

# 텍스트마이닝 기법을 활용한 허위·과장광고 관련 기사의 트렌드 분석 (1990-2019)

## Analyzing the Trend of False·Exaggerated Advertisement Keywords Using Text-mining Methodology (1990-2019)

김도희\*, 김민정\*\*

숙명여자대학교 빅데이터분석융합학\*, 숙명여자대학교 소비자경제학과\*\*

Do-Hee Kim(kimasdohee@sookmyung.ac.kr)\*, Min-Jeong Kim(min-jeong.kim@sookmyung.ac.kr)\*\*

### 요약

본 연구는 텍스트마이닝 기법을 사용하여 1990년부터 2019년까지 5,141건의 신문기사에서 '허위·과장광고' 용어의 트렌드를 분석하였다. 우선 전체 신문 기사를 대상으로 빈도 분석을 통해 허위·과장광고의 최빈 키워드와 추출된 키워드 간의 맥락을 확인하고자 하였다. 다음으로 허위·과장광고가 어떻게 변화해왔는지에 대해 고찰하기 위해 10년 단위로 기사를 분리하여 빈도 분석을 수행하였고, 연도별 최빈 키워드를 주제로 한 학술논문 수와 비교하여 해당 시기에 이슈가 된 키워드가 연구로까지 이어진 경향성을 파악하였다. 마지막으로 토픽모델링 분석을 통해 토픽 내 세부 키워드를 바탕으로 허위·과장광고의 동향을 제시하였다. 연구 결과, 특정 시점에 이슈가 되었던 주제가 최빈 키워드로 추출되었고 시대별 키워드 트렌드는 사회적, 환경적 요인과 연관되어 변화함을 확인하였다. 본 연구는 소비자들이 부당광고에 대한 배경지식을 함양함으로써 현명한 소비를 이어 나갈 수 있도록 도움을 주는 데 의의가 있다. 더욱이 핵심 키워드 추출을 통해 위법행위를 저지른 기업 및 관련 종사자들에게 광고의 참된 목적을 제시하고, 시사점을 전달할 수 있을 것이라 기대한다.

■ 중심어 : | 허위광고 | 과장광고 | 텍스트마이닝 | 빈도분석 | 토픽모델링 | 신문기사 |

### Abstract

This study analyzed the trend of the term 'false and exaggerated advertisement' in 5,141 newspaper articles from 1990 to 2019 using text mining methodology. First of all, we identified the most frequent keywords of false and exaggerated advertisements through frequency analysis for all newspaper articles, and understood the context between the extracted keywords. Next, to examine how false and exaggerated advertisements have changed, the frequency analysis was performed by separating articles by 10 years, and the tendency of the keyword that became an issue was identified by comparing the number of academic papers on the subject of the highest keywords of each year. Finally, we identified trends in false and exaggerated advertisements based on the detailed keywords in the topic using the topic modeling. In our results, it was confirmed that the topic that became an issue at a specific time was extracted as the frequent keywords, and the keyword trends by period changed in connection with social and environmental factors. This study is meaningful in helping consumers spend wisely by cultivating background knowledge about unfair advertising. Furthermore, it is expected that the core keyword extraction will provide the true purpose of advertising and deliver its implications to companies and related employees who commit misconduct.

■ keyword : | False Advertisements | Exaggerated Advertisements | Text Mining | Frequency Analysis | Topic Modeling | News Articles |

## I. 서론

최근 SNS (Social Networking Service) 사용자가 급격하게 증가하고 해당 매체의 영향력이 확대됨[1]에 따라 인플루언서를 활용한 제품 리뷰형 SNS 광고, 유튜브를 활용한 영상 콘텐츠 광고 등 다양한 형식의 온라인 및 모바일 광고가 출현하게 되었다[2]. 또한 기업에서도 SNS를 활용하여 오락성과 친근함을 바탕으로 한 광고 채널을 운영하고 있다[3]. 하지만 팬덤을 통해 빠르게 퍼져나가는 과급력에 비해 광고 현행 지침은 소비 환경을 뒤쫓아가지 못해 소비자의 불만과 피해가 확대되고 있다[3].

한국소비자원에 따르면 페이스북, 인스타그램, 트위터 등 SNS 이용자의 47%가 하루 평균 6건 이상의 광고에 노출되는 것으로 나타났다[4]. 더욱이 온라인 광고비가 오프라인 광고비를 추월한[5] 현 시점에서 허위·과장광고로 인한 피해 유형의 변화를 과거부터 현재까지 알아보고자 한다.

현재까지는 허위·과장광고의 역사를 파악하기 위해 관련 논문 분석과 설문을 통한 연구 등 주관적으로 해석한 연구가 주를 이루었고, 정량적인 수치에 의거하여 객관화한 연구는 없었다. 하여 본 연구에서는 허위·과장광고 용어 관련한 비정형 텍스트 자료에 근거하여 정량적 분석을 진행하고자 한다. 이 연구 과정 속에서 유의미한 결과를 발견한다면 광고의 사회적 책임을 제고하고 기업의 윤리 의식을 일깨워줄 자료가 될 것이라 사료된다. 또한 소비자가 상품 및 서비스를 선택하기 위해 도움을 주어야 하는 광고가 오히려 소비자를 기만하는 문구로 알 권리[6]를 침해하지 않도록 하는 것에 도움을 줄 수 있다고 판단된다.

따라서 본 연구는 텍스트마이닝 기법을 통한 최빈 키워드 추출로 1990년부터 2019년까지 종합한 결과를 알아보고, 키워드 사이의 관계를 파악하고자 한다. 또한 1990년대, 2000년대, 2010년대로 분류하여 시대별 분석을 통해 키워드의 트렌드를 파악하고, 토픽모델링을 기반으로 허위·과장광고의 주요 동향을 확인하려 한다.

## II. 선행 연구

### 1. 주제적 측면

기존 선행 연구의 경우, 허위·과장광고로 인한 피해 및 문제점에도 불구하고 광고의 긍정적인 면모를 부각하여 광고 효과에 관한 연구가 이루어졌으며[7], 허위·과장광고 사례를 통한 규제 분석 연구가 주를 이루고 있다[8-13]. 하지만 김동규는 역사적 흐름(17세기 중엽~1910년대) 안에서 과장광고의 형성과 발전 양식을 추적하기 위해 역사적 연구 방법과 사례연구 방법을 병용하였는데 과장광고의 역사를 종단적으로 분석할 수 있었기에 의미 있는 시사점을 제공하였다고 볼 수 있다[14]. 그럼에도 불구하고 비정형 텍스트 자료에 근거하여 정량적 분석을 통해 역사적 맥락 속에서의 허위·과장광고를 파악한 선행연구가 없기에 주제적 측면과 분석 기법적 측면으로 나누어 알아보하고자 한다.

먼저, '허위광고'와 '과장광고'는 부당광고에 속하는 세부 개념이다. 정용수와 강창경은 부당광고를 "소비자의 알권리와 합리적인 선택권을 침해하고, 기업 간 공정한 경쟁을 저해하는 정도가 사회통념에 어긋나는 광고"로 정의하고 있다[8]. 즉 부당광고는 공정거래법과 소비자보호법에 근거하여 소비자의 합리적인 판단을 방해하고, 공정한 거래를 저해할 수 있는 광고 행위이다[9]. 부당광고의 하위 개념으로 '과장광고', '허위광고', '기만광고'로 세분화할 수 있는데 본 연구에서는 하나의 단어처럼 혼용하고 있는 허위·과장광고의 사례만을 분석하기 위해 '기만광고' 키워드가 포함된 기사는 제외하였다. 또한 '과장광고'의 동의어인 '과대광고'를 핵심 키워드에 포함하여 기사를 추출하였다.

허위·과장광고의 법적 판점 및 규제를 중심으로 다룬 연구 또한 많이 발견되었는데 그 중 마운성, 황호영은 온라인상에 게재되고 있는 기사형 허위·과장광고에 대하여 법률적으로 접근하였다[10]. 객관적인 접근을 위해 법률과 실판례를 기반으로 허위·과장광고를 고찰한 연구도 있는데 홍태석은 허위·과장광고의 추상적인 정의를 파악하기 위해 관련 판례를 통한 사회적 기준을 확인하였으며[11], 이현규는 고등법원 판결에 기반하여 표시광고법의 방향성과 기준을 제시하고 있다[12]. 반면, 탁진영은 과장광고의 사회적 유해성을 설득효과를 중심으로 제3자 효과를 고찰하였는데 소비자들이 과장광고가 지닌 유해성보다 부정적인 인식을 하고 있기에

법적 규제가 심화되고 있다는 점을 파악하였다는 점에서 의의가 있다[13].

## 2. 분석 기법적 측면

정보통신환경이 발전함에 따라 스마트폰의 보급으로 모바일 환경이 구축되고 영상, 문자, 이미지를 자유롭게 공유하면서 데이터 트래픽이 기하급수적으로 증가하였다[15]. 이러한 빅데이터의 종류는 정형화 종류에 따라 정형, 비정형, 반정형 데이터로 구분할 수 있으며[16] 이러한 데이터를 처리하기 위해 텍스트마이닝(Text Mining), 소셜네트워크 분석(Social Network Analytics), 군집분석(Cluster Analysis) 등 다양한 기법들이 활용되고 있다[16]. 본 연구에서는 비정형 데이터인 신문 기사를 수집하여 자연어 분석을 통해 말뭉치 속의 의미를 찾아내는 텍스트마이닝 분석을 수행하고자 한다.

해당 기법을 사용하여 많은 연구가 진행되고 있는데 김용희, 한창근은 '사회적 가치' 관련 이슈를 다룬 언론 보도 기사를 대상으로 의미연결망 분석을 수행하여 시간의 흐름에 따라 사회적 가치가 언론을 통해 해석되고 있는 바를 확인하였다[17]. 또한 감미아, 송민은 주요 신문사별 텍스트마이닝을 통해 긍정 및 부정 논조를 분석하여 유의미한 결과를 얻었고[18], 강주연 외 2인은 2010년부터 2019년까지 Z세대 관련 뉴스 데이터 분석을 통해 다양한 의미구조와 단어 연결성을 확인하였다[19]. 김태중은 최근 이슈가 되고 있는 코로나19를 주제로 한 뉴스 빅데이터를 감염병 위기경보 4단계 기준에 따라 토픽모델링 분석을 진행하였고, 추후 언론보도의 방향을 제시했다는 점에서 학술적·정책적 의의가 높다[20].

## III. 연구방법

### 1. 분석 도구

본 연구에서는 텍스트마이닝을 위해 오픈소스 프로그램인 R 3.6.3 버전을 사용하였다. 한글로 작성된 기사만을 수집하여 분석하였기에 한글 자연어 분석 패키지인 KoNLP (Korean Natural Language

Processing)와 텍스트마이닝 수행에 필요한 tm (text mining) 패키지를 활용하였다. 형태소 분석을 수행하기 위한 레퍼런스 사전으로는 세종(sejong) 사전을 선택하였으며, 토픽모델링의 경우 topicmodels와 lda (Latent Dirichlet allocation) 패키지 설치를 통해 분석에 활용하였다. 또한 분석한 결과값을 시각화하기 위해 wordcloud2 패키지를 인스톨하여 사용하였다.

## 2. 분석 과정

1990년 1월 1일부터 2019년 12월 31일까지 30년간 허위·과장광고 관련 기사를 수집하기 위해 한국언론진흥재단이 운영하는 뉴스 빅데이터 분석 서비스 빅카인즈 (BIG KINDS)를 활용하였다. 분석 과정은 다음과 같다.

첫 번째, 허위광고(1,474건), 과장광고(1,875건), 과대광고(2,215건) 키워드가 1개 이상 포함되어 있는 중앙지(경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보) 기사를 수집하였다. 이 중 중복되는 기사 및 포토 뉴스를 제외하고 최종적으로 5,141건의 기사만을 분석에 활용하였다. 본 연구의 분석 대상을 연도별로 나누어 요약하면 [그림 1]로 표현되며, 허위·과장광고 관련 기사가 2017년에 감소했지만 지속적으로 증가하는 추세인 점을 확인하였다.

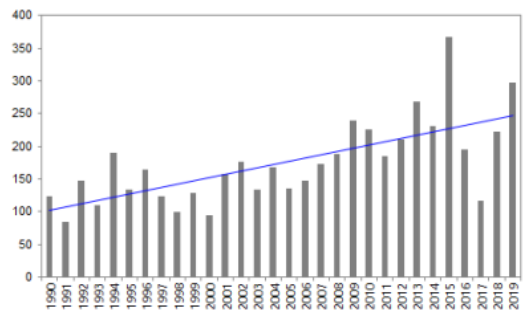


그림 1. 허위·과장광고 관련 기사수 및 추세선

두 번째, 수집된 기사 5,141건에 대해서 전처리 과정을 진행하였는데 숫자 및 공백을 제거하는 말뭉치 수정 작업을 우선적으로 수행하였다. 다음으로 빈도수가 높은 순서대로 키워드를 추출하였으나 연구에 불필요한

단어가 발견되어 이를 제거하는 작업을 진행하였다. 반대로, ‘식품의약품안전처’, ‘SK’ 등과 같은 기관명과 브랜드명을 사전에 추가하였다. 또한 ‘식품의약품안전처’와 같은 단어의 경우, ‘식약처’로 축약하여 사용되고 있으며, ‘공정거래위원회’와 ‘지방검찰청’ 또한 각각 ‘공정위’와 ‘지검’으로 줄여 쓰이고 있기에 합수를 사용하여 이를 통합하는 과정을 거쳤다.

세 번째, 전처리 과정이 끝난 후, 100회 이상 출현한 고빈도어 100개를 내림차순으로 선정하여 키워드 빈도 분석을 수행하였다. 이때 허위·과장광고 관련 기사에서 나타난 최빈 키워드를 업종, 단어의 의미, 단속 기관, 피해 대상 등 4가지 기준을 세워 분석한 결과 허위·과장광고 기사의 전반적인 구조를 파악할 수 있었다.

네 번째, 시간적 경과 속에서 허위·과장광고의 전개 과정을 알아보고자 기사를 시기별로 나누어 키워드 빈도 분석을 진행하였다. 10년 주기로 나누어 시간적 흐름에 따라 연대별 최빈 키워드 분석을 진행하여 키워드 변화 양상을 확인한 선행연구[21]를 기반으로 연구 분석 범위를 제한하고, 빈도 분석을 수행하였다. 또한 허위·과장광고 기사에서 언급되는 키워드들이 사회적 또는 환경적 요인에 의해 영향을 받는다는 사실을 뒷받침하고자 연도별 최빈 키워드를 주제로 한 학술논문 수를 파악하였고, 교차 비교하여 관련성을 파악하였다.

마지막으로 잠재 디리클레 할당(LDA) 모델을 활용하여 토픽모델링 분석을 수행하였다. 토픽모델링은 텍스트마이닝 기법의 하나로 빅데이터 동향을 파악하기에 적합한 기법이다[22]. 특정 단어가 모여 하나의 문서가 만들어진다고 가정하고 문서에 출현하는 단어의 수치를 측정하여 본문의 숨겨진 토픽을 발견하는 통계적 모델이다[23]. 본 연구에서는 토픽모델링 기법 중 Blei, Ng, Jordan(2003)이 제안한 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 모델을 사용하고자 한다[22]. LDA는 방대한 양의 비정형 텍스트로부터 토픽을 추출하기에 유리한 모델이기에[24] 해당 방식을 활용하였다. 토픽의 수를 결정하기 위해 토픽의 수를 3개~8개로 변경하며 분석을 반복하였고, 키워드 중복이 최소화되어 구성 단어를 통해 토픽 유추가 가능한 5개로 추출되도록 설정하였다[25]. 최종적으로 20개씩 배정된 하위 구성 단어를 고려하여 토픽명을 부여하는 과

정을 통해 허위·과장광고의 동향을 확인할 수 있었다. 본 연구의 자료 수집 및 분석 과정은 [그림 2]와 같다.

자료수집	<ul style="list-style-type: none"> <li>분석 대상: 1990년 1월부터 2019년 12월까지 허위·과장광고 관련 중앙지 기사 5,141건</li> <li>자료 수집처: 빅키인즈 (BIG KINDS)</li> </ul>
전처리과정	<ul style="list-style-type: none"> <li>숫자 및 공백 제거 등 말뭉치 수정</li> <li>축약 단어 정제</li> </ul>
전체 키워드 대상 빈도분석	<ul style="list-style-type: none"> <li>100회 이상 출현한 최빈 키워드 100개 선정</li> <li>4가지 기준을 세워 키워드를 분류하고 기사의 전반적인 구조 파악</li> </ul>
시대별 키워드 대상 빈도 분석	<ul style="list-style-type: none"> <li>10년 주기로 키워드 빈도 분석 수행</li> <li>각 연도별 최빈 키워드를 주제로 한 학술논문수 파악하여 비교</li> <li>허위·과장광고의 전개 과정 및 연도별 특성 파악</li> </ul>
시대별 키워드 대상 토픽모델링 분석	<ul style="list-style-type: none"> <li>LDA 모델을 활용하여 5개의 토픽과 20개씩의 세부 키워드 추출</li> <li>토픽명을 추론하여 허위·과장광고의 동향 확인</li> </ul>

그림 2. 자료 수집 및 연구 과정

## IV. 연구 결과

### 1. 전체 키워드 대상

#### 1.1 빈도 분석

1990년부터 2019년까지 ‘과장광고’, ‘허위광고’, ‘과대광고’ 키워드를 포함하고 있는 중앙지 기사 5,141건을 바탕으로 키워드 빈도 분석을 시행한 결과 중 ‘과장광고’, ‘허위광고’, ‘과대광고’를 제외한 키워드를 [표 1]에 요약하였다. [표 1]은 빈도수를 기준으로 100개의 키워드를 내림차순으로 정리한 것이며, 허위·과장광고 기사에서 가장 빈번하게 쓰인 키워드는 ‘공정거래위원회’로 나타났다. 이는 해당 기관에서 부당한 표시 및 기만적인 내용이 담긴 광고에 대해 제재를 가하고, 심사를 진행하기 때문에 상위 랭크되었다고 판단할 수 있다. 이밖에 ‘식품의약품안전처’, ‘경찰’, ‘지방검찰청’, ‘정부’ 등 조사 및 단속기관이 고빈도 키워드로 추출된 것과 더불어 ‘식품위생법’, ‘시정명령’, ‘과징금’, ‘구속’, ‘고발’, ‘불법’, ‘적발’, ‘규제’, ‘위반’ 등 담당 기관과 관련된 법률적인 부분도 함께 보도되었음을 알 수 있다. 또한 ‘부작용’, ‘부당’, ‘금지’ 등 피해 관련한 부정적인 단어가 다수 추출된 것을 확인하였다. 앞선 단속기관과 결합하여 해석해본다면 이는 허위·과장광고로 인해 소비자가

표 1. 허위·과장광고에서 추출된 고빈도 키워드

공정거래위원회	1540	경찰	406	신문	280	성분	218	방송	168
적발	1343	표시	398	건강식품	275	TV	208	백화점	166
식품	943	가입	395	구속	275	건강보조식품	203	국민	164
인터넷	800	홍소핑	393	시장	274	과징금	197	수사	163
분양	695	회사	388	투자	273	주의	197	환자	163
식품의약품안전처	691	단속	376	식품의약품	266	의료기기	195	시민	159
소비자	688	정부	373	부당	262	기업	194	식품위생법	159
효과	604	건강기능식품	372	강화	261	교육	192	업소	158
위반	579	수입	357	정보	256	노인	192	자동차	157
소비	563	다이어트	343	시정	254	현혹	191	개발	156
건강	493	고발	340	피해	241	고객	182	대학	155
조사	492	의약품	339	금융	238	구입	182	서비스	154
혐의	491	보험	337	모집	235	소송	180	검사	153
안전	488	검찰	335	사람	234	규제	178	인기	152
제조	483	은행	332	상가	229	지방검찰청	174	홍보	150
시정명령	462	부동산	321	쇼핑	226	자격증	175	공정거래	147
화장품	459	사이트	318	온라인	224	경쟁	173	대출	147
학원	447	유통	287	LG	222	부작용	172	거래	144
아파트	441	치료	286	병원	222	문구	170	여성	143
불법	413	효능	286	가격	220	금지	168	SK	141

피해를 받았고, 담당 기관에서 조사 및 제재를 가하는 과정이 주로 기사화되었다고 판단된다.

이 외에도 최상위 랭크된 ‘식품’, ‘분양’, ‘건강’, ‘학원’, ‘대학’, ‘아파트’, ‘부동산’, ‘건강기능식품’, ‘건강보조식품’, ‘의료기기’, ‘병원’, ‘치료’ 등의 키워드를 통해 금융, 교육, 식품, 의료업종의 부당광고로 인한 소비자 피해가 빈번하게 발생했다는 것을 추측할 수 있다. 더 나아가 추출된 키워드 분석을 통해 피해 대상 유형 또한 확인하였다. ‘시민’, ‘고객’, ‘사람’, ‘소비자’ 등 넓은 의미의 일반적인 피해자부터 허위·과장광고의 주된 피해자인 ‘노인’, ‘여성’, ‘환자’ 등 특정 피해 대상까지 세분화할 수 있다. 또한 ‘신문’, ‘방송’과 같은 4대 매체부터 ‘인터넷’, ‘사이트’, ‘온라인’ 등 BTL (Below The Line)까지 광고를 접할 수 있는 다양한 플랫폼이 최빈 키워드로 추출되었다.

## 2. 시대별 키워드 대상

### 2.1 빈도 분석

빠르게 변화하는 미디어 환경과 발전하는 기술 속에서 허위·과장광고 관련 기사에도 변화가 있는지 파악하고자 10년 단위로 최빈 키워드 비교 분석을 하였다. 1990년 1월부터 1999년 12월까지의 기사는 1990년대(1,300건)로 분류하고, 2000년 1월부터 2009년 12월까지는 2000년대(1,592건), 2010년 1월부터 2019

년 12월까지의 2010년대(2,249건)로 누적 합산하였다. 10년치씩의 기사 데이터의 전처리 과정의 경우, 전체 키워드 빈도 분석에서 진행한 순서대로 동일하게 진행하였다.

[표 2]는 연도별 허위·과장광고 주요 키워드 리스트이다. 전체 빈도 분석에서도 최빈 키워드 1위로 랭크된 ‘공정거래위원회’는 1990년대, 2000년대에는 1위로, 2010년대 역시 최상위권에 속하며 존재감을 나타내고 있다. 그러나 시간의 흐름에 따라 사회적, 환경적 요인에 의해 시기별로 키워드 트렌드가 다르게 나타나는 것을 볼 수 있다.

우선 1990년대의 경우, IMF (International Monetary Fund) 이슈로 인해 ‘분양’, ‘아파트’, ‘투자’, ‘금융기관’, ‘건설’ 등 투자 관련 키워드가 상위권을 차지한 것을 알 수 있다. 특히 함께 추출된 ‘투신사’는 ‘투자 신탁 회사’를 줄여 일컫는 단어이자[26] IMF 당시 금융시장의 안정을 위해 1996년 소개된 단기금융 펀드 명으로[27] 사회적인 특성을 잘 반영한 단어이다. 따라서 금융업종에서 불공정 거래가 많았음을 유추할 수 있다. 또한 ‘약사법’ 키워드와 ‘건강보조식품’이 상위권에 추출된 것을 보아 1990년대에는 의약품이 아님에도 불구하고 의약품으로 오인하거나 혼동할 우려가 있는 내용 혹은 의약품 효능이 있는 것처럼 과장 또는 기만하는 내용을 담아 광고한 제품이 다수 적발되었음을 추측하였다[28]. 1990년대에는 ‘일간지’ 키워드가 추출되어 피해

플랫폼이 오프라인 중에서도 신문이었음을 확인할 수 있었다.

2000년대에는 ‘금융감독원’ 및 ‘식품의약품안전처’ 키워드가 처음으로 나타났는데 ‘금융감독원’은 소비자를 보호하고, 공정한 시장 질서 확립에 힘쓰는 기관으로 1999년 1월 설립되었고[29], ‘식품의약품안전처’ 역시 식품 의약품의 안전 관리 체계를 구축 및 운영하고자 1998년 2월에 보건복지부 산하에 설립됨으로써[30] 이 두 기관은 1990년대와 2000년대를 구분할 수 있는 키워드이다. 그뿐만 아니라 1990년대에는 ‘건강보조식품’이 상위 키워드였으나 2000년대부터는 ‘건강기능식품’이 추출되었는데 2002년부터 건강기능식품이 식품의약품안전처에서 법적으로 관리[31]되고, 건강 기능 식품 광고 심의 절차가 강화[32]되며 허위·과장광고가 다수 적발되었음을 추측할 수 있다. 또한 ‘분양’, ‘아파트’, ‘부동산’, ‘상가’, ‘오피스텔’ 등 부동산 관련 키워드와 함께 ‘개발’, ‘대출’ 키워드가 상위 키워드로 추출된 것을 보아 부동산 분양 및 임대업자의 사기 분양으로 인한 피해[33]와 부동산 대출 허위·과장광고[34]가 빈번하여 공정거래위원회와 금융감독원에서 대대적인 모니터링을 시행했음을 파악하였다. 그리고 1994년 인터넷 상

용화 이후, 2000년대에는 인터넷 이용 수요가 급격하게 증가하여[35] ‘인터넷’은 상위권 키워드로 추출되었으나 ‘TV’, ‘방송’, ‘홈쇼핑’, ‘전화’, ‘신문’ 등 오프라인 매체 또한 큰 비중을 차지하며 추출되었기에 여전히 오프라인 매체가 허위·과장광고를 접하는 주요 플랫폼이었음을 추측할 수 있다.

2010년대에는 ‘사이트’, ‘온라인’, ‘초고속인터넷’ 등 온라인 광고비가 오프라인 광고비를 추월한 실제 광고 시장 상황과[5] 밀접하게 관련 있는 키워드가 나타났다. 온라인의 발달로 SNS를 활용한 인플루언서의 체험 후기형 허위·과장광고가 빈번하게 발생하였고, 식품의약품안전처는 안전에 취약한 다이어트, 미세먼지, 탈모, 여성 건강, 취약계층 등 5개 분야를 단속하였다[36]. 이로 인해 해당 분야의 허위·과장광고 사실이 다수 적발되었으며, 관련 주요 키워드들이 상위 랭크된 것을 확인하였다. 또한 ‘살균제’, ‘가습기’ 키워드가 추출된 이유는 2011년부터 사회적으로 이슈가 되었던 가습기 살균제 유해물질 사건으로[37] 인해 관련 내용이 담긴 기사가 많이 보도되었음을 유추할 수 있다. 인체에 유해한 가습기 살균제를 판매하면서 제품 광고 시, 안전하다고 허위 표시를 했기에 허위광고로 인한 다수의 피해자가

표 2. 연대별 허위·과장광고 주요 키워드 리스트

1990년대			2000년대			2010년대		
공정거래위원회	화장품	정지	공정거래위원회	신문	부당	적발	다이어트	시정명령
적발	지방검찰청	교육	적발	회사	금융감독원	식품의약품안전처	분양	KT
시정명령	신문	다이어트	분양	경찰	현혹	인터넷	정부	소송
식품	경찰	맥주	식품	다이어트	건강기능식품	공정거래위원회	노인	성분
제품	학원	부작용	인터넷	표시	성분	식품	시장	강화
수입	아파트	통신	홈쇼핑	TV	과징금	효과	단속	초고속인터넷
회사	효과	유통	아파트	제조	의료	가입	보험	회사
은행	고객	가격	안전	고발	피해	소비자	병원	사람
조사	규제	건설	학원	정부	은행	위반	정보	비교
소비	효능	금지	소비자	수입	업소	건강기능식품	학원	탈모
소비자	강화	위반형의	건강	쇼핑	전화	건강	의약품	마스크
구속	모집	투자	소비	효능	가격	사이트	검찰	헌자
고발	피해	수사	보험	모집	대출	소비	의료기기	은행
분양	거래	투신사	효과	의약품	독자	안전	홈쇼핑	가격
협의	경쟁	국민	위반	정보	시정	협의	투자	구매
시정	단속	동양	화장품	자격증	개발	경찰	건강식품	고발
부당	업소	건강	조사	치료	교육	화장품	아파트	피해
검찰	일간지	개정안	부동산	투자	병원	불법	부동산	수입
공정거래	기업	시민	불법	사이트	보장	온라인	살균제	위원회
백화점	현혹	금융기관	단속	사람	자동차	유통	가습기	구입
정부	금융	사람	식품의약품	시장	문구	조사	미세먼지	부작용
의약품	폭리	약사법	시정명령	강화	오피스텔	표시	쇼핑	주의
표시	건강식품	활인	상가	주의	등록	제조	금융	여성
건강보조식품	교재	불공정	협의	건강식품	방송	치료	효능	구속
위반	불법	자격증	식품의약품안전처	검찰	서비스	LG	SK	인기



그림 3. 연대별 허위·과장광고 키워드 비교 분석 결과

발생한 사건이다. 사회적 이슈뿐만 아니라 꾸준히 환경적인 요인으로 심각성이 강조되었던 미세먼지 문제 [38]도 관련 단어인 '마스크', '미세먼지'가 고빈도 키워드로 선정되며 어떠한 환경 문제가 불거지면 관련 제품에 대한 허위·과장광고가 증가한다는 것을 분석을 통해 확인할 수 있었다. 'KT', 'SK', 'LG' 기업명과 함께 '초고속인터넷', '가입' 키워드가 동시에 추출된 것을 종합적으로 분석한다면 3대 통신사의 인터넷 가입 관련 허위·과장광고가 다수 발생했다는 사실을 유추할 수 있다. 최근, 방송통신위원회는 인터넷과 방송 서비스 가입 시 할인 혜택을 제공한다는 문구를 사용하거나 이용조건을 제외하고 중요 혜택만 표시하여 소비자를 기만한 통신사 4곳에 대해 과징금을 부과하였다[39]. 연대별 상위 75개 키워드를 시각화한 결과는 [그림 3]과 같다.

2.2 핵심 키워드와 학술논문과의 교차 분석

허위·과장광고 관련 학술 논문의 트렌드 분석 결과와 뉴스 데이터에서 추출한 허위·과장광고 관련 키워드 간의 공통점을 확인함으로써 해당 시기 쟁점이 되었던 논의주제를 중심으로 정보가 생산되며, 시대 변화 양상에 따라 키워드 트렌드가 변화한다는 사실을 뒷받침하고자 한다. 따라서 네이버 학술정보 사이트를 활용하여 연도별 최빈 키워드를 주제로 한 학술논문을 연대별로 조사하였다.

앞선 시대별 빈도 분석으로 추출된 128개의 고유의 키워드 중 연대별 모두 나타난 일반적인 키워드는 제외하고, 특정 시점을 기준으로 키워드가 표출되어 시대의 특색이 뚜렷하게 나타나는 핵심 키워드 10개를 선정하였다. 예를 들어 '투신사'의 경우, 1990년대에만 나타난

키워드이며, 사회적 배경을 분석한 결과 IMF 여파로 투신사 키워드가 쟁점이 되었으나, 외환 위기가 극복된 이후에는 영향력이 미미해졌기에 시대상을 적절하게 반영한 키워드라고 판단되어 선정되었다. [그림 4]는 선정된 주요 키워드 10개를 주제로 게재된 학술논문을 연대별로 분류하여 도표화한 것이다. '미세먼지'와 '마스크', '가습기살균제'의 경우, 해당 이슈가 수면 위로 올랐던 2010년대에 압도적으로 많은 연구 논문 발행 비율을 차지하고 있다. 또한 건강에 대한 소비자의 관심이 증가하고 시장이 확대됨에 따라 '건강기능식품' 관련 연구 비중 또한 점진적으로 커지는 양상을 확인하였다. 일반 식품인 '건강보조식품'과 식품의약품안전처로부터 기능성과 안전성을 인정받은 '건강기능식품' 간의 차이[40]를 이용한 허위·과장광고로 인해 피해가 발생했다는 점을 앞선 빈도 분석과 학술논문 교차 분석을 통해 파악하였다. 또한 '온라인', '인터넷', '초고속인터넷', '홈쇼핑'의 경우, 인터넷 상용화 전인 1990년대에는 수치가 낮지만, 인터넷 이용자가 급격하게 증가한 2000년대 이후부터 연구가 급증하는 추세를 보인다.

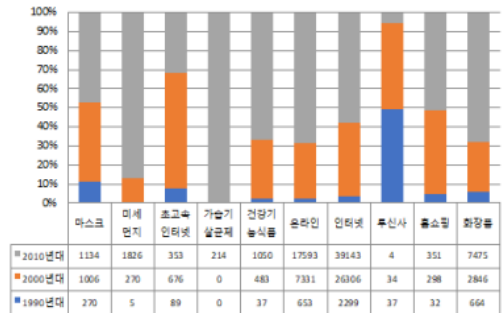


그림 4. 연도별 주요 키워드의 논문수 시각화

하지만 반대로 현재는 대체로 사용하지 않는 ‘투신사’ 키워드의 경우 1990년대 이후 점진적으로 감소하며 1990년대의 특성을 가장 두드러지게 보여주는 키워드임을 확인하였다.

### 2.3 토픽모델링 분석

본 연구에서는 허위·과장광고 관련 뉴스를 대상으로 토픽모델링 분석을 수행하여 소비자들이 실질적으로 마주한 허위·과장광고 피해의 주요 토픽과 변화 과정을 알아보고자 하였다. 분석 결과, 연대별 빈도 분석을 통해 추출된 최빈 키워드와 토픽은 해당 시기의 사회적, 환경적 상황을 반영하며 변화하고 있다는 점에서 일맥상통한다. 시대별 5개의 주제와 토픽별 주요 키워드 20개는 [표 3-표 5]와 같다.

1990년대의 토픽은 ‘금융업계의 피해’, ‘공정거래 위원회의 움직임’, ‘투자업계의 타격’, ‘다이어트 식품의 효능 문제’, ‘식품제조업계의 허위·과장광고 현황’으로 지정하였다. 앞선 빈도 분석을 통해 한국 외환 위기로 투자 관련 파생어들이 상위권에 위치한 것을 확인하였는데 토픽모델링 분석을 통해서도 금융, 투자, 식품 등 IMF 여파로 피해를 받은 업계가 쟁점이 되었음을 보여주는 바이다. 또한 1990년대부터 2010년대까지 최상위 키워드로 추출된 공정거래위원회의 경우, 1990년대에 하나의 토픽으로 선정되며 존재감을 나타냈다.

표 3. 1990년대 기사의 토픽별 세부 키워드 및 토픽명

토픽 1	토픽 2	토픽 3	토픽 4	토픽 5
은행	수입	분양	구속	공정거래위원회
금융	의약품	공정거래위원회	혐의	적발
고객	소비	시정명령	지방검찰청	고발
교재	공정거래위원회	아파트	식품	백화점
투신사	정부	건설	검찰	시정명령
경쟁	소비자	공정거래	회사	제조
금융기관	강화	상가	제조	식품
정부	부당	면적	적발	시정
자격증	화장품	경찰	경찰	업소
감독원	조사	구속	다이어트	부당
소비자	표시	투자	건강식품	조사
수익률	제조	시정	건강보조식품	위반
규제	개정안	수입	효과	표시
피해	정지	방송	효능	건강보조식품
소비	규제	회사	폭리	교육
회사	개정	현대	위반혐의	학원
현혹	공정거래	경품	수사	소비

저축	여행사	조사	맥주	검찰
시험	해외여행	기업	수입	화장품
보장	회사	소비자	일간지	할인
금융업계의 피해	공정거래위원회 움직임	투자업계의 타격	다이어트 식품의 효능 문제	식품제조업계의 허위·과장광고 현황

2000년대의 주요 토픽은 [표 4]과 같다. 각 토픽은 ‘셀럼을 활용한 다이어트 식품 광고의 실태’, ‘홈쇼핑 채널을 통한 보험 상품의 현황’, ‘인터넷 발달로 인한 다양한 피해 발생’, ‘인터넷 상용화로 카드 이용 카테고리의 확대’, ‘온라인으로 유통되는 건강식품의 효능 문제’로 명명하였다. 5개의 토픽 중 3개의 토픽이 인터넷과 관련되며, 1990년대에는 토픽으로 추출되지 않았던 홈쇼핑 연관 주제 또한 2000년대에서 볼 수 있다. 이는 빈도 분석에서 2000년대에 온라인과 오프라인 매체가 피해 플랫폼으로 함께 나타난 양상과 동일하다. 해당 시기의 인터넷 사용자 수가 급증하였고[34], 새로운 피해 플랫폼이 출현하면서 기사로 많이 보도되었음을 유추할 수 있는 부분이다.

표 4. 2000년대 기사의 토픽별 세부 키워드 및 토픽명

토픽 1	토픽 2	토픽 3	토픽 4	토픽 5
공정거래위원회	보험	분양	화장품	식품
홈쇼핑	홈쇼핑	공정거래위원회	아파트	적발
시정명령	TV	학원	공정거래위원회	안전
조사	소비	부동산	소비자	식품의약품
효과	은행	아파트	카드	인터넷
다이어트	사람	상가	교복	식품의약품안전처
명령	금융	적발	자동차	건강
결혼	소비자	자격증	분양	효과
건강	협회	인터넷	대학	불법
신문	시장	과징금	학교	제조
소비자	생명	오피스텔	인터넷	의약품
공표	보험사	단속	교육	혐의
소비	상조	경찰	입주자	사이트
시정	건강	위반	태반	치료
건강보조식품	금융감독원	등록	신용카드	수입
부당	방송	수강료	적발	위반
단속	가입	시정명령	현대	단속
위반	감독	조사	소송	건강기능식품
연예인	대출	신문	회사	건강식품
LG	모집	불법	면적	효능
셀럼을 활용한 다이어트 식품 광고의 실태	홈쇼핑 채널을 통한 보험 상품의 현황	인터넷 발달로 인한 다양한 피해 발생	인터넷 상용화로 카드 이용 카테고리의 확대	온라인으로 유통되는 건강식품의 효능 문제



2010년대의 토픽은 '3대 통신사의 초고속 인터넷 가입 관련 문제', '허위·과장광고의 주요 피해 대상과 채널', '다이어트 및 뷰티 업계의 광고 현황', '성형외과 불법 시술 실태', '가습기살균제 사건'으로 정하였으며, [표 5]와 같다. 해당 시기에 큰 논쟁거리가 되었던 가습기살균제 사건 및 통신사 3사의 초고속 인터넷 허위광고 사실의 경우, 최빈 키워드 추출 및 각각의 토픽으로 나타나며 사건의 심각성을 보여주는 결과이다. 이와 더불어 탈모, 여성 등 주요 피해 대상과 다이어트 관련 주제가 표출되며 식품의약품안전처가 단속한 분야와 연관성을 보인다. 또한 1990년대, 2000년대, 2010년대 모두 다이어트 식품의 효능 관련한 토픽이 지속적으로 발견되었고, 2010년대에는 성형외과 관련 토픽이 추출되며 미용에 대한 소비자의 관심이 증가하고 있다는 점을 유추해볼 수 있다.

표 5. 2010년대 기사의 토픽별 세부 키워드 및 토픽명

토픽 1	토픽 2	토픽 3	토픽 4	토픽 5
인터넷	보험	적발	부동산	공정거래위원회
가입	학원	식품	병원	분양
소비자	은행	식품의약품안전처	홈쇼핑	살균제
LG	금융	안전	의원	아파트
사이트	환자	효과	불법	가습기
미세먼지	남성	건강기능식품	소비	혐의
SK	탈모	식품의약품	의사	검찰
KT	치료	위반	국회	시정명령
소비	성장	경찰	신고	소송
초고속인터넷	교육	화장품	투자	조사
비교	여성	다이어트	성형외과	법원
마스크	대학	건강	대부	과징금
공정거래위원회	방법	노인	위원회	상가
통신사	홈쇼핑	유통	중고	기소
쇼핑	고민	제조	부작용	회사
온라인	적발	의약품	제주	수사
전자	투자	의료기기	건강기능식품	대법원
선택	학생	건강식품	기업	정보
요금	저축	효능	사회	원고
효과	건강	성분	소비자	선고
초고속인터넷 가입 문제	주요 피해 대상과 채널	다이어트 및 뷰티업계의 광고 현황	성형외과 불법 시술 실태	가습기 살균제 사건

[표 6]은 연대별 토픽모델링 변화추이를 요약한 것이다. 우선 다이어트 식품으로 포장된 건강기능식품은 연대별로 지속적으로 이슈화되었으나, 금융, 부동산 관련 토픽은 1990년대와 2000년대에 주로 이슈화되었고 2010년대는 초고속인터넷, 성형외과 불법시술, 가습기

살균제 같은 사회적, 환경적 이슈가 허위·과장광고의 토픽으로 도출되었다. 또한 허위·과장광고의 매체 변화도 나타나는데 1990년대는 주로 방송 광고가, 2000년대 이후에는 홈쇼핑과 인터넷 기반 광고가 도출되었다.

표 6. 연대별 토픽모델링 변화추이

구분	1990년대	2000년대	2010년대
주요 이슈	다이어트 식품	다이어트 식품	다이어트 식품
	저축은행	보험상품	초고속인터넷
	아파트 분양	부동산 분양	성형외과 부작용
매체	방송	홈쇼핑, 인터넷	홈쇼핑, 인터넷

## V. 결론 및 시사점

본 연구는 1990년부터 2019년까지 30년간 보도된 5,141건의 기사를 바탕으로 허위·과장광고 관련 핵심 키워드를 추출하여 트렌드를 파악하고자 하였다. 연구 결과와 시사점은 다음과 같이 요약할 수 있다.

먼저, 5,141건의 전체 기사에서 최빈 키워드 100개를 추출하여 허위·과장광고 관련 핵심 키워드를 확인하고, 추출된 키워드 간의 맥락을 파악하였다. '식품의약품안전처', '공정거래위원회'와 같은 단속 기관이 고빈도어로 추출되었으며, 이 키워드와 함께 출현한 피해 관련 키워드들을 파악하였다.

다음으로 1990년부터 2019년까지 10년 단위로 데이터를 구분하여 시간의 흐름에 따라 변화하는 키워드 동향을 분석하였다. 조사기관과 함께 '적발', '부당', '불법'과 같은 피해 키워드가 상위권에 동일하게 랭크된 것을 통해 과거부터 현재까지 허위·과장광고로 인한 피해는 꾸준히 지속되어 왔다는 것을 유추할 수 있다. 그러나 정부 부처의 이름이 변경되고 신설됨에 따라 추출되는 기관 관련 키워드들에 변화가 발생하게 되었다. 또한 연대별로 피해 업종과 피해 대상이 다르게 나타나는 것을 확인할 수 있었다.

앞선 빈도 분석으로 추출된 키워드 중 연도별 특징이 확실하게 나타나는 키워드 10개를 선정하여 해당 키워드를 주제로 게재된 학술논문을 연대별로 분류 및 분석하였다. 특정 시기에 이슈가 되었던 주제의 경우, 허위·과장광고 관련 기사에서 최빈 키워드로 추출되었을 뿐

만 아니라 전반적인 연구 동향과도 직결된다는 사실을 발견하였다.

마지막으로 토픽모델링 분석을 통해 허위·과장광고 기사에서 언급되는 키워드들이 사회적 또는 환경적 요인에 의해 영향을 받는다는 시사점을 발견하였다. 1990년대의 세부 키워드들은 IMF 사태로 인해 타격을 입은 투자업계, 금융업계와 연관되어 있었으며, 인터넷 상용화 시기인 2000년대에는 온라인 판매와 관련된 토픽이 다수 포착되었다. 더욱이 2010년대에는 통신업계의 허위·과장광고 실태와 가습기살균제 사건이 주요 토픽으로 추출되며 해당 시기에 이슈가 되었던 토픽들과 강한 연관성을 가졌다는 점을 확인할 수 있었다.

본 연구는 허위·과장광고가 변화하는 과정을 텍스트 마이닝 분석 기법을 적용하여 새로운 연구 방법을 시도했다는 것에 의의가 있다. 과장광고의 형성 및 전개 과정에 대해 해석적으로 접근한 연구는 많았지만 30년 역사에 걸쳐 허위·과장광고가 발생했던 기사를 기반으로 텍스트마이닝 분석 기법을 활용하여 객관적으로 분석했다는 점에서 함의가 적지 않다.

이번 연구를 통해 허위·과장광고가 현대사회와 밀접하게 연관되어 있다는 점을 파악하였다. 따라서 대용량의 비정형 데이터 분석을 진행한 본 연구의 결과를 활용하여 부당광고를 심의하고 단속하는 정부 부처에서는 빅데이터, AI 등 4차 산업 고도화에 따라 함께 발전하는 온라인 광고 관련 소비자 보호 정책을 구체화하는데 적용할 수 있다. 추가적으로 뉴스 빅데이터를 통해 추출된 키워드는 광고 산업 관련 종사자 및 기업의 임원진에게 광고 윤리 의식을 재조명하게 하여, 소비자 기반 광고 관련 교육에 대한 기초 자료로 활용될 수 있을 것이다.

그러나 본 연구의 한계점으로는 연구논문과의 교차 분석을 진행했으나 신문기사를 기반으로 텍스트마이닝을 수행했기 때문에 허위·과장광고를 직접적으로 마주하는 소비자의 의견 파악이 어렵다는 점이다. 따라서 후속 연구에서는 소셜네트워크서비스 및 블로그 데이터를 추가하여 미디어 기반 데이터와 함께 소비자 기반 데이터를 함께 분석하여 균형적이고 다각적인 시각을 제공하는 것이 필요하다.

참고 문헌

- [1] 서주연, 박민정, “SNS 텍스트마이닝을 중심으로 한 브랜드 목표 이미지와 브랜드 광고모델의 대중적 이미지 일치에 관한 연구,” 브랜드디자인학연구, 제16권, 제3호, pp.219-234, 2018.
- [2] 김은재, 황상재, “인플루언서 마케팅에서 정보원 유형과 경제적 대가 표시에 따른 광고 효과 연구,” 한국디지털콘텐츠학회 논문지, 제20권, 제2호, pp.297-306, 2019.
- [3] 나윤빈, “SNS 광고속성이 구매의도 및 브랜드 태도에 미치는 영향 - 설득지식의 조절효과를 중심으로-,” 한국콘텐츠학회논문지, 제19권, 제8호, pp.58-68, 2019.
- [4] <https://www.kca.go.kr/kca/sub.do?menukey=5084&mode=view&no=1001891152&page=34>.
- [5] <http://onlinead.or.kr/17/?idx=3291169&bmode=view>.
- [6] <https://terms.naver.com/entry.nhn?docId=3656521&cid=42131&categoryId=42131>.
- [7] 정수정, “허위광고의 표현양식과 브랜드에착에 따른 소비자 기만효과 연구,” 광고학연구, 제22권, 제1호, pp.303-333, 2011.
- [8] 정용수, 강창경, 부당광고 규제의 실효성 제고 방안 - 추천·보증에 관한 광고 및 광고실증제를 중심으로, 한국소비자원, 2011.
- [9] 손수진, “부당광고의 규제에 관한 연구 - 독점규제 및 공정거래에 관한 법률을 중심으로,” 광고연구, 제32호, pp.255-297, 1996.
- [10] 마운성, 황호영, “온라인 기사형 광고의 허위 및 과대·과장광고로써의 법률적 접근과 해석,” 디지털융복합연구, 제14권, 제5호, pp.391-402, 2016.
- [11] 홍태석, “광고에서의 허위 과장광고의 판단기준 - 대법원 2008. 8. 21 선고 2007도7415 판결을 대상으로 -,” 법이론실무연구, 제7권, 제3호, pp.239-257, 2019.
- [12] 이현규, “표시광고법상 부당한 표시광고의 합헌적 판단 기준,” 경쟁법연구, Vol.23, pp.381-407, 2011.
- [13] 탁진영, “과장광고의 설득효과와 제3자 효과 - 규제에 대한 소비자의 태도를 중심으로,” 언론과학연구, 제10권, 제1호, pp.398-441, 2010.
- [14] 김동규, “과장광고의 형성과 전개에 관한 역사적 연구 : 17세기 중엽~1910년대까지를 중심으로,” 광고

- PR실학연구, 제10권, 제3호, pp.9-40, 2017.
- [15] 장재혁, 박승근, “모바일 트래픽 동향,” 전자통신동향 분석, Vol.34, No.3, pp.106-113, 2019.
- [16] 김정숙, “빅 데이터 활용과 관련기술 고찰,” 한국콘텐츠학회지, 제10권, 제1호, pp.34-40, 2012.
- [17] 김용희, 한창근, “사회적 가치 이슈를 다룬 언론기사의 의미연결망 분석 - 2006년부터 2019년까지 국내 중앙지 언론보도를 중심으로 -,” 한국사회복지학, 제72권, 제2호, pp.201-229, 2020.
- [18] 감미아, 송민, “텍스트 마이닝을 활용한 신문에 따른 내용 및 논조 차이점 분석,” 지능정보연구, 제18권, 제3호, pp.53-77, 2012.
- [19] 강주연, 이이든, 김지수, “텍스트 마이닝을 활용한 ‘Z세대’ 관련 뉴스데이터 의미연결망 분석,” 미래청소년학회지, 제17권, 제2호, pp.25-48, 2020.
- [20] 김태중, “뉴스 빅데이터를 활용한 코로나19 언론보도 분석 :토픽모델링 분석을 중심으로,” 한국콘텐츠학회논문지, 제20권 제5호, pp.457-466, 2020.
- [21] 김민정, 김철주, “키워드 분석 기반 ‘전통’ 용어의 트렌드 분석 (1920~2017),” 한국콘텐츠학회논문지, 제18권, 제12호, pp.421-431, 2018.
- [22] 노철현, “토픽모델링을 활용한 인공지능 관련 이슈 분석,” 디지털융복합연구, 제18권, 제5호, pp.75-87, 2020.
- [23] 우창우, 이종연, “LDA 토픽모델링을 통한 ICT분야 국가연구개발사업의 주요 연구토픽 및 동향 탐색,” 한국융합학회논문지, 제11권, 제7호, pp.9-18, 2020.
- [24] 서보순, “장애아동 권리에 관한 뉴스 기사의 토픽모델링 분석,” 교육혁신연구, 제30권, 제2호, pp.31-50, 2020.
- [25] 박상현, 문현실, 김재경, “토픽 모델링에 기반한 온라인 상품 평점 예측을 위한 온라인 사용 후기 분석,” 한국IT서비스학회지, 제16권, 제3호, pp.113-125, 2017.
- [26] [https://stdict.korean.go.kr/search/searchView.do?word\\_no=497766&searchKeywordTo=3](https://stdict.korean.go.kr/search/searchView.do?word_no=497766&searchKeywordTo=3).
- [27] <https://dic.hankyung.com/economy/view/?seq=2437>.
- [28] <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=001&oid=001&aid=0004137982>.
- [29] [https://www.doopedia.co.kr/doopedia/master/master.do?\\_method=view&MAS\\_IDX=101013000868431](https://www.doopedia.co.kr/doopedia/master/master.do?_method=view&MAS_IDX=101013000868431).
- [30] <https://terms.naver.com/entry.nhn?cid=40942&docId=1184530&categoryId=34593>.
- [31] [https://www.seoul.co.kr/news/newsView.php?id=20200916018003&wlog\\_tag3=naver](https://www.seoul.co.kr/news/newsView.php?id=20200916018003&wlog_tag3=naver).
- [32] <https://terms.naver.com/entry.nhn?cid=42266&docId=2274625&categoryId=51094>.
- [33] <http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=001&oid=014&aid=0000203142>.
- [34] <https://news.mt.co.kr/mtview.php?no=2009052010243171939&outlink=1&ref=https%3A%2F%2Fsearch.naver.com>.
- [35] <http://www.korea.kr/news/reporterView.do?newsId=148807811>.
- [36] <https://www.fnnews.com/news/202001090915078772>.
- [37] <https://terms.naver.com/entry.nhn?cid=43667&docId=1847190&categoryId=43667>.
- [38] [http://news.khan.co.kr/kh\\_news/khan\\_art\\_view.html?art\\_id=201903061006001](http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=201903061006001).
- [39] <http://news.kbs.co.kr/news/view.do?ncd=5000597&ref=A>.
- [40] [https://www.mfds.go.kr/brd/m\\_629/view.do?seq=16](https://www.mfds.go.kr/brd/m_629/view.do?seq=16).

#### 저 자 소 개

김 도 희(Do-Hee Kim)

준회원



- 2019년 8월 : 숙명여자대학교 홍보 광고학과(문학사)
- 2020년 3월 ~ 현재 : 숙명여자대학교 빅데이터분석융합학 석사과정

〈관심분야〉 : Advertisement & Data Analysis

김민정(Min-Jeong Kim)

정회원



- 1991년 2월 : 연세대학교 응용통계학과(경제학사)
- 1993년 2월 : 서울대학교 산업공학과(공학석사)
- 2012년 2월 : 서울대학교 산업공학과(공학박사)
- 2015년 3월 ~ 현재 : 숙명여자대학교

교 소비자경제학과 부교수

〈관심분야〉 : User & Data Analysis