

머신러닝 프로젝트

MUSINSA

소비자 후기를 다각도로 분석하여
서비스와 제품의 개선점을 찾아보자!

머신러닝-주말오전-2조

이진주/윤의정/채유진/최선아/최승혁



목차

1. 주제 선정 배경
2. 프로젝트 주제
3. 필요 자료 및 데이터 수집 방법
4. 감성분석이란
5. 진행 계획

1. 주제 선정 배경

후기를 참고하여 구매 의사를 결정하는 경우가 많다.

이에 따라 **제품 후기의 내용을 감성분석**하여

제품 누적 판매량과의 상관관계를 알아보고자 한다

2. 프로젝트 주제

무신사의 최근 1년간의 후기를 감성분석하고,
누적판매량과의 상관관계를 연구하여
제품 판매량을 개선할 수 있는 요소를 알아낸다.

*최근 1년이란, 데이터를 수집하는 일자를 기준으로 1년
대략 2022.12월 2주차 - 2023.12월 2주차 사이의 일자로 결정할 예정

유용한 순

레이어드 크루 넥 반팔 티셔츠_일반 기장

D

LV.3 뚜왓뚜지

2023.01.18

남성 · 178cm · 90kg · 신고



레이어드 크루 넥 반팔 티셔츠_일반 기장 [화이트]
S 구매

★★★★★

사이즈 보통이에요

밝기 보통이에요

색감 보통이에요

두께감 보통이에요

이너로 입기 너무 좋고 여름이 벌써부터 기대됩니다~ 레이어드로 짱!



👍 도움돼요 223

😊 스타일 좋아요 526

댓글 1개

각 제품별 리뷰 내 긍/부정 언어를 추출하여 감성분석 진행



LV.6 세오니ㅎ

남성 · 177cm · 76kg · 신고

2022.03.17



레이어드 크루 넥 반팔 티셔츠_일반 기장 [화이트]

M 구매



사이즈 보통이에요

밝기 보통이에요

색감 보통이에요

두께감 보통이에요

177/ 75 입니다

맨투맨이나 니트에 레이어드 하실거면 m

가디건이나 바람막이 아우트에 레이어드 하실거면 L 사세요

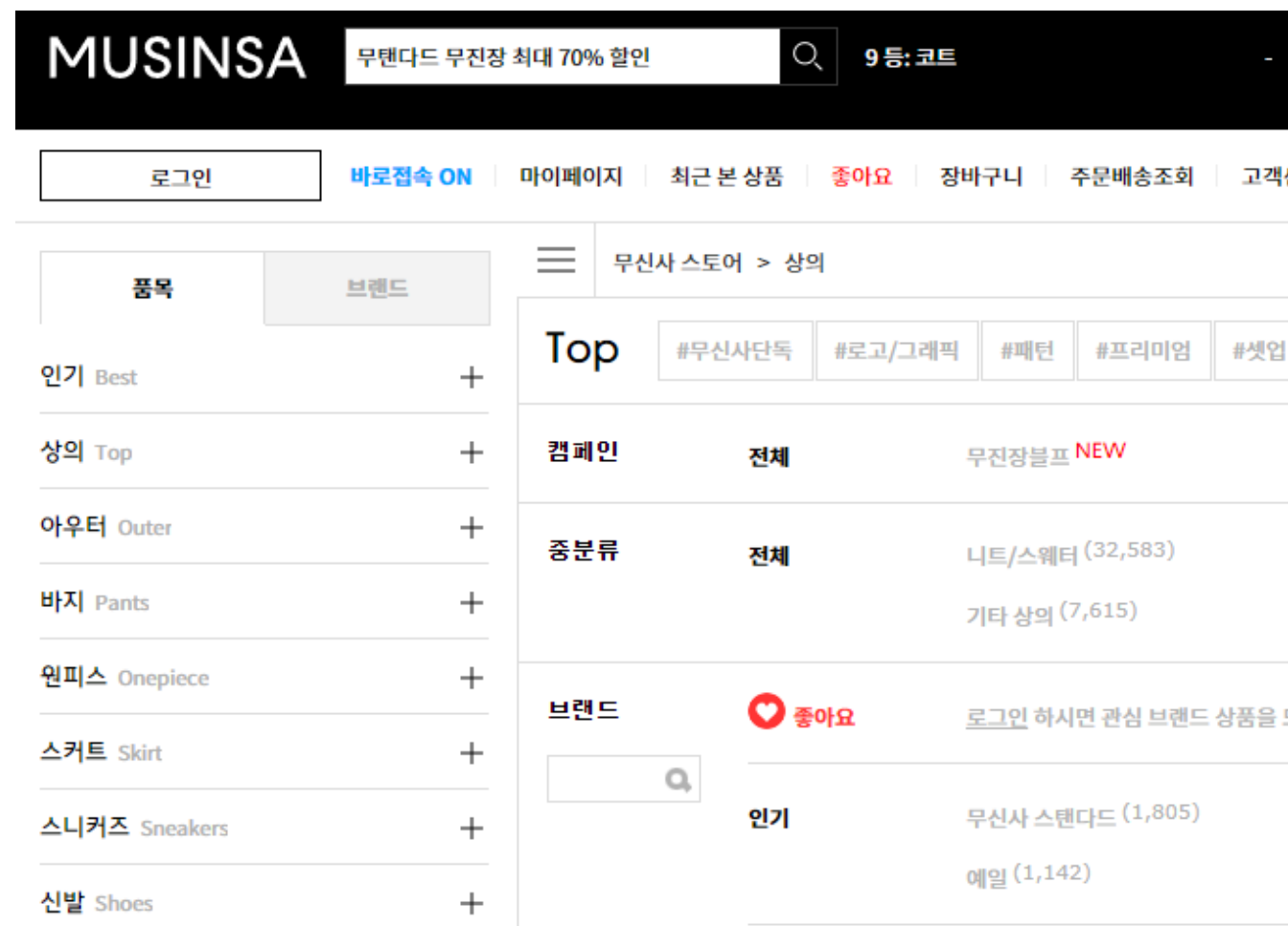


👍 도움돼요 93

😊 스타일 좋아요 18

사이즈 등 상세 정보와 관련한 후기를
단순 카운트 및 감정 분석을 진행하여
누적 판매량과 비교 후 상관관계를 찾아본다.

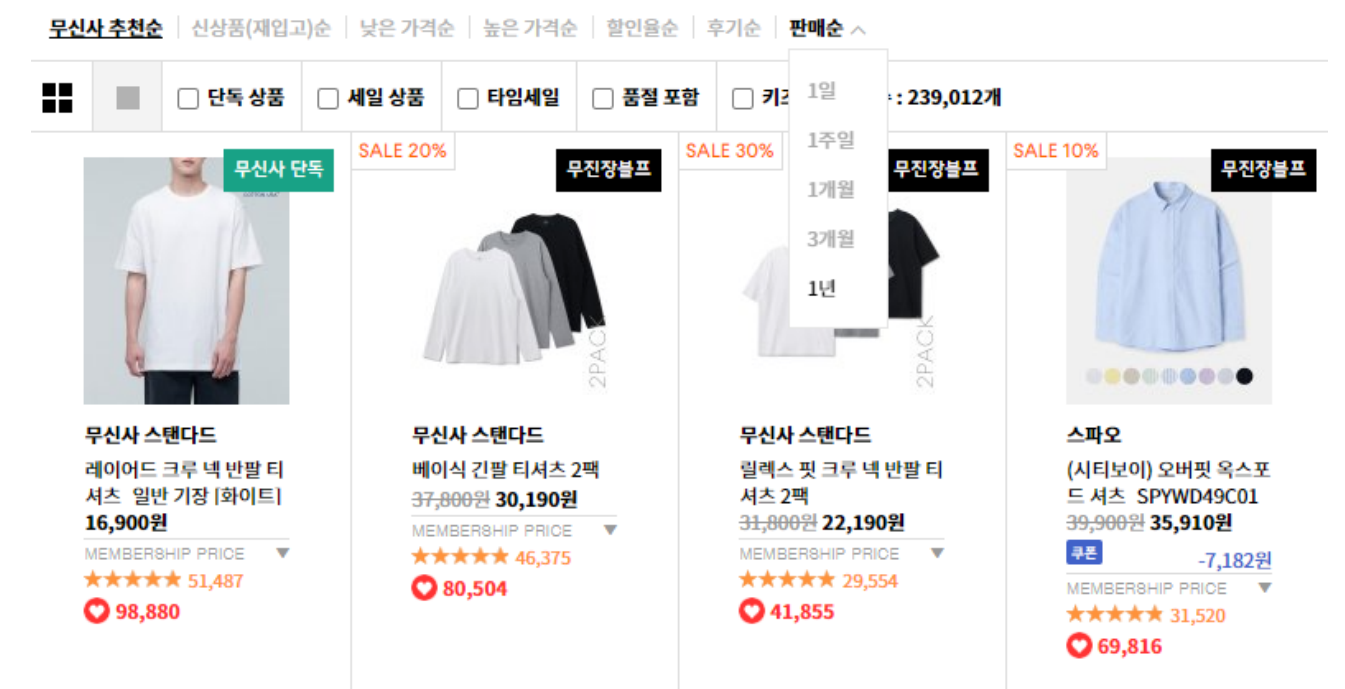
3. 필요 자료 및 수집 방법



무신사 플랫폼 품목 중 대표 N가지 선정

*1안) N = 첫 페이지 노출 상품 수

*2안) N = 50개



'판매순' 정렬을 1년으로 세팅하여
상위 순위 N개 선정

4. 감성분석이란

텍스트에 나타난 주관적 요소인 긍정/부정 감정을 판별해 수치/도식/등급 등으로 정량화 하는 작업.
단순 긍정/부정을 추출하는 것이 아닌 대상이 되는 개체나 속성을 추출하거나,
감성을 표현하는 이들의 감정, 감성, 태도, 입장 등을 분석함

```
sentiment_predict('이 상품 진짜 좋아요... 저는 강추합니다. 대박')
```

98.88% 확률로 긍정 리뷰입니다.

```
sentiment_predict('진짜 배송도 늦고 개짜증나네요. 뭐 이런 걸 상품이라고 만듦?')
```

99.31% 확률로 부정 리뷰입니다.

참고) 위키독스- 딥 러닝을 이용한 자연어 처리 입문
<https://wikidocs.net/94600>

4. 감성분석 단계

1단계 데이터 수집

- 웹사이트 게시판 선별, 상품평 데이터 수집

2단계 데이터의 전처리

자연어 데이터 -> 컴퓨터 인식할 수 있는 단어 변환하기 위해 토큰화/정수 인코딩/패딩 등 전처리

3단계 형태소 사전의 구축

해당 산업(이커머스, 패션)에 특화된 형태소 사전 구축

4단계 속성어 사전 구축

화자가 어느 속성에 대해 어떻게 서술하는지 파악

5단계 감성어 사전 구축

6단계 감성 분석 모형의 구축 및 결과 도출

참고) 위키독스- 딥 러닝을 이용한 자연어 처리 입문
<https://wikidocs.net/94600>

5. 추후 과제 진행 계획

12월 1주차 | 주제 선정 및 기획 발표(12/1)

12월 2주차 | 주제 및 계획 확정, 데이터 수집

12월 3주차 | 데이터 전처리, 모델 선택 및 훈련, 중간 발표(12/13)

12월 4주차 | 모델 평가 및 결과 시각화

12월 5주차 | 프로젝트 성능 평가 및 최종 발표(12/27)

6. 당면 과제

이커머스 현업과 밀접한 프로젝트를 진행하고 싶은데,

현재 설정한 기업인 무신사는 **구체적인 매출 데이터를 구할 수 없음.**

매출액 대신 1년 판매량 기준으로 구매후기수, 좋아요수 등의 요소와 상관관계를 찾으려 함.

과연 <매출> 데이터를 분석하는 것만큼 유의미한 결과가 나올 수 있을까?

*현업과 유사한 프로젝트를 진행하고 싶은데, kaggle 등에 등재된 정제된 데이터를 사용하는 게 좋을까요?

감사합니다