



Sentistock 포트폴리오

# 백테스트



뉴스, 차트, 수많은 지표 속에서 “지금 무엇을 믿어야 할지” 막막했던 순간들.

우리는 그 질문에 답하기 위해, 우리 서비스를 기반으로 수행된 백테스트 결과를 데이터 중심의 포트폴리오 리포트로 담아냈습니다.

Explore Now



# Index

## 1 백테스트 목적

sentistock 서비스 검증

## 2 감정분석 데이터 활용 방법

KR-FinBERT와 감정지수 산출식을 활용한 감정분석

## 3 상관관계 분석

감정분석 결과와 실제 주가 간 상관관계 분석

## 4 마무리

결론 및 시사점

# 백테스트 목적

## 「뉴스 감성 분석을 이용한 딥러닝 기반 주가 예측에 대한 연구」

강두원, 유소엽, 이하영 and 정옥란. (2022). 뉴스 감성 분석을 이용한 딥러닝 기반 주가 예측에 대한 연구. 한국컴퓨터정보학회논문지, 27(8)

기업 이익의 근본은 상품을 제공하는 대상인 소비자이고, 기업의 흥망성쇠는 소비자들의 기업에 대한 인식에 따라 결정되며, 이를 통하여 기업의 이익에 민감하게 반응하는 주식 시장에서의 기업의 가치를 결정할 수 있다는 것이다.

..하지만 주식 시장은 매우 복잡하고 다양한 정치, 경제적 요인, 정권 교체, 투자 심리 등 다양한 요인에 따라 변동성이 커서 예측이 어렵다. 효율적 시장 가설(Efficient Market Hypothesis)은 자산 가격이 금융 뉴스 기사, 소셜 미디어, 블로그 등과 같은 새로운 정보에 반응한다.

-p.33

뉴스는 금융 시장 가격에 영향을 미치는 요소 중 하나이다. ...기업과 관련된 뉴스 기사와 주가의 변동성 사이에 가한 상관관계가 있음이 드러났다.

-p.33

감성 지표의 분석은 외부 데이터로 수집한 뉴스 데이터를 사용한다. 뉴스 데이터는 기사의 제목과 본문 등으로 이루어져 있다. 경제 분야의 기사들은 기업에 대한 단순한 정보를 포함하기도 하지만 특정 기업에 대한 대중들의 인식과 같은 감정적인 정보를 포함하고 있다.

-p.34

# 감정을 ‘배제’가 아니라 ‘활용’으로

선행연구를 통해 투자자들은 기업 관련 뉴스 정보에 크게 영향을 받는다는 점을 확인했습니다.

사람들은 뉴스 기사를 읽으며 기업에 대한 감정을 형성하고, 그 감정을 주식 투자 판단에 활용합니다.

하지만 대부분의 투자 서비스는 감정을 배제하고 객관적인 지표 중심의 의사결정을 돋는 데 초점을 두고 있습니다.

우리는 이와 반대로, 감정을 제거할 대상이 아니라 의미 있는 신호로 볼 수 있다는 관점에서 출발했습니다.

그래서 뉴스에서 발생하는 감정을 데이터로 정량화해 투자 판단을 보조하는 방식으로 활용합니다.

⋮

즉, 감정 데이터는 우리의 분석과 서비스 설계를 관통하는 핵심 데이터입니다.

# 감정분석 데이터 활용 방법

## KR-FinBERT

KR-FinBERT에서 받은  
감정별 확신 확률값(score)를  
저희의 감정 수치 변환식으로  
데이터를 정교화합니다

## 감정 수치 변환식

### KR-FinBERT 모델이란?



서울대 연구실에서 만든 모델로, 금융 분야의 텍스트 처리에  
특화되도록 만들어진 한국어 특화 자연어 처리 모델입니다.

### KR-FinBERT의 기본 기능

한국어 금융 분야 텍스트를 읽은 다음, 이의 종합적인 센티먼트를  
[positive, neutral, negative]란 하나의 결론으로 정리하는 것이다.  
그리고 이 긍정/중립/부정에 대해서 확신 정도를 확률값으로 제시합니다.

입력된 텍스트에 대해 '감정 분류 결과'(label 항목)와 '확신 확률값'(score  
항목)을 빠르게 도출해서,  
주식 시장 방향성 판단, 종목 간 투자 순위 결정 등 다양한 작업을 할 수  
있게 됩니다.



$$\begin{aligned} \text{base} &= p_{\text{pos}} - p_{\text{neg}} \\ \text{confidence} &= 1.0 - p_{\text{neu}} \\ s &= \text{base} * \text{confidence} \\ \text{raw\_score} &= (s + 1.0) / 2.0 * 100.0 \\ &\text{return max}(0.0, \min(100.0, \\ &\quad \text{raw\_score})) \end{aligned}$$

설명

0~100 점수 계산:

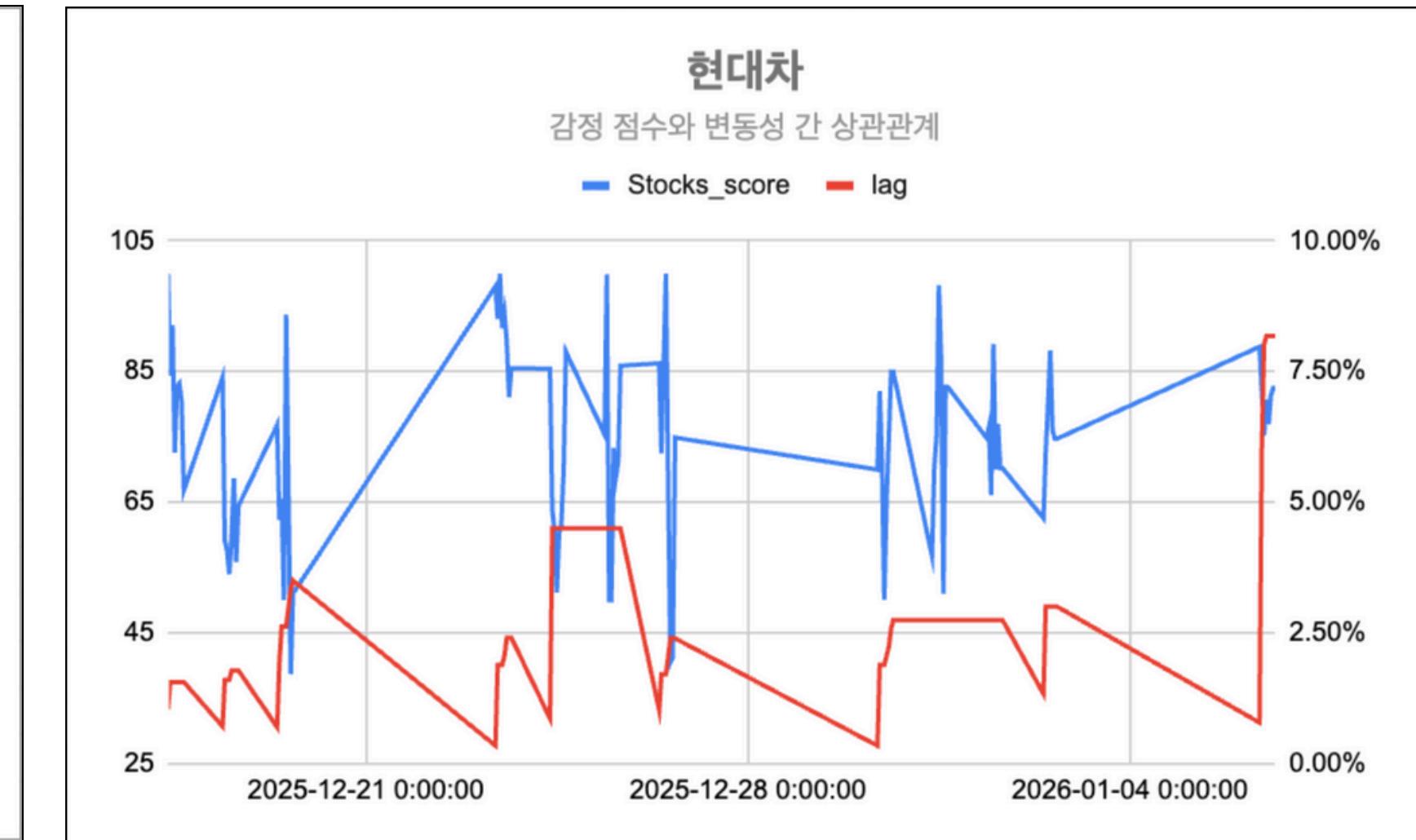
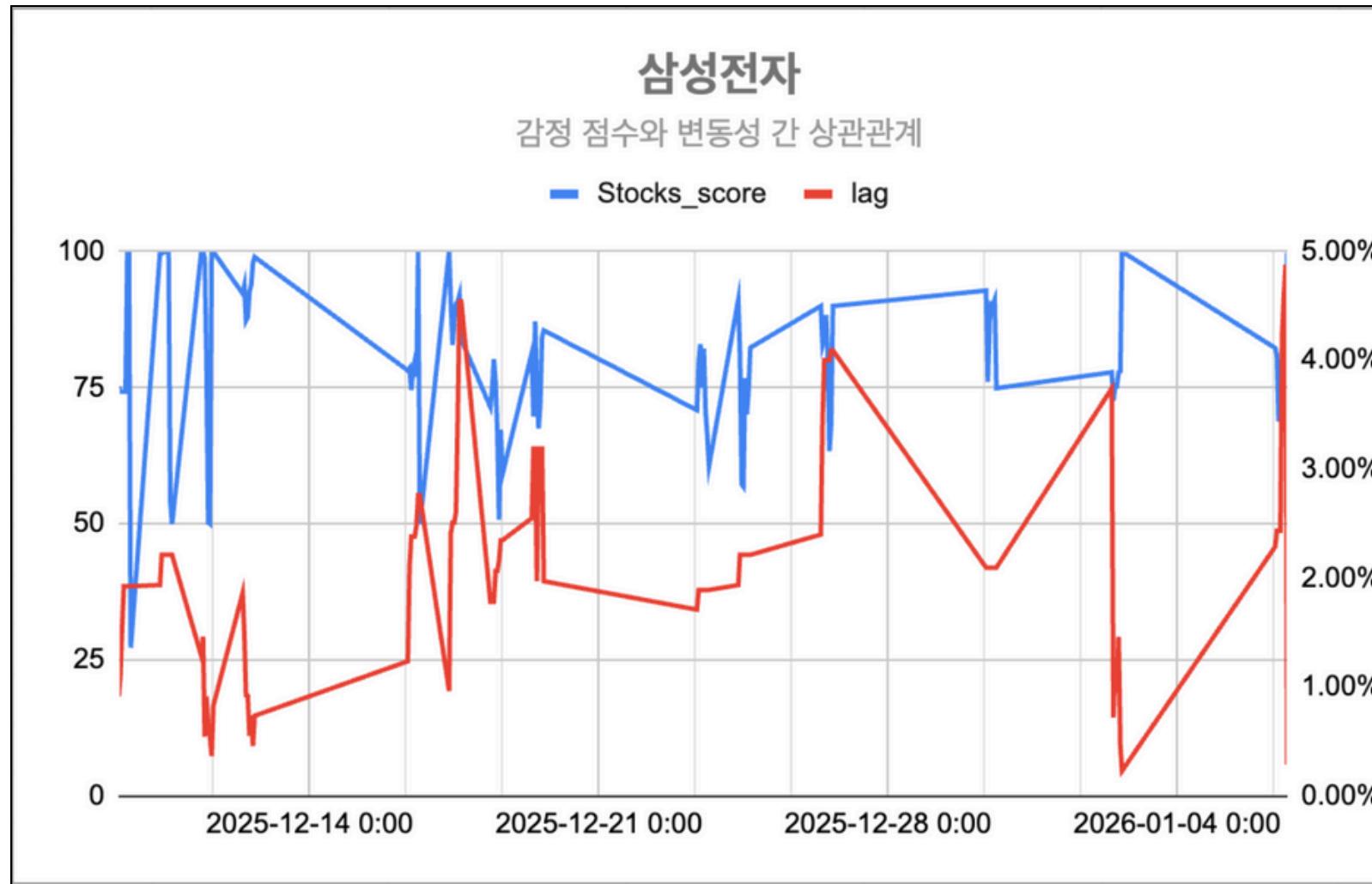
$$\begin{aligned} \text{base} &= \text{긍정} - \text{부정} \\ \text{confidence} &= 1 - \text{중립} \\ S &= \text{base} * \text{confidence} \\ \text{score} &= (S + 1) / 2 * 100 \end{aligned}$$

# 상관관계 분석

## 뉴스 감정 변화와 주가 변동성 흐름 비교

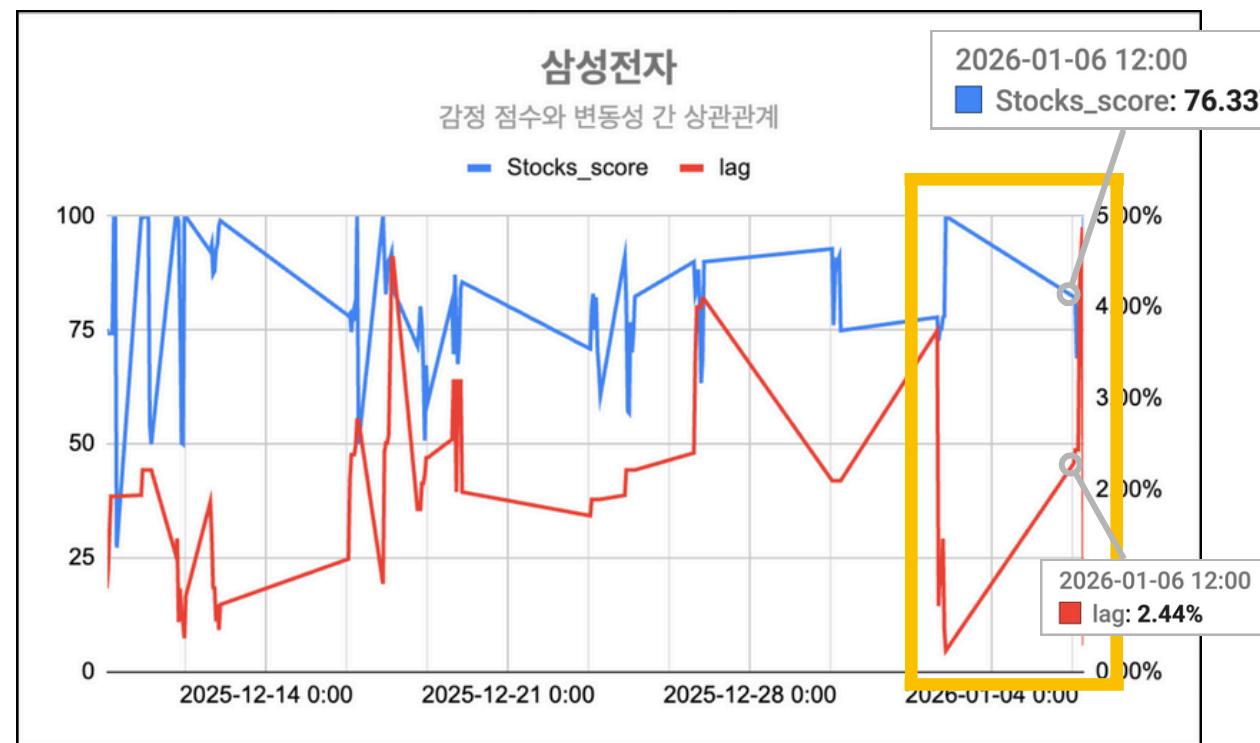
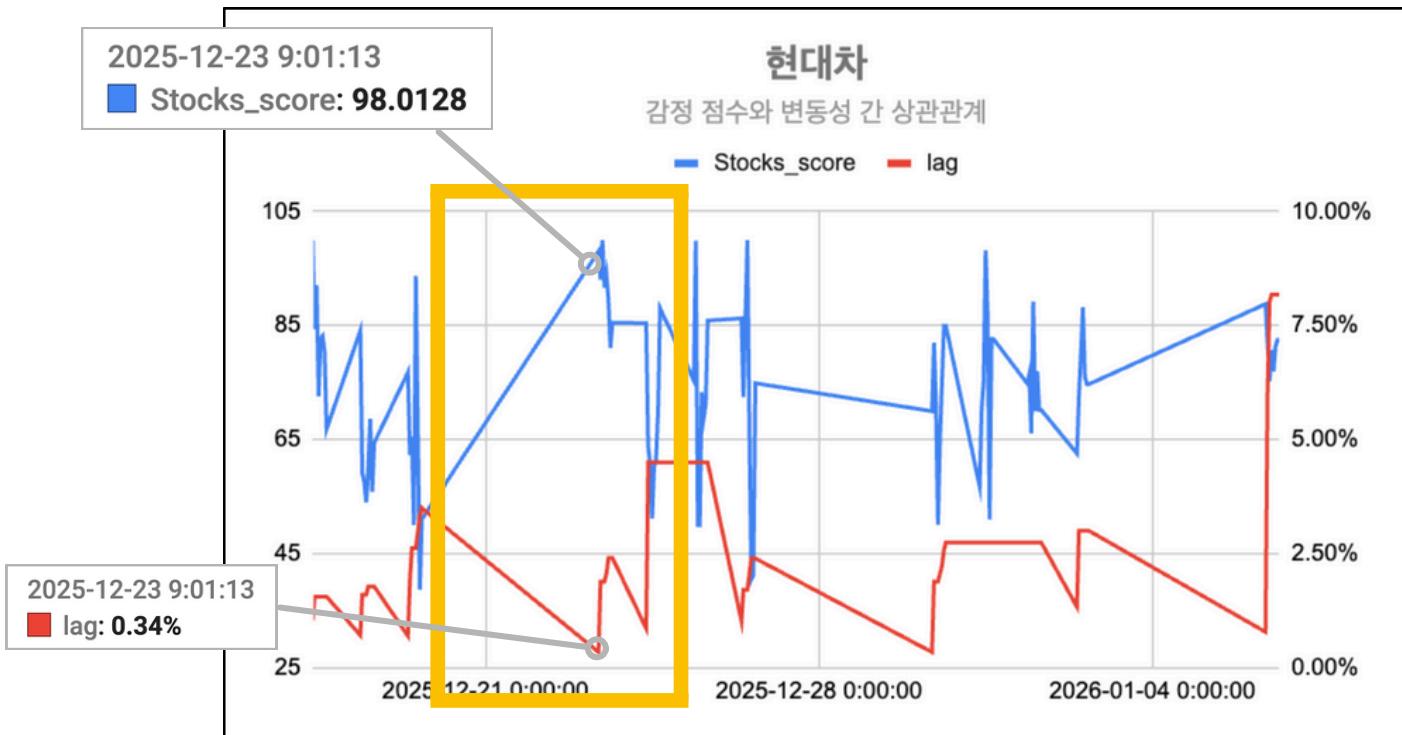
2025년 12월 9일부터 2026년 1월 9일 한달 간 오전 9시 ~ 오후 4시까지

뉴스 기사에서 추출한 감정 점수와 동일 시간대 주가 변동성 흐름을 함께 시각화한 결과입니다.



\* 변동성 = (주식 최고가 - 주식 최저가) / 시가  
 $= (\text{stck_hgpr} - \text{stck_lwpr}) / \text{stck_oprc}$

# 상관관계 분석



## 그래프 분석

특정 구간에서 **감정 점수가 단기간 급격히 상승함에도 불구하고 주가 변동성이 감소하는 현상이 관찰됩니다.**

이는 시장 참여자들의 감정이 분산된 상태에서 특정 방향으로 빠르게 정렬되며, **불확실성이 해소된 결과로 해석할 수 있습니다.**

즉, 감정 점수 상승이 항상 변동성 확대로 이어지는 것은 아니며, 감정이 한 방향으로 수렴할 경우 오히려 변동성은 감소하는 양상을 보입니다.

## 감정 점수와 변동성 관계

**감정 점수는 시장 분위기가 어느 쪽으로 얼마나 기울어 있는지를 나타내고, 변동성은 그 분위기에 대해 투자자들의 판단이 얼마나 서로 엇갈리고 있는지를 보여줍니다.**

| 상태    | 감정     | 변동성 |
|-------|--------|-----|
| 불확실   | 낮거나 혼재 | 높음  |
| 확신 형성 | 높음     | 낮아짐 |

# 상관관계 분석

## 과열(80-100) 구간은 '좋아 보이는 구간'이 아니라, 실제로 '최악의 급락'이 발생했던 위험 구간

2025년 12월 9일부터 2026년 1월 9일 한달 간 오전 9시 ~ 오후 4시까지  
뉴스 기사에서 추출한 감정 점수와 동일 시간대 주가 변동성 흐름을 함께 시각화한 결과입니다.

주식이나 암호화폐 등에서  
'나만 소외될까 봐' 급하게 투자하는  
행동을 유발



\* 감정점수 구간: 0-20, 20-80, 80-100

\*  $dd_5$ : 기준일 이후 5거래일의 최대낙폭(Max Drawdown)이며  
단기 급락 위험을 나타내는 지표  
→ 값이 더 음수일수록 단기 구간에서 더 큰 하락이 발생했음을 의미

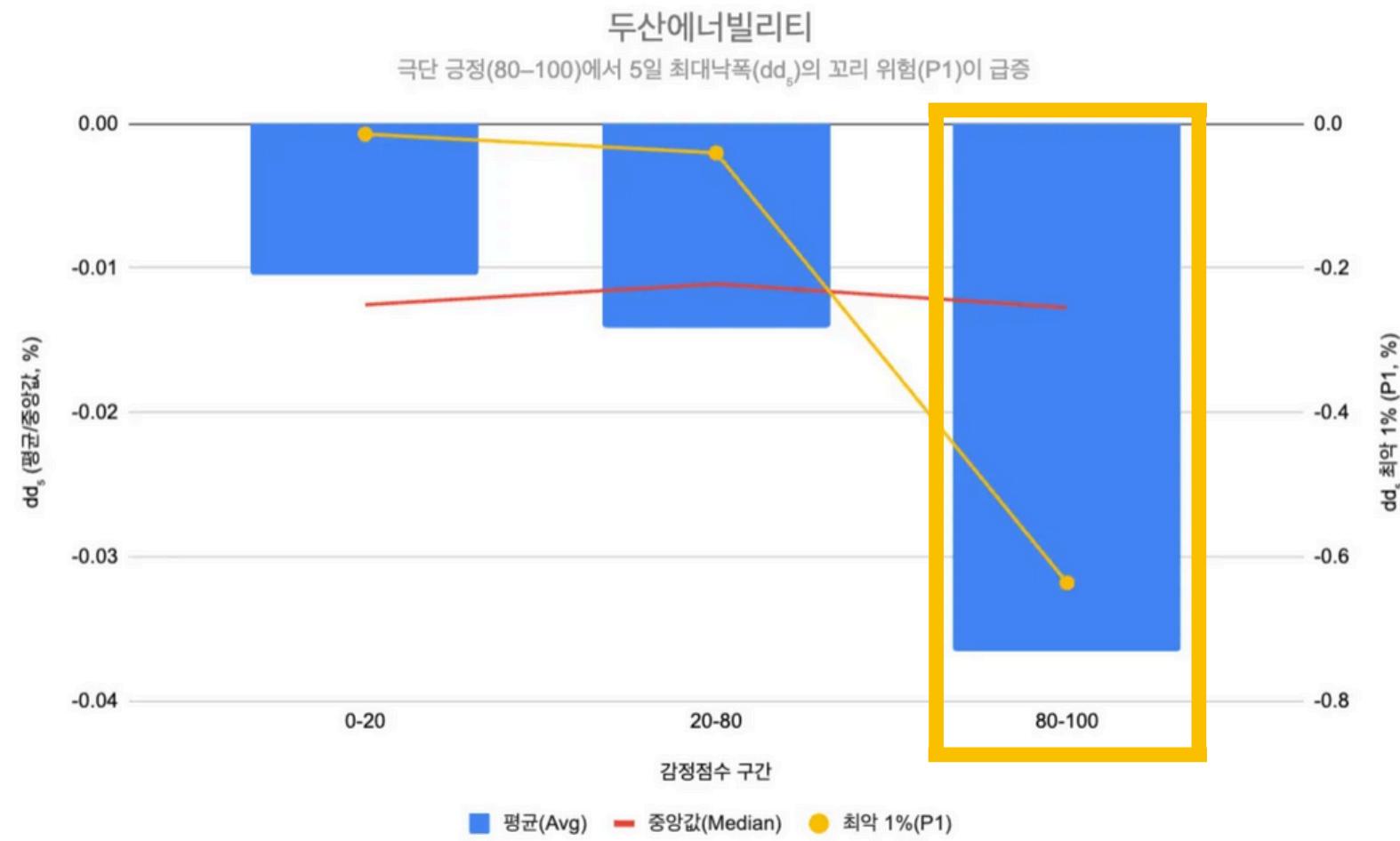
\* 평균(Avg) : 기준일 이후 5거래일 동안의 최대낙폭( $dd_5$ )을  
전체 표본에서 평균낸 값(일반적인 기대 손실 수준)

\* 중앙값(Median) : 기준일 이후 5거래일 최대낙폭( $dd_5$ )의 중앙값  
(극단값 영향이 적은 전형적인 손실 수준)

\* 최악 1%(P1) : 기준일 이후 5거래일 최대낙폭( $dd_5$ )의 하위 1% 분위값  
드물게 발생하는 극단 급락(꼬리 위험)을 대표하는 값

↑  
80-100 구간 → FOMO 유발 가능

# 상관관계 분석



## 그래프 분석

가장 중요한 관측은 감정점수 **80-100 구간에서  $dd_5$ 의 P1(최악 1%)**이 다른 구간 대비 크게 악화되었다는 것입니다. 이는 감정이 과열된 시점에서 항상 위험한 것은 아닐 수 있으나, 급락 시나리오의 규모가 커질 수 있음을 시사하는 결과입니다.

또한 **80-100 구간에서 평균(avg)도 음수 방향으로 악화되는 모습이 관측되고 있는데** 이는 일부 큰 하락 사례가 섞이면서 구간의 평균 위험을 끌어내리는 패턴이 나타난 결과입니다.

서비스의 과열은 감정점수 80-100 구간에서 “지금은 긍정이 극단화되어 **FOMO**가 유발될 수 있는 상태”임을 알려 주고, 사용자가 충동적 의사결정을 하지 않도록 리스크를 환기하는 것이 목적입니다.

**두산에너빌리티** 결과는 감정점수 80-100 과열 구간에서  $dd_5$ 의 최악 1%(P1) 급락이 실제로 관측되었다는 점에서, **과열 구간이 위험 구간이라는 경고의 타당성을 데이터로 뒷받침하는 근거가 되는 사례가 됩니다.**

# 상관관계 분석

감정 점수

99

점

긍정 우세

실제 주가

141,000

원

▲ +1.51%

분석 시간

7

시간

9 ~ 16시

일정 분석 시간 동안

뉴스 기사 기반 감정 점수 평균과 실제 주가 흐름을 비교한 결과,  
감정 변화에 따라 주가 반응이 달라지는 경향이 확인되었습니다.

긍정 감정 구간에서의 주가 상승 비중이 상대적으로 높게 나타났으며,  
부정 감정 구간에서는 주가 하락 또는 정체 구간이 더 자주 관찰되었습니다.

특히 감정 점수가 단기간에 상승한 시점 이후 주가는 동일 또는  
다음 시간대에서 반등 혹은 상승 흐름을 보이는 경우가 많았습니다.



이는 뉴스 기반 감정 지표가 시장 참여자의  
단기 심리 변화를 반영하는 신호로 작용할 수 있음을 시사합니다.

# 마무리

## 결론 및 시사점

뉴스 기사에서 추출한 감정 지표는 특정 시점의 주가 변동을 예측하기보다는 시장 심리의 변화 방향을 정량적으로 이해하는 보조 지표로 활용될 수 있음을 확인했습니다.

## 서비스 활용 관점에서의 의미

감정 데이터는 투자 결정 대체 지표가 아닌 시장 흐름을 해석하는 인사이트를 제공합니다. 복잡한 지표가 아닌 직관적인 감정 기반 신호 제공에 초점을 둡니다.

**본 백테스트 결과는 감정 데이터가  
단기 시장 흐름과 상관관계를 가질 가능성을 시사합니다.**



# SENTISTOCK

## 백테스트



본 프레젠테이션에 대해 문의할 점이 있으시면 언제든 문의주세요.



<https://sentistock.store>