

## 1. 주제

아마존 제품 데이터 기반 소비자 반응 분석 및 경쟁사와의 비교 시스템

분반, 팀, 학번, 이름

1반, 1팀, 20243302, 최연하

## 2. 요약

본 프로젝트는 Amazon 제품 데이터를 활용하여 소비자 반응에 영향을 미치는 핵심 요인을 분석하고, 경쟁사 대비 자사 제품의 시장 포지션을 파악함으로써 데이터 기반 의사결정을 지원하는 데이터 분석 프로그램을 구현하는 것을 목표로 한다.

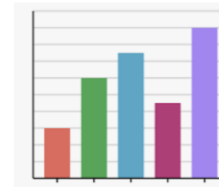
이 데이터 분석 프로그램은 두 개의 핵심 모듈로 구성된다. 첫째, 소비자 요인 분석 모듈은 머신러닝 기반 요인 분석, 감성 분석, 회귀 모델링을 통해 가격, 리뷰 특성, 감성 점수 등이 평점과 구매 의도에 미치는 영향을 정량화한다. 둘째, 경쟁사 분석 모듈은 카테고리별 가격-평점 포지션 매핑, 키워드 트렌드 분석을 통해 경쟁 환경 내 자사 제품의 강점과 약점을 제시한다. 이 두 모듈의 통합을 통해 "왜 잘 팔리는가"와 "경쟁사 대비 어떤 점이 우수한가"를 동시에 파악할 수 있는 종합적인 분석 기능을 제공한다.

전자상거래 시장에서 데이터 기반 의사결정은 더 이상 선택이 아닌 필수 요소가 되었다. 본 시스템은 제품 개발팀에게 소비자 반응 데이터 기반의 제품 개선 근거를 제공하고, 마케팅팀에게는 소비자 선호 요소를 제시하여 더 좋은 홍보 전략을 세우는 것을 돕는다. 또한, 경영진에게는 경쟁사 대비 전략적 포지셔닝과 투자 방향 수립을 위한 객관적 지표를 제시한다. 이를 통해 제품 경쟁력 강화, 마케팅 효율성 증대, 시장 점유율 확대라는 실질적 비즈니스 성과를 창출할 수 있을 것으로 기대된다.

## 3. 대표 그림

### A사 카메라 분석 결과

#### 1. 소비자 반응 분석



리뷰분석:

긍정적 리뷰

1. 빠른 배송(21.2%)

부정적 리뷰

1. 배송상태 지적(10.3%)

2. 늦은 문의 답변 지적(4.1%)

소비자가 자사 제품을 선택한 이유:

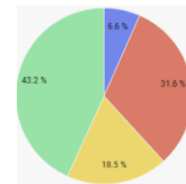
1. 가격(34.6%)

2. 품질(27.3%)

### A사 카메라 분석 결과

#### 2. 경쟁사 분석

자사 포지션 매핑



키워드 트렌드 매핑



경쟁사 대비 강점

1. 가격

2. 품질

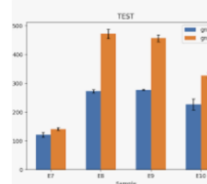


그림 1 예상 UI

## 4. 서론

### 4.1 배경 설명 및 사례 분석

전자상거래 시장의 급속한 성장과 함께 소비자 리뷰 데이터는 구매 의사결정의 핵심 요소로 자리잡았다. Amazon만 해도 하루 수십만 건의 리뷰가 생성되며, 이러한 비정형 데이터에는 소비자의 선호, 불만, 구매 동기 등 귀중한 정보가 내포되어 있다. 그러나 대부분의 기업들은 평균 평점이나 리뷰 개수와 같은 단순 지표에만 의존하며, 방대한 데이터의 잠재력을 충분히 활용하지 못하고 있다.

기업들의 사례를 살펴보면, Netflix는 시청 패턴 분석을 통해 콘텐츠 추천과 제작 의사결정에 활용하고 있으며, Airbnb는 호스트 리뷰의 감성 분석으로 서비스 품질을 관리한다. 이처럼 데이터 기반 의사결정은 현대 비즈니스의 필수 역량이 되었다. 특히 Amazon과 같은 플랫폼에서는 수많은 경쟁 제품이 공존하기 때문에, 단순히 좋은 제품을 만드는 것을 넘어 "소비자가 왜 특정 제품을 선호하는지", "경쟁사 대비 우리의 강점은 무엇인지"를 정확히 파악하는 것이 생존의 조건이 되었다. (이효정, 2018)

### 4.2 문제 정의

현재 전자상거래 환경에서 기업들이 직면한 주요 문제는 다음과 같다:

첫째, 소비자 반응 요인의 불명확성이다. 평점이 높은 제품과 낮은 제품의 차이가 단순히 품질 때문인지, 가격 때문인지, 아니면 다른 요인 때문인지 명확히 파악하기 어렵다. 이로 인해 제품 개선이나 가격 정책 수립 시 객관적 근거 없이 직관에 의존하는 경우가 많다.

둘째, 경쟁사 분석의 한계이다. 기존 방식은 주로 가격 비교나 평점 비교에 그치며, 리뷰 내용의 감성, 소비자가 언급하는 핵심 키워드, 카테고리별 트렌드 등 심층적인 경쟁 정보를 파악하지 못한다. 이는 시장 내 자사 포지션을 정확히 인식하지 못하고 잘못된 전략적 의사결정으로 이어질 수 있다.

셋째, 비정형 데이터 처리의 기술적 장벽이다. 리뷰 텍스트는 비정형 데이터로, 이를 분석하려면 자연어 처리(NLP), 감성 분석, 머신러닝 등 전문적 기술이 필요하다. 대부분의 기업, 특히 중소기업은 이러한 기술 역량과 인프라가 부족하여 데이터 활용에 어려움을 겪는다.

(WanniPreeti, 2025) (이수민, 2024)

### 4.3 극복 방안

본 프로젝트는 위 문제들을 해결하기 위해 다음과 같은 접근 방식을 제시한다:

첫째, 머신러닝 기반 요인 분석 자동화를 통해 평점에 영향을 미치는 요인들의 중요도를 정량화한다. Random Forest, XGBoost와 같은 앙상블 모델과 SHAP, LIME 같은 설명 가능한 AI 기법을 활용하여, "가격이 10% 하락하면 평점이 0.3점 상승"과 같은 구체적 인사이트를 도출한다. 이를 통해 의사결정의 객관성과 예측 가능성을 확보한다.

둘째, 감성 분석 및 키워드 추출을 통해 리뷰 텍스트에 숨겨진 소비자 의견을 구조화한다. BERT 기반 감성 분석 모델과 TF-IDF 키워드 추출 기법을 사용하여 긍정/부정 리뷰의 핵심 키워드를 파악하고, 이를 경쟁사와 비교함으로써 차별화 포인트를 발견한다.

셋째, 직관적인 시각화를 통해 편한 UI를 구현한다. Streamlit이나 Plotly를 활용하여 인터페이스를 구축하여 비전문가도 쉽게 분석 결과를 이해하고 활용할 수 있도록 한다.

넷째, 모듈형 설계를 통해 확장성과 유지보수성을 확보한다. 소비자 요인 분석 모듈과 경쟁사

분석 모듈을 독립적으로 개발하되 통합 인터페이스를 제공하여, 향후 새로운 분석 기능 추가나 데이터 소스 확장이 용이하도록 설계한다.

## 5. 본론

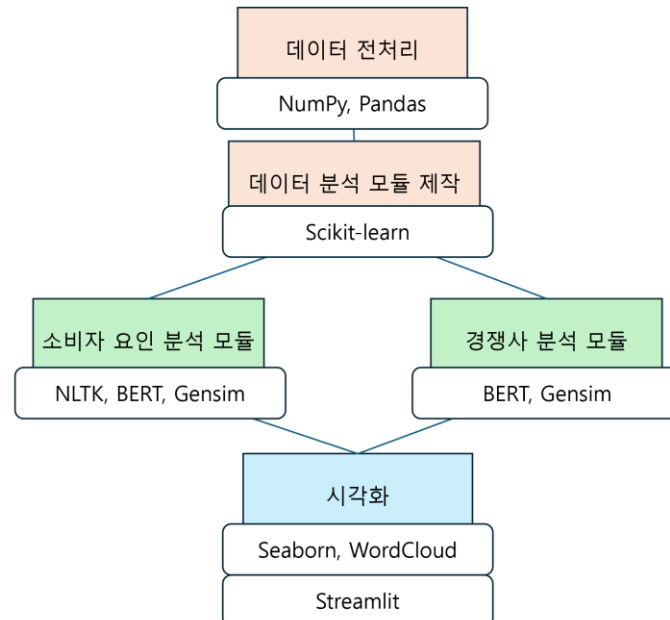


그림 2 시스템 개요 그림

### 5.1 시스템 개요 그림

본 시스템의 구현에는 데이터 처리, 자연어 처리, 설명 가능한 AI, 시각화, 웹 애플리케이션 개발 등 다섯 가지 핵심 기술 요소가 요구된다. 우선 데이터 처리 및 분석 단계에서는 Pandas와 NumPy를 이용해 대용량 데이터를 효율적으로 가공하고, Scikit-learn를 활용해 예측 모델을 구축한다. 이를 통해 제품 평점에 영향을 미치는 주요 요인을 정량적으로 파악한다.

자연어 처리(NLP) 영역에서는 NLTK를 통해 리뷰 데이터를 전처리하고, BERT 기반 모델을 이용해 감성 분석을 수행한다. 추가적으로 Gensim을 사용하여 주요 키워드와 토픽을 추출함으로써 소비자 인식의 패턴을 분석한다.

### 5.2 필요한 기술요소

설명 가능한 AI 기술로는 SHAP과 LIME을 도입하여 각 변수의 영향력을 시각적으로 설명함으로써 모델의 해석 가능성을 높인다. 시각화 단계에서는 Plotly, Seaborn, WordCloud 등을 활용하여 감성 분포, 주요 요인, 경쟁사 비교 결과를 직관적으로 표현한다. 마지막으로 인터페이스 구현에는 Streamlit을 사용하여 데이터 분석 결과를 대시보드 형태로 제공하고, 비전문가도 손쉽게 분석 결과를 탐색할 수 있는 환경을 구축한다.

### 5.3 구현 방법

개발은 단계적으로 진행된다. 초기 단계에서는 Kaggle 등에서 확보한 Amazon 제품 데이터를 전처리하고 품질을 검증한다. 이후 소비자 요인 분석 엔진을 개발하여 감성 분석과 평점 예측 모델을 구축하고, 경쟁사 분석 대시보드를 통해 시장 내 포지션과 주요 경쟁 요인을 시각화한다. 마지막 단계에서는 Streamlit을 기반으로 통합 대시보드를 완성하고, 테스트 및 최적화를

거쳐 사용자 친화적인 인터페이스를 구현한다. 전체 과정은 데이터 기반 분석의 정확성과 시각화의 직관성을 동시에 확보하는 방향으로 진행된다.

## 6. 결론

본 제안서는 Amazon 제품 데이터를 활용한 데이터 분석 프로그램을 제시하였다. 이 시스템은 두 개의 핵심 모듈로 구성된다:

1. 소비자 요인 분석 모듈: 머신러닝과 자연어 처리 기술을 활용하여 평점에 영향을 미치는 핵심 요인을 분석하고, 소비자 감성을 정량화한다. 이를 통해 제품 개선과 마케팅 전략 수립을 위한 데이터 기반 전략을 제공한다.
2. 경쟁사 분석 모듈: 경쟁사 제품과의 비교 분석을 통해 시장 내 자사 포지션을 파악하고, 가격, 평점, 감성 등 다양한 방면에서의 분석을 통해 개선 방향을 제공한다.

기술적으로는 scikit-learn, XGBoost, BERT 등의 머신러닝 및 NLP 기술을 활용하며, SHAP과 LIME을 통해 설명 가능한 AI를 구현하여 사용자의 이해를 돕는다. Streamlit 기반의 시각화로 쉽게 활용할 수 있는 사용자 경험을 제공한다.

향후에는 Kaggle에서 제공되는 Amazon 제품 리뷰 데이터를 수집하여 분석의 기반을 마련할 예정이다. 수집된 데이터는 결측치 처리, 텍스트 정제, 변수 변환 등 체계적인 전처리 과정을 거친 후 데이터 분석을 할 예정이다.

## 7. 출처

### 인용 자료

WanniPreeti. (2025). "감정 분석 시장 규모 및 전망." Research Nester.  
<https://www.researchnester.com/kr/reports/sentiment-analytics-market/7054>에서 검색됨

"사용자 행동 데이터 분석: 넷플릭스와 아마존은 데이터 분석을 어떻게 할까요?" (2022년 7월 18일). 요즘 IT: <https://yozm.wishket.com/magazine/detail/1588/>에서 검색됨

이수민. (2024). 쇼핑몰 리뷰 데이터 감성 분석 및 키워드 추출. "2024 한국컴퓨터종합학술대회", (페이지: 1,693 - 1,695).

이효정. (2018). "비즈니스 기회 창출을 위한 AI알고리즘의 활용." 삼성KPMG.