로고, 그래픽, 폰트, 상징이(가) 표시된 사진

자동 생성된 설명

[프로그래머스 데이터 분석 데브코스 4기]

**통합 주가 예측 모델 보고서**

: Capm의 베타 및 개별 주식의 기술적 분석, 거시지표, 재무제표 분석을 통해 주가를 예측하는 통합 모델 구현

폰트, 그래픽, 로고, 타이포그래피이(가) 표시된 사진

자동 생성된 설명

2024. 11. 26.

팀 명 : QuantVerse

팀 원 : 백준원, 조재상, 조혜진, 최동주

**목 차**

[1. 프로젝트 개요 4](#_Toc183548741)

[1.1. 프로젝트 목적 4](#_Toc183548742)

[2. 데이터 및 모델 구성 5](#_Toc183548743)

[2.1. 데이터 소스 5](#_Toc183548744)

[2.2. 사용 칼럼 6](#_Toc183548745)

[2.3. 모델 구성 6](#_Toc183548746)

[2.3.1. 경기선행지수를 기반으로 코스피 지수 예측 6](#_Toc183548747)

[2.3.2. CAPM 모델을 활용한 주식 가격 평가 7](#_Toc183548748)

[2.3.3. 산업군 분석 및 개별 기업 분석. 7](#_Toc183548749)

[2.3.4. 기술적 분석 기반 평가 7](#_Toc183548750)

[2.4. 모델 흐름 8](#_Toc183548751)

[3. 분석 과정 8](#_Toc183548752)

[3.1. 코스피 지수 예측 8](#_Toc183548753)

[3.2. 개별 주식 분석 8](#_Toc183548754)

[3.3. 산업군 분석 9](#_Toc183548755)

[3.4. 기술적 분석 12](#_Toc183548756)

[4. 모델링 16](#_Toc183548757)

[4.1. 경제 예측 모델링 16](#_Toc183548758)

[4.2. 주식 가격 평가 모델링 17](#_Toc183548759)

[4.3. 산업별 모델링 18](#_Toc183548760)

[4.4. 기술적 분석 모델링 21](#_Toc183548761)

[5. 결과 및 논의 23](#_Toc183548762)

[5.1. 경기선행지수를 활용한 코스피 지수 예측 23](#_Toc183548763)

[5.2. 개별 주식 가격 평가 23](#_Toc183548764)

[5.3. 산업분석 결과 25](#_Toc183548765)

[5.4. 기술적 분석 결과 26](#_Toc183548766)

[6. 결론 및 향후 과제 28](#_Toc183548767)

# 1. 프로젝트 개요

## 1.1. 프로젝트 목적

본 프로젝트는 다양한 기법을 활용하여 경제상황을 예측하고, 이를 바탕으로 통합적인 주가 예측 모델을 개발하는 것을 목표로 한다. 구체적으로 주식 자체의 자료만을 사용하여 분석하는 기술적 분석과 개별 기업의 재무제표를 분석하는 기본적 분석, 개별 주식의 수익률과 위험에 기반한 자산가격결정모형(CAPM) 및 기타 거시경제 지표를 활용한 주식 가격 예측 모델을 만들 예정이다.

본 프로젝트에 앞서 기존에 진행된 다양한 주가 예측 모델에 대해 소개하고 그 한계를 확인했다. 먼저 ARIMA모델을 활용한 주가 예측 방법론이 있다. 시계열 데이터를 분석하고 미래를 예측하는데 사용되는 통계적 기법으로서, 자기회귀, 차분, 이동평균을 결합한 모델이다. 해당 모델을 활용하여 주가의 등락을 예측한 결과 가장 높은 값으로 60%를 보고하는 논문이 있었다. 하지만 선형적인 결과값을 보여주기 때문에 정확한 가격을 제시하기 어렵고, 일반적으로 상승장이 오래 지속되었던 과거 데이터를 감안하면 지금과 같은 하락 추세를 예측하기 어렵다는 단점이 있다.

다음으로 LSTM 및 GRU를 활용한 주가 예측 방법론이 있다. 이러한 딥러닝 모델은 입력 데이터의 특징 패턴을 학습하여 기존 통계 및 머신러닝 기법보다 주가 예측 성능이 우수하다고 평가된다. 선행 연구 결과에 따르면 주가 예측 모델이 수익이 같은 기간 전체 시장의 수익대비 미미하였다. 또한 동 모델을 활용하였을 때 데이터 수의 제한으로 모델의 복잡성으로 인한 과적합 문제가 발생하고 ARIMA와 마찬가지고 상승장의 데이터를 학습하였을 때 지금과 같은 하락 추세 예측이 어렵다는 단점이 있다.

이 외에도 SVM, HMM 등의 모델을 활용한 주가예측 기법이 있었다. 위의 모델을 활용한 논문 및 프로젝트에서는 모델이 어떤 데이터를 학습하는 지와 외부의 경제 충격을 어떤 방식으로 모델에 적용했는지에 따라 예측 정확도가 달라졌지만, 대부분의 프로젝트에서 한 개 내지 다섯 개의 종목 주가 예측을 하여 일반화시키기 어렵다는 한계점이 존재하였다. 이에 본 프로젝트에서는 KOSPI 200 (시가총액 기준 상위 200개 기업)의 기업을 선정하여 INPUT 값으로 KOSPI200에 해당하는 종목명을 넣으면 한 달 후의 주가를 예측하는 조금 더 일반화된 모델을 더 높은 확률로 예측하고자 한다. 이에 아래의 다섯가지 기준을 본 모델에 적용하여 예측 모델을 구현하고자 한다.

* 경기선행지수
* CAPM 기반 개별 종목의 민감도
* 특정 산업의 민감도
* 재무제표 지표(수익성, 안정성, 성장성 등)
* 기술적분석

**역할 분담**

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

# 2. 데이터 및 모델 구성

## 2.1. 데이터 소스

**경기선행지수**: 향후 경제 활동의 전반적인 방향을 예측하기 위한 지표로, 통계청에서 제공하는 경기종합지수 데이터를 활용했다.

**개별 주식 데이터**: 한국거래소(KRX) 통계자료를 통해 특정 종목별로 데이터 수집이 가능했다. 기간을 정할 수 있어 시장 동향과 특정 주식의 과거 흐름을 분석할 수 있도록 구성했다.

**기업별 재무제표** : Dart 전자공시 통계자료로 개별 기업의 분기, 반기, 사업 보고서를 활용하여 재무분석 자료를 활용하였다.

**추가 지표**: 월별/연도별 주요 경제 지표와 기존 데이터 소스에서 가져온 자료를 통해 만든 보조지표(기술적 분석 지표)이다.

**데이터 소개**

|  |  |  |  |
| --- | --- | --- | --- |
| 데이터 이름 | 행 수 | 데이터 소스 | 설명 |
| kospi\_1year\_data.csv | 11421 | KRX정보데이터시스템 API | 1년간 일별 코스피 지수 데이터 |
| krx\_data.csv | 232017 | KRX정보데이터시스템 API | 1년간 일별 종목의 데이터 |
| rf.csv | 1 | KRX정보데이터시스템 API | 무위험 이자율 값 |
| leadingIndex.csv | 35 | ECOS API | 3년간 경기선행지수 값 |
| final\_result\_with\_beta.csv | 200 | rf.csv, kospi\_1year\_data.csv  krx\_data.csv | CAPM 모델의 결과 값(베타). |
| result\_data.csv | 3369 | Dart API | 산업군별 재무비율 데이터 |

**데이터 전처리**

본 프로젝트에 사용할 데이터는 주식시장의 일별 데이터 및 전자공시가 된 분기별 자료 등으로 결측치가 없어 전처리 과정에서 string으로 가져온 날짜 데이터(BAS\_DD)에 대해 시계열 자료(datetime)로 변환하는 작업을 수행하였다. 또한 날짜, 종목 이름, 지수 이름을 제외한 나머지 칼럼은 모두 정형화된 수치형 변수로 추가적인 전처리 작업은 없었다.

## 2.2. 사용 칼럼

|  |  |  |  |
| --- | --- | --- | --- |
| 칼럼명 | 내용 | 칼럼명 | 내용 |
| BAS\_DD | 시점 | CLSPRC\_IDX | 지수 종가 |
| ISU\_NM | 종목이름(주식/국채) | Market\_Returns | 지수 수익률 |
| Stock\_Returns | 종목 수익률 | BND\_EXP\_TP\_NM | 국채만기 |
| GOVBND\_ISU\_TP\_NM | 지표여부 | ISU\_CD | 종목코드 |
| TDD\_CLSPRC | 개별 종목 종가 | 'TDD\_LWPRC | 저가 |
| ACC\_TRDVOL | 거래량 | IDX\_NM | 지수이름 |
| MKTCAP | 시가총액 | CLSPRC\_YD | 이자율 |
| LIST\_SHRS | 발행주식수 | TDD\_HGPRC | 고가 |
| 공시날짜 | 전자공시 날짜 | 시계열 | 보고서 발표 연/분기 |
| ROE | 자본대비 수익률 | 영업이익률 | 영업이익/매출액 |
| 순이익률 | 당기순이익/매출액 | 매출총이익률 | 매출총이익/매출액 |
| 부채비율 | 부채/자기자본 | 유동비율 | 유동자산/유동부채 |
| 매출액\_증감률 | 분기별 매출 증감률 | 영업이익\_증감률 | 불기별 영업이익 증감률 |

## 2.3. 모델 구성

### 2.3.1. 경기선행지수를 기반으로 코스피 지수 예측

통계청에서 제공하는 경기선행지수를 기반으로 가까운 미래의 경제상황을 예측하였다. 경기선행지수란 7가지의 세부 지표를 바탕으로 산출한 값으로 100을 기준점으로 삼아 더 높은 값이 나오면 미래의 경제상황이 좋아지는 것을 의미하고 더 낮은 값이 나오면 미래의 경제상황이 나빠지는 것을 의미한다.

본 모델은 개별 주식의 가격이 시장의 흐름을 추종한다는 가정을 세우고 경기선행지수가 주가의 흐름을 미리 반영한다는 점을 활용하여 미래의 주식가격의 추세를 예측하고자 하였다.

### 2.3.2. CAPM 모델을 활용한 주식 가격 평가

CAPM은 포트폴리오의 자산가격을 결정하는 이론적 모형으로 해당 이론의 베타(B)는 개별 종목이 전체 시장의 수익률에 대비해 얼마나 민감한지를 나타낸다. 이에 따라 전체 시장의 흐름에 대비해 개별 종목이 어느 정도 상승, 하락하는지 알려주는 지표로써 활용된다.

이에 따라 CAPM모델 활용을 위해 1년간의 지수 수익률을 기준으로 베타 값을 개별 주식의 예상 주가를 평가한다. 이러한 메커니즘을 활용하기 위해 미래의 지수가 어떻게 움직일지 예측해야 한다. 이를 위해 경제상황을 예측하는 모델과 합쳐서 개별주식의 평가를 수행했다.

### 2.3.3. 산업군 분석 및 개별 기업 분석.

Dart에서 수집한 5개년 재무제표 데이터를 바탕으로 주요 재무 지표(매출액, 영업이익, ROE, 부채총계 등)를 계산하여 각 산업별 주가 민감도 분석 결과를 기반으로 산업군을 분류하고, 주요 변수에 따라 가중치를 설정한다.

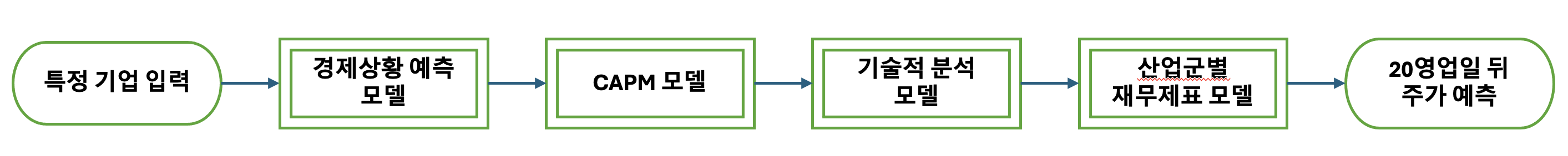
KOSPI 200 기업의 주가 데이터와 재무제표 데이터를 병합하여 시계열 데이터셋을 구축하였으며 업별 가중치를 기반으로 개별 기업의 주가 민감도 및 예측 가능성을 평가하였다.

### 2.3.4. 기술적 분석 기반 평가

기술적 분석은 금융 시장을 분석하고 예측하는 기법 가운데 하나로 주로 시세 동향 그래프를 이용해 분석했다. 이동평균선, RSI, MACD등의 많은 보조지표와 도구를 사용하여 주가의 변동성 분석한다. 이를 통해 저평가 및 고평가 여부를 판단하고 주가를 예측했다.

## 2.4. 모델 흐름

위의 모델을 각각 구현하여 개별 모델의 유효성을 검증한 후 아래의 흐름으로 본 프로젝트의 목적에 따라 20영업일 후의 주가를 구한다. 흐름은 아래와 같다.



# 3. 분석 과정

## 3.1. 코스피 지수 예측

경기선행지수를 분석하여 6개월 이내의 코스피 지수값을 예측한다. 통계청에서 제공하는 경기선행지수 순환변동치를 활용한다. 먼저 코스피와 경기선행지수 순환변동치의 상관관계를 계산하였다. 코스피 지수데이터는 경기선행지수 순환변동치와 상관관계가 대략 0.43~0.48 사이로 나왔으며 이는 상관관계가 높은 편으로 경기선행지수를 활용하여 미래의 주식가격 예측을 하는 것이 관련성이 있다고 판단되었다.

텍스트, 폰트, 화이트, 타이포그래피이(가) 표시된 사진

자동 생성된 설명

그림 1) 경기선행지수와 코스피 지수의 상관관계 분석 결과

최근 3년간 데이터를 기반으로 경기 변화 패턴 탐지하여 추후 코스피의 가격이 어떻게 변화할지 선형 그래프로 예측하고 본 데이터를 활용하여 향후 CAPM 모델에서 개별 주식가격의 민감도를 의미하는 베타값에 예측값을 가중치로 활용하여 모델을 구현한다.

## 3.2. 개별 주식 분석

CAPM모델에 포함되어 있는 시장수익률, 무위험이자율, 개별 종목의 수익률을 구하기 위해 KRX API에 있는 KOSPI지수의 수익률, 국채 3년물의 이자율, 종목별 수익률을 사용했다.  
먼저 국채 3년물 이자율과 KOSPI지수의 수익률은 API에서 제공하는 값을 그대로 사용하였다. 해당 값은 파라미터로 바로 전달되어 미래 가격 예측에 쓰이도록 구성했다.  
CAPM을 구현하기 위해 개별 종목의 베타값을 구한 후 기대수익률을 계산하였다. 이후 출력된 기대수익률에 1달 후의 값을 계산하기 위해 20영업일을 곱하여 계산했다.

## 3.3. 산업군 분석

영업이익, 부채총계와 같은 절대적인 크기의 지표는 기업 간 규모 차이로 인해 모델링 시 불리할 수 있다는 점을 고려하여, 다음과 같은 상대적 지표를 계산하여 사용하였다

**(1) 수익성 지표**

- ROE (자기자본이익률): 기업이 주주의 투자금을 얼마나 효율적으로 사용하여 이익을 창출하는지 보여준다.

- 영업이익률: 기업의 주요 영업활동에서의 수익성을 나타낸다.

- 순이익률: 모든 비용과 세금을 고려한 후의 최종적인 수익성을 보여준다.

- 매출총이익률: 제품이나 서비스의 기본적인 수익성을 나타낸다.

**(2) 안정성 지표**

- 부채비율: 기업의 재무 레버리지와 장기 지급능력을 평가한다.

- 유동비율: 단기 부채 상환 능력을 나타내며, 기업의 단기 재무 건전성을 평가한다.

- 고정비율: 자기자본으로 비유동자산을 어느 정도 충당하고 있는지 보여준다.

**(3) 성장성 지표**

매출액 증감률: 기업의 시장 점유율 변화를 나타낸다.

영업이익 증감률: 수익성 개선 여부를 나타낸다.

Dart에서 수집한 5개년 재무제표 데이터를 바탕으로 위 지표를 계산하였다.

각 산업군에 대해 주요 재무지표와 종가 간 상관관계를 분석했다. 이를 통해 각 산업별로 어떤 재무지표가 주가에 큰 영향을 미치는지를 확인할 수 있었다. 산업특성과 데이터 부족의 이유로 보험, 은행, 증권업은 분석에서 제외하였다.

**주요 산업별 상관 분석 결과**

1. **의약품**

영업이익률과 순이익률의 상관관계가 강함(0.67), 이는 의약품 산업에서 수익성 지표가 밀접하게 연동된다. ROE와 순이익률의 높은 상관관계(0.59), 수익성이 주가에 긍정적인 영향을 미칠 가능성이 높다.

1. **운수창고업**

영업이익률과 순이익률의 상관관계가 매우 강하다.(0.86) 이는 운수창고업에서도 수익성이 중요한 지표임을 나타낸다. 종가와 영업이익률의 강한 양의 상관관계(0.42), 운수창고업의 주가가 영업수익에 민감하게 반응한다. ROE, 순이익률도 각각 0.28 0.32 로 종가와 양의 상관관계를 가지고 있으나 매출 총이익률과는 -0.22의 상관관계를 가지는데, 이는 운수창고업에서 단순한 매출 마진보다는 전반적인 운영 효율성과 자본 활용도가 더 중요함을 의미한다.

1. **건설업**

종가와 영업이익률의 강한 양의 상관관계를 나타낸다(0.70). 영업수익이 주가에 직접적으로 영향을 미친다. 부채비율과 ROE 간 음의 상관관계를(-0.59)보임에 따라 부채 관리가 중요하다는 걸 알 수 있다. 매출액 증감률과 영업이익 증감률이 강한 양의 상관관계를(0.6) 보인다

1. **전기가스업**

종가와 연관 있는 지표가 많은데 영업이익률, 순이익률, 유동비율과 한 양의 상관관계를 가진다(각각 0.49, 0.43, 0.51). 즉 전기가스업은 이익지표의 수치가 클수록 주가가 상승한다. 매출총이익률과 종가의 강한 양의 상관관계(0.52), 매출총이익이 주가에 긍정적 영향을 미친다.

1. **기타제조업**

ROE와 영업이익률의 높은 상관관계를 가진다(0.73). 자본 효율성이 수익성과 밀접하다. 종가와 영업이익률, roe, 순이익률, 부채비율, 고정비율 모두 대략 0.5이상으로 양의 상관관계를 보였으며 매출총이익률과 유동비율과는 -0.47의 음의 상관관계를 보였다.

1. **유통업**

순이익률과 유동비율의 강한 상관관계를 보인다(0.71). 유통업의 수익성과 유동성 간 밀접한 연관성을 보인다. 종가와 매출총이익률이 양의 상관관계를 보인다(0.3).

1. **철강금속**

ROE와 순이익률의 강한 양의 상관관계를(0.65) 나타낸다. 자본 효율성과 수익성이 핵심 변수이다.

1. **전기전자**

종가와 ROE의 양의 상관관계(0.15), 자본 효율성이 주가에 미치는 영향은 상대적으로 낮았다.

또한 산업별로 각 상대적 재무지표의 평균을 구해서 비교분석 하였다. 결과는 다음과 같다.

텍스트, 스크린샷, 소프트웨어, 번호이(가) 표시된 사진

자동 생성된 설명

이러한 결과를 바탕으로 산업군을 분류하였다.

## 3.4. 기술적 분석

**목적**: 주가의 흐름을 예측하거나 매수/매도 시점을 판단한다.

**방법**: 다양한 지표를 조사 및 분석하고 기대 효과 분석한다.

**이동평균선 (Moving Average)**

특정 기간의 평균 주가를 선으로 나타내 주가의 추세를 시각화. 일정 기간 동안의 종가의 평균을 계산하여 구함. (예시: 5일 이동평균선은 경우 최근 5일 간의 종가의 평균임) 장기 및 단기 추세 분석에 유용하며, 골든크로스나 데드크로스를 통해 매수/매도 신호를 제공

**RSI (Relative Strength Index)**

주가의 과매수/과매도 상태를 판단하기 위해 0~100 사이의 값으로 표시되는 모멘텀 지표. 매도/매수 타이밍을 잡는데 활용. 일정 기간 동안의 상승폭과 하락폭을 이용하여 계산하며 보통 14일을 기준으로 상승폭의 평균과 하락폭의 평균을 이용해 현재 주가의 상대적인 강도를 나타냄. 70 이상일 경우 과매수 구간, 30 이하일 경우 과매도 구간

**볼린저 밴드 (Bollinger Bands)**

주가의 변동성을 시각화한 밴드로, 이동평균선을 중심으로 상단과 하단에 위치. 밴드의 확장과 축소를 통해 변동성을 파악 및 밴드 경계에서의 반등을 매매 전략으로 활용 가능. 이동평균선과 표준편차를 이용해 상단 밴드와 하단 밴드를 계산하며, 밴드 경계에서의 반등을 매매 전략으로 활용할 수 있음.

**PSAR (Parabolic Stop And Reverse)**

주가가 추세 전환 시점을 나타내는 점으로 표시. 추세의 방향성과 추세 종료 지점을 예측해 매수/매도 타이밍을 결정

**ATR (Average True Range)**

주가의 변동성을 측정하는 지표로, 평균 진폭을 나타냄. 변동성이 큰 시점을 파악하여 리스크 관리에 활용. 특정 기간 동안의 진폭(고가와 저가의 차이, 혹은 전일 종가와의 차이 등)을 이용해 평균 변동성을 계산함.

**MACD (Moving Average Convergence Divergence)**

이동평균선의 차이를 활용해 추세의 강도와 방향성을 나타냄. MACD선과 시그널 선의 교차를 통해 매매 신호 포착 가능. 단기 이동평균선과 장기 이동평균선의 차이를 구하여 계산함. MACD선과 시그널 선의 관계를 통해 상승하락을 예측함.

**오실레이터 (Oscillator)**

일정 범위 내에서 움직이며, 주가의 과매수/과매도 상태를 나타냄.. 주가의 전환 시점을 예측하여 단기 매매 전략에 활용. 오실레이터 값에 따라 상승하락을 예측함.

**하한선/상한선**

주가의 일정 범위 내 상한과 하한을 나타내는 선. 돌파 여부에 따라 매매/매도 신호를 제공. 선을 넘어서는걸 기준으로 상승하락을 예측함.

**저항선/지지선 (Resistance/Support Levels)**

주가가 상승하거나 하락을 멈추는 경계선. 주가가 저항선을 돌파하거나 지지선을 이탈할 때 중요한 매매 신호를 제공. 선을 넘어서는걸 기준으로 상승하락을 예측함.

**%K, %D (Stochastic Oscillator)**

주가의 현재 위치를 일정 기간 동안의 고점과 저점 사이에서 비율로 나타내는 지표. 둘의 교차를 통해 매매 타이밍을 파악 가능. %K는 현재 주가가 일정 기간 동안 고점과 저점 사이에서 어디에 위치하는지를 나타내며, %D는 %K의 3일 이동평균선임. %K와 %D의 값과 관계를 통해 상승하락을 예측함.

**CCI (Commodity Channel Index)**

주가의 추세와 평균 가격 사이의 차이를 측정. 주가의 과매수/과매도 상태를 파악해 매매 기회를 제공. CCI 값에 따라 상승하락을 예측함.

**ADX (Average Directional Index)**

추세의 강도를 나타내는 지표, 방향성과는 무관함. 강한 추세와 약한 추세를 구분해 추세 기반 전략 수립에 도움을 줌. PDI(Positive Directional Indicator)와 NDI(Negative Directional Indicator)를 이용해 추세의 강도를 측정하며, 100에 가까울수록 강한 추세, 0에 가까울수록 약한 추세임. ADX의 값을 어느 기점으로 잡냐에 따라 상승하락을 용이하게 예측할 수 있음.

**이격도 (Price Oscillator)**

주가와 이동평균성 간의 괴리율을 나타내는 지표. 주가가 평균에서 얼마나 떨어져 있는지 파악하여 과매수/과매도 상태를 분석. 주가가 이동평균선에서 얼마나 떨어져 있는지를 백분율로 나타낸 값. 이격도의 비율을 통해 상승하락을 예측함.

**Envelope**

이동평균선을 기준으로 일정 비율만큼 상하로 이동한 선으로 구성된 지표. 이동평균선에서 일정 비율 만큼 상하로 이동한 상한선과 하한선으로 구성되며, 주가의 추세와 반전 가능성을 파악해 매매 전략을 도출

**OBV (On-Balance Volume)**

거래량과 주가의 움직임을 결합한 지표로, 거래량의 흐름을 추적. 거래량 변화를 통해 주가의 상승/하락 가능성을 예측. 주가가 상승할 때 거래량을 더하고, 주가가 하락할 때 거래량을 빼는 방식으로 계산됨. 값의 기울기에 따라 상승하락을 예측함.

**Klinger (Klinger Oscillator)**

거래량 흐름과 주가의 추세를 결합한 지표. 장기적인 자금 흐름을 파악해 추세 변화 신호를 감지. 값이 양수에서 음수, 또는 음수에서 양수로 전환되는 것에 따라 상승하락을 예측.

**PDI (Positive Directional Indicator)**

일정 기간 동안의 상승폭을 기반으로 계산된 지표, 상승 추세의 강도를 나타냄 일정 기간 동안의 고가와 저가를 비교하여 상승폭을 측정하고, 강한 상승 흐름일 때 매수 신호를 제공

**NDI (Negative Directional Indicator)**

일정 기간 동안의 하락폭을 기반으로 계산된 지표, 하락 추세의 강도를 나타냄 일정 기간 동안의 고가와 저가를 비교하여 하락폭을 측정하고, 강한 하락 흐름일 때 매도 신호를 제공

# 4. 모델링

## 4.1. 경제 예측 모델링

**목적**: 경기선행지수를 활용하여 코스피 지수 예측.

**방법**: 통계적 회귀분석 및 시계열 예측 모델 적용. 통계적인 회귀분석을 통해 그린 경기선행지수와 코스피지수의 분석을 통해, 본 데이터가 모델에 반영되기에 적합한지 판단하여 모델링 과정에 접합한다.

**결과**: 예측 모델의 성능 지표(RMSE)와 실제 경기 데이터 비교 및 회귀분석 진행 결과 경기실사지수는 코스피 지수의 단기적인 미래 예측에 설명력이 낮아 모델링에서 제외하기로 했다.

텍스트, 폰트, 화이트, 영수증이(가) 표시된 사진

자동 생성된 설명

본 예측 방법에 따라 RMSE를 도출했을 때, 실제 지수와 약 150포인트 가량 차이가 남을 알 수 있었다. 이는 실제 코스피 지수의 변동폭에 대해 의미있는 값을 나타내지 못한다고 추측된다. 특히 경기선행지수 자체에 코스피 지수의 값이 들어있기 때문에 종속변수의 값이 다시 독립변수에 영향을 미치는 피드백 효과가 나타나고, 이는 경기선행지수와 코스피 지수가 등락을 대체로 동시에 하는 그래프를 그림으로써 나타나고 있다.  
텍스트, 라인, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

아래는 경기선행지수와 코스피 종가의 회귀분석을 3,4,5,6개월 후의 지수값과 비교하여 회귀분석한 결과 그래프이다. 일부 선형적인 관계가 나타나는 것을 확인할 수 있다.

|  |  |
| --- | --- |
| 텍스트, 라인, 스크린샷, 그래프이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 라인, 스크린샷, 그래프이(가) 표시된 사진  자동 생성된 설명 |
| 텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진  자동 생성된 설명 |

텍스트, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

또한 회귀분석을 진행 후 경기선행지수와 코스피 지수의 관계를 값을 도출하여 어느정도 설명력을 갖추었는지 확인했다. 그 결과 대략 20% 내외의 값을 보였고, 이는 독립변수인 경기선행지수가 종속변수인 수개월 후 코스피 지수의 변동성에 대략 20% 정도만을 설명할 수 있다는 것을 의미한다. 경기선행지수가 개별 주식 지수의 변동을 예측하는데 사용되는 지수가 아닌 전반적인 경제상황에 대한 흐름을 가늠하는 지표로 사용되기에 개별 주식시장에 대한 예측력은 제한적인 것으로 판단된다.

## 4.2. 주식 가격 평가 모델링

**목적**: CAPM의 베타를 활용한 개별 주식의 민감도를 파악하고 한달 후의 주식가격을 예측하고자 했다.

**방법**: 1년간의 기간을 두고 코스피 지수 수익률 및 개별 주식의 수익률을 계산했다. 또한 한달 이후의 주식 가격을 예측하는 방식으로 모델을 구현중이므로 데이터는 1년 1개월 이전부터 1개월 이전까지의 데이터를 바탕으로 현재의 주가를 예측해보는 방식으로 진행하였다. 이에 따라 이론적 주가와 실제 주가를 비교 분석 할 수 있었다. 이에 따라 모델링에는 11월 20일의 주가를 예측하고 실제 주가와 비교분석 하도록 2023년 10월 22일부터 2024년 10월 22일까지의 자료를 추출하여 종가, 연간 수익률, 베타값을 종목별로 추출하였고 코스피의 연간 수익률 및 무위험수익률(국채 3년물)을 가져왔다.

**결과** : 이에 따라 추출된 아래의 데이터를 파라미터로 해당 주식에 대한 선형 그래프를 그림으로써 한 달 후 주식가격을 예측할 수 있다. 이에 대한 예측 값은 결과란에 기술하였다.  
텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명  
텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명 텍스트, 폰트, 화이트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

## 4.3. 산업별 모델링

**목적**: 산업별 가중치와 재무 지표를 반영하여 주가를 예측한다. 주가 예측값을 기반으로, 이를 통해 매수/매도 판단을 내린다.

**방법**: 분석 결과를 바탕으로 유사한 특성을 가진 산업들을 그룹화하고, 각 산업군의 가중치를 설정하여 분류 모델을 적용한다. 이를 위해 산업별 재무 지표의 중요도를 반영한 가중치를 기반으로 주가 예측 모델을 만들었다. 예측에 사용된 주요 특징들은 산업군, 재무 지표(영업이익률, ROE, 순이익률 등), 과거 주가 데이터이며, 이를 가중치 및 재무 데이터를 추가하여 분류 모델을 훈련하고, 매수/매도 판단을 위한 예측값을 도출한다.

시가총액이 높은 기업은 일반적으로 해당 산업을 대표하는 기업일 가능성이 높으며(대표성), 대기업은 보다 정확하고 신뢰할 수 있는 재무 데이터를 제공하는 경향이 있다. 또한 각 산업의 선도 기업들은 해당 산업의 동향을 주도하는 경향이 있어, 이들 기업의 분석은 산업 전체의 방향성을 예측하는 데 유용할 수 있다.

위와 같은 이유로 각 산업군에서 시가총액이 가장 높은 기업들을 샘플링하여 모델을 예측하였다. 아래는 각 산업군에 반영한 가중치이다.

1. **수익성 중심 산업군** - 음식료품, 의약품, 운수창고업, 화학

특징: 이익률(영업이익률, 순이익률)과 매출 성장률이 중요한 지표로 작용

주요 지표:

- 영업이익률 (가중치: 0.4): 수익성을 나타내는 대표적인 변수.

- 순이익률 (가중치: 0.3): 기업의 최종 이익률

- 매출액 증감률 (가중치: 0.1): 성장성을 평가.

- ROE (가중치: 0.2): 자본 효율성을 간접적으로 반영

1. **안정성 중심 산업군** - 철강금속, 전기가스업, 통신업

특징: 부채비율과 유동비율이 중요한 지표로 작용

주요 지표:

- 부채비율 (가중치: 0.4): 부채 관리를 통한 안정성 평가

- 유동비율 (가중치: 0.3): 단기 지급 능력.

- 고정비율 (가중치: 0.2): 자산 효율성.

- 영업이익률 (가중치: 0.1): 안정성 속에서의 수익성 확인.

1. **성장성 중심 산업군** - 의료정밀, 비금속광물, 전기전자

특징: 매출액 증감률과 영업이익 증감률이 높으며 중요한 지표로 작용

주요 지표:

- 매출액 증감률 (가중치: 0.4): 성장성을 나타내는 핵심 지표

- 영업이익 증감률 (가중치: 0.3): 수익을 동반한 성장을 평가

- ROE (가중치: 0.2): 성장 과정에서의 자본 효율성 평가

- 부채비율 (가중치: 0.1): 성장을 위한 재무 안정성 고려

1. **혼합형 산업군** -건설업, 운수장비, 기계, 기타제조업

특징: 매출액 증감률과 영업이익 증감률이 높으며 중요한 지표로 작용

주요 지표:

- 영업이익률 (가중치: 0.3): 주요 수익성 변수.

- 순이익률 (가중치: 0.2): 최종 수익성.

- 부채비율 (가중치: 0.3): 안정성 변수.

- ROE (가중치: 0.2): 수익성과 안정성 모두를 반영.

5, **소비자 중심 산업군**- 유통업, 섬유의복, 서비스업

특징: 매출총이익률과 주가 간의 강한 상관관계, 소비자 트렌드와 시장 변화에 민감

주요 지표:

- 매출총이익률 (가중치: 0.4): 성장성 변수.

- ROE (가중치: 0.3): 자본 효율성.

- 유동비율 (가중치: 0.1): 안정성.

- 순이익률 (가중치: 0.2): 최종 수익성.

## 4.4. 기술적 분석 모델링

**목적:** 주가 변동성을 분석하여 매수/매도 판단.

주식 시장에서 주가의 변동성 분석을 통해 향후 주가의 상승 또는 하락을 예측하며, 더 나아가 상세한 주가까지 예측하는 것이 이 모델의 주요 목표임. 이를 통해 투자자들이 주가 흐름을 예상하고, 그에 맞는 매매 전략을 수립할 수 있도록 지원함.

**방법:** 이동평균선, RSI, 볼린저 밴드, PSAR, ATR 지표 활용.

**이동평균선**

주가의 5일, 20일, 50일 이동평균선(MA5, MA20, MA50)을 분석해 각 이동평균선의 차이를 통한 단기 상승, 중간 상승, 강한 상승 등의 추세를 점수화 하여 상승 비율과 하락 비율을 계산함. MA5가 MA20보다 높은 상태에서, MA50보다도 높으면 강한 상승 추세로 판단하고 상승 비율에 0.12를 추가함. 그렇지 않다면 중간 상승 추세로 판단하고 상승 비율에 0.09를 추가함. MA5가 MA20보다 낮은 상태에서, MA50보다도 낮으면 강한 하락 추세로 판단하고 하락 비율에 0.12를 추가함. 그렇지 않다면 중간 하락 추세로 판단하고 하락 비율에 0.09를 추가함. MA5가 MA20보다 높다면 단기 상승 추세로 판단하여 상승 비율에 0.07을 추가함. 그 반대라면 단기 하락 추세로 판단하여 하락 비율에 0.07을 추가함.

**RSI (Relative Strength Index)**

과매수 및 과매도 상태를 판단하여, 매수/매도 추세를 확인함. RSI가 70 이상일 경우 과매수로 판단하여 하락 비율을 0.06 증가, 30 이하일 경우 과매도로 판단하여 상승 비율을 0.06 증가시킴

**볼린저 밴드**

주가의 20일 이동평균을 기준으로 위 아래로 표준편차를 더하고 뺀 밴드임. 상단 밴드를 돌파할 경우 과매수 상태로 판단하여 하락 비율을 0.04 증가. 하단 밴드를 돌파할 경우 과매도 상태로 판단하여 상승 비율을 0.04 증가.

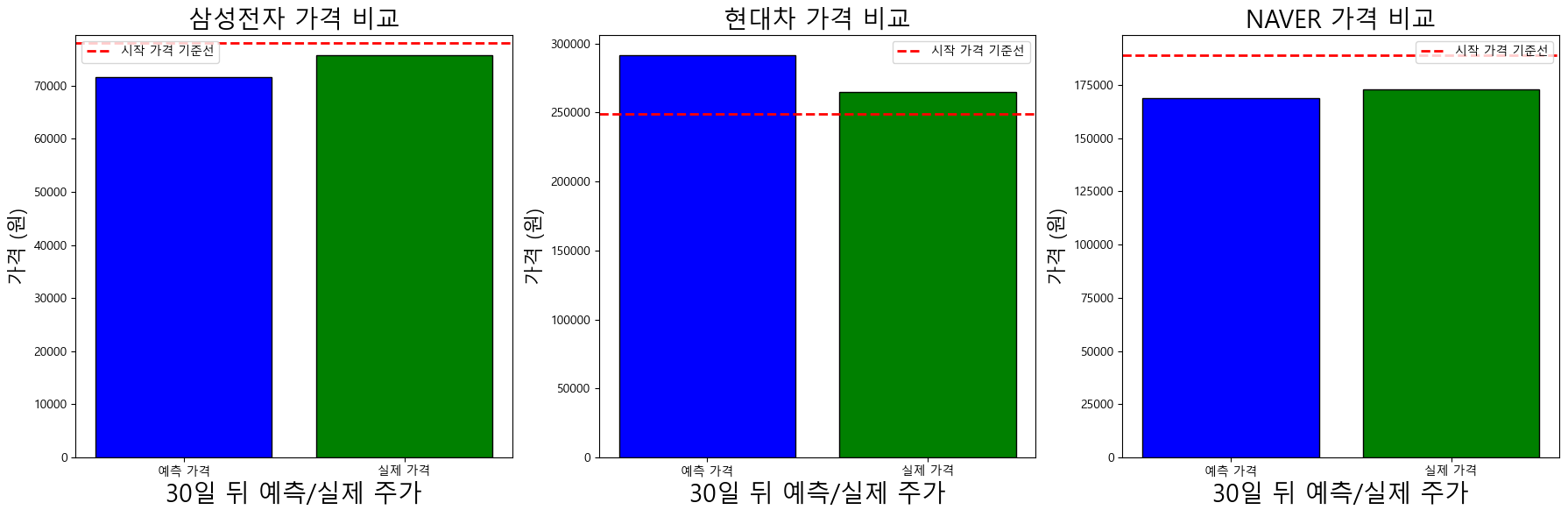
**PSAR (Parabolic Stop And Reverse)**

주가와의 위치 관계를 통해 매수 혹은 매도 신호를 확인하는 지표임. SAR 값이 현재 주가 아래에 위치할 경우 상승 추세로 판단하여 매수 신호를 나타내어 상승 비율을 0.05 증가시킴. SAR 값이 현재 주가 위에 위치할 경우 하락 추세로 판단하여 매도 신호를 나타내어 하락 비율을 0.05 증가시킴

**ATR (Average True Range)**

특정 기간 동안의 주가 변동 폭을 평균한 값으로 주가의 변동성을 측정하는 지표임. 변동성이 증가할 경우 하락 비율을 0.03 증가시키고, 감소할 경우 상승 비율을 0.03 증가시킴.

**결과**: 20 영업일 뒤의 주식 값과 상승/하락에 대한 정확도.



각 데이터는 5월 2일자의 krx open api 데이터를 바탕으로 20 영업일 뒤인 6월 3일의 주가를 예측한 결과이다. 삼성전자, 현대차, NAVER의 결과이고 파란색은 모델로 예측한 30일 뒤의 주가이며, 녹색은 실제 30일 뒤의 주가임. 정확한 주가를 예측하는 것은 실패하였지만, 해당 날짜 기준으로는 상승/하락의 방향성을 예측하는 것은 성공하였다. 특정 날짜만 예측하는 것이 아닌, 약 1년의 기간을 가지고 30일 뒤를 예측하고 실제와 비교하여 상승/하락의 방향성에 대한 정확도를 분석할 예정이다.

# 5. 결과 및 논의

## 5.1. 경기선행지수를 활용한 코스피 지수 예측

경기선행지수로부터 코스피 지수를 예측하기 위해 경기선행지수 순환변동치를 사용하였다. 일반적으로 경기선행지수는 단기간의 미래 경기를 예측하는데 사용되므로 3~6개월 후의 코스피 지수를 예측하는데 쓰일 것이라 가정하고 회귀분석 진행 후 성능을 확인하였다. 이에 RMSE값이 150가량으로 지수의 한달 변동폭이 과하게 크게 나왔다. 또한 값이 3개월부터 6개월 후의 값을 예측하는데 대략 20% 내외의 설명력을 지녔음을 확인하였다. 따라서 경기선행지수를 사용하여 코스피 지수를 예측하는 본 모델은 통합 모델에 사용하기 적합하지 않다 판단하여 제외시켰다.

## 5.2. 개별 주식 가격 평가

**이론적 주가와 실제 주가 비교 분석.** (2024년 11월 20일 종가 기준으로 비교하였음)

|  |  |
| --- | --- |
|  |  |
|  |  |

'종목명': '삼성전자', '현재 주가': 57700

'기대 수익률': 0.12186195859955674

'1개월 후 예상 주가': 58228

실제 주가 : 55300

'종목명': 'SK하이닉스', '현재 주가': 187800

'기대 수익률': 0.16564788503711314

'1개월 후 예상 주가': 190098

실제 주가 : 170600

'종목명': '현대차', '현재 주가': 234500

'기대 수익률': 0.11426169460464662

'1개월 후 예상 주가': 236522

실제 주가 : 218500

'종목명': '기아', '현재 주가': 92500

'기대 수익률': 0. 11494750700913818

'1개월 후 예상 주가': 93302

실제 주가 : 98100

**실제 주가와의 차이를 분석하여 개선 가능성 논의.**  
CAPM모델이 전체 지수에 대비한 개별 기업의 민감도를 통해 위험 대비 수익률을 계산하기 때문에, 실제 주가와 예상 주가의 차이인 잔차는 개별 산업군의, 기업의 변동성을 의미한다. 그렇기 때문에 산업분석, 기업분석, 기술적 분석을 통해 추가적인 보정을 하여 더 개선된 예측을 할 수 있으리라 예측한다.

## 5.3. 산업분석 결과

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명MAE: 13705.81

MSE: 1624363245.79

R^2: 0.97

**weighted\_score**는산업별 가중치를 반영하여 계산된 점수로, 각 기업의 재무 성과를 나타내는 종합적인 지표이다. 이 값은 특정 기업의 재무 건전성, 성장성, 수익성 등을 반영하며, 다양한 재무 지표들(예: 영업이익률, ROE, 부채비율 등)의 가중 평균으로 계산된다.

이 점수가 주가 예측에 비교적 큰 영향을 미친다는 것은, 기업의 종합적인 재무 상태가 주가 변동성을 예측하는 데 중요한 요소라는 것을 의미한다.

텍스트, 스크린샷, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 소프트웨어, 컴퓨터 아이콘이(가) 표시된 사진

자동 생성된 설명

공시날짜에서 1달 뒤의 주가 방향성 모멘텀이 크다고 판단하여 2024.08.14일 기준으로 힌달 뒤 주가를 예측하였다. 주가의 방향성 모멘텀은 반기보고서나 분기보고서를 기준으로 계산되며, 공시일 이후의 주가 변동을 반영하는 데에 중요한 요소로 된다.

**한계점**

데이터의 한계로 인해 예측 정확도가 다소 떨어질 수 있다. 특히 주가와 관련된 외부 요인(시장 상황, 정치 경제적인 요인 등)이 모델에 반영되지 않기 때문에, 재무 지표만으로는 한계가 있다.

예를 들어, 공시된 재무지표가 시장의 실시간 변화를 반영하기에는 한계가 있을 수 있다.따라서, 다양한 변수를 모델에 추가하여 예측의 정확도를 높이는 방법을 고려할 수 있다.

## 5.4. 기술적 분석 결과

**주요 기술적 지표의 변화와 향후 주가 흐름 예측 결과.**

A group of colorful bars

Description automatically generated with medium confidence

기술적 분석 기반 모델은 주식의 기술적인 지표 (이동평균선, RSI 등등) 5 가지를 활용하여 각 지표들이 주가의 상승/하락에 어느정도 영향을 미치는지를 분석하여 20 영업일 뒤에 몇 % 주가가 상승할지, 하락할지를 계산하여 예측 주가를 나타내는 모델이다. 모델에 사용된 지표들 중에서 이동평균과 PSAR이 주가 예측에 가장 큰 영향이 있었고, 다른 지표들은 영향이 비교적 미미하였다.

위 그래프는 삼성전자, NAVER, 현대건설, 현대차, LG화학, 삼성물산 총 6개의 24년 10월 16일자에서 24년 11월 13일의 주가를 예측한 결과이다. 삼성전자, 현대건설, 삼성물산의 경우 비교적 예측 주가와 실제 주가가 유사하게 나타났다. NAVER의 경우 30일 뒤에 주가가 상승한다는 것은 일치하나, 그 수치가 상대적으로 부정확했다. 현대차와 LG화학의 경우는 예측한 주가의 차이도 심하며, 기준일 대비 상승/하락의 방향 또한 일치하지 않았다. 기술적 분석 기반 모델의 정확도를 평가하기 위해 약 1년간의 데이터를 사용하여 30일 뒤 주가를 예측하였다. 예측 결과는 실제 주가와 동일한 상승/하락 방향을 가지는지를 확인하고, 예측값과 실제값 간의 RMSE를 통해 정확도를 측정하였다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 종목 | 삼성전자 | NAVER | 현대건설 | 현대차 | LG화학 | 삼성물산 |
| 정확도 | 43.97% | 48.16% | 43.45% | 45.54% | 38.74% | 36.64% |
| RMSE | 14001.78 | 30533.65 | 5231.23 | 49953.07 | 78307.36 | 30781.98 |

기술적 분석을 기반으로 20 영업일 뒤의 주가 상승 및 하락 방향성을 예측한 모델의 정확도는 종목별로 약간의 차이가 나타났으며 평균 정확도는 약 45%, 미래 주가 예측이 아직 불완전하다. 해당 모델이 이동평균선과, RSI 등 단기적인 변동성은 잘 반영하였으나, 30일 뒤의 주가이므로 외부적인 요인이나 비정상적인 시장 움직임을 충분하게 반영하지 못했다. 만약 머신러닝 기반의 비선형 예측 모델을 도입하여 지표 간의 상호작용을 더 정교하게 분석한다면 정확도를 높일 수 있을 것으로 기대된다.

# 6. 결론 및 향후 과제

**개별 모델 논의 결과 수정된 통합 모델 흐름**

개별 모델을 분석 및 결과 논의를 통해 모델에 통합하였을 때 위에서 논의한 바와 같이 부적합한 가정이 있었다. 먼저 경기선행지수로 코스피 지수 예측이 낮은 설명력을 지녔고, 산업군별 재무비율 분석 또한 적시성이 부족해 일별로 변화하는 주가를 예측하는데는 적합하지 않았음을 확인하였다. 이에 2번에서 다뤘던 모델의 흐름을 수정하여 최종 모델로 통합하였다.

텍스트, 폰트, 라인, 화이트이(가) 표시된 사진

자동 생성된 설명

특정 날짜를 입력하였을 경우 CAPM모델로 먼저 예상주가를 측량하고 개별기업의 기술적 분석기법을 활용하여 20일 후 주가를 예측하는 모델로서 최종 모델을 구현하였다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 종목 | 삼성전자 | NAVER | 현대건설 | 현대차 | LG화학 | 삼성물산 |
| 정확도 | 91.62% | 50.78% | 59.16% | 66.49% | 73.29% | 62.3% |
| RMSE | 10850.47 | 25629.68 | 4480.46 | 38103.2 | 54206.05 | 22064.83 |

**통합 주가 예측 모델의 성능 평가.** A graph of different colored bars

Description automatically generated

본 통합 모델은 특정 기준일에서 CAPM 기반의 예측 모델로 1차로 주가를 예측하고, 해당 예측 주가를 기술적 분석 모델로 2차로 한 번 더 처리하여 구체적인 주가를 예측하도록 모델링 하였다.

먼저 본 모델을 활용하여 주가의 상승/하락 방향성을 예측했다. 그 결과 각 종목에 대한 통합 모델은 평균 정확도가 60% 이상으로 상승하였다. 삼성전자의 경우 91.6%의 높은 정확도를 기록하였고, LG화학과 현대차 또한 각각 73.3%, 66.5%의 정확도를 나타낸다. 이는 통합 모델이 각각 개별 모델들에 비해 개선되었음을 보여준다. 하지만 일부 종목에서는 여전히 비교적 낮은 정확도를 보였으며, 특히 NAVER와 삼성물산의 경우 각각 정확도가 50.8%, 62.3%에 그쳤다. 이는 통합 모델이 모든 종목에 대해 일관된 성능을 보이지 않음을 시사하며, 특정 산업군이나 종목 특성에 따른 추가적인 분석과 모델 최적화가 필요한 것으로 생각된다.

결론적으로, CAPM 기반 분석과 기술적 분석을 결합한 통합 주가 예측 모델은 기존 개별 모델 대비 예측 정확도가 향상되었으며, 특히 주요 종목의 상승/하락 방향성 예측에서 우수한 성과를 보였다. RMSE 계산 결과 기술적 분석 모델로만 낸 결과대비 약 30% 정도 작은 RMSE 값이기에, 실제 주식 가격 또한 더 정확하게 예측하는 것으로 나타났다.

**모델 개선을 위한 추가적인 과제**

추가적인 데이터 소스와 다양한 경제 지표의 반영이 필요함. 특히 거시적인 경제지표를 활용하여 지수의 흐름을 예측하는 경우 모델의 예측성공률이 높아질 것으로 예상된다.

각 산업군에 맞게 세부적인 항목에 대한 가중치를 부여하여 모델을 보완이 필요하다.

머신러닝 기반의 비선형 모델을 도입하여 기술적 지표와 재무 지표 간의 상호작용을 정밀하게 분석이 필요하다.

**실제 투자 시 적용 가능성을 높이기 위한 실행 계획 마련**

모델의 실제 투자 적용 가능성을 위해 예측 결과를 기반으로 한 포트폴리오 전략을 수립하고, 리스크 관리 방안 또한 함께 고려해야 한다.

추가적으로, 산업별 특성에 맞춘 맞춤형 모델을 개발하여 다양한 투자 상황에서의 적용 가능성을 높이는 것이 필요하다.