

LDA 토픽 모델링을 활용한 무선이어폰 소비자 구매 의도 및 만족도 비교 분석: 쿠팡에서의 삼성과 애플 무선이어폰 리뷰를 중심으로

영덕톨¹, 강태구^{2*}

¹충북대학교 경영정보학과 박사수료, ²건양사이버대학교 IT비즈니스학과 교수

Wireless Earphone Consumers Using LDA Topic Modeling Comparative Analysis of Purchase Intention and Satisfaction: Focused on Samsung and Apple wireless earphone reviews in Coupang

Tuul Yondon¹, Tae-Gu Kang^{2*}

¹Student, Dept. of Management Information Systems, Chungbuk National University

²Professor, Dept. of IT Business Administration, Konyang Cyber University

요약 소비자 리뷰분석은 제품 개발, 고객 만족, 경쟁 우위 및 효과적인 마케팅을 위해 중요하다. 무선이어폰 사용 증가는 생활 방식의 성장과 함께 2026년까지 457억 달러에 이를 것으로 예상된다. 따라서 시장의 성장과 중요성을 고려하여, 애플과 삼성의 무선이어폰에 대한 소비자 리뷰를 분석하였다. 본 연구에서는 쿠팡에서 판매되고 있는 애플, 삼성의 무선이어폰 리뷰 11,320건을 수집하여 텍스트마이닝의 빈도, 감성, LDA 토픽 모델 분석을 통해 소비자의 구매의도를 분석하고 소비자 만족도를 분석에 대한 결과를 도출하였다. 토픽 모델링 결과 16개 토픽이 도출되어, 음질, 연결, 쇼핑몰 서비스, 구매의도, 배터리, 배송, 가격 등으로 분류되었다. 브랜드 비교 결과 삼성은 선물 목적으로 구매를 많이 하였고, 가격에 긍정 감정이 높았고, 애플은 배터리, 음질, 연결, 서비스, 배송에 대한 긍정 감정이 높은 결과로 나타났다. 본 연구의 결과는 제조업, 소매업, 마케터 및 소비자를 포함한 고객 만족도, 품질 및 시장 동향에 대한 개선과 통찰력을 얻을 수 있는 연구 결과로 관련 산업의 자료로 활용될 수 있을 것이다.

키워드 : 무선이어폰, 리뷰분석, 빈도 분석, 감성 분석, 토픽모델링, 애플, 삼성

Abstract Consumer review analysis is important for product development, customer satisfaction, competitive advantage, and effective marketing. Increased use of wireless earphones is expected to reach \$45.7 billion by 2026 with growth in lifestyle. Therefore, in consideration of the growth and importance of the market, consumer reviews of wireless earphones from Apple and Samsung were analyzed. In this study, 11,320 wireless earphone reviews from Apple and Samsung sold on Coupang were collected to analyze consumers' purchase intentions and analyze consumer satisfaction through analysis of the frequency, sensitivity, and LDA topic model of text mining. As a result of topic modeling, 16 topics were derived and classified into sound quality, connection, shopping mall service, purchase intention, battery, delivery, and price. As a result of brand comparison, Samsung purchased a lot for gift purposes, had a high positive sentiment for price, and Apple had a high positive sentiment for battery, sound quality, connection, service, and delivery. The results of this study can be used as data for related industries as a result of research that can obtain improvements and insights on customer satisfaction, quality and market trends, including manufacturing, retail, marketers, and consumers.

Key Words : Wireless earphones, Review analysis, Frequency analysis, Emotional analysis, Topic modeling, Apple, Samsung

*Corresponding Author : Tae Gu Kang(tgkang@kycu.ac.kr)

Received June 6, 2023

Accepted August 20, 2023

Revised June 26, 2023

Published August 28, 2023

1. 서론

소비자의 온라인 리뷰는 제품이나 서비스를 구매한 소비자들이 인터넷상에서 해당 제품이나 서비스에 대한 평가와 경험을 공유하는 것을 말한다[1]. 이는 소비자들이 제품이나 서비스를 선택하는 데 있어 중요한 정보를 제공하는 역할을 한다[2]. Bright Local(2023)이 2021년에 실시한 설문조사에서 소비자의 92%가 구매 결정을 내리기 전 온라인 리뷰를 읽는 것으로 나타났으며, 이 중 57%는 온라인 리뷰가 구매 결정에 상당한 영향을 미친다고 응답했다[3]. Power Reviews의 연구에 따르면 97%의 소비자가 고객 리뷰가 구매 결정에 영향을 준다고 답했으며, 95%의 소비자가 구매 전 리뷰를 적극적으로 찾아 읽었다고 하였다[4]. 온라인 리뷰는 매출에도 직접적인 영향을 미칠 수 있는 것을 알 수 있다. 슈피겔 리서치 센터의 연구에 따르면 리뷰가 있는 제품은 리뷰가 없는 제품보다 구매 가능성이 270% 더 높으며, 평점이 높고 리뷰가 많은 제품은 평점이 낮고 리뷰가 적은 제품보다 더 잘 팔리는 경향이 있었다. 따라서 기업들이 매출과 브랜드 이미지를 유지하거나 개선하기 위해서는 온라인 리뷰에 관심을 기울이고 온라인 평판을 적극적으로 관리하는 것이 필수적이다[5]. 쿠팡은 많은 온라인 쇼핑물 중에서도 한국에서 가장 대중 적으면서도 대표적인 온라인 쇼핑물 중 하나로 많은 소비자가 쿠팡을 이용하고 있으며, 다양한 브랜드와 제품을 보유하고 있다. 다양한 제품 중 무선이어폰 분야에서 삼성과 애플은 치열한 경쟁 업체로 쿠팡에서 많은 판매가 되고 있다. 쿠팡은 상품 리뷰를 실시간으로 반영하여 소비자들이 실제로 구매한 제품에 대한 의견을 공유할 수 있도록 하며, 빠른 변화와 트렌드의 추이를 파악하는 데 도움을 준다. 따라서 소비자가 제품을 구매하려는 의도를 비교 분석하기 위한 적합한 환경을 제공한다. 2022년 월평균 쇼핑 앱 사용자 수 1위는 쿠팡이며, 또한 온라인 쇼핑물 10대에서 60대 이상의 소비자들은 많은 쇼핑물 중에서 쿠팡을 최우선으로 선호하였다[6].

텍스트마이닝 분석은 소비자의 리뷰를 분석하는 데 유용한 방법이다[7]. 많은 소비자가 인터넷에 제품 리뷰를 작성하기 때문에 리뷰 분석 방법은 많은 양의 데이터를 처리할 수 있어 다양한 트렌드를 파악하는 데 유용하다[8]. 소비자 리뷰는 실제로 제품을 사용하면서 느낀 장·단점과 문제점을 나타내기에 제품의 실제 사용 경험을 반영할 수 있다. 리뷰는 제품의 기능, 디자인, 가격, 서비스 등

다양한 측면에서 평가할 수 있어서 제품에 대한 종합적인 평가를 할 수 있다[9]. 따라서, 리뷰 분석 방법은 대량의 데이터를 활용하여 제품에 대한 종합적인 평가 결과를 통해 소비자들의 실제 사용 경험으로 제품 개발 및 마케팅에 유용한 정보를 제공할 수 있다.

무선이어폰은 블루투스나 무선 기술을 사용하여 스마트폰, 태블릿, 노트북 등과 연결하여 음악을 듣거나 전화 통화를 할 수 있는 이어폰이다. 무선이어폰은 편리성, 기능 발전, 시장 수요, 경쟁 구도 등의 이유로 빠르게 발전하고 있다[10,11]. 무선이어폰 시장에서 삼성과 애플은 전 세계에서 가장 큰 시장 점유율을 보유하고 있으며, 강력한 입지를 갖고 있다. 이들 브랜드의 무선이어폰은 소비자들에게 매우 익숙하면서도 인지도가 높다. 따라서, 삼성과 애플의 무선이어폰에 대한 리뷰는 다른 브랜드와 비교하여 더 많은 관심과 참고 가치가 있다. 또한, 각각의 제품은 독특한 기술적 특징과 장점을 갖고 있다. 이들의 제품을 대상으로 비교 분석을 수행함으로써, 소비자들이 이들 제품 간의 차이를 인식하고 구매 결정에 영향을 받을 수 있는 요인들을 파악할 수 있다. 또한, 사용자 경험은 다른 브랜드보다 더욱 많기 때문에 소비자들이 자신의 경험을 리뷰로 남기는 경우 다른 브랜드에 비해 상대적으로 많다. 무선이어폰의 세계 시장은 건강, 게임, 가상현실, 음악 엔터테인먼트와 같은 생활 방식의 성장과 함께 지속적으로 성장하여 2026년까지 457억 달러에 이를 것으로 예상된다[12,13].

이에 본 연구에서는 무선이어폰 시장을 주도하고 있는 대표 제조업인 애플 및 삼성의 무선이어폰을 대상으로 온라인 쇼핑물 쿠팡에 작성된 리뷰를 수집한 데이터를 바탕으로 애플 및 삼성 무선이어폰의 긍정적, 부정적 단어 및 주제를 비교하여 구매의도를 도출하였다. 이를 위해 텍스트 마이닝 기법의 빈도 분석 및 감정분석을 통해 주요 키워드를 도출하여, LDA 토픽 모델링, 감정분석을 통해 무선이어폰에 관한 소비자 의견의 주요 요인을 살펴보고, 긍정적, 부정적 리뷰를 확인하여 소비자 구매의도 분석하고 소비자 만족도를 분석하고자 한다.

2. 이론적 배경

2.1 상품 리뷰

온라인 리뷰는 대개 온라인 쇼핑물, 소셜 미디어, 블로그 등 다양한 플랫폼에서 작성된다. 리뷰의 종류에는 별

점 평가, 글 리뷰, 이미지 또는 동영상 리뷰 등이 있으며, 이들은 다양한 소비자들의 의견을 수집하여 제품이나 서비스의 품질, 기능, 가격, 디자인, 서비스 등 다양한 측면을 평가하고 비교하는 데 유용한 정보를 제공한다. 소비자들은 온라인 리뷰를 통해 제품이나 서비스를 선택할 때 더욱 명확하고 정확한 정보를 얻을 수 있으며, 제조업체들은 이를 통해 제품 및 서비스의 개선점을 파악하고 고객들의 요구에 부합하는 제품을 출시할 수 있다[14,15]. 또한, 온라인 리뷰가 모바일 쇼핑의 잠재적 변수를 파악하는데 도움이 된다는 연구도 진행되었다[16,17].

2.2 텍스트 마이닝

텍스트 마이닝은 대량의 텍스트 데이터에서 의미 있는 정보를 추출하는 기술을 의미한다. 이는 자연어 처리와 데이터 마이닝 기술을 활용하여 텍스트 데이터를 분석 후 처리하여, 문서의 내용, 감성, 주제 등과 같은 정보를 파악할 수 있도록 한다[18]. 텍스트 마이닝은 정보 검색, 문서 분류, 감성 분석, 토픽모델링, 키워드 추출 등 다양한 분야에서 활용되고 있으며[19] 많은 연구가 진행되고 있다. 김도훈 등(2019)의 연구에서는 키즈폰(Kids Phone)에 대한 소비자 리뷰를 온라인 쇼핑몰에서 수집하고, TF/IDF, 감성 분석, 네트워크 분석을 통해 제품에 대한 서비스 개선 전략을 도출하였다[20]. 박경원 등(2020)은 '한식'에 대한 한국인 소비자 인식 파악을 위해 네이버 블로그와 뉴스 데이터에 대한 빈도, TF-IDF, CONCOR를 분석하여 한식에 대한 연상 이미지 강화 방안의 연구 결과로 마케팅 측면의 시각과 콘텐츠 개발을 제안하였다[21]. 강옥건 등(2018)의 패키징에 대한 소비자들의 주요 인식에 관한 연구에서는 네이버에서 '포장', '패키징'에 대한 데이터를 수집하고, 빈도, CONCOR 분석 방법을 통해 소비자 인식에 관한 연구를 진행하였다[22]. 또한, 노인 헬스케어 앱 활성화를 위한 기초자료를 제시한 연구에서는 네이버, 다음, 블로그 웹, 카페에서 데이터를 수집하여 TF-IDF, 워드 클라우드, 감성 분석, 의미 연결망 분석[23]과 코로나19 관련된 연구 주제를 탐색하고 연구의 동향을 검토한 연구에서는 디비피아에서 "코로나19", "COVID19"에 대한 데이터를 수집한 후 LDA 토픽모델링, 감성 분석[24]에 대한 다양한 연구가 진행되었다.

2.3 감성 분석

감성 분석은 자연어 처리 기술을 사용하여 텍스트에서

긍정, 부정인 감성을 자동으로 분석하는 기술이다. 이는 기계가 사람의 언어를 이해하고, 텍스트에 담긴 감성을 추론할 수 있는 능력을 갖추도록 하는 것을 의미한다[1]. 감성 분석을 활용하여 많은 연구가 진행되고 있으며, 온라인 리뷰에 감성분석을 적용한 국내 연구들을 살펴보면, 영화 리뷰 데이터의 감성 분석으로 평점을 예측하는 연구를 수행하였으며[25], 게임 리뷰의 수와 리뷰의 감성지수가 게임 판매변화량에 영향을 미친다는 연구 결과도 있었다[26]. 호텔 리뷰 데이터를 수집하여 호텔에 대한 감성 사전을 구축하고, 감성 분석을 통하여 리뷰의 감성지수와 호텔의 전체 평점과의 인과관계를 확인하였다[27]. 국외에서 진행된 감성 분석 연구들도 국내 연구들과 유사하게 책, 영화, 카메라 등의 리뷰를 수집하여 매출이나 매출 순위를 예측하는 연구들이 진행되었다[28-30]. 또한, 항공사 서비스 만족 요인 연구에서는 리뷰글에 있는 한국어 감성 사전을 이용해서 감성 분석에 관한 결과 도출과[31], 임영희 등(2019)은 외래관광객의 온라인 리뷰를 활용하여 감성 분석을 통해 만족과 불만족으로 강하게 나타난 단어들 추출하고 분석을 통한 시사점을 제언하였다[32].

2.4 LDA 토픽모델링

토픽 모델링은 대량의 텍스트 데이터에서 주제(topic)를 추출하여 분석하는 기술이다. 이를 통해 많은 양의 비정형 데이터에서 의미 있는 정보를 추출할 수 있다[33]. 따라서 쇼핑몰, 헬스케어, 관광, 의료서비스 등 분야에서 토픽 모델링 기법은 통해 대규모 데이터를 처리하면서, 고객의 의견과 트렌드, 만족 요인 및 요구사항을 파악할 수 있는 강력한 분석 도구로 인식되고 있다[34-36]. 김광국 등(2018)은 모바일 쇼핑 앱 사용자들의 직접적인 요구사항을 도출하기 위하여 사용자 리뷰 토픽모델링을 분석하여 고객만족도에 미치는 영향을 구체적이고, 다양한 요인들을 도출하였다[37]. 또한, 박재훈 등(2021)은 온라인 상에 저장되는 사용자 후기 정보를 텍스트마이닝의 감성 분석과 LDA분석을 활용하여 화장품 브랜드의 만족도와 결정요인을 분석하였다[38]. 토픽 모델링은 텍스트 마이닝 기법 중 하나로, 통계적 추론을 이용하여 문장들 속에 숨겨진 주제(topic)를 찾기 위한 방법이다. 이를 위해 일반적으로 LDA(Latent Dirichlet Allocation)라는 확률적 생성 모델을 사용한다. LDA는 각 문서가 다양한 주제들을 혼합하여 구성되어 있다고 가정한다. 즉, 각 문서에서

특정 주제가 나타날 확률이 존재함으로써 해당 주제에는 여러 개의 단어들의 조합으로 이루어진다[39]. Fig. 1에서 K 는 토픽의 개수, 말뭉치(D)는 문서들의 집합이고, 문서(N)는 단어들의 집합이라 할 수 있는데, 잠재된 토픽들은 추출된 단어($W_{d,n}$)들을 통해서 추론할 수 있다. 코퍼스 단위로 정해지는 값 α 와 β 중에서 α 는 θ 값에 영향을 주는 파라미터이고, β 는 토픽에 생성할 단어(W)를 결정하는 파라미터이다. 또한, η 는 β 값에 영향을 주는 파라미터이다. 즉, θ 는 문헌별 주제 분포(topic proportions)이고, Z 는 단어별 토픽 할당(topic assignment)이다. 또한, 주제(topic, β)는 추론하고자하는 잠재 구조(hidden structure)이며, 각각 문헌별 주제 분포인 θ 값과 주제인 β 값을 결정하는 파라미터는 α 와 η 이다. 이처럼 Z (토픽)는 문헌별 주제 분포인 값에 따라 결정되고, Z 값과 β 값에 따라 단어(W)가 결정된다고 할 수 있다[33].

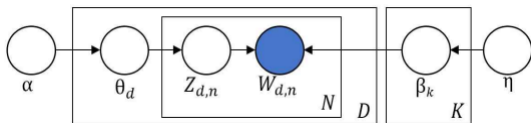


Fig. 1. LDA topic model

LDA는 사전학습 과정 없이 주어진 문서 속에서 확률과 추론을 통해 숨겨진 주제의 구조를 찾을 수 있는 특징이 있다[40,41]. LDA 토픽 모델은 문서 속에서 자주 함께 나타나는 단어들의 그룹을 하나의 토픽으로 인식한다. 이때, 토픽은 우리가 평소에 일상 대화에서 사용하는 주제와는 다른 개념이고, 실험 데이터에서 단어들의 출현 비

율에 따라서 정해지는 주제를 말한다.

3. 연구 방법

3.1 연구 모형 및 절차

본 연구의 연구 모형은 Fig. 2와 같다. 본 연구는 데이터 수집, 데이터 전처리, 데이터 분석, 분석을 통해 소비자의 구매의도에 대해서 도출하였다. 데이터 수집을 위해 온라인 쇼핑몰 쿠팡에서 애플 및 삼성의 리뷰 데이터를 수집하였다. 다음단계로 데이터 전처리를 위해 정제 및 정규화, 토큰화, 정수 인코딩, 패딩을 진행하였다. 데이터 분석 단계에서는 감성 분석, 빈도 분석, 토픽모델링을 진행하였다. 단계적 분석을 통해 긍정적, 부정적 단어 및 주제를 비교 및 도출하여, 절차에 따른 연구 진행을 통해 구매의도에 대한 데이터를 도출하였다.

3.2 데이터 수집

Python의 Requests를 사용하여 쿠팡에서 상품 페이지를 다운받고, BeautifulSoup을 사용하여 상품의 페이지 HTML에서 리뷰 내용을 구별할 때 활용하였다. 수집한 데이터는 2020년 8월 6일부터 2023년 4월 14일까지 작성된 애플, 삼성의 총 6개 무선이어폰의 11,320개의 리뷰를 확보하였다.

3.3 데이터 전처리

데이터 전처리 과정은 정규 표현식 사용으로 한글을 제외하고 모두 제거해준다. 또한 빈 샘플이 생기지는 않는지 확인한다. 토큰화 작업을 수행하면서 불용어(Stop

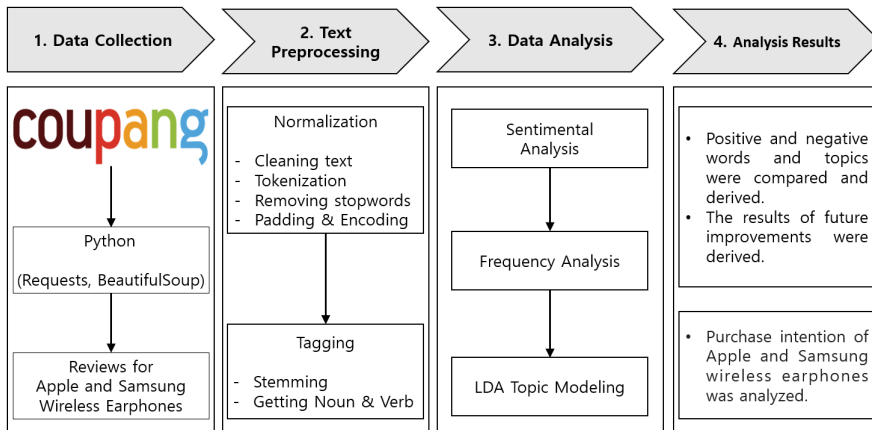


Fig. 2. Research model and progressing

words)를 지정하여 필요 없는 토큰들은 제거하도록 하고 훈련 데이터와 테스트 데이터에 대해서 동일한 과정을 거쳤다. 정수 인코딩에서는 훈련 데이터에 대해서 단어 집합(vocabulary)을 만들고 단어 집합이 생성되는 동시에 각 단어에 고유한 정수가 부여되었다. 다음과정으로 서로 다른 길이의 샘플들의 길이를 동일하게 맞춰주는 패딩 작업을 진행하였다. KoNLPy의 Okt를 활용하여 태깅하였다. 이는 각 단어가 어떤 품사를 나타내는지를 구분하는 작업이다. 이를 통해 단어의 의미를 파악하고, 문맥에 따라 다르게 해석되는 단어들을 구분할 수 있다.

3.4 데이터 분석

감성 분석, 빈도 분석, LDA 토픽 모델링 순서대로 진행하였다. 감성 분석을 위해 네이버 쇼핑물 데이터에서 학습한 deep learning 방법을 활용하였다[39]. 빈도 분석은 Python의 Sklearn의 CountVectorizerfunction를 활용하여 가장 많이 나타난 단어를 구별하여 WordCloud로 시각화하였다. 다음으로 빈도 분석을 바탕으로 LDA 토픽 모델링을 하였다. 마지막으로 감정분석, 토픽모델링의 결과 바탕으로 쿠팡에서 애플, 삼성의 무선이어폰을 구매한 소비자의 만족도를 분석하였다.

4. 분석 결과

4.1 데이터 대상

쿠팡의 무선이어폰 상품 대상으로 Table 1과 같이 애플 및 삼성의 리뷰를 수집하였다. 상품 리뷰별 시작 날짜가 달랐으며, 평점을 보면 상대적으로 별 5개인 긍정적인 리뷰가 많았다.

Table 1. Collection data content

	Samsung	Apple	Total
Number of products	3	3	6
Start Date	20.08.06	19.04.22	19.04.22
End Date	23.04.14	23.04.14	23.04.14
Number of reviews	5173	6147	11320
Product ratings	1	235	1287
	2	116	466
	3	318	1363
	4	869	1627
	5	3635	1404

4.2 빈도 및 감정분석 결과

본 연구에서 Fig. 3와 같이 무선이어폰 브랜드별 리뷰

중에 긍정적, 부정적인 상위 150개 키워드를 시각화 기법을 이용하여 워드 클라우드(Word Cloud) 형식으로 정리하였다.

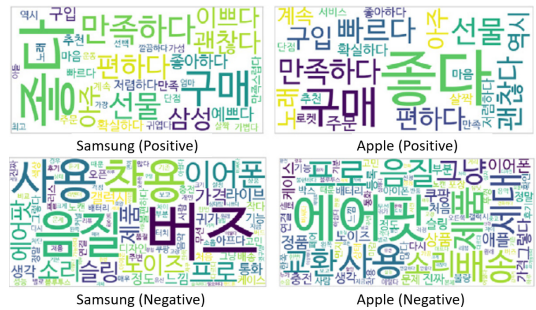


Fig. 3. Positive and negative word cloud by wireless earphone brand

삼성 및 애플의 무선이어폰 상위 10개의 긍정적 키워드 및 키워드별 감성값의 결과를 Table 2와 같이 삼성은 1순위 좋다, 2순위 만족, 3순위 선물 순의 결과가 도출되었고, 애플의 경우에는 1순위 좋다, 2순위 선물, 3순위 빠르다 순의 결과가 도출되었다.

Table 2. Positive keyword for wireless earphones

p o s i t i v e k e y w o r d s	Samsung			
	No.	Keyword	Frequency	Sentiment score
	1	Good	4644	0.869
	2	Satisfied	941	0.907
	3	Gift	773	0.618
	4	Comfortable	669	0.85
	5	Pretty	499	0.86
	6	Certainly	378	0.753
	7	Inexpensive	341	0.637
	8	Fast	288	0.82
9	Cute	216	0.873	
10	Light	194	0.735	
Apple				
No.	Keyword	Frequency	Sentiment score	
1	Good	3099	0.869	
2	Gift	485	0.618	
3	Fast	465	0.82	
4	Satisfied	440	0.907	
5	Comfortable	361	0.85	
6	Order	257	0.65	
7	Certainly	252	0.753	
8	Inexpensive	247	0.637	
9	Suggestion	199	0.523	
10	Rocket	185	0.556	

* negative<0.5 positive

삼성 및 애플의 무선이어폰 상위 10개의 부정적 키워

드 및 키워드별 감성값의 결과를 Table 3과 같이 삼성은 1순위 음질, 2순위 소음, 3순위 아프다 순의 결과가 도출되었고, 애플의 경우에는 1순위 음질, 2순위 교환, 3순위 소리 순의 결과로 도출되었다.

Table 3. Negative keywords for wireless earphones

negative keywords	Samsung			
	No.	Keyword	Frequency	Sentiment score
	1	Sound Quality	2331	0.311
	2	Noise	1155	0.144
	3	Sick	728	0.076
	4	Small	655	0.058
	5	Inconvenient	534	0.01
	6	Problem	263	0.073
	7	Worry	254	0.164
	8	Too Bad	254	0.068
9	Not Much	236	0.01	
10	Cut-Off	192	0.462	
Apple				
No.	Keyword	Frequency	Sentiment score	
1	Sound Quality	1104	0.311	
2	Exchange	1076	0.173	
3	Sound	956	0.264	
4	Noise	707	0.144	
5	Charging	639	0.181	
6	Problem	478	0.073	
7	Battery	441	0.16	
8	Faulty	422	0.022	
9	One Side	316	0.145	
10	Call	311	0.182	

* negative(0.5 positive)

4.3 LDA 토픽 모델링 및 감성분석 결과

4.3.1 토픽 수

LDA 토픽 모델링의 토픽 수를 정할 때 일최적의 토픽 수는 연구자가 결정해야 하는데 일반적으로 정합점수 (Coherence Score)를 이용한다. 토픽이 의미론적으로 얼마나 일관성이 있는지를 나타내는 점수로서 K 토픽의 수가 지속적으로 증가하고 높은 점수를 얻었다 토픽의 수에서 급격히 감소하는 K의 값을 결정한다. 본 연구의 토픽 수를 16으로 정하였다.

4.3.2 LDA 토픽 모델링 시각화

LDA 토픽 모델링 시각화는 Fig. 5와 같다. 결과에서 각 단어가 어떤 토픽에 속하는지 시각화하여 단어들의 분포를 파악할 수 있다. 이를 통해 각 단어가 어떤 주제와 관련이 있는지, 각 토픽의 대표적인 단어가 무엇인지 등을 파악할 수 있다.

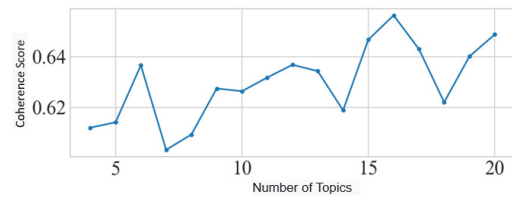


Fig. 4. Coherence score

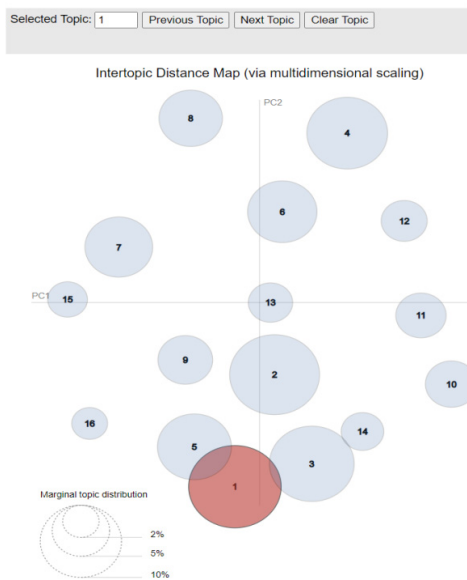
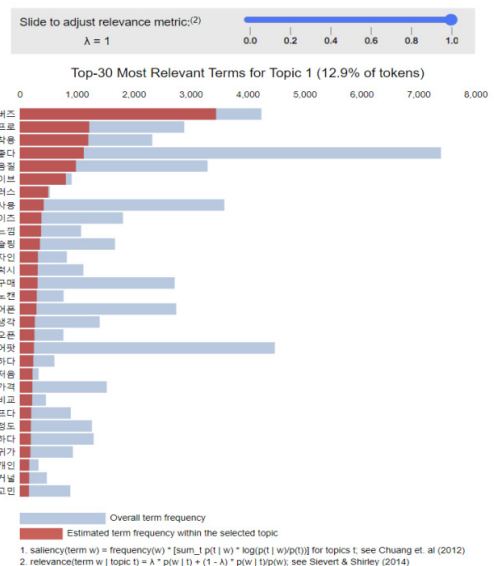


Fig. 5. LDA analysis results



1. saliency(term w) = frequency(w) * [sum_1 p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)p(w)$; see Sievert & Shirley (2014)

4.3.3 LDA 토픽 모델링 네이밍

LDA 토픽 모델링 네이밍은 결과에서 추출된 각 토픽에 대해 의미 있는 이름을 지어주는 것이다. Table 4와 같이 LDA 토픽 모델링 네이밍을 진행하였다. 토픽 1에서 선물, 좋아하다, 아들, 생일, 엄마, 남자친구, 남편, 동생, 아빠, 어머니, 친구, 여자친구, 부모님, 생신, 남친, 아버

지, 조카 등 키워드로 추출되었다. 토픽 1은 무선이러폰의 구매 의도를 보여주며 누군가에게 특별한 날에 선물 용도로 많이 구매가 이뤄진다는 결과를 예측 할 수 있으며, 감정값 0.618로 긍정적인 토픽이다. 이와 같이 토픽별 가장 높은 확률을 가지는 단어를 기반으로 이름을 지어왔다.

Table 4. Topic modeling naming and emotional values

Topic	Key Words	Topic Naming	Sentiment Score
Topic 1	Gift, Like, Son	Purchase intention	0.618
Topic 2	Like, Sound Quality, Call	Sound Quality	0.752
Topic 3	Like, Price, Purchase	Price	0.869
Topic 4	Wearing, Sound Quality, Sound	Sound Quality	0.516
Topic 5	Sick, Ear, Wearing	Sound Quality	0.483
Topic 6	Purchase, Center, Service	Service	0.616
Topic 7	Sound, Around, Sound Quality	Sound Quality	0.803
Topic 8	Long Time, Charge, Battery	Battery	0.907
Topic 9	Connection, Touch, Bluetooth	Connection	0.563
Topic 10	Satisfied, Exchange, Return	Service	0.594
Topic 11	Expensive, Prepare, Not Appropriate	Price	0.581
Topic 12	Box, Packaging, Delivery	Delivery	0.252
Topic 13	Fast, Battery	Battery	0.359
Topic 14	Wearing, Sound Quality, Function	Sound Quality	0.5
Topic 15	Exchange, Faulty, Application	Service	0.64
Topic 16	Delivery, Fast, Rocket	Delivery	0.748

* negative<0.5 positive

토픽 모델링 네이밍에 대한 결과는 Fig. 6와 같이 토픽 14, 7, 2, 4, 5는 음질, 토픽 15, 6, 10은 서비스, 토픽 3, 11은 가격, 토픽 12, 16은 배송, 토픽 8, 13은 배터리, 토픽 1은 구매의도, 토픽 7은 연결 등의 중요한 주제들이 도출되었다.

4.3.4 무선이러폰 브랜드별 토픽 차이

토픽별 삼성 및 애플의 차이는 Fig. 7과 같은 결과가 나타났다.

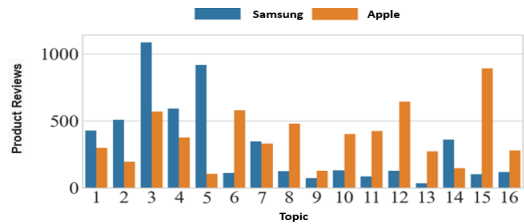


Fig. 7. Differences in topic models by brand

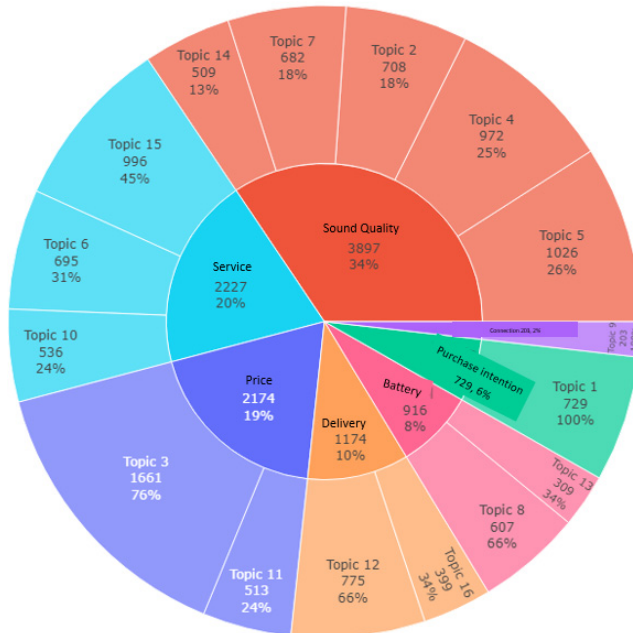


Fig. 6. Topic modeling naming

토픽1 구매 의도를 살펴보면 삼성은 애플보다 높게 나타났다. 토픽 2, 4, 5, 7, 14번 음질을 살펴보면, 토픽2에서 삼성이 애플보다 거의 2배 높으며, 토픽 2는 긍정적인 감정을 나타낸다. 토픽4에서 삼성이 애플보다 높게 나타났으며, 토픽 4는 부정적인 감정을 나타낸다. 토픽5에서 삼성이 애플보다 급격히 높게 나타났으며, 토픽 5는 부정적인 감정을 나타낸다. 토픽7에서 삼성이 애플 거의 비슷하게 나타났으며, 토픽 7은 긍정적인 감정을 나타낸다. 토픽 14에서 삼성이 애플보다 급격히 높게 나타났으며, 토픽 14는 부정적인 감정을 나타낸다. 따라서 무선이어폰 음질의 만족도는 삼성보다 애플이 높게 나타났다. 토픽 3, 11번 가격을 살펴보면, 토픽 3에서 삼성이 애플보다 2배 가까이 높으며, 토픽 3는 긍정적인 감정을 나타낸다. 토픽 11에서 삼성이 애플보다 낮게 나타났으며, 토픽 11은 부정적인 감정을 나타낸다. 따라서 무선이어폰의 가격은 애플보다 삼성이 더 만족도가 높게 나타났다. 토픽 6, 10, 15번 쇼핑물 서비스를 살펴보면, 토픽6에서 애플이 삼성보다 급격하게 높으며, 토픽 6은 긍정적인 감정을 나타낸다. 토픽 10에서 애플이 삼성보다 높게 나타났으며, 토픽 10은 긍정적인 감정을 나타낸다. 토픽 15에서 애플이 삼성보다 급격히 높게 나타났으며, 토픽 15는 긍정적인 감정을 나타낸다. 따라서 무선이어폰의 쇼핑물 서비스는 애플이 삼성보다 불만족도가 높지만, 또한 만족도는 삼성보다 높게 나타났다. 토픽 8, 13번 배터리를 살펴보면, 토픽 8에서 애플이 삼성보다 급격하게 높으며, 토픽 8은 긍정적인 감정을 나타낸다. 토픽 13에서 애플이 삼성보다 급격하게 높게 나타났으며, 토픽 13은 부정적인 감정을 나타낸다. 따라서 무선이어폰의 배터리에 대한 만족은 삼성보다 애플이 더 만족도가 높게 나타났다. 토픽 9번 연결을 살펴보면, 토픽 9에서 애플이 삼성보다 높으며, 토픽 9는 긍정적인 감정을 나타낸다. 따라서 무선이어폰의 연결에 대한 만족은 삼성보다 애플이 더 만족도가 높게 나타났다. 토픽 12, 16번 온라인 쇼핑물의 배송을 살펴보면, 토픽 12에서 애플이 삼성보다 급격하게 높으며, 토픽 12는 부정적인 감정을 나타낸다. 토픽 16에서 애플이 삼성보다 높게 나타났으며, 토픽 16은 긍정적인 감정을 나타낸다. 따라서 무선이어폰의 쇼핑물 배송은 애플이 삼성보다 불만족도가 높지만, 또한 만족도는 삼성보다 높게 나타났다.

4.3.5 연구결과

삼성의 긍정적 주제로 작용 편함, 예쁘고 귀여운 디자인, 저렴한 가격, 빠른 배송, 가벼운 무게가 결과로 나타났고 애플은 빠른 로켓 배송, 작용 편안함, 저렴한 가격 등이 나타났다. 반면 부정적 감정을 나타내는 것으로 삼성의 경우 음질, 노이즈 불편함, 귀 아픔, 작은 크기의 결과가 나타난 반면 애플의 경우 음질, 노이즈, 소리, 배터리, 통화, 한쪽이 불량 등의 결과를 보였다. 부정적 감정은 소비자들의 불만을 사는 부분으로 앞으로 개선해야 할 부분에 대하여 예측할 수 있다. 다음으로 브랜드별로 비교한 결과 삼성은 선물 목적으로 구매를 많이 하고 있고, 가격에 긍정 감정이 높았고, 애플은 쇼핑물 서비스, 배터리, 음질, 연결, 배송에 대한 긍정 감정이 높았다. 토픽 모델링 결과 16개 토픽이 나타났으며, 음질, 연결, 쇼핑물 서비스, 구매의도, 배터리, 배송, 가격 등으로 분류할 수 있었다.

5. 결론

본 연구에서는 쿠팡에서 판매되고 있는 애플과 삼성의 무선이어폰에 대한 소비자 리뷰 11,320건을 쿠팡에서 수집하여, 빈도 분석, 감성 분석, LDA 토픽 모델링을 통해 구매의도를 분석하였다. 분석 결과 삼성 및 애플의 긍정, 부정적 감정을 나타내는 주제를 추출하여, 소비자들의 불만을 사는 부분으로 앞으로 개선해야 할 부분에 대하여 예측할 수 있을 것이다. 다음으로 토픽 모델링 결과를 통해 브랜드별 비교를 통한 차이점을 제시하였다. 본 논문은 텍스트마이닝 분석을 통해 무선이어폰 제품의 소비자 리뷰를 분석하여 애플과 삼성 제품의 소비자 구매 의도의 차이점과 경쟁력을 파악함으로써 무선이어폰 시장에 대한 이해를 높이고 기업의 제품 개발과 마케팅 전략 수립에 도움을 주는 것을 목표로 하였다. 이를 통해 제품의 강점과 약점, 소비자의 니즈와 요구사항을 파악하여 기업에 중요한 정보를 제공할 수 있다.

따라서 본 연구는 무선이어폰 제품을 다루면서도 텍스트마이닝 분석 기법의 유용성을 보여준다는 차별적인 학술적 시사점을 가지고 있다. 또한, 소비자들이 무선이어폰을 선택하고 구매하는 과정에서 어떤 주제와 특징에 주목하는지, 어떤 요소가 만족도에 영향을 미치는지 등을 탐구할 수 있으며, 브랜드 간의 시장 경쟁력을 평가할 수 있다. 이는 기업 전략 수립 및 제품 개발에 도움을 줄 수

있다는 학술적 시사점이 있다. 실질적인 시사점으로는 소비자 구매 의도의 결과를 통해 제조업체와 소매업체가 제품 개발 및 품질 개선에 대한 피드백을 얻을 수 있다. 이를 통해 고객 만족도를 높여 경쟁 우위를 위해 새로운 마케팅 전략을 수립할 수 있다. 마케팅 담당자는 소비자들의 리뷰를 통해 제품의 특징, 가격, 서비스 등에 대한 긍정적인 평가를 파악하고 이를 활용하여 제품 홍보 및 마케팅 전략을 구성할 수 있다. 소비자는 자신이 구매하려는 제품에 대한 다양한 정보를 손쉽게 수집하고 비교할 수 있으며, 이를 바탕으로 최종 결정을 내릴 수 있다. 이러한 분석 결과는 시장 동향과 소비자 행동에 대한 통찰력을 제공하며, 이를 활용하여 기업 전략 수립 및 마케팅 방안에 활용할 수 있다.

하지만 본 연구의 한계점으로는 분석한 리뷰가 쿠팡에서만 수집되었기 때문에 다른 쇼핑몰에서 판매되는 제품들에 대한 소비자 리뷰를 고려하지 않았다는 점이 있다. 따라서 본 연구 결과는 쿠팡에서만 판매되는 애플과 삼성의 무선이어폰에 대한 소비자들의 의견을 대표하는 것으로 한계가 있다.

이에 향후 연구 방향으로는 다음과 같은 내용을 고려할 수 있다. 첫째, 다양한 쇼핑몰(예: 옥션, 지마켓, 티몬, 롯데몰, SSG 등)과 브랜드의 리뷰 데이터를 포함하여 더 포괄적인 분석을 수행할 수 있으며, 이를 통해 다른 쇼핑몰에서의 소비자 경험과 트렌드를 파악하고, 브랜드 간의 차이점을 비교하여 더 포괄적인 시장 분석을 수행할 수 있다. 둘째, 무선이어폰의 다양한 제품 속성에 대한 분석을 추가할 수 있다. 예를 들어, 음질, 배터리 수명, 편의성, 디자인 등과 같은 속성에 대한 소비자들의 평가와 선호도를 파악하여, 이러한 제품 속성이 구매 의도와 만족도에 미치는 영향을 조사할 수 있다. 셋째, 리뷰 데이터의 시계열 특성을 고려하여 시장 트렌드를 분석하고, 향후 트렌드 예측을 수행하는 것이 의미가 있을 수 있다. 이를 통해 소비자들의 관심사와 요구사항이 어떻게 변화하는지, 시장에서 어떤 제품과 기능이 주목받고 있는지 등을 파악할 수 있다. 넷째, 소비자 리뷰 데이터를 보다 효과적으로 분석하여 제품 개발, 마케팅 전략, 고객 만족 등에 대한 통찰력을 얻기 위해 연구를 진행할 수 있다. 이러한 연구를 통해 소비자 리뷰 데이터를 더욱 효과적으로 분석하여 제품 개발, 마케팅 전략, 고객 만족 등에 대한 통찰력을 얻을 수 있을 것이다.

REFERENCES

- [1] K. G. Jang, S. H. Park. & W. J. Kim. (2015). Automatic Construction of a Negative/positive Corpus and Emotional Classification using the Internet Emotional Sign. *Journal of KIISE*, 42(4), 512-521. DOI : 10.5626/JOK.2015.42.4.512
- [2] N. K. Jang & M. J. Kim. (2020). Material as a Key Element of Fashion Trend in 2010~2019: Text Mining Analysis. *Fashion & Textile Research Journal*, 22(5), 551-560. DOI : 10.5805/SFTI.2020.22.5.551
- [3] Brightlocal. (2023). Local Consumer Review Survey 2023. Retrieved from <http://www.brightlocal.com/research/local-consumer-review-survey/>
- [4] Power Reviews. (2018). New Data: 97% of Consumers Depend on Reviews for Purchase Decisions. Retrieved from <http://www.powerreviews.com/events/consumers-depend-on-reviews/>
- [5] Spiegel Research Center. (2021). How Online Reviews Influence Sales. Retrieved from http://spiegel.medill.northwestern.edu/wp-content/uploads/sites/2/2021/04/Spiegel_Online-Review_eBook_Jun2017_FINAL.pdf
- [6] WISEAPP. (2022). Mobile App Total Settlement Report 2022. Retrieved from <http://www.wiseapp.co.kr/insight/detail/379>
- [7] K. R. Kim & S. S. Kim. (2016). Trend Analysis of Research on Inclusive Education in Korea Using Text Mining. *Korean Journal of Physical, Multiple & Health Disabilities*. 59(4), 133-157. DOI : 10.20971/kcpmd.2016.59.4.133
- [8] G. George, M. Haas, R & A. Pentland. (2014). Big data and management. *Academy of Management Journal*, 57(2), 321-326. DOI : 10.5465/amj.2014.4002
- [9] K. H. Kim & H. W. Byun. (2020). The Analysis of Fashion Trend Cycle using Big Data. *Journal of the Korea Convergence Society*, 11(2), 113-123. DOI : 10.15207/JKCS.2020.11.12.113
- [10] M. K. Nam & E. T. Jung. (2020). A Study on Mobile Hardware User Interfaces of a Manufacturing Data Analysis Platform for SMEs: Focusing on Development of Field-friendly Wearable Wireless Osteoporosis Earphone Devices. *Journal of the Korean Society of Design*

- Culture*, 26(3), 115-127.
DOI : 10.18208/ksdc.2020.26.3.115
- [11] H. C. Kim & H. S. Kim. (2021). A Study on the Growth Stage of City Center According to Residential Development: Focusing on the case of Daegu City. *Journal of the Korean Housing Association*, 32(2), 37-45.
DOI : 10.6107/JKHA.2021.32.2.037
- [12] Arizton Advisory & Intelligence. (2021). True Wireless Headphones Market - Global Outlook and Forecast 2021-2026. Retrieved from <http://shorturl.at/ayFTW>
- [13] Korea Costumer Agency. (2020). Comparative survey of domestic and foreign online sales prices of wireless sound equipment directly purchased overseas. Korea Costumer Agency an investigative report.
- [14] H. J. Jung & K. W. Oh. (2016). Analysis of Outdoor Wear Consumer Characteristics and Leading Outdoor Wear Brands Using SNS Social Big Data. *Fashion & Textile Research Journal*, 18(1), 48-62. DOI : 10.5805/SFTI.2016.18.1.48
- [15] E. M. Kang. (2021). Changes in Fashion Perception Due to the Spread of COVID-19: Focusing on the Comparison between 2019 and 2020. *Journal of the Korean Society of Design Culture*, 27(1), 15-27.
DOI : 10.18208/ksdc.2020.27.1.15
- [16] A. Rese, D. Baier, A. Geyer-Schulz & S. Schreiber. (2017). How augmented reality apps are accepted by consumers : A comparative analysis using scales and opinions. *Technological Forecasting and Social Change*, 124, 306-319.
DOI : 10.1016/j.techfore.2016.10.010
- [17] M. Y. C. Yim, S. C. Chu & P. L. Sauer. (2017). Is Augmented Reality Technology an Effective Tool for E-commerce? An Interactivity and Vividness Perspective. *Journal of Interactive Marketing*, 39, 89-103.
DOI : 10.1016/j.intmar.2017.04.001
- [18] J. M. Lee, J. H. Lee & M. J. Kim. (2017). A Study on Perception of Ski Resort Using Big Data Analysis. *The Korean Journal of Physical Education*, 56(4), 415-430.
DOI : 10.23949/kjpe.2017.07.56.4.27
- [19] J. S. Kim. (2018). A Study on the Perception of Fashion Streaming Service Using Text Mining Analysis: Focused on PROJECT ANNE. *Journal of Fashion Design*, 18(1), 107-118.
DOI : 10.18652/2018.18.1.7
- [20] D. H. Kim & K. J. Cha. (2019). Formulating Strategies from Consumer Opinion Analysis on AI Kids Phone using Text Mining. *The Journal of Society for e-Business Studies*, 24(2), 71-89.
DOI : 10.7838/jsebs.2019.24.2.071
- [21] K. W. Park. & H. K. Yun. (2020). A study on the Domestic Consumer's Perception of "Hansik" with Big Data Analysis: Using Text Mining and Semantic Network Analysis. *Journal of the Korea Convergence Society*, 11(6), 145-151.
DOI : 10.15207/JKCS.2020.11.6.145
- [22] W. G. Kang, E. S. Ko, H. R. Lee & J. N. kim. (2018). A Study of the Consumer Major Perception of Packaging Using Big Data Analysis: Focusing on Text Mining and Semantic Network Analysis. *Journal of the Korea Convergence Society*, 9(4), 15-22.
DOI : 10.15207/JKCS.2018.9.4.015
- [23] H. Byun, S. W. Jeon & E. S. YI. (2022). A Study on the Current Situation and Trend Analysis of The Elderly Healthcare Applications Using Big Data Analysis. *Journal of The Korea Convergence Society*, 13(5), 313-325.
DOI : 10.15207/JKCS.2022.13.05.313
- [24] S. M. Heo & J. Y. Yang. (2021). A Convergence Study on the Topic and Sentiment of COVID19 Research in Korea Using Text Analysis. *Journal of the Korea Convergence Society*, 12(4), 31-42.
DOI : 10.15207/JKCS.2021.12.4.031
- [25] J. T. Jo & S. H. Choi. (2015). Sentiment Analysis of movie review for predicting movie rating. *Management & Information Systems Review*, 34(3), 161-177.
DOI : 10.29214/damis.2015.34.3.009
- [26] K. W. Jung & J. U. Kim. (2018). A Study on the Effects of Online Word-of-Mouth on Game Consumers Based on Sentimental Analysis. *Journal of Digital Convergence*, 16(3), 145-156.
DOI : 10.14400/JDC.2018.16.3.145
- [27] Y. H. Lim & H. B. Kim. (2019). A study on the sentiment analysis using big data of hotels online review. *Korean Journal of Hospitality and Tourism*, 28(7), 105-123.
DOI : 10.24992/KJHT.2019.10.28.07.105

- [28] N. Archak, A. Ghose & P. G. Ipeirotis. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485-1509. DOI : 10.1287/mnsc.1110.1370
- [29] P. K. Chintagunta, S. Gopinath & S. Venkataraman. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5), 783-961. DOI : 10.1287/mnsc.1100.0572
- [30] N. Hu, N. S. Koh & S. K. Reddy. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems*, 57, 42-53. DOI : 10.1016/j.dss.2013.07.009
- [31] S. W. Lee, H. Y. Lee. (2017). Using Korean natural language processing, LDA, and sentiment analysis A Study on Satisfaction Factors for Airline Service. *The Korea Society of Management information Systems*, 323-330.
- [32] Y. H. Lim & H. B. Kim. (2019). A study on the sentiment analysis using big data of hotels online review. *Korean Hospitality and Tourism Academe*, 28(7), 105-123. DOI : 10.24992/KJHT.2019.10.28.07.105
- [33] M. Song. (2017). Text Mining. Seoul, Chunglam Publishing Co. 112-117.
- [34] S. H. Lee, A. R. Jo & H. Y. Lee. (2017). The Medical Service Customer's Satisfaction Factors Extracted from Online Hospital Review Data Using Latent Dirichlet Allocation Method. *Korea Service Management Society*, 18(5), 23-44. DOI : 10.15706/jksms.2017.18.5.002
- [35] K. K. Kim, Y. H. Kim & J. H. Kim. (2018). A Study on Customer Satisfaction of Mobile Shopping Apps Using Topic Analysis of User Reviews. *Journal of Korea Society of Electronic Commerce*, 23(4), 41-62. DOI : 10.7838/jsebs.2018.23.4.041
- [36] H. J. Lee. (2018). Keyword identifications on dimensions for service quality of Healthcare providers. *The Knowledge Management Society of Korea*, 19(4), 171-185. DOI : 10.15813/kmr.2018.19.4.010
- [37] K. K. Kim, Y. H. Kim & J. H. Kim. (2018). A Study on Customer Satisfaction of Mobile Shopping Apps Using Topic Analysis of User Reviews. *The Journal of Society for e-Business Studies*, 23(4), 41-62. DOI : 10.7838/jsebs.2018.23.4.041
- [38] J. H. Park, Y. R. Kim & S. B. Kang. (2021). Customer Satisfaction Analysis for Global Cosmetic Brands: Text-mining Based Online Review Analysis. *Journal of Korean Society for Quality Management*, 49(4), 595-607. DOI : 10.7469/JKSQM.2021.49.4.595
- [39] D. M. Blei. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. DOI : 10.1145/2133806.2133826
- [40] D. M. Blei, T. L. Griffiths & M. I. Jordan. (2010). The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM(JACM)*, 57(2), 1-30. DOI : 10.1145/1667053.1667056
- [41] S. J. Ahn. & W. J. Yoo. (2023). Introduction to Natural Language Processing Using Deep Learning. Retrieved from <https://wikidocs.net/book/2155>

영 덩 톨(Tuul Yondon)

[정회원]



- 2021년 2월 : 충북대학교 경영정보학과(경영학석사)
- 2023년 2월 : 충북대학교 박사수료

- 관심분야 : 온라인 쇼핑몰, 소비자 행동, 텍스트마이닝
- E-Mail : tuul.yo@chungbuk.ac.kr

강 태 구(Tae-Gu Kang)

[정회원]



- 2007년 2월 : 충북대학교 경영정보학과(경영학석사)
- 2021년 2월 : 충북대학교 경영정보학과(경영학박사)
- 2017년 4월~현재 : 건양사이버대학교 IT비즈니스학과 교수

- 관심분야 : MIIS, SAP, e-비즈니스, 빅데이터, 클라우드
- E-Mail : tgkang@kycu.ac.kr