

# 거대언어모델을 이용한 뉴스 감정 분석 및 이를 이용한 주가 예측\*

김명섭<sup>1</sup>, 유태훈<sup>1</sup>, 정영민<sup>2</sup>, 이영민<sup>3</sup>

<sup>1</sup>서강대학교 컴퓨터공학과, <sup>2</sup>서강대학교 융합교육원, AI·SW교육센터, <sup>3</sup>서강대학교 인공지능학과  
{kyh147159,taehun2608,ymchung,ymyi}@sogang.ac.kr

## Stock Price Prediction Using News Sentiment Analysis Leveraging Large Language Model

Myeongseop Kim<sup>1</sup>, Taehun Yoo<sup>1</sup>, Youngmin Chung<sup>2</sup>, Youngmin Yi<sup>3</sup>

<sup>1</sup>Sogang University, Dept. of Computer Science and Engineering, <sup>2</sup>Sogang University, AI·SW Education Center, Sogang Institute for Convergence Education, <sup>3</sup>Sogang University, Dept. of Artificial Intelligence  
{kyh147159,taehun2608,ymchung,ymyi}@sogang.ac.kr

### 요약

주식 시장에 참여하는 시장 참여자들은 이용 가능한 정보를 기반으로 투자를 위한 결정을 내린다. 따라서 이용 가능한 정보 중 가장 대표적인 정보인 뉴스는 주가의 등락에 영향을 미친다. 본 연구에서는 시장 참여자가 1,336개의 뉴스 기사를 긍부정(Valence), 강도(Arousal), 중요도(Importance) 차원에서 분류하여 1,076일의 데이터로 Random Forest 모델에 학습을 진행한 후, 265일 기간의 주가를 예측해 성능을 비교 분석했다. 레이블링에는 OpenAI사의 o4-mini-high와 Google사의 gemini-2.5-pro-preview-03-25 모델을 활용했으며, 레이블링을 하지 않은 모델과 함께 실제 데이터와의 평균 제곱근 편차(Root Mean Squared Error, RMSE)를 측정하였다. 그 결과, gemini-2.5-pro-preview-03-25 모델은 4.2% 성능이 향상되었으나, o4-mini-high는 7.8% 하락된 것을 확인하였다.

### 1. 서론

금융시장에서는 금융자산가격의 변동을 시장기초요인만이 아닌, 새로운 뉴스에 대한 시장 참여자의 심리 변화로 설명할 수 있는 경우가 존재한다[1].

특히, 개별 주식 대신 S&P 500과 같은 시장 전체의 추이를 나타내는 지수를 추종하는 ETF의 경우, 시장 참여자의 시장에 대한 전망은 곧 이를 반영하는 지수를 추종하는 ETF에 대한 투자 심리에 영향을 미쳐 가격에 반영이 된다[2].

본 연구에서는 이를 반영하여 예측하기 위하여 뉴스 기사를 긍부정(Valence)과 강도(Arousal), 그리고 중요도(Importance)로 나누어서 감정 분석을 진행한 다음, 이를 모델에 학습시켜서 뉴스 기사를 바탕으로 주가를 예측하는 방안을 제안한다. 또한 금융자산가격의 변동에 영향을 미치는 시장 참여자의 심리를 2020년 1월부터 2025년 4월까지의 기간 동안의 총 1,336개의 뉴스 기사를 OpenAI사의 o4-mini-high 모델[3]과 Google사의 gemini-2.5-pro-preview-03-25 모델[4]을 기반으로 정량적으로 분석하였으며, 그 이후 1,076 개장일의 데이터를 학습한 이후, 265 개장일의 주가 변동을 예측하도록 하여 그 결과를 비교하였다.

### 2. 관련 연구

#### 2.1 주가 예측

관련된 연구를 종합하여 분석한 문헌에서는 딥러닝 모델을 사용한 경우, 그렇지 않은 경우보다 주가 예측에 있어 더 높은 성능을 보이고, 기술적 분석 및 감정 분석을 활용하여 주가를 예측할 때, 그렇지 않을 때보다 더 좋은 성능을 보이며, 또한 단일 모델보다는 하이브리드 모델이 주가 예측에 더 뛰어나다고 보고하였다[5].

### 2.2 감정 분석

#### 2.2.1 LLM을 이용한 감정 분석

8종의 거대언어모델(LLM)과 33명의 인간 작업자의 레이블링을 비교하여 분석한 연구에서는 거대언어모델이 인간보다 감정의 밀집도(Emotional intensity)를 인간보다 더 정확하게 측정하고, 또한 정치적 중립성 및 긍부정에 있어서도 인간만큼의 신뢰도를 보여준다는 결론을 도출한 바 있다[6].

### 2.3 뉴스 정보를 이용한 주가 예측

[7]에서는 긍부정의 정도를 BERT로 예측해 0과 1 사이의 값으로 나타낸 다음, LSTM 및 CNN으로 학습을 시킨바 있으며, 이 과정에서 한국 주요 4개 기업에 대해 감정 분석이 유의미한 정확도 증가로 이어짐을 보였다.

[8]에서는 호가창과 뉴스 헤드라인을 이용한 다음, Word2Vec을 이용해 그 특징을 추출한 다음 CNN을 이용해 모델을 학습시킨 다음, 모의 투자를 진행해 최대 \$2,840의 수익을 얻은 바 있다.

### 3. 데이터 수집, 모델 설계 및 학습 방법

#### 3.1 개요

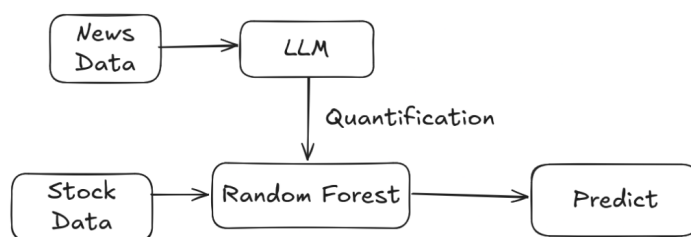


그림 1. 모델 설계도

본 연구에서는 뉴스 데이터를 가져와 LLM을 거쳐 긍부정

\* 본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (2024-0-00043)

(Valence), 강도(Arousal), 중요도(Importance)로 나누어 정량화를 하고, 수집한 주식 데이터와 함께 Random Forest 기반 모델을 학습시켜 주가 등락폭을 예측할 수 있도록 한다.

공부정(Valence)은 기사에 대한 긍정적·부정적 반응을 나타내는 지표이며, 심리학계에서 널리 통용되는 리커트 척도를 활용한다. 강도(Arousal)의 경우, 반응의 정도를 나타내는 지표이며, 1점부터 5점까지의 척도로 나타낸다. 또한 중요도(Importance)의 경우, 기사가 관심 있는 주제와 어느 정도의 연관성이 있는지, 또한 해당 주제에 대한 영향이 어느 정도인지를 나타내는 지표이며, 1점부터 5점까지로 나타내어 진다.

표 1. 공부정, 강도, 중요도의 레이블링 예시[9]

TOPIC: KOSPI
TITLE: 대한상공회의소, “4분기 경기도 어려울 것”
Article: 대한상공회의소가 최근 10개 업종별 단체와 공동으로 ‘2012년 4분기 산업기상도’를 조사한 결과를 발표한 보도 자료에 따르면, ‘맑음’으로 전망된 업종은 하나도 없었다. 정보통신·기계·석유화학·섬유·자동차 등 5개 업종은 ‘구름조금’, 철강·정유·의류·건설 등 4개 업종은 ‘흐림’으로 전망되었으며, 조선업종은 ‘비’로 예보됐다.
LLM Labeling: Valence 2, Arousal 2, Importance 6
Researcher Labeling: Valence 2, Arousal 2, Importance 4

표 1에서는 LLM과 인간 연구자 모두 기사의 논조에 대해 ‘약간 부정’으로 판단하여 공부정(Valence)에 대해 5점 만점 중 2점을 주었으며, 강도(Arousal)에 있어서는 ‘약간의 충격’을 의미하는 5점 만점 중 2점을 주었으나, 중요도(Importance)에 있어서는 LLM은 ‘다소 중대함’을 의미하는 6점, 인간 연구자는 ‘약간 중대함’을 의미하는 4점을 부여하였다.

### 3.1 데이터 수집 및 전처리

구글 RSS 피드를 이용해 최신 뉴스 기사 5000개 가량을 수집한 다음, 개장일의 주요 뉴스 하나만을 선정하여 1,336개의 기사를 수집한 다음 BeautifulSoup[10] 라이브러리로 후처리하고, 그 이후 야후 Finance에서 주가 정보를 수집하였다. 그 이후, LLM 모델을 이용해 감정분석을 진행한 다음, 50개의 샘플 데이터를 추출해 연구자가 감정 분석을 한 것과 비교하였다.

### 3.3 모델 설계 및 학습 방법

#### 3.3.1 모델 설계 방법

Ensemble Model을 비롯한 고전적 머신 러닝을 사용한 기법은 아직까지도 주가 예측의 방법론에서 배제되지 않고 있다[5]. 따라서, 본 연구에서는 Ensemble Model의 대표적인 사례인 Random Forest를 채택하였다.

#### 3.3.2 모델 학습 방법

레이블링한 데이터를 실험군으로, 그것이 없는 것을 대조군으로 한 다음, 시계열 데이터의 학습할 윈도우 사이즈를 Feature

에 넣은 다음, RMSE(Root Mean Squared Error)를 손실 함수로 하여 Random Forest 모델을 학습시켰다.

## 4. 실험 결과 및 분석

### 4.1 실험 설계

실험 설계 과정에서는 훈련 데이터와 추론 데이터를 8:2 비율로 나누어 1,336일 개장일의 데이터에 대해 1,076개 데이터와 265 데이터로 분리하였으며, 훈련 과정에서는 훈련 데이터셋을 5분할 하여 교차 검증(Cross Validation)을 하는 방식으로 진행하였다. 학습 과정에서는 skforecast[11] 및 scikit-learn[12] 라이브러리를 활용하였다. 또한, 최적의 파라미터를 찾기 위해 Optuna[13] 라이브러리로 파라미터를 찾은 다음, 대조군과 실험군 각각에 1,076개 데이터를 학습시켰다.

실험군은 o4-mini-high 모델로 레이블링한 데이터로 학습한 것과 gemini-2.5-pro-preview-03-25로 레이블링한 데이터로 학습한 것으로 하였고, 대조군은 레이블링 데이터가 없이 S&P 500 종가 데이터만 있는 것으로 하였다.

### 4.2 실험 결과

#### 4.2.1 데이터셋 전처리 결과

데이터셋에 대해 50개의 샘플을 추출하여 o4-mini-high 모델과 gemini-2.5-pro-preview-03-25 모델을 이용하여 레이블링한 것과, 연구자가 직접 레이블링을 한 것을 비교한 결과는 아래 표와 같다. 레이블링의 경우, 1부터 5까지 나타내지는 공부정(Valence), 강도(Arousal)의 경우, 20%를 넘는 오차가 나는 경우 음성으로 판단하여, 아래와 같은 식이 만족되면 진양성으로 판단하였다.

$$|\text{model predict}_{\text{scale}} - \text{human labeling}_{\text{scale}}| \leq 1 \quad (1)$$

1부터 10으로 나타내지는 중요성(Importance)의 경우, 아래와 같은 식이 만족되면 진양성으로 판단하였다.

$$|\text{model predict}_{\text{scale}} - \text{human labeling}_{\text{scale}}| \leq 2 \quad (2)$$

레이블링 결과를 위에서 상기 수식에 따라 정확도를 계산한 결과, gemini-2.5-pro-preview-03-25 모델의 경우, 공부정(Valence), 강도(Arousal), 중요도(Importance)의 정확도(Accuracy)는 86%, 44%, 92%였으며, o4-mini-high 모델의 경우, 90%, 52%, 92%였다. 표2, 표3, 표4에서 확인할 수 있듯이 gemini-2.5-pro-preview-03-25 모델이 뉴스 기사의 공부정(Valence)에 대해서는 연구자와 o4-mini-high 모델보다 낙관적으로 레이블링하는 경향이 있었지만, 강도(Arousal)과 중요도(Importance)에 대해서는 반대의 경향이 확인된다.

#### 4.2.2 모델 예측 실험 결과

실험 결과는 아래 표와 같다. 대조군의 경우, Baseline로 표기하였다. 실험 결과 표 5에서 확인할 수 있듯이, o4-mini-high의 경우 대조군보다 오히려 성능이 나빠지는 결과가 보여졌지만, gemini-2.5-pro-preview-03-25의 경우 4.2% 성능이 향상되는

모습이 확인되었다.

표 2. 긍부정(Valence)에 대한 레이블링 결과

	gemini-2.5-pro-preview-03-25	o4-mini-high	Human labeling
1	3	3	9
2	2	13	16
3	45	33	22
4	0	1	3
5	0	0	0

표 3. 강도(Arousal)에 대한 레이블링 결과

	gemini-2.5-pro-preview-03-25	o4-mini-high	Human labeling
1	32	28	7
2	13	11	11
3	2	7	22
4	0	2	10
5	3	2	0

표 4. 중요도(Importance)에 대한 레이블링 결과

	gemini-2.5-pro-preview-03-25	o4-mini-high	Human labeling
1	40	28	28
2	3	17	10
3	2	0	4
4	1	2	4
5	1	0	2
6	0	1	2
7	1	0	0
8	2	1	0
9	0	0	0
10	0	1	0

표 5. 추론 데이터과의 RMSE 오차

	gemini-2.5-pro-preview-03-25	o4-mini-high	Human labeling
RMSE	257.68	278.03	247.04

5. 모델 예측 실험 결과

금융시장의 시장 참여자는 이용 가능한 정보에 반응하여 투자한다. 따라서 본 연구에서는 이를 고려하기 위해 LLM을 이용해 긍부정(Valence), 강도(Arousal), 중요도(Importance)을 레이블링한 다음 이를 기반으로 주가를 예측할 수 있고, 또한 LLM의 모델에 따라서 감정 분석을 진행하지 않은 것보다 성능이 악화되는 경우도, 개선되는 경우도 있음을 확인하였다. 향후 연구에서는 이러한 경향이 LSTM, CNN 등 딥러닝 모델에 기반한 주가 예측에도 동일하게 나타날 것인지, 모델의 어떤 요인이 이런 결과에 영향을 미칠지를 분석할 예정이다.

참고 문헌

[1] 송치영, “뉴스가 금융시장에 미치는 영향에 관한 연구,” 국제경제연구, vol. 8, no. 3, pp. 1-34, 2002.

[2] F. Chau, R. Deesomsak, and M. C. K. Lau, “Investor sentiment and feedback trading: Evidence from the exchange-traded fund markets,” Int. Rev. Financ. Anal., vol. 20, no. 5, pp. 292-305, Oct. 2011, doi: 10.1016/j.irfa.2011.06.006.

[3] “OpenAI o3 and o4-mini System Card.” Accessed: Apr. 29, 2025. [Online]. Available: <https://openai.com/index/o3-o4-mini-system-card/>

[4] “Gemini 2.5: Our most intelligent AI model,” Google. Accessed: Apr. 29, 2025. [Online]. Available: <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/>

[5] K. W. M., S. Allagi, and M. Laddi, “A Survey of Stock Market Prediction - Based on Machine Learning Techniques,” in 2024 Global Conference on Communications and Information Technologies (GCCIT), Oct. 2024, pp. 1-5. doi: 10.1109/GCCIT63234.2024.10862395.

[6] L. Bojić et al., “Comparing large Language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm,” Sci. Rep., vol. 15, no. 1, p. 11477, Apr. 2025, doi: 10.1038/s41598-025-96508-3.

[7] 강두원, 유소엽, 이하영, and 정옥란, “뉴스 감성 분석을 이용한 딥러닝 기반 주가 예측에 대한 연구,” 한국컴퓨터정보학회논문지, vol. 27, no. 8, pp. 31-39, Aug. 2022.

[8] 류의림, 이기용, and 정연돈, “호가창과 뉴스 헤드라인을 이용한 딥러닝 기반 주가 변동 예측 기법,” 한국전자거래학회지, vol. 27, no. 1, pp. 63-79, Feb. 2022.

[9] “대한상공회의소, ‘4분기 경기도 어려울 것’ - 위키뉴스.” Accessed: Apr. 30, 2025. [Online]. Available: [https://ko.wikinews.org/wiki/대한상공회의소\\_4분기\\_경기도\\_어려울\\_것](https://ko.wikinews.org/wiki/대한상공회의소_4분기_경기도_어려울_것)

[10] “Beautiful Soup Documentation - Beautiful Soup 4.13.0 documentation.” Accessed: May 02, 2025. [Online]. Available: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

[11] “scikit-learn: machine learning in Python - scikit-learn 1.6.1 documentation.” Accessed: Apr. 30, 2025. [Online]. Available: <https://scikit-learn.org/stable/index.html>

[12] “Welcome to skforecast - Skforecast Docs.” Accessed: Apr. 30, 2025. [Online]. Available: <https://skforecast.org/0.15.1/index.html>

[13] “Optuna - A hyperparameter optimization framework,” Optuna. Accessed: May 02, 2025. [Online]. Available: <https://optuna.org/>