

딥러닝 이론과 실습 연구 계획서

제목: *Text & Image* 기반 패션 리뷰 분석 자동화

모델 개발

2023 년 10 월 11 일

연세대학교

정보대학원 비즈니스 빅데이터 분석

프로젝트 팀

팀장: 황성아 (2023522119)

박세연 (2023522114)

목 차

1. 연구의 필요성 및 배경	-----
2. 연구의 목표 및 내용	-----
3. 연구 방법	-----
4. 연구 데이터	-----
5. 연구 결과의 중요성 및 기대효과	-----
6. 연구 진행 개요 (WBS)	-----
7. 참고 문헌	-----

1. 연구의 필요성 및 배경

■ 연구의 필요성

패션 산업은 쇼핑물 시장의 중요한 부분으로, 이 분야는 사용자들의 열정적인 참여와 피드백에 의존하고 있다. 특히 코로나 19 이후 쇼핑물 시장의 경쟁이 성장함과 동시에 치열해지고 있으며, 이에 따라 패션 쇼핑물 회사들은 사용자들의 피드백을 빠르게 반영하여 서비스를 개선하는 것이 중요하다.

패션 쇼핑물 플랫폼 시장에서 큰 점유율은 가진 '무신사'에서는 수많은 사용자 리뷰가 생성되며, 이는 제품 개선과 소비자의 의사결정에 중요하다. 사용자 리뷰는 제품에 대한 직접적인 사용자 경험과 반응을 담고 있어, 제품의 장단점을 파악하고 의사결정의 방향성을 찾을 수 있다. 패션 분야에서는 사용자 리뷰가 색상, 사이즈, 매칭 가능한 스타일 등 실제로 착용했을 시의 다양한 요소에 대한 평가를 포함하므로 이를 분석하는 것은 유용하다.

그러나 현재까지는 많은 패션 쇼핑물 회사들이 각자 독립적으로 데이터를 처리하고 분석하는 데에 한계가 있어 구체적인 제품의 상태나 스타일을 파악하기 어렵다. '무신사' 플랫폼에서 제공되는 리뷰와 구매 만족도 점수 시스템만으로는 많은 리뷰를 파악하는 데의 한계점이 존재하며, 높은 점수와 다르게 텍스트에 단점이 수록된 경우가 존재한다는 점, 매칭 가능한 스타일 탐색에 어려움이 존재한다는 점에서 개선이 필요하다.

■ 연구 배경

영화 리뷰 요약 분석을 진행한 선행연구¹에서는 온라인 사용자들이 매일 수천 개의 영화 리뷰를 게시하기 때문에 리뷰를 수동적으로 요약하기는 어렵다고 판단하여 자동으로 긴 영화 리뷰를 요약해주는 연구를 진행하였다. 영화 리뷰에서 특징을 추출하고 이를 벡터 공간 모델 또는 특징 벡터로 표현한 후 나이브 베이즈 머신러닝 알고리즘을 사용하여 긍정과 부정으로 리뷰를 분류했다. 그런 후 가중 그래프 기반 알고리즘을 적용하여 각 리뷰 문장에 대한 순위 점수를 계산해 높은 순위 점수를 기준으로 선택하여 추출 요약을 진행했다. 본 연구에서는 딥러닝 사전학습 모델들을 활용하고 모델에서 새로운 텍스트를 생성해내야 하기에 말이 되지 않는 표현이 만들어질 가능성이 존재하나 좀 더 유연한 접근이 가능한 생성 요약을 진행하여 딥러닝 기반 생성 요약의 우수성을 입증하고자 한다.

¹ Khan, A., Gul, M.A., Zareei, M., Rajesh, R.B., Zeb, A., Naeem, M., Saeed, Y., & Salim, N. (2020). Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm. *Computational Intelligence and Neuroscience*, 2020.

음식점 리뷰 감정 분석을 진행한 선행 연구²에서는 한국어로 작성된 음식점 리뷰를 대상으로, 감성분석을 수행하여 평가 항목별로 세분화된 평점을 제공 가능한 예측 방법론을 제안했다. 이를 위해, 음식점의 주요 평가항목으로 '음식', '가격', '서비스', '분위기'를 선정하고, 평가항목별 맞춤형 감성사전을 구축했다. 또한 평가항목별 리뷰 문장을 분류하고 감성분석을 통해 세분화된 평점을 예측하여 소비자가 의사결정에 활용 가능한 추가적인 정보를 제공했다. 이러한 선행연구를 바탕으로, 본 연구에서는 패션 도메인에 적용시켜 발전시키고자 한다.

이외에도 패션 스타일을 군집화하는 방법을 제시한 선행 연구³에서는 패션 이미지 데이터셋에 대해 centroid 기반 밀도 중심 클러스터링 알고리즘을 통하여 패션 스타일을 군집화해 표현한다. 스타일 종류로는 Preppy, Goth, Hipster, Bohemian 등이 있으며, Preppy x Goth 와 같이 두 가지 스타일을 믹스한 스타일로도 표현한다는 점에 있어 강점을 가진다. 패션 이미지를 다양한 스타일로 군집화한다는 면에 있어서 본 연구에서 원하는 태스크와 일치하지만, 해당 모델은 패션 쇼 등의 이미지를 사용해 클러스터를 구성하였기 때문에 우리가 분석하려는 리뷰 이미지와는 거리가 있는 데이터셋이며, style 속성 이외에도 texture, fabric, part 속성을 함께 고려해 클러스터를 구성한다는 점에 있어 차이를 보인다.

패션 이미지의 색상을 분류하는 태스크에 대한 선행 연구⁴에서는 K-Means 클러스터링을 통해 이미지 내 옷의 색상을 구분한다. 이미지에서 MaskRCNN 모델을 이용해 이미지 내 옷 부분의 색상만을 추출하고, 한 이미지 내에서 옷 색상을 5 가지 등으로 세부적으로 분류할 수 있는 모델을 구축하였다. 그러나 학습과 테스트에 사용된 이미지가 상의 또는 하의 중 하나에 해당되는 옷 하나에 대한 이미지를 이용했다는 점에 있어, 전신 리뷰 이미지를 분석하려는 우리의 태스크에는 정확히 일치하지 않으며, 한 이미지 내에서 상의와 하의를 각각 하나의 색상으로 나타내야 하는 모델을 구축할 필요가 있을 것으로 보인다.

² SoJinSoo and Shin, Pan-Seop. (2020). Rating Prediction by Evaluation Item through Sentiment Analysis of Restaurant Review. Journal of The Korea Society of Computer and Information, 25(6), 81-89.

³ Chen, J., Yuan, H., Fang, F., Peng, T., & Hu, X.R. (2023). Unsupervised Fashion Style Learning by Solving Fashion Jigsaw Puzzles. 2023 IEEE International Conference on Multimedia and Expo (ICME), 1847-1852.

⁴ 장혜림, 손봉기, 허권, 이재호. K-Means 클러스터링 기반 패션 이미지 색상 분류 구현. 한국통신학회 학술대회논문집.

2. 연구의 목표 및 내용

■ 연구 목표

본 연구는 패션 플랫폼 산업 전반에 대한 리뷰 분석을 기반으로 고객에게 가장 적합한 소비자 행동을 제안하는 것이 주된 목적이다. 구체적으로 Python 언어와 Pytorch 프레임워크 등으로 딥러닝 사전 학습 모델들을 활용하여 fine-tuning 을 거쳐 다양한 리뷰 분석 딥러닝 모델을 개발한다. 이는 소비자들의 구매 실패율을 줄이고, 구매 시간 단축으로 소비자들의 만족도를 높이며 판매자들에게도 매출 상승 등 긍정적인 영향을 미쳐 패션 플랫폼 시장에서 유의미한 결과를 이끌어 낼 것으로 기대한다.

■ 내용

본 연구는 '무신사' 플랫폼에서 사용자들의 텍스트와 이미지 형태의 리뷰 데이터를 분석하는 것을 중심으로 진행한다. 이 데이터들은 사용자들의 반응을 직접적으로 보여주므로, 해당 제품에 대한 사용자들의 의견을 파악하는 데 유용하다.

수집 된 데이터를 바탕으로 네 가지 리뷰 분석 자동화 모델을 개발한다. 첫 번째, 텍스트 리뷰 데이터의 자주 언급되는 색상, 사이즈 등의 단어 키워드를 분석한 후 키워드 별 리뷰 요약을 제공한다. 두 번째, 텍스트 리뷰 데이터를 통해 해당 제품의 긍정/부정 정도를 표현한다. 세 번째, 이미지 리뷰 데이터로 사용자들의 착용 스타일을 클러스터링 한다. 마지막으로, 이미지 리뷰 데이터에서 해당 제품과 함께 매치한 옷이나 제품을 탐지하여 해당 옷이나 제품들의 색상을 RGB 값을 기반으로 분류한다. 이 네 가지 분석 모델을 종합하여 소비자들이 제품 리뷰를 간편하게 접하게 한다.

나아가, 본 연구는 실제 서비스 플랫폼 상에서 적용 가능한 웹 User Interface 예시를 제안함으로써 다른 제품이나 플랫폼으로의 확장을 위한 인사이트를 제공한다.

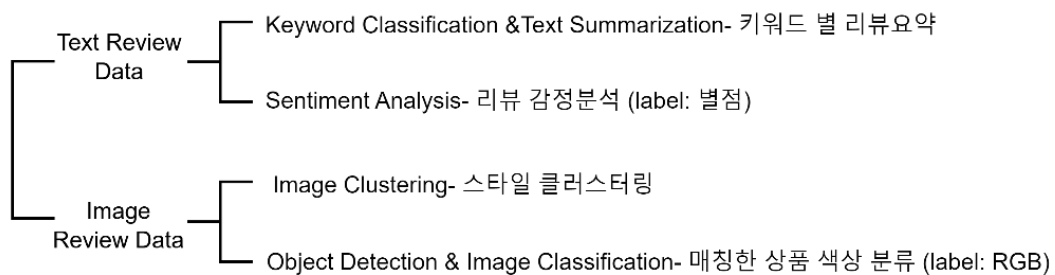
3. 연구 방법

본 연구에서 활용 되는 System Architecture 는 아래와 같다.

System Architecture	
Data	MUSINSA text & image
Database	MYSQL
Language(framework)	Python(Pytorch, Tensorflow)
Web	HTML, CSS, javascript

■ 연구 방법론

본 연구에서 진행되는 연구 방법론을 요약하면 아래와 같다.



● 텍스트 기반 리뷰 분석

- Keyword Classification

키워드 분석은 불용어 제거와 어간 추출 및 형태소 분석 등의 자연어 처리를 시행한 후 텍스트에서 많이 등장하는 형태소 등장 빈도를 분석함으로써 핵심 키워드를 추출한다. 특정 텍스트 자료에 많이 나타나는 형태소가 그 텍스트 주제를 표출할 가능성이 높다는 가정에 기초한다. 빈도 분석에서 조사와 같이 의미를 별로 담고 있지 않은 불용어는 제거한다. 키워드 분석은 텍스트 주제 추정, 텍스트 유사도, 검색 엔진의 검색 결과 우선 순위 측정 등 다양하게 사용된다. 이를 기반으로 맞춤형 키워드 사전을 제작하여 키워드 별로 텍스트 리뷰를 분류한다.

- Text Summarization

텍스트 요약은 상대적으로 큰 원문을 핵심 내용만 간추려서 작은 요약문으로 변환한다. 일반적으로 텍스트 요약(Text Summarization)은 Extractive Summarization 과 Abstractive Summarization 로 구분된다. 본 연구에서 활용할 방법은 Abstractive Summarization 이며, Extractive Summarization 보다 모델에서 새로운 텍스트를 생성해내야 하기에 말이 되지 않는 표현이 만들어질 가능성이 존재하나 좀 더 유연한 접근이 가능하다.

- **Sentiment Analysis**

감정 분석이란 텍스트에 들어있는 의견이나 감성.평가.태도 등의 주관적인 정보를 컴퓨터를 통해 분석하는 과정이다. 감정 분석은 크게 두 가지 단계로 이루어져 있다. 첫 번째 단계로는 문서의 어떤 부분에 의견이 담겨있는 지를 정의하고 그 다음 단계에서는 이를 바탕으로 요약한다. 텍스트 내의 감정을 분류하거나 긍정과 부정의 정도를 점수화 한다.

- 이미지 기반 리뷰 분석

- **Image Clustering**

이미지 클러스터링이란 이미지 데이터에 대해 질적 상황 별 특징이 비슷한 이미지끼리 군집화하는 기법이다. 특성이 유사한 데이터는 같은 군집에 해당하게 되기 때문에 클러스터링 결과를 통해 해당 군집 간의 특성을 파악할 수 있게 된다. 본 연구에서는 패션 아이템 리뷰 이미지를 이용하여 비슷한 스타일이나 분위기를 띄는 리뷰 이미지끼리 묶어 주기 위하여 이미지 클러스터링 기법을 이용한다.

- **Object Detection**

객체 인식이란 한 이미지 내에서 찾고자 하는 객체의 위치와 그 객체가 어떤 것인지를 탐지하는 것을 말한다. 본 연구에서는 패션 아이템 전신 리뷰 이미지를 매치한 상하의 색상에 따라 분류할 때, 이미지에서 입고 있는 상의와 하의를 object detection 기법을 사용해 검출한다.

- **Image Classification**

이미지 분류 기법이란 주어진 이미지 데이터에 대해 사전에 정의된 label 집합의 값 중 하나로 할당하는 방법이다. 본 연구에서는 패션 아이템 전신 리뷰 이미지를 매치한 상하의 색상에 따라 이미지를 분류하는 문제에서 크롭된 상의, 하의 이미지를 image classification 기법을 활용해 해당 상의와 하의 객체의 색상을 파악한다. 상의와 하의의 색상은 크롭된 이미지의 RGB 색상값을 이용해 하나의 색상 단어값으로 나타낸다. 예를 들어, 상의의 RGB 값이 [250, 0, 0]인 경우 상의 색상은 'Red'가 되는 형식이다.

■ 연구 프로세스

● 데이터 수집

패션 쇼핑 플랫폼 사이트 '무신사' 에서 크롤링하는 방식으로 데이터를 수집한다. 이 과정에서는 Python 언어와 BeautifulSoup 라이브러리 등을 활용하여 웹 스크래핑을 진행한다. 이 과정에서 수집할 정보는 리뷰 텍스트와 이미지 데이터, 기존 구매 만족도 점수 데이터이다.

● 데이터 탐색, 정제

수집된 텍스트 데이터는 Konlpy 의 Okt 모듈을 활용하여 표현 방법이 다른 단어들을 통합시켜서 같은 단어로 만드는 정규화, 말뭉치를 더 이상 나눌 수 없는 단위로 나누는 토큰화, 텍스트를 숫자로 처리하도록 정수 인코딩 수행, 하나의 행렬로 만들기 위해 서로 다른 길이의 샘플들의 길이를 동일하게 만드는 패딩 등의 적절한 정제 작업을 거쳐 모델 개발에 적합한 형태로 변환한다.

수집된 이미지 데이터 중 전신 및 상·하반신이 나온 이미지만을 추려낸다. 이후 초상권 문제를 최소화하기 위해 얼굴이 나와있는 이미지에 대해서는 얼굴 부분에 모자이크 처리를 진행한다.

● 모델 Fine-tuning

키워드 별 리뷰요약

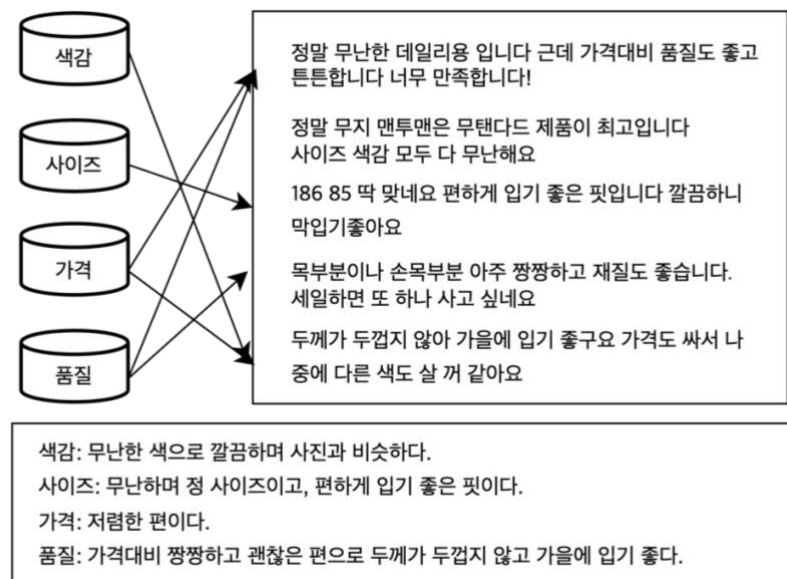


그림 1. 키워드 별 리뷰요약 결과 예시

- 색감, 사이즈, 가격, 품질 등의 키워드 별로 텍스트 리뷰를 분류하기 위하여 패션 상품을 고르는데 있어서 필요한 정보를 평가항목에 포함하여 맞춤형 키워드 사전을 제작한다. 해당 어휘가 리뷰 문장 텍스트에 출현하면, 리뷰 문장을 해당 평가항목을 평가하는 문장으로 판단하고 해당 평가 항목으로 분류한다. 각 평가항목별 감성 어휘를 기준으로 리뷰 문장 분류를 수행하면 각 평가항목을 평가하는 문장들로 리뷰 문장이 재구성된다.
- 한국어 텍스트 요약에서 좋은 성능을 자랑하는 mT5(101 개 언어를 포함하는 다국어 Common Crawl corpus(mC4)를 이용해 사전 학습된 T5 의 다국어 버전), mBART-50(50 개 언어로 사전 학습된 BART 의 다국어 버전), KoBERT(한국어 버전의 BERT 사전학습 모델), KoBART(한국어 버전의 BART 사전학습 모델), KoELECTRA(한국어 버전의 ELECTRA 사전학습 모델) Fine-tuning 하여 생성 요약을 각각 진행한다.

리뷰 감성분석



Review	Sentiment	Total Sentiment
와 기본으로 최고입니다 다른것도 사야지 굿굿 무난하니 최고입니다	긍정 95%	 $(95\% + 80\%)/2 = 87.5\%$
사기치냐 2만 9천원 짜리를 2만원에 팔아 왜 도대체 얼탱이가 없네	부정 95%	
두께도 얇고 맨투맨이라고 하기에는 쫘 그렇네요	부정 70%	 $(95\% + 70\%)/2 = 82.5\%$
목부분이나 손목부분 아주 째깍하고 재질도 좋습니다. 세일하면 또 하나 사고 싶네요	긍정 80%	

그림 2 리뷰 감성분석 결과 예시

- 한국어 감성분석에 적합하면서도 높은 성능을 자랑하는 KoBERT(한국어 버전의 BERT 사전학습 모델), KoELECTRA (한국어 버전의 ELECTRA 사전학습 모델), XLM-RoBERTa(XLM 에서 성능 향상을 위해 몇 가지 보완한 확장 버전), DistilKoBERT(KoBERT 에서 레이어 수를 줄이면서 성능 저하를 최소화 하는 사전학습 모델), Multilingual-BERT(다국어 버전의 BERT 사전학습 모델)를 사용하여 감정의 긍정, 부정 정도를 나타낸다. 라벨 값은 기존 '무신사' 사이트에서 해당 상품이 가지고 있는 텍스트 리뷰의 각 구매 만족도 점수로 한다.
- 추가적으로, '키워드 별 리뷰 요약' 부분에서 나눈 키워드 별로도 리뷰 감성분석을 제공할 예정이다.

스타일 클러스터링



그림 3. 스타일 클러스터링 결과 예시

- 리뷰 이미지의 스타일을 나누기 위해서는 image clustering 기법을 이용한다. 사용할 클러스터링 모델은 ResNet18 모델을 백본으로 하고 있는 SCAN, 밀도 기반 클러스터링 기법인 DBSCAN, single-stage ConvNet 기반 DAC 을 이용한다. 세 가지 모델을 이용해 리뷰 이미지의 스타일을 군집화해보고, 결과를 비교해 보았을 때 더 군집화가 잘 된 모델을 최종 모델로 선택해 결과를 반환한다.

매칭한 상품 색상 분류

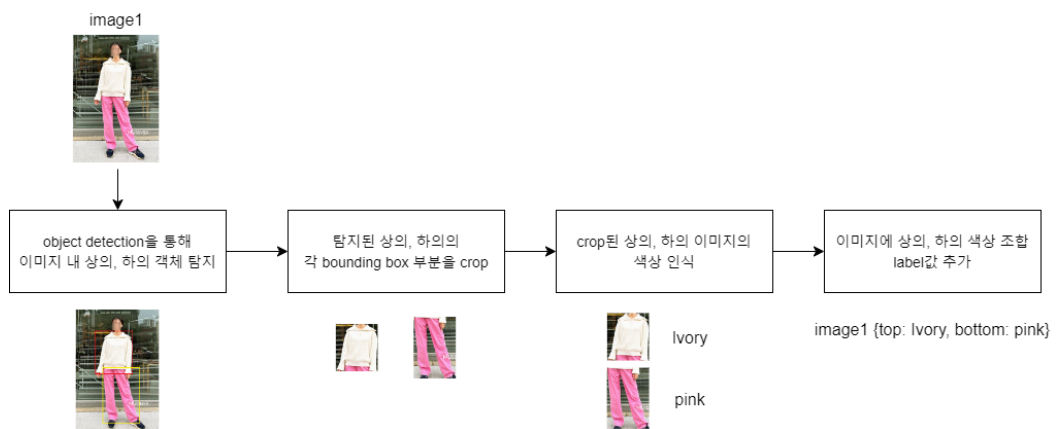


그림 4. 한 이미지 내 매칭된 상하의 색상 정보 파악 과정

- 한 이미지 내에서 상의와 하의를 탐지하기 위해서는 object detection 기법을 사용한다. 사용할 pretrained 모델은 기존의 2-stage 기반의 기법보다 빠르고 정확도가 높다는 특징을 가진 YOLOv8 이다. 해당 모델에 대한 fine-tuning 을 진행한다. 이때 label 집합은 [top, bottom]이 되며 fine-tuning 시 학습할 데이터셋은 우선 Fashion-MNIST 이미지 데이터셋을 이용해 학습한다. Fine-tuning 시 라벨값은 각각 해당되는 라벨에 맞추어 top 과 bottom 으로 바꿔준다.
- 실제 리뷰 이미지 데이터셋을 fine-tuning 한 모델에 테스트해 결과값으로 각 이미지에서 top 과 bottom 값을 검출하면, 검출한 객체의 영역에 대해 crop 을 통해 이미지를 잘라낸다.
- Crop 한 top 과 bottom 이미지는 image classification 기법을 사용해 해당 상의, 하의 색상을 구분해준다. 이때 사용할 모델은 KNN 으로 이미지의 RGB 값을 기준으로 거리를 구해 어떤 색상 label 값에 해당되는지 분류한다. 이때 색상 label 값은 [black, white, ivory, khaki, navy, gray, brown, pink, blue, green, denim, purple, yellow, red, orange] 15 가지로 구분한다.

● 모델 개발 및 검증

키워드 별 리뷰요약

전체 (713)	5점 (530)	4점 (145)	3점 (32)	2점 (4)	1점 (2)
주제전체	만족도	사이즈	디자인	무게	색상
활용도	보온성	코디	충전재	품질	가격
					라인
					착용감
					사용성
					소재
					향기
					마감처리

그림 5. 네이버 쇼핑에서 제공하는 리뷰 키워드

- 위의 사진은 네이버 쇼핑 페이지에서 제공하는 리뷰의 키워드이다. 해당 키워드를 참고하여 본 논문에서 활용할 키워드를 지정한다. 키워드 간 유사도 계산을 통해 검증한다. Max Sum Similarity 문서와 가장 유사한 단어를 2 개 사용하여 이 두개의 코사인 유사도를 구한다. 여기서 유사도가 가장 낮은 조합을 추출한다.
- 텍스트 요약에서 검증 지표로는 요약본의 일정 부분을 비교하는 지표인 ROUGE (Recall-Oriented Understudy for Gisting Evaluation)를 이용한다. ROUGE 는 성능을 평가하는 방법에 따라 ROUGE-1,-2, ROUGE-N, ROUG-L, ROUGE-S 와 같이 여러 종류의 지표로 나뉜다. 예를 들어, ROUGE-N unigram, bigram, trigram 등 문장 간

중복되는 n-gram 을 비교하는 지표이다. 본 연구에서 사용하는 태스크에 가장 적합한 ROUGE 지표를 찾아 성능을 평가한다.

리뷰 감정분석

- 감정 분석 모델 개발을 통해 각 리뷰의 긍정과 부정의 정도를 percentage 로 나타낸다. 본 논문에서 감정분석의 검증 지표로는 정확도(Accuracy)와 f1 score($2 \times (\text{recall} \times \text{precision}) / (\text{recall} + \text{precision})$) 을 사용한다.

스타일 클러스터링

- 리뷰 이미지가 얼마나 잘 분류되었는지에 대해서는 태스크에 따라 다른 평가지표를 사용한다. Image clustering 태스크 시 사용한 모델에 대해서는 accuracy, NMI, ARI, silhouette 평가지표를 이용한다.

매칭한 상품 색상 분류

- Image classification 태스크 시에는 confusion matrix 의 accuracy(정확도), precision(정밀도), recall(재현율), f1 score 을 이용해 성능을 평가한다.

● 웹 페이지 개발

HTML, CSS, Javascript 을 이용하여 '무신사' 사이트 한 상품에 대한 리뷰 페이지의 프론트 부분을 구현하여 다른 제품이나 플랫폼으로의 확장 가능성을 제공한다.

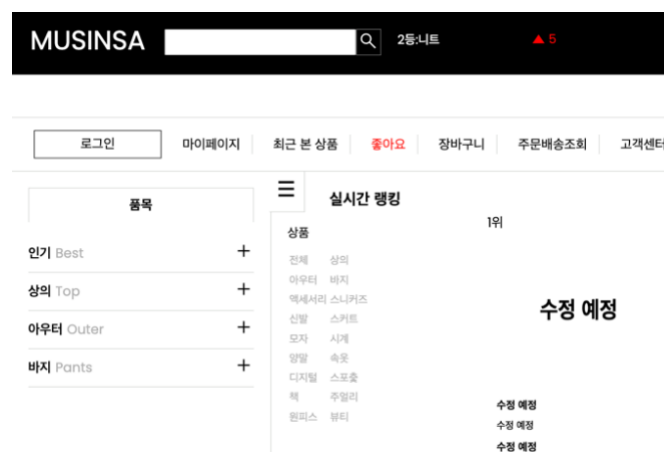


그림 6. 웹 페이지 개발 결과 예시

위 사진은 예시 상품의 메인 '무신사' 클론 페이지이다. '수정 예정' 이라고 작성한 부분을 분석 시점 활용할 상품의 정보로 수정하고 해당 상품의 리뷰 페이지를 연구 결과를 반영하여 개발한다.

4. 연구 데이터

본 연구에서는 패션에 대한 정보를 패션 쇼핑 플랫폼 '무신사' 사이트에서 텍스트와 이미지 형태로 수집한다. 데이터 수집일 기준(2023.10 월 중순) 리뷰 개수가 가장 많은 패션 상품을 선택하여 텍스트와 이미지 형태의 구매자 리뷰를 분석의 라벨(label) 값으로 활용하기 위한 기준 구매 만족도 점수 데이터를 크롤링 한다.

자료 1) 패션 분야 쇼핑물 리뷰에 대한 텍스트 데이터는 '무신사' 사이트에서 임의로 수집되었으며, 수집 예정인 리뷰 텍스트 데이터는 다음과 같이 구성된다.

이 가격에 이 재질이라니!! 언제나 믿고 있는 우텐다드입니다!
배송도 빠르고 포장 깔끔하고 핏도 오버하게 잘 떨어집니다 저번에 구매하고 두 번째 구매입니다 ^^
안에 이너티로 입었습니다! 너무 입기 좋고 평평하네요 가성비 정말 좋습니다
사이즈도 딱 좋고 원단도 좋아서 잘 입고 있습니다
사진처럼 사이즈는 적당한 오버핏이고 이미 몇번 빨았는데도 하나도 안늘어났습니다 *
안에 레이어드 해서 입기 좋은 기장이예요. 믿고 사는 우텐다드입니다!
배송도 빠르고, 가격도 너무 좋아요 일단 재질 높은 부분이 원단이 좋다는 것입니다! 가격도 천대 재질도 너무 좋아서 만족스럽네요. 길이가 고민이어서 될 줄라야 할지 몰랐는데 딱 맞게 잘 고른 것 같아요
아주 좋습니다 너무 만족하면서 애용하고 있어요
레이어드티 필요해서 샀는데 그냥저냥 잘 입을 것 같아요
레이어드용으로 화이트 블랙 1개씩 구매했는데 기장도 적당해서 좋네요
전에 입던게 누락해서 하나 더 샀는데 레이어드 하기 좋고 편해요
탄탄하고 입기 편하네요 역시 믿고 사는 우텐다드입니다!
주거적으로 구입하는 무신사 반물티예요. 가성비로 따지면 따라올 티가 없네요
행사기간에 저렴하게 구매했습니다 레이어드인줄 몰랐거나 길이면에서 제일 만족스럽네요 배송도 하루만에 왔네요 만족합니다
합리적인 가격이라 기본 원색 반팔로 매년 구매하고 있습니다. 옷감도 얇지도 않고 늘 만족하게 입고 있습니다.
소재가 탄탄하고 묵도 잘 안 늘어날 것 같아요 좋습니다! 이너로 쓰려고 샀는데 두 볼을 입으면 너무 더워서 단품으로 입습니다
옷 디자인도 이쁜데 핏이나 재질도 맘에 들어서 굿
탄탄하고 두터운 재질의 면티입니다. 두꺼워서 그런지 비침 없습니다. 여름에 단독으로 입기엔 덥기도하고 애매해서 이너로 잘 활용할 거 같습니다.
기본 이너로 입기에 가성비도 좋고 최고예요! 벌써 4벌 짝 구매했습니다
옷감이 좋구요. 속이 레이어드 하러구 샀는데 길이도 좋고 편해요
레이어드로 입으려구 크게 샀는데 오버핏으로 좋아요
레이어드 할때 길이감 너무 좋아요 한장으로 부족해서 추가 구매했습니다
레이어드하기 좋은 무난한 티입니다. 기장도 적당한 길이요.
사이즈가 루즈핏해서 여유가 있어요 여름에 얹지 않고 시원해서 다시 구매하고 싶어요! 여름에 추천하는 티입니다
정석적인 기장과 핏의 레이어드 티셔츠입니다. 레이어드 스타일링 하실거라면 한개쯤 있으면 요긴하게 쓸 수 있어요

그림 7. 텍스트 데이터 예시

자료 2) 패션 분야 쇼핑물 리뷰에 대한 이미지 데이터는 '무신사' 사이트에서 임의로 수집되었으며, 수집 예정인 리뷰 이미지 데이터 샘플은 다음과 같이 구성된다.

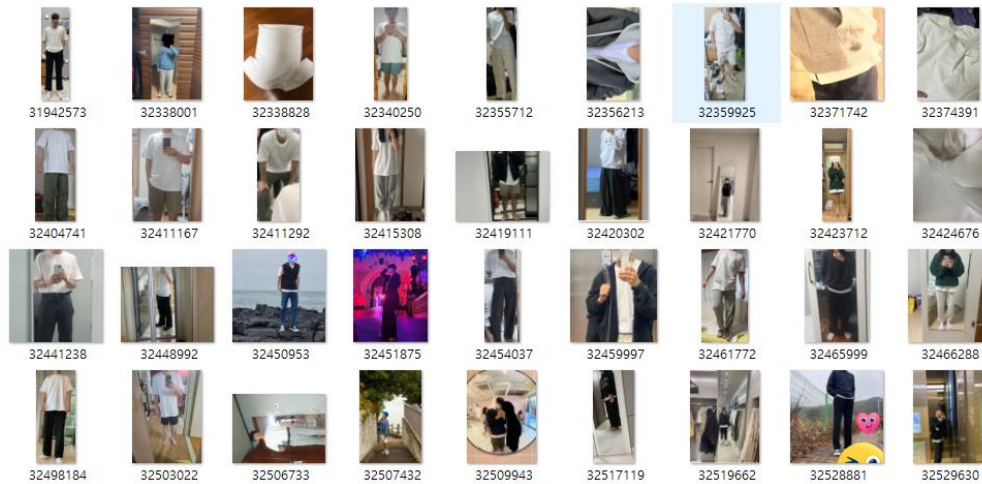


그림 8. 이미지 데이터 예시

상기한 두 자료를 활용해 패션 리뷰 분석 자동화 모델을 개발한다.

5. 연구 결과의 중요성 및 기대효과

■ 연구결과 중요성 및 기대효과

본 연구는 사용자 리뷰 활용 강화에 많은 기여를 할 것으로 기대된다. 본 연구의 분석 방법 및 개발된 모델은 해당 상품 뿐만 아니라 전체 상품, 그리고 '무신사' 플랫폼 뿐 아니라 다른 플랫폼으로의 확장성을 가진다. 따라서 실무자가 본 연구의 결과를 바탕으로 효율적인 서비스를 제공할 수 있을 것으로 기대된다.

소비자 입장에서는 단순히 나열된 리뷰들과 구매 만족도 점수만으로 구매 의사결정을 해야 한다는 한계점을 극복하여 상품 구매 결정에 도움을 줄 것이다. 또한, 판매자 입장에서는 소비자들의 구매 만족도와 활용 사례를 간편하면서도 구체적으로 파악할 수 있게 됨에 따라 소비자의 요구를 더욱 세밀하게 반영하는 상품 개발 전략을 세울 수 있게 된다. 이러한 과정은 소비자의 구매 만족도 상승, 판매자의 매출 상승과 패션 플랫폼 시장 전반에 긍정적인 영향을 미칠 것이라 기대된다.

6. 연구 진행 개요(WBS)

내용 \ 주차	1	2	3	4	5	6	7	8	9	10
주제 선정 및 데이터 수집										
데이터 탐색, 전처리										
모델 탐색										
Fine-tuning										
중간 보고서										
모델 개발 및 검증										
웹 페이지 개발										
최종 결과 도출										
최종 발표										

7. 참고 문헌

Sizov, G. (2010). Extraction-Based Automatic Summarization: Theoretical and Empirical Investigation of Summarization Techniques.

Khan, A., Gul, M.A., Zareei, M., Rajesh, R.B., Zeb, A., Naeem, M., Saeed, Y., & Salim, N. (2020). Movie Review Summarization Using Supervised Learning and Graph-Based Ranking Algorithm. *Computational Intelligence and Neuroscience*, 2020.

Liu, Y., & Lapata, M. (2019). Text Summarization with Pretrained Encoders. *ArXiv*, abs/1908.08345.

Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., & Huang, X. (2020). Extractive Summarization as Text Matching. *Annual Meeting of the Association for Computational Linguistics*.

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *ArXiv*, abs/1708.07747.

Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., & Huang, X. (2020). Extractive Summarization as Text Matching. *Annual Meeting of the Association for Computational Linguistics*.

Chung, J. (2023). Proposal of Customer Experience-based Service Innovation Framework through Case Analysis of Domestic Unicorn Service Companies. *Korea Institute of Design Research Society*.

Lee, D., Jo, J.-C., & Lim, H.-S. (2017). User sentiment analysis on Amazon fashion product review using word embedding. *Journal of the Korea Convergence Society*, 8(4), 1–8. doi:10.15207/jkcs.2017.8.4.001

소진수 and 신판섭. (2020). 음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측. *한국컴퓨터정보학회논문지*, 25(6), 81-89.

[bert] keybert 로 리뷰 키워드 추출하기. (n.d.). Retrieved from <https://velog.io/@mare-solis/BERT-keyBert%EB%A1%9C-%ED%82%A4%EC%9B%8C%EB%93%9C-%EC%B6%94%EC%B6%9C%ED%95%98%EA%B8%B0>

Text summarization. (n.d.). Retrieved from <https://www.sciencedirect.com/topics/computer-science/text-summarization>

Supersimples. (n.d.). Retrieved from <https://github.com/supersimples/musinsaclone>

(N.d.-a). Retrieved from <https://m.ddaily.co.kr/page/view/2023042609504238764>

Grootendorst, M. P. (n.d.). Retrieved from <https://maartengr.github.io/KeyBERT/index.html>

Lee, D., Jo, J.-C., & Lim, H.-S. (2017). User sentiment analysis on Amazon fashion product review using word embedding. *Journal of the Korea Convergence Society*, 8(4), 1–8. doi:10.15207/jkcs.2017.8.4.001

Wankhade, M., Rao, A.C., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55, 5731 - 5780.

Wouter, V. G. (2020, May 25). SCAN: Learning to Classify Images without Labels. arXiv.org. <https://arxiv.org/abs/2005.12320>

Ester, M. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. <https://www.semanticscholar.org/paper/A-Density-Based-Algorithm-for-Discovering-Clusters-Ester-Kriegel/5c8fe9a0412a078e30eb7e5eeb0068655b673e86>

Deep Adaptive Image Clustering. (2017, October 1). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/8237888>

<https://github.com/zalandoresearch/fashion-mnist>