

머신러닝을 통한 상권분석

‘신규 업종 추천 서비스’ & ‘창업 위험지수 예측 서비스’

팀명 : 비바(VIVA)

팀원 : 송재성, 홍석현, 박동민

멘토 : 정좌연

INDEX

001 프로젝트 개요

- 배경 및 필요성
- 목표

002 신규 업종 추천 서비스

- 구현과정
- 활용방안 및 기대효과

003 창업 위험지수 예측 서비스

- 구현과정
- 활용방안 및 기대효과

004 결론

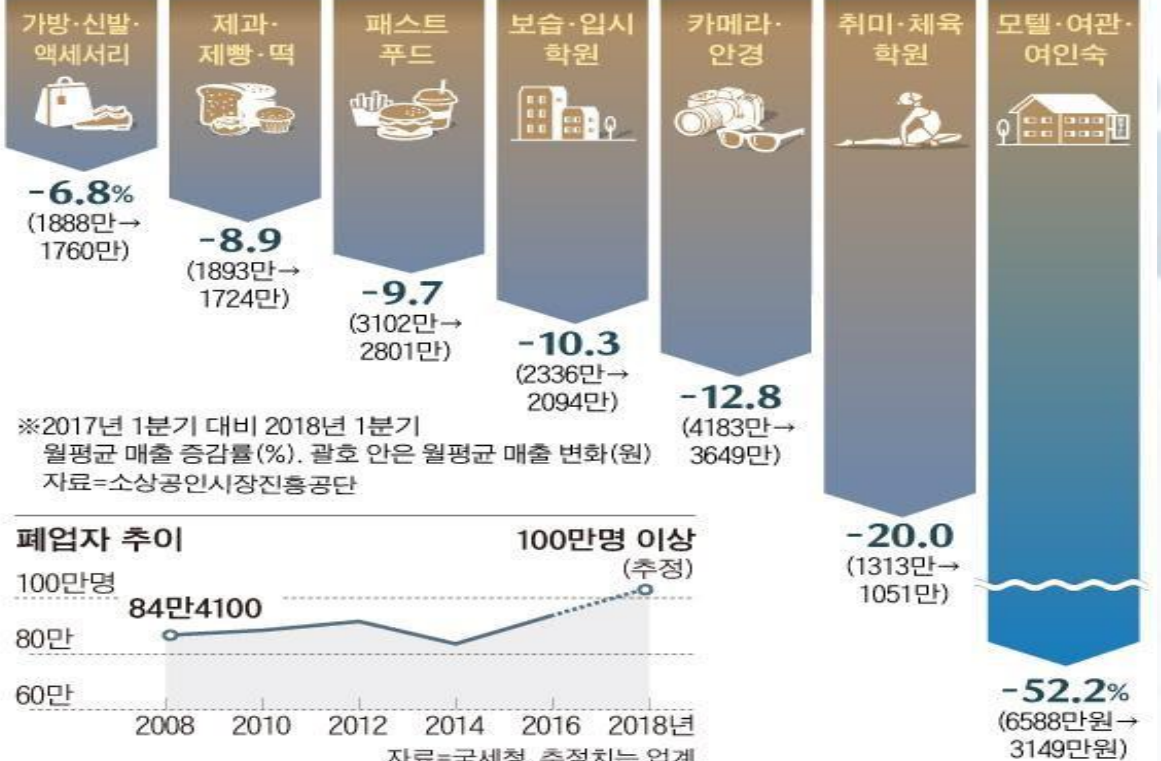
1. 프로젝트 개요

< 배경 및 필요성 >

소상공인의 창업 실패

1. 장기 경기침체에 따라 소상공인 매출이 급감하는 현실
2. 최저 임금 인상으로 영세업자들에게 직면한 위기
3. (예비)자영업자들에게는 전문 기업인들이 활용하는 빅데이터 분석 자료에 대한 진입장벽이 높음

매출 급감하는 주요 소상공인 업종



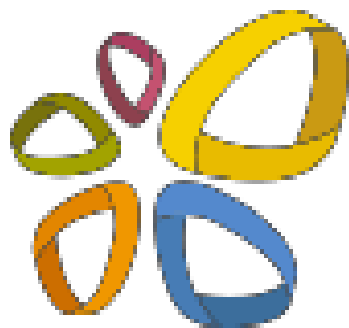
» (예비)자영업자들에게 객관적인 빅데이터 분석을 통한 전략적 계획수립 필요

궁극적 목표

1. 업종별 매출액의 상관관계 분석을 통해 특정 지역 내
신규 업종을 추천
2. 위험지수 예측 모델링 개발을 통해 특정 지역 내
신규 창업자에게 창업 위험도를 제공

신규 업종 추천 서비스- 데이터 준비

-서울 열린데이터 광장에서 제공하는 '서울시 골목상권 프로파일링 정보(상권배후지-추정매출)' 중
기준연월(2018년 4월)에 대해 상권코드/업종코드/당월매출/점포수 등을 분석데이터로 활용



서울 열린데이터 광장

기준_연월	상권_코드	상권_코드	서비스_업	서비스_업	당월_매출	시간대_00	시간대_06	시간대_11	시간대_14	시간대_17	시간대_21	남성_매출	여성_매출	연령대_10	연령대_20	연령대_30	연령대_40	연령대_50	연령대_60
201803	1744	희우정로1	CS300011	화장품	1.65E+08	0.7	2.4	17	30.6	42.9	6.5	24	76	1.8	25.4	18.5	21	17.8	15.4
201803	1744	희우정로1	CS300010	문구점	1.52E+08	0	4.3	14.4	33.5	42.7	5.1	35.2	64.8	7.4	19.2	14.8	41.9	13.1	3.6
201803	1744	희우정로1	CS300009	약국	3.79E+08	0	15.6	23	28.2	31.7	1.4	44.3	55.7	0.7	13.1	27.6	23.5	17.1	18.1
201803	1744	희우정로1	CS300008	패션잡화	56700320	0	0.7	15.2	41.6	38.8	3.7	23.5	76.5	2.7	34.3	18.4	11.3	16.7	16.7
201803	1744	희우정로1	CS300007	의류점	6.33E+08	0	2.8	20.2	37.9	36.9	2.1	25.4	74.6	1.8	14	14.7	16.8	22.7	30
201803	1744	희우정로1	CS300006	과일채소	8.42E+08	0	6.4	19.3	33.6	38.6	2.1	37.2	62.8	0.2	7.6	18.9	25.2	24.3	23.9
201803	1744	희우정로1	CS300005	정육점	7.55E+08	0	5.9	18.3	30.6	43.1	2.1	37.4	62.6	0.3	7.6	17.6	26.1	24.7	23.8
201803	1744	희우정로1	CS300004	휴대폰	45452742	0	2.3	23.3	34.5	35.2	4.7	43.8	56.2	3.3	20.9	14.8	17.5	24.9	18.6
201803	1744	희우정로1	CS300003	컴퓨터판매	80770411	0	28	23.1	39.6	9.3	0	43.2	56.8	5.2	10.4	10.4	28.6	24.7	20.8
201803	1744	희우정로1	CS300002	편의점	1.41E+09	14.4	13.8	11.8	12.8	25.5	21.7	64.5	35.5	2.1	31.4	32.2	20.7	9.9	3.7
201803	1744	희우정로1	CS300001	슈퍼마켓	2.65E+09	0.7	7.9	16.5	23.2	40.1	11.7	45.2	54.8	0.9	14.7	22.7	26.9	19.1	15.6
201803	1744	희우정로1	CS200022	자동차미용	16212851	0	12.8	21.7	44.6	20.8	0	74.9	25.1	0	11.1	34.6	29.3	18.8	6.3
201803	1744	희우정로1	CS200021	자동차수리	2.7E+08	0	11.5	23.5	34.4	28.3	2.3	75.1	24.9	0.8	10.4	14.7	33.1	23.8	17.1
201803	1744	희우정로1	CS200020	세탁소	1.72E+08	0	13.2	22.7	28	34.7	1.2	46.9	53.1	0.5	11.3	27.9	34	18.1	8.2
201803	1744	희우정로1	CS200019	여관업	43409529	15.3	22	17.8	16.8	16.2	12	55.9	44.1	0.8	27.8	16.5	23.7	19.4	11.7
201803	1744	희우정로1	CS200018	피부관리	86685422	1.7	4	22	31.2	37.4	3.9	38.3	61.7	0.8	12.8	12.8	20.1	26.3	27.3
201803	1744	희우정로1	CS200017	네일숍	17749383	0	0	15	26.6	55.1	3.3	6.3	93.7	0	33.9	20.6	24.6	8	13
201803	1744	희우정로1	CS200016	미용실	4.28E+08	0	2.9	22	30.5	42.8	1.8	48.2	51.8	2.2	16	22	29.4	18.7	11.6
201803	1744	희우정로1	CS200014	헬스클럽	1.72E+08	0	10.4	11.9	20.8	37.8	19.1	46.8	53.2	1.5	25.8	24	27.6	13.9	7.2
201803	1744	희우정로1	CS200013	골프연습장	15985068	0	17.9	10.3	48.7	17.9	5.1	71	29	0	6.7	40	16.7	36.7	0
201803	1744	희우정로1	CS200012	당구장	28290960	13.3	0	3.1	10	34.8	38.8	95.4	4.6	4.9	14.3	16.4	26.5	27.1	10.7
201803	1744	희우정로1	CS200011	PC방	31355344	24.2	7.5	6.6	13.1	23.7	25	86	14	6.5	67	21.9	3.1	1.1	0.3
201803	1744	희우정로1	CS200010	노래방	3.26E+08	45.4	0.9	0	1.3	11.9	40.5	73.9	26.1	3.1	23.9	23.2	25.8	15.7	8.2

신규 업종 추천 서비스- 전체업종 매출 상관관계 분석

- 서울시 공목상권 프로파일링 정보(상권배후지-추정매출)의 1744개 상권코드별 업종별 매출 집계
- 매출 데이터가 현저히 부족한 2개 업종(PC방, 부동산)은 분석 배제
- 1000개 미만의 데이터가 있을 경우 feature 제거 (그 외 결측치 발생시 0으로 대체)

	A	C	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1		화장품	편의점	슈퍼마	자동차	자동차	세탁소	여관업	피부관	네일샵	미용실	헬스클럽	골프연	양구장	노래방	인테리어	보육시	일반의
3	휴대폰	0.07	0.16	0.06	0.06	0.30	0.03	0.03	0.01	0.04	-0.01	0.05	-0.01	0.10	0.12	0.04	-0.11	-0.02
4	화장품	1.00	0.65	0.24	0.12	0.04	0.26	0.24	0.66	0.49	0.72	0.50	0.17	0.62	0.46	0.18	-0.10	0.35
5	호프간이주점	0.71	0.74	0.20	0.07	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
6	헬스클럽	0.50	0.46	0.26	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
7	한의원	0.53	0.34	0.17	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
8	한식음식점	0.74	0.78	0.18	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
9	피부관리실	0.66	0.53	0.17	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
10	편의점	0.65	0.65	0.31	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
11	패스트푸드점	0.65	0.72	0.31	0.06	0.06	0.16	0.22	0.57	0.54	0.72	0.39	0.13	0.69	0.67	0.19	-0.12	0.34
12	패션잡화	0.49	0.49	0.13	0.04	0.24	0.01	0.11	0.32	0.41	0.48	0.19	-0.06	0.42	0.39	0.06	-0.16	0.17
13	컴퓨터판매수리	0.04	0.12	-0.04	-0.03	0.05	0.10	0.04	0.04	0.01	0.04	0.08	0.14	0.02	-0.04	-0.01	-0.03	-0.03
14	커피음료	0.72	0.78	0.18	0.11	0.14	0.24	0.31	0.63	0.46	0.71	0.48	0.18	0.57	0.41	0.17	-0.19	0.37
15	치킨집	0.57	0.65	0.42	0.12	0.19	0.36	0.13	0.53	0.43	0.62	0.51	0.10	0.60	0.55	0.14	0.18	0.30
16	치과의원	0.62	0.49	0.28	0.14	-0.01	0.28	0.16	0.56	0.54	0.61	0.51	0.09	0.60	0.43	0.12	0.06	0.48
17	중국집	0.43	0.46	0.09	0.09	0.16	0.14	0.18	0.40	0.23	0.31	0.25	0.14	0.39	0.55	0.11	-0.12	0.23
18	제과점	0.63	0.67	0.35	0.10	0.00	0.35	0.38	0.57	0.49	0.65	0.50	0.12	0.50	0.35	0.07	-0.02	0.36
19	정육점	-0.02	0.04	0.16	-0.01	0.01	0.13	-0.02	-0.06	0.08	0.05	-0.02	0.06	0.12	0.05	0.06	0.23	-0.03
20	자동차수리	0.04	0.19	0.12	0.31	1.00	0.12	-0.01	0.09	0.06	-0.04	0.08	0.14	0.15	0.20	0.31	0.01	0.02
21	자동차미용	0.12	0.15	0.05	1.00	0.31	0.17	-0.03	0.14	0.16	0.16	0.12	0.10	0.14	0.16	0.16	0.06	0.03
22	임시보습학원	0.10	0.09	0.13	0.14	0.08	0.17	0.02	0.15	0.03	0.16	0.16	0.11	0.05	-0.01	0.03	0.04	0.04
23	일식집	0.70	0.68	0.16	0.17	0.11	0.23	0.23	0.65	0.49	0.70	0.47	0.17	0.66	0.60	0.23	-0.10	0.33
24	일반의원	0.35	0.31	0.05	0.03	0.02	0.08	0.14	0.31	0.23	0.24	0.28	0.07	0.25	0.17	0.11	-0.10	1.00
25	인테리어	0.18	0.17	0.05	0.16	0.31	0.17	0.02	0.27	0.14	0.14	0.08	0.06	0.17	0.24	1.00	0.00	0.11
26	이음점	0.55	0.59	0.13	0.07	0.11	0.11	0.22	0.47	0.40	0.56	0.27	0.07	0.47	0.38	0.13	-0.11	0.18

편의점 매출(열 값), 패스트푸드 매출(행 값)의
상관관계가 0.72라는 의미

신규 업종 추천 서비스- 전체업종 매출 상관관계 분석 결과 해석



특정 업종과 매출 관련성이 높은 업종이
(상관계수 0.7 이상)
다수 존재한다는 사실을 발견 !

신규 업종 추천 서비스- 업종별 1:1 매출 상관관계 분석 결과 해석

- Ex) 커피음료 업종과 한식음식 업종이 매출 관련성이 높다는(상관계수 0.7 이상) 것의 의미

→ '한식음식점 매출액이 큰 곳의 커피음료 매출액 또한 비례하여 크다' 또는
'커피음료 매출이 증가 할 때 한식음식점 매출도 비례하여 증가 할 수 있다' 고 볼 수 있다.

A 상권

한식음식점 매출이 100 이면
커피음료 매출이 100 근방일 가능성이 높다

B 상권

한식음식점 매출이 50 이면
커피음료 매출이 50 근방일 가능성이 높다

신규 업종 추천 서비스- 커피업종분석, 활용방안 및 기대효과

상위 10개 업종		상권코드	상권개수	0.89	0.87	0.82	0.79	0.78	0.74	0.73	0.71	0.71	0.68	가중평균매출액	가중평균매출액 대비 커피매출액 비율
중구골목상권		상권코드	커피음료(매출액)	한식음식점	양식집	분식집	일식집	편의점	호프간이주점	패스트푸드점	미용실	제과점	피부관리실	가중평균매출액	가중평균매출액 대비 커피매출액 비율
명동	명동길	564	39,973,211,210	24,260,038,977	6,607,533,508	3,517,835,847	4,442,979,091	3,776,901,976	2,877,084,499	1,439,476,584	1,262,993,753	1,133,383,068	1,204,105,275	42,380,630,825	0.943
동입-동역사	퇴계로 56길	1570	1,144,773,644	9,008,195,291	2,089,970,369	750,700,661	828,127,215	2,241,178,873	631,314,367	318,133,556	91,829,898	1,612,671,320	104,794,403	14,834,348,779	0.077
충무로	퇴계로 46길	1567	1,388,690,915	10,290,164,266	1,697,901,347	681,757,883	705,767,443	1,345,464,786	654,128,736	523,876,130	84,521,623	979,356,452	13,554,082	14,432,535,694	0.096
충무로	퇴계로 32길	1566	1,112,067,202	9,208,840,518	918,708,775	822,772,965	873,980,309	1,448,278,080	868,840,860	503,765,612	77,252,967	330,552,486	42,774,275	12,819,239,385	0.087
충무로	퇴계로 28길	1565	1,389,741,969	8,349,322,064	779,710,619	817,424,290	779,710,619	1,354,021,962	752,687,685	517,962,713	184,485,220	433,559,990	89,973,063	11,886,736,762	0.117
동입-악수	동호로 15길	487	636,428,317	5,442,359,946	542,375,329	585,173,680	652,119,860	1,495,084,300	570,313,968	422,872,560	413,437,258	1,327,773,904	163,005,513	9,554,582,728	0.067
명동인근	퇴계로 20길	1564	1,078,792,079	4,823,061,961	2,008,825,958	1,002,684,055	546,537,540	1,087,420,363	224,786,251	465,661,413	670,357,128	276,843,470	267,019,583	9,502,717,592	0.114
동입-악수	동호로 14길	486	640,675,578	5,309,787,292	489,467,701	572,687,502	664,649,784	1,515,062,938	501,602,763	425,465,079	325,118,962	1,261,328,253	153,605,151	9,240,578,340	0.069
동입-악수	동호로 11길	485	549,483,266	4,513,152,927	1,002,868,481	589,743,376	590,819,252	1,510,103,123	475,151,558	365,769,740	469,787,728	728,348,040	166,585,760	8,599,997,672	0.064

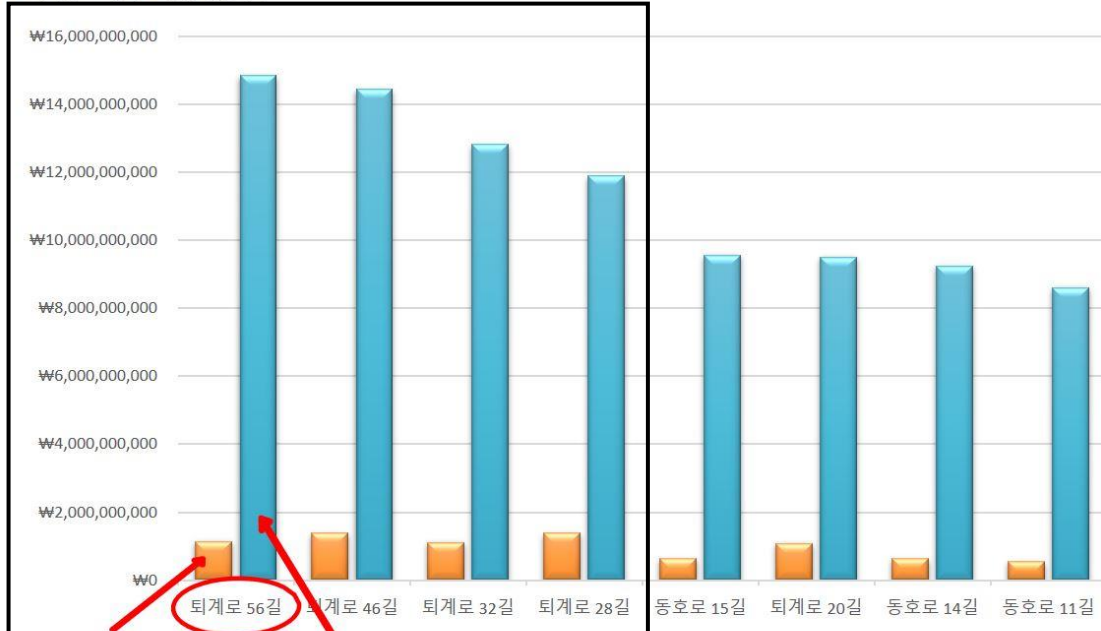


가중평균 매출액을 바탕으로 일정 액수 이상 상권이
형성되어 있는 지역 중,
가중평균매출액 대비 커피음료매출액이 가장 작은
지역 선정



신규 업종 추천 서비스- 커피업종분석, 활용방안 및 기대효과

■ 커피음료 매출 ■ 업종연관지수



커피음료의
현재 매출 규모

가중평균 매출액

커피업종 분석결과, **퇴계로56길**이
업종 관련성 지표는 중구 내 1위로 최상위이며
커피음료 매출은 약 11 억원으로 상대적으로 낮아
향후 성장성이 높다고 볼 수 있음

가중평균매출액	가중평균매출액 대비 커피매출액 비율
42,380,630,825	0.943
14,834,348,779	0.077
14,432,535,694	0.096
12,819,239,385	0.087
11,886,736,762	0.117
9,554,582,728	0.067
9,502,717,592	0.114
9,240,578,340	0.069
8,599,997,672	0.064

즉, 해당 지역은 커피 전문점의 시장 규모가 더 커질
가능성이 있으므로, 신규 개업을 추천함

창업 위험지수 예측 서비스- 데이터 전처리



서울시 공공데이터를 활용하여 프로젝트를 진행하였고
각 데이터시트에 담겨진 속성들을 상관 명과 서비스 업종 별로 통합.

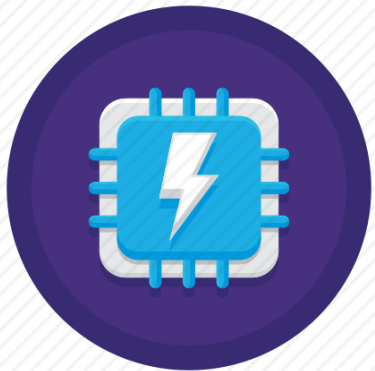
창업과밀지수



위험도를 보여주기 위한 타겟 변수로 '과밀지수' 속성 사용.

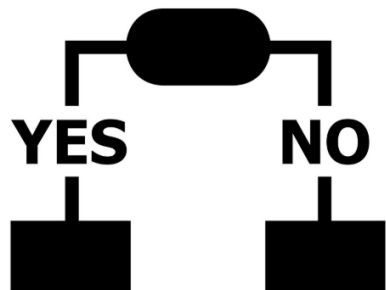
1742개의 상관 X 11개의 업종 ➡ 1.8만 여개 인스턴스 & 64개의 속성

창업 위험지수 예측 서비스- 특징 선택



학습 및 테스트에 64개의 속성 모두 사용하는 것?

시간 & 컴퓨팅 파워 **Efficiency** ↓ **Performance** ↓



Decision tree

‘결정 나무 알고리즘’을 활용한 Feature selection을 통해
데이터를 가장 잘 분류하는 9개의 속성을 추출

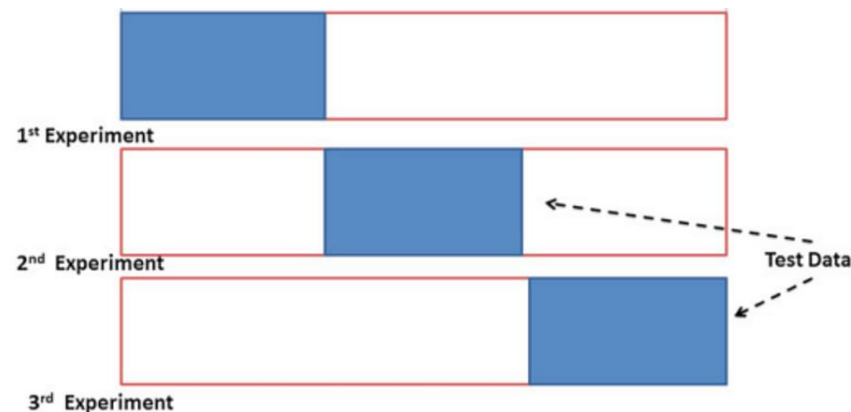
‘과밀 지수’ 속성을 포함한 10개의 속성으로 학습 진행.

창업 위험지수 예측 서비스- 모델링

Algorithms used for learning →

SVM
XGBoost
Neural net
Random Forest

Performance
Validation →



창업 위험지수 예측 서비스- 성능평가

SVM – 71.1%

XG Boost – 74.2%

Neural net – 71.7%

Random Forest – 70.4%

*Confusion matrix를 통해 창업을 진행하기에 실제로 위험하지만
안전하다고 예측한 부분에 집중*

창업 위험지수 예측 서비스- 활용방안 및 기대효과

- 복잡한 입지선정 과정의 간소화
- 폐업의 가장 큰 원인인 '경쟁'을 최소화
- 상권 별 결제 데이터 추가 확보를 통한 향후 발전 가능성 농후
- 수평적 확장을 통한 국내 전 지역 대상 옴니 서비스 영역 확보