

2022년도 종합 설계 프로젝트 최종 발표

크라우드펀딩제품 분석을 통한 성공 예측 및 유사도 기반 비교 서비스

담당 교수님: 배혜림 교수님

산업공학과	201627537	이재헌
산업공학과	201727508	김민규
산업공학과	201727534	이가원
산업공학과	201727539	이찬영

CONTENTS

01. 주제소개

- 주제 요약
- 기존 연구와 차별성

02. 분석내용

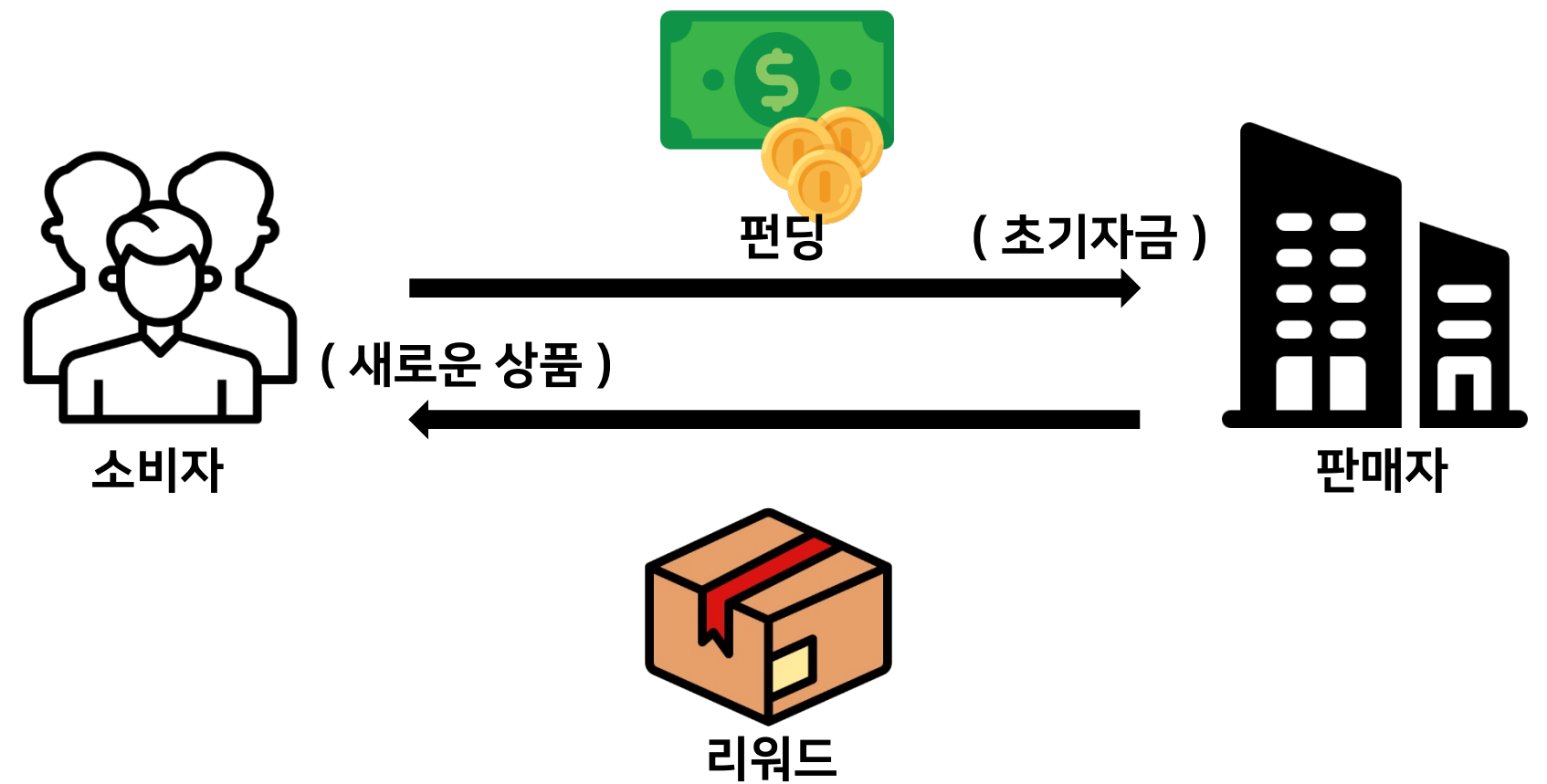
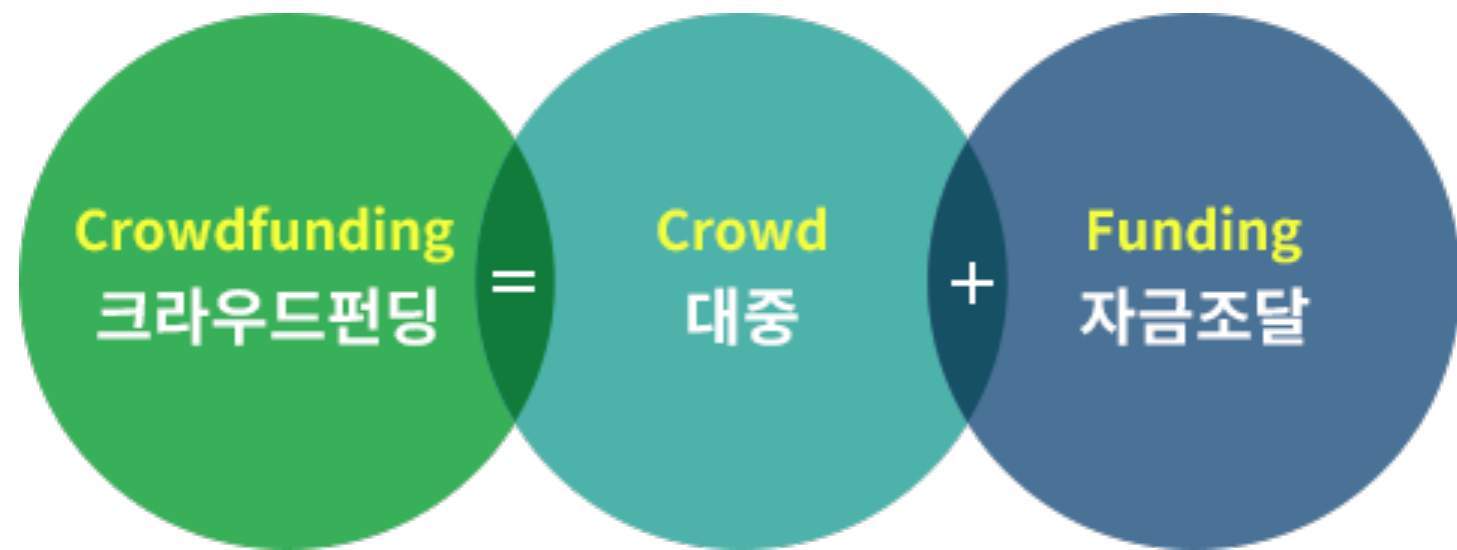
- 웹 크롤링
- Modeling
- Word Embedding
- K-Means

03. 결과

- 분석 결과 정리
- 제공 서비스 예시
- 확장성

크라우드 펀딩이란?

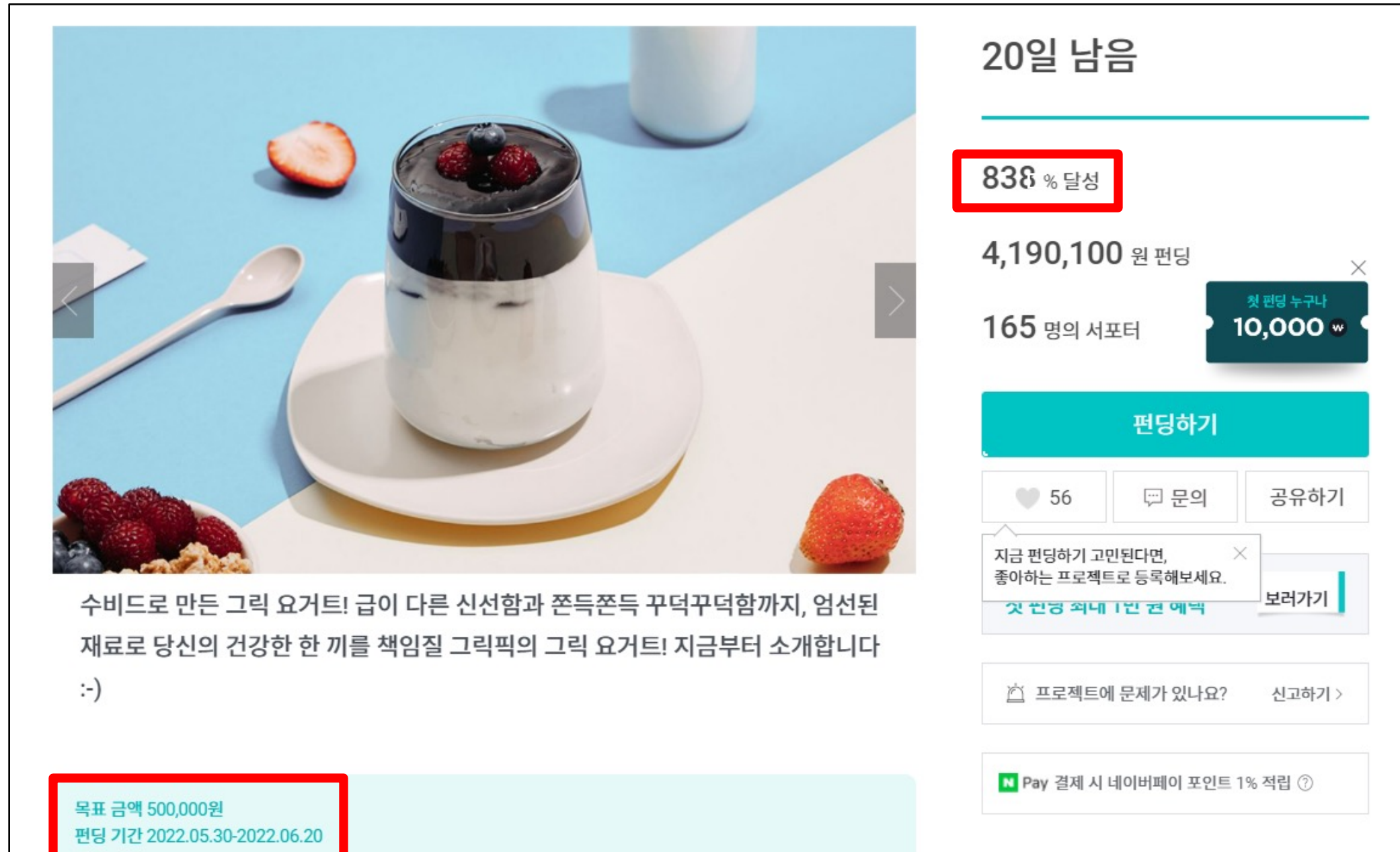
주제 소개
01



온라인 플랫폼을 이용해 다수의 대중으로부터 자금을 조달하는 방식

크라우드 펀딩이란?

주제 소개
01

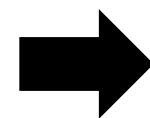


The screenshot shows a crowdfunding campaign for a yogurt product. The main image is a glass of yogurt topped with raspberries and blueberries. The text on the page includes:

- 20일 남음
- 83% 달성 (highlighted with a red box)
- 4,190,100 원 펀딩
- 165 명의 서포터
- 첫 펀딩 누구나 10,000 ₩
- 펀딩하기
- 56
- 문의
- 공유하기
- 지금 펀딩하기 고민된다면, 좋아하는 프로젝트로 등록해보세요.
- 첫 펀딩 최대 1만 원 혜택
- 보러가기
- 프로젝트에 문제가 있나요? 신고하기 >
- Pay 결제 시 네이버페이 포인트 1% 적립
- 목표 금액 500,000원 (highlighted with a red box)
- 펀딩 기간 2022.05.30-2022.06.20

< 와디즈 크라우드 펀딩 플랫폼 >

펀딩 총 금액 > 목표금액



성공



프로젝트 **성공 여부**를 **예측**할 수 있다면
소비자와 판매자의 **기회비용**을 줄여줄 수 있지 않을까?

기존 연구와 차별성

주제 소개
01

	기존 연구	본 연구
연구 내용	소비자 반응 분석	제품 정보 분석
예시	커뮤니티 댓글 수, 서포터 수, 새 소식 수, 페이스북 공유 건수	제품 카테고리, 목표금액, 리워드 옵션 개수, 제품 정보

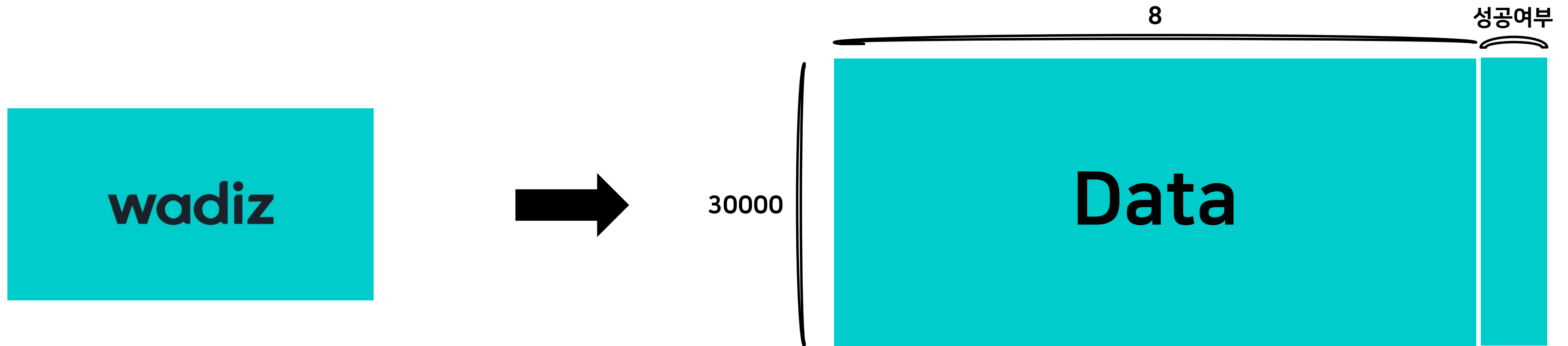
기존 연구의 **한계점** : 펀딩이 종료된 시점에서 제품에 대한 **소비자 반응**을 분석. **결과론적인 분석**

본 연구의 **차별점** : 소비자의 반응에 대한 정보를 제외함. **제품 정보**를 분석해 유의미한 정보 제공

Appendix 1 참고

데이터 수집

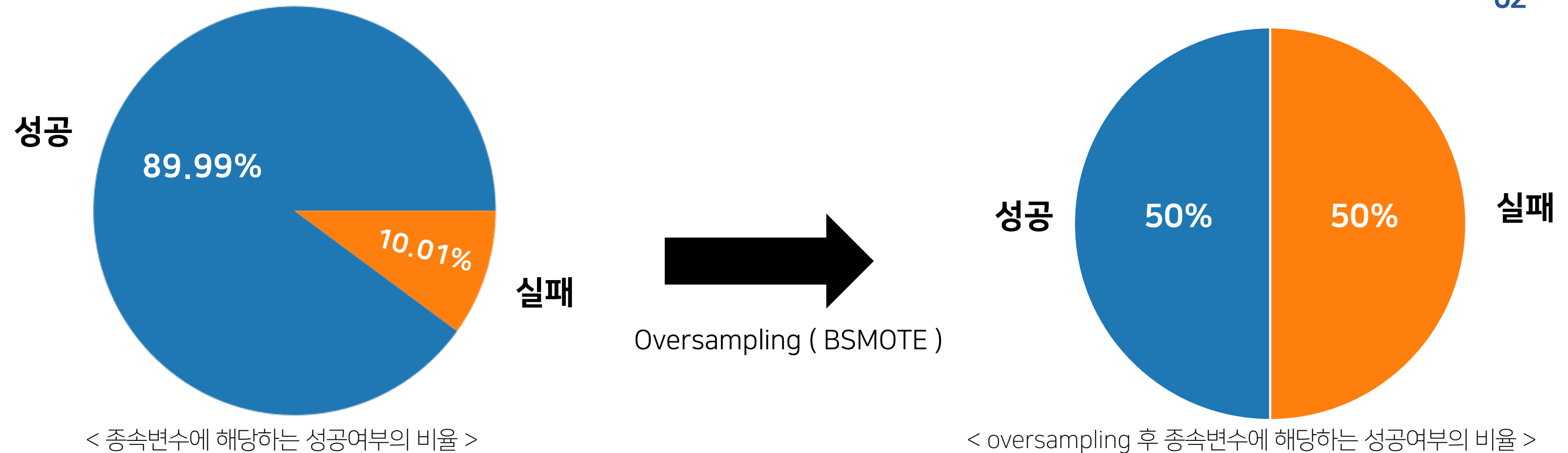
진행 과정
02



- 웹 크롤링 -> 30000개의 데이터 수집
- 수집한 정보 : 제품 제목, 제품 소개글, 펀딩 기간, 카테고리, 상세 카테고리, 본문 이미지개수, 제목 이미지 개수, 리워드 옵션 개수, 제품 가격, 목표금액 (Appendix 2 참고)
- 예측에 사용된 피처 : 카테고리, 상세 카테고리, 리워드 옵션 개수, 제품 가격, 목표금액, 목표 판매 개수, 제목 이미지 개수, 리워드 옵션 개수

Modeling

진행 과정
02



모델링 결과 (Appendix 3 참고)

Accuracy	Precision	Recall	F1_score
0.9658	0.9747	0.9866	0.9806

* MLPClassifier 로 train set : test set = 7 : 3 , 학습 예측한 결과

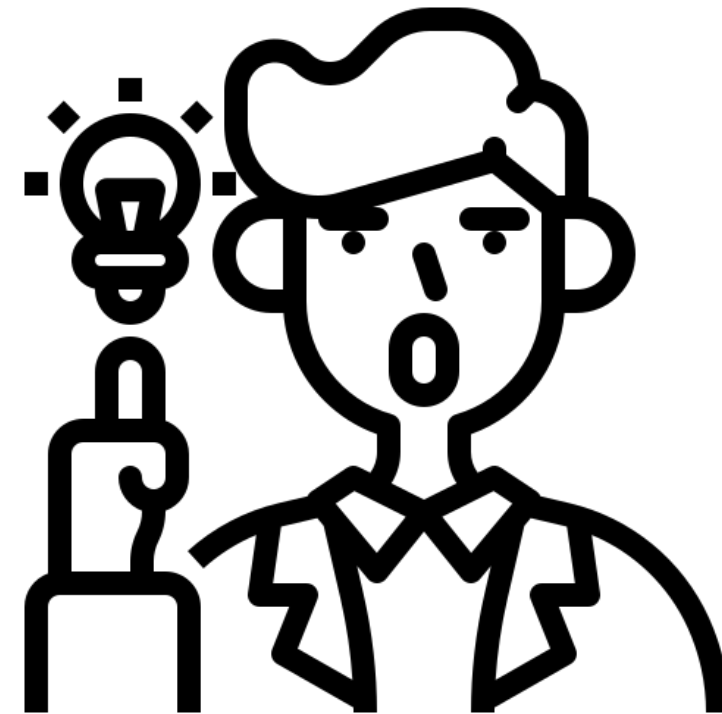
모델링 결과 (Appendix 3 참고)

Accuracy	Precision	Recall	F1_score
0.9811	0.9831	0.9791	0.9811

* LGBMClassifier 로 train set : test set = 7 : 3 , 학습 예측한 결과

Modeling

진행 과정
02



성공 예측만으로는 부족하다고 판단

예시: 실패할 것 같다고 그래서?

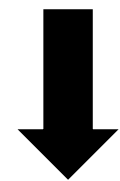
성공 예측 + 유사도 기반 비교 서비스 제공

예시: 유사도 기준 상위 10개의 제품들에 대한 정보를 제공

자연어 처리

진행 과정
02

근성장을 위해 섭취하세요! 백백하지 않은 닭가슴살, 정말 맛있습니다.



근성장, 섭취, 백백, 닭, 닭가슴살, 맛있



word	1	2	3	4	5	6
근성장	0.31	-0.21	0.13	0.55	0.39	-1.97
섭취	1.2	0.3	2.3	-9.1	1.2	0.9
백백	0.1	-0.8	0.1	0.5	0.3	-1.9
닭	0.33	-0.22	0.16	0.52	0.31	-1.92
닭가슴살	0.32	-0.23	0.11	0.52	0.37	-1.95
맛있	1.0	-0.18	1.9	-2.3	1.8	0.2

- 제목과 본문을 사용하기 위해서는 벡터화시킬 필요
(Stanza package를 이용해 키워드 추출)

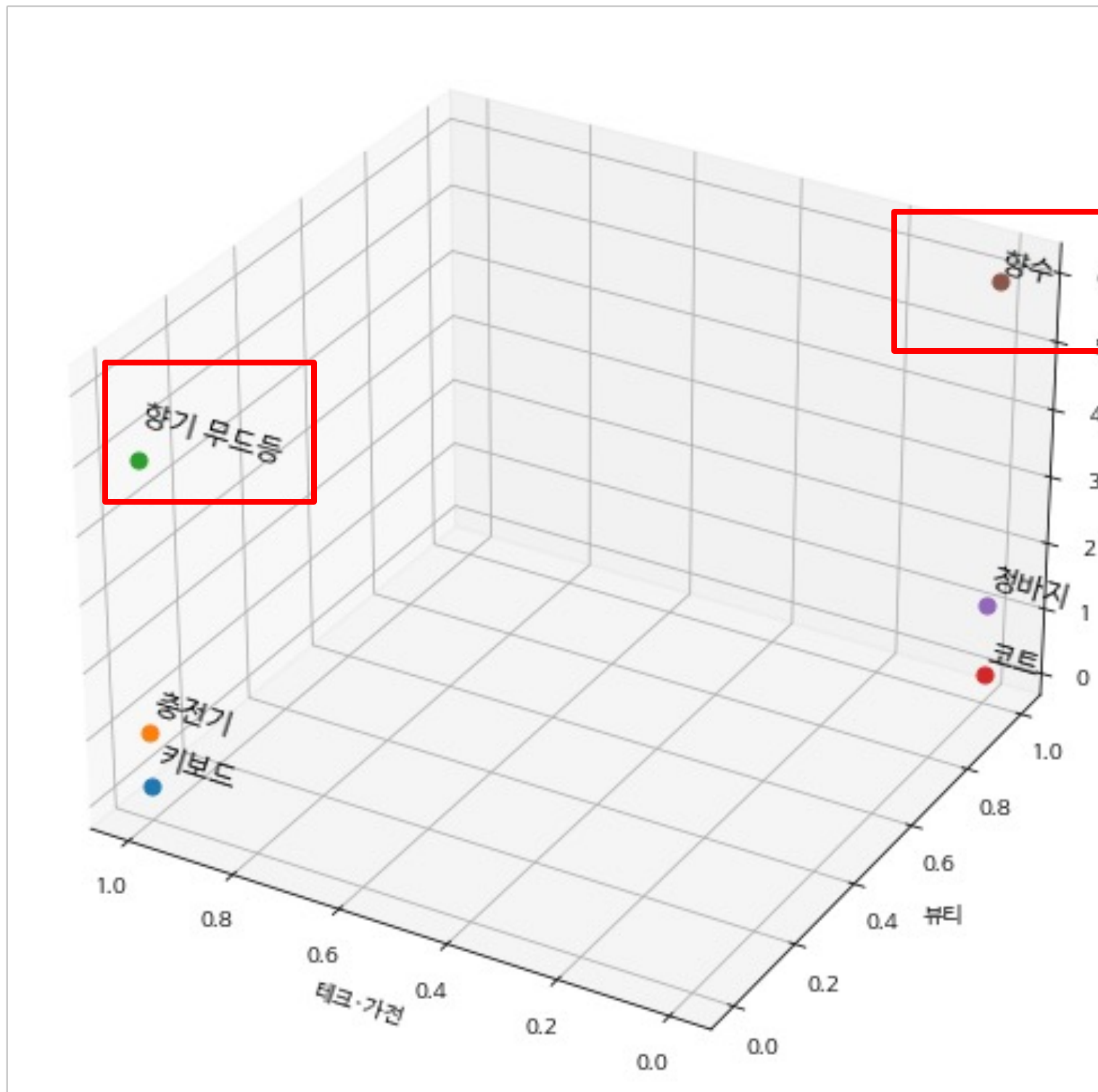
- Bert model을 사용하여 word embedding 진행

- 구글에서 개발, 배포한 사전 훈련 언어 모델
Word2Vec, Glove, Fasttext 방식보다 더 성능이 뛰어나다고 알려짐

* 임의로 설정한 값입니다.

K-means clustering

진행 과정
02



< 클러스터링 예시 >

- 예측에 사용된 피쳐 + 제목 본문 벡터화 -> 클러스터링

- 다른 카테고리더라도,

제품의 또 다른 특성을 고려해 같은 군집으로 속할 수 있음

예시: 향기 무드등 \subset (테크,가전), 향수 \subset (뷰티)

향기 무드등과 향수가 같은 그룹에 속할 가능성 존재

- 군집을 통해 정보제공

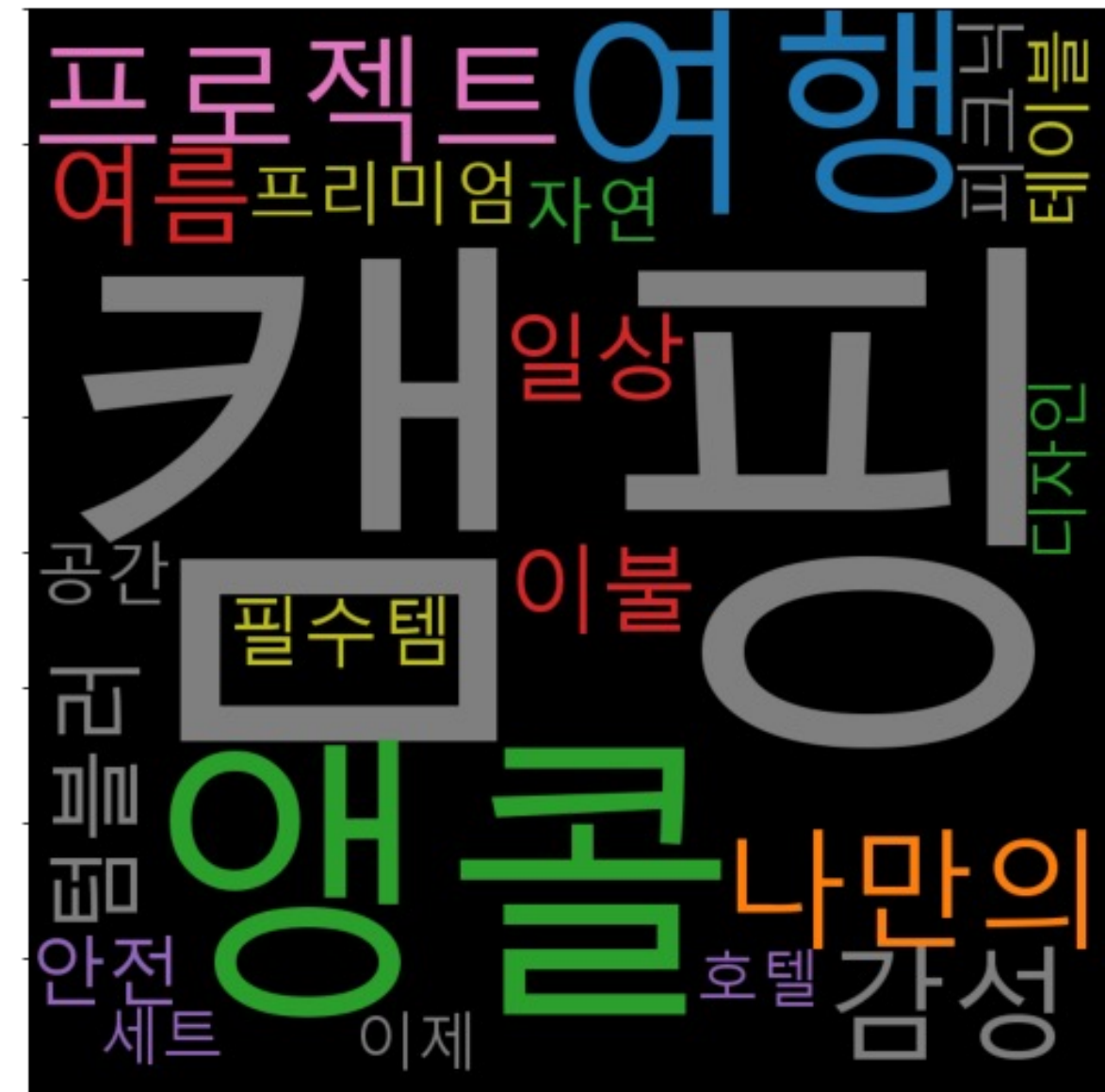
군집의 수는 17개로 설정, 이는 Silhouette_score를 참고하여 설정한 값

또한 군집내 데이터가 어느 정도 고르게 분포하도록 고려

코사인 유사도로 유사도 계산. 이를 이용해 군집내 제품 정렬

Word Cloud

진행 과정
02



< 홈 리빙 카테고리 군집 결과 >

- 왼쪽 군집은 청결에 대한, 오른쪽 군집은 나들이에 대한 군집이라고 판단가능
- '앵콜'은 인기있는 제품 펀딩 재진행 → 해당 군집에 성공적인 펀딩 多
(추가적인 결과는 Appendix 4 참고)

분석결과

03

2. 군집화를 통한 정보제공

Wordcloud를 통해 목표에 맞는 군집화 결과임을 확인 가능
코사인 유사도를 기반으로 상위 n개 추출 가능

* LGBMClassifier 로 train set : test set = 7 : 3 , 학습 예측한 결과



최종 제공 서비스

분석결과
03

첫 번째 창

펀딩 오픈 예정 상품의 URL을 입력해주세요.

`https://www.wadiz.kr/web/wcomingsoon/rwd/144479?acid=10017093&_refer_section_st=RCP_0`

성공확률은 ~~%입니다.
※ 유사도가 높은 상위 10개 제품에 대한 정보입니다.

1. 펀딩 성공한 제품의 평균 판매 개수

2. 펀딩 성공한 제품의 평균 판매 금액

3. 키워드 추천

4. 상위 10개 제품 image 및 URL

⋮

두 번째 창

감사합니다.

QnA

Appendix 1: 선행연구 정리자료

1. 머신러닝 기반의 보상형 크라우드펀딩 성공 예측 모델링, 한국경영학회, 2020

- 관련성: 우리 연구와 동일한 목적의 달성을 꾀함
- 활용·차별화 방안: 크라우드 펀딩 성공 자체에 대한 예측에 대해 학습하고 소비자 반응에 의해 형성된 속성들을 제외하고 상품 자체에 대한 데이터만을 이용하여 성공 예측이 가능한지에 대해 연구할 계획.

2. 보상형 크라우드펀딩 플랫폼에서 후원자의 참여의도 결정 요인에 관한 연구, 한남대학교, 2021

- 관련성: 크라우드 펀딩의 성공 여부와 후원자의 참여의도 결정 요인은 상관성이 높은 연구임으로 동일한 목적의 달성을 꾀함
- 활용·차별화 방안: 후원자의 참여 의도 결정 요인을 참고하여 사이트 내 추출해야 할 데이터를 선별하는데 활용할 예정. 해당 연구의 경우, 데이터를 활용하는 것이 아닌 소비자들의 심리, 설문조사 등을 이용한 정성적 연구로써의 한계를 가지고 있음.

3. 투자형 크라우드펀딩의 투자자 등급별 투자결정요인 분석, 성균관대학교, 2021

- 관련성: 보상형 크라우드 펀딩이 아닌 투자형 크라우드 펀딩의 특성에 대해 학습하고 보상형 크라우드 펀딩과의 연관성, 공통점에 대해 파악
- 활용·차별화 방안: 투자자를 등급별로 나누는 과정에 대해 학습하고 투자자에 대한 정보를 수집 시, 각 부류에 따른 모델 생성에 활용될 예정

4. 시민적 크라우드 펀딩의 성공 요인: 비영리조직의 '와디즈'이용 사례 분석, 시민사회와 NGO, 2020


- 관련성: 우리 연구와 동일한 목적의 달성을 꾀함
- 활용·차별화 방안: 시각적 디지털 콘텐츠 활용 여부에 따른 성공 여부에 대해 학습할 수 있지만 디지털 콘텐츠 정보에 대한 분석이 아닌 콘텐츠 개수, 동영상 개수 등에 대한 정보만을 활용하였기 때문에 본 연구에서는 제품 설명에 대한 정보분석을 통해 성공 여부를 예측하는데 활용할 예정.

Appendix 2 : 데이터 수집 예시

테크·가전

[이어폰에서 완전 해방] 글로벌 평점 4.5 | 제대로 만든 사운드 선글라스

스토리 반환·정책 **새소식 4** 커뮤니티 13 서포터 396 리워드 선택



여름 내내 음악과 함께 할 겁니다. 단, 무선 이어폰 없어요. 글로벌 평점 4.5로 검증 끝! 세상 작은 블루투스 스피커를 장착한 사운드 선글라스를 대한민국 단독으로 소개합니다.

목표 금액 3,000,000원

펀딩 기간 2022.05.18-2022.06.20


100% 이상 모이면 펀딩이 성공되며, 펀딩 마감일까지 목표 금액이 100% 모이지 않으면 전체가 진행되지 않습니다.

26일 남음


2445 % 달성

73,370,000 원 펀딩


396 명의 서포터


 글로벌 프로젝트 ✓ 국내 공식 선정 ✓ 인증 보증된

예약 주문하기

♡ 217  문의 공유하기

펀딩이 처음이라면 누구나!
첫 펀딩 최대 1만 원 혜택 [보러가기](#)

 프로젝트에 문제가 있나요? [신고하기 >](#)

71명에게 확산 중입니다 

199,000원 펀딩

[슈퍼 얼리버드] 사운드코어 선글라스 세트 (블랙)

*프레임 (블랙) + 브릿지 세트 / 펀딩할인율: 25% (소비자가 249,000원)

배송비 3,000원

리워드 발송 시작일 2022년 07월 초 (1~10일) 예정

제한수량 300개 **현재 1개 남음!**

총 299개 펀딩완료

209,000원 펀딩

[얼리버드] 사운드코어 선글라스 세트 (블랙)

*프레임 (블랙) + 브릿지 세트 / 펀딩할인율: 19% (소비자가 249,000원)

배송비 3,000원

리워드 발송 시작일 2022년 07월 초 (1~10일) 예정

제한수량 190개 **현재 176개 남음!**

총 14개 펀딩완료

< 웹 크롤링 예시 >

Appendix 2 : 데이터 수집 예시

min_price	target	cnt_main_images	cnt_title_images	start_date	end_date	category	code	result	success	cnt_sell	subcategory	title	explain	main
11900	500000	5	2	2022-04-20	2022-04-29	푸드	145312	435	1	42.01681	건강 기능 식품	몸에도 좋고 맛도 좋은 백숙	백숙을 통해 단백질을 섭취하세요.	본 백숙은 10년치의 노하우를 가진 특별한 음식
35900	500000	7	3	2022-04-15	2022-05-09	푸드	145276	395	1	13.92758	식품(농수축산물)	생략	생략	생략
22000	500000	31	4	2022-04-19	2022-04-25	푸드	145132	442	1	22.72727	식품(농수축산물)	생략	생략	생략
41900	500000	16	1	2022-04-17	2022-04-25	게임·취미	145082	278	1	11.93317	기타 재화	생략	생략	생략
98000	500000	12	1	2022-04-18	2022-05-13	스포츠·모빌리티	145037	0	0	5.102041	구두/신발	생략	생략	생략
89000	500000	10	1	2022-04-17	2022-05-20	스포츠·모빌리티	144930	126	1	5.617978	구두/신발	생략	생략	생략
99000	500000	61	1	2022-04-21	2022-04-28	패션·잡화	144921	953	1	5.050505	패션잡화 (모자,벨트,액세서리)	생략	생략	생략
17900	500000	33	1	2022-04-14	2022-04-27	푸드	144912	327	1	27.93296	가공식품	생략	생략	생략
98000	500000	11	1	2022-04-18	2022-05-11	스포츠·모빌리티	144910	0	0	5.102041	구두/신발	생략	생략	생략

< 수집 데이터 예시 >

피처 설명 : num_reward=리워드 개수 / min_price=최소금액 / target=목표금액 / cnt_main=본문이미지개수 / cnt_title=본문이지미개수 / start_date=프로젝트 시작날짜 / end_date=끝난 날짜 / category=카테고리 / code=프로젝트번호 / result=달성량 / success=성공여부 / cnt_sell=목표달성하기위한최소판매개수 / subcategory=상세카테고리 / title=제목 / explain=설명란 / main=본문

Appendix 3 : 모델링 결과 비교표

Model	Accuracy	Precision	Recall	F1_score
KNeighborsClassifier (n_neighbors=2)	0.9358125318390219	0.9750297265160524	0.951276102088167	0.9630064591896653
SVC(C=0.025, kernel='linear')	0.9617931737137035	0.9598438371444507	0.9982598607888631	0.978675007108331
SVC(C=1, gamma=2)	0.9205298013245033	0.9188034188034188	0.9976798143851509	0.9566184649610678
GaussianProcessClassifier (kernel=1**2 * RBF(length_scale=1))	0.9653591441670912	0.9747706422018348	0.9860788863109049	0.9803921568627451
DecisionTreeClassifier (max_depth=5)	0.9612837493632196	0.9708571428571429	0.9854988399071926	0.9781232009211283
RandomForestClassifier (max_depth=5, max_features=1)	0.8782475802343352	0.8782475802343352	1.0	0.9351776512069433
MLPClassifier (alpha=1, max_iter=1000)	0.9658685685175752	0.974785100286533	0.9866589327146171	0.9806860766791583
AdaBoostClassifier()	0.9653591441670912	0.9747706422018348	0.9860788863109049	0.9803921568627451
LogisticRegression (max_iter=5000, solver='saga')	0.9653591441670912	0.9709897610921502	0.990139211136891	0.980470993681792
LGBMClassifier (n_estimators=2000)	0.9597554763117677	0.973517559009787	0.9808584686774942	0.9771742271019935
XGBClassifier (n_estimators=2000)	0.9592460519612838	0.9724137931034482	0.9814385150812065	0.976905311778291

* 기존 데이터 / train set : test set = 7 : 3 / 측정값 별로 상위 높은 값을 bold 표시

Appendix 3 : 모델링 결과 비교표

- MLPClassifier (alpha=1, max_iter=1000)

적용기법	Accuracy	Precision	Recall	F1_score
None	0.9658685685175752	0.974785100286533	0.9866589327146171	0.9806860766791583
SMOTE	0.9673968415690269	0.9742857142857143	0.9889791183294664	0.9815774323546345
BSMOTE	0.947215777262181	0.9415807560137457	0.9535962877030162	0.947550432276657
ADS	0.9465116279069767	0.9567022538552787	0.935614849187935	0.9460410557184751

- LogisticRegression (max_iter=5000, solver='saga')

적용기법	Accuracy	Precision	Recall	F1_score
None	0.9653591441670912	0.9709897610921502	0.990139211136891	0.980470993681792
SMOTE	0.9463457076566125	0.9291689905186837	0.9663573085846868	0.9473983508672164
BSMOTE	0.9405452436194895	0.9298245614035088	0.953016241299304	9412775708965913
ADS	0.9267441860465117	0.9088888888888889	0.9489559164733179	0.9284903518728717

- LGBMClassifier (n_estimators=2000)

적용기법	Accuracy	Precision	Recall	F1_score
None	0.9597554763117677	0.973517559009787	0.9808584686774942	0.9771742271019935
SMOTE	0.9782482598607889	0.9779710144927536	0.978538283062645	0.9782545665410264
BSMOTE	0.9811484918793504	0.9831100757134537	0.9791183294663574	0.9811101424004649
ADS	0.9758720930232558	0.9773123909249564	0.974477958236659	0.9758931164681964

* 상위 3개 모델에 대해 oversampling 적용 후 / train set : test set = 7 : 3 / 측정값 별로 상위 높은 값을 bold 표시

Appendix 4 : 군집화 결과 _ WordCloud



< 뷰티 카테고리 군집 결과 >

- 왼쪽 군집에는 향수에 대한 군집이 오른쪽 군집에는 피부에 대한 군집이 형성되었음을 확인할 수 있다.
- 같은 카테고리임에도 다른 군집이 형성되었음을 확인할 수 있음.
- 양 쪽 군집 모두 앵콜이라는 키워드가 차지하는 것으로 보아 인기가 많은 군집이라고 판단할 수 있다.