델-금융 뉴스 문장 감성 분석

모델	정확도	특징
LSTM	77.95%	순환 신경망 기반, 기본 구조
BERT	85.82%	Transformer 기반, 사전학습 모델



사용자 기반 유사 종목 추천 - 유사도 분석

투자 성향을 기반으로 선호 종목 유형 파악
즐거찾기 종목의 공통 산업군 및 테마 분석
포트폴리오 보유 종목과 매매 이력 분석
유사 사용자 그룹의 투자 패턴을 참고하여 추천



X

stockr님의 투자자 성향은 **적극투자형** 이에요!

🕒 2027년 7월 8일까지 유효해요

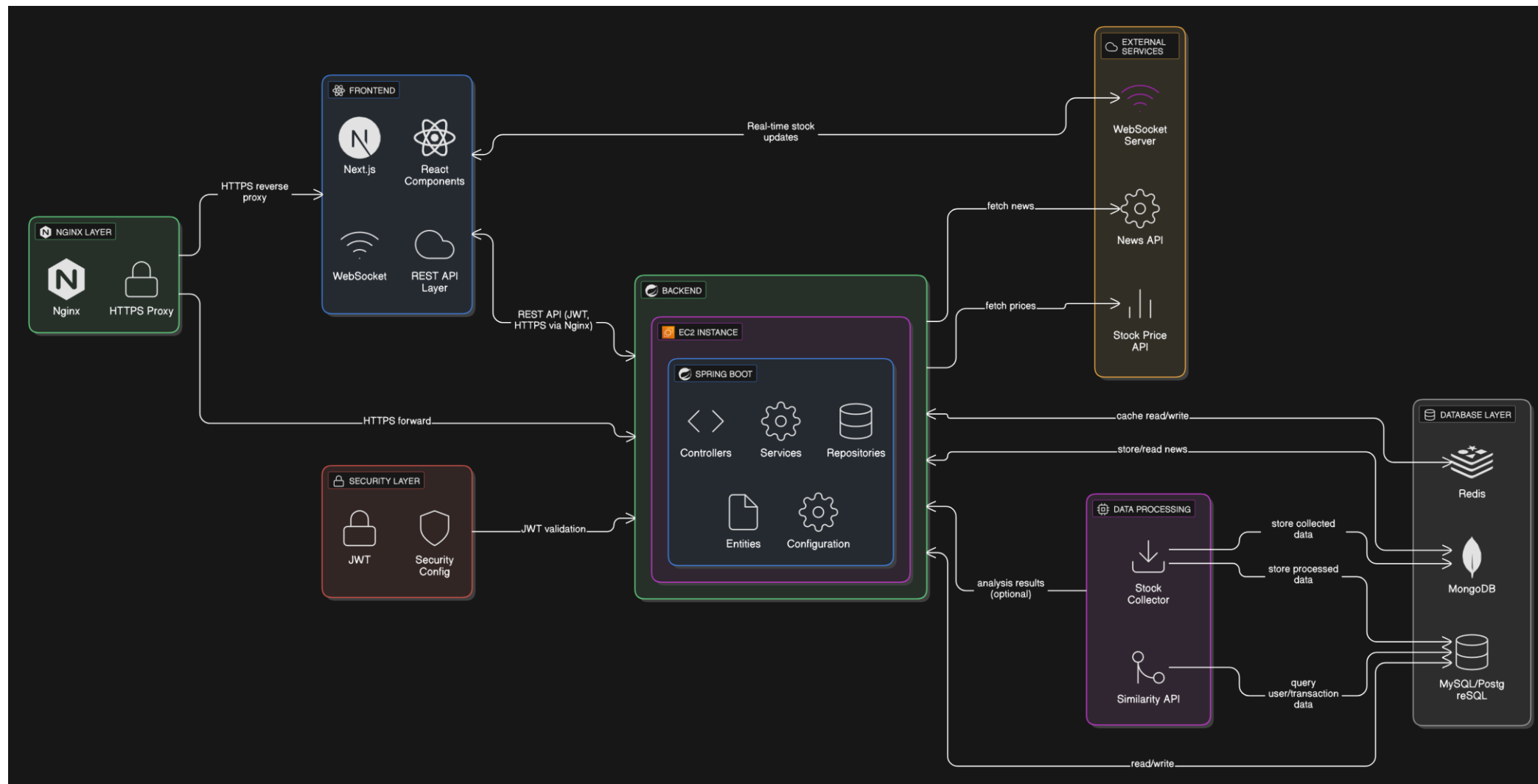
📊 내 투자성향 분석



📌 등록된 투자자 성향 정보는 **2년 동안** 유효해요

완료

아키텍처



감사합니다