

금융 빅데이터 활용 아이디어 공모전 보고서

스타트업 및 크라우드 펀딩 지원 신규 플랫폼 서비스 제안서

남 정 재

이 현 우

이 세 혁

데이터를 줍다! Zoup!

CONTENTS

01

팀 소개

02

아이디어
제안

03

플랫폼 소개

04

플랫폼 기능

05

기대 효과

06

시도 및 한계

07

후기 및
참고 문헌

01

팀 소개

01

“ 팀 소개 ”

데이터를 줍다 ! Zoup !

		
남 정 재	이 현 우	이 세 혁
고려대학교 세종캠퍼스 응용통계학과	고려대학교 세종캠퍼스 컴퓨터융합소프트웨어학과	고려대학교 세종캠퍼스 디지털경영학과
njj97@naver.com 010 - 3033 - 9273	msj11188i@gmail.com 010 - 4078 - 3930	1227peter@korea.ac.kr 010 - 7734 -2344

02

아이디어 제안

02

“ 아이디어 제안 ”

상권분석 기반  창업지원 정보 제공

창업자 데이터 기반  클라우드 펀드 상품 추천



빅데이터 기반의 **AI** 를 통한 **상품성 예측**
클라우드 펀드 투자자 **매칭** 서비스 제안 !

03

플랫폼 소개

03

“ 플랫폼 소개 ”

{ 소 개 }

빅데이터를 저장, 정제하는 플랫폼
사용자의 입력값을 시로 분석하여 지표 및 시각화 정보 제공

{ 추가 데이터 사용 }

창업자와 금융상품 고객과의 클라우드 펀딩 매칭
시 모델을 이용한 상품 추천
기존의 흩어진 창업 및 상권 정보를 결집 및 정형화

{ 발전 가능성 }

상권에서 일어나는 거래의 중심이 되는 플랫폼으로 도약

03

“ 플랫폼 소개 ”

1. 인사이트 서치

{ 데이터 접근 }

‘**창업**’ 이란 키워드와 직접적으로 연관된,
제공된 데이터와 그 외에 영향이 있는 외부데이터를 다루었다.

03

“ 플랫폼 소개 ”

2. lightGBM AI 모델

창업 상품의 상품성을 예측하는 모델

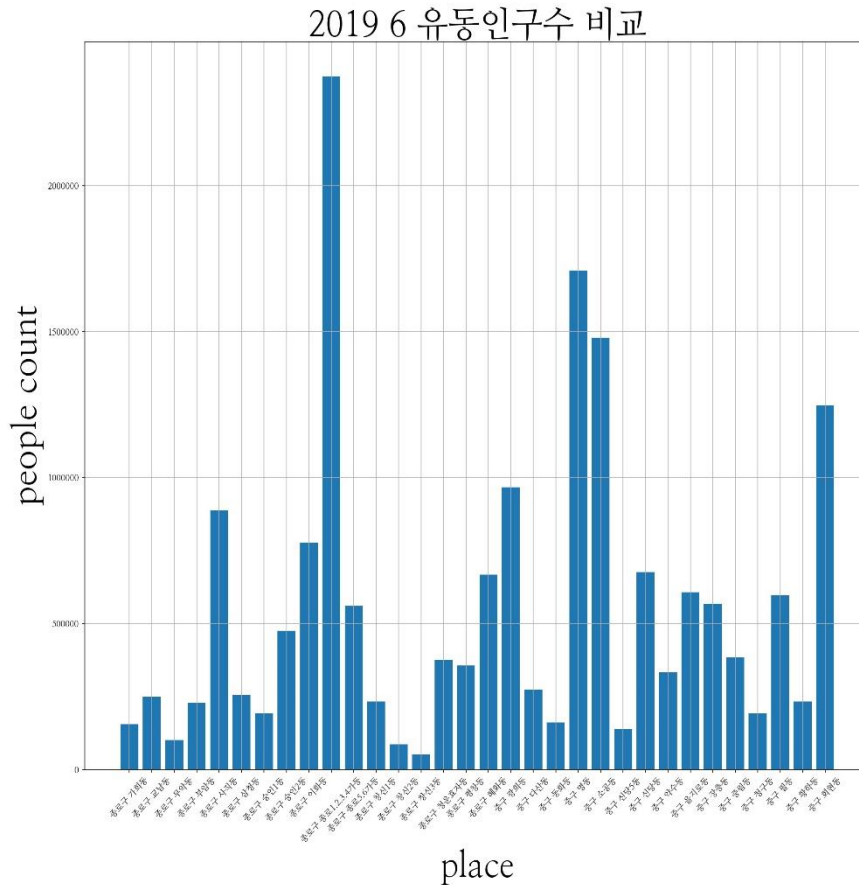
고객데이터를 토대로 투자자의 크라우드 펀딩 상품 추천

유통업체와 창업자와의 매칭 추천

☞ ‘ 창업 상품의 상품성을 예측하는 모델 ’ 구현하기로 결정

03

“ lightGBM AI 모델 ”



유동인구가 **증감**

But,

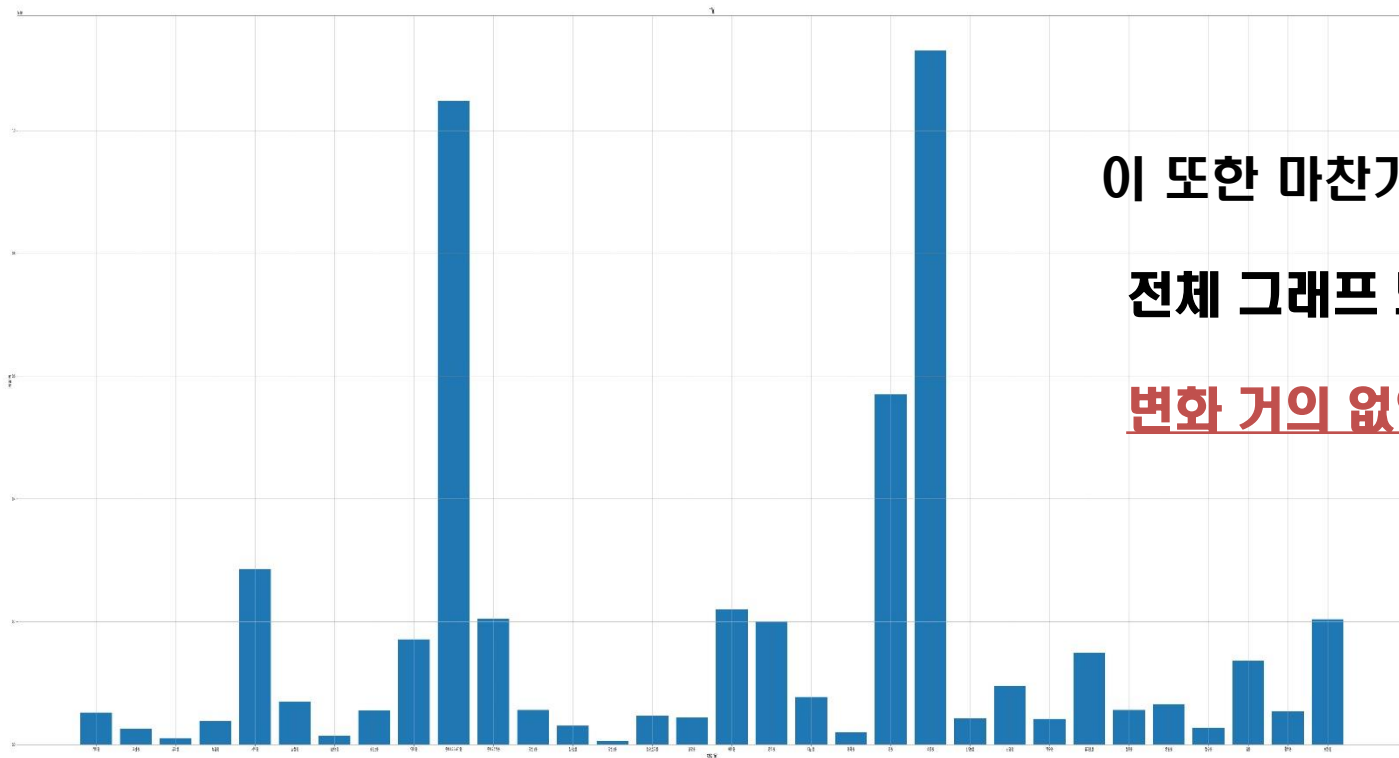
전체 그래프 모양 **변화 없음**

그래프 . 지역의 특성을 조사한 결과

03

“ lightGBM AI 모델 ”

그래프 . 월별 매출액을 지역(동)끼리 비교한 그래프



이 또한 마찬가지로
전체 그래프 모양
변화 거의 없었음

03

“ lightGBM AI 모델 ”

	Unnamed: 0	movie	big_mart	cafe	hospital_L	hospital_s	m10_max	m10_min	m20_max	m20_min	...			
Unnamed: 0	1.000000	0.169423	0.381392	0.257306	0.392509	0.283557	0.088095	0.029792	0.136015	0.097349	...			
movie	0.169423	1.000000	0.576665	0.673676	-0.080453	0.262779	0.640949	0.676267	0.784649	0.797985	...			
big_mart	0.381392	0.576665	1.000000	0.844257	0.595752	0.746849	0.451051	0.521987	0.577333	0.564063	...			
cafe	0.257306	0.673676	0.844257	1.000000	0.431331	0.803010	0.573893	0.562842	0.733368	0.694463	...			
hospital_L	0.392509	-0.080453	0.595752	0.431331	1.000000	0.636224	-0.081352	-0.026728	-0.065300	-0.084688	...			
hospital_s	0.283557	0.262779	0.746849	0.803010	0.636224	1.000000	0.186742	0.286379	0.344231	0.336169	...			
		place	price_range_code		menu_text		review_count							
		1	종구 필동		3		11		7					
	place	etl_ymd	m10	m20	m30	m40	m50	m60	f10	f20	f30	f40	f50	f60
0	종로구 가회동	20190601	2887	16819	18011	14594	13676	10008	3825	22438	17366	14225	13636	8012

상관분석을 통해 주변 인프라의 수가 유동인구에 영향이 있음을 수치적으로 보였음

03

“ lightGBM AI 모델 ”

[장소(범주)] [유동인구(성별, 나이) 고/저] [가맹점 정보(범주)]

☞ Input Data

각 창업 아이템의 가치를 나타내는 데이터는 찾기 힘들었다.

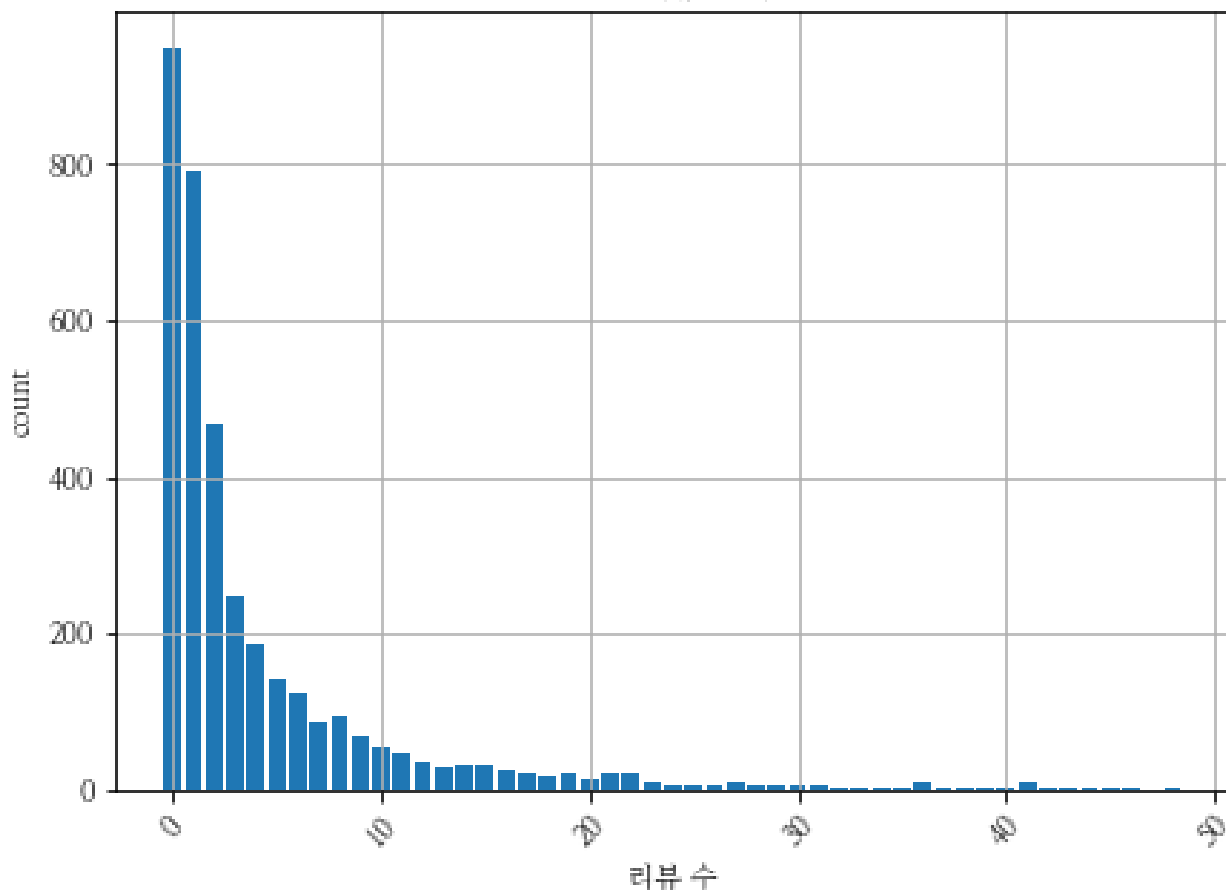
배달 어플을 토대로 리뷰 개수가 가게의 인지도를 절대적임은 아니
지만 나타내는 하나의 지표로 어느정도 작용할 수 있음을 보였다.

☞ Output Data는 ' 리뷰의 개수 ' 로 설정 후 학습 시도

03

“ lightGBM AI 모델 ”

2019 6 리뷰 수 비교



극단적으로 편향
넓은 스펙트럼
⇒ 학습에 부적절

⇒ 일정 기준을 두고
범주형으로 처리

lightGBM은
xboost와 달리
범주형 데이터
처리 가능 !

03

“ lightGBM AI 모델 ”

```
print("accuracy: ", fo
```

accuracy: -23.377 %



```
print("accuracy: ", format
```

accuracy: 43.833 %

이전의 정확도와 비교 결과 **매우 큰 향상**을 보임

50%에 미치지 못하는 정확도

⇒ 각각의 점포를 대표하기엔 부족한 데이터로 한 학습이 원인

⇒ 모호한 데이터의 유형에도 불구하고 43 %의 정확도

⇒ 더 세밀하고 정교한 데이터로 모델 구축 시
충분히 좋은 모델 구축 기대할 수 있음

03

“ lightGBM AI 모델 ”

bc 플랫폼 창업지수

지역검색

Enter 행정동:

중구 신당동

submit

View details »

검색된 리스트 출력

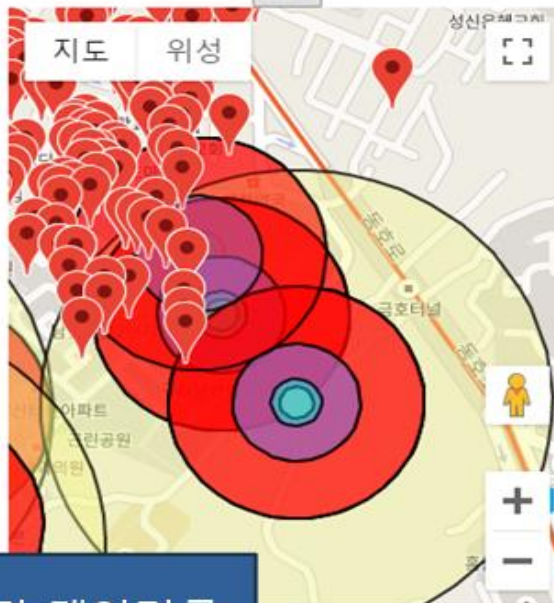
남산타운
남산타운
남산타운
남산프라임
남산타운
남산타운

인프라 데이터를
통해 영향요소와
영향력 확인 가능

google api

Search Box

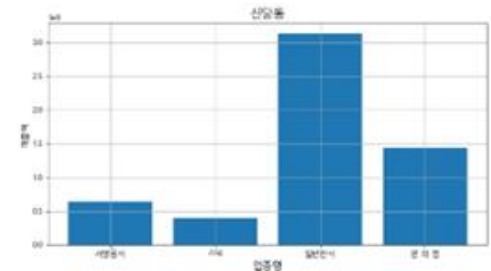
find



창업분포

그래프

ai 연산 후 결과
출력 부분



AI 모델서 입력 값을 부여하였을 때 나오는 결론을 출력

임의로 정해진 인프라의 영향을 받는 공간상 반경을
구글 API를 통해 시각적 자료로 제공이 가능한 UI 구성 결과임

04

플랫폼 기능

04

“ 플랫폼 기능 ”

1

창업 아이템의 **상품성** 시로 예측

2

BC 협력 은행은 **사업대출의 승인 여부**를 판단할 **지표** 제공

3

플랫폼 내 창업자 **정보 관리** 가능 및
창업에 영향 주는 **다른 분야의 데이터** 저장 가능

4

클라우드 펀딩 활성화 & 시를 통해 클라우드 펀딩 **종목 추천** 가능

5

데이터 플랫폼을 넘어 **소상공인 플랫폼**의 역할 가능
>>> **유통업체 추천과 연결**

05

기대 효과

05

“ 플랫폼 구조 ”



[그림1. 스타트업 및 클라우드 펀딩 지원 Platform 구조화]

05

“ 기대 효과 ”



06

시도 및 한계

06

“ 시도 ”

**1. 여러가지 통계를 통해 넓은 주제로
창업 관련 데이터 도출(주변 인프라 데이터 등)**

다변량, 범주형 데이터를 다룸 [범주에 따라 다른 분기문]

2. Output 데이터로 선정한 값(리뷰 수) 정규화 어려움

편향된 수, 넓은 스펙트럼 모델 예측 곤란

Softmax로 범주화

일정 이상은 한 범주에 포함되도록 처리하려 정확도 증가시킨 방법론

3. 2019년도 기준 최신 AI 알고리즘 'lightGBM'을 적용

2017년에 발표된 현재까지 기준으로 최신 AI 알고리즘으로 모델 구현
기존의 프로젝트의 정확도를 보여 예측모델에 적합함을 보이는 것을 확인함
주어진 데이터를 감안하여 성공적인 구축이라고 할 수 있음

06

“ 한 계 ”

1. 창업 주제를 대표하는 데이터 즉, **점포 고유의 데이터** 없기에 **AI 예측 모델 구축 한계**(세부적인 데이터 필요)
2. 데이터가 적은 **마이너한 분야**의 경우 예측 성공이 보장되기에 힘들다.
3. 동일한 주제임에도 불구하고 데이터로 나타내기 힘든
[맛], [홍보 정도] 와 같은 **데이터의 누락**
⇒ **다른 결과를 보이는 데이터 존재**
(ex) 동일업종, 동일가격 수치의 데이터임에도 불구하고 리뷰 수 다름

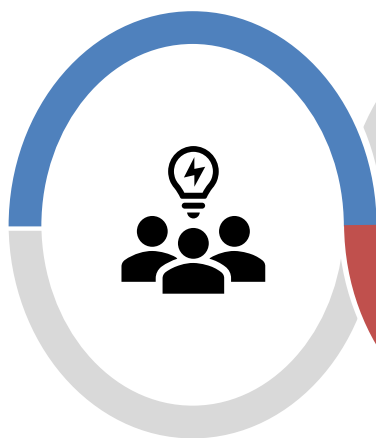
07

후기 및 참고 문헌

“ 후기 및 참고문헌 ”

1. 짧은 시간동안 매우 다양한 데이터를 분석
2. 창업 즉, 스타트업의 성공 판별 기준에 대해서 깊게 생각해볼 수 있었음
3. 비록 AI의 정확도가 수치상으로 높은 값은 아니었지만 비교적 불확실한 데이터를 갖고 모델을 구현하였으므로 꽤나 유의한 수치라고 판단되어 기대됨
4. AI 와 빅데이터 , 창업과 펀드 , 그리고 플랫폼 관련 논문과 서적, 기사들을 조사하고 읽어보면서 이에 대한 안목과 역량을 키울 수 있었던 시간

“ 후기 및 참고문헌 ”



**Lig htGBM : A Highly
Efficient Gradient
Boosting Decision Tree**

[논문]
NeurIPS (2017)



**국내 사업자
현황 데이터**

국세청 국세통계
(2016 ~ 2019)



**국내 폐업자
현황 데이터**

국세청 국세통계
(2013 ~ 2017)



**창업 장애요인
[2013년 이후]**

KOSIS
창업기업실태조사
[2013 ~]



**신규 자영업자
사업자금 조달방법**

KOSIS
경제활동 인구조사

{ 기본 데이터 출처 }

[출처] 공모전 제공 데이터, [BC 카드], [KT], [망고플레이트], [다음소프트]

데 이 터 를 줍 다 ! Z o u p !

**THANK
YOU**