

# 상품군별 소비자의 주요 고려 요소 파악: 쿠팡 리뷰 데이터를 중심으로

팀 번호 10

팀원 1: 석누리, 2020118013, 문헌정보학과, nuriyonsei@yonsei.ac.kr

팀원 2: 허원재, 2020195020, 계량위험관리학과, hushpond@gmail.com

## 1. 서론 Introduction

### 1) 배경 Background

기업이 신제품을 출시하기 위해서, 혹은 이미 출시되어 있는 자사 상품이 시장에서 어떤 경쟁력을 가지고 있는지 조사하기 위해서 소비자들의 성향과 니즈를 분석하는 것은 필수적이다. 이렇게 소비자들이 원하는 것이 무엇인지 직접 파악하는 방법에는 설문조사, 리뷰 분석 등이 있다.

한편, 소비자는 상품을 구매함에 있어서 원하는 상품을 구매하기 위해 상당히 많은 시간을 투자하여 정보를 탐색하고 이를 소비자의사결정에 반영한다. 이 과정에서 제품 설명란, 리뷰, 사용설명서, 쇼핑몰이 자체 구축한 상품 비교 서비스 등 소비자가 이용하는 정보원은 상당히 많다.

소비자는 구매의사를 판단하는 '최종 의사결정' 단계에서 다른 사람들이 소비한 상품에 대해 어떻게 생각했는지 확인하고 추측한 후 최종적인 판단을 내린다. 특히 상품에 대한 객관적인 판단 자료가 충분히 확보되지 않은 경우 소비자는 판단을 위한 정보를 찾게 되는데, 이 과정에서 사람들은 '다른 사람들이 그 상품을 어떻게 평가했는지'에 의존한다. 이러한 이론적 배경에 근거하여, 우리 조는 '리뷰'라는 정보원을 분석할 필요가 있음을 느꼈다. '별점'이라는 수치화된 리뷰와 함께, 상품에 대한 소비자의 주관적인 평가가 반영되어 있는 리뷰 분석을 통해 기업은 소비자의 니즈를 분석할 수 있고, 시장을 분석할 수 있으리라 판단했다.

### 2) 목적 및 필요성 Purpose

쿠팡에서 판매되고 있는 특정 상품군들 중 판매량 상위 10 개 상품을 선정하여 리뷰와 별점을 분석함으로써 소비자들이 공통적으로 언급하고 있는 '상품의 장점'과 '상품의 아쉬운 점'을 파악할 수 있다. 이를 통해 해당 상품군에서 시장 경쟁력을 확보하기 위해 기업이 고려하여야 할 요소가 무엇이 있는지 파악할 수 있을 것이다.

### 3) 프로젝트 내용 및 의의 Significance

쿠팡에서 특정 상품군을 선정하여 그 중 판매량 상위 10 위에 속하는 상품들의 리뷰를 분석한다. 쿠팡이 국내 최대 규모의 소셜커머스 업체인만큼 다른 업체의 리뷰들에 비해 방대한 양의 리뷰에 접근 및 분석함으로써 기업이 시장 경쟁력을 가진 상품을 출시하기 위해 무엇을 고려해야 하는지 제언 및 시각화할 수 있을 것이라고 판단했다.

## 2. 데이터 Data

## 1) 데이터 수집 Data collection

쿠팡에서 많이 판매되는 특정 상품군을 키워드로 검색한 후, (가령 '물'일 경우 '물'을 검색하여 나오는 상품들 중 판매량 기준 상위 10 개 상품) 각 상품별로 나타나는 '별점'과 소비자가 직접 글로 작성한 '후기'가 우리 조가 추출할 데이터이다. 쿠팡의 Open API에서는 리뷰 데이터를 제공하지 않기 때문에 상품의 카테고리 분류에서는 API를 일부 사용하되, 일반 이용자가 접근할 수 있는 페이지를 대상으로 한 웹크롤링을 데이터 수집에서의 주된 도구로 이용할 것이다.

## 2) 데이터 설명 Data description

각 리뷰 한 단위가 곧 샘플 하나가 되고, 변인은 상품 종류, 상품명(또는 코드), 별점, 리뷰 raw text 정도가 될 것이다. 수집한 데이터를 어떤 형식으로 저장할 것인지 문제 되는데, 계획 단계에서는 일단 상품 분류별로 JSON 파일을 만들어 이를 처리하는 형태를 상정하였다.

## 3. 방법 Methods

### 1) 주제 유형 Type of the analysis

프로젝트가 다루고자 하는 주요 분석 유형은 Aspect-based Sentimental Analysis이다. 프로젝트 구현 과정에서, 이미 쿠팡의 카테고리 분류로 제품이 제시되고, 별점을 통해서 전체 리뷰에서의 감정 분석이 어느 정도 되어 있기는 하나, 리뷰 데이터의 문장마다 Sentiment classification이 수행되어야 할 것이다.

### 2) 예상되는 난제 Challenges

본 프로젝트에서 가장 중요한 문제는 팀원 모두가 전문적으로 프로그래밍을 배우지 않은 상태이기에, 구현 그 자체에서 어려움을 겪을 가능성이 매우 높다. 나아가 실제 프로젝트 개발이 어느 정도 선에서 이루어지는지 파악되지 않으면, 설계 단계에서도 비현실적인 목표 설정과 같은 문제가 생길 우려가 있다.

데이터의 수집 난이도는 쿠팡에서 물리적으로 막지 않는 한, 어렵지는 않을 것으로 예상된다. 리뷰 데이터는 대체로 정형화되어 있지 않기 때문에, 여기서 원하는 정보를 추출하는 작업은 간단하지 않다. Aspect term 설정 시, 미리 정해진 리스트를 만들 것인지, 아니면 자동으로 추출된 것을 기반으로 만들 것인지 확실하지 않다.

이 프로젝트는 각기 다른 제품마다 소비자들의 다양한 선호 방식이 나타날 것이라고 가정하고 진행되지만, 분석 수행 결과 카테고리 사이에서 주목할 만한 차이점을 보이지 않는다면 프로젝트 결과의 의미가 퇴색될 수 있다. 아마존, 네이버, 쿠팡 등 여러 이커머스 업체들이 많은 연구를 통해 이미 그 결과를 실제 서비스에 반영하고 있는 바, 본 프로젝트의 목표인 기업 또는 소비자들에게 제공될 제품 영역별 주요 관심 사항이 현재 서비스에 드러나지 않는 경우, 지금 상용화되어 있는 서비스와의 구별점이 문제 된다.

## 4. 참고문헌 References

김영신, 강이주, 이희숙, 정순희, & 허경옥. (2009). 새로 쓰는 소비자의사결정. 주) 교문사, 90-92.

안은미. (2013). 소비자 심리학. 서울: 박학사. 90.

Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. Expert systems with applications, 118, 272-299.

[E-commerce] 아마존, 네이버, 쿠팡, 11 번가는 리뷰를 어떻게 이해하고 보여줄까? (Aspect-based), 2022 년 4 월 30 일 검색, <https://velog.io/@jonas-jun/ABSA> 사례