



자영업 폐업률 예측 모델

곽민지

문제제기: 자영업 포화 시장



- 서울의 변화가 : 한집 건너 한집이 치킨집
- 맛있었던 그집... 일년 후에 찾아가보면? 망했다!
- “5년 만에 최대, 자영업자 붐과 가속화! 치킨 커피 요식업 과잉 상태 심각하다”
- 치킨집 등 자영업자 1인당 빚 4300만원
- 너도나도 치킨, 김밥집, 작년 자영업자 8만 9000명 눈물의 폐업

인사이트

지역(상권)
+
업소의 유형
+
동종 업소의
밀집도



2년 이내에
폐업할 확률
(0~100점)

비즈니스 액션

- ✓ 모델을 통해 분석한 결과
- ✓ 2년 내에 해당 사업이 폐업할 확률을 알 수 있다

✓ 묻지마 창업 방지 효과!

확보 가능한 데이터

서울시 우리마을가게 상권분석 서비스
(golmok.seoul.go.kr)



낙성대로 골목상권

- 상권유형 준주거지역
- 상권면적 35410 m²
- 최대업종 한식음식점41.4%
- 점포수 72 개 (전체:174개)
- 관련키워드 한식음식점, 관악구, 낙성대동, 낙성대로

상권 테이블

상권 번호

배후지 가구수
상권 내 유동인구
배후지 직장인구

매출 테이블

상권 번호

업종소분류코드

평균 매출
최소 매출
최대 매출

공공 데이터 포털
(data.go.kr)

☐ ZIP 상가업소정보 (2016년 9월)
다운로드 상세정보 오류신고 ★

☐ ZIP 상가업소정보 (2016년 12월)
다운로드 상세정보 오류신고 ★

☐ ZIP 상가업소정보 (2016년 3월)
다운로드 상세정보 오류신고 ★

☐ ZIP 상가업소정보 (2017년 6월)
다운로드 상세정보 오류신고 ★

상가 정보 테이블

상가 업소 번호

업종 소분류 코드
경도
위도
프랜차이즈 여부

상권정보업종 코드집

업종 중분류 명

업종 소분류 코드

업종 소분류 명

D	E	F
업종중분류명	업종소분류코드	업종소분류명
대중목욕탕/휴게	F09A04	휴게텔
한식	Q01A01	한식/백반/한정식
분식	Q04A01	라면김밥분식
분식	Q04A02	국수/만두/칼국수
일식/수산물	Q03A11	갈치/생선구이
중식	Q02A00	중국음식/중국집
중식	Q02A01	탕수육전문점
일식/수산물	Q03A02	음식점-일식

포물레이션

상권 테이블

상권 번호

배후지 가구수
상권 내 유동인구
배후지 직장인구

Clustering

군집 테이블

Cluster#

Avg(배후지 가구수)
Avg(상권 내 유동인구)
Avg(배후지 직장인구)

- ✓ 상권: 위도, 경도를 이용하여 어느 상권에 속하는지를 알아낸다.
- ✓ 경쟁지수: 근접한 동일 업종 업소의 개수 지수를 계산
 - 1) 동일 상권 = $x1$ (가중치 $\alpha = 1$)
 - 2) 50m 이내 = $x2$ (가중치 $\beta = 10$)
 - 3) 10m 이내 = $x3$ (가중치 $\gamma = 100$)
 '경쟁지수' = $x1 * \alpha + x2 * \beta + x3 * \gamma$
- ✓ 폐점유무: 2년이상 폐점하지 않은 경우 0, 폐점한 경우 1

상가 정보 테이블

상가 업소 번호

업종 분류

.....

위치정보

프랜차이즈 정보

상권 번호

경쟁 지수

폐점 유무 (출력변수)

상가 + 상권

상가 업소 번호

상권 번호

업종 분류

배후지 가구수 평균

위치정보

상권 내 유동인구 평균

경쟁 지수

배후지 직장인구 평균

폐점 유무

Cluster#

전처리	군집 테이블과 상가 테이블을 merge 후 상권별로 '경쟁지수'와 '폐점유무'의 mean값을 구한다.
입력변수	[폐점유무]를 제외한 모든 변수 X
출력변수	[폐점유무] y
알고리즘	Random Forest Decision Tree Regression
모델 학습	각 상권 cluster X 업종 분류에 대해 세 가지 모델을 학습. 가장 예측이 잘 되는 모델을 채택

모델 적용 방법



1. 입점하고자 하는 위치 또는 모델로 군집 찾기

2. 원하는 업종

3. 근처의 경쟁 업소 -> 경쟁지수 구하기

3. 모델에 test

4. 폐업률의 기댓값을 확인

한계점/문제점

- ✓ 상점이 폐업한 이유를 정확히 알 수 없다.
- ✓ 상권에 대한 정보는 비교적 다양한 것에 비해, 상점에 대한 변수는 너무 빈약하다. 기껏해야 상점의 층. 사장님이 얼마나 능력 있는가는 무시된다.
- ✓ 또한, 매출에 대한 정보가 부족하다. 주로 관심있는 것은 매출이 부진하여 폐업한 상점인데, 매출이 부진하지 않더라도 다른 사정에 의해 폐업한 상점까지 데이터에 포함된다.