

최종 발표

클러스터링 기법을 활용한
e커머스 사용자 리뷰에 따른 시장 세분화 연구

A Study on Market Segmentation Based on E-Commerce User Reviews
Using Clustering Algorithm

2017314005 김민경
2017314020 사애진
2018316027 이한별
2018316043 전아름

목차

01 연구 개요

02 클러스터링

03 군집 분석

04 연구 결론

05 참고 문헌



연구 배경 및 목적

- 코로나로 인한 **e커머스 시장 확대** 및 인터넷 쇼핑물 이용률 증가
- **고객 리뷰**가 소비자 구매 의사결정에 미치는 영향력 증대
- 다양한 형태의 고객 리뷰어들에 대한 **고객 분석 및 관리** 필요



연구 단계



연구 단계 <데이터 수집>

eun***** TOP500 랭킹 180 도움 수 211 작성한 후기 281

★★★★★ 5.0
163/56/ 66free 지금 입기 좋아요. 바람 송송 ...
정상가 : 99800원 / 베이직 언발 ...

★★★★★ 5.0
이 옷은 시스루예요. 티가 굉장히 많이 나니까...
소인더를 마이추 소프트 긴팔티

★★★★★ 5.0
163/56 66free 다 잘 맞는 평범한 몸이에요 이...
소인더를 워를 아가일 배색조끼

★★★★★ 5.0
오트밀색 사고 싶었는데 편찮네용 작지도 크지...
소인더를 햄스베어 프린팅 맨투맨

★★★★★ 5.0
옷 색깔은 실명이랑 같고 셔링도 잘 잡혀 있는데...
소인더를 헬렌 어 캐서링 티셔츠

★★★★★ 5.0
이렇게 글쭙 물랐어요.ㅎ 저는 이런 핏 좋아...
소인더를 스탠드 칼라발포 맨투맨

eun ***** TOP500
★★★★★ 5.0 | 2021.09.11 | 신고
정상가 : 99800원 / 베이직 언발 투웨이 셔츠, BL(41), F1
163.56/ 66free
지금 입기 좋아요.
바람 송송 통하고
사실 생각했던건 아닌데 쉬폰같은 재질 이예요.
그래서 비침 있어요.
색상은 작샹중 써서입은 사진 이랑 비슷하고용
중간에 단추3개가 사선으로 붙어있어요
3개중 어느곳이든 바지에 넣어입어도 써서 입어도 예쁠니다
저 처럼 혹정에 입으셔도 되고
와이드슬렉스에 입으시면 예쁠거 같아요
2명에게 도움이 되었습니다. 😊 도움 됨 😞 도움 안 됨

출처 : 위메프

수집 요소

- 사용자 이름
- 랭킹
- 전체 도움이 돼요 수
- 작성 리뷰 수
- 리뷰 상품명
- 리뷰 평점
- 리뷰 작성 날짜
- 이미지 여부
- 리뷰 내용
- 각 리뷰의 도움이 돼요 수

연구 단계 <전처리>

기준	변수명		
감성	긍정 리뷰 비율	분기	2020년 1월 - 2020년 3월 작성 비율
	부정 리뷰 비율		2020년 4월 - 2020년 6월 작성 비율
	중립 리뷰 비율		2020년 7월 - 2020년 9월 작성 비율
	전체 리뷰 긍정 비율		2020년 10월 - 2020년 12월 작성 비율
	전체 리뷰 부정 비율		2021년 1월 - 2021년 3월 작성 비율
	전체 리뷰 중립 비율	계절	3월 - 5월 작성 비율
	내용 없는 리뷰 비율		6월 - 8월 작성 비율
	평점 - 댓글 불일치 비율		9월 - 12월 작성 비율
			12월 - 2월 작성 비율
별점	별점 1 비율	작성주기	중복 날짜 제거 작성 비율
	별점 2 비율		작성 주기 평균
	별점 3 비율		리뷰 작성하지 않은 기간
	별점 4 비율	기타 사용자 정보	리뷰어 랭킹 점수
	별점 5 비율		전체 리뷰 개수
	별점 평균		'도움이 돼요' 받은 리뷰 비율
리뷰 내용	이미지 첨부 리뷰 비율		총 '도움이 돼요' 개수
	리뷰 길이 평균		

32개 입력변수 전처리 및 정규화

연구 단계 <전처리>

텍스트 마이닝(감성분석)

lee*****

★★★★★ 5.0 | 2021.04.30 | 신고

반려견타입 중형견

강아지이름 (기본사이즈) 반려동물인식표, 강아지인식표(기본사이즈) 푸들(블랙), 형광다홍

눈에 확 띄네요
 고급재 이라고 하셨는데 가격에 비해 품질소재질을 사용하신 느낌이랄까요? 더
 좋았을 것 같아요.
 아담한 사이즈 포장이 쫄지도 보송주사워 너무 감동!
 항상 산책 다닐때 기본에 쓰고있는데 인식표가 가슴줄 사이드에서 다보니 이름을 물어보셨는
 데 등에 딱 매달아주면 확실히 선명해서 너무 좋을것같아요
 아무래도 검둥이라ㅋㅋㅋ 밝은 나시 하나 입혀야 할것같네요ㅋㅋㅋ 재구매 의사있어요~~~

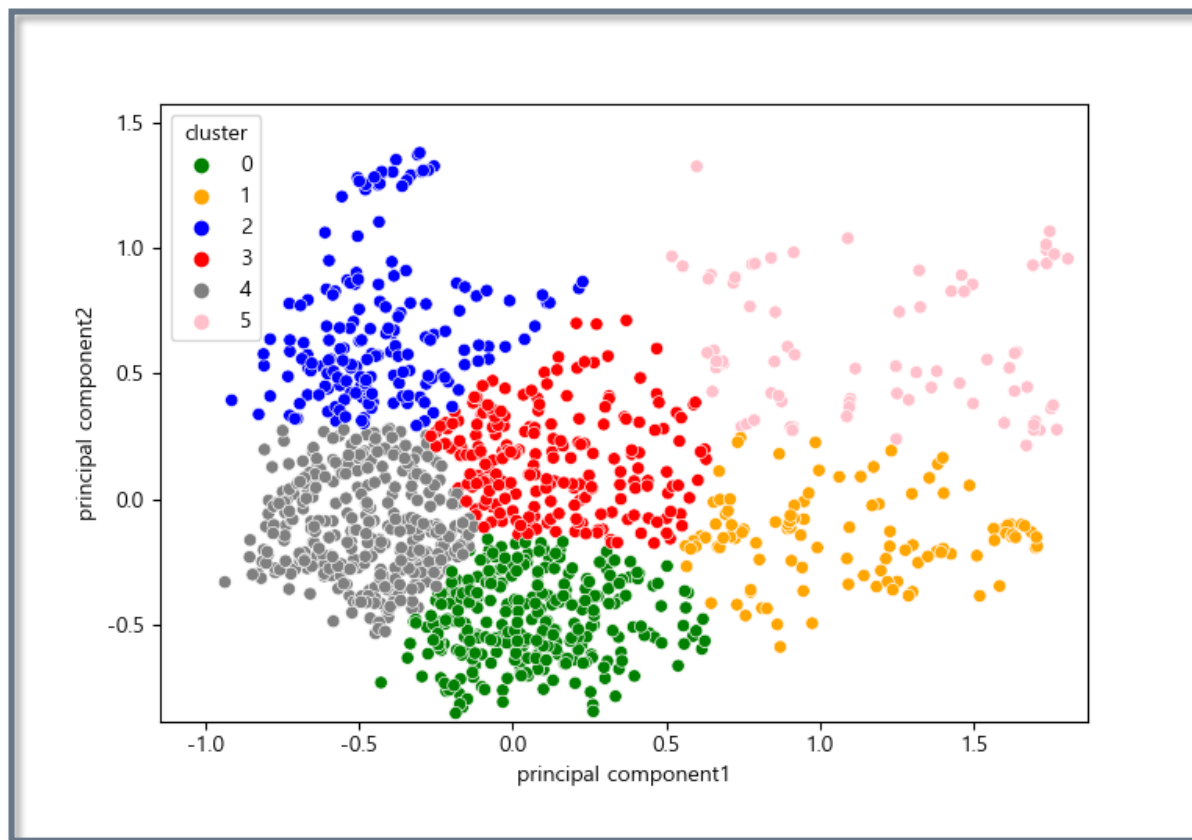


[감성분석 과정]

1. '리뷰내용' 칼럼 사용
2. 데이터 정제
 - 특수문자 및 불용어 제거, 형태소 분리 등
3. KOSAC 감성사전 변형 및 구축
4. 감정 사전을 통한 각 리뷰의 긍/부정 여부 판단

K-means 클러스터링 기법을 이용한 군집화 진행

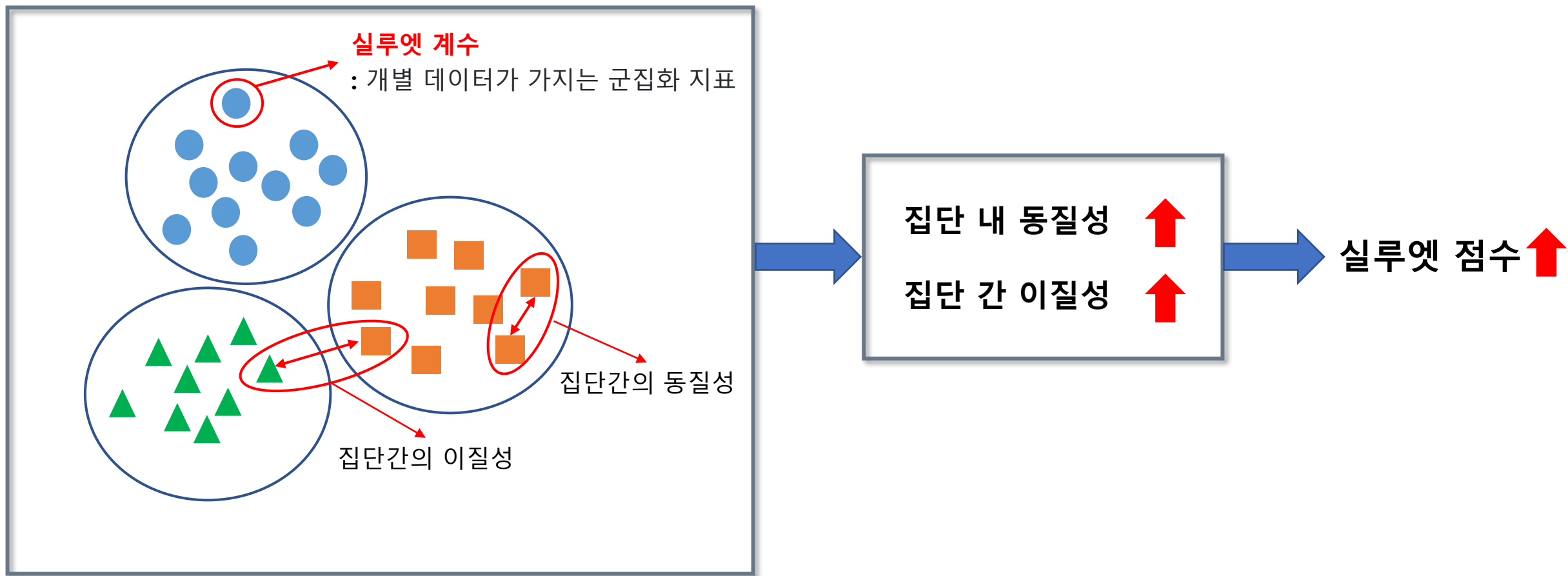
모델 시각화 및 성능향상을 위해 **PCA (차원 축소)**, 입력변수 간 **상관계수 분석**



클러스터링 2차원 시각화 그래프

실루엣 점수*

: 클러스터링 성능 평가 지표로 각 군집 간의 거리가 얼마나 효율적인지를 나타내는 점수 (-1.0 ~ 1.0)



* 참고문헌 : 김성수, 백준영, 강범수, (2017), "실루엣을 적용한 그룹탐색 최적화 데이터클러스터링", 한국경영과학회지, pp.25-34

모델 성능 향상을 위한 입력변수 조정

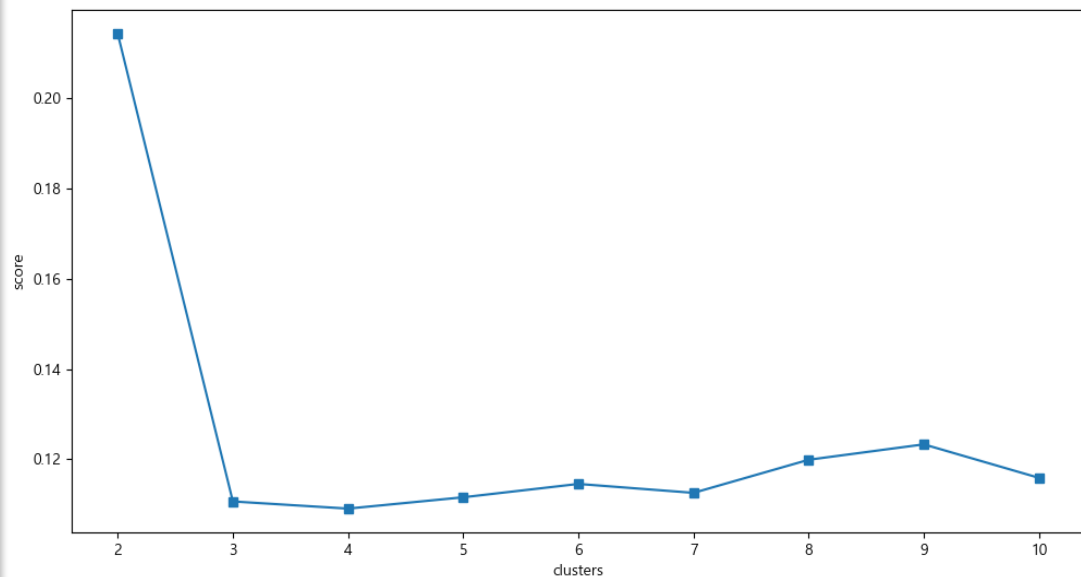
분기와 계절에
대한 입력변수

기준	변수명		
감성	긍정 리뷰 비율	분기	2020년 1월 - 2020년 3월 작성 비율
	부정 리뷰 비율		2020년 4월 - 2020년 6월 작성 비율
	중립 리뷰 비율		2020년 7월 - 2020년 9월 작성 비율
	전체 리뷰 긍정 비율		2020년 10월 - 2020년 12월 작성 비율
	전체 리뷰 부정 비율		2021년 1월 - 2021년 3월 작성 비율
	전체 리뷰 중립 비율	계절	3월 - 5월 작성 비율
	내용 없는 리뷰 비율		6월 - 8월 작성 비율
	평점 - 댓글 불일치 비율		9월 - 12월 작성 비율
			12월 - 2월 작성 비율
별점	별점 1 비율	작성주기	중복 날짜 제거 작성 비율
	별점 2 비율		작성 주기 평균
	별점 3 비율		리뷰 작성하지 않은 기간
	별점 4 비율	기타 사용자 정보	리뷰어 랭킹 점수
	별점 5 비율		전체 리뷰 개수
	별점 평균		'도움이 돼요' 받은 리뷰 비율
리뷰 내용	이미지 첨부 리뷰 비율		총 '도움이 돼요' 개수
	리뷰 길이 평균		

→ 제거

모델 성능 향상을 위한 입력변수 조정

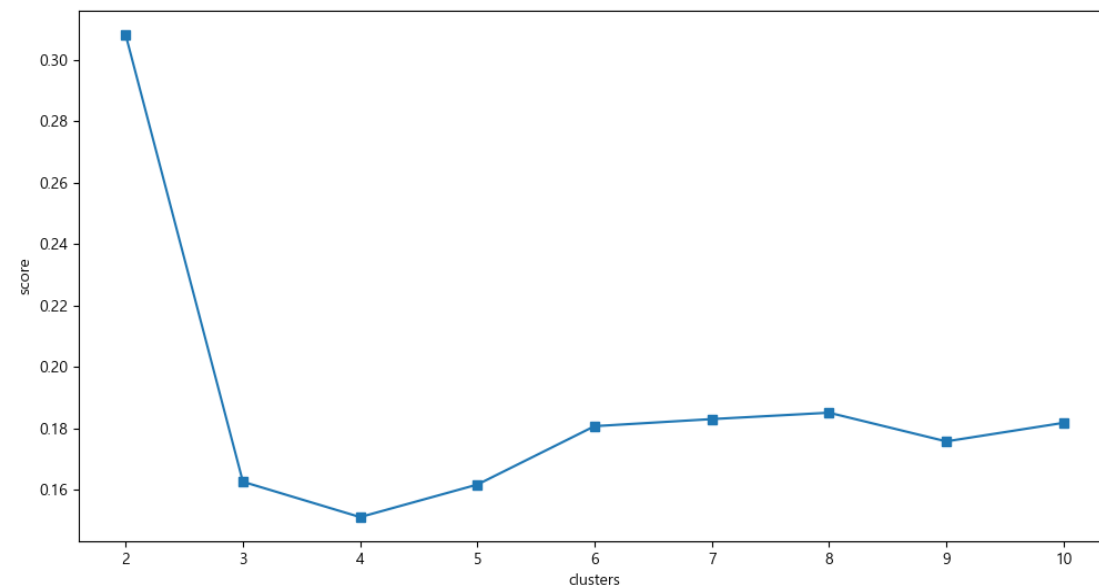
32개의 변수 모델



실루엣 점수
모두 0.5이상 증가

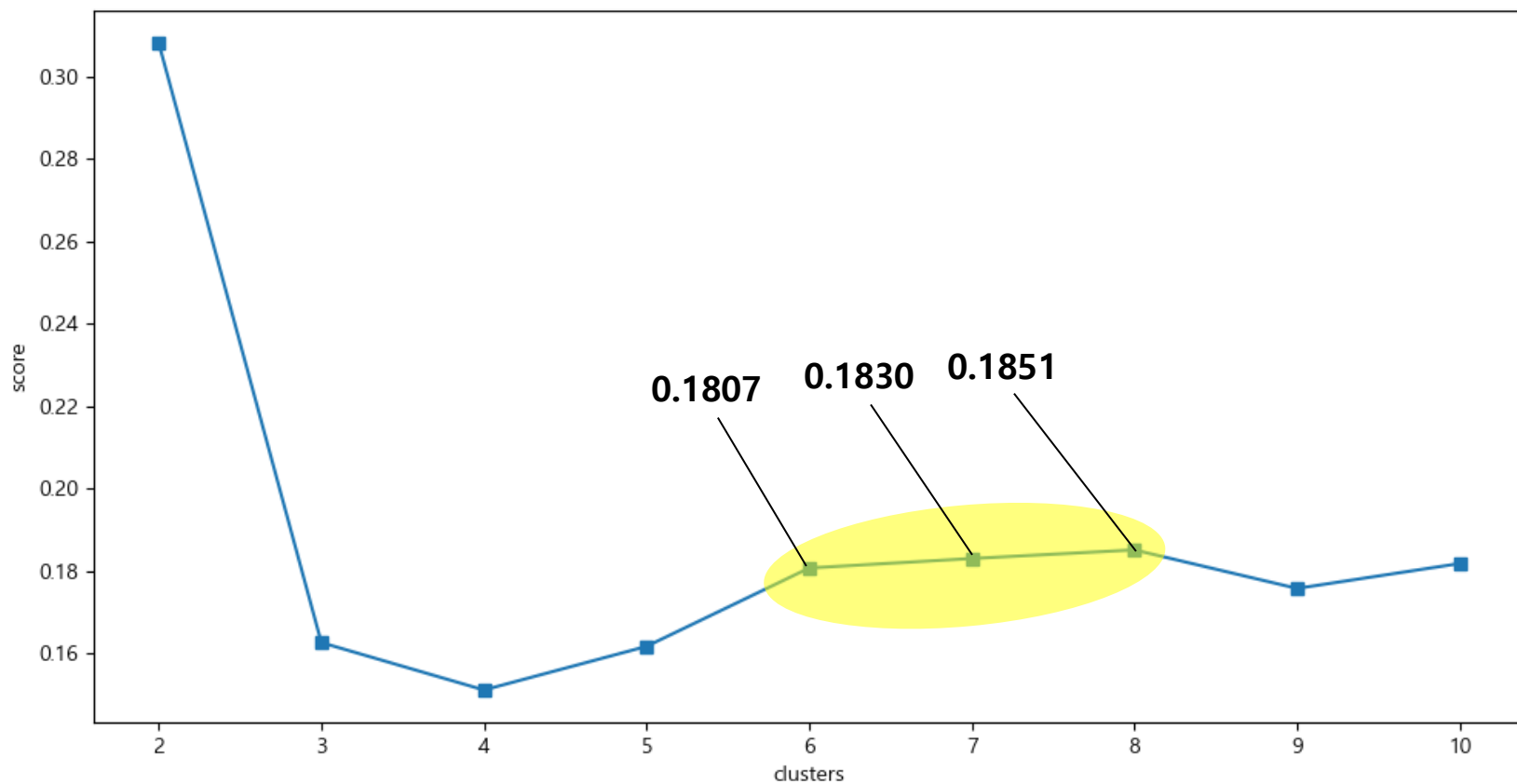


23개의 변수 모델
(분기와 계절 입력변수 제거)



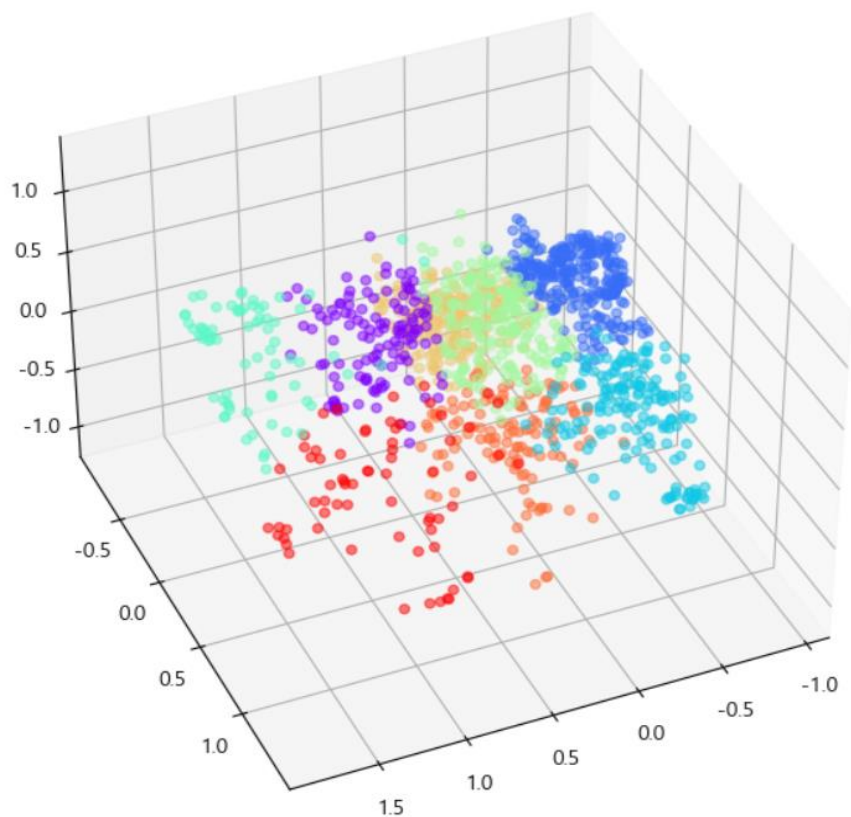
최종모델 선정: 실루엣 점수 비교 분석

23개의 변수 모델(분기와 계절 입력변수 제거)



최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 8



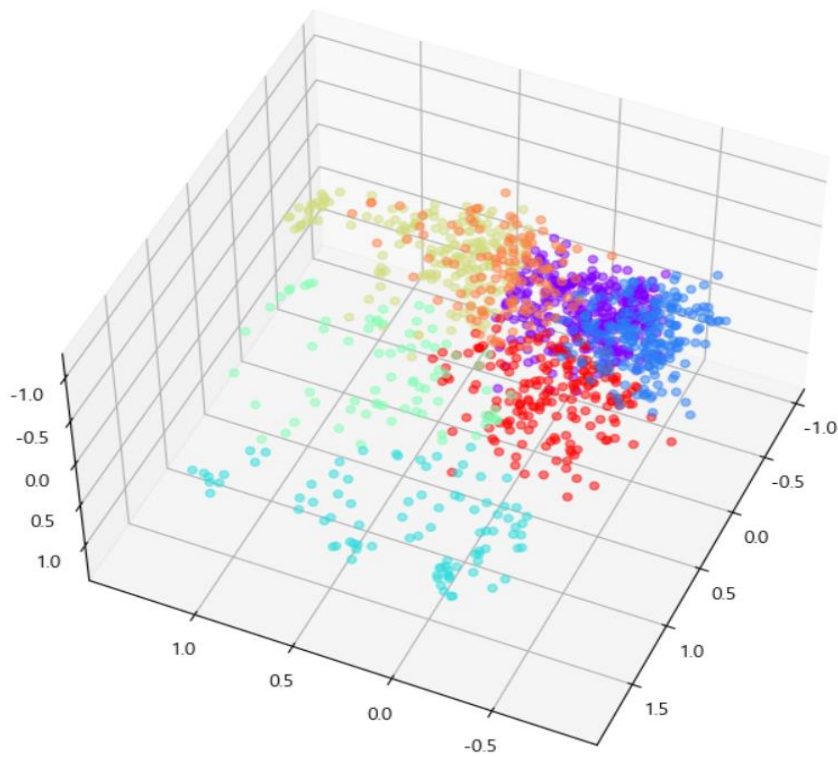
[모델 내 군집 별 실루엣 점수]

1	0.265
2	0.404
3	0.392
4	0.445
5	0.188
6	0.382
7	0.307
8	0.150

군집5과 군집8의 실루엣 점수가 현저히 낮음

최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 7



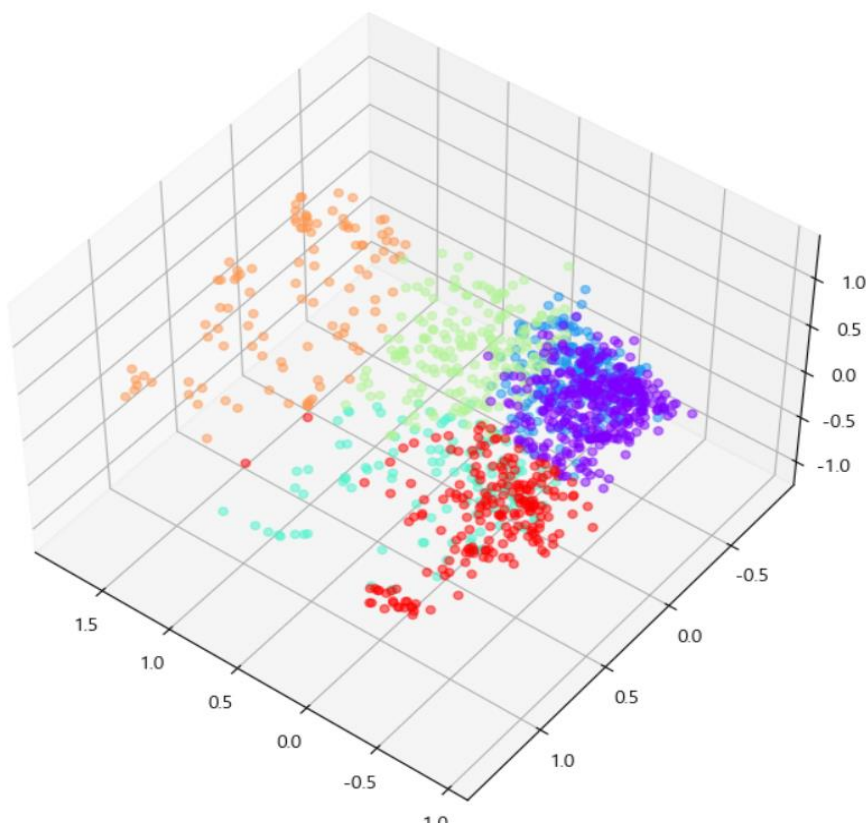
[모델 내 군집 별 실루엣 점수]

1	0.405
2	0.387
3	0.400
4	0.189
5	0.356
6	0.314
7	0.213

군집4과 다른 군집들 간의 실루엣 점수 차이가 크며
현저히 낮은 점수를 가짐

최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 6



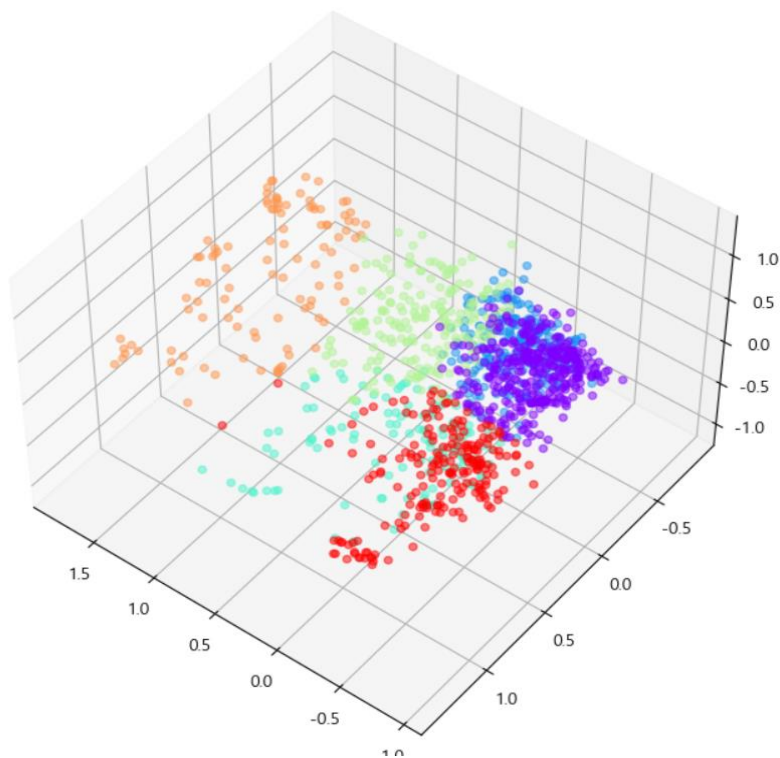
[모델 내 군집 별 실루엣 점수]

1	0.407
2	0.394
3	0.235
4	0.202
5	0.392
6	0.340

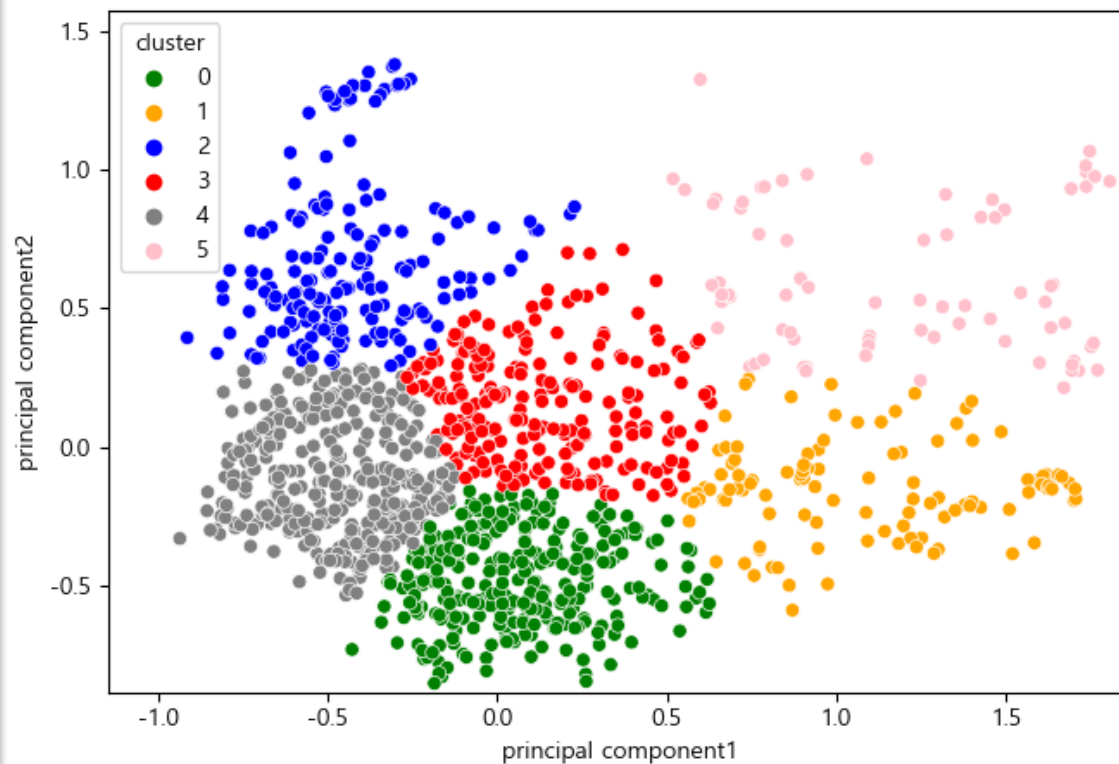
군집 label 별 비교적 비슷한 실루엣 점수
군집 개수 6,7,8개 중 가장 높은 실루엣 점수 가짐

최종 모델

PCA = 3, Clustering = 6
실루엣 점수: 0.18



PCA = 2, Clustering = 6



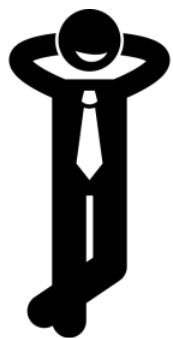
군집 분석: 입력변수 별 각 군집 기술통계 비교분석

	군집 1	군집 2	군집 3	군집 4	군집 5	군집 6
리뷰어 랭킹 점수				값이 0	값이 0	가장 높음
도움이 돼요 수			0 값이 많음		가장 낮음	가장 높음
별점 1 비율	가장 낮음			값 골고루 분포	가장 높음 (격차 큼)	
별점 2 비율				값 골고루 분포	가장 높음	가장 낮음
별점 3 비율		가장 높음 (골고루)			평균 대비 1 많음	가장 낮음
별점 4 비율		높은 편	가장 높음		가장 낮음	
별점 5 비율	높은 편				가장 낮음	가장 높음
별점 평균	높은 편	높은 편	높은 편	높은 편	가장 낮음	가장 높음
중복 날짜 제거 작성 기간		가장 낮음	높은 편	높은 편	가장 높음 (격차 큼)	높은 편
작성 주기 평균				가장 높음	가장 낮음	
리뷰 작성하지 않은 기간	가장 낮음				가장 높음 (골고루)	
이미지 댓글 비율	낮은 편	가장 낮음	높은 편	낮은 편		가장 높음 (격차 큼)
리뷰 길이 평균	가장 낮음	낮은 편				가장 높음 (비슷함)
리뷰어의 전체 댓글 수		가장 높음 (비슷함)			가장 낮음	
도움이 돼요 수 받은 댓글		낮은 편	높은 편	낮은 편	높은 편	가장 높음
None 리뷰 비율	높은 편	높은 편	낮은 편	높은 편	값이 0	낮은 편
긍정 댓글 비율	높은 편	높은 편			가장 낮음	가장 높음 (격차 큼)
부정 댓글 비율	가장 낮음		낮은 편	높은 편	가장 높음	낮은 편
중립 댓글 비율				가장 높음 (비슷함)		가장 낮음 (비슷함)
전체 긍정 비율	높은 편	높은 편	높은 편		가장 낮음	가장 높음
전체 부정 비율	낮은 편	낮은 편			가장 높음 (격차 큼)	낮은 편
전체 중립 비율						가장 낮음 (격차 큼)
평점 댓글 불일치	낮은 편		높은 편	가장 높음		낮은 편

시장세분화



군집 1



무던한 귀차니즘 고객

대충대충 좋게 별점만 !

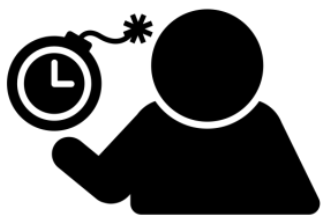
리뷰 특징

- 별점 5 점만을 많이 부여
- 리뷰 길이 짧음
- None 리뷰 비율 높음
- 도움이 돼요 수 낮음
- 반복되고 성의 없는 리뷰 많음

마케팅 방안

- 리뷰 자동완성 기능 제공
- 리뷰 10글자 이상 글자 수 제한

군집 2



마감임박 베틀치기 고객

빨리빨리 리뷰 복붙 !

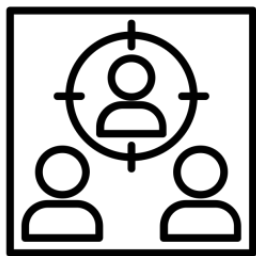
리뷰 특징

- **별점 2,3,4** 점을 다양하게 부여함
- 이미지 댓글 비율 낮음
- **리뷰를 몰아서 작성함**

마케팅 방안

- 리뷰 작성 기간 세분화 및 **기간에 따른 포인트 차등 지급**
- 리뷰 **이미지 유/무**에 따른 포인트 차등 지급

군집 3



미리 잡아 잠재 고객

좋은데 별점 5점은
안줄래 !

리뷰 특징

- 별점 3,4점을 많이 부여함
- 별점 5점에 인색함
- 도움이 돼요 받은 댓글 비율 높은 편
- 이미지 리뷰 많은 편

마케팅 방안

- AI 기반 상품 추천 시스템을 통한 1인 맞춤형 마케팅

군집 4



논리정연 스마트 고객

내 주관대로 꼼꼼히 열심히 !

리뷰 특징

- 긍정/부정 댓글 골고루 작성
- 상품 구매에 대한 주관 뚜렷
- 별점 1~5점 다양
- 작성주기 높음

마케팅 방안

- 제품 체험단으로 선정하여 충성고객으로의 유도
- 우수 체험단 선정 후 추가 보상 및 혜택 제공

군집 5



주의경보 이탈 고객

상품 구매가 실망스러워 !

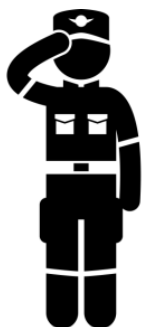
리뷰 특징

- 별점 1,2점을 많이 부여함
- 부정 댓글 작성 비율 높음
- 리뷰작성 자주 안함
- 중복날짜 가장 높음
- 전체 리뷰 수 적음
- 리뷰를 1개만 작성한 사람이 많음

마케팅 방안

- VOC(Voice of Customer)를 통한 실시간 피드백 제공
- 최근 1년 미구매 고객 대상 할인쿠폰 및 포인트 지급

군집 6



한결같은 충성 고객

상품도 서비스도 좋아 ! 충성 !

리뷰 특징

- 별점 5점을 많이 부여함
- 긍정 댓글 작성 비율 높음
- 리뷰어 랭킹 달성 고객 많음
- 전체리뷰 수 많음
- 도움이 돼요 수 많음
- 이미지 댓글 비율 높음

마케팅 방안

- 멤버십 서비스 혜택 제공
- 고객 등급 관리제 도입 및
매월 등급에 따른 혜택 제공

✓ 연구 결과

- 이커머스 시장의 리뷰고객을 총 6개의 군집으로 세분화할 수 있음
- 시장세분화 결과를 통해 고객 관리 방안을 제시함

✓ 기대효과 및 시장성

- e커머스 시장의 고객 유형 파악 및 고객 관리 용이
- 잠재 고객 발굴을 통한 맞춤형 고객 관리 서비스 제공
- 고객 군집별 타겟 마케팅을 통한 기업의 마케팅 비용 절감 및 수익 창출

강진구, 이기성, (2019), "텍스트마이닝(Text Mining)을 통해 본 제주 예멘 난민 -네이버 뉴스 댓글을 중심으로", 다문화콘텐츠연구, pp.103-135.

김담희, 안가경, (2018), "머신러닝을 이용한 고객세분화에 관한 연구", 융복합지식학회논문지, 6(2), 융복합지식학회, pp.115-120.

김성수, 백준영, 강범수, (2017), "실루엣을 적용한 그룹탐색 최적화 데이터클러스터링", 한국경영과학회지, pp.25-34

김수연, 정유경, 송민, (2015), "한글 감성 분류를 위한 감성 사전 구축에 관한 실험적 연구", 한국도서관정보학회 동계 학술발표회, pp.143-150.

김문형, 장하연, 조유미, 신호필, (2013), "KOSAC(Korean Sentiment Analysis Corpus): 한국어 감정 및 의견 분석 코퍼스", 한국정보과학회, pp.650-652.

김종원, 최형준, (2021), "군집분석을 통한 K리그 축구팀 플레이스타일 분류", 한국체육측정평가학회지, 23(1), pp.1-9.

이민철, 윤현식, (2020), "머신러닝을 활용한 가짜리뷰 탐지 연구 : 사용자 행동 분석을 중심으로", 지식경영연구, 21(3), pp.177-195.

이소영, 인관호, (2013), "오피니언 마이닝을 이용한 인터넷 쇼핑몰의 화장품 상품평 분석", 한국정보과학회, pp. 362-364.

이찬도, (2020), "문화관광형시장의 고객 소비형태별 군집분석과 시장 세분화 전략 - 고양시 일산 및 원당 시장을 중심으로 -", 문화산업연구, 20(4), pp.87-96.

이태규, 이경호, (2018), "Filter Method와 Classification 알고리즘을 이용한 전자상거래 블랙컨슈머 탐지에 대한 연구", 정보보호학회논문지, pp.1499-1508.

장하원, 이수번, (2021), "레스토랑의 서빙로봇 속성 평가에 따른 고객 세분화", 호텔경영학연구, 30(1), pp.49-63.

최경빈, 남기환, (2019), "쇼핑 웹사이트 탐색 유형과 방문 패턴 분석", 지능정보연구, 한국지능정보시스템학회, 25(1), pp.85-107.

최자영, 김현아, 김용범, (2020), "온라인 리뷰가 매출에 미치는 영향력 분석: 텍스트기반 감성지수를 중심으로", 한국유통학회, 流通研究 25(3), pp.1-21.

감사합니다