

석사학위논문

거시경제변수와 뉴스 감성 분석을 활용한 LSTM 시계열  
모델 기반 업종별 주가 예측 성능 비교 연구

A Comparative Study on the Performance  
of LSTM Time Series Models for Stock Price Forecasting  
by Industry Using Macroeconomic Variables and  
News Sentiment Analysis

이 정 현

한양대학교 대학원

2025년 2월

석사학위논문

거시경제변수와 뉴스 감성 분석을 활용한 LSTM 시계열  
모델 기반 업종별 주가 예측 성능 비교 연구

A Comparative Study on the Performance  
of LSTM Time Series Models for Stock Price Forecasting  
by Industry Using Macroeconomic Variables and  
News Sentiment Analysis

지도교수 황 승 준

이 논문을 경영학 석사학위논문으로 제출합니다.

2 0 2 5 년 2 월

한양대학교 대학원

경영컨설팅학과

이 정 현

이 논문을 이정현의 석사학위 논문으로 인준함

2025년 2월

심사위원장 : 구재훈



심사위원 : 백승현



심사위원 : 황승준



한양대학교 대학원

## 목 차

목 차 .....	i
표 목 차 .....	iv
그림목차 .....	vi
국문요지 .....	viii

### 제 1장 서론

제 1절 연구의 배경 및 필요성 .....	1
제 2절 연구의 구성 .....	3

### 제 2장 선행연구

제 1절 주가 예측 모델 관련 연구 .....	4
제 2절 뉴스 감성 분석을 통한 주가 예측 연구 .....	6
제 3절 거시경제변수와 주가와의 상관관계 연구 .....	8

### 제 3장 이론적 배경

제 1절 LSTM(Long Short-Term Memory) 모델 .....	9
제 2절 감성 분석(Sentiment Analysis) .....	11

## 제 4장 연구방법

제 1절 실험모형 .....	12
제 2절 데이터 수집 및 전처리 .....	14
제 3절 LSTM 모델 학습 구성 .....	20
제 4절 성능지표 및 평가 .....	22

## 제 5장 실험결과 및 분석

### 제 1절 10개 업종별 주가 예측 성능 비교

1. 전체 설명 .....	23
2. 반도체 .....	25
3. 자동차 .....	25
4. 바이오 .....	26
5. 금융 .....	27
6. 방산 .....	28
7. 조선 .....	28
8. 석유·화학 .....	29
9. IT .....	30
10. 철강 .....	30
11. 건설 .....	31

## 제 2절 두 모델 간의 대응표본 t-검정 결과 분석 비교

1. 전체 설명 .....	32
2. 반도체 .....	33
3. 자동차 .....	33
4. 바이오 .....	34
5. 금융 .....	35
6. 방산 .....	35
7. 조선 .....	36
8. 석유·화학 .....	37
9. IT .....	37
10. 철강 .....	38
11. 건설 .....	39
제 3절 종합 실험 결과 분석 .....	40

## 제 6장 결론 및 시사점

제 1절 연구 결과 요약 및 시사점 .....	47
제 2절 연구의 한계 및 향후 연구 방향 .....	48

참고문헌 .....	50
------------	----

Abstract .....	53
----------------	----

## 표 목 차

[표 1] 업종 및 기업 종목 목록 .....	14
[표 2] 뉴스 기사 데이터 예시 .....	17
[표 3] LSTM 모델 매개변수 설정 .....	21
[표 4] 주가 데이터 + 거시경제변수 모델 성능 지표 값 .....	24
[표 5] 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 모델 성능 지표 값 .....	24
[표 6] 반도체 업종 대응표본 t-검정 결과 .....	33
[표 7] 자동차 업종 대응표본 t-검정 결과 .....	34
[표 8] 바이오 업종 대응표본 t-검정 결과 .....	34
[표 9] 금융 업종 대응표본 t-검정 결과 .....	35
[표 10] 방산 업종 대응표본 t-검정 결과 .....	36
[표 11] 조선 업종 대응표본 t-검정 결과 .....	36

[표 12] 석유·화학 업종 대응표본 t-검정 결과 .....	37
[표 13] IT 업종 대응표본 t-검정 결과 .....	38
[표 14] 철강 업종 대응표본 t-검정 결과 .....	38
[표 15] 건설 업종 대응표본 t-검정 결과 .....	39



## 그림 목 차

[그림 1] LSTM Model 구조 .....	9
[그림 2] 주가 데이터 + 거시경제변수 모델 .....	12
[그림 3] 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 모델 .....	13
[그림 4] 주가 데이터셋 예시 .....	15
[그림 5] 주가 + 거시경제변수 병합 데이터셋 예시 .....	16
[그림 6] VADER 감성 분석 점수 .....	19
[그림 7] 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 병합 데이터셋 예시 .....	19
[그림 8-1] 삼성바이오로직스 주가 예측 그래프 .....	41
[그림 8-2] 셀트리온 주가 예측 그래프 .....	42
[그림 8-3] 유한양행 주가 예측 그래프 .....	43

[그림 9-1] 네이버 주가 예측 그래프 .....	44
[그림 9-2] 카카오 주가 예측 그래프 .....	45
[그림 9-3] KT 주가 예측 그래프 .....	46

## 국 문 요 지

주가는 기업의 재무 상태나 경영전략 같은 내부 요인뿐만 아니라 거시경제변수나 산업 동향, 정치·사회적 변화 등 다양한 외부 요인의 영향을 받는다. 이러한 요인들로 인해 주가는 변동성을 가지며, 이는 정부와 기업 정책 등 경제 전반에 중대한 영향을 미친다. 이에 따라, 주가를 예측하기 위한 다양한 분석 모델을 활용한 연구가 지속적으로 이루어지고 있다. 이러한 주가 예측 모델에 대한 기존 연구에서는 머신러닝이나 딥러닝 등 인공지능망을 활용한 분석이 주로 이루어졌으며, 특히 거시경제변수와 뉴스 감성 분석을 결합한 주가 예측 연구가 진행되었다. 그러나 기존 연구는 주로 개별 기업이나 특정 산업군에 집중된 주가 예측에 중점을 두었으며, 업종별로 구분하여 외부 요인이 주가 예측에 미치는 영향을 분석한 연구는 드물다.

이에 따라 본 연구에서는 KOSPI 상장 기업을 대상으로 반도체, 자동차, 바이오, 금융 등 10개의 주요 업종에서 대표하는 주요 기업 3개로 선정하여 표본을 구성하고, 거시경제변수와 뉴스 감성 분석을 결합한 LSTM 시계열 모델을 활용해 두 가지 모델로 나누어서 주가 예측 성능을 비교하였다. 이를 통해 외부 요인이 주가 예측에 미치는 영향을 업종별로 분석함으로써 기존 연구의 한계를 보완하고자 하였다. 뉴스 감성 분석에는 VADER 감성분석을 사용하였으며, 주가 예측 모델로는 첫 번째 모델로 주가 데이터와 거시경제변수를 조합한 모델, 두 번째 모델은 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 모델 두 가지 모델로 나누어서 딥러닝 알고리즘 RNN의 한 종류인 LSTM을 적용하였다. 본 연구 결과를 통해, 뉴스 감성 분석이 모든 업종에서 일관되게 예측 성능을 향상시키지는 않았으며, 특히 거시경제변수가 경제 지표와 시장 동향을 반영하고 주가의 변동성을 설명하는 데 중요한 핵심 요인임을 확인하였다. 이러한 결과는 각 업종의 특성에 맞는 맞춤형 주가 예측 모델을 개발하는 데 필요하다는 점을 설명하고 있다.

주제어: 주가 예측, 거시경제변수, 뉴스 감성 분석, LSTM, 대응표본 t-검정

# 제 1장 서론

## 제 1절 연구의 배경 및 필요성

주가는 기업의 내부 요인뿐만 아니라 외부 요인, 즉 타 기업들의 영향, 국제 유가, 환율, 금리, 다른 국가의 주가지수 및 경제 상황에 상호 연관성을 가지며 이들의 변동에 민감하게 영향을 받는다. 이러한 외부 요인들이 상호작용하며 주가 변동에 큰 영향을 미치고, 이는 정부와 기업 정책 등 전반적으로 중대한 영향을 미치므로 다양한 분석 모델들을 이용하여 주가 예측이 지속적으로 이루어졌다(박강희, 신현정, 2011). 주식은 다양한 종속변수에 의해 움직이며, 많은 잡음과 비정상성, 비선형성으로 인해 가격변동이나 기대수익을 예측하는 것은 매우 어려운 작업으로 기존의 분석 방법론으로 주가를 예측하는 것은 매우 회의적이다(신동하, 최광호, 김창복, 2017). 그래서 이를 위해 주가 예측 연구를 수행하기 위해 산업계에서 학계까지 주목을 받고 있으며, 다양한 기계학습 알고리즘들이 활용되고 있다(K.Chen et al, 2015). 특히, 그 중에 LSTM(Long Short-Term Memory)과 같은 딥러닝 모델은 과거 데이터를 기반으로 미래 주가를 예측하는 데 뛰어난 성능을 보이며, 타 모델에 비해 성과가 좋으며 본질적으로 주가 예측에 적합하다고 평가하고 있다(T. Fischer & C. Krauss, 2018). 더불어, 거시경제 지표와 뉴스 감성 분석을 결합한 주가 예측 연구도 활발히 진행 중이다. 거시경제 지표는 경제 상황을 반영하고, 뉴스 감성 분석은 시장 참여자들의 감정 변화를 통해 예측에 중요한 추가 정보를 제공할 수 있다. 그래서 이 두 가지 변수를 결합하여 딥러닝 알고리즘을 통해 주가 예측을 수행하는 연구들이 진행되고 있다(성노운, 남기환, 2017). 기존 연구들로 거시경제변수와 뉴스 감성 분석을 결합한 주가 예측 연구가 진행되었지만 주로 개별 기업이나 특정 산업에 집중하여 주가 예측을 수행해 왔으며, 업종별로 거시경제변수와 뉴스 감성 분석을 결합해서 함께 외부 요인들

의 통합적 분석을 다루고 이에 대한 영향력 차이를 고려한 주가 예측 연구는 극히 드물었다.

이에 본 연구는 KOSPI 상장 기업들을 각 국내 산업 업종별로 구분하고, 거시경제변수와 뉴스 감성 분석을 결합한 주가 예측 모델을 제안하여 이를 실증적으로 평가하고자 한다. 기존 연구가 주로 개별 기업이나 특정 산업에 집중한 것과 달리, 본 연구는 외부 요인들을 통합 분석함으로써 업종별 주가 예측에서 나타나는 차이에 중점을 두고 있다. 이를 위해 성능 지표를 사용해 모델의 예측 정확성을 평가하고, 이 분석이 다양한 산업 업종에 일반화될 수 있는지를 검증할 것이다. 또한, 거시경제변수와 뉴스 감성 분석을 결합한 모델과 거시경제변수만을 사용하는 모델의 예측 성능 차이를 비교 분석함으로써 두 모델의 성능 차이가 특정 표본에만 국한되지 않고 전체 업종에 걸쳐 일반화될 수 있는지를 통계적으로 검토할 예정이다. 이를 통해 업종별 주가 예측에 큰 영향을 미치는 외부 요인을 파악하고자 하며, 이러한 통합적 접근은 기존 주가 예측 연구와의 차별성을 부각하는 동시에, 통계적 검증을 통해 새로운 연구 방향을 제시하는 데 기여할 것이다.

## 제 2절 연구의 구성

본 연구는 앞서 언급한 연구 배경과 필요성을 바탕으로, 기존 연구에서 다루지 않았던 업종별로 외부 요인들의 통합적인 분석을 수행하고, 이러한 요인이 주가 예측에 미치는 영향을 통계적으로 검증하는 것을 목표로 한다.

연구의 구성은 다음과 같다. 제2장에서는 주가 예측 모델, 뉴스 감성 분석, 거시경제변수와 주가의 관계에 대한 기존 연구들을 분석 검토하여, 본 연구의 타당성을 확보하고자 한다. 제3장에서는 LSTM(Long Short-Term Memory) 모델과 감성 분석의 이론적 배경을 설명함으로써 연구의 주요 방법론을 뒷받침하는 기초 이론을 확립한다.

제4장에서는 10개의 주요 업종을 선정하여 KOSPI 상장 기업의 주가 데이터를 수집하고, 각 업종에서 대표 기업 3개씩을 표본으로 삼아 LSTM 모델을 구성한 후 실험 설계, 데이터 수집 및 전처리 과정 등을 상세히 기술하였다. 제5장에서는 ‘주가 데이터 + 거시경제변수’ 모델과 ‘주가 데이터 + 거시경제변수 + 뉴스 감성 분석’ 모델의 성능을 비교하고, 대응표본 t-검정을 통해 두 모델 간 예측 성능 차이가 통계적으로 유의미하고 업종 전반에 걸쳐 일반화할 수 있는지를 검증하였다. 마지막으로 제6장에서는 연구의 주요 결과를 종합적으로 요약 및 정리하고, 연구의 한계와 향후 연구 방향을 논의하여 본 연구의 완성도를 높였다.

## 제 2장 선행연구

### 제 1절 주가 예측 모델 관련 연구

주가 예측 연구는 꾸준히 발전해 왔으며, 기존의 예측 방법으로는 ARIMA와 같은 시계열 분석 기법이 널리 사용되었다. 그러나 최근 인공지능(AI)과 딥러닝과 같은 방법론으로 주가를 예측하는 연구가 활발히 이루어지고 있다. 특히, LSTM(Long Short-Term Memory)과 같은 순환신경망이 시계열 데이터를 처리하는 데 우수한 성능을 보이면서 많은 연구에서 활용되고 있다.

본 절에서는 이러한 주가 예측 모델 관련해서 주요 선행 연구들을 살펴보고자 한다. 신동하, 최광호, 김창복(2017)은 RNN과 LSTM을 활용한 주가 예측 모델을 제안하면서, 전통적인 DNN(Deep Neural Network)보다 RNN과 LSTM이 시계열 데이터를 더 잘 처리해 주가 예측 성능을 높인다고 설명하였다.

주일택, 최승호(2018)는 시계열 데이터인 주가의 변동 패턴을 학습하고 주가를 예측하기 위한 딥러닝 모델로 양방향 LSTM 순환 신경망을 이용하여 주가 예측 모델을 구현하여 낮은 예측 오차를 발생시키면서 주가 예측 정확성을 향상시켰다.

한태동(2021)은 기술적 지표, 거시경제 지표, 시장 심리 지표 등을 모두 포함하는 통합적 주가 예측 모델로 LSTM을 사용하였다. 이 연구에서는 모든 종목에 적용할 수 있는 범용 모델을 찾을 수 없지만, 업종이나 가격대에 따라 예측력을 높이는 조합이 존재하며, 현재 시점에서 먼 과거의 데이터보다 최근 데이터를 사용하는 것이 주가 예측에 더욱 유용하다는 결론을 도출하였다.

박강희, 신현정(2011)은 그래프 기반 준지도 학습 알고리즘을 사용하여 KOSPI에 상장된 개별 회사의 주가와 외부 경제 지표 간의 상호 관계를 고려한 시계열 예측 모델을 통해, 기존의 ANN(인공신경망)과 SVM(서포트 벡터 머신) 모델에 비해 더 높은 성능을 보였다.



T. Fischer & C. Krauss(2018)는 S&P 500 지수를 대상으로 LSTM을 사용한 주가 예측 모델을 연구하였고, 이 모델이 랜덤 포레스트나 로지스틱 회귀 등 다른 머신러닝 기법보다 높은 예측 정확도와 수익률을 보인다고 입증하였다. 이와 같은 연구들은 LSTM과 같은 시계열 모델이 기존의 전통적 방법론보다 주가 예측에서 더 나은 성능을 발휘할 수 있음을 시사하며, 주가 예측 모델 연구에서 큰 비중을 차지하고 있다.

## 제 2절 뉴스 감성 분석을 통한 주가 예측 연구

최근 주가 예측 연구에서는 뉴스와 소셜 미디어와 같은 비정형 데이터를 분석하는 방법이 활발히 활용되고 있다. 특히, 감성 분석을 통해 대중의 정서가 시장에 미치는 영향을 파악하고 이를 주가 예측에 반영하려는 시도들이 주목을 받고 있다. 본 절에서는 뉴스 감성 분석을 활용한 주가 예측에 관련된 주요 선행 연구들을 살펴보고자 한다.

J. Bollen et al. (2011)은 Twitter 데이터를 활용해 대중의 감정이 주가에 미치는 영향을 분석하였다. 행동 경제학적 관점에서 긍정적 및 부정적 감정이 주식 시장 방향성에 중요한 역할을 한다고 주장했다.

성노윤, 남기환(2017)은 온라인 뉴스 데이터와 거시경제 변수를 결합한 주가 예측 모델을 제시하였으며, 뉴욕타임스 뉴스 데이터를 Word2Vec으로 분석하고, 금값과 환율 등의 거시경제지표를 LSTM 모델과 결합하여 주가 변동성을 예측하였다. 그 결과, 두 데이터를 함께 사용할 때 단독으로 사용할 때보다 예측 성과가 향상됨을 입증하였다.

장은아, 최희련, 이홍철(2020)은 BERT와 NLTK VADER를 활용한 뉴스 감성 분석과 거시경제지표를 결합하여 다우존스 지수 예측 모델을 제안하였다. 뉴욕타임스 뉴스 데이터를 BERT와 NLTK로 분석하고, 금, 유가, 환율 등의 지표와 결합하여 효과적인 지표 조합을 찾고자 LSTM 모델로 주가 변동성을 예측한 결과, EUR/USD 환율, NLTK VADER, BERT가 결합된 모델이 가장 높은 예측 성능을 보였다.

김유신, 김남규, 정승렬(2012)은 오피니언 마이닝 기법으로 뉴스의 긍정/부정 감성을 분석해 주가지수 변동성을 예측하는 지능형 투자 의사결정 모형을 제안했으며, 감성 분석이 주가 변동에 유의미한 영향을 미친다는 것을 확인하였다.

김동영, 박제원, 최재현(2014)은 뉴스와 트위터 데이터를 결합하여 감성 분석 및 기계학습 기법을 활용한 주가 예측 모델을 개발하였고, 로지스틱 회귀와

서포트 벡터 머신(SVM)을 통해 80% 이상의 예측 정확도를 달성하였다.

이와 같은 연구들은 뉴스 및 소셜 미디어 감성 분석이 주가 예측 모델에 유용하며, 비정형 데이터가 주가 예측 연구에 효과적으로 활용될 수 있음을 보여준다. 앞으로도 이러한 접근법은 주가 예측 연구에서 더욱 중요한 역할을 할 것으로 기대된다.

### 제 3절 거시경제변수가 주가에 미치는 영향 관련 연구

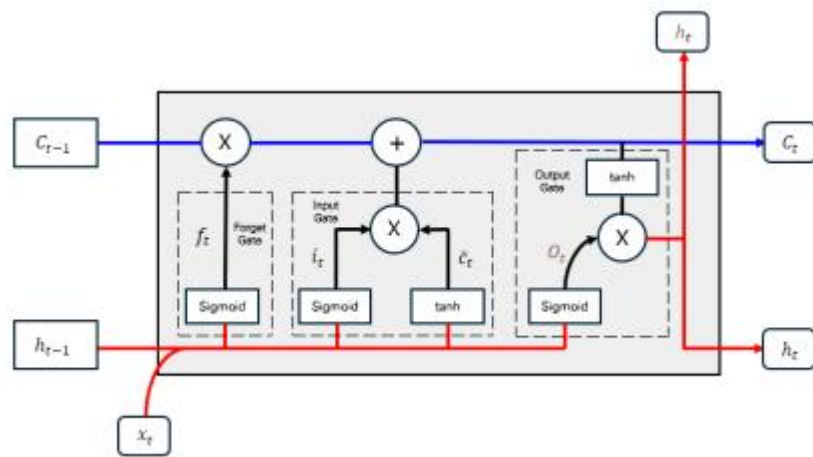
거시경제변수는 금융 시장과 주가에 영향을 미치는 중요한 외부 요인 중 하나로, 많은 연구에서 주가와와의 관계를 분석해 왔다. 금리, 환율, 유가 등의 거시경제변수는 국내외 경제 상황을 반영하는 지표로, 주식 시장의 변동성에 직접적인 영향을 미칠 수 있다. 이들 변수는 경기 사이클에 따라 변화하여 주가 예측의 주요 요소로 작용한다. 본 절에서는 거시경제변수가 주가에 미치는 영향을 분석한 주요 선행 연구들을 살펴보고자 한다. 감형규, 신용재(2017)는 회귀분석과 VAR 모델을 사용하여 금리, 산업생산지수증가율, 경기선행종합지수증가율, 월/미국달러환율증가율 등의 거시경제변수가 코스피 지수에 유의미한 영향을 미친다는 결과를 도출하였다. 이를 통해 거시경제변수들이 주가 예측에 중요한 역할을 한다는 것을 입증하였다. 송영렬, 김흥기, 한동협(2009)은 다요인 모형을 기반으로 금리, 환율, 산업생산지수, 물가, 유가 등과 같은 여러 거시경제변수가 코스피와 코스닥 시장 수익률에 미치는 영향을 비교 분석하였으며, 여러 거시경제변수들이 주가에 미치는 유의미한 영향을 실증적으로 밝혔다. 윤일현(2016)의 연구에서는 금리, 물가상승률, 외환보유액, 실업률, 고용지수, 환율, 국제유가와 같은 거시경제변수가 코스피 지수에 미치는 영향을 분석하였으며, GARCH 모형을 사용하여 환율과 국제유가가 한국 주식 시장 수익률에 유의미한 영향을 미친다는 결론을 도출하였다. 이는 환율과 유가가 한국 주식 시장 변동성에 대해 설명할 수 있는 중요한 요소임을 보여주고 있다. K. S. Sujit & B. Rajesh Kumar(2011)은 금값, 유가, 환율과 주식 시장 간의 상호작용을 VAR 및 공적분 분석을 통해 환율이 다른 변수들에 큰 영향을 받으며, 금값과 유가의 변동이 서로 밀접한 상관관계를 가진다는 것을 확인하였다. 이와 같은 연구들은 거시경제변수가 주가에 미치는 영향을 심층적으로 분석하고, 주가 예측에 있어 핵심적인 요소임을 보여주고 있다.

## 제 3장 이론적 배경

### 제 1절 LSTM(Long Short-Term Memory) 모델

LSTM(Long Short-Term Memory) 모델은 Hochreiter와 Schmidhuber(1997)가 RNN의 한계를 극복하기 위해 제안한 모델로, 주로 기울기 소실 문제를 해결하여 장기 의존성 문제를 처리하는 데 적합하다. RNN은 순차적인 데이터 처리에 유리하지만, 장기적인 정보가 기울기 소실 현상으로 인해 잊혀지는 한계를 가진다. 이를 해결하기 위해 LSTM은 정보의 흐름을 조절할 수 있는 게이트 구조와 메모리 셀을 도입하였다.

LSTM은 망각 게이트(Forget Gate), 입력 게이트(input Gate), 출력 게이트(Output Gate)의 세 가지 주요 게이트로 구성되어 있다.



<그림 1> LSTM Model 구조

첫째, 망각 게이트  $f_t$  는 현재 시점의 입력값  $x_t$  와 전 시점의 은닉값  $h_{t-1}$  에 각각 가중치를 곱한 후, 시그모이드 함수를 통해 0에서 1 사이의 값을 출력한다. 이 값은 이전 셀 상태  $C_{t-1}$  에 곱해져 얼마나 이전 정보를 유지할지를 결정하며, 출력값이 1에 가까울수록 이전 정보가 많이 유지하고, 0에 가까울수록 더 많은 정보가 잊혀진다. 둘째, 입력 게이트  $i_t$  는 새로운 입력값  $x_t$  의 정보를 셀 상태에 얼마나 반영할지를 결정한다. 이 값 또한 시그모이드 함수를 통해 계산되며, 새로운 정보를 셀 상태에 반영되는 정도를 조정한다. 이와 함께 새로운 셀 상태 후보 값  $\tilde{C}_t$  는 tanh 함수를 통해 계산되며, 이는 입력 게이트의 출력값과 결합되어 현재 셀 상태에 반영된다. 셋째, 출력 게이트  $O_t$  는 셀 상태에서 출력할 정보를 결정한다. 시그모이드 함수로 셀 상태에서 선택된 정보를 조정한 후, 현재 셀 상태  $C_t$  는 tanh 함수를 통해 -1에서 1 사이의 값으로 변환되며, 이 값이 출력 게이트에서 계산된 값  $O_t$  와 곱해져 현재 시점의 은닉 상태  $h_t$  로 출력된다. 이 은닉 상태  $h_t$  는 LSTM 셀의 최종 출력값이 된다.

이처럼 LSTM은 망각 게이트와 입력 게이트를 통해 이전 정보를 잊거나 새 정보를 받아들이도록 조정함으로써, 장기적인 정보 흐름을 유지하는 동시에 불필요한 정보를 제거한다. 이러한 구조는 RNN에서 발생하는 기울기 소실 문제를 해결하여 긴 시퀀스 데이터도 정보 손실 없이 안정적으로 학습할 수 있도록 한다.

## 제 2절 감성 분석(Sentiment Analysis)

감성 분석(Sentiment Analysis)은 자연어 처리 기법을 활용해 텍스트 데이터를 긍정, 중립, 부정 감정으로 자동 분류하고 분석하는 방법이다. 주로 뉴스 기사나 소셜 미디어, 온라인 리뷰 등과 같은 비정형 텍스트의 감정을 파악하는 데 유용하게 활용되고 있다. 감성 분석 방법론으로 크게 어휘(Lexicon-based) 기반 감성 분석과 기계 학습(Machine Learning-based) 기반 감성 분석으로 나누어져 있으며, 본 연구에서는 어휘(Lexicon-based) 기반 감성 분석을 활용하였다. 어휘(Lexicon-based) 기반 감성 분석은 명사, 형용사, 동사를 대상으로 모든 단어에 감성을 부여한 감성 사전을 구축한 후에, 이를 기반으로 텍스트 내 단어들의 감성 분석을 수행하는 방식이다. 감성 사전은 직접 구축하거나 외부의 감성 사전을 활용할 수 있으며, 본 연구에서는 VADER(Valence Aware Dictionary and sEntiment Reasoner)를 사용해 뉴스 데이터의 긍정적, 부정적, 중립적 감정을 분석하였다. VADER는 소셜 미디어 텍스트 분석에 최적화된 패키지로, 대규모 텍스트 데이터도 빠르게 분석할 수 있다는 장점이 있다.

VADER의 주요 기능 중 하나인 SentimentIntensityAnalyzer 클래스는 텍스트 내 감정의 강도를 네 가지 극성값으로 제공한다. ‘pos’는 긍정, ‘neg’는 부정, ‘neu’는 중립, 그리고 compound는 이 세 값을 결합한 전체 감정 점수로 -1에서 1 사이의 값을 가진다. 이 compound 점수에 따라 감정이 긍정(0.05 이상), 중립(-0.05에서 0.05 사이), 부정(-0.05 이하)으로 평가된다.

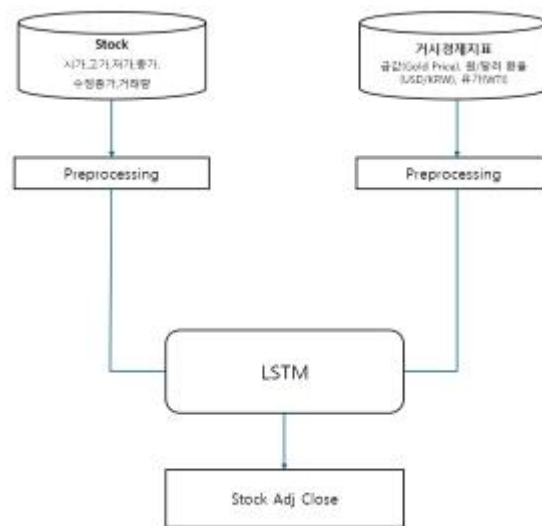
이는 대용량 텍스트 데이터를 효과적으로 분석할 수 있으며, 특히 소셜 미디어나 뉴스 기사처럼 다량의 데이터에서 감정을 추출할 때 유용하게 사용된다.

## 제 4장 연구방법

### 제 1절 실험모형

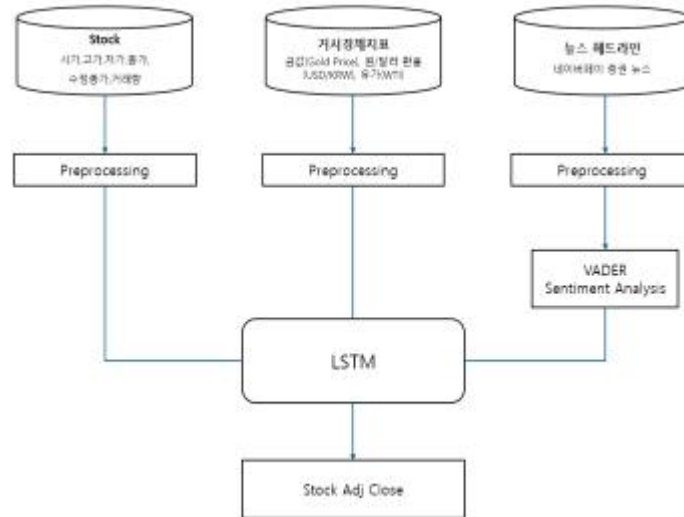
본 연구에서는 LSTM(Long Short-Term Memory) 시계열 모델을 기반으로 하여, 10개의 업종(반도체, 자동차, 바이오, 금융, 방산, 조선, 석유·화학, IT, 철강, 건설)에서 각각 3개의 대표 기업 종목을 선정하고, 코스피(KOSPI)시장에 상장된 기업들의 주가 예측 성능을 비교하였다. 성능 차이 비교를 위해 주가 데이터와 거시경제변수를 조합한 모델과 주가 데이터와 거시경제변수, 그리고 뉴스 감성 분석을 조합한 모델 두 가지 모형을 실험하였다.

<그림 2>에서 보면 첫번째 모형은 주가 데이터와 거시경제변수를 조합한 모델로, 주가에 영향을 미치는 주요 거시경제변수들인 금값(Gold Price), 원/달러 환율(USD/KRW), 유가(WTI Crude Oil Price)를 독립변수로 사용하여 예측을 수행하였다.



<그림 2> 주가 데이터 + 거시경제변수 모델





<그림 3> 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 모델

<그림 3>에서 보면 두번째 모형은 거시경제변수에 더해 코스피(KOSPI)를 키워드로 한 뉴스 감성 분석 결과를 추가한 모형이다. 이 모형은 뉴스 헤드라인 데이터를 수집하고 VADER 감성 분석 도구를 사용하여 각 뉴스의 감정(긍정적, 중립적, 부정적)을 분석한 뒤, 이를 종합한 compound 점수로 나타내었다. 본 연구에서 사용한 두 가지 모형은 동일한 시계열 데이터를 기반으로 학습하였고 두 모형 간의 성능 차이를 분석하기 위해 대응표본 t-검정을 수행하였다. 특히, 각 업종별로 대표 표본 3개를 추출하여 분석을 진행하였기 때문에, 이러한 표본이 전체 업종을 일반화하는 데 적절한지 검증하고자 대응표본 t-검정을 적용하였다. 이를 통해 뉴스 감성 분석의 도입이 업종별 주가 예측 성능에 미치는 영향을 통계적으로 검토하였다.

## 제 2절 데이터 수집 및 전처리

본 연구는 2019년 7월 1일부터 2024년 6월 28일까지의 데이터를 수집하여 실험에 활용하였다. 각 데이터는 코스피(KOSPI)시장에 상장된 기업들의 주가 데이터와 거시경제변수 및 뉴스 감성 분석 데이터로 구성된다. 본 연구에서는 10개의 주요 업종에서 각 업종별로 3개의 대표 기업 종목을 선정하였다. 선정된 업종과 대표 기업 종목들은 <표 1>과 같다.

업종	기업	업종	기업
반도체	삼성전자	조선	HD한국조선해양
	SK하이닉스		한화오션
	DB하이텍		삼성중공업
자동차	현대차	석유·화학	SK이노베이션
	기아		LG화학
	KG모빌리티		한화솔루션
바이오	삼성바이오로직스	IT	네이버
	셀트리온		카카오
	유한양행		KT
금융	KB금융지주	철강	포스코홀딩스
	신한금융지주		현대제철
	하나금융지주		KG스틸
방산	한화에어로스페이스	건설	현대건설
	한국항공우주산업		대우건설
	LIG넥스원		GS건설

<표 1> 업종 및 기업 종목 목록

주가 데이터는 각 업종의 3개 대표 기업종목의 시가(Open Price), 고가(High Price), 저가(Low Price), 종가(Close Price), 수정종가(Adjusted Close Price), 거래량(Volume) 데이터를 Yahoo Finance에서 제공하는 ‘yfinance’ 라이브러리를 활용해서 수집하였으며 아래의 <그림 4>에서 데이터셋의 예시를 볼 수 있다.

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-07-01	47350.0	47400.0	46250.0	46600.0	40888.687500	11383522
2019-07-02	46200.0	46900.0	45850.0	46250.0	40581.570312	8463073
2019-07-03	45750.0	46350.0	45200.0	45400.0	39835.757812	9669368
2019-07-04	45250.0	46200.0	45250.0	46000.0	40362.222656	6365573
2019-07-05	45950.0	45950.0	45250.0	45650.0	40055.109375	7235395
...	...	...	...	...	...	...
2024-06-24	79700.0	80900.0	79500.0	80600.0	80242.109375	15454227
2024-06-25	80600.0	81800.0	80100.0	80800.0	80441.218750	19088458
2024-06-26	80100.0	81400.0	79900.0	81300.0	80939.000000	17783242
2024-06-27	81300.0	81600.0	80500.0	81600.0	81600.000000	11739720
2024-06-28	81900.0	81900.0	80800.0	81500.0	81500.000000	9455929

<그림 4> 주가 데이터셋 예시

주가예측에 영향을 미칠 수 있는 주요 거시경제변수 데이터로는 금값(Gold Price), 원/달러 환율(USD/KRW),유가(WTI)데이터를 선정하여 Yahoo Finance에서 제공하는 ‘yfinance’ 라이브러리를 활용해서 수집하였다. 이 데이터들은 주가 데이터와 함께 분석에 포함하였고, 동일한 날짜를 기준으로 각 데이터셋을 병합하는 방식으로 통합하였다. 병합 과정은 Python 프로그래밍을 활용하여 날짜별로 데이터를 정렬한 후, 동일한 날짜의 데이터만을 기준으로 결측값을 제거하고, 데이터의 일관성을 유지하면서 하나의 분석용 데이터셋으로 결합하였다. <그림 5>에서 주가 데이터와 거시경제변수를 병합한 데이터셋의

예시를 볼 수 있다.

Date	Open	High	Low	Close	Adj Close	Volume	Gold	USD/KRW	OIL
2019-07-01	47350.0	47400.0	46250.0	46600.0	40888.687500	11383522	1385.599976	1155.599976	59.090000
2019-07-02	46200.0	46900.0	45850.0	46250.0	40581.582031	8463073	1404.599976	1159.750000	56.250000
2019-07-03	45750.0	46350.0	45200.0	45400.0	39835.753906	9669368	1417.699951	1165.579956	57.340000
2019-07-05	45950.0	45950.0	45250.0	45650.0	40055.121094	7235395	1396.699951	1167.130005	57.509998
2019-07-08	44750.0	44800.0	44350.0	44400.0	38958.316406	7823843	1397.000000	1174.020020	57.660000
...	...	...	...	...	...	...	...	...	...
2024-06-24	79700.0	80900.0	79500.0	80600.0	80242.109375	15454227	2330.000000	1388.390015	81.629997
2024-06-25	80600.0	81800.0	80100.0	80800.0	80441.218750	19088458	2316.600098	1386.109985	80.830002
2024-06-26	80100.0	81400.0	79900.0	81300.0	80939.000000	17783242	2299.199951	1390.890015	80.900002
2024-06-27	81300.0	81600.0	80500.0	81600.0	81600.000000	11739720	2324.500000	1393.380005	81.739998
2024-06-28	81900.0	81900.0	80800.0	81500.0	81500.000000	9455929	2327.699951	1386.949951	81.540001

<그림 5> 주가 + 거시경제변수 병합 데이터셋 예시

뉴스 감성 분석에 사용할 뉴스 기사 데이터는 네이버페이 증권 뉴스에서 ‘코스피’ 검색어에 대한 뉴스 기사 헤드라인 9,183개를 크롤링해서 수집하였다. 본 연구의 분석 대상 기업이 KOSPI 상장 기업임에 따라, 감성 분석을 위한 데이터 수집 시 적합한 키워드로 ‘코스피’로 선정하였다. 이를 통해 KOSPI 시장과 관련된 감성 변화를 효과적으로 반영하고자 하였다. 이 때 뉴스 데이터는 주식 시장의 장 마감 시간인 당일 16시부터 다음날 15시 59분까지 수집된 뉴스 기사들이 다음날의 주가에 영향을 미친다고 판단하여, 해당 시간대의 데이터를 기본으로 날짜를 조정하여 바뀌주었다. 수집한 뉴스 데이터는 한글로 작성되었기 때문에, Googletrans 라이브러리를 사용하여 뉴스 기사를 영어로 번역하였다. 감성 분석에 사용된 VADER(Valence Aware Dictionary and sEntiment Reasoner)와 같은 감성 분석 도구는 영어 텍스트를 기반으로 최적화되어 있으므로, 번역 과정을 통해 분석의 정확성을 높이하고자 했다. <표 2>에서 이에 대한 뉴스 기사 데이터 예시를 볼 수 있다.

날짜	제목	영문 번역 제목
2019-07-03	하반기 무역분쟁 타결 유력... 코스피 2350선까지 오른다	KOSPI rises to the likely level of trade dispute settlement in the second half of the year
2019-07-03	증권업계 미·중 무역분쟁 전망 엇갈려 “연내 타결 가능성 커” VS “연내 완전...”	Securities industry: The outlook for the US-China trade dispute is mixed; there is a high possibility that it will be resolved within the year.
2019-07-03	외화채에 몰린 개미... 올해 41.9조 사들었다	Ants flocking to foreign currency bonds were surveyed and purchased this year.
2019-07-05	'왕따' 韓증시... 주식형펀드 수익률도 꼴찌	Outcast stock market fund returns ranked last
2019-07-05	환율도, 기업실적도 불안... 올들어 신용거래 확 줄었다	The exchange rate and corporate performance are unstable, and credit transactions have decreased significantly this year.
2019-07-05	빚내 주식 산 개미 느는데... 손실 눈덩이	Buying stocks with debt, the loss is snowballing.

<표 2> 뉴스 기사 데이터 예시

VADER 감성 분석 도구를 활용해 긍정/중립/부정을 조합한 점수(compound score)를 산출하였다. 또한, 한 날짜에 수집된 여러 기사들의 감성 점수(compound score)를 평균하여 해당 날짜의 대표 감성 점수로 사용하였다. 이렇게 산출된 해당 날짜의 대표 감성 점수를 주가 예측 모델의 독립 변수로 사용하였으며, 감성 점수에 대한 예시는 <그림 6>에서 볼 수 있다. 수집한 주가 데이터, 거시경제변수, 뉴스 감성 분석한 데이터를 하나의 데이터셋으로 병합하였다. 날짜를 기준으로 각 데이터를 정렬하여 병합하였으며, 이를 통해 수정종가, 거시경제변수, 뉴스 감성 점수가 동일한 시점의 데이터로 결합되었고 이렇게 병합된 데이터는 모델 학습 및 테스트에 활용되었다. 수집된 데이터는 결측값을 제거한 후, MinMaxScaler를 사용하여 0과 1사이의 값으로 정규화하였다. 병합된 데이터셋의 예시는 <그림 7>에서 볼 수 있다.

compound_score	
Date	
2019-06-29	-0.00770
2019-06-30	-0.13816
2019-07-01	0.23546
2019-07-02	-0.01418
2019-07-03	0.17574
...	...
2024-06-25	0.00972
2024-06-26	0.25960
2024-06-27	0.33934
2024-06-28	0.19330
2024-06-29	0.23494

1828 rows × 1 columns

<그림 6> VADER 감성 분석 점수

Date	Open	High	Low	Close	Adj Close	Volume	Gold	USD/KRW	Oil	compound_score
2019-07-01	47350.0	47400.0	46250.0	46600.0	40888.691406	11383522	1385.599976	1155.599976	59.090000	0.23546
2019-07-02	46200.0	46900.0	45850.0	46250.0	40581.582031	8463073	1404.599976	1159.750000	56.250000	-0.01418
2019-07-03	45750.0	46350.0	45200.0	45400.0	39835.750000	9669368	1417.699951	1165.579956	57.340000	0.17574
2019-07-05	45950.0	45950.0	45250.0	45650.0	40055.117188	7235395	1396.699951	1167.130005	57.509998	-0.04838
2019-07-08	44750.0	44800.0	44350.0	44400.0	38958.316406	7823843	1397.000000	1174.020020	57.660000	0.23902
...	...	...	...	...	...	...	...	...	...	...
2024-06-24	79700.0	80900.0	79500.0	80600.0	80242.109375	15454227	2330.000000	1388.390015	81.629997	0.43698
2024-06-25	80600.0	81800.0	80100.0	80800.0	80441.218750	19088458	2316.600098	1386.109985	80.830002	0.00972
2024-06-26	80100.0	81400.0	79900.0	81300.0	80939.000000	17783242	2299.199951	1390.890015	80.900002	0.25960
2024-06-27	81300.0	81600.0	80500.0	81600.0	81600.000000	11739720	2324.500000	1393.380005	81.739998	0.33934
2024-06-28	81900.0	81900.0	80800.0	81500.0	81500.000000	9455929	2327.699951	1386.949951	81.540001	0.19330

<그림 7> 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 병합 데이터셋 예시

### 제 3절 LSTM 모델 학습 구성

본 연구에서 사용한 LSTM 모델의 매개변수 설정과 학습 과정은 다음과 같다. 최적화 알고리즘으로는 Adam 옵티마이저를 사용하였으며, 이는 경사 하강법 기반으로 학습 속도를 조정하여 효율적인 학습을 지원한다. 활성화 함수는 LSTM 레이어의 기본 활성화 함수인 Tanh를 적용하여 입력값을 -1에서 1 사이로 변환함으로써 비선형성을 부여하고 복잡한 데이터 패턴을 학습할 수 있도록 했다. 손실 함수는 예측 값과 실제 값 간의 차이를 최소화하기 위해 평균 제곱오차(Mean Squared Error, MSE)를 사용하였으며, 오차가 클수록 큰 페널티를 부여하는 방식으로 학습을 진행하였다. Epoch 수는 총 60으로 설정하였고, 이는 데이터셋을 60회 반복 학습하는 것을 의미한다. 과적합 방지를 위해 Early Stopping 기법을 적용하여, 검증 손실이 10 Epoch 동안 개선되지 않으면 학습이 자동으로 중단되도록 설정하였다. Early Stopping은 과적합을 방지하고 모델 성능을 유지하는 데 기여한다. 배치 크기(Batch Size)는 32로 설정되어, 모델이 각 Epoch에서 한 번에 학습하는 데이터 샘플 수를 의미하며, 이는 메모리 효율과 학습 속도에 중요한 영향을 미친다.

추가 예측 모델의 정확성을 평가하기 위해 전체 데이터를 학습용(train) 데이터와 테스트용(test) 데이터로 분리하여 분석을 수행하였다. 학습용 데이터는 모델이 패턴을 학습하는 데 사용되며, 테스트용 데이터는 학습되지 않은 데이터에서 예측 성능을 검증하기 위해 활용된다. 데이터는 학습용 80%, 테스트용 20%의 비율로 분할되었으며, 학습용 데이터는 2019년 7월 1일부터 2023년 6월 30일까지, 테스트용 데이터는 2023년 7월 1일부터 2024년 6월 28일까지의 기간에 해당하는 데이터를 사용하였다.

본 연구에서 사용한 LSTM 모델의 매개변수 설정은 아래 <표 3>에 정리되어 있다.



매개변수	설정 값
Optimizer	Adam
Activation Function	Tanh
Loss Function	Mean Squared Error
Epoch	60
Batch Size	32

<표 3> LSTM 모델 매개변수 설정

## 제 4절 성능지표 및 평가

본 연구에서는 주가 데이터와 거시경제변수를 조합한 모델과 주가 데이터, 거시경제변수, 뉴스 감성 분석을 결합한 모델의 예측 성능을 비교하기 위해 성능 평가 지표로 MAE(Mean Absolute Error), MSE(Mean Squared Error), RMSE(Root Mean Squared Error)를 사용하였다.

MAE는 예측 값과 실제 값 간의 절대 차이 평균을 구하여 오차의 크기를 직관적으로 측정하는 값이다. MSE는 예측 값과 실제 값 간의 차이를 제곱한 후 평균을 구하는 값으로, 큰 오차에 더 큰 패널티를 부여하는 특징을 가지고 있다. RMSE는 MSE의 제곱근을 취한 값으로, 이상치에 민감하게 반응하여 오차 크기를 평가하는 지표이다. 이 세 가지 성능 지표를 통해 두 모델 간의 예측 성능을 비교하였다.

두 모델 간의 예측 성능 차이의 통계적 유의성을 검증하기 위해 대응표본 t-검정을 수행하였다. 대응표본 t-검정은 동일한 데이터셋에서 두 모델의 성능 지표 평균 차이를 비교하는 방법으로, 본 연구에서는 유의수준 0.1을 기준으로 통계적 유의미성을 판단하였다. 각 업종별로 전체 기업이 아닌 3개 대표 기업을 표본으로 선정하였기 때문에, 이 표본이 업종의 일반적 특성을 대변할 수 있는지 확인하고자 대응표본 t-검정을 활용하였다. 업종별로 두 모델의 성능 차이를 평가한 후, 그 차이가 유의미한지를 분석하였다.

## 제 5장 실험결과 및 분석

### 제 1절 10개 업종별 주가 예측 성능 비교

#### 1. 전체 설명

본 연구에서는 주가 데이터와 거시경제변수를 조합한 LSTM 모델과 주가 데이터, 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델을 비교 분석하였다. 편의상, 전자를 거시경제변수 모델, 후자를 뉴스 감성 분석 모델로 명명하였다. 반도체, 자동차, 바이오, 금융, 방산, 조선, 석유·화학, IT, 철강, 건설 등 10개 주요 업종에서 대표 기업 3개씩을 표본으로 선정하여 예측 성능을 평가하였으며, 정규화된 MAE(Mean Absolute Error), MSE(Mean Squared Error), RMSE(Root Mean Squared Error) 지표를 통해 두 모델 간 예측 성능 차이를 정량적으로 분석하였다.

<표 4>와 <표 5>에는 각 모델의 예측 성능 지표 값이 제시되어 있으며, 이를 기반으로 업종별 모델 예측 성능의 차이를 평가하고자 한다. 이후 뉴스 감성 분석이 포함된 모델이 주가 예측에 미치는 영향을 분석하고, 업종별 특성에 따른 모델 성능의 차이를 논의할 것이다.

업종	종목	MAE	MSE	RMSE
반도체	삼성전자	0.02592	0.00115	0.03392
	SK하이닉스	0.0467	0.0042	0.06482
	DG하이텍	0.02173	0.00095	0.03086
자동차	현대차	0.02303	0.00101	0.03191
	기아	0.03296	0.00185	0.04304
	KG모빌리티	0.01486	0.00036	0.01899
바이오	삼성바이오로직스	0.01898	0.00059	0.02423
	셀트리온	0.01863	0.00066	0.02675
	유한양행	0.04824	0.00479	0.06946
금융	KB금융지주	0.0287	0.00162	0.04022
	신한금융지주	0.039	0.00311	0.0558
	하나금융지주	0.02137	0.00069	0.02987
방산	한화메트로스페이스	0.03264	0.00159	0.04106
	한국항공우주산업	0.02189	0.00063	0.02874
	LIG넥스원	0.04884	0.00527	0.0726
조선	HD한국조선해양	0.02733	0.00118	0.03442
	한화오션	0.02789	0.00143	0.03777
	삼성중공업	0.03917	0.00245	0.04947
석유·화학	SK이노베이션	0.0191	0.00068	0.02601
	LG화학	0.02476	0.00081	0.03017
	한화솔루션	0.02249	0.00086	0.02937
IT	네이버	0.01626	0.00042	0.02061
	카카오	0.00941	0.00015	0.01208
	KT	0.02633	0.00118	0.03437
철강	포스코홀딩스	0.04381	0.00453	0.06728
	현대제철	0.02042	0.00061	0.02472
	KG스틸	0.00426	0.00004	0.00677
건설	현대건설	0.01642	0.00048	0.02129
	대우건설	0.01531	0.00037	0.0193
	GS건설	0.02882	0.00111	0.03334

<표 4> 주가 데이터 + 거시경제변수 모델 성능 지표 값

업종	종목	MAE	MSE	RMSE
반도체	삼성전자	0.02546	0.00106	0.03251
	SK하이닉스	0.03514	0.00229	0.04789
	DG하이텍	0.02064	0.00092	0.03026
자동차	현대차	0.02563	0.00127	0.0357
	기아	0.03761	0.00245	0.04958
	KG모빌리티	0.01389	0.00034	0.01831
바이오	삼성바이오로직스	0.03473	0.00096	0.031
	셀트리온	0.02025	0.00086	0.02939
	유한양행	0.05601	0.00555	0.07448
금융	KB금융지주	0.03676	0.00261	0.05108
	신한금융지주	0.03599	0.00257	0.05068
	하나금융지주	0.02505	0.00117	0.03416
방산	한화메트로스페이스	0.03352	0.00188	0.04337
	한국항공우주산업	0.02025	0.00076	0.02751
	LIG넥스원	0.05391	0.00623	0.07893
조선	HD한국조선해양	0.03119	0.00159	0.03985
	한화오션	0.02733	0.0014	0.03738
	삼성중공업	0.05311	0.00419	0.0647
석유·화학	SK이노베이션	0.01861	0.00071	0.02659
	LG화학	0.02301	0.00086	0.02934
	한화솔루션	0.02309	0.00088	0.02968
IT	네이버	0.01707	0.00048	0.02145
	카카오	0.01103	0.0002	0.01412
	KT	0.02716	0.00129	0.03584
철강	포스코홀딩스	0.04306	0.00473	0.06874
	현대제철	0.01865	0.00054	0.0232
	KG스틸	0.00514	0.00006	0.00753
건설	현대건설	0.01636	0.00049	0.02182
	대우건설	0.01466	0.00034	0.01857
	GS건설	0.0218	0.00091	0.03013

<표 5> 주가 데이터 + 거시경제변수 + 뉴스 감성 분석 모델 성능 지표 값

## 2. 반도체

반도체 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교 분석하였다.

실험 대상 기업으로는 삼성전자, SK하이닉스, DB하이텍을 선정하였다.

삼성전자의 경우, 뉴스 감성 분석 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 거시경제변수 모델보다 낮은 값을 나타내어, 뉴스 감성 분석이 예측 성능을 약간 향상시켰다. SK하이닉스와 DB하이텍의 경우에서도 뉴스 감성 분석 모델이 거시경제변수 모델 대비 모든 지표에서 더 낮은 오차를 보이며, 삼성전자와 유사한 경향을 보였다.

반도체 업종의 전반적으로는 뉴스 감성 분석 모델이 거시경제변수 모델보다 일관되게 더 나은 예측 성능을 보였으며 이는 반도체 업종에서 뉴스 감성 분석이 주가 예측에 유용한 정보로 작용할 가능성을 나타내고 있음을 보여주고 있다.

## 3. 자동차

자동차 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교 분석하였다. 실험 대상 기업으로는 현대차, 기아, KG모빌리티를 선정하였다.

현대차와 기아의 경우, 모든 성능 지표(MAE, MSE, RMSE)에서 거시경제변수 모델이 뉴스 감성 분석 모델보다 낮은 값을 보여, 뉴스 감성 분석이 예측 성능을 오히려 저하시킨 것으로 나타났다. KG모빌리티의 경우에는 뉴스 감성 분석 모델이 거시경제변수 모델보다 모든 성능 지표에서 약간 더 낮은 오차 값을

보이며 더 우수한 성능을 나타냈다.

자동차 업종 전반적으로는 뉴스 감성 분석 모델이 필수적으로 예측 성능 개선에 크게 기여하지 않았으며, 일부 기업에서는 거시경제변수 모델이 더 높은 예측 정확성을 보였다. 이는 자동차 업종에서 뉴스 감성 분석이 주가 예측에 미치는 영향이 제한적임을 시사한다.

#### 4. 바이오

바이오 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 삼성바이오로직스, 셀트리온, 유한양행을 선정하였다.

삼성바이오로직스, 셀트리온, 유한양행 모두 거시경제변수 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 뉴스 감성 분석 모델보다 낮은 값을 나타내어, 뉴스 감성 분석이 예측 성능을 오히려 저하시킨 것으로 나타났다.

바이오 업종 전반적으로는 뉴스 감성 분석 모델이 모든 기업에서 예측 성능을 개선하지는 않았으며, 오히려 거시경제변수 모델이 더 나은 성능을 보였다. 이는 바이오 업종에서 뉴스 감성 분석이 주가 예측에 미치는 영향이 제한적임을 보여주고 있다.

## 5. 금융

금융 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 KB금융지주, 신한금융지주, 하나금융지주를 선정하였다.

KB금융지주의 경우, 거시경제변수 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 뉴스 감성 분석 모델보다 낮은 값을 나타내어, 뉴스 감성 분석이 예측 성능을 오히려 저하시킨 것으로 나타났다. 신한금융지주에서도 MAE 지표에서는 뉴스 감성 분석 모델이 약간 우수한 성능을 보였으나, MSE와 RMSE 지표에서는 거시경제변수 모델이 더 나은 성능을 나타내어 뉴스 감성 분석의 효과가 제한적임을 보여준다. 하나금융지주의 경우에도 거시경제변수 모델이 모든 지표에서 더 나은 성능을 보여, 뉴스 감성 분석이 금융 업종 예측 성능에 기여하지 않는 것으로 나타났다.

금융 업종에서는 뉴스 감성 분석을 포함한 모델이 예측 성능을 개선하지 못했으며, 거시경제변수 모델이 더 나은 성능을 보였다. 이는 금융 업종에서 뉴스 감성 분석이 주가 예측에 미치는 영향이 제한적임을 보여주고 있다.

## 6. 방산

방산 업종의 주가 예측 성능을 비교하기 위해, 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터, 거시경제변수, 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)을 사용하였다. 실험 대상 기업으로는 한화에어로스페이스, 한국항공우주산업, LIG넥스원을 선정하였다.

한화에어로스페이스와 LIG넥스원에서는 모든 성능 지표(MAE, MSE, RMSE)에서 거시경제변수 모델이 뉴스 감성 분석 모델보다 낮은 오차 값을 보여, 뉴스 감성 분석이 예측 성능을 오히려 저하시킨 것으로 나타났다. 반면, 한국항공우주산업에서는 뉴스 감성 분석 모델이 모든 지표에서 거시경제변수 모델보다 낮은 오차 값을 보이며 예측 성능을 향상시켰다.

방산 업종 전반적으로는 뉴스 감성 분석 모델이 모든 기업에서 예측 성능을 개선하지는 않았으며, 일부 기업에서는 오히려 거시경제변수 모델이 우수한 성능을 보였다. 이는 방산 업종에서 뉴스 감성 분석이 예측 성능에 미치는 영향이 제한적임을 보여주고 있다.

## 7. 조선

조선 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 HD한국조선해양, 한화오션, 삼성중공업을 선정하였다.

HD한국조선해양과 삼성중공업의 경우, 거시경제변수 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 뉴스 감성 분석 모델보다 낮은 값을 보여, 뉴스 감성 분석이 예측 성능을 오히려 저하시킨 것으로 나타났다. 한화오션의 경우에는



뉴스 감성 분석 모델이 모든 지표에서 약간 더 나은 성능을 보였으나 그 향상 정도는 미미한 수준이었다.

조선 업종 전반적으로는 뉴스 감성 분석 모델이 예측 성능에 크게 기여하지 않았으며, 오히려 거시경제변수 모델이 더 우수한 예측 성능을 보였다. 이는 조선 업종에서 뉴스 감성 분석이 주가 예측에 긍정적인 영향을 미치지 않음을 보여주고 있다.

## 8. 석유·화학

석유·화학 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 SK이노베이션, LG화학, 한화솔루션을 선정하였다.

SK이노베이션의 경우, MAE 지표에서는 뉴스 감성 분석 모델이 거시경제변수 모델보다 낮은 값을 보였다. 그러나 MSE와 RMSE 지표에서는 거시경제변수 모델이 더 나은 성능을 보여, 뉴스 감성 분석이 일관된 성능 향상을 제공하지 못했음을 확인할 수 있었다. LG화학의 경우 뉴스 감성 분석 모델이 모든 지표에서 약간 더 나은 성능을 보였으며, 예측 성능 개선에 기여하였다. 반면, 한화솔루션에서는 거시경제변수 모델이 모든 지표에서 더 나은 성능을 보여, 뉴스 감성 분석이 예측 성능에 기여하지 못했다.

석유·화학 업종 전반적으로 기업별로 상이한 결과를 보였으며, 뉴스 감성 분석이 모든 기업에서 예측 성능을 향상시키지 않는 것으로 나타났다.

이는 석유·화학 업종에서 뉴스 감성 분석이 일부 기업에만 예측 성능 개선에 기여할 가능성을 시사한다.

## 9. IT

IT 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 네이버, 카카오, KT를 선정하였다.

네이버, 카카오, KT 모두 거시경제변수 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 뉴스 감성 분석 모델보다 더 낮은 값을 보여 예측 성능이 우수한 것으로 나타났다. 이는 IT 업종에서 뉴스 감성 분석이 예측 성능 향상에 기여하지 못했음을 보여주고 있다.

IT 업종에서는 뉴스 감성 분석을 포함한 모델이 예측 성능을 개선하지 못했으며, 거시경제변수 모델이 더 우수한 성능을 보였다. 이는 IT 업종의 주가 예측에 뉴스 감성 분석이 유의미한 정보를 제공하지 못할 가능성을 시사한다.

## 10. 철강

철강 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터, 거시경제변수, 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 포스코홀딩스, 현대제철, KG스틸을 선정하였다.

포스코홀딩스와 KG스틸의 경우, 모든 성능 지표(MAE, MSE, RMSE)에서 거시경제변수 모델이 더 낮은 오차 값을 보여 뉴스 감성 분석이 예측 성능에 기여하지 못한 것으로 나타났다. 반면, 현대제철의 경우 뉴스 감성 분석 모델이 모든 지표에서 더 낮은 오차 값을 보이며 예측 성능을 개선하였다.

철강 업종 전체적으로는 기업별로 예측 성능 차이가 상이하게 나타났으며, 뉴스 감성 분석이 일관되게 성능 향상을 제공하지 않았다. 이는 철강 업종에서

뉴스 감성 분석이 일부 기업에서만 예측 성능 개선에 기여할 가능성을 시사한다.

## 11. 건설

건설 업종에 대해 주가 데이터와 거시경제변수를 조합한 LSTM 모델(이하 거시경제변수 모델)과 주가 데이터와 거시경제변수 및 뉴스 감성 분석을 조합한 LSTM 모델(이하 뉴스 감성 분석 모델)의 예측 성능을 비교하였다. 실험 대상 기업으로는 현대건설, 대우건설, GS건설을 선정하였다.

현대건설의 경우, MAE 지표에서는 뉴스 감성 분석 모델이 더 나은 성능을 보였으나, MSE와 RMSE 지표에서는 거시경제변수 모델이 우수한 성능을 나타내어, 뉴스 감성 분석이 전체적인 예측 성능 향상에 크게 기여하지는 않았다. 반면, 대우건설과 GS건설에서는 뉴스 감성 분석 모델이 모든 성능 지표(MAE, MSE, RMSE)에서 거시경제변수 모델보다 낮은 오차 값을 보이며 예측 성능을 개선하였다.

건설 업종 전반적으로는 뉴스 감성 분석이 일부 기업에서 예측 성능 향상에 있어 긍정적인 영향을 미쳤음을 보여주고 있다.

## 제 2절 두 모델 간의 대응표본 t-검정 결과 분석 비교

### 1. 전체 설명

본 연구에서는 두 가지 예측 모델(거시경제변수 모델, 뉴스 감성 분석 모델)의 성능 차이를 업종별로 평가하기 위해 대응표본 t-검정을 수행하였다.

동일한 집단에서 측정된 두 변수의 평균 차이가 유의미한지를 검증하는 통계 기법인 대응표본 t-검정을 활용하여, MAE, MSE, RMSE의 세 가지 예측 성능 지표에 대해 두 모델 간의 성능 차이를 분석하는 데 사용되었다.

파이썬의 `ttest_rel` 함수를 활용하여 MAE, MSE, RMSE 각각에 대해 대응표본 t-검정을 수행하였다. 각 대응표본 t-검정의 결과로 도출된 t-통계량과 p-값을 바탕으로 두 모델 간의 성능 차이가 통계적으로 유의미한지 판단하였다. 본 실험에서는 유의수준을 0.1로 설정하였으며, p-값이 0.1 미만일 경우 두 모델 간의 성능 차이가 통계적으로 유의미한 것으로 해석하였다. 반면, p-값이 0.1 이상일 경우 두 모델의 성능 차이는 통계적으로 유의미하지 않은 것으로 간주하였다. 이 과정을 통해, 각 업종별로 두 모델 간의 성능 차이가 전체 모집단에도 적용 가능한지를 검토하였다.

## 2. 반도체

반도체 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 지표 모두에서 두 모델 간 성능 차이가 통계적으로 유의미하지 않은 것으로 나타났다. MAE의 p-값은 0.3669, MSE는 0.3871, RMSE는 0.3568로 모두 유의수준 0.1을 초과하여, 반도체 업종에서는 거시경제변수 모델과 뉴스 감성 분석 모델 간에 유의미한 예측 성능 차이가 없는 것으로 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	-1.1567	0.3669	유의미하지 않음
MSE	-1.0969	0.3871	유의미하지 않음
RMSE	-1.1878	0.3568	유의미하지 않음

<표 6> 반도체 업종 대응표본 t-검정 결과

## 3. 자동차

자동차 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 지표 모두에서 두 모델 간 성능 차이가 통계적으로 유의미하지 않은 것으로 나타났다. MAE의 p-값은 0.3404, MSE는 0.2587, RMSE는 0.2625로, 모든 지표가 유의수준 0.1을 초과하였다. 따라서 자동차 업종에서는 거시경제변수 모델과 뉴스 감성 분석 모델 간 예측 성능에 유의미한 차이가 없는 것으로 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	1.2410	0.3404	유의미하지 않음
MSE	1.5620	0.2587	유의미하지 않음
RMSE	1.5444	0.2625	유의미하지 않음

<표 7> 자동차 업종 대응표본 t-검정 결과

#### 4. 바이오

바이오 업종에 대한 대응표본 t-검정 결과, MAE와 RMSE 지표에서는 두 모델 간 성능 차이가 통계적으로 유의미한 것으로 나타났다. MAE의 t-통계량은 2.9542, p-값은 0.0980으로 유의수준 0.1 이하에서 유의미한 차이를 보였고, RMSE의 경우 t-통계량은 4.1518, p-값은 0.0534로 유의미한 성능 차이를 확인할 수 있었다. 반면, MSE 지표는 p-값이 0.2046으로 유의수준을 초과하여 유의미한 차이가 없는 것으로 나타났다. 따라서 바이오 업종에서는 MAE와 RMSE 지표에서 거시경제변수 모델과 뉴스 감성 분석 모델 간 예측 성능에 유의미한 차이가 있는 것으로 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	2.9542	0.0980	유의미함
MSE	1.8559	0.2046	유의미하지 않음
RMSE	4.1518	0.0534	유의미함

<표 8> 바이오 업종 대응표본 t-검정 결과

## 5. 금융

금융 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다. MAE의 p-값은 0.4614, MSE의 p-값은 0.6373, RMSE의 p-값은 0.5458로 나타났으며, 모두 유의수준 0.1을 초과하였다. 따라서 금융 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	0.9041	0.4614	유의미하지 않음
MSE	0.5505	0.6373	유의미하지 않음
RMSE	0.7210	0.5458	유의미하지 않음

<표 9> 금융 업종 대응표본 t-검정 결과

## 6. 방산

방산 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다. MAE의 p-값은 0.5393, MSE의 p-값은 0.3645, RMSE의 p-값은 0.3754로 나타났으며, 모두 유의수준 0.1을 초과하였다. 따라서 방산 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	0.9041	0.4614	유의미하지 않음
MSE	0.5505	0.6373	유의미하지 않음
RMSE	0.7210	0.5458	유의미하지 않음

<표 10> 방산 업종 대응표본 t-검정 결과

## 7. 조선

조선 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다. MAE의 p-값은 0.3124, MSE의 p-값은 0.3154, RMSE의 p-값은 0.2764로 나타났으며, 모두 유의수준 0.1을 초과하였다. 따라서 조선 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	1.3393	0.3124	유의미하지 않음
MSE	1.3282	0.3154	유의미하지 않음
RMSE	1.4825	0.2764	유의미하지 않음

<표 11> 조선 업종 대응표본 t-검정 결과



## 8. 석유·화학

석유·화학 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다. MAE의 p-값은 0.5052, MSE의 p-값은 1.0, RMSE의 p-값은 0.9673으로 나타났다. 모두 유의수준 0.1을 초과하였다. 따라서, 석유·화학 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	-0.8051	0.5052	유의미하지 않음
MSE	0.0	1.0	유의미하지 않음
RMSE	0.0463	0.9673	유의미하지 않음

<표 12> 석유·화학 업종 대응표본 t-검정 결과

## 9. IT

IT 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미한 것으로 분석되었다.

MAE의 p-값은 0.0448, MSE의 p-값은 0.0928, RMSE의 p-값은 0.0512로 나타났으며, 모두 유의수준 0.1 이하에서 유의미한 차이를 보였다. 따라서, IT 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 있다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	4.5664	0.0448	유의미함
MSE	3.0500	0.0928	유의미함
RMSE	4.2489	0.0512	유의미함

<표 13> IT 업종 대응표본 t-검정 결과

## 10. 철강

철강 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다.

MAE의 p-값은 0.5558, MSE의 p-값은 0.6190, RMSE의 p-값은 0.8196으로 나타났다으며, 모두 유의수준 0.1을 초과하였다. 따라서, 철강 업종에서 거시경제 변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	-0.7013	0.5558	유의미하지 않음
MSE	0.5828	0.6190	유의미하지 않음
RMSE	0.2594	0.8196	유의미하지 않음

<표 14> 철강 업종 대응표본 t-검정 결과

## 11. 건설

건설 업종에 대한 대응표본 t-검정 결과, MAE, MSE, RMSE 모든 지표에서 두 모델 간의 성능 차이가 통계적으로 유의미하지 않은 것으로 분석되었다.

MAE의 p-값은 0.3295, MSE의 p-값은 0.4352, RMSE의 p-값은 0.4298로 나타났으며, 모두 유의수준 0.1을 초과하였다. 따라서, 건설 업종에서 거시경제변수 모델과 뉴스 감성 분석 모델 간의 예측 성능에 유의미한 차이가 없다고 판단된다.

성능지표	t-statistic	p-value	유의미성
MAE	-1.2782	0.3295	유의미하지 않음
MSE	-0.9679	0.4352	유의미하지 않음
RMSE	-0.9815	0.4298	유의미하지 않음

<표 15> 건설 업종 대응표본 t-검정 결과

### 제 3절 종합 실험 결과 분석

주가 예측 성능 비교와 대응표본 t-검정 결과를 종합적으로 고려한 결과, 바이오와 IT 업종에서는 뉴스 감성 분석 모델과 거시경제변수 모델 간의 평균 차이가 통계적으로 유의미하게 나타났다. 이러한 차이는 <그림 8>의 바이오 업종과 <그림 9>의 IT 업종 종목들의 주가 그래프를 통해 확인할 수 있다. 그러나 예측 성능 향상에 대한 다음과 같은 제한된 결과가 확인되었다. 첫째, 예측 오차 측면에서 거시경제변수 모델이 더 나은 성능을 보였으며, 뉴스 감성 분석 추가가 예측 성능을 일관되게 개선하지 않았다. 둘째, 뉴스 감성 분석이 주가 예측에 미치는 영향은 제한적이며, 모든 업종에서 일관된 성능 개선을 보이지 않았다. 예를 들어, 바이오와 IT 업종에서는 통계적으로 유의미한 차이가 있었지만, 예측 성능 면에서는 거시경제변수 모델이 더 우수한 결과를 보였다. 셋째, 바이오와 IT 업종을 제외한 다른 업종에서는 두 모델 간 성능 차이가 통계적으로 유의미하지 않아, 뉴스 감성 분석이 모든 업종에서 일관된 성능 향상을 제공하지는 않는다는 점을 시사한다. 이러한 결과는 거시경제변수가 주가 변동의 주요 요소로서 장기적 패턴을 더 안정적으로 반영한다는 점에 있고, 뉴스 감성 분석은 KOSPI 전반을 아우르는 뉴스 데이터가 특정 업종이나 기업에 맞춘 예측에는 제한적이다 보니 전체적인 예측 성능 개선에는 기여하지 못한 것으로 보인다.

## 삼성바이오로직스



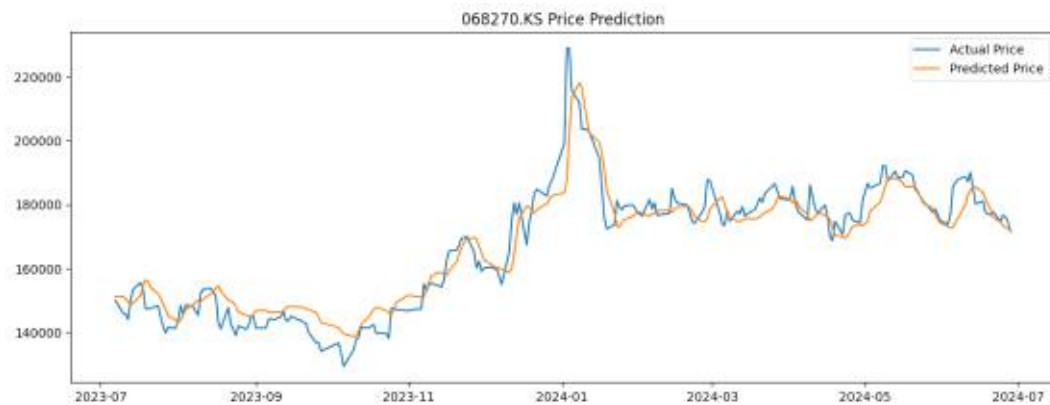
거시경제변수 모델



뉴스 감성 분석 모델

<그림 8-1> 삼성바이오로직스 주가 예측 그래프

## 셀트리온



거시경제변수 모델



뉴스 감성 분석 모델

<그림 8-2> 셀트리온 주가 예측 그래프

## 유한양행



거시경제변수 모델



뉴스 감성 분석 모델

<그림 8-3> 유한양행 주가 예측 그래프

## 네이버



거시경제변수 모델

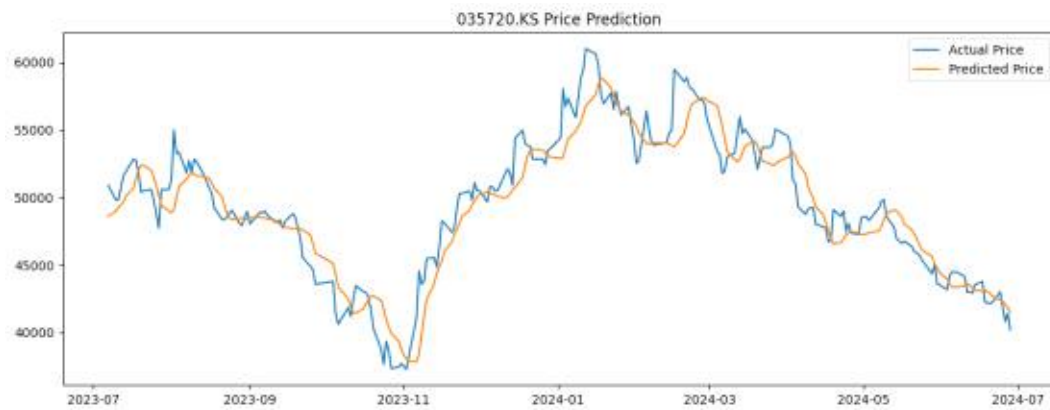


뉴스 감성 분석 모델

<그림 9-1> 네이버 주가 예측 그래프



## 카카오



거시경제변수 모델



뉴스 감성 분석 모델

<그림 9-2> 카카오 주가 예측 그래프

KT



거시경제변수 모델



뉴스 감성 분석 모델

<그림 9-3> KT 주가 예측 그래프

## 제 6장 결론 및 시사점

### 제 1절 연구 결과 요약 및 시사점

본 연구에서는 LSTM 시계열 모델을 기반으로 10개 업종(반도체, 자동차, 바이오, 금융, 방산, 조선, 석유, 화학, IT, 철강, 건설)의 주가 예측 성능을 평가하고, 뉴스 감성 분석의 추가가 예측에 미치는 영향을 분석하였다. 연구 결과와 시사점은 다음과 같다.

첫째, 거시경제변수만을 사용한 모델과 뉴스 감성 분석을 결합한 모델을 비교한 결과, 바이오와 IT 업종에서 두 모델 간 통계적으로 유의미한 차이가 발견되었다. 다만, 예측 오차 측면에서는 거시경제변수 모델이 뉴스 감성 분석 모델보다 더 낮은 오차를 보였다.

둘째, 예측 오차를 기준으로 보면, 거시경제변수만을 사용한 모델이 더 낮은 오차를 보여 모든 업종에서 뉴스 감성 분석이 주가 예측 성능을 일관되게 향상시키지 않는 것으로 확인되었다.

셋째, 거시경제변수는 전체 시장 동향과 주요 경제 지표를 반영하여 주가 예측에 핵심적인 요소로 작용하는 것으로 나타나, 주가 예측에서 안정적이고 유의미한 변수임을 보여준다. 또한, 예측 결과 그래프에서 실제 주가의 변동성과 유사한 패턴이 관찰되어 주가의 큰 흐름을 예측하는 데 적용 가능성이 있음을 확인하였다. 이러한 특성을 활용하면 기업의 의사 결정, 리스크 관리, 투자 전략 수립 등에 실질적인 참고자료로 활용될 수 있을 것으로 기대한다.

마지막으로, 향후 연구에서는 업종별 특성에 맞춘 맞춤형 주가 예측 모델 개발이 필요하다. 업종별 특성을 반영한 예측 모델을 구축함으로써 주가 예측의 정확성과 신뢰성을 한층 높일 수 있을 것이다.

## 제 2절 연구의 한계 및 향후 연구 방향

본 연구는 뉴스 감성 분석과 LSTM 모델을 활용하여 주가 예측 성능을 평가하였으나, 다음과 같은 한계가 있으며 이를 보완하기 위한 향후 연구 방향을 제시하고자 한다.

첫째, 감성 분석에 사용된 뉴스 데이터의 출처가 제한적이어서 분석 결과의 신뢰성과 대표성에 영향을 미칠 수 있다. 향후 다양한 뉴스 소스를 통해 폭넓은 데이터를 수집하여 분석의 정확성과 일반성을 높일 필요가 있다.

둘째, 본 연구는 LSTM 모델을 사용해 주가 예측을 수행하였으나, 다른 머신러닝 및 딥러닝 모델과의 성능 비교는 이루어지지 않았다. 향후 연구에서는 다양한 모델을 탐색하여 예측 성능을 개선하고, 모델 간 성능을 비교하는 연구가 필요하다.

셋째, ‘코스피’ 키워드만을 사용해 뉴스 데이터를 수집했으므로, 일부 업종이나 기업의 감성 정보가 충분히 반영되지 않았을 가능성이 있다. 이를 보완하기 위해 향후 연구에서는 다양한 키워드를 포함하여 업종별로 세밀한 감성 정보를 반영해야 할 것이다.

넷째, 본 연구는 VADER 감성 사전을 사용하여 감성 분석을 수행하였다. 감성 사전 기반 접근은 간편하지만, 보다 정교한 분석을 위해 머신러닝 기반 감성 분석 기법을 적용할 필요가 있다.

본 연구는 LSTM 모델과 뉴스 감성 분석을 결합하여 업종별 주가 예측 성능을 평가하고, 감성 분석이 예측에 미치는 영향을 탐구하였다. 그 결과, 감성 분석이 일부 업종에서는 예측 성능에 긍정적인 영향을 미쳤으나, 모든 업종에서 일관된 성능 향상을 제공하지는 않았다. 이는 감성 분석이 주가 예측에 보조적으로 활용될 수 있으나, 거시경제변수가 여전히 주가 예측에서 핵심적인 요소임을 재확인시켜준다.

향후 연구에서는 더 다양한 데이터 출처와 분석 기법을 통합하여 주가 예측의 정확성을 높이고, 업종별 특성에 맞춘 예측 모델 개발에 기여할 수 있기를 기

대한다. 본 연구가 감성 분석 기반 주가 예측 연구의 기초 자료로 활용되고, 후속 연구를 통해 주가 예측 모델의 성능이 한층 개선되기를 기대한다.

## 참 고 문 헌

[1] S.Hochreiter, J.Schmidhuber (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

[2] 송영렬, 김흥기, 한동협 (2009). 다요인모형을 기초로 한 거시경제변수와 주가간의 영향에 관한 연구. e-비즈니스연구, 10(3), 97-128.

[3] K.S.Sujit, B.Rajesh Kumar (2011). Study on dynamic relationship among gold price, oil price, exchange rate and stock market returns. International Journal of Applied Business and Economic Research, 9(2), 145-165.

[4] 박강희, 신현정 (2011). 시계열 네트워크에 기반한 주가예측. 경영과학, 28(1), 53-60.

[5] Johan Bollen, Huina Mao, Xiaojun Zeng (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

[6] 김유신, 김남규, 정승렬 (2012). 뉴스와 주가 : 빅데이터 감성분석을 통한 지능형 투자의사결정모형. 지능정보연구. 18(2), 143-156.

[7] 김동영, 박제원, 최재현 (2014). SNS와 뉴스기사의 감성분석과 기계학습을 이용한 주가예측 모형 비교 연구. 한국IT서비스학회지, 13(3), 211-233.

[8] Kai Chen, Yi Zhou, Fangyan Dai (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. 2015 IEEE International Conference on Big Data (Big Data), 2823-2824.

[9] 윤일현 (2016). 거시경제변수가 국내주가에 미치는 영향에 관한 연구. 지역산업연구, 39(2), 37-53.

[10] 신동하, 최광호, 김창복 (2017). RNN과 LSTM을 이용한 주가 예측을 향상을 위한 딥러닝 모델. 한국정보기술학회논문지, 15(10), 9-16.

[11] 성노윤, 남기환 (2017). 온라인 뉴스 및 거시경제 변수를 활용한 주가예측. Entru Journal of Information Technology, 16(2), 41-54.

[12] 감형규, 신용재 (2017). 거시경제변수가 주식수익률에 미치는 영향에 관한 연구. 대한경영학회지, 30(1), 33-52.

[13] Thomas Fischer & Christopher Krauss (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

[14] 주일택, 최승호 (2018). 양방향 LSTM 순환신경망 기반 주가예측모델. 한국정보전자통신기술학회논문지, 11(2), 204-208.

[15] 장은아, 최희련, 이홍철 (2020). BERT를 활용한 뉴스 감성분석과 거시경제지표 조합을 이용한 주가지수 예측. 한국컴퓨터정보학회 논문지, 25(5), 47-56.

[16] 한태동 (2021). LSTM을 이용한 주가 예측: 기술 지표, 거시 경제 지표, 시장 심리의 조합을 중심으로. 융복합지식학회논문지, 9(4), 189-198.

[17] 박상언, 강주영 (2023). 파이썬 텍스트 마이닝 완벽 가이드 - 자연어 처리 기초부터 딥러닝 기반 BERT와 트랜스포머까지.



# Abstract

## A Comparative Study on the Performance of LSTM Time Series Models for Stock Price Forecasting by Industry Using Macroeconomic Variables and News Sentiment Analysis

Lee, Jeong Hyun

Dept, of Management Consulting

The Graduate School, Hanyang University

Stock prices are affected not only by internal factors such as a company's financial position and management strategy, but also by various external factors such as macroeconomic variables, industrial trends, and political and social changes. Due to these factors, stock prices are volatile, which has a significant impact on the overall economy, including government and corporate policies. Accordingly, research using various analysis models to predict stock prices is being conducted continuously. Previous studies on these stock price prediction models mainly focused on analysis using artificial neural networks such as machine learning and deep learning, and in particular, stock price prediction studies combining macroeconomic variables and news sentiment analysis were conducted. However, existing studies have mainly focused on stock price prediction, which is focused on individual companies or specific industrial groups, and few studies have analyzed the effect of external factors on stock price prediction by classifying them by industry. Accordingly, this study selected three

major companies representing 10 major industries, including semiconductor, automobile, bio, and finance, for KOSPI-listed companies, and compared the stock price prediction performance by dividing them into two models using an LSTM time series model that combines macroeconomic variables and news sentiment analysis. Through this, we tried to supplement the limitations of existing studies by analyzing the effects of external factors on stock price prediction by industry. VADER sentiment analysis was used for news sentiment analysis, and the first model was a combination of stock price data and macroeconomic variables, and the second model was divided into two models that combined stock price data, macroeconomic variables, and news sentiment analysis, and LSTM, a type of deep learning algorithm, was applied. The results of this study confirmed that news sentiment analysis did not consistently improve predictive performance in all industries, and that macroeconomic variables are particularly important key factors in reflecting economic indicators and market trends and explaining stock price volatility. These results explain that it is necessary to develop a customized stock price prediction model for the characteristics of each industry.

Keywords: Stock Price Prediction, Macroeconomic Variables, News Sentiment Analysis, LSTM, Paired sample t-test

## 연구 윤리 서약서

본인은 한양대학교 대학원생으로서 이 학위논문 작성 과정에서 다음과 같이 연구 윤리의 기본 원칙을 준수하였음을 서약합니다.

첫째, 지도교수의 지도를 받아 정직하고 엄정한 연구를 수행하여 학위논문을 작성한다.

둘째, 논문 작성시 위조, 변조, 표절 등 학문적 진실성을 훼손하는 어떤 연구 부정행위도 하지 않는다.

셋째, 논문 작성시 논문유사도 검증시스템 "카피킬러"등을 거쳐야 한다.

2024년12월09일

학위명 : 석사

학과 : 경영컨설팅학과

지도교수 : 황승준

성명 : 이정현

한 양 대 학 교 대 학 원 장 귀 하

## Declaration of Ethical Conduct in Research

I, as a graduate student of Hanyang University, hereby declare that I have abided by the following Code of Research Ethics while writing this dissertation thesis, during my degree program.

"First, I have strived to be honest in my conduct, to produce valid and reliable research conforming with the guidance of my thesis supervisor, and I affirm that my thesis contains honest, fair and reasonable conclusions based on my own careful research under the guidance of my thesis supervisor.

Second, I have not committed any acts that may discredit or damage the credibility of my research. These include, but are not limited to : falsification, distortion of research findings or plagiarism.

Third, I need to go through with Copykiller Program(Internet-based Plagiarism-prevention service) before submitting a thesis."

DECEMBER 09, 2024

Degree : Master  
Department : DEPARTMENT OF MANAGEMENT CONSULTING  
Thesis Supervisor : Seung-June Hwang  
Name : LEE JEONG HYUN

(Signature)