

제 1 회

금융 빅데이터

활용 아이디어 공모전 보고서

데이터를 줍다! Zoup!

남정재 이현우 이세혁

## 요 약

### “빅데이터를 기반으로 한 상품성 예측과 클라우드 펀드 투자자 매칭”

최근 창업을 하는 사람들이 점점 증가하고 있다. 그와 동시에 창업을 하고자 하는 예비 창업자들 또한 늘고 있다. 하지만 대부분의 예비 창업자들은 좋은 사업 아이템을 들고 있음에도 창업 자금 확보와 경영에 대한 지식의 부족으로 인해 창업에 어려움을 겪고 있다. 이 문제에 대해 빅데이터 분석으로 상품의 마케팅을 지원하고, 사업에 맞는 투자자를 매칭해주는 플랫폼을 제안한다.

이 플랫폼의 작동방식은 다음과 같다. 우선, 사용자들을 분류 한 뒤(창업자, 투자자) 그들의 정보를 모은다. 창업 예정자의 경우에는 본인의 상품에 대해 정보를 기입한다. 그리고 투자자는 본인의 투자 경험이나, 본인의 관심사 등에 대해 기입한다. 플랫폼은 창업자가 희망하는 사업에 대해 자체 알고리즘을 통하여 분석하고, 그것의 상품성을 반환한다. 이것을 바탕으로 사업을 결정하고, 결정된 사업은 투자자에게 노출된다. 이에 대한 사업을 보고 투자자는 투자를 결정한다.

이 방식으로 인하여 창업자는 본인의 사업에 대한 수치적인 상품성을 직접 확인 할 수 있게 되고, 이것에 플랫폼은 추가적인 데이터를 제시하여 사업에 대한 전반적인 안목을 제시해준다. 반면에 투자자는 간결한 데이터를 통해 쉽게 투자 할 수 있게 되고, 전망이 불확실한 사업은 피할 수 있게 된다.

또한, 축적된 고객들의 빅데이터를 바탕으로 ‘사용자별 맞춤 펀드 서비스’, ‘유통 구조의 개선’, ‘투자의 난이도에 대한 인식 개선’ 등을 이끌어 낼 수 있을 것으로 기대된다.

# 목차

1. 팀 소개
2. 사전 조사 및 아이디어 제안
3. 모델 설명 및 플랫폼 소개
  - 1) 모델 설명
    - a. 구현 방법
    - b. 한계
  - 2) 플랫폼 소개
    - a. 구성요소
    - b. 기능
    - c. 기대효과
4. 결론 및 후기
5. 참고문헌

## 1. 팀 소개

**데이터를 줍다!**

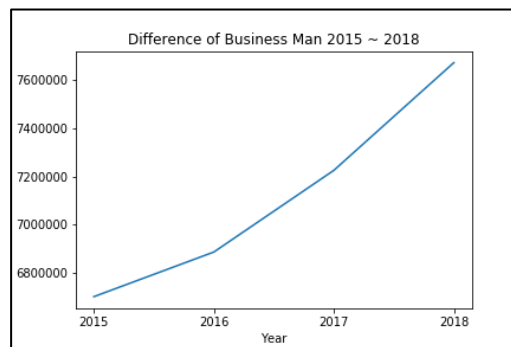
***Zoup!***

		
남정재	이현우	이세혁
고려대학교 세종캠퍼스 응용통계학과	고려대학교 세종캠퍼스 컴퓨터융합소프트웨어학과	고려대학교 세종캠퍼스 디지털경영학과
njj97@naver.com 010 - 3033 - 9273	<a href="mailto:msj11188i@gmail.com">msj11188i@gmail.com</a> 010 - 4078 - 3930	<a href="mailto:1227peter@korea.ac.kr">1227peter@korea.ac.kr</a> 010 - 7734 -2344

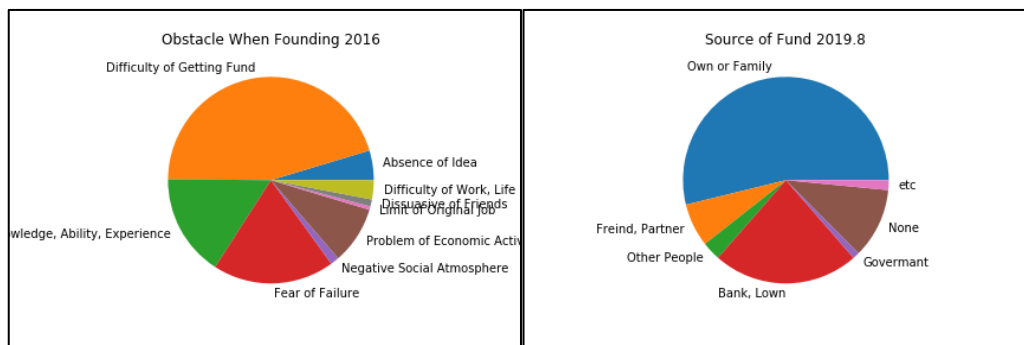
## 2. 사전 조사 및 아이디어 제안

### “빅데이터 분석을 통한 창업 서포팅과 투자자와 창업자를 연결하는 플랫폼”

처음 주제를 받고 나서, 조원들과 상의해본 결과 비금융데이터와 금융데이터를 결합하면 주변의 상권 분석을 통해 더 쉽게 사업을 전개할 수 있을 것이라 생각했다. 따라서 우선 창업 시장 현황을 알아보려고 했다. 국세청의 데이터를 확인해본 결과 2013년부터 2018년까지 신규 사업자가 꾸준히 증가했음을 확인했고, “빅데이터를 통해 창업자를 지원하는 플랫폼”으로 방향을 잡았다.



위와 같은 주제로 선정한 후, 그에 따라 창업을 어렵게 하는 요인들을 분석했다. 그 결과 창업자들은 ‘창업 자금 확보’가 가장 큰 걸림돌이었고, ‘창업에 대한 전반적인 지식, 능력, 경험 능력 부족’, ‘창업에 대한 막연한 두려움’이 뒤를 이었다.



위의 조사 이후 창업자들이 가장 어려워하는 사업 자금 확보를 해결한 사람들은 어떻게 해결했는지 의문점이 생겼고, 대부분의 창업자들이 ‘본인 또는 가족들이 마련한 목돈’ 혹은 ‘은행, 보험 회사, 상호 신용 금고’에서 자금을 마련하는 것을 발견했다. 특히 후자의 경우 은행이 일반적으로 창업 전 담보만 보고 대출해주기 때문에 자영업자의 실패를 유도한다는 기사를 발견했다.<sup>1</sup>

즉, 예비 창업자가 겪는 주요 문제는 ‘창업 자금 확보’와 ‘창업에 대한 전반적인 지식과 능력 부족’이며, 이를 보완하기 위해서는 창업 아이템의 상품성에 대한 적합한 분석과 투자를 더 간단히 받을 수 있는 방법이 필요하다고 생각했고, 기존의 아이디어를 보완했다

<sup>1</sup> <http://www.hani.co.kr/arti/economy/finance/860675.html>

### 3. 예측 모델 설명 및 플랫폼 소개

→ 모든 데이터를 조합하여 제시하는 창업 아이템의 상품성을 AI로 예측한다.

#### 1) 모델 설명

→ 플랫폼은 창업자가 제시한 창업 주제에 대해 AI 모델을 통해 상품성에 대한 데이터를 제공하고, 그것과 창업에 영향을 주는 데이터를 함께 제공한다. 단, 기존의 타 플랫폼과 달리 더 넓은 주제의 데이터를 제공한다.

- 기존의 플랫폼은 지역의 상품 특성이나 업종 분석 등, 기업 하나하나를 분석하는 우리의 플랫폼과 달리 지역별로 분석을 했다.

##### a. 구현 방법

1. 여러가지 통계를 통해 넓은 주제로 창업관련 데이터를 도출한다.
2. Output 데이터로 선정한 값을 softmax<sup>2</sup>로 범주화 시킨다.
3. AI 알고리즘인 lightGBM<sup>3</sup>을 적용시킨다.

##### b. 한계

- 제공된 데이터와 추가로 찾은 데이터의 경우, 창업 주제를 대표하는 데이터와 창업하는 가게 고유의 데이터가 없었기 때문에 AI 예측 모델을 구축하는데 한계가 있었다.
- 데이터가 적은 마이너한 분야의 사업인 경우 예측 성공이 보장되기엔 힘들다.
- 동일한 주제임에도 불구하고 수치로 나타내기 힘든 데이터<sup>4</sup>의 누락으로 인해 다른 결과를 보이는 데이터가 존재한다.<sup>5</sup>

#### 2) 플랫폼 소개

→ 빅데이터를 저장, 정제하는 플랫폼으로 사용자의 입력값을 AI로 분석하여 지표 제공을 할 수 있으며, 시각화 정보를 제공하는 데이터 센터이다. 여기에 추가로 데이터를 이용하여 창업자와 금융상품 고객과의 클라우드 펀딩 매칭, AI 추천 기능 등을 하며, 기존의 흩어진 창업 및 상권 정보를 결집시키고 정형화하여 이 데이터를 기반으로 상권 거래의 중심이

---

<sup>2</sup> 일정치 이상은 한 범주에 포함되도록 처리를 하여 정확도를 증가시킨 방법론이다.

<sup>3</sup> 기존의 프로젝트에서 예측모델에 적합성을 보였다.

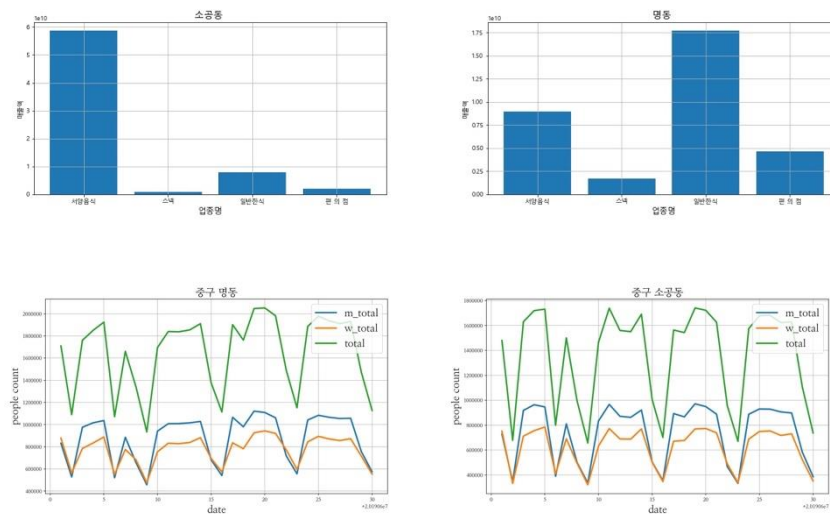
<sup>4</sup> 맛, 가게 별 광고의 질과 양 등등

<sup>5</sup> 동일 업종, 동일 가격 수치의 데이터임에도 불구하고 리뷰 수가 달리 나오는 것들

되기도 하는 플랫폼이다.

#### a. 플랫폼 구성 요소

- **인사이트 서치**: 데이터 접근을 '창업'이란 키워드와 직접적으로 연관된 제공받은 데이터와 그 외에 영향이 있다고 판단되는 외부데이터를 다루었다.
- **lightGBM AI 모델**: 창업 상품의 상품성을 예측하는 모델이다. 모델을 구현하기 위해 각 지역의 특성을 조사했다. 이를 통해 각 지역마다 소비자의 특성이 다르다는 것을 발견했다.



이것에 추가로 인프라 정보를 추가하고, 상관분석을 통해 주변 인프라의 수가 유동인구에 영향이 있음을 수치적으로 확인했다.

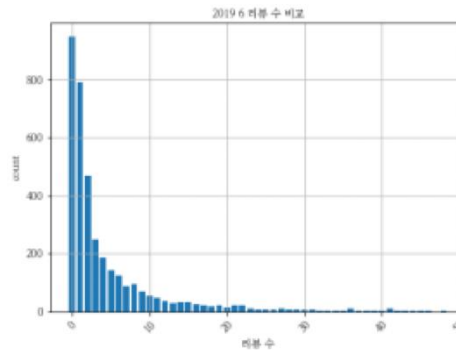
Unnamed: 0	movie	big_mart	cafe	hospital_L	hospital_s	m10_max	m10_min	m20_max	m20_min	...
Unnamed: 0	1.000000	0.169423	0.381392	0.257306	0.392509	0.283557	0.088095	0.029792	0.136015	0.097349 ...
movie	0.169423	1.000000	0.576665	0.673676	-0.080453	0.262779	0.640949	0.676267	0.784649	0.797985 ...
big_mart	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 ...

그 외의 다른 요소 (KT유동인구, 망고플레이트의 가맹점 정도 등)들을 여러 방면으로 조합했고, 결국 '[장소(범주), [유동인구(성별, 나이) 고/저], 가맹점 정보(범주)]'의 형식으로 데이터를 입력했다.

각 창업 아이템의 가치를 나타내는 구체적인 데이터로는 배달 어플에서 리뷰 개수가 각 가게의 인지도를 나타내는 하나의 지표로 활용 할 수 있다는 점에서 착안하여

출력 데이터는 리뷰의 개수로 설정한 후 학습을 시도했다.

하지만, 리뷰 수는 극단적으로 편향 되어 있고, 넓은 스펙트럼을 보이는 특징을 가졌기에 AI 학습에 부적절했다. 따라서, 일정 기준을 두고 범주형으로 처리했다.



정확도가 범주형으로 처리 이전과 비교했을 때 매우 큰 향상을 보였다. 이 모델에 주어진 데이터는 중복요소에 비슷한 데이터를 대표 할 순 있지만, 점포 하나하나의 특성을 대표하기에는 부족한 데이터로 학습을 하였다. 모호한 데이터였음을 감안했을 때 이정도 정확도가 나왔다면, 더 세밀하고 정교한 데이터로 모델을 구축하면 충분히 좋은 정확도를 보여줄 것이라 예측한다.

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

accuracy: -23.377 %

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

accuracy: 43.833 %

- **지도 데이터를 통한 수치 데이터의 산출:** AI 모델에서 입력값을 주었을 때 나오는 결론을 출력하고, 임의로 정해진 인프라가 영향을 미치는 공간상 반경을 구글API를 통해 시각적 자료로 제공이 가능한 UI를 구성해보았다.





## b. 플랫폼의 기능

- 모든 데이터를 조합하여, 제시하는 창업 아이템의 상품성을 AI로 예측
- 예측한 상품성을 토대로 BC협력 은행에 사업 대출의 승인 여부를 판단할 지표 제공
- 플랫폼을 이용한 창업자와 플랫폼에서 얻은 창업자 정보를 한 곳에서 관리
- '창업'이라는 키워드와 직접적으로 연관되지는 않지만, 창업에 영향을 주는 다른 분야의 데이터 관리
- 신규 창업자와 소상공인을 대상으로 한 클라우드 펀딩의 장



## c. 플랫폼 기대효과

- 상품성 등 여러 지표를 AI로 예측하여 판단에 도움을 줄 것이다.
- 전문적이 투자자가 아닌, 일반 투자자의 경우에도 간결한 수치로 데이터를 제공받기 때문에 펀딩 시장을 더 넓힐 수 있을 것이다.
- 이용자들에게 여러 데이터를 줌으로서 인사이트 제공할 수 있을 것이다.
- 최대의 데이터 저장소로써 창업 외 유통이나 다른 분야의 중심 플랫폼으로도 활용할 수 있을 것이다.
- 좋은 창업 아이디어가 있지만, 창업에 관한 지식이나, 능력이 없어 창업을 시작하지 못하던 사람들도 쉽게 창업을 할 수 있을 것이다.
- 이 플랫폼을 중심으로 여러 사업으로 확장 할 수 있을 것이다. 예를 들어, 이용자 간의 새로운 커뮤니티의 생성, 창업자의 사업 홍보 수단 등으로 활용이 가능 할 것이다.
- 사업에 대한 전망 데이터를 통해 은행의 대출 리스크를 줄여 줄 수 있을 것이다.

## 4. 결론 및 후기

이번 프로젝트에서 저희의 아이디어를 실현시키기 위해 짧은 시간 동안 매우 다양한 데이터를 분석해야 했습니다. 특히, 그 사업이 성공한 것을 무슨 기준으로 판별 할 지 결정하는 문제가 가장 어려웠던 것 같습니다. 비록 AI 의 정확도가 수치상으로 높은 값은 아니었지만, 짧은 시간 동안 비교적 불확실한 데이터를 갖고 꽤나 유의미한 수치를 이끌어냈다고 생각합니다.

프로젝트를 진행하는 과정에서 다양한 논문과 서적을 읽어보면서, 빅데이터로 정말 다양한 일을 할 수 있다고 느꼈습니다.

감사합니다.

**데이터를 줘다! Zoup! 일동**

## 5. 참고문헌

1. [출처] Lig htGBM: A Highly Efficient Gradient Boosting Decision Tree  
NeurIPS 2017 • Guolin Ke • Qi Meng • Thomas Finley • Taifeng Wang • Wei Chen • Weidong Ma • Qiwei Ye • Tie-Yan Liu
2. [출처] 국세청 국세통계: 국내 사업자 현황 데이터 2016 ~ 2019
3. [출처] 국세청 국세통계: 국내 폐업자 현황 데이터 2013 ~ 2017
4. [출처] KOSIS: 창업기업실태조사: 창업 장애요인 (2013 년 이후)
5. [출처] KOSIS: 경제활동인구조사: 신규 자영업자 사업자금 조달방법
6. [출처] 공모전 제공 데이터 , 비씨카드, KT, 망고플레이트, 다음소프트