

실적에서 기대감으로

: AI 반도체 주가 급등이 원인 및 프레임 전이 분석



21010430 수학통계학부 수학전공 고은경

21010469 수학통계학부 응용통계전공 임주영

목 차

1. 서론	4
1.1 연구 배경	4
1.2 선행 연구	5
1.2.1 선행 연구 분석	5
1.2.2 본 연구의 차별성	5
2. 연구 설계	6
2.1 연구 문제 및 가설 설정	6
2.2 연구 흐름도	7
2.3 데이터	7
2.3.1 데이터 수집	7
2.3.2 데이터 전처리	8
2.4 방법론	8
2.4.1 텍스트 분석	8
2.4.2 통계 분석	9
2.4.3 인과 분석	10

목 차

3. 연구 결과	10
3.1 텍스트 분석	10
3.1.1 토픽 모델링	10
3.1.2 감성 분석	13
3.2 통계 분석	19
3.2.1 차이 검정	19
3.2.2 상관관계 분석	21
3.3 인과 분석	24
3.3.1 그레인저 인과 분석	24
4. 결론	26
4.1 연구 요약 및 결론	26
4.2 연구의 시사점	27
4.2.1 학술적 시사점	27
4.2.2 실무적 시사점	27
4.3 한계점 및 향후 연구 방향 제안	28
4.3.1 연구의 한계점	28
4.3.2 향후 연구 방향 제안	29
Reference	30

Abstract

본 연구는 2025년 AI 반도체 섹터의 주가 급등기에서 관찰된 기업 실적과 주가의 괴리 (Decoupling) 현상을 규명하고, 시장을 주도하는 정보 프레임의 구조적 전이를 실증적으로 분석하였다. 이를 위해 뉴스 미디어(공급자)와 유튜브 댓글(수요자) 데이터를 수집하고, BERTopic, KR-FinBERT, 사전 기반 분석을 결합한 텍스트 마이닝 방법론을 적용하였다.

연구 결과, 첫째, 주가 급등 시기에 미디어의 지배적 프레임이 ‘확정된 실적’에서 ‘미래의 기대감’으로 전이되는 구조적 변화가 확인되었다. 둘째, 뉴스 보도량의 폭발적 증가에도 불구하고 텍스트 감성 점수는 하락하거나 중립화되는 ‘감성의 역설’ 현상이 관측되어, 주가 상승이 정보의 질보다 양과 화제성에 의해 견인되었음을 입증하였다. 셋째, 개인 투자자들은 심층적으로는 불안을 느끼면서도, 행동적으로는 탐욕적 언어를 표출하는 ‘인지부조화’ 상태에서 추격 매수에 동참한 것으로 나타났다. 마지막으로 그레인저 인과검정을 통해 ‘기대감’ 정보가 실제 주가 수익률을 약 2~3일 선행함을 통계적으로 증명하였다. 본 연구는 텍스트 데이터를 통해 내러티브가 유동성 장세를 견인하는 핵심 원인임을 정량화하고, 투자자 행동 심리를 입체적으로 규명했다는 점에서 의의를 갖는다.

주제어: AI 반도체, 프레임 전이, 감성의 역설, 텍스트 마이닝, 내러티브 경제학, 행동 재무학

1. 서론

1.1 연구 배경

최근 국내 증시(KOSPI), 특히 AI 반도체 섹터의 급격한 상승 랠리 과정에서 기업의 재무적 성과와 주가 흐름이 엇갈리는 디커플링(Decoupling) 현상이 심화되고 있다. 호실적 발표에도 주가가 하락하는 뉴스에 팔고, 반대로 실체가 없는 기대감에는 폭등하는 소문에 사는 이례적인 현상은 주가 동력이 객관적 실적에서 막연한 기대감으로 구조적으로 이동했음을 시사한다. 이와 같은 비이성적 시장 과열은 효율적 시장 가설에 기반한 기존의 재무 데이터 중심 분석만으로는 설명하는 데 한계가 있다. 이에 본 연구는 단순한 등락 분석을 넘어, 미디어와 대중의 인식 프레임 전이를 정량적으로 포착함으로써 2025년 반도체 랠리의 실체와 급등 원인을 규명하고자 한다.

본 연구는 AI 반도체 섹터의 주가 급등 현상을 입체적으로 분석하기 위해 다음과 같은 세 가지 구체적인 연구 목적을 갖는다.

1) 주가와 실적의 괴리(Decoupling)

기존 이론으로 설명하기 어려운 비이성적 시장 과열 현상을 분석한다. 특히, 실적 호조와 무관하게 주가가 변동하는 시장 메커니즘을 정량적으로 입증하여, 재무 데이터와 주가 간의 괴리 원인을 규명한다.

2) 시장 주도 프레임의 구조적 전이 포착

주가 상승의 핵심 동력이 과거의 확정된 실적에서 미래의 불확실한 기대감으로 이동하는 구조적 변화 과정을 시계열적으로 추적한다. 이를 통해 시장 참여자들이 실제 어떤 정보 프레임에 반응하여 매수 의사결정을 내리는지 파악한다.

3) 정보와 주가 간의 시차 및 인과관계 검증

뉴스와 유튜브 댓글 분석을 통해 시장의 감성 역설과 탐욕을 지표화한다. 나아가 특정 정보가 실제 주가 수익률을 며칠 선행하는지 시차 효과를 분석함으로써 정보의 예측력을 검증한다.

1.2 선행 연구

1.2.1 선행 연구 분석 및 한계점

본 연구는 주가 급등의 원인을 규명하기 위해 기존의 주가 결정 이론을 편더멘털 중심 연구와 감성 분석 중심 연구로 나누어 고찰하고, 그 한계점을 분석하였다.

[1세대] 재무 데이터 중심 연구

- 핵심 연구: An Empirical Evaluation of Accounting Income Numbers (Ball & Brown, 1968)
- 주요 내용: 기업의 이익정보가 주가에 미치는 영향을 실증 분석하여, 어닝 서프라이즈가 주가의 즉각적인 상승을 견인한다는 실적 기반 모형을 확립한다.
- 한계점: 시장이 효율적이라는 가정하에 숫자에만 집중함으로써, 최근 시장에서 나타나는 스토리텔링 중심의 주가 급등이나 투기적 과열 현상을 설명하는 데 한계가 있다.

[2세대] 단순 텍스트 마이닝 연구

- 핵심 연구: Twitter mood predicts the stock market (Bollen et al, 2011)
- 주요 내용: 텍스트 마이닝을 통해 대중의 감정을 6가지 차원으로 분류하고, 이러한 집단 심리가 다우존스 산업평균지수 등 주식 시장의 등락을 유의미하게 선행함을 증명한다.
- 한계점: 단순한 긍정 및 부정 단어 빈도 분석에 의존하여, 감정 변화의 원인이 실적 때문인지 단순 루머 때문인지 구체적인 맥락을 식별하지 못한다.

기존 연구들은 재무적 편더멘털과 투자자 심리를 분리하여 접근하는 경향이 있다. 이에 따라 본 연구는 텍스트의 '맥락'을 파악하여 실적에서 테마로 이동하는 프레임 전이를 포착하고, 주가 변동의 인과성을 통합적으로 규명하고자 한다.

1.2.2 본 연구의 차별성

본 연구는 기존 선행 연구들이 가진 방법론적 한계를 극복하고, 주가 급등의 원인을 보다 입체적이고 인과적으로 규명했다는 점에서 다음과 같은 세 가지 차별성을 갖는다.

첫째, 금융 도메인에 특화된 하이브리드 텍스트 분석 방법론을 적용하여 분석의 정밀도를 극대화하였다. 기존 연구들은 단순 키워드 빈도수 분석에 그치거나, 문맥을 파악하지 못하는 구형 모델인 LDA에 의존하여 토픽 도출의 정확도가 낮다는 한계가 있었다. 이에 반해 본 연구는 문맥을 이해하는 BERTopic과 금융 데이터에 특화된 언어 모델인 KR-FinBERT를 결합하여 텍스트 분석의 깊이를 더하였다. 이를 통해 단순한 단어의 나열이 아닌, 시장 내 존재하는 미묘한 뉘앙스와 구체적인 의도를 파악할 수 있었다.

둘째, 주제별 감성 입체 분석을 통해 감성의 역설 현상을 규명하였다. 기존 연구는 뉴스의 주제를 구분하지 않고 전체 시계열을 통합하여 감성을 분석함으로써, 해당 감정이 구체적으로 무엇에 대한 반응인지 식별하기 어렵다는 한계가 존재했다. 본 연구는 이를 보완하여 토픽을 실적과 기대감으로 명확히 분리하고, 각 주제별 감성 점수와 주가 간의 관계를 교차 분석하였다. 이를 통해 전체 감성 점수는 하락함에도 주가는 상승하는 이른바 감성의 역설을 포착해냈다.

셋째, 단순 상관관계를 넘어 프레임 전이의 인과성을 통계적으로 검증하였다. 선행 연구들이 감정과 주가의 단순 동행 여부만을 파악하는 정적인 상관관계 분석에 머물러, 시장의 구조적 변화나 인과적 전이 과정을 설명하지 못했다면, 본 연구는 시계열적 인과 분석을 시도했다는 점에서 차별화된다. 구체적으로 주가 급등 이벤트를 전후로 한 시장 인식의 전환을 Welch's T-test를 통해 통계적으로 검증하고, 그레인저 인과검정을 수행하여 단순한 상관성을 넘어 정보가 주가를 선행하는 확산 과정을 입증하였다.

2. 연구 설계

2.1 연구 문제 및 가설 설정

최근 주가 급등 현상에서 나타난 정보와 주가의 비선형적 관계를 규명하기 위해, 다음과 같은 연구 문제와 가설을 설정하였다. 가설은 크게 미디어 프레임의 구조적 전이(H1), 시장 감성의 역설(H2, H3), 그리고 정보의 인과적 선행성(H4)을 검증하는 데 목적을 둔다.

문제1. 주가 급등 시기에 시장을 지배하는 정보 프레임은 어떻게 변화하는가?

기존의 효율적 시장 가설과 달리, 특정 국면에서 시장의 관심사가 기업의 내재 가치에서 미래의 불확실한 성장성으로 이동하는 현상이 존재하는지 묻는다.

- H1. 주가 급등 이벤트 발생 후, 미디어 내 실적 관련 토픽의 비중은 감소하고 기대감 관련 토픽의 비중은 유의미하게 증가할 것이다.

문제2. 펀더멘털과 무관한 주가 급등은 어떤 감성적 특징과 인과관계를 갖는가?

정보의 양적 증가가 질적 긍정으로 이어지는지, 그리고 실제 주가를 견인하는 정보의 원천이 무엇인지 규명한다.

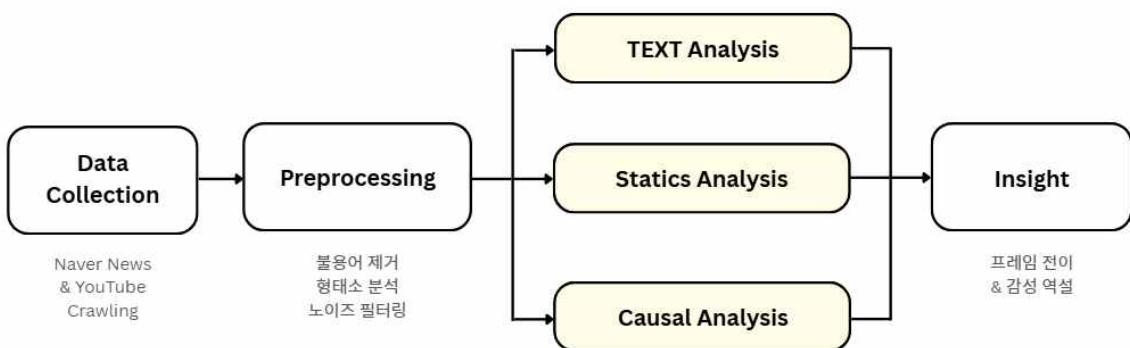
- H2. 주가 급등기에는 뉴스 보도량이 증가함에도 불구하고, 텍스트의 평균 감성 점수는 하

락하거나 중립화되는 감성 곤리 현상이 나타날 것이다.

- H3. 뉴스의 감성과 달리, 개인 투자자의 탐욕 키워드 빈도는 주가 상승 추세와 유의미한 양의 상관관계를 보일 것이다.
- H4. 실적 정보는 주가 수익률을 선행하지 못하는 반면, 기대감 정보는 유의미한 시차를 두고 주가 수익률을 그레인저 인과 할 것이다.

2.2 연구 흐름도

본 연구의 흐름은 다음 도식표와 같다.



2.3 데이터

본 연구는 2025년 반도체 섹터의 주가 급등 현상을 입체적으로 분석하기 위해, 시장 정보 공급자인 뉴스 미디어와 정보 수용자인 개인 투자자의 데이터를 이원화하여 수집하였다.

2.3.1 데이터 수집

분석 기간은 2025년 9월 26일부터 11월 6일까지(총 6주)로 설정하였다. 이 기간은 SK하이닉스와 삼성전자를 중심으로 반도체 섹터의 급격한 랠리가 관측된 시기이다. 연구의 객관성을 확보하기 위해, 반도체 대표 종목인 삼성전자와 SK하이닉스의 종가 표준화(Z-score) 합산 점수가 임계치 3.0을 초과하는 10월 16일을 급등 이벤트 기준일로 정의하고, 이를 기준으로 전후 3주를 수집 범위로 확정하였다.

1) 뉴스 데이터

언론의 이성적 판단 근거를 파악하기 위해 네이버 금융의 주요 기사를 수집하였다. Python 기반의 웹 크롤링을 통해 반도체 관련 기사 약 12,000건을 확보하였으며, 이는 시장에 공급되는 공식적 정보를 대변한다.

2) 유튜브 댓글 데이터

개인 투자자의 투기적 심리를 포착하기 위해 주요 경제 유튜브 채널의 댓글을 수집하였다. YouTube API를 활용하여 약 5,000건의 댓글을 추출하였으며, 이는 시장 내 잠재된 투기적 심리를 측정하는 대리 변수로 활용된다.

2.3.2 데이터 전처리

수집된 비정형 텍스트의 품질을 확보하고 분석 정밀도를 높이기 위해, 데이터 소스별 특성에 맞춘 차별화된 전처리를 수행하였다.

1) 뉴스 데이터 정제

기사 제목, 작성 일시, 본문을 수집하였으며, 송출 과정의 노이즈 제거를 위해 다음 과정을 거쳤다.

- 중복 기사 제거: 재송출 기사를 제거하여 특정 키워드의 과대평가를 방지하였다.
- 불필요한 정보 삭제: 분석의 초점을 내용에 맞추기 위해 언론사명, 기자명, 이메일 등 불필요한 정보를 패턴 매칭으로 삭제하였다.

2) 유튜브 댓글 데이터 정제

댓글 본문, 작성 일시, 좋아요 수 등을 수집하였다. 특히, 좋아요 수는 공감도 가중치로 활용된다. 구어체와 은어가 많은 특성을 고려하여 다음과 같이 처리하였다.

- 투자 은어 정규화: ‘가즈아’, ‘풀매수’ 등 투자 은어를 식별 가능한 표준어로 변환하거나 감성 사전에 등재하여 유의미한 변수로 활용하였다.
- 스팸 및 노이즈 필터링: 투자 심리와 무관한 광고성 댓글이나 자동 생성된 도배성 봇 댓글을 필터링하여 데이터의 순도를 높였다.

2.4 분석 방법론

본 연구는 주가 급등 이면의 비정형 텍스트 정보를 정량화하고, 이것이 실제 시장 데이터와 맺는 인과관계를 규명하기 위해 텍스트 마이닝, 통계적 검정, 인과 분석으로 구성된 다층적 분석 프레임워크를 적용하였다.

2.4.1 텍스트 분석

비정형 데이터인 뉴스 기사와 유튜브 댓글에서 시장의 주요 토픽과 감성을 추출하기 위해, 최신 딥러닝 기법인 BERTopic과 KR-FinBERT, 그리고 사전 기반 분석인 Lexicon-based Analysis을 상호 보완적으로 활용하였다.

1) 토픽 모델링

- **BERTopic:**

시장의 관심사가 실적에서 기대감으로 이동하는 구조적 변화를 정밀하게 포착하기 위해, 기존의 LDA 모델 대신 대규모 언어 모델 기반의 BERTopic을 채택하였다. LDA는 단어의 등장 횟수에 기반하여 문맥을 파악하지 못하고 노이즈 데이터까지 토픽에 포함하는 한계가 있다. 반면, 2020년 Maarten Grootendorst에 의해 제안된 BERTopic은 트랜스포머 임베딩을 통해 단어의 문맥과 의미를 깊이 있게 이해하며, 의미 없는 문서를 노이즈로 분류하여 제거하는 소프트 클러스터링을 지원한다. 이러한 특성은 미세한 뉘앙스 차이가 중요한 금융 텍스트 분석에 최적화된 성능을 제공한다.

2) 감성 분석

- **KR-FinBERT:**

뉴스와 유튜브 텍스트의 감성 분석을 위해 KR-FinBERT 모델을 적용하였다. KR-FinBERT는 서울대학교 NLP 연구실 등에서 2020년대 초반 한국어 금융 도메인에 특화되어 개발한 BERT 기반 언어 모델이다. 일반적인 언어 모델과 달리 금융 시장 특유의 문맥을 학습하여 호재와 악재를 판별하는 데 우수한 성능을 보인다. 본 연구에서는 해당 모델을 활용하여 수집된 뉴스 헤드라인의 감성을 -1(부정)에서 +1(긍정) 사이의 확률 점수로 산출하였으며, 이를 통해 뉴스 텍스트의 감성 점수와 실제 주가 흐름 간의 관계를 정량적으로 측정하였다.

- **Lexicon-based Analysis:**

비속어와 은어가 빈번한 유튜브 댓글 분석을 위해, 딥러닝 모델과 별도로 고전적으로 널리 통용되는 사전 기반 분석을 병행하였다. 사전 기반 분석은 전통적인 기법이나, 기존 모델의 노후화 문제를 보완하고자 본 연구는 2025년 급등 당시 시장에서 통용되던 최신 투자 은어를 반영한 탐욕 및 공포 사전을 자체 구축하여 적용하였다. 탐욕 지수 산출 방법은 ‘매수’, ‘급등’과 같은 탐욕 키워드와 공포 키워드의 단순 출현 빈도를 카운트하여, 전체 댓글 대비 탐욕 어휘의 비율 $\frac{Greed}{Greed + Fear}$ 을 계산하였다. 이는 개인 투자자들의 투기적 과열 심리를 가장 직관적으로 보여주는 대리 변수로 활용된다.

2.4.2 통계 분석

텍스트 마이닝을 통해 도출된 지표(토픽 빈도, 감성 점수)와 시장 데이터(주가 수익률, 변동성) 간의 관계를 검증하기 위해 다음의 두 가지 통계 기법을 적용하였다.

1) 상관관계 분석

두 변수 간 선형적 관계의 강도를 측정하기 위해 피어슨 상관계수를 산출하였다.

- Topic vs Price: 특정 토픽의 빈도 증가가 주가 상승과 양의 상관관계를 갖는지 확인한다.
- Sentiment Paradox: 뉴스 감성 점수는 하락함에도 주가는 상승하는 비동조화 구간을 식별한다.

2) 차이 검정

주가 급등 이벤트를 기점으로 시장의 관심도와 감성이 구조적으로 변화했는지 확정하기 위해 독립표본 T-검정을 실시하였다. 단, 금융 시계열 데이터의 특성상 급등 전후의 데이터 분산이 다를 가능성이 높으므로, 등분산을 가정하는 일반 T-test 대신 Welch's T-test를 채택하여 검정의 오류를 최소화하였다.

- Volume Effect: 이벤트 전후, 시장의 관심도가 유의미하게 폭발했는가?
- Sentiment Shift: 이벤트 전후, 시장의 평균 감성이 긍정 또는 부정 방향으로 유의미하게 이동했는가?

2.4.3 인과 분석

단순한 상관관계는 변수 간의 선후 관계를 설명하지 못한다는 한계가 있다. 이에 본 연구는 정보가 가격을 선행하는지 규명하기 위해 그레인저 인과관계 분석(Granger Causality Test)을 수행하였다.

• 분석 목적

특정 뉴스 토픽(X)의 발생이 실제 주가 수익률(Y)의 변동을 예측하는 데 통계적으로 유의미한 정보를 제공하는지 검증한다. 이를 통해 주가가 상승해서 뉴스가 나온 것인지(후행), 뉴스가 보도되어 주가가 오른 것인지(선행)에 대한 방향성을 판단하고, 정보가 시장에 반영 되기까지 소요되는 시차를 측정한다.

• 검증 절차

- 1단계: 시계열 데이터의 안정성을 확보하기 위해 정상성 검정(Augmented Dickey-Fuller Test)을 선행한다.
- 2단계: 정보 확산의 잠복기를 고려하여 1일에서 5일 간의 시차를 설정하여 인과성을 테스트한다.
- 3단계: 유의수준 0.05 하에서 귀무가설(H_0 : 뉴스 토픽은 주가 수익률의 선행 원인이 아니다)을 기각할 경우, 해당 정보가 주가 변동을 통계적으로 선행한다고 결론 내린다.

3. 연구 결과

3.1 텍스트 분석

본 연구는 주가 급등기를 전후하여 미디어와 시장 참여자가 시장을 바라보는 관점이 어떻게 변화했는지 규명하기 위해, 텍스트 마이닝 기법을 활용하여 주요 토픽과 감성을 분석하였다.

3.1.1 토픽 모델링

1) 주요 토픽 도출 및 키워드 식별

뉴스 데이터에 BERTopic 알고리즘을 적용하여 토픽 모델링을 수행한 결과, 시장을 설명하는 8개의 주요 토픽이 도출되었다. [그림 1]은 각 토픽을 구성하는 핵심 단어와 그 중요도를 나타내는 c-TF-IDF Score를 시각화한 것이다. 분석 결과, 시장의 담론은 크게 ‘시황/랠리’, ‘산업/실적/펀더멘털’, ‘글로벌 거시 정치’, ‘협력 기대감’ 등의 주요 주제로 분류되었다. 그 외 ‘정책’, ‘이벤트’, ‘경영 승계/보상’, ‘노이즈’ 등의 토픽이 부수적으로 식별되었다.

2) 핵심 대립 프레임 식별

도출된 토픽 중, 본 연구는 주가 변동의 핵심 원인이 되는 두 가지 상반된 프레임으로 ‘산업/실적/펀더멘털’과 ‘협력 기대감’ 토픽에 집중하였다. 두 토픽의 키워드 구성은 시장의 관심사가 ‘이성적 판단’에서 ‘투기적 기대’로 이동하는 구조적 변화를 명확히 보여준다.

- 산업/실적/펀더멘털: 이성적 판단의 프레임

해당 토픽은 ‘반도체’, ‘메모리’, ‘분기’, ‘3분기’ 등의 키워드로 구성되어 있다. 이는 기업의 구체적인 실적 수치와 제품 경쟁력에 기반한 것으로, 투자자들의 이성적 펀더멘털 판단을 대변하는 프레임이다.

- 협력 기대감: 투기적 기대의 프레임

반면, 협력 기대감 토픽은 ‘AI’, ‘엔비디아’, ‘젠슨’ 등 특정 인물과 미래 성장성에 대한 스토리텔링이 중심 키워드로 나타났다. 이는 구체적인 실적보다는 AI 생태계 편입에 대한 막연한 기대와 특정 CEO와의 협력 관계를 강조하는 내러티브 중심의 프레임으로, 투자자들의 기대 심리를 자극하는 요인으로 작용한다.



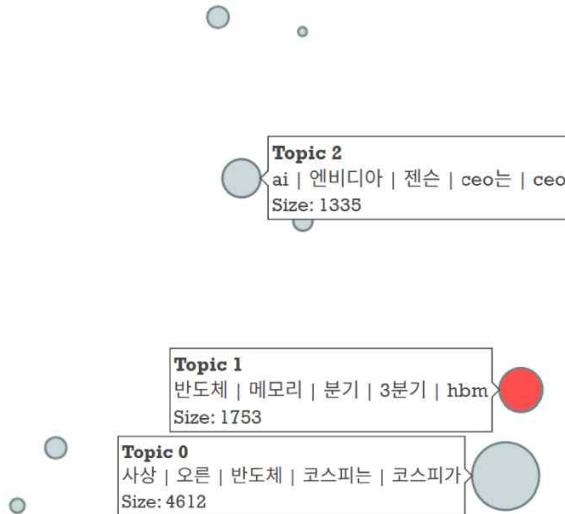
[그림 1] BERTopic 기반 토픽 별 핵심 단어 구성비

이러한 명확한 키워드의 분리는 실제 데이터상에서 ‘확정된 과거(실적)’와 ‘불확실한 미래(기대감)’라는 두 가지 프레임이 뚜렷하게 대립하고 있음을 시사한다. 특히, 주가 급등 구간에서 나타난 이러한 토픽 구조는 시장의 주도권이 숫자에 근거한 펀더멘털에서 이야기에 근거한 모멘텀으로 전이되었음을 뒷받침하는 정량적 증거라 할 수 있다.

3) 토픽 구조 분석

[그림 2]는 BERTopic 기반 Intertopic Distance Map으로, 각 토픽 간의 문맥적 유사도를 2 차원 공간에 시각화한 것이다. 토픽 간의 거리가 가까울수록 의미적으로 연결되어 있음을, 멀 수록 독자적인 맥락을 형성하고 있음을 의미한다. 본 연구는 이 공간적 배치를 통해 시장 정보를 구성하는 ‘시장 전체의 움직임(Beta)’과 ‘초과 수익 창출(Alpha)’을 구조적으로 식별하였다.

Intertopic Distance Map



[그림 2] BERTopic 기반 Intertopic Distance Map

- **시황/랠리(Topic 0)와 편더멘털(Topic 1)의 근접성: Beta**

분석 결과, ‘시황/랠리’와 ‘산업/실적/편더멘털’은 매우 근접한 위치에 분포하고 있다. 이는 텍스트 데이터 상에서 시장 전반의 상승을 논할 때 반도체 실적이 필연적으로 함께 언급됨을 의미하며, 두 주제 간의 문맥적 유사도가 높음을 시사한다. 실적 뉴스는 주가를 지탱하는 필요조건이지만, 시장 전체의 흐름(Beta)과 차별화되지 않는 일반적인 정보로 작용 한다. 즉, 투자자들에게 실적 정보는 이미 시장 가격에 반영된 상수로 인식되어, 단기 급등을 유발하는 폭발력은 부족한 것으로 해석된다.

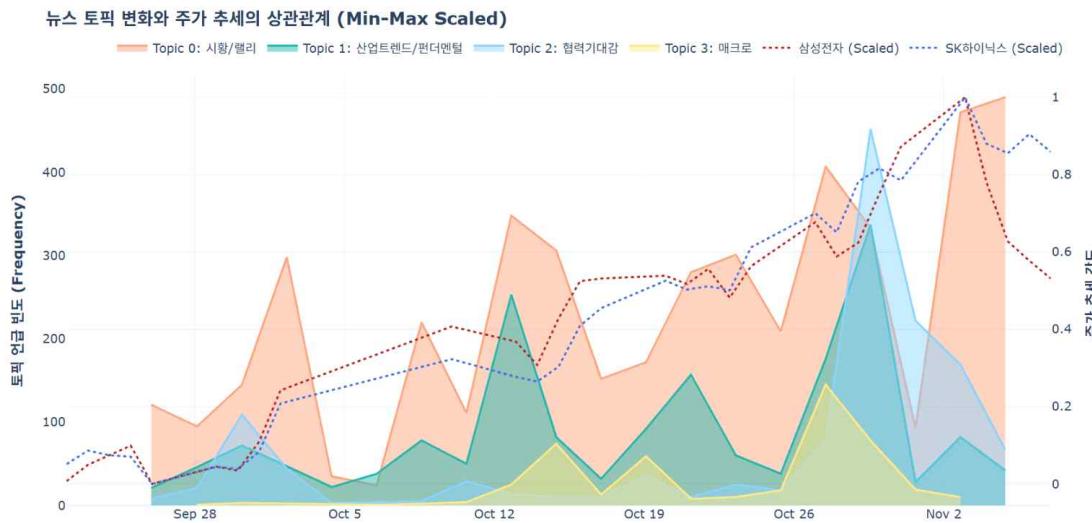
- **협력 기대감(Topic 2)의 독립성: Alpha**

‘협력 기대감’은 ‘시황/랠리’ 및 ‘편더멘털’과 독립적인 위치에 존재한다. ‘AI’, ‘엔비디아’, ‘젠슨 황’ 등의 키워드는 기존의 국내 증시 시황이나 일반적인 제조업 실적과는 문맥적으로 섞이지 않는 별도의 구조를 형성하고 있다. 투자자들에게 협력 이슈는 실적 정보와는 구별되는 독자적인 모멘텀으로 작용하였다. 이는 개별 종목의 초과 수익(Alpha)을 창출하는 충분조건이 되었으며, 정보의 희소성과 신규성이 투자자들의 기대 심리를 자극하여 주가 급등이라는 비이성적 과열을 견인한 핵심 원인으로 분석된다.

4) 시계열 토픽 추세와 주가 상관관계

도출된 토픽의 시계열 빈도 변화와 실제 주가 추세(삼성전자, SK하이닉스)를 중첩 분석하여, 시장의 지배적 프레임이 전환되는 결정적 시점을 포착하였다. [그림 3]은 뉴스 토픽의 빈도와

표준화된 주가 추세를 시각화한 것으로, 크게 두 가지 국면으로 구분된다.



[그림 3] 뉴스 토픽 변화와 주가 추세 비교

• 실적 중심의 관망세 (10월 초 ~ 10월 중순)

그래프의 초반부인 10월 초부터 중순까지는 ‘산업/실적/펀더멘털’이 큰 비중을 차지하였습니다. 이 시기 삼성전자와 SK하이닉스의 주가는 횡보하거나 완만한 등락을 반복하는 제한적인 흐름을 보였다. 이는 투자자들이 3분기 확정 실적 발표를 기다리며 기업의 펀더멘털을 확인하려는 관망 심리가 시장을 지배했음을 의미한다. 즉, 실적 관련 뉴스가 다수 보도되었으나, 이것이 즉각적인 매수세나 주가 급등으로 이어지지는 않는 이성적 탐색 구간으로 정의할 수 있다.

• 프레임의 교차와 주가 급등 (10월 19일 이후)

10월 19일을 기점으로 시장의 정보 구조는 급격한 변곡점을 맞이하였다. 실적 토픽의 비중이 급감하는 동시에, ‘협력 기대감’의 비중이 폭발적으로 증가하는 프레임 골든크로스 현상이 발생하였다. 해당 시점부터 주가 추세선은 협력 기대감 토픽의 상승 곡선과 동조화되며 수직 상승하는 양상을 보였다. 이는 주가 급등의 핵심 동력이 펀더멘털이 아닌, AI 협력 및 앤비디아 벤류체인 편입이라는 재료의 확산에 있었음을 시사한다. 즉, 시장 참여자들의 의사결정 기준이 실적에서 기대감으로 완전히 전이되었으며, 이러한 내러티브의 확산이 실제 유동성을 자극하여 랠리를 견인했음을 시계열적으로 입증하였다.

3.1.2 감성 분석

1) 뉴스 헤드라인 감성 분석

금융 특화 언어 모델인 KR-FinBERT를 활용하여 뉴스 헤드라인의 감성 점수를 산출하고, 각 토픽별 시장의 심리적 반응을 정량화하였다. [그림 4]는 도출된 주요 토픽별 평균 감성 점수를 시각화한 것이다.

토픽별 평균 감성 점수 (Market Sentiment by Topic)



[그림 4] 토픽별 평균 감성 점수

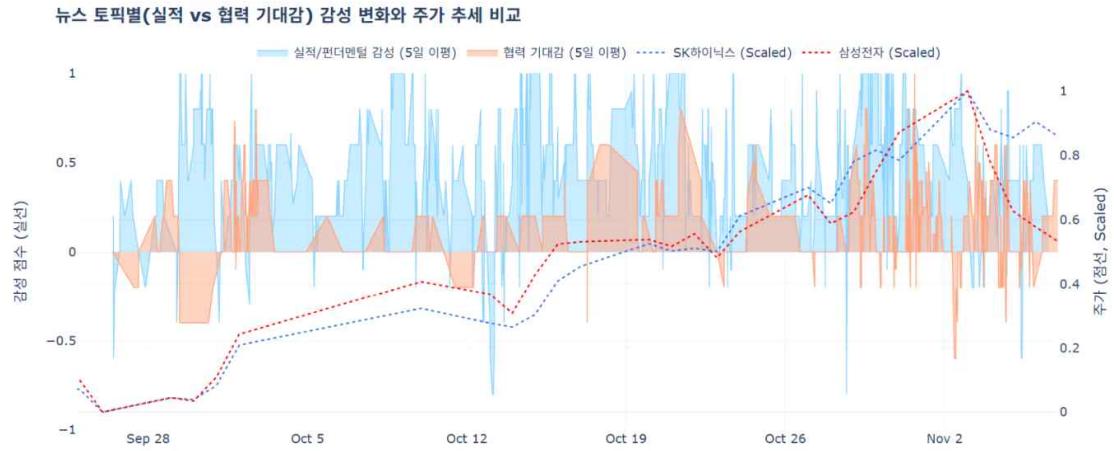
분석 결과, ‘산업/실적/펀더멘털’은 0.556의 매우 높은 긍정 점수를 기록하였다. 반면, 주가 급등의 주재료였던 ‘협력 기대감’은 0.156으로, 상대적으로 낮은 약한 긍정 또는 중립 수준에 머물렀다. 일반적인 상식으로는 뉴스의 감성이 긍정적일수록 주가 상승폭이 커야 한다. 그러나 감성 점수가 낮은 기대감 토픽이 주가 폭등을 견인하는 감성의 역설 현상이 관측되었다. 이에 대한 금융학적 해석은 다음과 같다.

- 확정된 호재의 선반영: ‘실적’의 높은 감성 점수는 이미 확인된 과거의 성과에 기반한다. 효율적 시장 가설에 따르면, 이러한 확정된 정보는 즉시 가격에 반영되어 추가적인 상승 모멘텀을 제공하는 데 한계가 있다. 즉, 시장 참여자들에게는 이미 알고 있는 좋은 뉴스로 인식되어 감성적 포화 상태에 이른 것이다.
- 불확실성에 기인한 투기적 반응: ‘협력 기대감’의 낮은 감성 점수는 해당 정보가 아직 실현되지 않은 미래의 기대임을 의미한다. ‘협력 논의’, ‘가능성’ 등은 KR-FinBERT 모델상에서 명확한 긍정으로 분류되지 않고 중립 영역에 위치한다. 역설적이게도 급등장에서는 이처럼 해석이 분분하고 불확실성이 내포된 정보가 투자자들의 상상력과 기대를 자극하여 거래량을 폭발시키는 기제로 작용하였다.

결론적으로 이번 상승장에서 시장은 명확한 사실보다 모호한 기대에 더 민감하게 반응하였다. 이는 투자자들의 의사결정이 기업의 현재 가치가 아닌, 미래의 성장 스토리에 베팅하는 투기적 성향으로 기울어져 있음을 시사한다.

2) 시계열 감성 분석

앞선 토픽별 평균 감성 분석에서 확인된 감성의 역설을 규명하기 위해, 뉴스 토픽별 감성 점수의 시계열 변화와 주가 추세를 비교 분석하였다. [그림 5]는 ‘실적/펀더멘털’과 ‘협력 기대감’의 일별 감성 점수에 5일 이동평균을 적용하여 노이즈 제거 후, 삼성전자 및 SK하이닉스의 표준화된 주가 추세와 중첩하여 시각화한 것이다.



[그림 5] 토픽별 감성 변화와 주가 추세 비교

- **실적/펀더멘털**

파란색 영역으로 표시된 실적/펀더멘털 감성 지수는 분석 기간 전반에 걸쳐 양의 값을 유지하며 상대적으로 높은 수준을 기록하였다. 그러나 시계열적으로 볼 때, 잣은 변동성을 보이며 주가의 단기 급등 패턴과는 뚜렷한 상관관계를 보이지 못했다. 이는 양호한 실적 뉴스가 주가의 급락을 방어하는 기반 역할을 수행했음을 의미한다. 즉, 투자자들에게 실적 호조는 주식을 보유해야 할 기본적 근거는 되었으나, 추가적인 매수를 자극하여 가격을 밀어 올리는 강력한 트리거로는 작용하지 못한 것으로 해석된다.

- **협력 기대감**

주황색 영역인 협력 기대감 감성 지수는 주가가 수직 상승하는 구간(10월 중순 이후)과 시각적으로 명확하게 동행하는 모습을 보인다. 감성 점수의 절대적인 수치는 실적 토픽보다 낮을지라도, 해당 토픽의 감성 변화 추세가 실제 주가의 방향성과 일치한다는 점은 주목할 만하다. 이는 ‘협력’이라는 키워드가 내포한 미래 성장성이 시장에 움직임을 제공했음을 시사한다. 투자자들은 확정된 실적보다는, 미래 비전에 더 즉각적이고 민감하게 반응하였다.

두 토픽의 시계열 감성 패턴을 종합하면, 이번 반도체 랠리의 성격은 ‘실적 뉴스는 주가의 하단을 지지하며 기반을 다졌고, 협력 기대감 뉴스는 상단을 뚫고 주가를 폭등시켰다’고 요약할 수 있다. 이는 주가 형성 과정에서 펀더멘털 정보와 모멘텀 정보가 각기 다른 층위에서 작동하고 있음을 보여주는 실증적 증거이며, 시장의 관심이 ‘실적’에서 ‘기대감’으로 이동했을 때 주가가 폭발적인 상승 탄력을 받는다는 사실을 입증한다.

3) 유튜브 댓글 감성 분석

개인 투자자들의 실제 심리를 파악하기 위해, 유튜브 댓글에 대해 KR-FinBERT 감성 분석을 수행하였다. [그림 6]은 주가의 흐름과 댓글 수, 그리고 감성 점수의 상관관계를 시계열로 나타낸 것이다.

주가 흐름(삼성전자/SK하이닉스) vs 투자자 감정(FinBERT Score)



[그림 6] KR-FinBERT 기반 주가 흐름 vs 투자자 감정 비교

분석 결과, 주가는 지속적으로 상승했음에도 불구하고 대중의 심리는 ‘환호’보다는 ‘불안’과 ‘의심’이 공존하는 복합적인 양상을 보였다. 기간 내 전체적인 탐욕 지수는 0.65로 측정되었는데, 시장이 과열권에 진입했음을 의미하는 약한 탐욕 단계이지만 통상적인 베블 정점에서 관측되는 극단적 탐욕(0.8 이상)에는 미치지 못했다. 이는 개인 투자자들이 주가의 상승 랠리를 즐기면서도, 학습 효과로 인해 언제든 하락장이 올 수 있다는 위기관리 심리를 깔고 있었음을 시사한다.

시계열 국면별 심리 변화는 다음과 같다.

- Phase 1. 관심의 폭발과 초기 상승기 (9월 말 ~ 10월 초)

SK하이닉스의 상승세가 본격화되던 10월 2일, 댓글 수는 378건으로 기간 내 최고치를 기록하였다. 주가 상승 초기에서 시장의 관심도가 폭발적으로 증가하며 대중의 시선이 집중되었다. 이때 감성 지수는 양의 방향성을 보이며, 상승에 대한 기대감이 시장을 주도하는 모습을 보였다.

- Phase 2. 불안한 상승과 심리적 괴리 (10월 중순 ~ 10월 말)

주가는 지속적으로 우상향 곡선을 그렸으나, 10월 14일 감성 지수는 -0.31(부정)까지 급락하는 심리적 저점을 형성하였다. 이는 전형적인 고점 공포 현상으로 해석된다. 주가가 단기간에 급등하자 댓글창에서는 ‘과도한 상승이다’, ‘곧 조정이 올 것이다’라는 부정적 전망과 불안감이 팽배해졌다. 즉, 실제 가격은 오르지만 투자 심리는 냉각되는 괴리현상이 발생하며, 상승장이 불안 속에서 진행되었음을 입증한다.

- Phase 3. 고점 형성 및 차익 실현 충격 (11월 초)

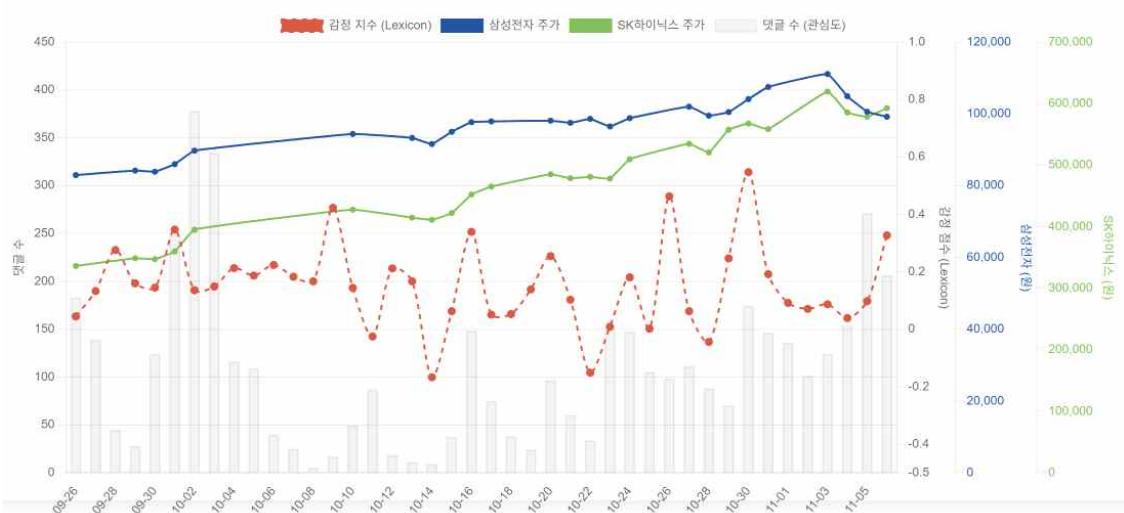
주가가 최고점에 도달했다가 꺾이기 시작한 11월 3일~5일 구간, 댓글 수는 271건으로 다시 급증했으나 감성 지수는 0.005로 사실상 중립에 수렴하였다. 주가가 하락세로 전환되자, 시장 참여자들은 즉각적인 공포를 느끼기보다 차분히 상황을 관망하거나 정보를 공유하는 태도를 보였다. 감성 점수의 중립화는 감정적 동요보다는 팩트 체크와 매도 타이밍을 재는 이성적 탐색이 주를 이루었음을 의미한다.

유튜브 감성 분석 결과, 이번 상승장은 맹목적인 광기에 의한 랠리라기보다는, 불안해하면서도 매우 버튼을 누르는 인지부조화 상태에서 진행된 랠리임이 확인되었다. 이는 앞서 뉴스 분석에서 도출된 기대감 토픽이 개인 투자자들에게 희망을 줌과 동시에, 확신 부족이 불안감을 유발했기 때문으로 풀이된다.

4) 유튜브 댓글 Lexicon 분석

딥러닝 모델(KR-FinBERT)이 포착하기 어려운 개인 투자자들의 직관적인 욕망과 투기 심리를 측정하기 위해, 탐욕/공포 사전을 활용한 Lexicon 기반 분석을 수행하였다. [그림 7]은 주가 흐름과 Lexicon 기반 감성 지수의 추이를 시각화한 것이다.

[그림 7] 주가 흐름(삼성전자/SK하이닉스) vs 투자자 감정(Lexicon Score)



[그림 7] Lexicon 기반 주가 흐름 vs 투자자 감정 비교

분석 기간 동안의 평균 탐욕 지수는 0.68로 ‘약한 탐욕’ 단계로 측정되었다. 흥미로운 점은 앞서 KR-FinBERT 분석에서는 고점에서의 불안이 감지되었으나, 은어와 비속어를 직접 카운트 한 Lexicon 분석에서는 ‘지금이 기회다!', '사야 한다!'와 같은 긍정적이고 공격적인 매우 신호가 압도적이었다는 점이다. 이는 투자자들이 심리적으로는 불안해하면서도, 행동적으로는 ‘가즈아’, ‘풀매수’를 외치며 시장에 참여하는 인지부조화 상태였음을 시사한다.

시계열 국면별 심리 변화는 다음과 같다.

- Phase 1. 논쟁과 관심이 급증 (9월 26일 ~ 10월 5일)

시장에 대한 관심은 폭발했으나, 긍정/부정 단어가 팽팽하게 대립하며 감성 점수는 약한 긍정 수준에서 등락을 거듭했다. 이는 상승 초입 단계에서 투자자들 간의 의견이 엇갈리며 방향성을 탐색하는 간보기 장세로 해석된다. 긍정적인 기대감이 존재했으나, 확신 부족으로 인해 탐욕 점수가 상쇄되는 구간이었다.

- Phase 2. 공포의 바닥 확인 (10월 14일 ~ 10월 22일)

10월 14일 감성 점수가 -0.16까지 하락했으나, 불과 이틀 뒤인 10월 16일 +0.33으로 급 반등하는 패턴(V자)이 관측되었다. 이 시기 공포 키워드는 약 40건 감소한 반면 탐욕 키워드는 유지되었다. 이는 투자자들이 일시적인 주가 조정을 공포가 아닌 저가 매수 기회로

인식했음을 의미한다. 전형적인 놀림목 매수 심리가 작동하며 하락을 대기 매수세 유입의 기회로 활용한 것이다.

- Phase 3. 탐욕의 고점: 축제의 시작 (10월 26일 ~ 10월 31일)

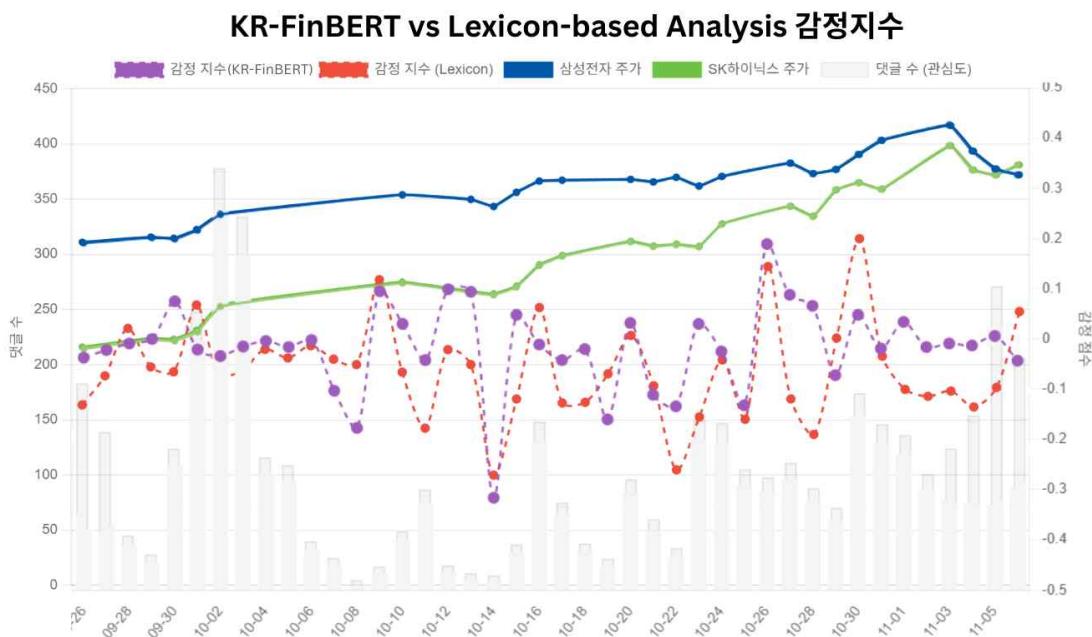
댓글량이 가장 많았던 이 시기에 감성 점수는 +0.547로 기간 내 심리적 최고점을 기록했다. ‘가즈아’, ‘돌파’, ‘상한가’ 등 탐욕 단어의 사용량이 폭증했다. 앞선 KR-FinBERT 분석에서는 이 구간에서 고점 공포가 감지되었으나, Lexicon 분석에서는 정반대로 축제 분위기가 점수에 반영되었다. 이는 문맥 속에 숨겨진 불안감을 곁으로 드러나는 투기적 언어가 덮어버린 현상으로, 이성적 판단이 마비되고 군중 심리가 극에 달한 구간으로 정의할 수 있다.

- Phase 4. 고점 피로감과 이탈 (11월 1일 ~ 11월 6일)

11월 5일 댓글 수는 271건으로 다시 폭증했으나, 감성 점수는 0.09로 급락했다. 주가가 흔들리기 시작하자 시장의 관심은 여전히 뜨거웠으나, 맹목적인 환호는 사라지고 눈치 보기 장세가 시작되었다. 이는 고점 피로감 누적으로 인해 투자자들이 이탈 조짐을 보이며, 탐욕의 열기가 급격히 식어가는 국면임을 보여준다.

5) 감성 지표의 교차 검증

방법론적 특성이 상이한 두 가지 모델의 감성 지수를 중첩 분석함으로써, 상승장 이면에 존재하는 투자자들의 심리적 과리를 규명하였다. [그림 8]은 두 감성 지수와 주가 흐름을 하나의 시계열 선상에 비교 시각화한 것이다.



[그림 8] KR-FinBERT vs Lexicon 감성 지수 비교

분석 결과, 동일한 댓글 데이터를 분석했음에도 불구하고 두 모델의 감성 궤적은 뚜렷한 차이를 보였다.

- KR-FinBERT: 내재된 불안

문맥을 파악하는 모델인 KR-FinBERT의 감성 지수는 주가 급등기에도 큰 폭으로 상승하

지 않고 다소 냉정함을 유지하거나 오히려 하락하는 경향을 보였다. 이는 문장 속에 숨겨진 ‘언제 떨어질지 모른다’는 불안감과 위기관리 심리를 모델이 포착했기 때문이다.

- Lexicon: 표출된 탐욕

직관적인 키워드를 카운트하는 Lexicon 분석은 10월 말 주가 고점 부근에서 폭발적인 상승세를 기록했다. 이는 투자자들이 ‘매수 기회’, ‘가즈아’와 같은 긍정적이고 투기적인 단어를 적극적으로 사용하며, 표면적으로는 축제 분위기를 형성했음을 의미한다.

이러한 두 지표의 엇갈림은 투자자들의 심리 상태가 전형적인 인지부조화 상태에 있었음을 시사한다. 딥러닝 모델이 읽어낸 투자자의 심층 심리는 고점에 대한 두려움으로 인해 위축되어 있었다. 그러나 사전 기반 분석이 포착한 언어적 행동은 추격 매수로 나타났다. 결론적으로 이번 상승장에서 개인 투자자들은 이성적으로는 리스크를 인지하고 두려워했으나, 행동적으로는 군중 심리에 휩쓸려 투기적 매수에 동참하는 이중적인 태도를 보였다.

본 연구는 서로 다른 두 가지 분석 방법론을 결합함으로써, 단순히 ‘시장이 좋았다’는 일차원적 결론을 넘어 ‘불안해하면서도 매수 버튼을 누르는’ 상승장 특유의 복합적인 투자 심리를 정량적으로 분리하였다.

3.2 통계 분석

3.2.1 차이 검정

1) 이벤트 전후 담론 양상 변화

주가 급등 이벤트를 기점으로 뉴스의 지배적 담론이 어떠한 양상으로 재구성되는지 확인하기 위해, 이벤트 전과 후의 토픽 비중 변화를 추적하였다. [그림 9]는 시기별 토픽 구성비의 변화를 시각화한 것이다.



[그림 9] 이벤트 전후 뉴스 토픽 구성 변화

분석 결과, 시장의 관심사가 확정된 실적에서 미래의 기대로 이동하는 구조적 변화가 관측되

었으며, 세부적인 변화 양상은 다음과 같다.

- [주황/빨강색] 펀더멘털 프레임의 퇴조

가장 눈에 띄는 변화는 기업의 성과나 업황을 다루는 ‘실적/펀더멘털’ 토픽의 비중 감소다. 해당 토픽은 이벤트 전 25.8%를 차지하며 주요 담론을 형성했으나, 급등 구간인 이벤트 후에는 16.9%로 약 8.9%p 급감하였다. 이는 주가가 폭등하는 구간에서 시장이 기업의 객관적인 성과나 펀더멘털에 대한 관심을 줄이고 있음을 보여준다. 즉, 투자 판단에 있어 이성적 지표의 영향력이 약화되고, 다른 요인이 의사결정의 중심에 들어섰음을 시사한다.

- [하늘색] 기대감 프레임의 폭발

감소한 펀더멘털의 빈자리는 미래 성장성을 다루는 ‘협력 기대감’ 토픽이 대체하였다. 해당 토픽의 비중은 이벤트 전 9.2%에 불과했으나, 이벤트 후 17.3%로 약 2배 가까이 급증하였다. 이는 시장 참여자들이 과거의 확정된 실적이 아닌, 불확실하지만 높은 수익이 기대되는 미래의 내러티브를 주가 상승의 핵심 근거로 삼기 시작했음을 통계적으로 입증한다.

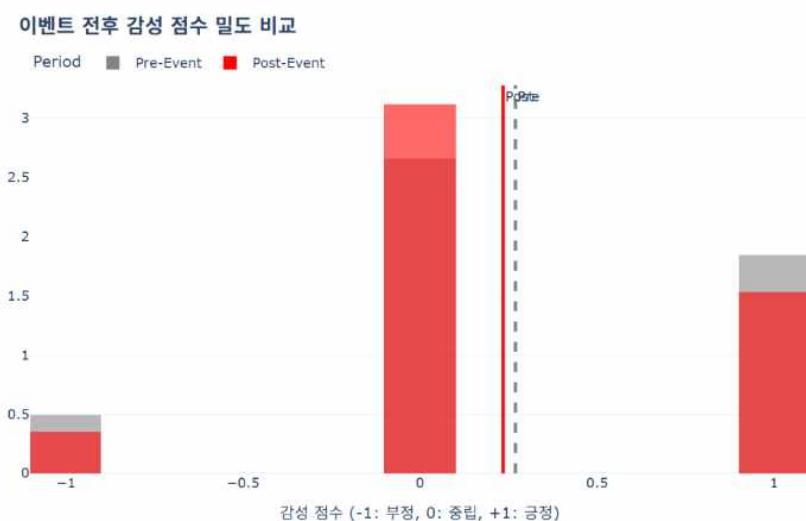
- [보라색] 촉매제의 등장

급등 이전에는 전혀 관측되지 않았던 ‘이벤트’ 토픽이 급등 기간 동안 5.2%의 비중으로 새롭게 등장하였다. 이는 특정 이슈나 사건이 잠재되어 있던 기대감을 폭발시키는 트리거 역할을 수행했음을 의미하며, 시장이 특정 재료에 의해 민감하게 반응하는 국면으로 진입했음을 보여준다.

뉴스 토픽 구성 변화는 시장을 지배하는 프레임이 이성에서 감성으로, 즉 사실에서 기대로 전이되었음을 보여주는 명확한 증거이다. 이번 주가 급등이 펀더멘털의 재평가가 아닌, 내러티브의 확산에 기인한 유동성 랠리임을 시사한다.

2) 이벤트 전후 정보 특성 변화

주가 급등 이벤트를 기점으로 시장에 공급되는 정보의 성격이 유의미하게 변화했는지 검증하기 위해, 보도량과 감성 점수에 대한 Welch's T-test를 실시하였다. [그림 10]은 이벤트 전후의 감성 점수 밀도 변화를 시각화한 것으로, 붉은색 막대가 중립 구간에서 폭발적으로 증가한 양상을 보여준다.



[그림 10] 이벤트 전후 감성 점수 비교

통계 검정 결과, 이번 상승장은 관심도 폭발과 감성의 희석이라는 두 가지 상반된 현상이 동시에 발생한 것으로 나타났다.

- **관심도의 폭발**

분석 결과, 일평균 뉴스 보도량은 이벤트 이전 기간 1.2건에서 이벤트 이후 기간 1.4건으로 증가하였으며, ($t = 6.74$, $p < 0.001$)로 통계적으로 매우 유의미한 차이를 보였다. 높은 t값은 급등 기간 동안 미디어의 관심이 폭발적으로 집중되었음을 의미한다. 이는 주가 급등 현상이 대중 및 미디어의 관심도 증가와 밀접하게 연동되어 있음을 시사하며, 시장의 유동성이 화제성에 의해 견인되었음을 뒷받침한다.

- **감성의 희석**

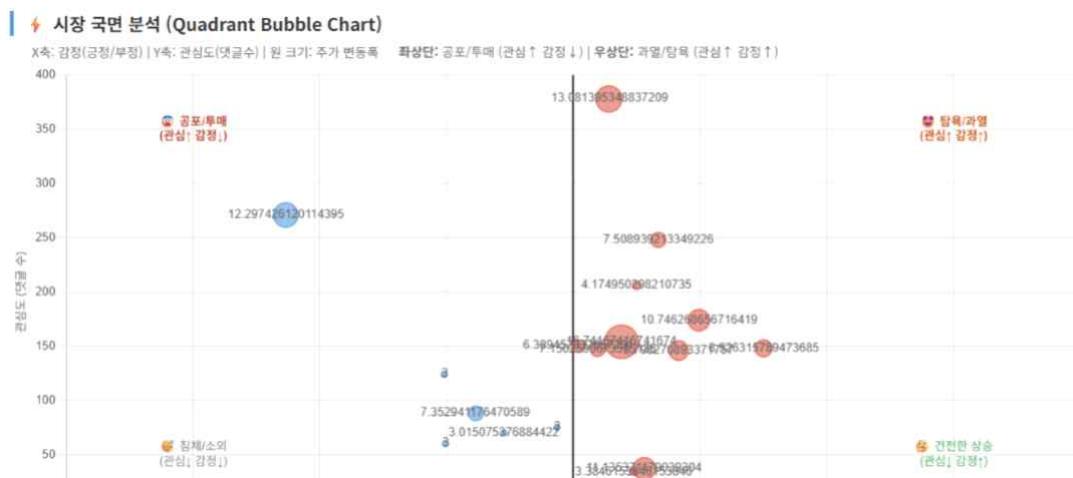
정보의 양이 증가했음에도 불구하고, 정보의 질적 지표인 평균 감성 점수는 오히려 하락했음을 발견했다. 평균 감성 점수는 이벤트 이전 0.27에서 이후 0.23으로 유의미하게 감소하였다. 이는 [그림 10]의 밀도 그래프에서도 확인되듯, 상승 초기에는 실적 호조에 기반한 뚜렷한 긍정 뉴스가 주를 이루었으나, 주가 급등기에는 기대감과 추측이 섞인 중립 성향의 뉴스가 대량 생산되면서 전체적인 감성의 평균을 끌어내렸기 때문이다.

두 검정 결과를 종합하면, 급등 기간의 시장 정보는 양은 늘어났으나, 감성의 순도는 희석되었다고 요약할 수 있다. 이는 이번 반도체 랠리가 편더멘털의 질적 개선보다는 정보의 양과 화제성이 투기적 수요를 자극하여 만들어낸 유동성 장세였음을 통계적으로 입증한다.

3.2.2 상관관계 분석

1) 시장 국면 분석

유튜브 댓글 데이터의 감성(X축)과 관심도(Y축), 그리고 주가 변동성을 3차원으로 결합한 시장 국면 분석을 수행하여, 이번 상승장의 성격을 입체적으로 바라보았다. [그림11]은 해당 분석 결과를 4분면으로 시각화한 것이다.



[그림 11] 시장 국면 분석

- **상승장의 역설**

분석 결과, 주가 상승기임에도 불구하고 상당수의 데이터 버블이 좌상단의 공포/투매(관심

↑, 감정 ↓) 영역에 집중되어 있는 현상이 관측되었다. 이는 이번 랠리가 투자자들의 신뢰와 환호 속에 진행된 것이 아니라, 의심과 공포 속에서 진행되었음을 의미한다. 즉, 시장이 악재와 불안심리를 딛고 상승하는 전형적인 불안 장세였음을 시각적으로 규명되었다.

- **소음과 변동성의 상관관계**

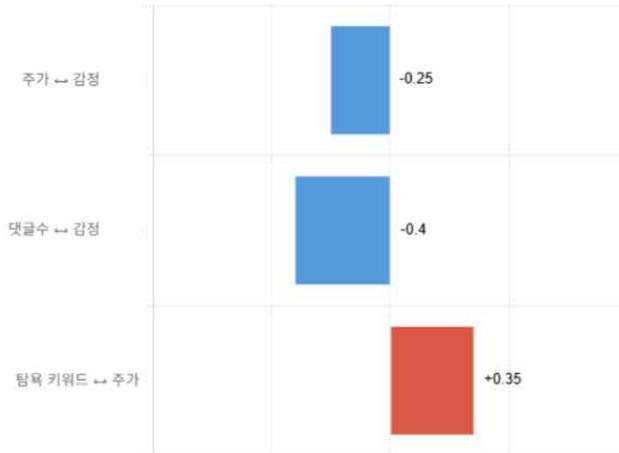
그래프의 Y축은 대중의 관심도를, 버블의 크기는 주가의 변동폭을 나타낸다. 시각화 결과, 버블의 크기가 클수록 그래프의 위쪽에 위치하는 경향이 뚜렷하게 나타났다. 이는 대중의 논쟁과 관심이 과열될수록 주가의 흔들림 또한 커진다는 것을 시사한다. 반면, 관심도가 낮은 하단부의 버블들은 상대적으로 크기가 작아 주가 흐름이 안정적이었다. 결론적으로 이번 상승장에서 대중의 과도한 관심은 주가의 안정적 상승보다는 변동성을 확대시키는 노이즈로 작용했음을 알 수 있다.

시장 국면 분석을 통해 확인된 이번 상승장의 실체는 대중이 불안해하며 시끄럽게 떠들수록, 주가는 더 크게 오르는 비이성적 유동성 장세로 요약할 수 있다. 이는 앞서 KR-FinBERT 분석에서 도출된 내재된 불안이 실제 시장 데이터의 분포에서도 동일하게 관측됨을 뒷받침한다.

2) 유튜브 상관관계 분석

유튜브 댓글 데이터에서 추출한 심층 감정(KR-FinBERT), 표면적 탐욕 키워드(Lexicon), 시장 데이터 간의 피어슨 상관분석을 수행하였다. [그림12]는 주요 변수 간의 상관계수를 시각화한 것이다.

| 변수 간 상관관계 계수 (Correlation Coefficient)



[그림 12] 상관계수 비교

분석 결과, 상승장 이면에는 감성의 역설과 투자자의 인지부조화가 통계적으로 유의미하게 존재함이 입증되었다.

- **주가와 심층 심리의 괴리 ($r = -0.25$)**

일반적으로 주가 상승은 투자자의 행복감을 고취시켜 감성 점수와 양의 상관관계를 가질 것으로 기대된다. 그러나 KR-FinBERT로 분석한 심층 감정 점수는 주가와 -0.25의 음의 상관관계를 보였다. 이는 주가가 상승할수록 오히려 대중의 심리는 부정적으로 변했음을 의미하며, 이번 상승장이 언제 폭락할지 모른다는 불안감과 고점 공포가 지배하는 가운데

진행되었음을 통계적으로 확인시켜 준다.

- **행동적 동조화 ($r = +0.35$)**

심리적 불안과는 별개로, 투기적 단어를 사용하는 빈도는 주가와 $+0.35$ 의 뚜렷한 양의 상관관계를 보였다. 이는 심층적인 감정은 위축되어 있음에도 표면적인 언어 행동은 상승 추세에 편승하여 탐욕을 드러내는 현상이다. 통계적으로 볼 때, 당시 투자자들은 불안해하면서도, 매우 버튼을 누르는 인지부조화 상태에 놓여 있었음을 방증한다.

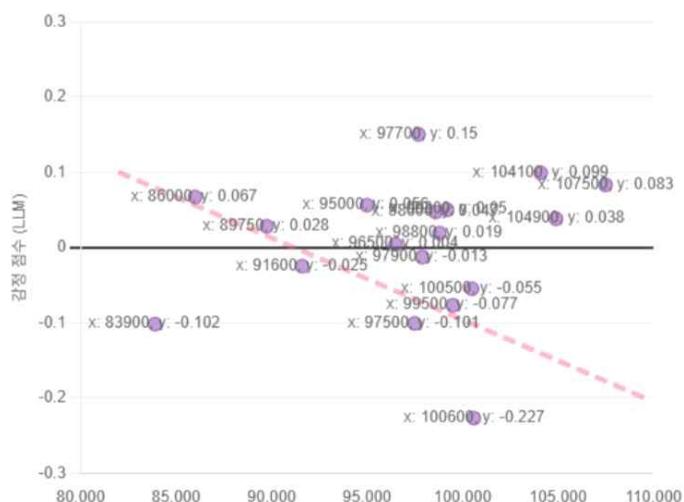
- **관심의 역설 ($r = -0.40$)**

댓글 수와 감정 점수 간에는 -0.40 의 강한 음의 상관관계가 관측되었다. 이는 대중의 관심이 폭발적으로 쏠릴수록 여론이 긍정적으로 통합되기보다는, 오히려 논쟁이 격화되고 부정적이며 공격적인 언어가 난무했음을 시사한다.

3) 상관관계 상세 분석

앞서 확인된 주가와 심리 간의 음의 상관관계를 정밀하게 파악하기 위해, 삼성전자 주가와 KR-FinBERT 감성 점수에 대한 산점도 및 회귀 분석을 수행하였다.

상관관계 상세: 주가 vs 심리 산점도 (Scatter)



[그림 13] 주가와 감성 점수 분포 비교

분석 결과, 주가가 고점 구간에 진입할수록 대중의 심리가 급격히 냉각되는 고점 공포 현상이 구조적으로 존재함이 시각적으로 나타났다.

- **우하향 회귀선: 가격 상승이 주는 심리적 압박**

데이터의 분포를 설명하는 회귀선은 뚜렷한 우하향 패턴을 보인다. 이는 주가가 상승할수록 투자자들의 심리는 반비례하여 낮아짐을 의미한다. 일반적으로 주가 상승은 투자자에게 수익을 안겨주어 긍정적 심리를 유발할 것으로 예상되나, 본 데이터는 정반대의 결과를 보여준다. 이는 급등장 속에서 가격 상승이 대중에게 환호의 대상이 아니라, 심리적 압박으로 작용했음을 시사한다.

- **10만 원 구간의 심리적 저항**

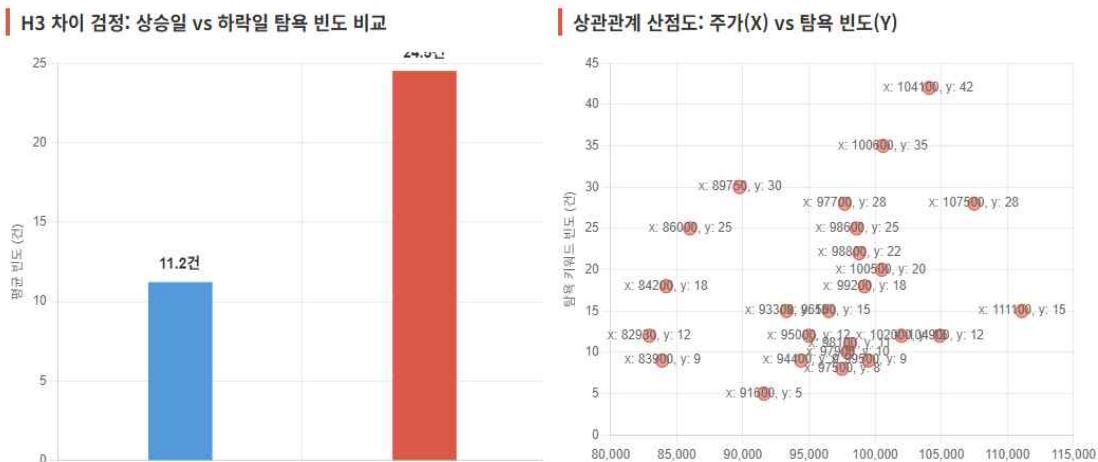
삼성전자 주가가 10만 원을 돌파하는 구간, 즉 X축 100,000원 이상의 구간에서 감정 점수

는 대부분 0 이하의 음수 영역에 밀집되어 있다. 이는 시장 참여자들이 차익 실현 욕구와 하락에 대한 공포가 충돌하는 위험 구간으로 인식했음을 보여준다.

회귀 분석 결과를 통해, 이번 상승장에서 개인 투자자들은 주가가 오를수록 안도하는 것이 아니라 오히려 더 큰 스트레스와 공포를 느꼈음이 확인되었다. 이는 주가 급등이 실적에 대한 확신 없이 유동성의 힘으로만 진행될 때 나타나는 불안한 강세장의 특징이라고 할 수 있다.

4) 유튜브 탐욕 빈도 분석

추가로, 주가 등락에 따른 개인 투자자들의 심리 변화를 행동적 차원에서 검증하기 위해, 유튜브 댓글 내 탐욕 키워드의 출현 빈도를 주가 상승일과 하락일로 나누어 차이 검정을 실시하였다.



[그림 14] 좌: 주가 상승일과 하락일 차이검정 / 우: 주가와 탐욕 간의 상관관계

주가 등락 여부에 따른 탐욕 키워드 평균 빈도를 분석한 결과, 하락일의 평균 빈도는 11.2건에 불과했으나, 상승일에는 24.5건으로 2배 이상(+119%) 폭증하는 현상이 관측되었다. 이는 개인 투자자들의 언어적 행동이 주가 방향성에 극도로 민감하게 반응함을 보여준다. 즉, 시장이 조금만 반등해도 투자자들은 즉각적으로 투기적 언어를 쏟아내며 시장 분위기를 과열시키는 경향이 있음이 통계적으로 확인되었다.

주목할 점은 주가와 탐욕 빈도 간의 산점도 분석 결과로, 두 변수는 +0.35의 뚜렷한 양의 상관관계를 보였다. 이성적인 투자자라면 주가가 고점에 도달할수록 리스크를 관리하고 탐욕을 절제해야 한다. 그러나 주가가 비싸질수록 투자자들의 탐욕도 함께 커지는 비이성적 패턴이 나타났다. 이는 고점에서 차익 실현을 하기보다 추격 매수에 나서는 개인 투자자 특유의 비이성적 과열 양상을 보여준다.

3.3 인과 분석

3.3.1 그레인저 인과 분석

본 연구는 상관관계 분석에서 확인된 변수 간의 연관성을 넘어, ‘뉴스의 발생이 실제 주가 변동을 시계열적으로 선행하는가?’를 규명하기 위해 그레인저 인과 검정을 실시하였다. 시장을 지배하는 두 가지 핵심 프레임인 ‘펀더멘털’과 ‘협력 기대감’이 주가 수익률에 미치는 인과적 영향력을 시차(Lag 1~5)별로 비교 분석하였다.



[그림 15] 그레인저 인과 검정 결과

- [실적/펀더멘털] 선행 지표로서의 기능 상실

기업의 재무적 성과와 업황을 다루는 실적 관련 뉴스는 주가 변동에 유의미한 영향을 미치지 못하는 것으로 나타났다. Lag 1일부터 5일까지의 모든 시차에서 p-value는 유의수준 0.05를 크게 상회($p > 0.13$)하여 귀무가설을 기각하지 못했다. 이는 이번 상승장에서 확정된 실적 정보는 이미 가격에 선반영 되었거나, 시장 참여자들의 매수 욕구를 자극하는 선행 지표로서의 기능을 상실했음을 시사한다. 실적은 주가를 지지하는 후행적 근거일 뿐, 상승의 동력은 아니었던 것이다.

- [협력 기대감] 주가 폭발의 선행 원인 (2~3일 시차)

미래 성장성과 협력을 다루는 기대감 관련 뉴스는 주가 수익률을 통계적으로 유의미하게 선행하는 것으로 확인되었다. 구체적으로 Lag 2일($p=0.0272$)과 3일($p=0.0237$)에서 유의수준 0.05 미만의 값을 기록하여 기대감 뉴스는 주가 수익률을 유의미하게 선행하며, 예측력을 갖는 것으로 나타났다. 즉, 협력 관련 이슈가 보도되면 즉각적인 반응보다는 약 2~3일의 잠복기를 거친 후 주가가 폭발적으로 상승하는 패턴이 확인되었다.

기대감 뉴스가 주가에 반영되기까지 2~3일의 시차가 존재한다는 사실은 정보 확산의 메커니즘을 시사한다. 이는 ‘뉴스 보도 → 유튜브/커뮤니티 확산 → 개인 투자자 심리 자극 → 매수’로 이어지는 일련의 정보 소화 과정에 물리적인 시간이 소요되기 때문으로 해석된다. 결론적으로 이번 반도체 랠리는 펀더멘털의 즉각적 반영이 아닌, 미래에 대한 기대감이 대중 사이로 확산되며 만들어낸 시차를 둔 유동성 폭발 현상이었음을 인과적으로 규명하였다.

4. 결론

4.1 연구 요약 및 결론

본 연구는 2025년 반도체 섹터의 주가 급등 현상을 대상으로 텍스트 마이닝과 통계적 분석 기법을 적용하여, 미디어 정보와 시장 심리가 주가에 미치는 영향을 실증적으로 분석하였다. 연구 결과, 본 연구가 설정한 4가지 핵심 가설은 모두 통계적으로 유의미하여 채택되었으며, 그 세부 검증 결과는 다음과 같다.

가설 1. 프레임 전이 가설

주가 급등 이벤트 발생 후, 미디어 내 실적 관련 토픽의 비중은 감소하고 기대감 관련 토픽의 비중은 유의미하게 증가할 것이다.

- 뉴스 토픽의 시계열 분석 결과, 이벤트 전 25.8%를 차지하던 편더멘털 토픽은 급등기 진입 후 16.9%로 급감한 반면, 협력 기대감 토픽은 9.2%에서 17.3%로 약 2배 폭증하였다. 이는 시장의 관심사가 ‘확정된 과거(실적)’에서 ‘불확실한 미래(기대)’로 구조적으로 전이되었음을 입증한다.

가설 2. 감성의 역설 가설

주가 급등기에는 뉴스 보도량이 증가함에도 불구하고, 텍스트의 평균 감성 점수는 하락하거나 중립화되는 감성 괴리 현상이 나타날 것이다.

- 주가 폭등과 함께 뉴스 보도량은 유의미하게 증가($t=6.74$)하였으나, 평균 감성 점수는 0.27에서 0.23으로 오히려 하락하였다. 이는 주가 상승이 정보의 긍정적 내용이 아닌, 압도적인 관심의 양과 화제성에 의해 견인되었음을 보여주는 ‘감성의 역설’을 확인시켜 준다.

가설 3. 투기적 동조화 가설

뉴스의 감성과 달리, 개인 투자자의 탐욕 키워드 빈도는 주가 상승 추세와 유의미한 양의 상관관계를 보일 것이다.

- KR-FinBERT 분석과 달리, Lexicon 기반의 탐욕 키워드 빈도는 주가 상승과 뚜렷한 양의 상관관계를 보였다. 이는 개인 투자자들이 고점에서 리스크를 경계하기보다 추격 매수에 나서는 비이성적 과열 및 동조화 현상을 보였음을 시사한다.

가설 4. 정보 선행성 가설

실적 정보는 주가 수익률을 선행하지 못하는 반면, 기대감 정보는 유의미한 시차를 두고 주가 수익률을 그레인저 인과 할 것이다.

- 그레인저 인과 검정 결과, 실적 토픽은 선행성이 기각된 반면, 기대감 토픽은 2~3일의 시차를 두고 주가 상승을 유의미하게 선행하는 것으로 나타났다. 이는 정보가 확산되는 물리적 시간을 거쳐 유동성 폭발로 이어지는 메커니즘을 통계적으로 규명한 것이다.

가설 검증 결과를 종합할 때, 2025년 반도체 섹터의 랠리는 기업의 내재 가치에 기반한 이성적 재평가가 아닌, 미디어 내러티브의 확산과 대중의 투기적 동조화가 결합된 유동성 장세로 결론지을 수 있다. 시장은 명확한 ‘숫자’보다 기대감을 불러일으키는 ‘이야기’에 반응하였으며, 이 과정에서 발생한 정보의 비대칭과 시차는 주가의 변동성을 확대시키는 요인으로 작용하였음을 알 수 있다.

4.2 연구의 시사점

본 연구는 텍스트 마이닝과 통계적 분석을 통해 반도체 섹터의 주가 급등 현상을 실증적으로 규명하였다. 이러한 연구 결과는 로버트 쉴러의 내러티브 경제학을 정량화했다는 학술적 의의 와, 투자자 및 금융 기관에 구체적인 대응 전략을 제시한다는 실무적 의의를 동시에 갖는다.

4.2.1 학술적 시사점

1) 내러티브 경제학의 정량적 실증

기존의 재무금융 연구는 효율적 시장 가설에 기반하여 주가 변동을 설명하려 했으나, 비이성적 과열 현상을 설명하는 데 한계가 있었다. 로버트 쉴러는 내러티브 경제학을 통해 이야기가 시장을 움직인다고 주장했으나, 이는 주로 정성적 이론에 머물러 있었다. 본 연구는 텍스트 마이닝과 시계열 분석을 결합하여, 시장의 주도권이 실적에서 기대감으로 넘어가는 구조적 전이 과정을 시각적 및 통계적으로 입증하였다. 이는 내러티브가 단순한 소음이 아니라, 주가를 견인하는 핵심 원인임을 정량적으로 객관화했다는 점에서 학술적 가치를 지닌다.

2) 이종 데이터 결합을 통한 하이브리드 분석 프레임워크 제시

본 연구는 정보의 공급자인 '제도권 뉴스(News)'와 정보의 수요자인 '개인 투자자(YouTube)' 데이터를 통합 분석하는 하이브리드 방법론을 시도하였다. 뉴스의 이성적 텍스트와 유튜브의 감성적 반응을 교차 분석함으로써, 기존의 단일 데이터 분석으로는 포착하기 어려웠던 '감성의 역설'이나 '인지부조화'와 같은 시장 미시 구조를 규명하였다. 이는 빅데이터 분석이 금융 시장의 비이성적 과열 현상을 해석하는 데 있어 강력한 도구가 될 수 있음을 시사하며, 행동 재무학 연구의 방법론적 지평을 확장하였다.

4.2.2 실무적 시사점

1) 투자자 관점: 감성 괴리 기반의 리스크 헤지 전략

본 연구 결과는 개인 투자자들에게 추격 매수의 위험을 경고하고 과학적인 매매 타이밍을 포착하는 데 활용될 수 있다.

- 감성 역설의 지표화: 주가는 상승하지만 뉴스의 감성 점수가 하락하거나 중립화되는 시점을 투기적 베팅 신호로 정의할 수 있다. 투자자들은 이를 추격 매수를 자제하고 차익 실현을 준비하는 선행 보조 지표로 활용하여, 고점 매수의 리스크를 최소화할 수 있다.
- 2차 확산 시차 활용: 뉴스 보도 후 개인 투자자들에게 정보가 확산되어 실제 주가 급등으로 이어지기까지 약 2~3일의 물리적 시차가 존재함이 규명되었다. 이는 뉴스 발생 직후가 아닌, 정보가 유튜브 등으로 확산되는 시점을 노리는 골든타임 전략 수립을 가능하게 한다.

2) 정책 및 규제 관점: 골든타임 확보를 통한 시장 건전성 제고

금융 기관 및 규제 기관은 본 연구의 모델을 활용하여 시장의 불공정 거래나 과열을 선제적으로 모니터링할 수 있다.

- 규제 골든타임 확보: 정보의 발생과 주가 폭등 사이에 존재하는 2~3일의 시차는 규제 당

국이 대응할 수 있는 골든타임이다. 이 기간 동안 불공정 거래나 허위 사실 유포를 집중 모니터링함으로써 시장 교란 행위를 조기에 차단할 수 있다.

- 탐욕 키워드 기반 경보 시스템: 기업의 펀더멘털과 무관하게 투기적 탐욕 키워드가 급증하는 구간을 투자 유의 구간으로 설정할 수 있다. 이를 통해 맹목적인 뇌동매매로 인한 개인 투자자들의 피해를 방지하고, 시장 경보를 발령하는 정량적 근거로 활용 가능하다.

4.3 한계점 및 향후 연구 방향

본 연구는 텍스트 마이닝을 통해 시장의 비이성적 과열과 정보 확산의 시차를 실증적으로 규명했다는 점에서 의의가 있으나, 표본의 특성 및 방법론적 측면에서 다음과 같은 한계점을 갖는다. 이에 본 연구는 식별된 한계를 극복하기 위한 향후 연구 방향을 제안한다.

4.3.1 연구의 한계점

1) 정보 확산 경로의 '블랙박스' 미규명

본 연구는 뉴스 발생과 주가 반영 사이에 약 2~3일의 시차가 존재함을 통계적으로 입증하였으나, 정보가 구체적으로 어떤 경로를 통해 증폭되는지에 대한 중간 전달 과정을 규명하지는 못했다. 즉, 뉴스가 유튜브 채널이나 특정 커뮤니티 인플루언서를 거쳐 확산되는 미시적 메커니즘이 여전히 블랙박스로 남아있어, 정보 확산의 경로보다는 결과를 확인하는 데 그쳤다는 한계가 있다.

2) 특정 국면 및 섹터에 국한된 결과의 일반화

문제 분석 대상이 2025년 반도체 섹터의 강력한 상승장에 국한되어 있어, 연구 결과를 일반화하는 데 제약이 따른다. 펀더멘털이 중시되는 하락장이나, 바이오 및 금융 등 타 산업 섹터에서도 본 연구에서 확인된 프레임 전이나 감성의 역설 현상이 동일하게 나타나는지 단정하기 어렵다.

3) 텍스트 분석 모델의 한계와 심리 해석의 불완전성

KR-FinBERT와 Lexicon 분석 결과가 엇갈리는 인지부조화 현상을 발견했으나, 오직 댓글 텍스트에만 의존한 분석으로는 투자자의 복잡한 심리를 완벽하게 분리해내기에 한계가 존재한다. 텍스트는 투자자의 의도를 간접적으로 보여줄 뿐, 실제 매수/매도 행위와 직결되는지 확증할 수 없기 때문이다.

4) 정량적 데이터의 신뢰성 및 통계적 착시

데이터 수집 과정에서 일일 댓글 수가 10건 미만인 날짜가 일부 존재하였다. 이러한 표본 과소 구간에서는 소수의 극단적인 의견이 전체 여론을 대변하는 것처럼 왜곡되는 소수 법칙의 오류나 통계적 착시가 발생할 가능성이 있다. 이는 시계열 분석 시 특정 시점의 감성 점수가 과도하게 튀는 현상을 유발하여, 전체 추세의 왜곡을 초래할 수 있다.

4.3.2 향후 연구 방향 제안

1) 사회 연결망 분석을 통한 슈퍼 전파자 식별

단순 시계열 분석을 넘어, 유튜브 채널 간의 인용 관계나 커뮤니티 내 공유 패턴을 분석하는 사회 연결망 분석 도입을 제안한다. 이를 통해 정보 확산의 허브 역할을 하는 슈퍼 전파자를 식별하고, 정보의 증폭 경로를 시각화하는 연구로 확장할 필요가 있다.

2) 산업별/국면별 비교 연구

연구의 외적 타당성을 높이기 위해, 실적 민감도가 높은 가치주 섹터와 기대감 의존도가 높은 성장주 섹터를 비교 분석하는 후속 연구가 필요하다. 또한 상승장과 하락장을 구분하여, 시장 국면에 따른 정보 반영 속도와 프레임 변화 양상의 차이를 규명하는 비교 연구를 제안한다.

3) 멀티모달 분석을 통한 교차 검증

텍스트 데이터의 한계를 보완하기 위해, 거래량, 회전율, 공매도 잔고 등 실제 시장 미시 데이터를 결합한 멀티모달 분석 도입을 제안한다. 심리가 실제 시장 행동으로 이어지는지 교차 검증함으로써, 분석 모델의 설명력을 높일 수 있을 것이다.

4) 데이터 전처리의 고도화 및 시계열 신뢰도 제고

향후 연구에서는 데이터의 신뢰성을 확보하기 위해 표본 과소 구간을 분석에서 제외하거나, 3 일/5일 이동평균선을 적용하여 노이즈를 제거하는 등 통계적 강건성을 높이는 전처리 과정을 강화해야 한다.

References

국내 문헌

- 서울대학교 컴퓨터공학부 자연어처리 연구실 (SNUNLP). (2021). KR-FinBERT: Pre-trained BERT model for Financial Domain in Korean. <https://github.com/snunlp/KR-FinBERT>

국외 문헌

- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2), 159-178.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171-4186.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424-438.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
- Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4), 967-1004.
- Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.

Our Code

https://drive.google.com/drive/folders/1cOJCeD5rqCVa0Da8Xj2trN781YkvjhVO?usp=drive_link