

대형 언어 모델을 활용한 한국어 식품 리뷰 분석: 감성분석과 다중 라벨링을 통한 식품안전 위해 탐지 연구

Korean Food Review Analysis Using Large Language Models: Sentiment Analysis and Multi-Labeling for Food Safety Hazard Detection

최은선¹ · 이경희² · 조완섭^{2*}

충북대학교 빅데이터협동과정¹, 충북대학교 경영정보학과²

요약

최근 온라인 플랫폼에서 구입한 육회를 섭취한 후 식중독 증상을 호소하거나 방울토마토에서 쓴맛이 난다는 리뷰가 뉴스에 등장한 사례가 있다. 이것은 정부 기관, 식품 제조업체나 유통업체가 온라인 플랫폼의 식품 리뷰를 분석하여 식품 위해를 탐지함으로써 소비자 식품안전 위험을 관리할 수 있음을 시사한다. 본 연구는 감성분석과 대형 언어 모델을 활용하여 식품 리뷰를 분석하고, 부정적인 리뷰를 탐지하여 주요 식품안전 위해(식중독, 변질, 화학적 이취, 이물질)를 다중 라벨링하는 분류 모델을 제안한다. 감성 분류 모델에서는 'funnel' 모델이 낮은 False Positive 비율로 부정 리뷰의 오분류 가능성을 최소화하는 데 효과적이었다. 식품안전 위해 다중 라벨링 모델은 GPT-3.5 보다 GPT-4 Turbo를 활용한 것이 재현율과 정확도 모두 96% 이상으로 높은 성능을 보였다. 정부 기관, 식품 제조업체나 유통업체는 제안된 모델을 사용하여 소비자 리뷰를 실시간으로 모니터링하고, 잠재적인 식품안전 문제를 조기에 탐지함으로써 위험을 관리할 수 있다. 이와 같은 시스템은 기업의 브랜드 평판을 보호하고, 소비자 보호를 강화하며, 궁극적으로는 소비자의 건강과 안전을 증진시키는 결과를 가져올 수 있다.

■ 중심어 : 감성분석, BERT, Fine Tuning, 대형언어모델, 다중라벨링, 식품안전

Abstract

Recently, there have been cases reported in the news of individuals experiencing symptoms of food poisoning after consuming raw beef purchased from online platforms, or reviews claiming that cherry tomatoes tasted bitter. This suggests the potential for analyzing food reviews on online platforms to detect food hazards, enabling government agencies, food manufacturers, and distributors to manage consumer food safety risks. This study proposes a classification model that uses sentiment analysis and large language models to analyze food reviews and detect negative ones, multi-labeling key food safety hazards (food poisoning, spoilage, chemical odors, foreign objects). The sentiment analysis model effectively minimized the misclassification of negative reviews with a low False Positive rate using a 'funnel' model. The multi-labeling model for food safety hazards showed high performance with both recall and accuracy over 96% when using GPT-4 Turbo compared to GPT-3.5. Government agencies, food manufacturers, and distributors can use the proposed model to monitor consumer reviews in real-time, detect potential food safety issues early, and manage risks. Such a system can protect corporate brand reputation, enhance consumer protection, and ultimately improve consumer health and safety.

2024년 05월 23일 접수; 2024년 06월 11일 수정본 접수; 2024년 06월 13일 게재 확정.

* 본 연구는 정부(식품의약품안전처)의 출연연구사업 지원을 받아 수행된 연구임(과제고유번호 : KMDF-RnD 21163수입안517-1)

† 교신저자 (wscho63@gmail.com)

■ Keyword : Sentiment Analysis, BERT, Fine Tuning, Large Language Model, Multi-labeling, Food Safety

I. 서론

2022년에 발표된 식품안전정보원의 불량식품 소비자신고 동향 보고서에 따르면 부정 및 불량 식품에 대한 소비자 신고 건수는 2만 1,088건으로, 전년 대비 5.9% 증가했다[1]. 특히, 코로나19로 인해 비대면 문화가 확산되면서 식품의 온라인 쇼핑 및 배달 서비스 같은 온라인 플랫폼의 활성화가 주요 요인으로 지목된다. 온라인 플랫폼을 통해 식품 구매한 소비자는 식품구입 후 식품에 대한 긍·부정 의견을 남기게 되며, 이 데이터를 수집하여 분석함으로써 식품안전 이슈를 모니터링 할 수 있다. 온라인 플랫폼의 식품 리뷰 데이터는 식품안전을 확보하는 새로운 접근법을 모색하는데 주요한 자산이 되고 있다. 이와 같은 추세에 발맞추어 신세계와 풀무원 등 일부 기업들은 소비자 리뷰를 분석하고 품질관리에 적용하기 위해 인공지능 리뷰 분석 시스템을 도입하고 있으며, 리뷰 분석 관련 연구도 활발하게 진행하고 있다. 그러나 한국어로 작성된 리뷰를 머신러닝 기법으로 분석하고 이를 활용하는 연구는 부족한 상황이다.

한국어 리뷰 분석의 어려움은 한국어의 독특한 언어적 특성으로 인해 발생한다. 한국어는 교착어로 단어의 형태가 문장 내에서 변화하며 다양한 문법 규칙에 따라 조사, 어미, 접두사, 접미사 등이 추가된다. 이 때문에 텍스트 분석 및 해석이 복잡해지며, 기존의 사전 기반 감성분석 방법으로는 작성자의 감성을 이해하는 데 한계가 있다[2,3]. 그리고, 감성분석에 필요한 라벨링 작업은 시간과 비용이 많이 소요되며, 소비자 리뷰에 포함된 별점은 리뷰 내용과 일치하지 않는 경우가 있어 별점을 감성분석 평가의 기준으

로 삼는 것은 한계가 있다.

본 연구는 네이버 쇼핑과 신세계몰에서 수집한 식품 리뷰 데이터를 분석하여, 식품 위해 관련 내용을 찾아내고 빠르게 위해 가능성이 있는 식품을 식별하는 데 목적이 있다. 이러한 목표를 달성하기 위한 1단계로 머신러닝을 활용한 감성분석 모델이 리뷰 내용의 긍부정을 분류한다. 2단계로 대형 언어 모델(Large Language Model, LLM)을 사용한 주요 식품 위해(식중독, 변질, 화학적 이취, 이물질)의 다중 라벨링(Multi-Labeling) 분류 모델을 구축한다.

본 연구의 공헌은 한국어 리뷰 분석에서 기존의 자연어 처리 기법(NLP)으로는 어렵고 정확도가 낮은 주제별 라벨링을 대형 언어 모델을 통해 극복하고, 위해도가 높은 리뷰와 상품을 정밀하게 식별하여 식품안전 관리에 활용할 수 있는 신뢰도 높은 방법론으로 제시한 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 정리하고, 3장에서는 분석에 사용한 데이터를 설명하고, 대형 언어 모델을 활용한 식품안전 위해 요소 다중 라벨링 모델구축 방법을 제시한다. 4장에서는 성능평가를 하고, 5장에서는 본 연구의 결론과 한계점에 대하여 기술한다. 이 연구는 한국어 소비자 리뷰를 통한 식품안전 위해 요소 탐지의 중요성을 인식하고, 이를 해결하기 위한 효과적인 방법을 모색함으로써 식품안전 관리에 새로운 방향을 제시하고, 향후 관련 연구에 기여할 수 있다.

II. 관련 연구

2.1 리뷰데이터 감성분석

감성분석(Sentiment Analysis) 또는 감정분석은 자연어 처리, 텍스트 분석, 계산 언어학의 하위 분야로써 주어진 텍스트 내에서 의견, 감정, 주관성 등을 자동으로 식별하고 분류하는 기술이다. 감성분석은 텍스트 작성자의 감정이 긍정, 부정 또는 중립 인지를 판단하는 데 주로 사용되며 소비자 리뷰, 소셜 미디어 메시지, 뉴스 기사 등 다양한 텍스트 데이터 소스에서 유용하게 활용되고 있다.

소진수와 신태섭[4]은 한국어로 작성된 음식점 리뷰를 대상으로 감성분석을 수행하고 세부 평가 항목별 평점을 예측했다. 최혜선과 연구필[5]은 밀키트 제품 리뷰에 SVM, 로지스틱 회귀, 랜덤 포레스트 같은 머신러닝 기법을 적용하여 리뷰에서 나타나는 긍정적 및 부정적 감성의 주요 요인들을 분석했다. 최준영 등 [6]과 정재윤 등[7] 연구에서는 한국어로 작성된 리뷰를 머신러닝 방법론으로 분석한 연구를 진행했다.

신비 등[8]은 Google Maps 상의 음식점 리뷰 데이터를 텍스트 마이닝을 통해 분석하였고, TF-IDF를 사용하여 단어의 중요도를 벡터화하고, 머신러닝 알고리즘인 랜덤포레스트를 사용하여 리뷰에서 사용된 단어의 긍정성 및 부정성 계수를 계산하여 감정 사전을 구축했다. 또한, 방글라데시의 온라인 음식 리뷰와 트위터, Yelp 같은 소셜 미디어 플랫폼에서 수집된 데이터에 머신러닝 및 LSTM과 word2sequence 같은 딥러닝을 적용한 감성분석 연구도 진행됐다[9,10]. 상품 또는 서비스 품질, 소비자 인식에 관련된 연구[11]들도 진행됐다.

백세희 등[12]은 코로나19 이후 지속가능한 관광에 대한 관심이 증가함에 따라, 관광객들의 수요를 충족시키기 위한 구체적인 프로그램과 정책의 부족을 지적하고, 이를 해결하기 위해

뉴스 데이터와 리뷰 데이터를 분석하였다.

2.2 파인튜닝(Fine-tuning)

파인튜닝은 머신러닝과 딥러닝 분야에서 중요한 개념으로 이미 훈련된 모델을 추가 데이터셋으로 재훈련하여 특정 작업의 성능을 향상시키는 과정이다. 이는 대규모 데이터셋을 처음부터 학습하는 것에 비해 시간과 자원을 절약하며 적은 데이터로도 효과적인 성능 향상을 가능하게 한다.

Tang, T. 와 Tang, X 등[13]은 다양한 감정을 포함한 복잡한 텍스트에서 불균형한 감정 분포를 처리하기 위해 BERT 모델을 파인튜닝하는 방법을 연구했다. 이를 위해 데이터 증강, 언더샘플링 및 앙상블 학습을 통합한 새로운 접근법을 사용하여 다수의 멀티 라벨 BERT 분류기를 훈련시키고, 분류 결과를 결합한 최종 결과를 도출함으로써 감정 분석에서 불균형 문제를 효과적으로 해결했다.

노동훈과 민재옥 등[14]은 특허 상담 텍스트를 BERT에 적용한 자동 분류의 성능 향상방안을 연구했다. 특허상담에 특화된 파인튜닝 모델을 적용하여 분류한 결과 기존의 분류 모델보다 우수한 성능을 확인했다.

2.3 대형 언어 모델을 활용한 데이터 처리기법

이원민과 온병원[15]은 감성 및 감정 단어 마스킹 기반의 BERT 모델을 사용하여 감정이 반영된 학습 데이터를 자동으로 생성하고, 이러한 학습 데이터로 GPT 모델을 학습시켜 감정이 잘 반영된 응답 문장을 생성하는 BERT와 GPT를 결합한 파이프라인 모델을 제안했다.

Peskine, Y., Korenčić, D., [16]의 연구에서는 GPT-3 모델을 사용하여 트위터에서 음모 이론과 같은 다양한 주제를 식별하고 분류하는 새로

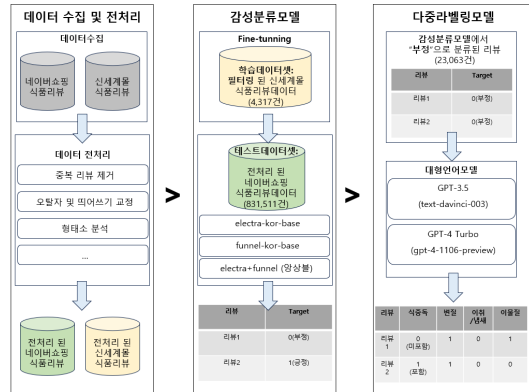
은 방법을 제안한다. 이 과정에서 제로샷 라벨링 기법을 적용하여 주제에 대한 정의를 생성하고, 이를 사용하여 분류 정확도를 향상시킨다. 이 접근법은 라벨링이 어려운 상황에서 유용하며, 새로운 주제가 추가될 때마다 모델이 즉시 적용할 수 있어 분류 시스템을 유연하고 확장성 있게 만들 수 있다.

Yang, S., Shang, Z[17]등은 사전 훈련된 대형 언어 모델(LLM)인 ChatGLM을 사용하여 훈련 데이터 없이 다중 라벨 이미지 인식을 수행했다. 언어적 설명을 기반으로 이미지 내 객체를 식별하고, 이 지식을 활용하여 사전 훈련된 비전-언어 모델(CLIP)을 프롬프트 튜닝함으로써, 시각적 및 언어적 임베딩 공간을 정렬하여 다중 라벨 분류의 효율성을 향상시키는 방법을 제안했다. 이 방법은 전통적인 데이터 중심 접근 방식에서 벗어나, 언어 모델이 제공하는 풍부한 지식을 활용하여 이미지 내의 다양한 객체를 효과적으로 분류하는 데 초점을 맞추고 있다.

III. 식품안전 위해 요소 탐지 모델

본 연구는 식품안전 위해 요소 탐지를 목표로 하는 분석 프로세스를 수행하며, 처리 방법은 <그림 1>과 같다.

프로세스는 데이터 수집 및 전처리 단계에서 시작한다. 이 단계에서는 네이버 쇼핑과 신세계몰에서 식품 리뷰 데이터를 수집하고, 중복된 데이터의 제거와 오타자 교정 등의 전처리 작업을 진행한다. 다음으로 감성 분류 모델 학습 및 평가 단계에서는 사전 학습된 모델을 적용하여 리뷰 데이터를 긍정과 부정으로 라벨링한다. 이 과정은 리뷰 데이터의 90%는 식품 위해보다는 관련이 없는 내용이므로 이를 필터링하기 위한 과정이다. 마지막 단계인 다중 라벨링에서는 ‘부정’으로 라벨링된 리뷰를 대상으로 대형 언어 모델을 활용하여 식품 위해 카테고리에 대한 다



<그림 1> 식품안전 위해요소 탐지 모델 순서도

<표 1> 다중 라벨링 프로세스 별 설명

분석 단계	설명
데이터 수집 및 전처리	플랫폼: 네이버쇼핑, 신세계몰
	도구: Python, Selenium을 활용한 웹 크롤링
	과정: 수집된 데이터에서 불필요한 정보 제거 및 분석에 적합한 형태로 정제
감성 분류 모델 학습 및 평가	모델: 사전에 학습된 electra-kor-base, funnel-kor-base, 앙상블모델(electra+funnel)
	목적: 데이터셋의 감성 분류 및 모델의 예측 능력 평가
다중 라벨링	모델: text-davinci-003(GPT-3.5), gpt-4-1106-preview(GPT-4 Turbo)
	목적: 식품위해 내용을 포함하는 리뷰 및 상품 도출

중 라벨링을 수행한다.

3.1 데이터 수집 및 전처리

본 연구에 활용한 식품 리뷰 데이터는 네이버 쇼핑, 신세계몰 등의 다양한 온라인 쇼핑 플랫폼을 대상으로 웹 크롤러를 이용하여 수집했다. 수집된 데이터는 MySQL 데이터베이스에 저장되며, 일별 데이터 수집의 정확성을 철저히 확인했다.

수집 데이터 중복을 제거하기 위해 ['userId',

‘content’, ‘product’, ‘category1’, ‘category2’, ‘category3’, ‘Date’] 등의 다양한 컬럼을 활용하여 중복 리뷰를 제외하였다. 리뷰의 품질 보증을 위해 마침표(.)를 제외한 모든 특수 문자를 삭제하고, 길이가 10글자 미만이거나 한글이 포함되지 않은 리뷰, 그리고 ‘아직’이라는 단어를 포함한 리뷰는 제거했다. 그리고 ETRI-et5 모델을 기반으로 파인튜닝한 한국어 구어체 전용 맞춤법 교정기인 “et5-typos-corrector” 모델을 활용하여 오타자 및 띄어쓰기 오류를 정정하여 데이터의 품질을 개선했다. 이 과정은 신조어나 띄어쓰기 오류가 자연어 처리(분석)의 정확도를 떨어뜨리기 때문에 이를 최소화하기 위한 것이다. 예를 들어, “아니 진짜 무너하냐고”라는 문장은 “아니 진짜 뭐 하냐고”로 교정되었다. 이러한 전처리 및 교정 작업은 모델이 정확한 예측을 수행할 수 있도록 돕고, 연구 결과의 신뢰성을 강화하는 데 중요한 기능을 한다[18].

3.2 네이버쇼핑 식품 리뷰 데이터 : 테스트데이터셋

이 연구에서 사용된 데이터는 Python 프로그래밍 언어와 Selenium 웹 드라이버를 활용한 자동화된 웹 크롤링 방법으로 수집하였다. Selenium은 웹 브라우저 자동화 도구로, 사용자의 브라우저 조작을 모방하여 데이터를 수집하는 기능을 제공한다. 이를 통해 각 온라인 쇼핑 사이트의 동적 웹 페이지에서도 실시간으로 데이터를 추출할 수 있다. 예를 들어, 신세계몰에서는 특정 URL을 입력하여 결과 페이지의 리뷰 섹션으로 이동하고, 리뷰 텍스트와 관련 메타데이터를 수집했다. 이 과정은 사이트의 구조와 인터페이스에 맞춰 세심하게 설정되었으며, 데이터 수집의 정확성과 신뢰성을 주기적으로 검증하였다. 네이버 쇼핑에서는 식약처에 신고된 수입식품 중 축산물 분류 제품, 축산가공 수입식품의 브랜드명, 상품명 키워드를 기반으로 데이터를 수

집하였다. 신세계몰에서 조리하지 않고 바로 섭취하는 육류(육회, 사시미, 몽티기)와 해산물(연어, 굴 등)의 리뷰 데이터를 수집했다. 수집된 데이터는 별점을 기준으로 라벨링하고 감성 분석을 통해 리뷰를 긍정·부정 감성으로 분류했다. 4점 이상은 긍정적인 감성으로, 3점 이하는 부정적인 감성으로 분류하였다. 이 데이터는 감성분류 모델의 학습 및 파인튜닝에 활용되어, 모델이 리뷰 텍스트에서 긍정적 또는 부정적인 감성을 정확히 인식하고 분류하는 성능을 개선한다. 각 사이트 별로 수집된 데이터는 문서 하단의 <표 2>에 자세히 나타나 있으며, <표 3>에서는 수집된 식품 리뷰 데이터 예시를 확인할 수 있다.

3.3 신세계몰 데이터 : 파인튜닝(Fine-tuning) 데이터

본 절에서는 식품위해 다중 라벨링 모델의 성능을 향상시키기 위하여 식품리뷰를 긍·부정으로 분류하는 감성 분류 모델 구현과정을 설명한다. 식품리뷰 데이터에서 긍정 리뷰가 92% 이상을 차지하는 반면, 부정 리뷰는 7%에 불과하다. 이러한 불균형을 고려하여, 대형 언어 모델을 사용할 때 전체 데이터를 분석하는 것은 시간과 비용 면에서 비효율적이다. 따라서, 우리는 부정적인 리뷰에 초점을 맞춘 감성분석을 통해

<표 2> 사이트별 수집 데이터 정보

	네이버쇼핑	신세계몰
수집 식품명	A	A

	B	C
수집품목 개수	13,158개	1,435개
수집기간	23.06.14~ 23.12.12	14.04.04~ 23.11.26
수집리뷰 개수	831,511 건	1,250,583 건
수집주기	일별수집	비정기적 수집

<표 3> 수집된 데이터 예시 : 네이버 쇼핑

품목명	category 1	category 2	category 3	...	star Score	content
A	식품	축산물	축산가공 식품	...	1	상품이 잘 안까지고 맛도 영 아니네요
B	식품	건강식품	영양제	...	3	차전자피 좋아요 매일 먹으면 소화기 안되서 과식 한날만 먹어요
C	식품	음료	우유/요구르트	...	2	일반토코 우유 생각하면 큰일나요 ㅠㅠ치약맛이 왜 날까요 민트초코도 아닌데 말이죠저걸 아맛게가 먹나요 ㅠㅠ
D	식품	냉동/간편조리식품	즉석국/즉석탕	...	5	유통기한 너무 짧아요.
E	식품	과자/베이커리	초콜릿	...	5	또또또 주문할까 말까 고민고민 중입니다. 살만 안찌면 고민안하고 주문했겠지만.

데이터 필터링을 우선적으로 수행하였다. 이 필터링 과정은 식품위해 요인을 포함하는 리뷰를 선별적으로 다중 라벨링하기 위한 전략으로, 다중 라벨링 작업의 효율성을 높이고, 리소스의 낭비를 줄이는데 기여한다.

특히 감성 분류 모델은 과적합의 위험이 있어, 모델의 긍·부정 분류 정확도를 높이기 위한 파인튜닝이 필수적이다. 효과적인 파인튜닝을 통해, 모델은 긍정 및 부정 리뷰를 보다 정확하게 구분할 수 있으며, 이는 식품위해 다중 분류 모델의 전반적인 성능을 향상시킬 수 있다.

식품리뷰 감성 분류 모델의 파인튜닝에 사용

된 데이터셋은 신세계몰에서 수집한 4,317건의 리뷰 데이터를 활용하였다. 이 데이터는 별점과 무관하게 식품에 대한 부정적인 리뷰를 정확하게 탐지하는 능력을 강화하는 데 사용되었다. 데이터 라벨링은 별점 3점 이하를 부정, 그 이상을 긍정으로 설정하였다.

특히, <표 4>의 3번 유형과 같이 별점은 높지만 리뷰 내용이 부정적인 리뷰의 예시는 <표 5>와 같으며, 감성 분류 모델의 정확도를 높이는 데 활용할 필요가 있다. 식품 온라인 쇼핑몰에서는 별점이 낮은 리뷰를 관리자가 삭제하여 식품에 대한 소비자의 부정적 의견이 삭제되기도

<표 4> 파인 튜닝을 위한 학습 데이터 유형

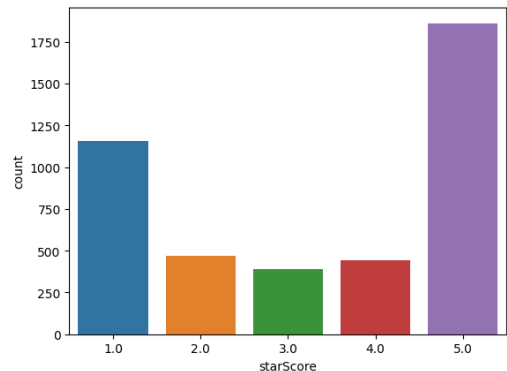
	파인 튜닝 데이터 유형	설명
1	식중독 관련 리뷰	식중독은 식품안전 위해 탐지에서 가장 주요한 이슈로써 식품리뷰 분석에서 매우 중요하므로, 식중독과 관련된 내용을 언급하거나, 관련 단어를 포함하는 리뷰를 선별하여 학습에 파인튜닝에 활용한다.
2	낮은 별점 리뷰	별점이 3점 이하인 리뷰는 식품에 대한 부정 감성을 담은 리뷰로 간주하여 해당 리뷰를 추출한다.
3	별점은 높지만 부정적인 리뷰	감성분석 모델의 결과가 부정(0)으로 분류된 리뷰를 확인하며, 이 중에서 긍정적인 단어가 포함되지 않은 리뷰를 선별한다.
4	특정 카테고리 관련 리뷰	식품안전 위해 탐지에서 중요한 카테고리인 변질, 화학적 이취, 이물과 관련된 단어가 포함된 리뷰를 선별하여 학습강화에 활용한다.
5	“아직”이 포함된 리뷰	“아직”이라는 단어가 포함된 리뷰는 해당 제품을 아직 사용하지 않았거나, 리뷰 보상을 위해 작성된 리뷰로 간주되므로 제외한다.

<표 5> 파인 튜닝에 사용된 데이터 예시

리뷰 예시	star Score	target
짱짱 육회 지방도 거의 없고 너무 맛나요. 가성비도 가성비지만 진짜 맛나게 먹었어요 ㅋㅋㅋ180g이 1인용으로 넉넉한 양이에요. 절대 적지 않음. 재구매 의사 있습니다 후딱먹고 또 주문할게요 다음에도 좋은 것으로 부탁드립니다	5.0	1
음식에 이렇게 나왔습니다. 하지만 판매자는 묵묵부답 이번이 두 번째 구매였습니다. 첫 번째로 육사 시미200g 육회200g을 주문했고 가격대비 퀄리티가 나쁘지 않다고 생각해서 이번에는 육회200g을 2두 개 주문 했습니다. 배송 온 첫 날 하나를 먹었고 다음 날 두 번째 팩 포장을 뜯고 양념 넣고 무치려는데 고기와는 다른 색감과 질감의 무언가가 보이더라고요. 꺼내보니 사진에 보이는 것 처럼 끈 줄 같은 것이 나왔습니다. 분명히 말씀드리지만 소고기의 힘줄은 아닙니다. 뭔가 목장갑 등의 면섬유 재질 같은 것 같아 보입니다. 판매자 분께 자초지종을 설명드리고 확인을 부탁드립니다 문의를 했으나 5 6일이 지나도록 아무런 답변이 없네요. 믿고 먹어야 하는 음식에서 먹어서는 안 될 물건이 나온 것도 충격이지만 이 업체의 대응 방식도 매우 실망스럽습니다.	5.0	1
신중히 생각하고 구매하시길 바랍니다 가성비비는 최고 안 좋은리뷰 있길래 불안했는데 냄새나 이런거도 없고 맛있었음 단지 소분해서 냉동보관하고 먹으려면 하루정도 해동시기필요하다는 귀찮음이 단점물론 오자마자 한번에 다먹으면 상관없음	4.0	1
맛있는데 뭔가 묽은 느낌이 나요	4.0	1
못 먹음 다시는 안시켜요 배송이 늦어져 영망 진창으로 배송와서 기분 찔찔해서 못먹었어요사진 업로드 시키는데 안올라가네요	3.0	0
후추냄새가 너무 심해서 석유냄새같이 느껴져요 ㅠㅠ 한번 냉동 들어갔다 나온 두부라그런지 퍼석퍼석하고 별로예요. 그래서 1 1인가ㅠㅠ 재구매는 안할듯 억지로먹는중여	2.0	0
최악. 다신 안 샅니다. 이걸 누가 먹으라는지. 처음엔 조리법대로 전자렌지 돌렸더니 용기가 찌그러질 정도로 변형이 왔고 시금치 손내가 잔뜩 남. 용기 때문인가 싶어 두번째는 적당히 자연해동하고 냄새맡으니 여전히 시금치 손내. 후라이팬에 조리 할 것도 없이 으스스 행 나물이 선거 같음 생산 라인에서 빼야죠 다른걸 넣거나 리뷰보고 좋다하길래 샀는데 입 다 버렸어요 시금치 잡채 처음 봄 - -	1.0	0

하여 부정 리뷰지만 별점은 5점으로 등록하는 사례가 실제로 많이 있다.

<그림 2>는 파인튜닝에 활용한 데이터 4,317건의 별점 분포를 보여준다. 낮은 별점의 리뷰만 사용할 경우 식품리뷰 감성 분류 모델이 부정적인 리뷰에 과적합 될 위험이 있어서 이를 방지하기 위하여 긍·부정 비율을 6:4로 조절하였다.



<그림 2> 파인튜닝에 활용한 데이터의 별점 분포

3.4 감성 분류 모델 학습 및 평가

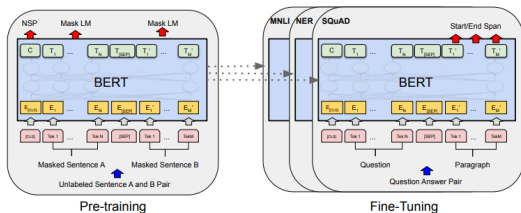
본 연구에서는 식품 리뷰 데이터를 긍·부정으로 분류하는 이진 분류 감성 분류 모델을 적용했다. 이진 분류 방식은 분석의 명확성을 제

공하고 데이터셋 내 클래스 간 불균형 문제를 효과적으로 해결하는 방안이다. 이를 통해 모델의 학습 효율성을 극대화하고 오류를 최소화하

는 데 중점을 두었다.

감성 분류에는 BERT 계열의 “funnel-kor-base”와 “electra-kor-base” 모델이 사용하였으며, 두 모델의 장점을 결합하기 위해 앙상블 기법을 적용했다. “funnel-kor-base”와 “electra-kor-base” 모델은 광범위한 한국어 텍스트 데이터셋을 기반으로 사전 학습되었으며, 비정형 언어 패턴을 효과적으로 처리할 수 있도록 설계되었다[19]. 사전 학습에 활용된 데이터셋은 화장품, 식품, 전자제품, 반려동물 등 다양한 카테고리의 텍스트로 구성되어 있어, 도메인 특화 언어 모델 학습에 매우 적합하다. 이러한 방대한 언어 데이터셋은 모델이 한국어의 다양한 비정형 언어 패턴을 효과적으로 처리하고, 실제 리뷰 데이터에서 나타나는 감성을 정밀하게 분석하고 예측하는 데 중요한 역할을 한다.

본 연구에서 사용된 데이터는 온라인 쇼핑몰 및 소셜커머스 플랫폼에서 수집된 식품 리뷰를 포함하며, 파인튜닝 과정에서도 이러한 식품 리뷰 데이터를 활용하여 자연어로 작성된 식품 리뷰 분석에 특화되어 있다. 감성분석에 대형 언어 모델을 사용하지 않은 이유는 이러한 대규모 모델들이 요구하는 시간과 비용이 많이 소요되기 때문이다. GPT와 같은 모델들은 훈련과 파인튜닝에 상당한 연산 자원을 요구하며, 특히 대용량 데이터셋에 대한 처리 시 상대적으로 높은 비용이 발생할 수 있다. 반면, “funnel-kor-base”와 “electra-kor-base”는 사전 훈련된 모델을 효율적으로 활용하여, 보다 저렴한 비용으로 빠르게 감성분석을 수행할 수 있기 때문에 이번 연



〈그림 3〉 BERT 적용 사전학습 및 파인 튜닝 절차

구에 적합하다고 판단된다.

BERT 모델의 파인튜닝 단계는 사전학습 된 모델을 특정 작업, 예를 들어 감성분석과 같은 태스크에 최적화하는 과정이다[20]. 이 과정에서 모델은 실제 사용 환경에 맞는 세밀한 언어 패턴을 학습하게 되며, 이는 한국어 감성분석의 정확성과 효율성 향상에 기여한다. 감성 분류 모델 구축에 적용된 주요 파라미터는 <표 6>에서 확인할 수 있다.

〈표 6〉 감성 분류 모델 파라미터

파라미터 이름	n
클래스 수	2
K-Fold 교차 검증 횟수	10
최대 시퀀스 길이	50
배치 크기	128
에포크(Epoch)	10
학습률	2e-5
라벨 스무딩 (Label Smoothing) 비율	0.05
난수 시드(SEED)	2,023

이러한 파라미터 설정은 모델이 데이터셋의 특성을 효과적으로 학습하고, 예측 능력을 최대화하는 데 중요한 역할을 한다. 에포크 수, 학습률, 배치 크기, 시퀀스 길이는 모델 학습의 깊이와 범위를 정의하며, 라벨 스무딩은 모델이 과대적합을 방지하고 일반화 능력을 향상시키는 데 도움을 준다. 난수 시드의 설정은 실험의 재현성을 보장하고, 연구의 신뢰성을 확보하는 데 중요하다.

감성 분류 모델의 평가는 다양한 평가지표를 통해 이루어진다. 이러한 지표들은 모델의 성능을 여러 관점에서 평가하고, 모델이 실제 리뷰 데이터에서 긍정 및 부정 리뷰를 얼마나 정확하게 예측하는지 정량적으로 파악하는 데 도움을 준다. 이러한 다양한 평가 메트릭을 통해 모델의 종합적인 성능을 평가하고, 연구의 신뢰성과 정확성을 높일 수 있다. 다만 데이터가 사전에

<표 7> 모델 평가 메트릭

평가 메트릭	설명
정확도	모델이 전체 데이터 중 올바르게 예측한 긍정 및 부정 사례의 비율.
정밀도	모델이 긍정으로 예측한 사례들 중 실제로 긍정인 경우의 비율.
재현율	모델이 실제 긍정 사례 중 올바르게 긍정으로 예측한 비율. 이는 모델이 실제로 긍정적인 사례를 얼마나 잘 감지하는지 나타내는 지표로, 부정 사례에도 적용되어 모델이 실제 부정 사례를 부정으로 얼마나 정확하게 예측하는지를 나타낼 수 있다.
F1 Score	정밀도와 재현율의 조화 평균을 나타내며, 이 두 메트릭의 균형을 평가하는 데 사용됨.
AUC	ROC 곡선 아래의 면적으로, 모델이 실제 긍정 사례를 무작위로 선택한 긍정 사례보다 높은 점수로 분류할 확률을 나타낸다.
Kappa 통계량	관찰된 정확도와 우연에 의한 정확도 사이의 일치 정도를 측정한다.

라벨링 되지 않았기 때문에 본 연구의 감성 분류 모델 평가에는 실제 부정적인 리뷰 중 모델이 부정으로 정확하게 예측한 비율인 재현율을 중점적으로 고려하여 평가한다.

3.5 대형 언어 모델을 사용한 다중 라벨링

본 연구의 마지막 단계는 감성 분류를 수행한 데이터 중 ‘부정’으로 분류된 데이터를 선택하고, 식품안전 관련 요소를 탐지하기 위한 다중 라벨링 기법을 적용하는 과정이다. 이 과정에서는 특정 기준에 따라 선별된 데이터에 대형 언어 모델인 GPT-3.5(text-davinci-003)와 GPT-4 Turbo(gpt-4-1106-preview)를 적용하여 각 리뷰에 다양한 카테고리의 라벨을 할당한다. 다중 라벨링 카테고리는 ‘식중독’, ‘변질’, ‘화학적이취’, ‘이물질’로 구성되어 있다. <표 8>를 통해 각 리뷰의 다중 라벨링 예시를 확인할 수 있으며, 각 리뷰가 어떻게 다양한 카테고리에 라벨

<표 8> 대형 언어 모델을 활용한 다중 라벨링 예시

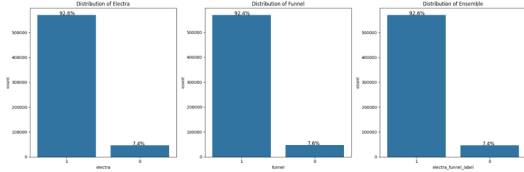
Review	식중독	변질	화학적이취	이물질
포장 신선도 다 좋아요 그런데 언어 맞아요? 정말 이리 맛있는 언어는 처음 TT TT회에 어찌 와서비도 안 넣었는지 TTT 다시는 구매 안함	0	0	0	0
숙성된 언어의 맛 괜찮아요 좋아요	0	0	0	0
녹지는 않았는데 스티로폼 박스를 여니까 고기냄새가 나네요	0	1	0	0
3개나 터져서 왔어요 터진줄 모르고 현관에 방치했다가 청국장 냄새가 나서 뜯어 보니 3개나 터져서 썩고 있었네요	0	1	0	0

링 되었는지 확인할 수 있다. 이러한 다중 라벨링은 각 리뷰의 내용을 더욱 깊이 있게 분석하고, 식품안전 위해 요소를 포괄적으로 이해하는데 중요한 기여를 한다. 다중 라벨링된 데이터들은 다양한 평가메트릭을 통해 성능을 검증했다.

IV. 모델 성능 평가

4.1 감성 분류 모델 성능 평가

본 연구에서는 부정 리뷰를 잘 분류할 목적으로 세 가지 감성 분류 모델(funnel, ko-electra, 앙상블)의 성능을 비교 분석하였다. 모델의 성능은 정확도, F1 Score, 정밀도, 재현율, AUC, 그리고 Kappa와 같은 지표를 사용하여 평가하였다. 각 모델마다 세 번의 실험을 거쳐 평균 성능을 산출하였다.



<그림 4> 감성 분류 모델 별 라벨링 결과

대규모 데이터셋의 수동 검토는 비효율적이므로, 모델의 실제 성능을 정확하게 반영하기 위해 적절한 샘플링 전략이 필수적이다. 본 연구에서는 부정 리뷰 비율을 높은 클래스별 샘플링 방법을 사용하여 모델의 부정적 리뷰 분류능력을 평가하였다. 총 300개의 리뷰를 랜덤 추출하여 구성된 데이터셋은 60%가 부정 리뷰, 40%는 긍정 리뷰로 구성된다. 데이터가 사전에 라벨링 되지 않았기 때문에 정확도는 적절한 평가 지표가 아니므로, 정밀도, 재현율, F1 Score와 같은 지표를 중점적으로 고려했다. 이와 같은 샘플링 전략을 통해 얻은 결과는 <표 9>에 정리되어 있으며, 각 모델의 장단점을 보다 명확하게 파악할 수 있었다.

<표 9>에 따르면 전체 평균 F1 Score는 0.76으로, 모델이 평균적으로 균형 잡힌 성능을 제공한다고 볼 수 있다. 전체 평균 정밀도는 0.86으로, 모델이 긍정적 결과를 예측할 때 높은 신

<표 9> 테스트 데이터의 감성 분류 모델 성능평가

		정확도	F1 Score	정밀도	재현율	AUC	Kappa
funnel	1	0.735	0.7665	0.8447	0.7016	0.7455	0.4662
	2	0.73	0.7632	0.8878	0.6692	0.756	0.4632
	3	0.73	0.7568	0.8571	0.6774	0.7466	0.4626
평균		0.73	0.76	0.86	0.68	0.75	0.46
ko-electra	1	0.705	0.7401	0.8155	0.6774	0.7137	0.4057
	2	0.75	0.7845	0.8922	0.7	0.7714	0.497
	3	0.72	0.7477	0.8469	0.6694	0.736	0.4427
평균		0.73	0.76	0.85	0.68	0.74	0.45
funnel + ko-electra	1	0.725	0.7511	0.8557	0.6694	0.7426	0.4539
	2	0.745	0.7713	0.9247	0.6615	0.7808	0.5005
	3	0.72	0.7431	0.8617	0.6532	0.7411	0.4479
평균		0.73	0.76	0.88	0.66	0.75	0.47
전체평균		0.73	0.76	0.86	0.67	0.75	0.46

뢰도를 보인다. 전체 평균 AUC는 0.75로, 모델이 양호한 분류 성능을 가지고 있음을 의미한다. 전체 평균 재현율이 0.67로 보고된 것을 고려할 때, 모델이 부정 리뷰를 식별하는 능력은 긍정 리뷰를 식별하는 것보다 약간 낮은 수준임을 알 수 있다.

결과적으로, 재현율과 F1 Score의 균형을 고려할 때, 'funnel' 모델은 부정 리뷰 탐지에 가장 적합한 모델로 나타났다. 이 모델은 높은 정밀도와 더불어 상대적으로 높은 재현율을 제공하며, 부정 리뷰 탐지의 정확성과 신뢰성을 동시에 제공한다.

4.2 식품위해 다중 라벨링 모델의 성능평가

본 연구에서는 2023년 6월부터 10월까지 수집된 총 599,321건의 식품 리뷰 데이터 중 부정적으로 분류된 23,063건을 대형 언어 모델을 활용하여 다중 라벨링을 수행하였다. 이러한 다중 라벨링의 결과는 <표 10>에서 각 카테고리별 빈도수를 통해 확인할 수 있다.

이를 통해 대형 언어 모델을 활용한 다중 라벨링 방식의 성능을 종합적으로 검증하고, 식품 안전 관련 위해 요소를 식별하는 데 있어 모델의 유효성과 한계를 명확히 파악할 수 있었다.

특히, 다중 라벨링을 통해 식중독 카테고리에 도출된 데이터 중 유당불내증 또는 유산균으로 인한 증상이 식중독과 유사하게 나타나는 경우가 관찰되었다. 이러한 혼동을 방지하고 연구의 정확성을 높이기 위해, 유당불내증이나 유산균 관련 증상을 유발할 수 있는 특정 식품 카테고리(우유/요구르트, 영양제, 단백질 보충제)는

<표 10> 다중 라벨링 카테고리 별 빈도 수

LLM 모델	식중독	변질	화학적 이취	이물질
GPT-3.5	831	5,618	20	1,140
GPT-4 Turbo	271	1,699	169	1,536

식중독 카테고리의 성능평가에서 제외하였다. 이 조치는 데이터의 오류를 최소화하고, 식중독과 유사한 증상을 유발할 수 있는 다른 원인의 영향을 감소시키는 데 중점을 두었다.

<표 11>에서는 GPT-3.5(text-davinci-003) 모델의 성능평가 결과를 확인할 수 있다.

<표 11> 다중 라벨링 성능 평가 : GPT-3.5

	검증 데이터 수	정확도	F1 Score	정밀도	재현율	AUC	Kappa	FNR
식중독	813	0.98	0.0	0.0	0.0	0.5	-0.01	1.0
변질	1,409	0.67	0.08	0.42	0.04	0.51	0.02	0.58
화학적 이취	1,409	0.96	0.06	0.04	0.22	0.59	0.05	0.96
이물질	1,409	0.88	0.11	0.1	0.12	0.53	0.05	0.90
평균	1,260	0.87	0.25	0.14	0.10	0.53	0.03	0.86

모델의 정확도는 대체로 높게 나왔으나, 다른 지표들은 낮은 값을 보인다. 특히 식중독의 False Negative Rate(FNR)가 100%로 모든 실제 식중독 사례를 놓쳤음을 의미한다. 이는 모델이 유당불내증이나 유산균 부작용과 같이 식중독 증상과 유사한 증상을 보이는 상황을 식중독으로 잘못 분류했기 때문이다.

<표 12> GPT-3.5 식중독 카테고리 오분류 예시

제품명	카테고리	리뷰
A	영양제	가루라서 별하나 뺐어요 남푼이 두드러기가 매일 나서 혹시나 하고 시켜봤어여 한달 뒤에 효과좀 있으면 리뷰 다시쓸게요
B	단백질 보충제	여기 제품만 2번째 먹는 건데 여기꺼먹을때마다 설사함 그래서 다른 제품도 먹어보고 식단도 다 해보고 여러가지해봤는데 여기 프로틴 먹을 때만 설사함 두번 다신 안시켜먹을꺼임 웬만한 헬창들은 알꺼임 삼대오백 마케팅빨인거
C	우유/요구르트	많이 마시면 설사. 대책없음. 중단. 감사합니다.

해당 리뷰는 주로 우유, 요구르트, 영양제, 단백질 보충제 카테고리에서 발견되었으며, 식중독 카테고리의 성능 평가에서 이러한 사례들을 제외했기 때문에 이 문제가 발생한 것으로 추정된다. 해당 카테고리에서 제외된 리뷰 예시는 <표 12>에서 확인할 수 있다.

<표 13> 다중 라벨링 성능 평가 : GPT-4 Turbo

	검증 데이터 수	정확도	F1 Score	정밀도	재현율	AUC	Kappa	FNR
식중독	813	0.99	0.87	0.77	1.0	1.0	0.87	0.23
변질	1,409	0.87	0.67	0.81	0.57	0.76	0.59	0.19
화학적 이취	1,409	0.96	0.68	0.57	0.86	0.91	0.66	0.43
이물질	1,409	0.94	0.64	0.48	0.96	0.95	0.61	0.52
평균	1,260	0.94	0.72	0.66	0.85	0.91	0.68	0.34

<표 13>에서는 GPT-4 Turbo(gpt-4-1106-preview)모델의 성능평가 결과를 확인할 수 있다. 식중독 카테고리에서는 정확도가 0.99와 F1 Score 0.87이 관찰되었다. 이는 모델이 식중독 위험 요소를 효과적으로 식별하고 있음을 의미한다. 정밀도는 0.77이며 재현율은 1.0으로 모든 실제 식중독 사례를 정확히 식별하였으며, AUC 값은 1.0, Kappa 값은 0.87로 일관된 판단력을 보여준다.

변질 카테고리에서는 정확도 0.87, F1 Score 0.67, 정밀도 0.81, 재현율 0.57로 관찰되었다. 이는 일부 변질 사례가 누락될 가능성이 있음을 시사한다.

화학적 이취 카테고리에서는 정확도 0.96, F1 Score 0.68을 기록했으며, 낮은 정밀도 0.57과 높은 재현율 0.86으로 많은 거짓 긍정(화학적 이취가 없는 리뷰를 화학적 이취가 있다고 잘못 분류하는 경우)이 발생했음을 나타내며, AUC는 0.91, Kappa는 0.66으로 비교적 높은 성능을 보였다.

마지막으로, 이물질 카테고리에서는 정확도

0.94, F1 Score 0.64, 정밀도 0.48, 재현율 0.96으로, 높은 재현율에도 불구하고 많은 거짓 긍정(실제로 이물질이 없는 리뷰를 이물질이 있다고 잘못 판단하는 경우)의 발생을 의미하며, AUC 0.95와 Kappa 0.61로 상당히 좋은 결과를 나타냈다.

이 결과들은 모델의 다양한 분류 성능을 반영하며, 특히 식중독과 이물질 분류에서 높은 재현율을 보이거나, 일부 카테고리에서 낮은 정밀도와 거짓 긍정의 문제가 확인되어, 알고리즘 조정이나 추가 데이터 수집을 통해 개선이 필요함을 나타낸다.

V. 결론

본 연구에서는 한국어 식품 리뷰를 분석을 통해 식품안전 위해 카테고리를 탐지하는 새로운 접근법을 제시하였다. 이를 위한 방법으로 감성 분류 모델과 대형 언어 모델을 사용한 다중 라벨링 방식을 적용하였다. 사용된 감성 분류 모델과 대형 언어 모델의 다중 라벨링 방식은 특히 유의미한 결과를 도출하였다. 감성 분류 모델에서는 ‘funnel’ 모델이 낮은 False Positive(FP) 비율을 보여주며 부정 리뷰의 오분류 가능성을 최소화하는 데 기여했다. 또한, 다중 라벨링 모델은 GPT-4 Turbo를 활용한 모델이 GPT-3.5를 활용한 것보다 좋은 성능을 보였다. 정확도, 정밀도, 재현율 등 모든 성능지표가 향상되었고, 특히, 본 연구의 목표를 달성하는데 주요 지표인 재현율이 대부분의 식품 위해 카테고리에서 평균 85%로 크게 향상되었다.

이 연구는 대형 언어 모델 기반의 생성형 AI 기술을 활용하여 식품안전 위협을 탐지하는 중요한 시도로써 식품 산업에서 소비자 리뷰 데이터의 활용 가능성을 보여준다. 한국어 리뷰 데이터의 분석을 통해 우리나라에 특화된 식품안전 모니터링 시스템 기반을 제공할 수 있다. 한국어의 복잡한 문법과 표현을 효과적으로 처리

할 수 있는 모델을 개발함으로써 분석의 정확성을 높였다. 그리고, 한국어 데이터 분석에 필요한 라벨링을 빠르고 정확하게 할 수 있는 기술로써 비용을 절감할 수 있다.

정부 기관, 식품 제조업체나 유통업체는 제안된 모델을 사용하여 소비자 리뷰를 실시간으로 모니터링하고, 잠재적인 식품안전 문제를 조기에 탐지함으로써 위험을 관리할 수 있다. 이와 같은 시스템은 기업의 브랜드 평판을 보호하고, 소비자 보호를 강화하며, 궁극적으로는 소비자의 건강과 안전을 증진시키는 결과를 가져올 수 있다.

또한, 제안된 모델은 안전 관련 분야에서 광범위하게 활용될 수 있다. 예를 들어, 제조 현장의 안전 리뷰, 환경 관련 리포트 그리고 공공 안전을 위한 감시 데이터 등에 모델을 적용하여, 위험 요소를 효과적으로 식별하고 관리할 수 있을 것으로 기대된다. 이는 식품안전 뿐만 아니라 다른 안전 관련 분야에서도 중요한 발전을 이룰 수 있는 기반이 될 것이다.

본 연구의 한계점은 사용된 언어 모델의 범위에 있다. 현재 연구에서는 ChatGPT 계열의 모델만을 활용하였으나, 최근 한국어 처리에 특화된 것으로 알려진 HyperCLOVA와 같은 국산 대형 언어 모델의 성능을 비교 분석하지 않았다. 향후 연구에서는 다양한 국산 모델을 포함하여 성능 비교를 진행함으로써, 한국어 데이터의 미묘한 표현을 더 정밀하게 처리할 수 있는 모델을 평가할 계획이다. 이러한 접근은 모델 선택의 다양성을 높이고, 식품안전 위협 탐지의 정확도를 개선하는 데 기여할 것으로 기대된다.

참 고 문 헌

- [1] 식품안전정보원. (2022). 부정·불량식품 소비자신고 동향보고서.
- [2] 이상아, 장한술, 백연미, 박수지, & 신효필.

- (2020). A small-scale Korean-specific BERT language model. *Journal of KIISE*, 47(7), 682-692. <https://doi.org/10.5626/JOK.2020.47.7.682>
- [3] 안정국 & 김희웅. (2015). 집단지성을 이용한 한글 감성어 사전 구축. *지능정보연구*, 21(2), 49-67. <https://doi.org/10.13088/jiis.2015.21.2.49>
- [4] 소진수 & 신판섭. (2020). 음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측. *한국컴퓨터정보학회논문지*, 25(6), 81-89.
- [5] 최혜선 & 연구필. (2022). A case study on text analysis using meal kit product review data. *한국콘텐츠학회논문지*, 22(5), 1-14. <https://doi.org/10.5392/JKCA.2022.22.05.001>
- [6] 최준영 & 임희석. (2020). 자연어처리 모델을 이용한 이커머스 데이터 기반 감성 분석 모델 구축. *한국융합학회논문지*, 11(11), 33-39. <https://doi.org/10.15207/JKCS.2020.11.11.033>
- [7] 정재운, 모경현, 서승완, 김창엽, 김해동, & 장필성. (2018). 워드 임베딩과 단어 네트워크 분석을 활용한 비지도학습 기반의 문서 다중범주 가중치 산출. *대한산업공학회지*, 44(6), 442-451. <https://doi.org/10.7232/JKIE.2018.44.6.442>
- [8] Shin Bee, Ryu Sohee, Kim Yongjun, & Kim Dongwhan.(2022). Analysis on review data of restaurants in Google Maps through text mining: Focusing on sentiment analysis. *Journal of Multimedia Information System*, 9(1), 61-68. <https://doi.org/10.33851/JMIS.2022.9.1.61>
- [9] M. I. Hossain Junaid, F. Hossain, U. S. Upal, A. Tameem, A. Kashim and A. Fahmin, “Bangla Food Review Sentimental Analysis using Machine Learning.” 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2022, pp. 0347-0353, doi: 10.1109/CCWC54503.2022.9720761.
- [10] N. Islam, N. Akter and A. Sattar, “Sentiment Analysis on Food Review using Machine Learning Approach,” 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 157-164, doi: 10.1109/ICAIS50930.2021.9395874.
- [11] 정민경, 권지윤, 이재우, 이유나, & 이새봄. (2021). 텍스트마이닝을 활용한 식품유통 플랫폼에 대한 소비자 인식 분석 - 토픽모델링 기법을 중심으로. *외식경영연구*, 24, 71-100. <https://doi.org/10.47584/jfm.2021.24.7.71>
- [12] 백세희, 김세형, 배미란, & 강주영. (2022). 제주시의 지속가능한 여행 활성화를 위한 지역화폐 ‘탐나는전’ 가맹점의 리뷰 데이터 분석. *한국빅데이터학회 학회지*, 7(2), 113-128. <https://doi.org/10.36498/KBIGDT.2022.7.2.113>
- [13] Tang, T., Tang, X., & Yuan, T. (2020). Fine-tuning BERT for multi-label sentiment analysis in unbalanced code-switching text. *IEEE Access*, 8, 193248-193256.
- [14] 노동훈, 민재옥, & 우소연. (2024). 특허상담 자동분류의 성능 향상 방안 연구: 트랜스포머 기반 인공지능 모델 버트(BERT)를 활용. *지식재산연구*, 19(1), 159-177. <https://doi.org/10.34122/jip.2024.19.1.7>
- [15] 이원민, & 온병원. (2021). 감성 및 감정 단어 마스킹 기반 BERT와 GPT 파이프라인 방식을 통한 감정 문장 생성. *한국정보기술학회논문지*, 19(9), 29-40. <https://doi.org/10.14801/jkiit.2021.19.9.29>
- [16] Peskine, Y., Korenčić, D., Grubisic, I., Papotti, P., Troncy, R., & Rosso, P. (2023). Definitions matter: Guiding GPT for multi-label classification. *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 4054-4063. <https://doi.org/10.18653/v1/2023.findings-emnlp.267>

- [17] Yang, S., Shang, Z., Wang, Y., Deng, D., Chen, H., Cheng, Q., & Wu, X. (2024). Data-free multi-label image recognition via LLM-powered prompt tuning. arXiv preprint arXiv:2403.01209.
- [18] 서혜진, & 신정아. (2020). 딥러닝을 활용한 감정 분석 과정에서 필요한 데이터 전처리 및 형태 변형. 영어학, 20, 42-63.
- [19] <https://github.com/kiyoungkim1/LMkor/blob/main/README.md>
- [20] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805v2. <https://doi.org/10.48550/arXiv.1810.04805>



조완섭(Wan-Sup Cho)

- 1987년 : 한국과학기술원 컴퓨터과학학과 (박사)
- 1996년~현재 : 충북대학교 경영정보학과(교수)
- <관심분야> 빅데이터, 비즈니스 인텔리전스, 머신러닝

저자 소개



최은선(Eun-Seon Choi)

- 2021.03-충북대학교 대학원
- 빅데이터협동과정 박사과정
- <관심분야> 빅데이터, 머신러닝



이경희(Kyung-Hee Lee)

- 2022.01-충북대학교 경영정보학과 초빙교수
- 2020.03-2021.12 ㈜힐링소프트 데이터사업부 이사
- 2016.04-2020.02 충북대학교 빅데이터학과 초빙교수
- <관심분야> 빅데이터, 데이터분석, 알고리즘