

품사 정보를 활용한 크라우드 펀딩 성공 여부 예측

계명대학교 통계학과 안유나 계명대학교 경영정보학과 김웅기
계명대학교 통계학과 최지우 계명대학교 경영정보학과 김서영

목차

연구 배경

연구 방법

연구 진행

결론 및 개선 방향

연구 배경

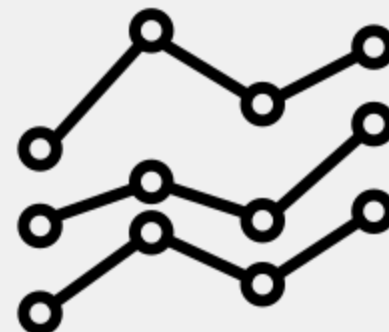


크라우드 펀딩의 등장으로 인한 투자 절차와 과정의 단순화



신뢰할 수 있는 투자 정보 필요 : 품사가 미치는 영향 분석

연구 방법



데이터 수집

와디즈 펀딩 : 패션잡화

2019년 9월~

2022년 8월

데이터 전처리

신규 메이커 여부,

성공 및 실패 여부 변수 생성

Konlpy 패키지 사용

제목별 품사 정보 생성

모델링

LightGBM,

GradientBoosting,

AdaBoost 3개 모델로

클라우드 펀딩 성공 예측

모델 성능 평가 및 결론

Accuracy, Precision, Recall,

F1-score 평가 지표 사용

연구 진행 - 데이터 수집

와디즈(<https://www.wadiz.kr/web/main>)의 패션잡화 카테고리에서
2019년 3월 ~ 2022년 8월까지 데이터 수집
BeautifulSoup 라이브러리와 Selenium 라이브러리사용

변수명	변수 설명	비고
goal	펀딩 목표 금액	
img	펀딩 게시글의 이미지 수	
length	펀딩 게시글 제목 길이	
funding_term	펀딩 기간	
pf	펀딩 성공 여부	1: 성공, 0: 실패

<표 1> 수집 데이터 변수 설명

연구 진행 - 데이터 전처리

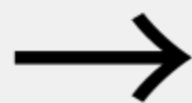
추가 변수 생성

수집한 데이터를 기반으로 메이커에 대한 신뢰성을 판단할 수 있는 지표 생성
(성공, 실패, 신규)

Konlpy를 활용한 품사 정보 생성

Konlpy의 Okt 분석기 사용

: 토큰화, 불용어 제거, 품사 태깅
(사전 추가, 불용어 사전 생성)



제목별 품사 개수 정보 생성

빈도 상위 5개 ('Noun', 'Josa', 'Verb',
'Adjective', 'Number') 품사 정보만을
활용하여 분석 진행

연구 진행 - 분석 데이터

구분		변수명	변수 설명	비고
종속변수		pf	펀딩 성공 여부	1: 성공 0: 실패
독립변수	펀딩 정보	sucess	펀딩 메이커의 이전 프로젝트까지 성공 횟수	
		fail	펀딩 메이커의 이전 프로젝트까지 실패 횟수	
		new	펀딩 메이커의 프로젝트 경험 여부	1: 신규메이커 2: 기존메이커
		goal	펀딩 목표 금액	
		img	펀딩 게시글의 이미지 수	
		length	펀딩 게시글 제목 길이	
		funding_term	펀딩 기간	
	게시글 제목 품사 정보	Josa	조사	
		Adjective	형용사	
		Noun	명사	
		Number	숫자	
		Verb	동사	

<표2> 분석 데이터 변수 설명

사용한 모델링 기법

A light green circle containing the text "LGBM" in a bold, dark gray sans-serif font.

LGBM

LightGBM

A yellow circle containing the text "GBT" in a bold, dark gray sans-serif font.

GBT

Grandient Boosting

A light orange circle containing the text "ADB" in a bold, dark gray sans-serif font.

ADB

AdaBoost

연구 결과 - 모델링 평가

사용변수	평가지표	LGBM	GBT	ADB
편딩정보	Accuracy	0.87	0.87	0.87
	Precision	0.85	0.86	0.86
	Recall	0.87	0.87	0.87
	F1-score	0.85	0.86	0.86
편딩정보 + 품사정보	Accuracy	0.87	0.88	0.88
	Precision	0.86	0.87	0.87
	Recall	0.87	0.88	0.88
	F1-score	0.86	0.86	0.87

<표3> 모델 성능 결과

결론 및 개선 방향

결론

프로젝트 제목의 품사 정보가
클라우드 펀딩 성공 여부 예측 모델
성능에 기여함

개선 방향

데이터 자체의 불균형으로
실패한 프로젝트 예측 어려움
실패 프로젝트의 충분한 데이터
확보 필요

감사합니다.