

쿠팡 세제 리뷰 감성분석 활용한 고객 상품 추천



Content

1. 배경

주제 선정
분석 프로세스

2. 데이터 정제

데이터 수집
데이터 가공
EDA

3. 모델링

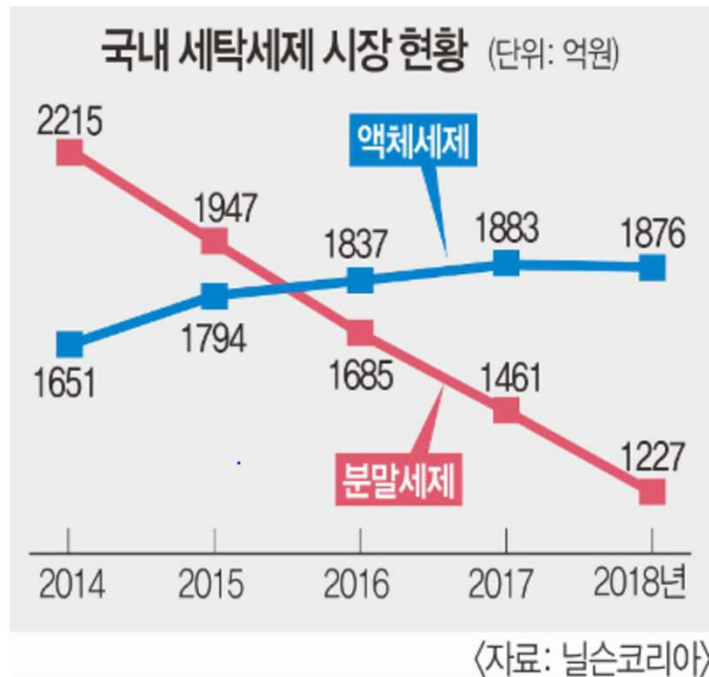
모델 소개
최적 모델 선택 / 활용

4. 결론

활용 방안
시사점
한계점

1. 주제 선정

유형별 세탁세제 시장 현황

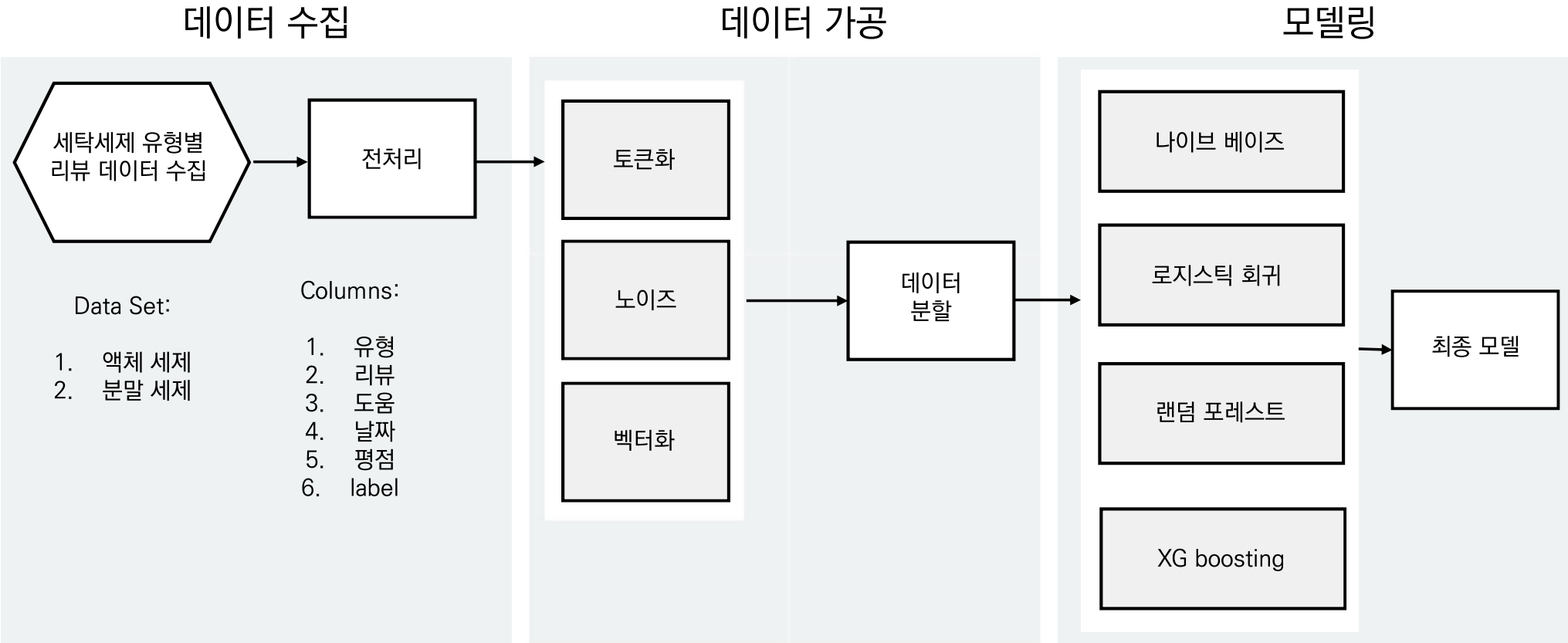


19일 관련 업계에 따르면 지난해 국내 세탁세제 시장은 5220억원으로 전년(4900억원)보다 320억원 커졌다. 업계 관계자는 "코로나19로 집에 머무는 시간이 길어지면서 빨래를 몰아서 하기보다는 자주 하는 경향이 나타났다"며 "세제 사용량이 늘고 시장 규모도 많이 커졌다"고 설명했다.

액체세제는 꾸준한 성장, 분말세제는 시장 침체

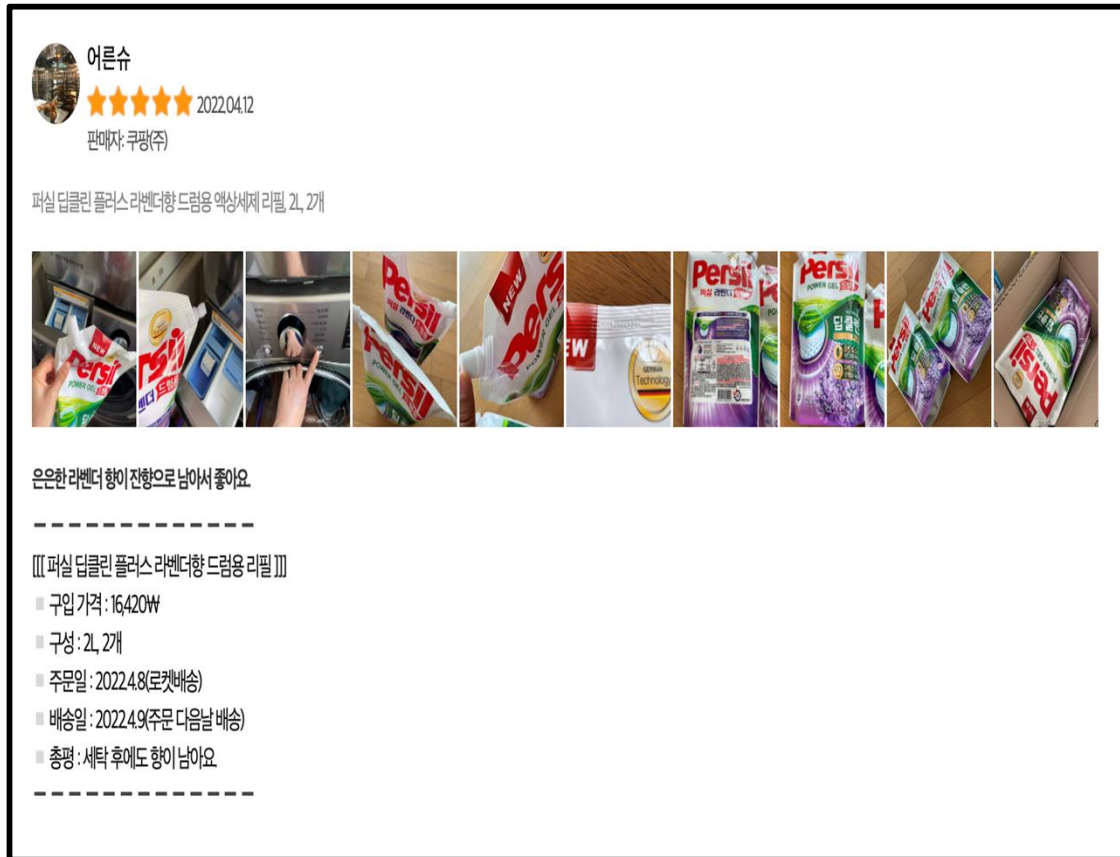
1. 분석 프로세스

분석 프로세스



2. 데이터 수집

세탁세제 유형별 리뷰 데이터 수집



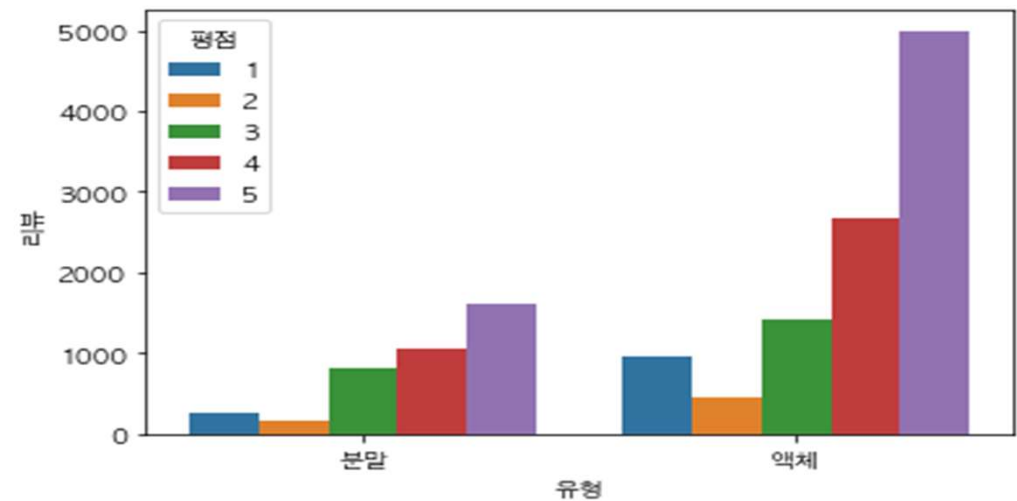
〈쿠팡 세탁 세제 리뷰〉

Under – Sampling

별점 3개 이하 리뷰의 개수가 4개 이상보다 현저히 적다.

별점 3개 이하 리뷰는 200 페이지까지 크롤링하고

별점 4개 이상 리뷰는 50 페이지까지 크롤링한다



〈평점별 구간 빈도〉

2. 데이터 수집

데이터 소개

리뷰 출처 : 쿠팡의 세탁세제 중 액체 •분말 세제 판매량 높은 순으로 약 50개 상품의 베스트 리뷰

Columns / Rows :

6 개 / 14,441 개

Columns :

유형 : 세탁 세제의 종류 (액체 •분말)

리뷰 : 쿠팡에서 세탁세제를 구매한 사람의 후기 중 베스트 리뷰에 선정된 것

도움 : 해당 리뷰가 타 구매자에게 도움이 된 수

날짜 : 구매자가 해당 리뷰를 남긴 날짜

평점 : 구매자가 해당 제품에 매긴 점수 (5점 만점)

Label: 평점 4점 이상은 긍정 리뷰, 4점 미만은 부정 리뷰로 간주

토큰화

Open Korea Text (OKT)

모든 형태소

노이즈

한글 및 공백 외 모두 제거

불용어 : 788 개

벡터화

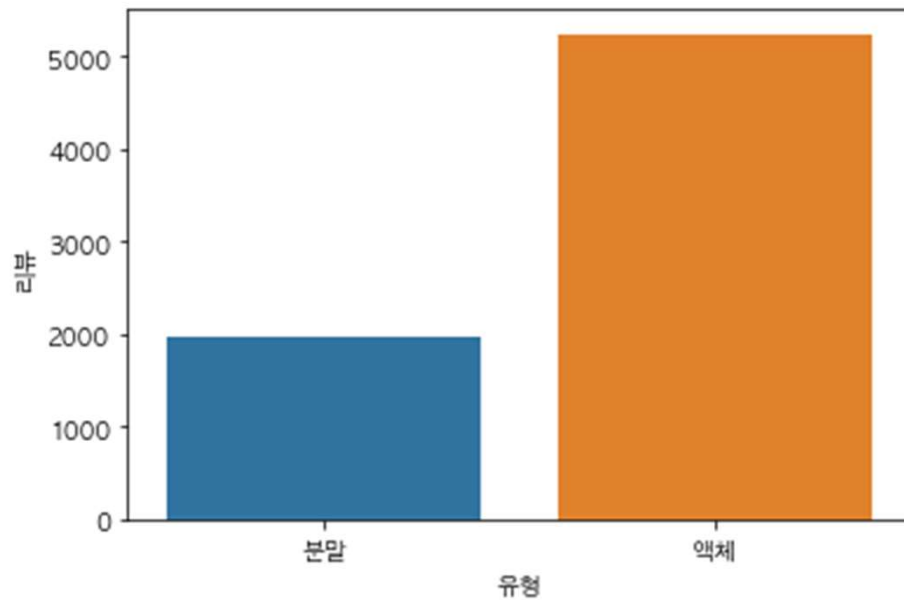
TF - IDF

ngram_range=(1, 3)

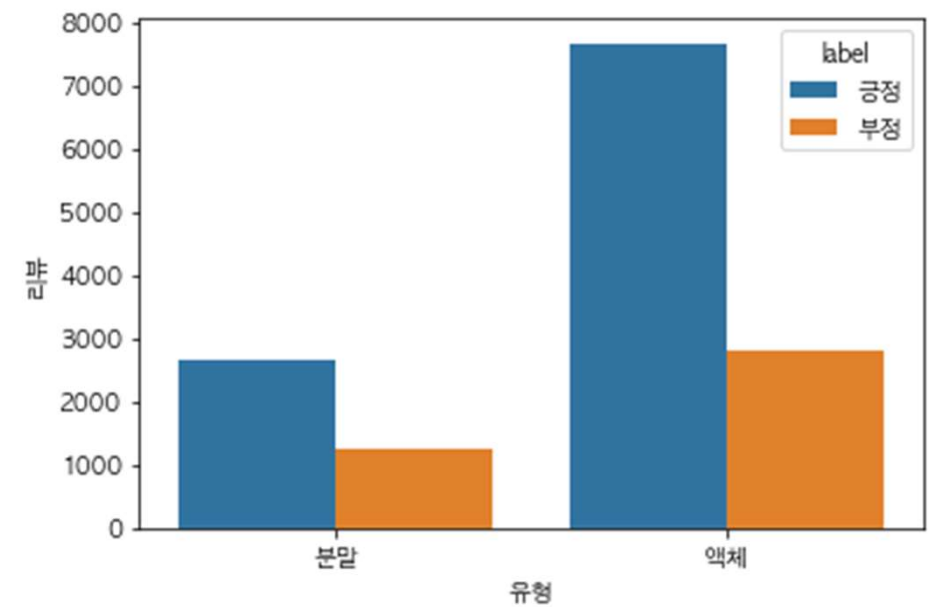
2. EDA

세탁세제 유형별 리뷰 수

유형별 리뷰 수



유형별 긍정/부정 리뷰 수



액체 세제 리뷰 수와 긍정 리뷰 수가 가장 많다

2. EDA

액체 서제 리뷰의 주요 키워드



긍정 리뷰
은은하다, 진드기, 건조기,
편이, 아기, 가성



부정 리뷰
작다, 뭉치다, 빨리, 무게, 가성비,
귀찮다, 헛금

2. EDA

분말 세제 리뷰의 주요 키워드



가성, 찌들다, 잘빠지다, 빨리, 브랜드, 삶다,
편이



뭉치다, 작다, 무게, 행굼, 빨리, 가성

3. 최적 모델 선택 / 활용

액체 세제 리뷰 최적의 머신러닝 모델

나이브 베이즈

정확도 : 73.38%
정밀도 : 73.49%
재현율 : 95.4%
F1 : 83.02%

랜덤 포레스트

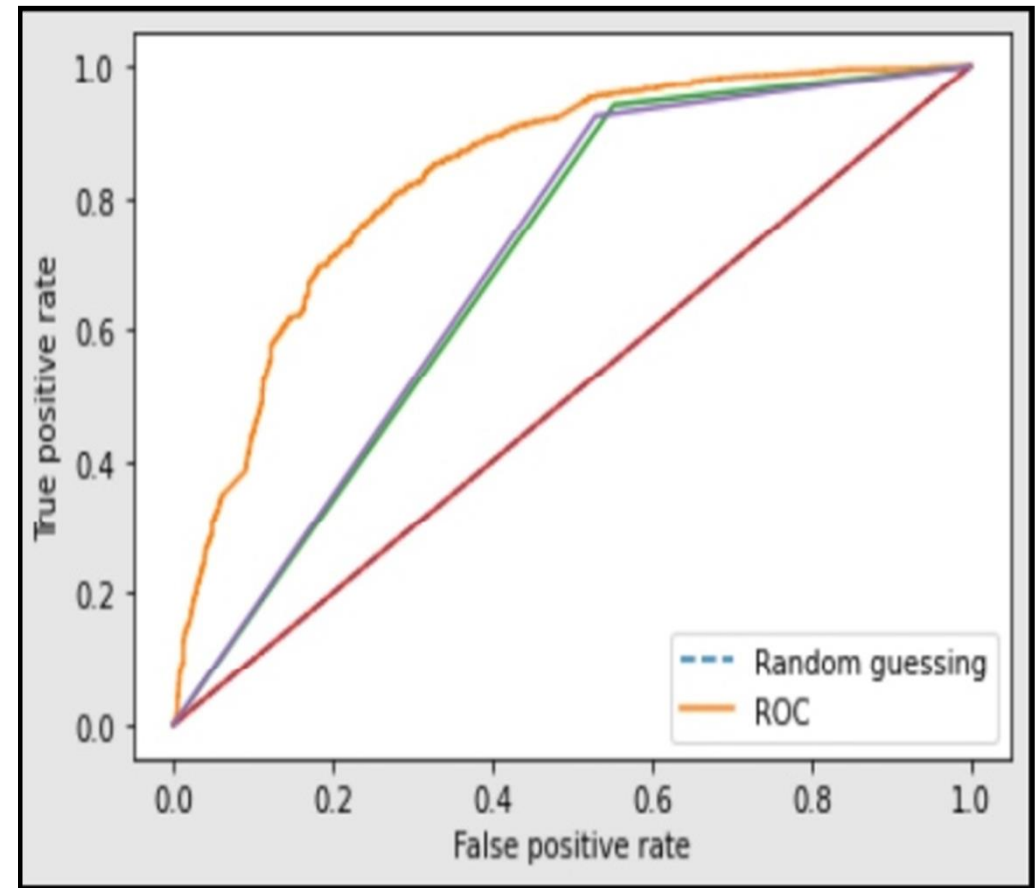
정확도 : 68.22%
정밀도 : 68.22%
재현율 : 100%
F1 : 81.1%

로지스틱 회귀

정확도 : 82.41%
정밀도 : 83.03%
재현율 : 95.68%
F1 : 88.91%

XGB boosting

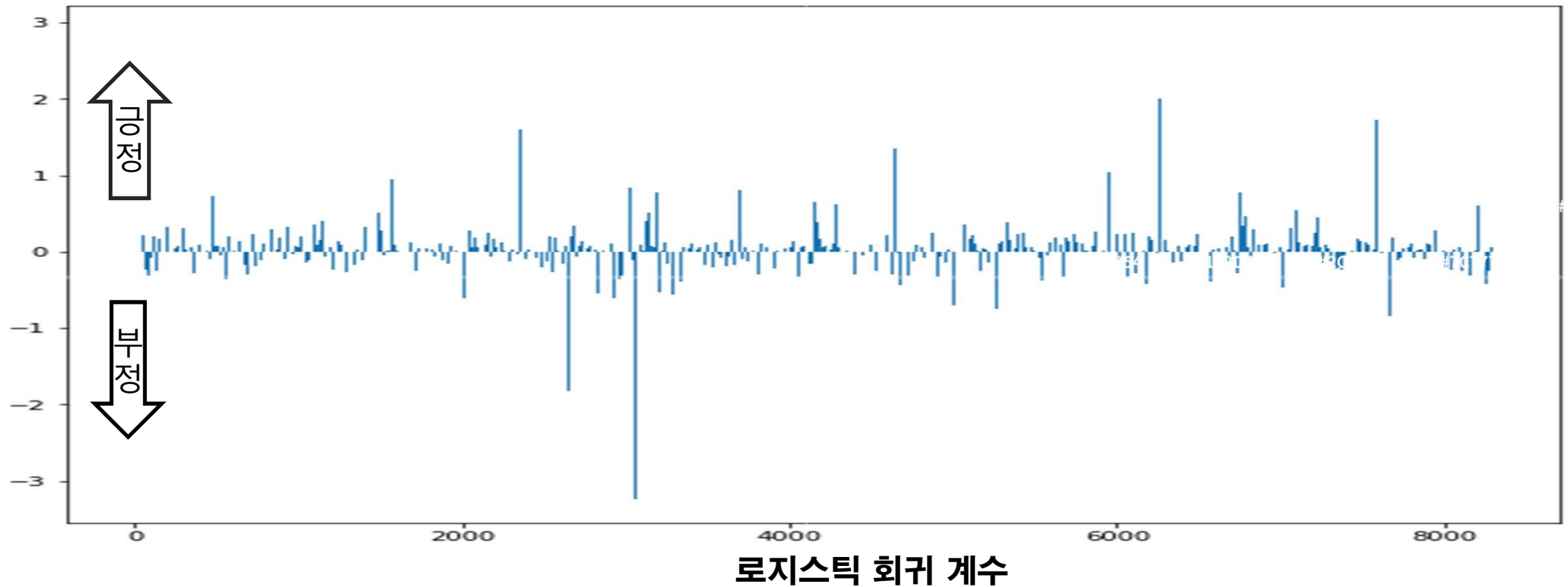
정확도 : 77.03%
정밀도 : 82.96%
재현율 : 83.47%
F1 : 83.21%



〈액체 세제 ROC 결과〉

3. 최적 모델 선택 / 활용

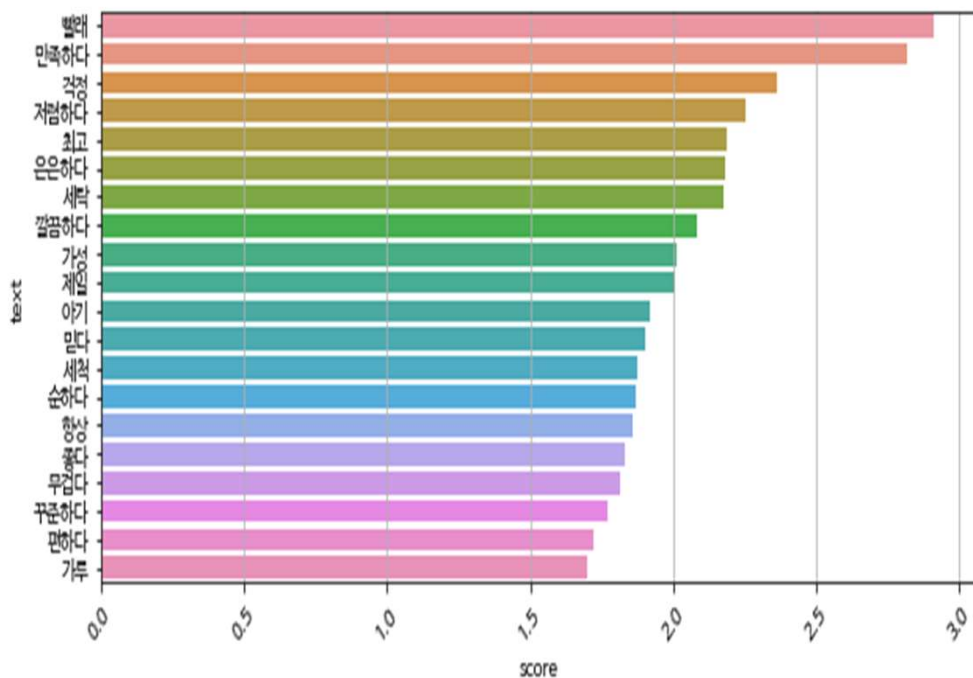
액체 세제 최종 모델 결과 및 시각화



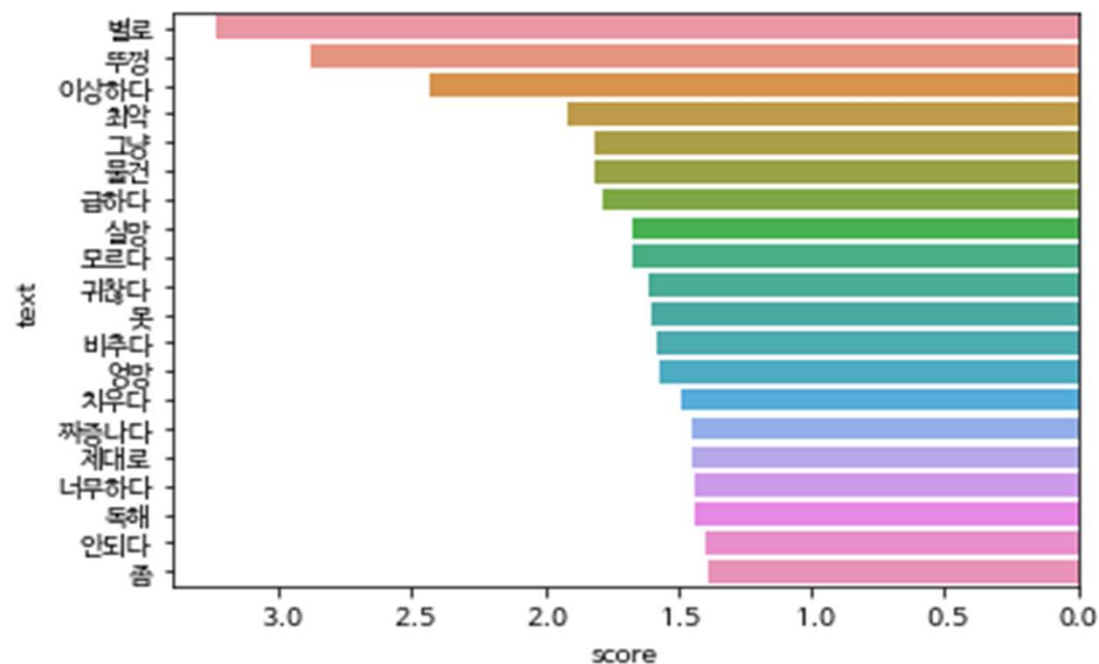
회귀 계수가 양수인 경우 긍정을 음수인 경우 부정을 나타낸다

3. 최적 모델 선택 / 활용

액체 세제 리뷰 최적의 머신러닝 모델



〈긍정 단어〉



〈부정 단어〉

빨래에 만족하고, 걱정이 없으며, 가격이 저렴하고 은은한 향과 아기 옷 세탁과 관련하여 만족하는 것으로 보인다. 두껍의 경우 별로 생각하며 향이 독한 것과 사용하기 귀찮은 경우가 있는 것으로 보인다.

3. 최적 모델 선택 / 활용

분말 세제 리뷰 최적의 머신러닝 모델

나이브 베이즈

정확도 : 73%
정밀도 : 73%
재현율 : 26%
F1 : 38%

랜덤 포레스트

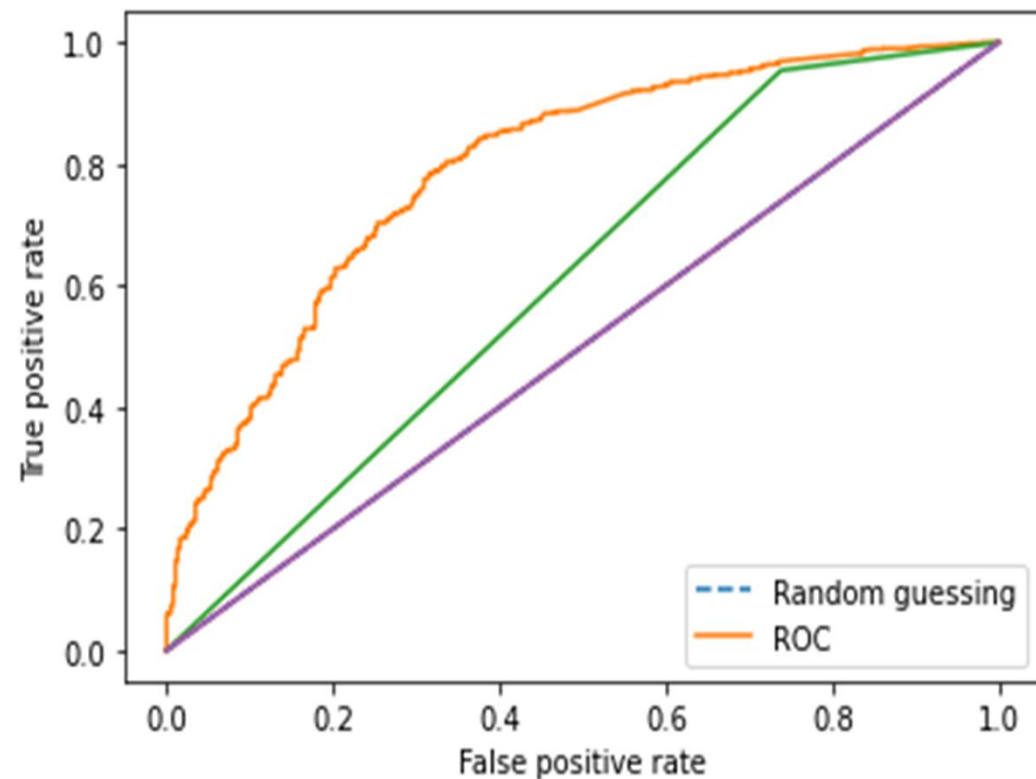
정확도 : 68.22%
정밀도 : 68.22%
재현율 : 100%
F1 : 81.1%

로지스틱 회귀

정확도 : 75.9%
정밀도 : 76.4%
재현율 : 93.5%
F1 : 84.1%

XGB boosting

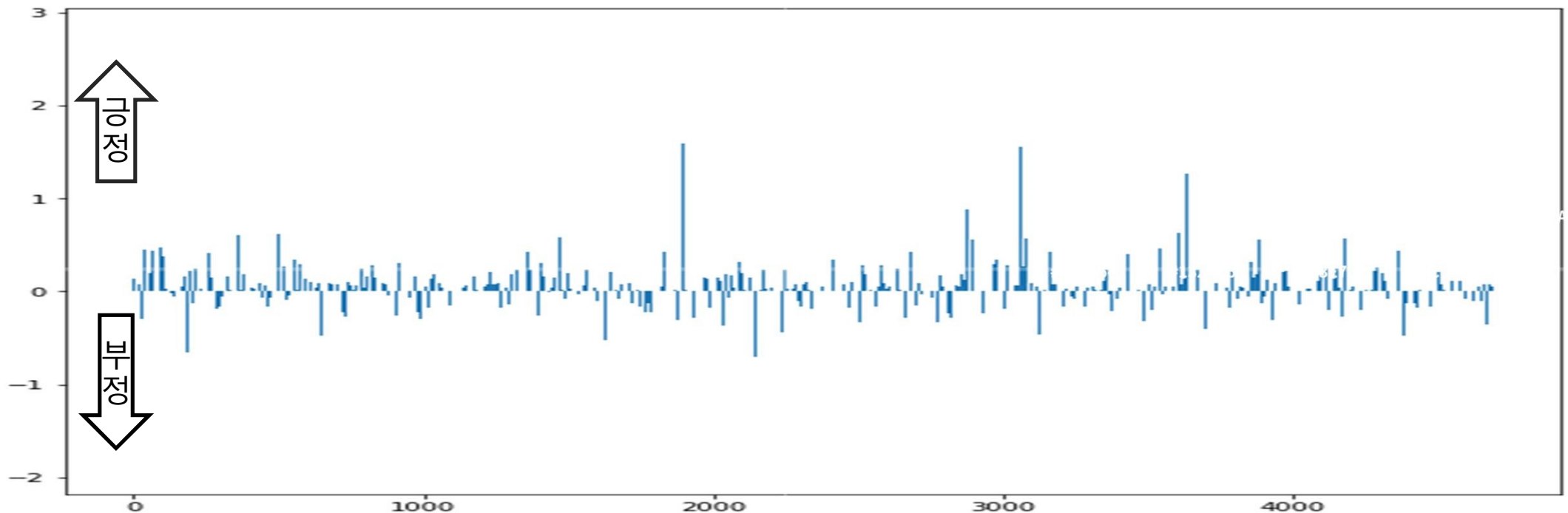
정확도 : 76.4%
정밀도 : 82.2%
재현율 : 83.4%
F1 : 82.28%



〈액체 세제 ROC 결과〉

3. 최적 모델 선택 / 활용

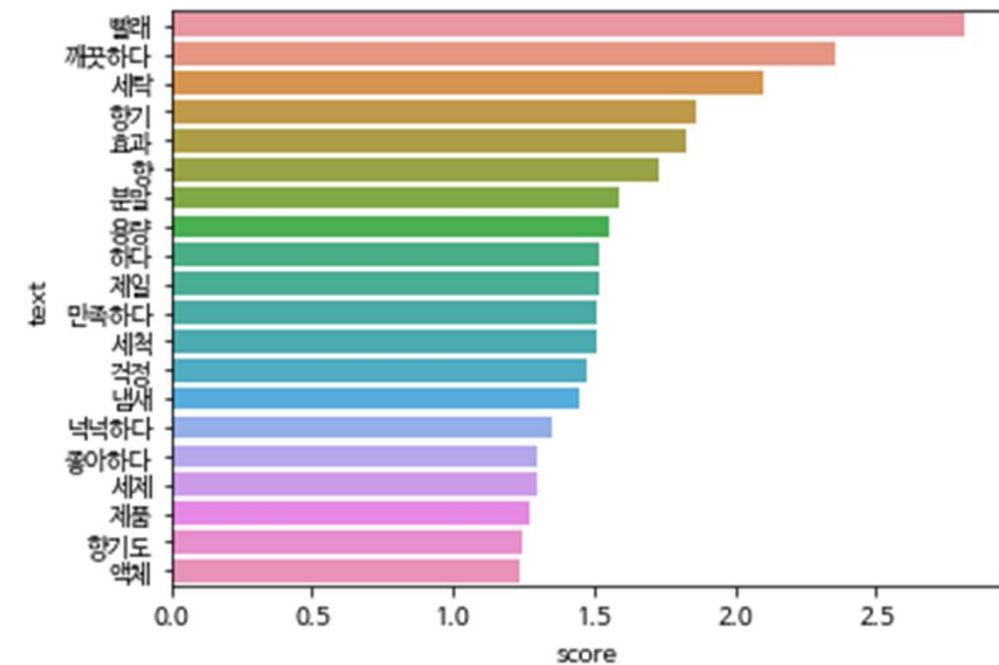
분말 세제 최종 모델 결과 및 시각화



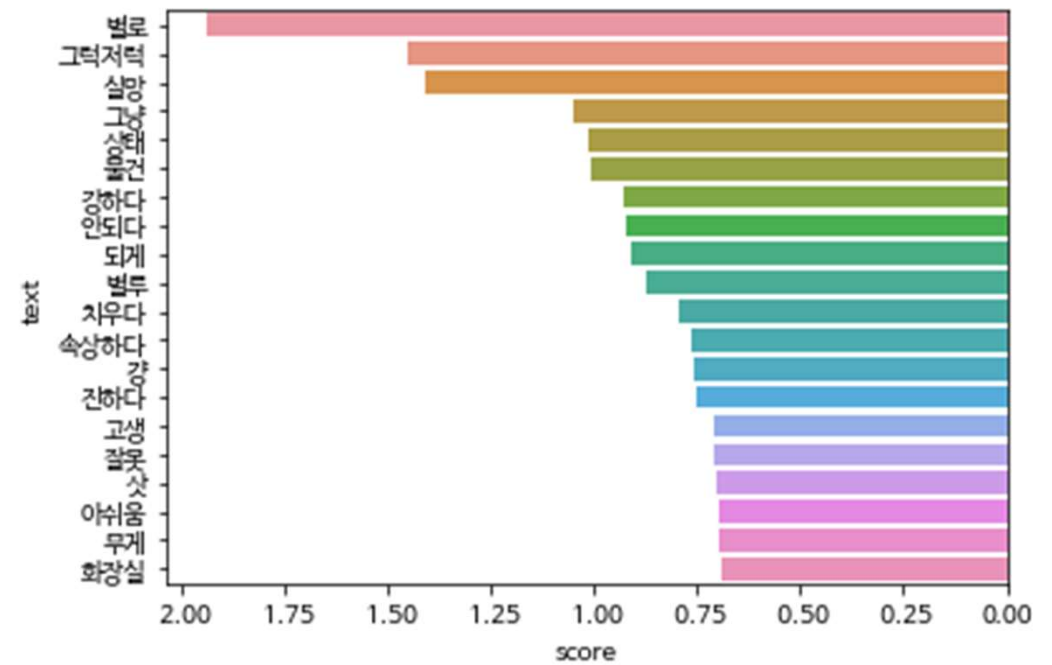
로지스틱 회귀 계수

회귀 계수가 양수인 경우 긍정을 음수인 경우 부정을 나타낸다

3. 최적 모델 선택 / 활용 분말 세제 리뷰 최적의 머신러닝 모델



〈긍정 단어〉



〈부정 단어〉

세탁 중 향기와 효과 용량 세척 등 만족하는 것을 보인다.

액체와 다르게 무게와 상태 등 또한 강한 향에 많은 아쉬움을 느끼는 것으로 보인다.

4. 활용 방안

액체 세제 활용방안



키워드	항목	추천여부
뚜껑	용이성	X
독해	용이성	X
은은한향	용이성	O
아기 / 민다	신뢰성	O
저렴하다 / 가성비	비용성	O
무겁다	용이성	X

〈액체 세제를 추천 / 비 추천 키워드〉

액체 세제의 경우 아기를 데리고 있고 가성비를 중요시 하는 고객에게 추천한다.
그렇지만 뚜껑과 향에 민감하고 무거운 것을 싫어하는 고객들에게는 추천하지 않는다.

4. 활용 방안

분말 세제 활용방안



키 워 드	항 목	추 천 여 부
용량 / 넉넉하다	용이성	O
향기 / 향	용이성	O
강하다(향)	용이성	X
고생 / 별루 / 안된다	용이성	X
무게	용이성	X
만족하다 / 제일	신뢰성	O

〈분말 세제를 추천 / 비 추천 키워드〉

분말 세제의 경우 용량과 강한 향에 민감하지 않은 고객에게 추천한다.
그렇지만 강한 향에 민감하고 무게와 편리함을 추구하는 고객들에게는 추천하지 않는다.

4. 시사점 및 한계점

시사점 및 한계점

시사점

액체 세제
아기 옷을 세탁하는데 만족하는
것으로 나타났다.
그렇기에 아이가 있는 부모를
타겟으로 홍보와 마케팅이
필요하다.

분말 세제
분말 세제를 사용함에 있어
용량과 넉넉함에 만족하는 것으로
나타났다 그래서
많이 사용하는 코인 세탁소 등 매장
및 1인 가구 위주로 타겟으로
홍보와 마케팅이 필요하다

한계점

세제 유형별 간 데이터
불균형

긍정/부정 리뷰 개수 불균형

평점을 통한 긍정/부정 리뷰
구분 한계

데이터의 품질 부족으로
인한 모델링 아쉬움

4가지 모델만 사용

사용한 Tool



Q & A

**Thank
you**