

네이버 뉴스 댓글을 이용한 산업 분야별 담론의 감성에 기반한 주제 트렌드 및 여론의 변화와 주가 흐름의 연관성 분석*

Analyzing Topic Trends and the Relationship between Changes in Public Opinion and Stock Price based on Sentiment of Discourse in Different Industry Fields using Comments of Naver News

오찬희 (Chanhee Oh)**

김규리 (Kyuli Kim)***

주영준 (Yongjun Zhu)****

초 록

본 연구에서는 대한민국 정부가 지정한 국가전략기술 사업인 반도체, 이차전지, 바이오 산업에 대한 여론을 파악하고 여론의 변화와 주가 흐름의 연관성을 분석하기 위해 각 산업별 대표 기업에 대한 기사의 댓글을 분석하였다. 반도체 산업에서 '삼성전자', 'SK하이닉스', 이차전지 산업에서 '삼성SDI', 'LG화학', 바이오 산업에서 '삼성바이오로직스', '셀트리온'을 선정하여 이를 제목에 포함하고 있는 2020년 1월 1일부터 2020년 12월 31일까지 발행된 네이버 뉴스 기사의 댓글 47,452개를 수집하고 분석하였다. 먼저, 해당 댓글을 긍정, 중립, 부정의 감성으로 나누고 각 감성 그룹에서의 시간의 흐름에 따른 댓글의 동적인 주제를 분석하여 각 산업별 여론의 트렌드를 파악하였다. 분석 결과 반도체 산업 분야의 경우 투자, 코로나19관련 이슈, 삼성전자라는 대기업에 대한 신뢰, 정부 정책 변화로 인한 타격에 대한 언급이 주제 토픽으로 나타났다. 이차전지 산업체의 경우 투자, 배터리, 기업 이슈에 대한 언급이 주제 토픽으로 나타났다. 바이오 산업체의 경우 투자, 코로나19 관련 이슈 및 기업 이슈에 대한 언급이 주제 토픽으로 나타났다. 다음으로, 댓글의 감성이 실제 주가와 연관성이 있는지를 알아보고자 각 대표 기업 별 주가의 변화와 댓글의 감성 점수 변화를 시각적 분석기법을 이용하여 비교 분석하였다. 분석 결과, 댓글의 감성 점수와 주가의 변화 흐름이 매우 유사하게 나타남을 통해 여론의 감성 점수 변화와 주가의 흐름에는 연관성이 있음을 확인하였다. 본 연구는 주가와의 연관성이 높은 뉴스 기사 댓글을 분석했다는 점, 수집 시기를 코로나19로 선정하여 코로나19라는 특수한 상황에서의 여론 트렌드 변화를 파악했다는 점, 국가전략기술제도에 속하는 산업 기업에 대한 여론을 분석하여 정부기관의 관련 정책 제정에 객관적인 근거를 제공하였다는 점에서 의의를 지닌다.

ABSTRACT

In this study, we analyzed comments on news articles of representative companies of the three industries (i.e., semiconductor, secondary battery, and bio industries) that had been listed as national strategic technology projects of South Korea to identify public opinions towards them. In addition, we analyzed the relationship between changes in public opinion and stock price. 'Samsung Electronics' and 'SK Hynix' in the semiconductor industry, 'Samsung SDI' and 'LG Chem' in the secondary battery industry, and 'Samsung Biologics' and 'Celltrion' in the bio-industry were selected as the representative companies and 47,452 comments of news articles about the companies that had been published from January 1, 2020, to December 31, 2020, were collected from Naver News. The comments were grouped into positive, neutral, and negative emotions, and the dynamic topics of comments over time in each group were analyzed to identify the trends of public opinion in each industry. As a result, in the case of the semiconductor industry, investment, COVID-19 related issues, trust in large companies such as Samsung Electronics, and mention of the damage caused by changes in government policy were the topics. In the case of secondary battery industries, references to investment, battery, and corporate issues were the topics. In the case of bio-industries, references to investment, COVID-19 related issues, and corporate issues were the topics. Next, to understand whether the sentiment of the comments is related to the actual stock price, for each company, the changes in the stock price and the sentiment values of the comments were compared and analyzed using visual analytics. As a result, we found a clear relationship between the changes in the sentiment value of public opinion and the stock price through the similar patterns shown in the change graphs. This study analyzed comments on news articles that are highly related to stock price, identified changes in public opinion trends in the COVID-19 era, and provided objective feedback to government agencies' policymaking.

키워드: 감성 분석, 다이나믹 토픽 모델링, 댓글, 소셜 네트워크 서비스, 시각적 분석, 여론 분석, 주가, 트렌드 분석, comments, discourse analysis, dynamic topic modeling, sentimental analysis, SNS, stock, trend analysis, visual analysis

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5C2A02088387).

** 성균관대학교 문헌정보학과 석사과정(010130@g.skku.edu) (제1저자)

*** 성균관대학교 문헌정보학과 석사과정(kyulikum@g.skku.edu) (공동저자)

**** 연세대학교 문헌정보학과 조교수(zhu@yonsei.ac.kr) (교신저자)

■ 논문접수일자: 2022년 2월 21일 ■ 최초심사일자: 2022년 3월 7일 ■ 게재확정일자: 2022년 3월 14일

■ 정보관리학회지, 39(1), 257-280, 2022. <http://dx.doi.org/10.3743/KOSIM.2022.39.1.257>

※ Copyright © 2022 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

2021년 12월, 대한민국 정부는 2022년 2월부터 '국가전략기술' 제도를 시행하겠다고 발표했다(박성우, 2021). 이는 글로벌 기술 패권과 공급망 경쟁으로 심화된 반도체, 이차전지, 바이오 3대 분야를 국가전략기술로 선정하고(박성우, 2021), 세액공제율에 우대를 적용해 세제지원을 강화하려는 방침이다(기획재정부, 2021). 정부는 국가전략기술의 연구개발(R&D)과 시설에 투자하는 기업에 세제 혜택을 제공하여(기획재정부, 2021), 반도체, 이차전지, 바이오 분야의 연구 및 사업의 성장을 장려하였다. 이에 따라 본 연구는 국가전략기술 사업의 본격적인 시행 전 각 산업 분야의 연구 개발과 시설 투자의 근거를 제공하기 위해 반도체, 이차전지, 바이오 분야의 대표 기업에 대한 여론을 파악해 보고자 한다. 이를 통해 각 산업별 대표 기업에 대한 사람들의 인식을 확인하고 사업별 연구 및 시설 개발의 기반을 제공할 수 있을 것이라 기대한다.

한편, 행동경제학에 따르면 '투자자 심리(sentiment)'가 주가에 영향을 미치는 것으로 보고되고 있다(김세완 외, 2020). 선행 연구에서는 투자자 심리가 반영되어 있는 곳을 찾아 주가와 금융시장의 연관성을 파악하기 위해 소셜 데이터, 온라인 뉴스, 페이스북 댓글과 같은 데이터를 분석하였다(김건태, 황수성, 2015; 김세완 외, 2020; 김유신, 김남규, 정승렬, 2012; 박성욱, 2017; 이미경, 최인호, 정세훈, 2013; 이병현 외, 2020; Fu et al., 2008; Tetlock, 2007). 이 중에서 특히 온라인 뉴스, 즉 웹뉴스는 투자자와 밀접한 관계가 있는데 이는 많은 투자자

들이 웹뉴스를 참고하여 주식거래를 하고 있기 때문이다(남달우 외, 2012). 이러한 웹뉴스에는 댓글 작성 구역이라는 항목이 생성되어 있으며 투자자들은 이를 통해 자신의 의견을 표출하고 있다(김지룡, 한은경, 2019).

따라서 본 연구에서는 뉴스 댓글을 분석함으로써 국가전략기술 사업인 반도체, 이차전지, 바이오 산업별 대표 기업에 대한 여론의 인식을 파악해보고자 한다. 감성 분석을 통해 댓글에 감성 점수 및 긍정, 중립, 부정의 감성을 부여한 뒤 감성 별로 나타나는 여론의 주제를 파악하고 여론의 감성 변화가 주가의 흐름과 연관성이 있는지 구체적인 해석을 통해 확인할 것이다.

2. 선행연구

뉴스와 주식시장 간의 연관성을 확인하기 위해 선행 연구에서는 뉴스 텍스트를 분석하여 분석 결과가 주식시장에 영향을 주는지를 알아 보았다(김유신, 김남규, 정승렬, 2012; 이병현 외, 2020; Fu et al., 2008; Tetlock, 2007). 김유신, 김남규, 정승렬(2012)은 오피니언 마이닝으로 뉴스 콘텐츠를 분석한 결과와 주가지수 등락 간의 관계를 통계 분석하여 뉴스 콘텐츠의 감성 분석 결과 값이 주가지수 등락과 유의미한 관계를 가지고 있음을 파악하였다. Tetlock(2007)은 뉴스의 내용이 투자자 심리를 반영하고 있고 이는 주가에 영향을 미치고 있음을 밝혀냈으며 구체적으로는 뉴스의 내용이 부정적인 경우 주가 또한 하락하는 경향을 보이기 때문에 뉴스의 내용과 주가의 변화는 서로 연관

성이 있다고 밝혔다. Fu et al.(2008)은 홍콩 뉴스 감성 분석으로 뉴스 내용이 담고 있는 분위기와 주가의 연관성을 분석하였으며, 긍정, 부정의 분위기에 따라 주가의 변동이 이루어지고 있음을 파악하였다. 이병현 외(2020)는 온라인 뉴스 텍스트 마이닝을 통해 회사 평판과 수입의 상관성을 분석하여 뉴스에서 CEO의 리더십 역량, 사회적 책임 평판 등에 대한 긍정적인 평판이 이어질수록 기업의 수익성이 증가함을 서술하였다.

뉴스와 주식시장의 최근 연구 동향은 뉴스 텍스트를 분석하는 방법에 딥러닝을 적용한 주가 예측 모델 개발에 맞춰져 있다(김다예, 이영인, 2018; 김동영, 박제원, 최재현, 2014; 장은아, 최희련, 이흥철, 2020). 김동영, 박제원, 최재현(2014)은 SVM으로 SNS와 뉴스 기사의 감성을 분석해 로지스틱 회귀 분석, 베이지안 네트워크, 인공신경망을 적용한 방법과 비교하여 가장 높은 예측 정확도와 성능을 보이는 주가 예측 모델을 개발했으며, 장은아, 최희련, 이흥철(2020)은 NLTK VADER(Valence Aware Dictionary and sentiment Reasoner)와 딥러닝 기반의 BERT 모델을 이용한 뉴스 정보의 감성 분석을 적용하여 효과적인 주가 예측 지표를 조합할 수 있다는 것을 제시하였다. 김다예, 이영인(2018)은 증시 뉴스 25,000여건에 Word2Vec을 진행한 주가지수 방향성 예측용 감성 사전을 구축하여 KOSPI 증가 지수를 예측하는 모델을 개발하였다. 이처럼 뉴스 데이터 분석을 통해 주가와 연관성을 알아보는 연구는 연관성을 통해 주가를 예측하는 수준까지 도달했으나 앞선 연구의 경우 연관성 존재 여부를 밝히는 것이나 주가 예측력을 높이는

것에만 초점이 맞춰져 있어 기업에 대한 실제 감성과 주가가 연관되어 있음을 구체적인 데이터를 통해 알아보는 연구는 미미한 상황이다.

한편, 대부분의 뉴스 기사가 온라인으로 발행되며 댓글 작성 부분이 생성됨에 따라 뉴스 댓글도 주목을 받기 시작했다. 김지룡, 한은경(2019)은 온라인 뉴스와 뉴스 댓글로 형성된 기업 평판이 그 기업의 수익성과 어떠한 관계가 있는지를 알아보려고 6개의 기업을 선정하여 빅데이터 분석을 실시하였고 그 중 5개의 기업에서 온라인 뉴스 및 뉴스 댓글로 형성된 기업 평판이 수익성에 영향을 미치는 것을 확인하였다. 김원정, 한정호(2016)는 조직의 사건사고 기사에 달린 댓글이 조직체의 위기책임성과 향후 제품 구매 의도에 영향력을 발휘한다는 연구 결과를 도출하였다. 배지양(2009)은 인터넷 뉴스 기사 댓글의 긍정 및 부정의 감정에 따라 달라진 응답자의 태도가 그 기업 제품에 대한 구매 의도에 영향을 미치는 것을 확인하였으며, 양혜승(2008)은 뉴스기사보다 댓글의 견해가 개인들의 찬반 태도에 더 큰 영향을 미치며, 낮은 품질의 댓글 또한 개인의 태도에 영향을 발휘하는 경향을 보였음을 확인하였다.

뉴스 댓글은 뉴스보다 개인들의 찬반 태도에 더 큰 영향력을 보이고 있다(양혜승, 2008). 댓글은 여론 형성의 주된 근원으로 작동할 수 있으며, 정제되지 않은 언어와 격한 감정을 분출하는 댓글 역시도 독자들의 태도에 적지 않은 영향을 미칠 수 있다(양혜승, 2008). 공중과의 쌍방향 커뮤니케이션 수단으로 인터넷이 각광받으며 온라인 환경에서의 소통의 중요성이 강조됨에 따라 공중의 반응을 즉각적으로 확인할 수 있는 댓글의 영향력이 더 커지고 있으며(조

수영, 장혜지, 권구민, 2012), 온라인 뉴스를 접한 뉴스 수용자들은 댓글을 통해서 사회적 이슈에 대한 자신의 생각을 표현하고 있기 때문에 (박미화, 김술, 2017) 본 연구에서는 뉴스 댓글을 통해 사람들의 여론을 파악하고자 한다. 또한, 뉴스 댓글은 주식 시장과 직·간접적으로 연관되어 있기 때문에(김유신, 김남규, 정승렬, 2012; 배지양, 2009; Fu et al., 2008; Tetlock, 2007), 댓글의 감성 변화가 주가의 흐름과 연관성이 있는지 알아보하고자 한다.

3. 데이터

본 연구는 대한민국 정부가 지정한 국가전략기술 사업인 반도체, 이차전지, 바이오 산업 분야에서 이를 대표하는 기업을 선정하고 기업과 관련된 뉴스 기사 댓글의 여론을 파악하여 기업에 대한 여론의 감성 점수 변화와 기업의 실제 주가의 흐름에 연관성이 있는지 알아보하고자 하였다. 각 산업 분야별 대표 기업은 2020년 8월 10일을 기준으로 한국거래소(KRX)의 코스피200 종목에서 반도체, 이차전지, 바이오 산업 분야 중 상장시가총액 상위2개 종목을 각각 선출하였다(한국거래소, 2020). 최종적으로 선정된 산업 분야별 대표 기업명은 <표 1>과 같다.

본 연구에서는 데이터 수집을 위해 <표 1>의 기업명을 질의어로 검색포털 사이트인 네이버에서 검색을 실시하였다. 수집 대상은 2020년 1월 1일부터 2020년 12월 31일 사이에 발행된 뉴스 기사 중 네이버 뉴스 기사만을 수집하여 그 댓글을 분석하였다. 파이썬 라이브러리인 BeautifulSoup 4.8.1 (Richardson, 2019)과 Selenium 3.141.0 (Muthukadan, 2018)을 사용하여 기사의 제목, 발행 날짜와 시간, url, 댓글의 발행 날짜와 내용을 크롤링하였다. 421,893개의 기사를 수집하였으나, 제목이 아닌 본문에 기업명이 존재하여 검색된 기사의 경우 해당 기업과의 연관성이 부족하다고 판단하여 검색된 기사 중 제목에 기업명을 포함하고 있는 기사 41,047개를 남기고 삭제하였다.

최종 기사의 댓글 중에서 url이나 '카카오톡 오픈 채팅방'을 지칭하는 '카톡방', '오픈방', '돈번방' 등을 포함하는 댓글의 경우 댓글의 목적이 광고인 것으로 판단하여 이를 삭제하였다. 따라서 기사에 달린 47,452개의 댓글을 최종적인 데이터 분석 대상으로 선정하였다. 최종적으로 선정된 기업별 뉴스 기사 및 댓글의 개수는 <표 2>와 같다.

선정된 47,452개의 댓글에는 산업 분야별로 감성 분석을 실시하였고, 감성 분석 결과값에 따라 긍정, 중립, 부정의 감성을 부여하였으며 그 결과는 <표 3>과 같다.

<표 1> 반도체, 이차전지, 바이오 산업 분야별 대표 기업명

산업 분야	기업명
반도체	삼성전자, SK하이닉스
이차전지	삼성SDI, LG화학
바이오	삼성바이오로직스, 셀트리온

〈표 2〉 기업별 뉴스 기사 및 댓글 개수

반도체	삼성전자	기사	8,465
		댓글	16,025
	SK하이닉스	기사	6,016
		댓글	3,301
이차전지	삼성SDI	기사	5,211
		댓글	1,013
	LG화학	기사	8,001
		댓글	9,221
바이오	삼성바이오로직스	기사	4,917
		댓글	3,702
	셀트리온	기사	8,437
		댓글	14,190

〈표 3〉 뉴스 기사 댓글의 산업 분야별 감성 분석 결과

반도체			이차전지			바이오		
긍정	중립	부정	긍정	중립	부정	긍정	중립	부정
3,319	10,756	5,251	1,935	5,933	2,366	4,084	10,301	3,507

각 산업 분야 모두 긍정 및 부정의 감성 보다는 중립의 감성이 차지하는 비율이 높았으며, 대부분의 경우 긍정 보다는 부정 감성 댓글의 비율이 더 높았으나, 바이오 분야의 댓글의 경우 부정 보다 긍정 댓글의 비율이 높은 결과가 나타났다.

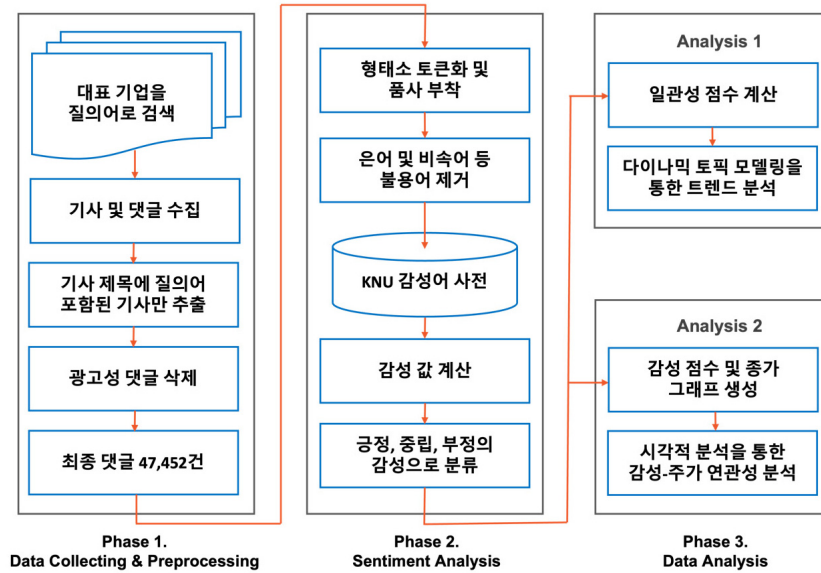
4. 연구 방법

본 연구는 반도체, 이차전지, 바이오 산업 분야에서 이를 대표하는 기업과 관련된 뉴스 기사 댓글을 수집하여 댓글에서 형성된 여론의 트렌드를 파악하고, 여론 감성의 변화와 실제 주가 흐름의 연관성을 파악하고자 하였다. 이를 위해 먼저 댓글의 감성 별 트렌드를 분석하기 위해

댓글의 감성 분석을 진행한 뒤 다이나믹 토픽 모델링을 실시하였다. 이후, 기업에 대한 여론의 감성 점수 변화와 기업의 주가 흐름이 연관성이 있는지를 알아보기 위해 감성 점수 그래프와 증가 그래프의 시각적 분석을 실시하였다. 전체 연구 과정은 다음 〈그림 1〉과 같다.

4.1 감성 분석을 통한 여론의 감성 점수 계산

본 연구에서는 수집된 기사의 댓글의 감성을 분석하였다. 파이썬 라이브러리인 KoNLPy(박은정, 조성준, 2014)의 오픈 소스 한국어 분석기 Okt를 사용하여 형태소를 토큰화하였다. 토큰화 된 각 단어에는 품사 정보를 부착하고 감성 분석을 실시하였다. 본 논문에서는 군산대



〈그림 1〉 전체 연구 과정

학교의 KNU 한국어 감성어사전을 이용하여 감성 점수를 부여하였는데, KNU 한국어 감성어 사전은 인간의 보편적인 기본 감정 표현을 나타내는 긍부정어로 구성되어 있다(박상민 외, 2018). 해당 사전에서 나타나는 보편적인 긍정 표현으로는 ‘감동받다’, ‘가치 있다’, 보편적인 부정 표현으로는 ‘그저 그렇다’, ‘~ 수 없다’ 등이 있다(박상민 외, 2018). 각 단어의 긍정 및 부정의 판별은 매우 부정(-2), 부정(-1), 중립(0), 긍정(1), 매우 긍정(2) 등의 리커트 척도로 이루어져 있다(박상민 외, 2018). 감성 점수 계산은 한 댓글에 부여된 점수를 모두 더하는 방식으로 진행하였으며 계산된 점수가 양수인 경우 긍정, 0인 경우 중립, 음수인 경우 부정의 감성을 부여하는 방식으로 진행했다. 예를 들어, “영업 이익률 최고네요. 엄청납니다. 매 분기 실적 증가 증이니 이번년도 기대됩니다.”라는 댓글의 경우 ‘이익’에 2점, ‘최고’에 2점, ‘기대’에

1점이 부여되어 감성 점수는 5점으로 합산되었고, 계산된 점수가 양수임에 따라 이 댓글은 긍정의 감성이 부여되었다. 감성 분류가 끝난 댓글에는 다이나믹 토픽 모델링을 이용하여 댓글의 시계열 주제 분석을 실시하였다. 이후 이를 바탕으로 시각적 분석 및 피어슨 상관계수를 이용한 댓글 감성 점수 변화와 주가 흐름 간의 연관성 분석을 진행하였다.

4.2 다이나믹 토픽 모델링을 이용한 산업별 트렌드 분석

다이나믹 토픽 모델링(Dynamic Topic Modeling)은 LDA 기반의 토픽모델링이다. LDA 기반의 토픽 모델링은 방대한 양의 비정형 문서 집합 내에서 핵심 토픽을 추출하는 확률 기반의 알고리즘(Wang & Blei, 2011)으로, 이에 시간이라는 변수를 추가한 것이 다이나믹

토픽 모델링이다(Blei & Lafferty, 2006). 따라서 다이나믹 토픽 모델링 기법을 통해 거시적인 차원의 주제별 분포를 비교적 쉽게 포착할 수 있으며(김일환, 2019), 시간의 흐름에 따라 두드러지는 토픽의 변화를 파악할 수 있다. 본 연구에서는 반도체, 이차전지, 바이오 산업 분야의 댓글이 감성 별로 어떠한 주제를 내포하고 있는지를 시간의 흐름에 따라 파악하기 위해 다이나믹 토픽 모델링을 실시하였다. 각 산업별 대표 기업의 기사에 달린 댓글의 감성을 분석하고 결과에 따라 댓글의 감성을 긍정, 중립, 부정으로 분류한 뒤 다이나믹 토픽 모델링 분석을 진행하였다. 다이나믹 토픽 모델링 분석 시 대입 해야할 일관성 토픽 개수를 산출하기 위해 파이썬 라이브러리인 gensim 3.8.1(Rehurek & Sojka, 2011)을 사용하여 2부터 10까지의 일관성을 계산하였다. 일관성 토픽 개수는 각 토픽의 개수 별 일관성 점수 그래프에서 기울기가 부정의 값이 되거나 0이 되기 전에 일관성 점수가 가장 높은 토픽의 개수로 채택하였다(Prabhakaran, 2018). 2개 이상의 토픽 개수가 적절하다고 판단되는 경우, 이를 모두 적용하여 다이나믹 토픽 모델링을 실시하고 분석 결과 문서 내에서 각 토픽의 독립성이 더 분명하게 나타나는 숫자를 일관성 토픽 개수로 채택하였다.

4.3 시각적 분석을 이용한 기업의 댓글 감성과 주가와의 연관성 분석

시각적 분석(Visual Analytics)은 인간의 지능을 이용해 데이터를 분석하여 복합적인 문제를 해결하는 추론 분석 방법이다(Thomas & Cook, 2006). 시각적 분석은 다양한 해석이 가

능한 모호한 데이터에서 납득할만한 결과를 추론해낼 수 있다(Keim et al., 2008). 본 연구에서는 댓글에서 나타나는 감성 점수가 실제 주가를 반영하는지를 알아보기 위해 댓글의 감성 점수와 기업의 실제 증가의 그래프를 그린 뒤 시각적 분석을 진행하였다. 시각적 분석을 위하여 감성 점수와 실제 주가 그래프를 작성한 후 첫째, 주가와 감성 점수의 상승 및 하락 트렌드가 일치하거나 일치하지 않는 구간을, 둘째, 감성 점수 또는 주가가 전일대비 매우 크게 상승하였거나 하락한 날짜를 파악하였다. 이후 해당 구간이나 날짜에 해당하는 기업의 사건 및 사고는 무엇이 있었는지를 뉴스, 기업 보고서, 정부 보고서 등을 통해 확인하였다. 마지막으로 연구원이 직접 해당 구간이나 날짜의 댓글을 확인함으로써 그에 해당하는 여론을 파악하여 데이터를 해석하였다.

5. 결과

5.1 다이나믹 토픽 모델링을 이용한 산업별 트렌드 분석

본 연구에서는 산업 분야별 대표 기업에 관련된 기사에 달린 댓글이 가지는 긍정, 부정, 중립의 감성 별 시계열 주제 흐름을 알아보기 위해 감성을 분석한 각 산업별 대표 기업 기사의 댓글에 다이나믹 토픽 모델링 기법을 실시하였다. 먼저 토픽 모델링을 돌리기 위한 토픽 개수 선정을 위해 각 분야의 감성 별 댓글의 일관성 토픽 개수 분석하였고, 그 결과는 <표 4>와 같이 나타났다.

〈표 4〉 각 산업 분야의 감성 별 일관성 토픽 개수 결과

산업 분야	감성	일관성 토픽 개수
반도체	긍정	3
	중립	4
	부정	3
이차전지	긍정	3
	중립	4
	부정	2
바이오	긍정	3
	중립	2
	부정	3

위의 감성 별 일관성 토픽 개수를 사용하여 반도체, 이차전지, 바이오 산업 분야의 댓글에 긍정, 중립, 부정의 감성 별 시계열 주제 분석을 진행하였다. 도출된 주제 토픽을 통해 각 댓글이 속할 확률 값을 도출하여 댓글이 가지는 최종 주제 토픽은 무엇인지를 측정하고 분기별로 각 주제 토픽이 차지하는 비율을 계산하였다.

수집한 댓글의 기간은 분기에 따라 네 구간으로 나누어 분기별 주제의 변화 흐름을 파악하고자 하였다. 1분기는 2020년 1월 1일부터 2020년 3월 31일까지, 2분기는 2020년 4월 1일부터 6월 30일까지, 3분기는 7월 1일부터 9월 30일까지, 4분기는 10월 1일부터 12월 31일까지로 설정하였다.

다이나믹 토픽 모델링으로 추출된 각 토픽 별 레이블링은 주영준 외(2019)의 연구에서 사용된 주제 토픽 선정 방법을 채택하였다. 주제는 주제어들과 각 주제어들이 해당 주제에 속할 수 있는 확률 값으로 구성되어 있기 때문에 각 키워드들은 모든 주제에 제각기 다른 확률 값을 가진 채 존재할 수 있게 된다(주영준 외, 2019). 주제 토픽을 구성하는 키워드가 다른 주제 토픽

에 중복으로 존재하는 것을 미연에 방지하기 위해 전체 키워드는 가장 높은 확률 값의 주제 토픽에 나타난 경우를 제외하고는 삭제하였다. 이후 모든 주제별 토픽의 상위 50개의 키워드에서 두 명의 연구자들이 독립적으로 레이블링을 실시한 뒤 상의를 통해 최종 주제 토픽을 선정하였다. 두 명의 연구자 간의 의견이 상이한 경우, 다른 연구자의 중재를 통해 합의를 진행하였다.

레이블링이 끝난 후, gensim으로 주제 토픽 별 확률 분포를 확인하여 가장 높은 확률 값을 각 댓글의 주제 토픽으로 선정하였다. 이를 통해 시간의 흐름에 따른 댓글 주제 분포를 확인하였다. 예를 들어, 이차전지 분야의 긍정 감성의 댓글 중 “우리나라에 전기차 배터리 최강 회사인 LG화학이 있어서 기대가 됩니다.”의 주제 토픽 확률을 확인해본 결과 확률 분포 값이 첫번째 토픽 “이차전지 관련 주 상승”은 약 0.0009%, 두번째 토픽 “최고의 기술력”은 약 99.8005%, 세번째 토픽 “개인 투자 성공”은 약 0.0009%로 나타남에 따라 이 댓글의 주제 토픽을 “최고의 기술력”으로 분류하였다.

이러한 과정을 통해, 각 산업 분야마다 댓글의

감성 별 주제 토픽을 산출하였다. 다음 <표 5>는 반도체 분야의 댓글에서 감성 별 주제 토픽 명 및 주요 키워드를 추출한 것이다.

추출된 주제 토픽을 통해 알아본 반도체 분야 기업 댓글의 기간에 따른 주제 토픽의 비율은 <그림 2>와 같다.

반도체 분야의 긍정 감성 댓글의 경우 “경제적 기여 호감”, “투자 성공 기쁨”, “국내 기업 자부심”의 주제 토픽이 나타났다. “경제적 기여 호감”이 전체 주제 분야 토픽의 모든 기간에서 1~2위로 나타남에 따라 국내 최고 수출액인 반도체 분야가 한국 경제에 기여를 하고 있다는 의견이 주를 이루고 있음을 파악할 수 있었다. “경제적 기여 호감”과 “국내 기업 자부

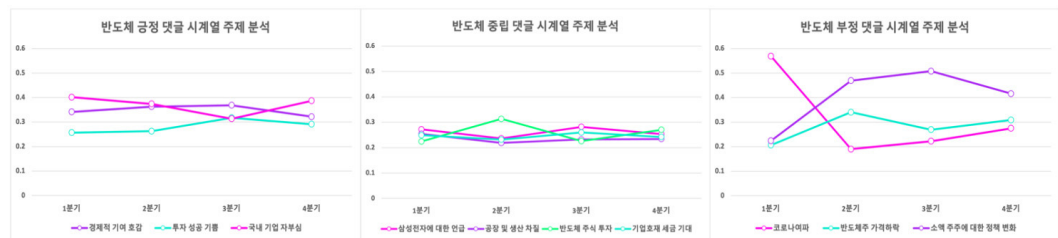
심”의 주제 비율은 반비례로 나타났는데 이는 기업의 반도체의 수출액 상승과 국내 기업이 세계 시장에서 갖는 위상이 국제 반도체 분야에서의 높은 매출이라는 같은 결과를 지칭하기 때문으로 유추하였다.

반도체 분야의 중립 감성의 댓글의 경우 “삼성전자에 대한 언급”, “공장 및 생산 차질”, “반도체 주 투자”, “기업 호재 세금 기대”의 주제 토픽이 나타났다.

전체 주제 토픽 중 “삼성전자에 대한 언급”이 대부분의 기간에서 가장 높은 비율로 나타났으며 총 댓글 건수도 가장 높았다. 이는 삼성전자가 국내 시가총액 1위기 때문에 사람들의 관심도가 높다는 점과 2020년 삼성전자 이재용

<표 5> 반도체 분야 기업 댓글의 감성 별 토픽 명, 주요 키워드

감성	토픽 명	주요 키워드
긍정	경제적 기여 호감	경제, 한국, 기술, 인정
	투자 성공 기쁨	수익, 최고, 주주
	국내 기업 자부심	좋은, 우리나라, 존경, 축하
중립	삼성전자에 대한 언급	삼성, 전자, 이재용, 반도체
	공장 및 생산 차질	중국, 공장, 노조
	반도체 주 투자	개인, 동학, 개미, 주식, 주가
	기업 호재 세금 기대	하이닉스, 세금, 경제, 정부
부정	코로나 여파	재앙, 중국, 코로나, 피해
	반도체 주 가격하락	개미, 삼전, 하이닉스, 가격
	소액 주주에 대한 정책 변화	세금, 투자, 3억, 정책, 주주



<그림 2> 반도체 분야 기업 댓글의 감성 별 시계열 주제 분석

부회장이 검찰에 소환되었던 사건이 영향을 미친 것이라고 유추하였다. “반도체 주식 투자”의 비율이 2분기에 급격히 증가하였는데, 이는 삼성전자와 SK하이닉스 모두 2분기 직전 낮아진 주가가 회복되지 않아 이에 대한 언급이 많아진 것으로 보인다.

반도체 분야의 부정 감성의 댓글의 경우 “코로나 여파”, “반도체 주 가격하락”, “소액 주주에 대한 정책 변화”의 주제 토픽이 나타났다. 특히 “코로나 여파”의 경우 코로나 바이러스가 막 유행하기 시작한 1분기엔 그 비율이 약 60%로 나타날 정도로 높게 나타났다. 삼성전자의 경우 1분기에 코로나19로 인한 수요 둔화가 반영되어 이에 대한 우려가 댓글에서도 나타난 것으로 보인다. 2분기부터는 마케팅 비 절감 등의 대안과 재택근무를 위한 모니터의 수요 증

가 등으로 영업 이익이 상승되는 결과를 통해 전체 부정 댓글에서 “코로나 여파”가 차지하는 비율이 10%대로 하락하였다. “소액 주주에 대한 정책 변화” 주제 토픽 비율은 지속적인 상승세를 보였는데, 이를 통해 주식 양도 소득세의 대주주 기준을 10억원에서 3억원으로 낮추고 과세 대상을 확산한다(손일선, 양연호, 2020)는 정부 방침 변화에 대한 반발 의견이 지속적으로 증가하고 있음을 유추하였다.

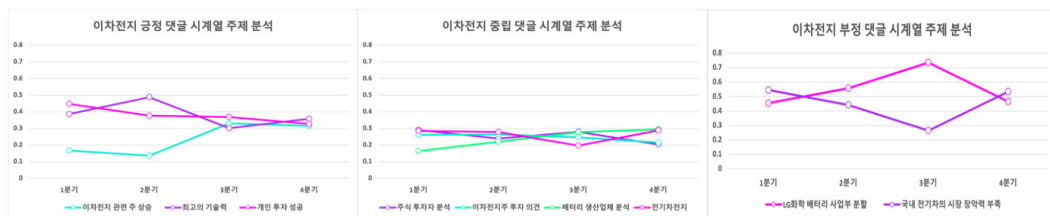
다음 <표 6>은 이차전지 분야의 댓글에서 감성 별 주제 토픽 명 및 주요 키워드를 추출한 것이다.

추출된 주제 토픽을 통해 알아본 이차전지 분야 기업 댓글의 기간에 따른 주제 토픽의 비율은 <그림 3>과 같다.

이차전지 분야의 긍정 감성의 댓글의 경우

<표 6> 이차전지 분야 기업 댓글의 감성 별 토픽 명, 주요 키워드

감성	토픽 명	주요 키워드
긍정	이차전지 관련 주 상승	주주, 주가, 투자, 이익, 가치
	최고의 기술력	배터리, 전기차, 기술, 세계, 최고
	개인 투자 성공	개미, 매수, 매도, 좋다, 감사합니다
중립	주식 투자자 분석	개미, 개인, 기관, 외국인, 증권사
	이차전지주 투자 의견	주가, 반대, 사라, 빨리
	배터리 생산업체 분석	LG화학, 삼성SDI, SK이노베이션, 배터리
	전기차전지	전기차, 테슬라, 현대자동차
부정	LG 화학 배터리 사업부 분할	LG화학, 기업, 사업, 분할, 분사
	배터리 기술력 의심	배터리, 문제, 기술, 전기차, 현대차, 테슬라



<그림 3> 이차전지 분야 기업 댓글의 감성 별 시계열 주제 분석

“이차전지 관련 주 상승”, “최고의 기술력”, “개인 투자 성공”의 주제 토픽이 나타났다. 주식 관련 주제 토픽인 “이차전지 관련 주 상승”과 “개인 투자 성공”의 증감이 비슷한 양상을 띄고 있었다. 실제 주가와 주제 토픽의 증감 시기를 비교해본 결과, 이차전지 분야의 경우, 주가가 하락하는 경우에는 긍정의 언급이 많아지고 주가가 상승하는 경우에는 부정의 언급이 많아짐을 알 수 있다. 이는 주가하락의 경우에는 주가 상승을 위한 응원과 긍정의 댓글을, 상승하는 경우에는 주가 하락에 대한 걱정과 우려의 댓글을 작성하기 때문으로 유추하였다. 이에 따라 가장 높은 주가 상승률을 보인 2분기에 “이차전지 관련 주 상승”과 “개인 투자 성공”의 비율이 가장 낮았으며, 대신 “최고의 기술력” 주제 토픽은 약 50%를 차지하며 가장 높은 비율을 차지했다.

이차전지 분야의 중립 감성의 댓글의 경우 “주식 투자자 분석”, “이차전지주 투자 의견”, “배터리 생산업체 분석”, “전기차전지”의 주제 토픽이 나타났다. “주식 투자자 분석”, “이차전지주 투자 의견”, “배터리 생산업체 분석”은 이차전지 관련 주식에 투자하기 위한 담론이 이루어진 것으로 보았다. “전기차전지”의 경우 LG 화학과 삼성SDI에서 제작한 배터리가 테슬라와 현대자동차 등의 전기차전지로 사용되어 이에 대한 언급이 나타난 것으로 유추하였다. 이는 이차전지 부정 댓글의 “배터리 기술력 의심”에 대한 댓글과 유사한 증감 추세를 보였는데, 배터리 관련 사건이 발생하면 그에 대한 언급 비율도 함께 증가하고 감소하기 때문으로 유추하였다.

이차전지 분야의 부정 감성의 댓글의 경우

“LG 화학 배터리 사업부 분할”, “국내 전기차 시장 장악력 부족”의 주제 토픽이 나타났다. “LG화학 배터리 사업부 분할”은 LG화학이 배터리 사업부문을 물적 분할한 사건에 대한 것으로 댓글에서 이에 대한 우려와 염려의 의견이 나타났다. 이는 실제 물적 분할이 이루어진 4분기까지 상승 추세를 보이다가 4분기인 12월 실제 물적 분할이 이루어지고 난 후엔 하락 추세를 보였다. 반면 “배터리 기술력 의심”의 경우 4분기에 높은 비율을 차지하였는데, 테슬라의 화재 소식과 현대차의 리콜 발표에 따라 LG화학 배터리에 대한 불신이 높아져 댓글에서 이에 대한 언급이 증가한 것으로 파악하였다.

다음 <표 7>은 바이오 분야의 댓글에서 감성 별 주제 토픽 명 및 주요 키워드를 추출한 것이다.

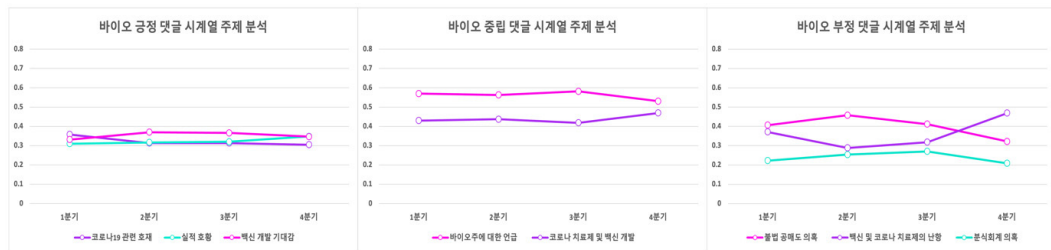
추출된 주제 토픽을 통해 알아본 반도체 분야 기업 댓글의 기간에 따른 주제 토픽의 비율은 <그림 4>와 같다.

바이오 분야의 긍정 감성 댓글의 경우 “코로나19 관련 호재”, “실적 호황”, “백신 개발 기대감”의 주제 토픽이 나타났다. 2019년 하반기에 코로나19의 대유행이 시작되고 국내 기업인 셀트리온이 코로나19 진단 키트 및 백신 개발 사업에 참여함에 따라 “코로나19 관련 호재”와 “백신 개발 기대감” 주제의 댓글 비율이 70%로 나타났다. 코로나19 상황으로 인해 의약품 매출이 증가에 대한 기쁨과 국내 1호 코로나19 백신 개발에 대한 기대가 댓글에서 나타남으로 유추하였다.

바이오 분야의 중립 감성 댓글의 경우 “바이오주에 대한 언급”, “코로나 치료제 및 백신 개발”의 주제 토픽이 나타났다. 코로나 19 상황으

〈표 7〉 바이오 분야 기업 댓글의 감성 별 토픽 명, 주요 키워드

감성	토픽 명	주요 키워드
긍정	코로나19 관련 호재	약, 승인, 진단, 키트
	실적 호황	이익, 영업, 매출, 실적
	백신 개발 기대감	코로나, 치료제, 개발, 항체, 백신, 화이팅, 응원, 성공
중립	바이오주에 대한 언급	주가, 주식, 주주, 개미, 매수
	코로나 치료제 및 백신 개발	코로나, 치료, 백신, 개발
부정	불법 공매도 의혹	공매도, 주가, 삼바, 주식, 셀트, 불법, 세력
	백신 및 코로나 치료제의 난항	코로나, 백신, 치료, 약, 효과, 부작용
	분식회계 의혹	삼성, 바이오로직스, 셀트리온, 분식회계, 사기, 범죄



〈그림 4〉 바이오 분야 기업 댓글의 감성 별 시계열 주제 분석

로 인해 바이오주에 대한 여론의 관심이 상승하였는데 그에 따라 “바이오주에 대한 언급”이 전체 기간 내내 50%가 넘는 비율로 나타났음을 유추하였다.

바이오 분야의 부정 감성 댓글의 경우 “불법 공매도 의혹”, “백신 및 코로나 치료제의 난항”, “분식회계 의혹”이 나타났다. “백신 및 코로나 치료제의 난항” 댓글의 경우 4분기에 급증하였는데, 백신 관련 긍정 주제 토픽인 “코로나19 관련 호재”, “백신 개발 기대감”의 비율은 하락함에 따라 연관성이 존재한다고 보았다. 이는 셀트리온에 대한 국내 코로나19 백신 1호 개발의 기대감이 높았는데, 화이자 및 모더나와 같은 해외 백신은 상용화 되고 국내 백신의 최종 승인은 이루어지지 않음에 따라 백신 개발에 대한 실망감이 반영된 것으로 유추하였다.

5.2 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 기업의 댓글 감성과 주가와와의 연관성 분석

본 연구는 먼저 반도체, 이차전지, 바이오 산업 분야 댓글의 감성이 실제 주가에 반영되었는지를 알아보기 위해 먼저 댓글의 감성 점수와 기업의 실제 증가에 대한 시각적 분석을 실시하였다. 댓글의 감성 점수는 중립 값 0(기준선)을 통해 긍정의 감성인 양수와 부정의 감성인 음수를 구분하였다. 각 날짜 별 감성 점수는 해당 날짜에 게시된 전체 댓글의 감성 점수를 모두 더한 것을 전체 댓글의 개수로 나눈 감성 점수의 평균 값으로 설정하였다. 계산된 감성 점수는 그래프로 나타낼 때 댓글이 존재하지 않아 공백으로 나타난 날짜는 감성 점수가 존재하는 날짜

의 그래프와 이어서 표현하여 그래프상으로 공백이 표현되지 않도록 하였다. 기업의 실제 증가는 한국 거래소(KRX)를 통해 각 기업의 종목 코드를 검색하여 2020년 1월 2일부터 12월 30일까지의 증가 데이터를 수집하였다. 수집된 데이터로 감성 점수의 그래프와 증가 그래프를 생성하여 이를 비교하였다. 증가의 경우 기본 축으로 지정하여 각 증가의 가격에 맞추어 축의 범위를 설정하였으며, 감성 점수의 경우 보조 축으로 지정하여 10과 -10 사이의 값으로 범위를 설정하였다. 주식 시장의 특성에 따라 공휴일의 주식 가격은 존재하지 않아 증가의 경우도 공백이 발생하였는데, 감성 점수와 마찬가지로 증가가 존재하는 날짜의 그래프와 존재하지 않는 날짜의 그래프를 이어서 표현하여 그래프 상으로 공백이 표현되지 않도록 하였다. 감성 점수와 증가의 그래프를 작성한 뒤에는 두 명이 연구원이 각자 시각적 분석을 실시하여 증가에 감성 점수가 반영되었다고 분석되는 구간을 분석한 뒤, 이후 각 연구원이 지정한 기간 중 중복되는 구간을 감성 점수가 증가에 반영되는 구간

으로 선정하였다. 도출된 유사 구간에는 이동 평균에 대한 피어슨 상관계수를 이용하여 감성 점수와 주가 흐름의 연관성을 확인하였다. 이동 평균은 10일을 간격으로 분석하였으나 10일의 이동 평균 값의 상관 분석으로 연관성을 보기 어려운 경우에는 이동 평균 기간을 늘려 상관 분석을 진행하였다. 주식시장의 특성상 공휴일과 주말에는 장을 열지 않고 댓글도 기업 관련 뉴스가 발행되지 않는 날이 존재했기 때문에 구간을 정확히 일치시켜 비교하기 어려운 경우엔 가장 근접한 날짜에 맞춰 분석을 실시하였다.

5.2.1 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 삼성전자의 댓글 감성과 주가와의 연관성 분석

삼성전자 댓글 및 실제 증가의 시각적 분석 결과는 <그림 5>와 같다.

시각적 분석을 통해 삼성전자는 2020년 1월 초부터 2020년 4월 말까지, 2020년 6월 초부터 2020년 12월 말까지의 감성 점수와 증가 그래프의 변화 흐름이 유사한 구간으로 도출하였다.



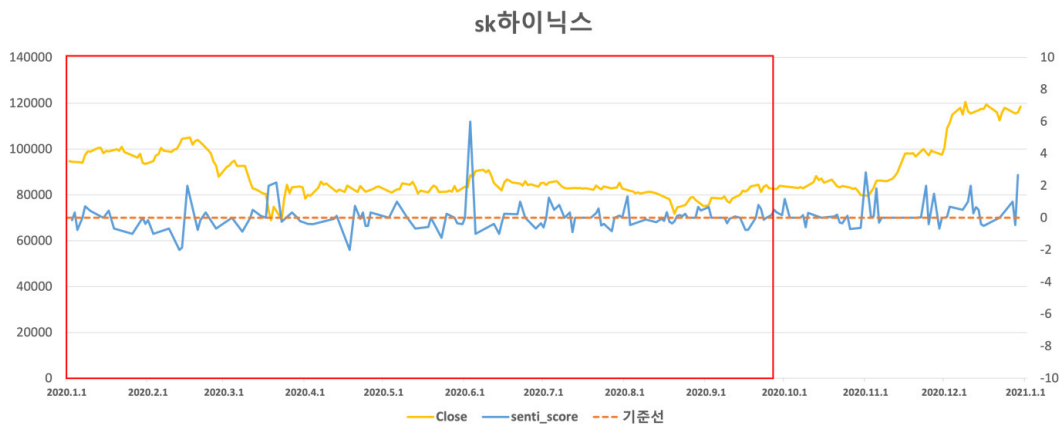
<그림 5> 삼성전자에 대한 댓글의 감성과 삼성전자의 실제 증가 비교

시각적 분석에서의 유사성은 세부적인 그래프의 항목 변화가 아닌 전체적인 흐름을 중심으로 확인하였다. 예를 들어 2020년 1월부터 4월 중순까지의 그래프는 완전히 일치하진 않아도 상승과 하락을 반복하는 유지하는 기간이 시각적으로 보았을 때 비슷하게 나타났다. 1월 초에 감성 점수의 그래프의 경우 기준선인 0에서 양의 방향으로 이동하였으며, 주가 또한 5만원 중반에서 6만원 초반으로 상승하였다. 1월 중반 이후 2월 초까지 살펴보았을 때, 감성 점수는 상승과 하락을 반복하다가 양수에서 음의 방향으로 이동하였으며, 주가 또한 상승과 하락 끝에 6만원 초반에서 5만원 대로 하락하였다. 이러한 양상에서 볼 때, 두 그래프의 모습이 크게 다르지 않았기 때문에 해당 기간의 변화 흐름이 유사하다고 보았다. 뒤이어 SK하이닉스, 삼성SDI, LG화학, 삼성바이오로직스, 셀트리온 역시 같은 과정으로 시각적 분석을 실시하였다. 삼성전자의 경우 10일을 기준으로 하는 이동 평균 상관분석에서 상관계수가 낮은 것으로 파악 되어 30일을 기준으로 피어슨 상관 계수를

산출하였다. 2020년 1월 초부터 2020년 4월 말까지 이동 평균을 분석한 결과 상관계수가 약 0.71로 나타났고, 2020년 6월 초부터 2020년 12월 말까지의 상관계수는 약 0.76으로 나타나 시각적 분석으로 추출한 유사 구간 모두에 강한 상관관계가 있는 것으로 보였다. 유사 구간과 연구자의 지식을 통해 그래프를 해석한 결과 삼성전자는 대부분의 감성 점수가 2와 -2 값 사이에서 나타나는 완만한 흐름을 보였다. 증가가 하락하는 시기에도 대부분의 경우 감성 점수가 크게 변화하지 않았는데 이는 삼성전자가 높은 상장시가총액을 보유하고 있다는 점과 삼성전자는 우량주에 대한 믿음이 작용하여 주가가 하락해도 여론이 크게 동요하지 않음으로 확인하였다.

5.2.2 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 SK 하이닉스의 댓글 감성과 주가와의 연관성 분석

SK 하이닉스 댓글 및 실제 증가의 시각적 분석 결과는 <그림 6>과 같다.



<그림 6> SK하이닉스에 대한 댓글의 감성과 SK하이닉스의 실제 증가 비교

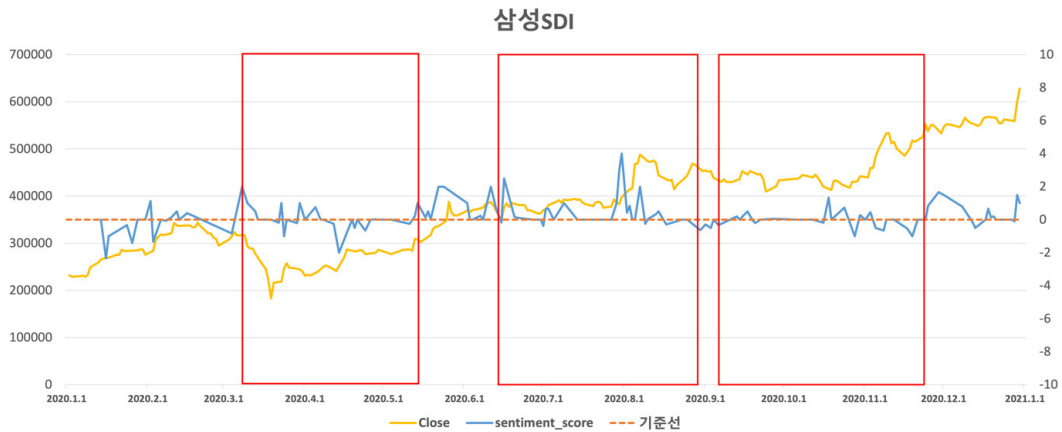
시각적 분석을 통해 SK하이닉스는 2020년 1월 초부터 2020년 9월 말까지의 감성 점수와 증가 그래프의 변화 흐름이 유사하다고 보았다. SK하이닉스의 경우 10일을 기준으로 하는 이동 평균 상관분석에서 상관계수가 낮은 것으로 파악 되어 20일을 기준으로 피어슨 상관계수를 산출하였다. SK하이닉스의 이동 평균에 따른 상관관계 분석으로 2020년 1월 초부터 2020년 9월 말까지 이동 평균을 분석한 결과 상관계수가 약 0.75로 나타나 이 구간에 강한 상관관계가 있는 것으로 보았다. 유사 구간과 연구자의 지식을 통해 그래프를 해석한 결과 전체적으로 증가가 하락할 때에도 감성의 변화가 크지 않는데 이는 삼성전자와 마찬가지로 기업의 높은 상장시가총액으로 인한 신뢰로 유추하였다. 6월 3일 자에 나타난 높은 긍정 감성 점수의 경우, 댓글을 통해 SK하이닉스의 NAND 사업에 대한 기대치가 높게 평가되었기 때문으로 확인하였다. 2020년 11월 1일자의 경우 주가는 하락하였지만 감성 점수는 상승하였는데, 이는 주가 하락에 대한 응원의 댓글이 감성 점

수에 반영되었음으로 파악하였다.

5.2.3 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 삼성SDI 댓글 감성과 주가와의 연관성 분석

삼성SDI 댓글 및 실제 증가의 시각적 분석 결과는 <그림 7>과 같다.

시각적 분석을 통해 삼성SDI는 2020년 3월 중반부터 2020년 5월 말까지, 2020년 6월 중반부터 8월 말까지, 2020년 9월 말부터 2020년 11월 말까지의 감성 점수와 증가 그래프의 변화 흐름이 유사하다고 보았다. 삼성SDI의 경우 10일을 기준으로 이동 평균에 따른 피어슨 상관계수를 산출하였다. 2020년 3월 중반부터 2020년 5월 말까지는 약 0.76의 상관계수가 나타났고, 2020년 6월 중반부터 2020년 8월 말까지는 약 0.78의 상관계수가 나타났으며, 2020년 9월 말부터 2020년 11월 말까지는 약 0.82의 상관계수가 나타나 시각적 분석으로 추출한 유사 구간 모두에 강한 상관관계가 있는 것으로 보았다. 유사 구간과 연구자의 지식을 통해 그래



<그림 7> 삼성SDI에 대한 댓글의 감성과 삼성SDI의 실제 증가 비교

프를 해석한 결과 삼성전자와 마찬가지로 2와 -2값 사이의 높지 않는 완만한 감성 변화 폭을 보였다. 이는 삼성 계열의 삼성SDI까지도 삼성 전자에 대한 신뢰성이 작용하고 있음으로 유추하였다. 실제로 주가가 4만원 가까이 하락한 3월 19일자 댓글의 감성 점수의 경우 0(중립)으로 큰 동요가 없으므로 확인하였다.

5.2.4 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 LG화학 댓글 감성과 주가와의 연관성 분석

LG화학 댓글 및 실제 증가의 시각적 분석 결과는 <그림 8>과 같다.

시각적 분석을 통해 LG화학은 2020년 2월 초부터 5월 말까지, 2020년 7월 중반부터 10월 말까지의 감성 점수와 증가 그래프의 변화 흐름이 유사하다고 보았다. LG화학의 경우 10일을 기준으로 하는 이동 평균 상관분석에서 상관계수가 낮은 것으로 파악되어 20일을 기준으로 피어슨 상관 계수를 산출하였다. LG화학의 이동 평균에 따른 상관관계 분석으로 2020년 2

월 초부터 2020년 5월 말까지는 상관계수가 약 0.7로 나타났고, 2020년 7월 중반부터 2020년 10월 말까지는 상관계수가 약 0.77로 나타나 시간적 분석으로 추출한 유사 구간 모두에 강한 상관관계가 있는 것으로 보였다. 유사 구간과 연구자의 지식을 통해 그래프를 해석한 결과 LG화학의 감성 점수는 다른 기업에 비해 변화가 잦았는데, 이는 이차 전지의 대표적인 상품인 배터리가 시장의 흐름에 영향을 많이 받는 제품이기 때문에 주주들 또한 배터리 뉴스 기사에 즉각적으로 반응하기 때문이라고 추측하였다. 20만원 상당의 큰 주가 하락과 상승이 발생한 2020년 3월 중반과 2020년 8월 초의 경우, 하락과 상승이 발생하기 전 감성의 변화가 다른 기간에 비해서 특히 잦으며 그 폭도 매우 컸다. 전지는 LG화학의 사업부 물적 분할 및 회장 사임의 사건이 발생하여 하락세를 보였으며, 후자는 LG화학이 2분기 실적에서 2019년 대비 131% 급증하고 세계 전기자동차 배터리 시장에서 1위를 차지하는 사건을 통해 상승세를 보였다. 이를 통해 2020년의 LG 화학 주가는 20



<그림 8> LG화학에 대한 댓글의 감성과 LG화학의 실제 증가 비교

만원 상당의 주가 하락 및 상승이 발생하기 전에 담론 감성 점수 그래프의 방향 변화가 찾아짐을 확인할 수 있었다.

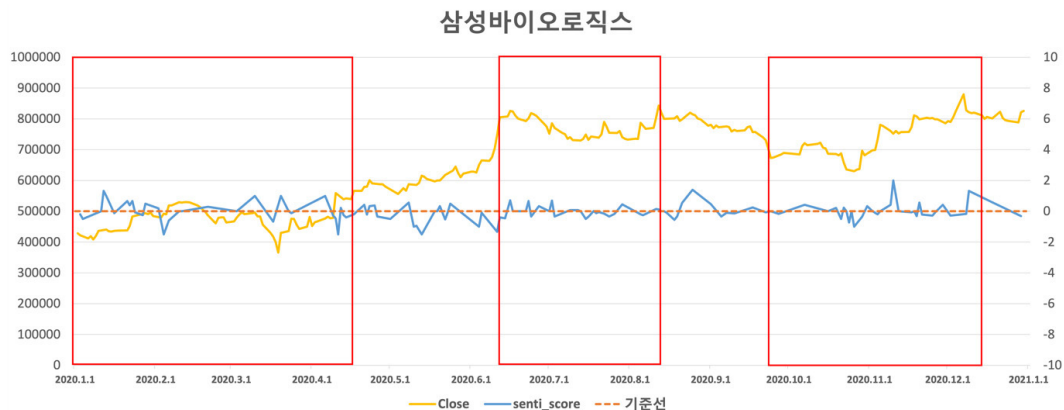
5.2.5 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 삼성바이오로직스 댓글의 감성과 주가와의 연관성 분석
 삼성바이오로직스 댓글 및 실제 증가의 시각적 분석 결과는 <그림 9>와 같다.

시각적 분석을 통해 삼성바이오로직스는 2020년 1월 초부터 4월 중반, 2020년 6월 중반부터 8월 중반, 2020년 9월 말부터 12월 중반까지의 감성 점수와 증가 그래프의 변화 흐름이 유사하다고 보았다. 삼성SDI의 경우 10일을 기준으로 이동 평균에 따른 피어슨 상관계수를 산출하였다. 2020년 1월 초부터 2020년 4월 중반까지는 약 0.64의 상관계수가 나타나 이 구간에는 상관관계가 있음으로 보았으며, 2020년 6월 중반부터 2020년 8월 중반까지는 약 0.86의 상관계수가 나타나 이 구간에는 강한 상관관계가 있음으로 보았다. 2020년 9월 말부터 2020년 12월 중반까

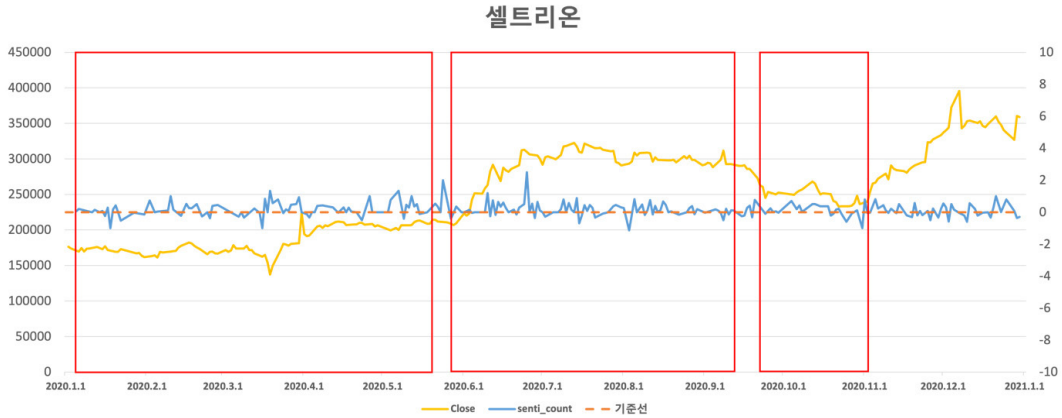
지는 상관계수가 약 0.56으로 나타나 이 구간에 상관관계가 있는 것으로 보았다. 유사 구간과 연구자의 지식을 통해 그래프를 해석한 결과 삼성바이오로직스도 삼성전자, 삼성SDI와 마찬가지로 전체 감성 점수가 2와 -2값 사이의 높지 않는 완만한 감성 변화 폭을 보였다. 또, 증가에 비해 감성 점수가 높게 측정된 1월 13일의 댓글에서는 삼성에 대한 지지와 신뢰의 의견으로 감성 점수가 높게 나타났다. 이는 삼성 계열사인 삼성SDI까지도 삼성 전자에 대한 신뢰성이 작용하고 있기 때문으로 유추하였다.

5.2.6 시각적 분석 및 이동 평균에 따른 피어슨 상관계수를 이용한 셀트리온 댓글의 감성과 주가와의 연관성 분석
 셀트리온 댓글 및 실제 증가의 시각적 분석 결과는 <그림 10>과 같다.

시각적 분석을 통해 셀트리온은 2020년 1월 초부터 5월 후반, 2020년 6월 초부터 9월 중반, 2020년 10월 초부터 11월 초까지의 감성 점수와 증가 그래프의 변화 흐름이 유사



<그림 9> 삼성바이오로직스에 대한 댓글의 감성과 삼성바이오로직스의 실제 증가 비교



〈그림 10〉 셀트리온에 대한 댓글의 감성과 셀트리온의 실제 증가 비교

하다고 보았다. 셀트리온의 경우 10일을 기준으로 이동 평균에 따른 피어슨 상관계수를 산출하였다. 2020년 1월 초부터 5월 후반까지는 상관계수가 약 0.86으로 나타나 강한 상관관계를 보였으며, 2020년 6월 초부터 9월 중반까지는 상관계수가 약 0.77로 나타나 상관관계가 있는 것으로 보았다. 마지막으로 2020년 10월 초부터 2020년 11월 초까지의 경우 상관계수가 약 0.77로 나타나 이 구간에는 강한 상관관계가 있는 것으로 보았다. 유사 구간과 연구자의 지식을 통해 그래프를 해석한 결과 셀트리온의 감성 점수는 다른 기업과 비교하여 잦은 변화를 보였는데, 이는 코로나19라는 특수한 상황 속에서 코로나19 백신 및 치료제 개발 사업을 진행하고 있는 셀트리온에 사람들의 관심이 상승하였기 때문이라고 추측하였다. 2020년 하반기 말인 11월 중반부터 감성 점수의 변화가 특히 잦았는데, 이는 셀트리온이 코로나 항체 치료제인 '렉키로나주'의 승인을 신청함에 따라 코로나 치료제 개발에 대한 기대감이 반영된 것으로 유추하였다.

6. 결론 및 제언

본 연구는 대한민국 정부가 지정한 국가 전략 기술 사업인 반도체, 이차전지, 바이오 산업 분야에서 이를 대표하는 기업을 선정하고 기업과 관련된 뉴스 기사 댓글의 여론을 파악하여 기업에 대한 여론의 주제 트렌드 흐름 및 여론의 감성과 실제 주가와의 연관성을 알아보려고 하였다. 이를 위해 포털사이트인 네이버에서 반도체, 이차전지, 바이오 분야를 대표하는 기업을 키워드로 검색하고 전처리를 거쳐 총 421,893개의 기사를 수집하였다. 기사에 달린 41,047개의 댓글을 대상으로 데이터 분석을 실시하였고 이를 통해 첫번째, 감성 분석을 통한 현재 트렌드를 살펴본 결과 국내에서는 삼성이라는 대기업에 대한 신뢰 여론이 형성되어 있음을 알 수 있었다. 삼성SDI나 삼성바이오로직스가 큰 수익을 내지 못하는 시기에도 삼성 계열사라는 이유로 신뢰를 가지는 여론이 생성되어 감성 폭이 크게 동요하지 않는 것이 나타났다. 또한, 삼성전자의 경우 그 위상이 전세계적이며 삼성전

자의 본사가 한국에 위치하고 있다는 것으로 큰 자부심을 가지고 있음이 긍정 감성 주제 트렌드 분석에서 나타났다. 이를 통해 수집 기간의 여론의 경우 삼성이라는 대기업에 대한 신뢰가 형성되어 있음을 유추할 수 있었다. 두번째, 시각적 분석기법을 통한 댓글의 감성과 주가의 비교를 통해 댓글의 감성 점수 변화와 주가의 흐름이 연관성을 띠고 있음을 유추할 수 있었다. 이를 통해 감성 점수와 주가는 서로 연관성이 있다고 분석하였다. 예외적으로 바이오 산업의 기업 셀트리온의 경우 2020년 하반기 말에 주가의 변화에 비해서 감성 점수의 변화가 매우 잦았는데, 이는 셀트리온이 개발 중인 코로나 치료제에 대한 댓글에서의 갑론을박이 심하게 나타났기 때문에 예외적인 현상이 발생했으므로 유추하였다.

본 연구에서는 다음과 같은 한계점을 지니고 있다. 첫째, 코로나19라는 특수한 상황 기간 내에 댓글의 주제 분석을 실시하였기 때문에 기업에 대한 지엽적인 여론의 흐름만을 분석하였다. 그렇기 때문에 조사 기간을 확장한 데이터 수집을 진행하여 각 산업별 대표 기업에 대한 일반적인 여론을 파악하는 것이 필요하다.

둘째, 시각적 분석을 이용하여 감성 점수와 주가의 연관성 구간을 도출했기 때문에 분석 결과가 다소 주관적일 수 있다는 한계점을 지

닌다. 본 연구의 경우, 댓글의 내용 주제를 바탕으로 댓글의 감성 점수와 주가의 흐름이 연관성이 있음을 파악하였다. 이를 위해서는 연구자의 도메인 지식이 필요하였기에, 빅데이터 분석 결과에 대한 해석을 통해 연관성을 확인한 뒤 그 구간에 대한 이동 평균의 상관관계를 분석하였다. 따라서 매일 종가에 대한 상관관계 분석을 진행했을 때와 비교하여 세밀성과 신뢰도 부분에서 한계점을 지닌다.

그러나 본 연구는 다음과 같은 기존 연구와의 차별성이 존재한다. 첫째, 기존의 연구에서는 트위터 또는 커뮤니티 중심의 분석이 진행되었으나 본 연구에서는 대표 기업에 대한 기사의 댓글 분석으로 실시하였다. 기사 댓글의 특성상 기사 내용과 연관성 있는 주제로 구성되어 있어 기존 연구에서 주로 사용된 트위터 및 커뮤니티 데이터보다 주제에 적합한 데이터로 분석을 진행하였다는 의의를 지닌다. 둘째, 본 연구는 코로나19 기간에 속하는 여론을 파악하였다. 따라서 코로나19라는 특수한 상황 속에서의 여론 트렌드 변화를 파악할 수 있다는 의의를 지닌다. 세번째, 국가전략기술제도에 속하는 반도체, 이차전지, 바이오 산업 분야의 대표 기업에 대한 여론을 분석함으로써 국가전략기술 연구의 근거를 제공하였다는 의의를 지닌다.

참 고 문 헌

- 기획재정부 (2021). 2022년부터 이렇게 달라집니다. 출처: <https://whatsnew.moef.go.kr/mec/ots/dif/main.do>
- 김건태, 황수성 (2015. 5). 국내 일간신문이 주가 수익률에 미치는 영향 및 투자자별 성과 분석. 한국재

- 무학회 학술대회, 1810-1841.
- 김다예, 이영인 (2018). Word2Vec을 활용한 뉴스 기반 주가지수 방향성 예측용 감성 사전 구축. 한국 빅데이터학회 학회지, 3(1), 13-20. <https://doi.org/10.36498/kbigdt.2018.3.1.13>
- 김동영, 박제원, 최재현 (2014). SNS와 뉴스기사의 감성분석과 기계학습을 이용한 주가예측 모형 비교 연구. 한국IT서비스학회지, 13(3), 211-233. <https://doi.org/10.9716/KITS.2014.13.3.221>
- 김세완, 박지원, 김영민, 함회경 (2020). 소셜 감성이 개별 기업 주식수익률에 미치는 비대칭적 영향 분석. Information Systems Review, 22(4), 59-74. <https://doi.org/10.14329/isr.2020.22.4.059>
- 김원정, 한정호 (2016). 조직 위기상황 시 보도기사에 달린 댓글들이 조직 위기책임성과 구매의도에 미치는 영향. 광고학연구, 27(8), 109-137.
- 김유신, 김남규, 정승렬 (2012). 뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형. 지능정보연구, 18(2), 143-156. <https://doi.org/10.13088/jiis.2012.18.2.143>
- 김일환 (2019). 인문학을 위한 신문 빅 데이터와 텍스트 마이닝. 어문론집, 78, 41-62. <https://doi.org/10.15565/jll.2019.06.78.41>
- 김지룡, 한은경 (2019). 기업평판과 수익성에 관한 연구 온라인 뉴스와 뉴스댓글을 중심으로. 디지털융복합연구, 17(9), 399-406. <https://doi.org/10.14400/JDC.2019.17.9.399>
- 남달우, 박진우, 김민경, 조현, 김성희 (2012). 인터넷 주식게시판을 통한 집단지성과 주식시장과의 상관관계 연구. 인터넷전자상거래연구, 12(2), 149-164.
- 박미화, 김솔 (2017). 온라인 뉴스 댓글 분석을 통해 본 한국 사회의 중국동포 혐오증: 오원춘 사건 보도 전후를 중심으로. 다문화와 평화, 11(3), 92-117. <https://doi.org/10.22446/mnpisk.2017.11.3.005>
- 박상민, 나철원, 최민성, 이다희, 온병원 (2018). Bi-LSTM 기반의 한국어 감성사전 구축 방안. 지능정보연구, 24(4), 219-240. <https://doi.org/10.13088/jiis.2018.24.4.219>
- 박성우 (2021. 12. 31). 반도체 · 배터리 등 '국가전략기술' 신설... 대기업 투자공제 2배↑. 조선비즈, 출처: https://biz.chosun.com/policy/policy_sub/2021/12/31/3RAH5ZVKWNFDDCXCRGNKPWE6PU/
- 박성욱, 권오병, 나형중 (2017). Big Data중 비정형Data인 소셜미디어가 기업의 매출액에 미치는 효과. 회계정보연구, 35(2), 321-342.
- 박은정, 조성준 (2014). KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지. 제26회 한글 및 한국어 정보처리 학술대회 논문집, 4-7.
- 배지양 (2009). 댓글의 품질 및 방향성이 영리기업의 사회공헌활동에 대한 공중의 책임성인식, 태도, 구매의도, 구전활동의도에 미치는 영향. 광고학연구, 20(5), 7-37.
- 손일선, 양연호 (2020. 10. 8). 대주주 주식 양도세 3억 기준...동학개미 반발에 "2년 유예". 매일경제, 출처: <https://www.mk.co.kr/news/economy/view/2020/10/1033433/>

- 양혜승 (2008). 인터넷 뉴스 댓글의 견해와 품질이 독자들의 이슈에 대한 태도에 미치는 영향. *한국언론학보*, 52(2), 254-281.
- 이미경, 최인호, 정세훈 (2013). 기업 페이스북에서 긍정·부정 댓글이 소비자들의 태도, 규범 및 구매 행동에 미치는 효과. *한국언론학보*, 57(4), 51-71.
- 이병현, 최일영, 전정우, 최이권, 김재경 (2020). 온라인뉴스 텍스트 마이닝을 이용한 기업 평판과 수익성에 관한 연구. *경영과학*, 37(4), 55-66. <https://doi.org/10.7737/KMSR.2020.37.4.055>
- 장은아, 최희련, 이홍철 (2020). BERT를 활용한 뉴스 감성분석과 거시경제지표 조합을 이용한 주가지수 예측. *한국컴퓨터정보학회논문지*, 25(5), 47-56. <https://doi.org/10.9708/jksoci.2020.25.05.047>
- 조수영, 장혜지, 권구민 (2012). 기업 위기에 대한 온라인 공중 반응. *한국언론학보*, 56(4), 311-338.
- 주영준, 김동훈, 이창호, 이용정 (2019). 페이스북 그룹 게시물 분석을 통한 우울증 관련 주제에 대한 고찰. *한국문헌정보학회지*, 53(4), 171-187. <https://doi.org/10.4275/KSLIS.2019.53.4.171>
- 한국거래소 (2020. 8. 10). 지수구성종목. 출처: <http://www.krx.co.kr>
- Blei, D. M. & Lafferty, J. D. (2006). Dynamic Topic Models. *Proceedings of the 23rd International Conference on Machine Learning*, 113-120.
- Fu, T. C., Lee, K. K., Sze, D., Chung, F. L., & Ng, C. M. (2008). Discovering the correlation between stock time series and financial news. *Proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 880-883. <https://doi.org/10.1109/WIAT.2008.228>
- Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: definition, process, and challenges. In *Information Visualization*, 154-175. https://doi.org/10.1007/978-3-540-70956-5_7
- Muthukadan, B. (2018). Selenium 3.141.0 [Online App]. Available: <https://selenium-python.readthedocs.io/>
- Prabhakaran, S. (2018). Topic Modeling with Gensim (Python). Available: <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/>
- Rehurek, R. & Sojka, P. (2011). Gensim-python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).
- Richardson, L. (2019). BeautifulSoup (4.8.1) [Online App]. Available: <https://beautiful-soup-4.readthedocs.io/>
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Thomas, J. J. & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and*

Applications, 26(1), 10-13. <https://doi.org/10.1109/MCG.2006.5>

Wang, C. & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 448-456. <https://doi.org/10.1145/2020408.2020480>

<p>• 국문 참고문헌에 대한 영문 표기 (English translation of references written in Korean)</p>

Bae, Ji Yang (2009). The influence of online news replies on netizens' CSR perception, attitude, purchase intention, and word of mouth intention: focusing on corporate social responsibility activity. *The Korean Journal of Advertising*, 20(5), 7-37.

Jang, Eun A, Choi, Hoe Ryeon, & Lee, Hong Chul (2020). Stock prediction using combination of BERT sentiment analysis and macro economy index. *Journal of the Korea Society of Computer and Information*, 25(5), 47-56. <https://doi.org/10.9708/jksci.2020.25.05.047>

Jo, Su Young, Jang, Hye Ji, & Kwon, Gu Min (2012). Netizens' reactions toward corporate crises: content analysis of crisis news stories and online readers' comments. *Korean Journal of Journalism & Communication Studies*, 56(4), 311-338.

Kim, Da Ye & Lee, Young Gin (2018). News based stock market sentiment lexicon acquisition using Word2Vec. *The Korea Journal of BigData*, 3(1), 13-20.
<https://doi.org/10.36498/kbigdt.2018.3.1.13>

Kim, Dong Young, Park, Je Won, & Choi, Jae Hyeon (2014). A comparative study between stock price prediction models using sentiment analysis and machine learning based on SNS and news articles. *Journal of Information Technology Services*, 13(3), 221-233.
<https://doi.org/10.9716/KITS.2014.13.3.221>

Kim, Gun Tae & Hwang, Su Seong (2015, May). Korea daily newspaper impact on the price-earnings ratio and investment outcome of investors. Conference on The Korean Finance Association, 1810-1841

Kim, Il Hwan (2019). Newspaper big data and text mining for digital humanities. the society of chung-ang lang. & Lit, 78 41-62. <https://doi.org/10.15565/jll.2019.06.78.41>

Kim, Ji Ryong & Han, Eun Kyung (2019). A study on corporate reputation and profitability focus on online news and comments. *Journal of Digital Convergence*, 17(9), 399-406.
<https://doi.org/10.14400/JDC.2019.17.9.399>

Kim, Se Wan, Park, Ji Won, Kim, Young Min, & Ham, Hee Kyung (2020). Asymmetric effect

- of social sentimental on an individual stock price return. *Information Systems Review*, 22(4), 59-74. <https://doi.org/10.14329/isr.2020.22.4.059>
- Kim, Won Jung & Han, Jeong Ho (2016). The effects of posted comments tagged to newspaper articles to readers' perceived crisis responsibility and their purchase intention. *The Korean Journal of Advertising*, 27(8), 109-137.
- Kim, Yu Shin, Kim, Nam Gyu, & Jeong, Seung Ryeol (2012). Stock-index invest model using news big data opinion mining. *Korea Intelligent Information Systems Society*, 18(2), 143-156. <https://doi.org/10.13088/jiis.2012.18.2.143>
- Korea Exchange. (2020, August 10). Index composition. Available: <http://www.krx.co.kr>
- Lee, Byung Hyun, Choi, Il Young, Jeon, Jeong Woo, Choi, Ikwon, & Kim, Jae Kyung (2020). A study on corporate reputation and profitability using online news text mining. *Korean Management Science Review*, 37(4), 55-66. <https://doi.org/10.7737/KMSR.2020.37.4.055>
- Lee, Mi Kyung, Choi, In Ho, & Jeong, Se Hoon (2013). The effect of positive and negative comments on consumers' attitudes, norms, and purchasing behavior on corporate Facebook. *Korean Society for Journalism and Communication Studies*, 57(4), 51-71.
- Ministry of Economy and Finance (2021). It's going to change like this starting 2022. Available: <https://whatsnew.moef.go.kr/mec/ots/dif/main.do>
- Nam, Dal Woo, Park, Jin Woo, Kim, Min Kyung, Jo, Hyeon, & Kim, Seong Hee (2012). A study about correlation between collective intelligence on the internet stock message board and stock market. *Korea Internet Electronic Commerce Association*, 12(2), 149-164.
- Park, Eun Jeong & Cho, Sung Zoon (2014). KoNLPy: Korean natural language processing in python. *Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology*, 4-7.
- Park, Mi Hwa & Kim, Sol (2017). A study on analysis of korean society's xenophobia towards korean-chinese through comments of online news: focused on before and after oh won-chun news report. *Multiculture & Peace*, 11(3), 92-117. <https://doi.org/10.22446/mnpisk.2017.11.3.005>
- Park, Sang Min, Na, Cheol Won, Choi, Min Seong, Lee, Da Hee, & On, Byeong Won (2018). KNU Korean sentiment lexicon: Bi-LSTM-based method for building a Korean sentiment lexicon. *Journal of Intelligence and Information Systems*, 24(4), 219-240. <https://doi.org/10.13088/jiis.2018.24.4.219>
- Park, Seong Woo (2021, December 31). New 'national strategic technology' such as semiconductors and batteries... Double the investment deduction for large companies ↑. *Chosunbiz*, Available:

https://biz.chosun.com/policy/policy_sub/2021/12/31/3RAH5ZVKWNFDDCXCRGNKPWE6PU/

- Park, Seong Wook, Kwon, Oh Byeong, & Na, Hyung Jong (2017). The effect of social media related to informal data of big data on sales. Korea Accounting Information Association, 35(2), 321-342.
- Son, Il Sun & Yang, Yeon Ho (2020, October 8). Based on the stock transfer tax of 300 million won...It was called a "two-year postponement" due to the opposition of Donghak Ant. Maeil Business Newspaper, Available:
<https://www.mk.co.kr/news/economy/view/2020/10/1033433/>
- Yang, Hye Seung (2008). The effects of the opinion and quality of user postings on internet news readers' attitude toward the news issue. Korean Society For Journalism And Communication Studies, 52(2), 254-281.
- Zhu, Yong Jun, Kim, Dong Hoon, Lee, Chang Ho, & Lee, Yong Jeong (2019). Investigating major topics through the analysis of depression-related facebook group posts. Journal of the Korean Library and Information Science, 53(4), 171-187.
<https://doi.org/10.4275/KSLIS.2019.53.4.171>