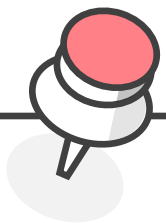


소상공인 활성화를 위한 상권 분석 서비스

인공지능학과 김재경

인공지능학과 맹준영

인공지능학과 신유승



Contents

- 프로젝트 개요

- 주제 선정 배경

- 데이터 구축

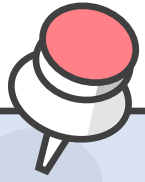
- 사용 데이터
- 분석 대상

- 아이디어 제안

- 예측 모델
- 모델링 과정

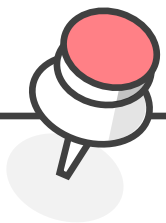
- 결론 및 제언





프로젝트 개요





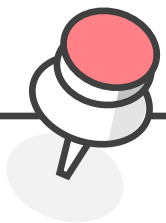
프로젝트 개요

주제 선정 배경

우리나라 자영업자 비중은 2021년 기준 **24.6%**로 OECD 회원국 중 6위로 높은 수준임.

자영업자 비율이 높은 만큼 경쟁 또한 치열하기 때문에 충분한 수익을 얻기 위해서는 **상권 분석이 중요**함.

젊은 사람들이 많이 모여들고 어떤 상권에 유동 인구가 많은지 객관적인 근거자료 기반으로 창업을 하는 방법으로 서울시에서 제공하는 '**서울시 우리마을가게 상권분석 서비스**'가 있음.



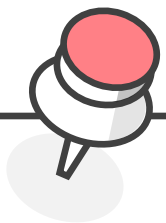
프로젝트 개요

주제 선정 배경



'서울시 우리 마을 가게 상권분석 서비스'는 다양한 빅데이터를 분석하여 원하는 상권을 찾아주고, 다양한 정보를 알기 쉽게 설명하는 맞춤형 리포트는 제공하는 서비스.

플랫폼을 통해 창업을 준비하는 시민과 기존 영세 소상공인이 누구나 쉽게 정확하고 유용한 빅데이터 기반의 상권 별 생존률, 평균 매출액, 임대 시세 등 다양한 상권 및 경영 정보를 무료로 제공 받을 수 있음.



프로젝트 개요

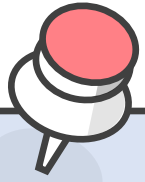
주제 선정 배경

하지만 제공된 상권에 대한 분석 정보를 토대로 개인이 직접 판단해야 하는 어려움과 IT 서비스를 다루기 어려워하는 부모 세대 연령층의 소상공인에게는 서비스 이용에 제약이 있음.

또한 현재 상권에 대한 분석이 가능하기 때문에 앞으로의 상권에 대한 전망은 예측할 수 없음.

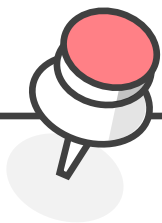
따라서 소상공인 활성화를 위한 방안으로, 유동인구, 매출 등의 빅데이터 기반 예측 알고리즘을 활용한 AI 모델을 통해 개인이 원하는 지역을 선택하면 해당 지역 상권이 앞으로 좋을지 안좋을지 종합적으로 판단해주는 쉽고 간편한 **"로케이션 원클릭 서비스 (Location one-click service)"**를 제안함.

이러한 AI 서비스를 통해 IT 분석 서비스에 어려움을 겪는 소상공인 세대에게도 한번의 클릭만으로 미래 경영 환경에 대비할 수 있도록 손쉬운 상권 분석이 가능해질 수 있을 것이라 생각함.



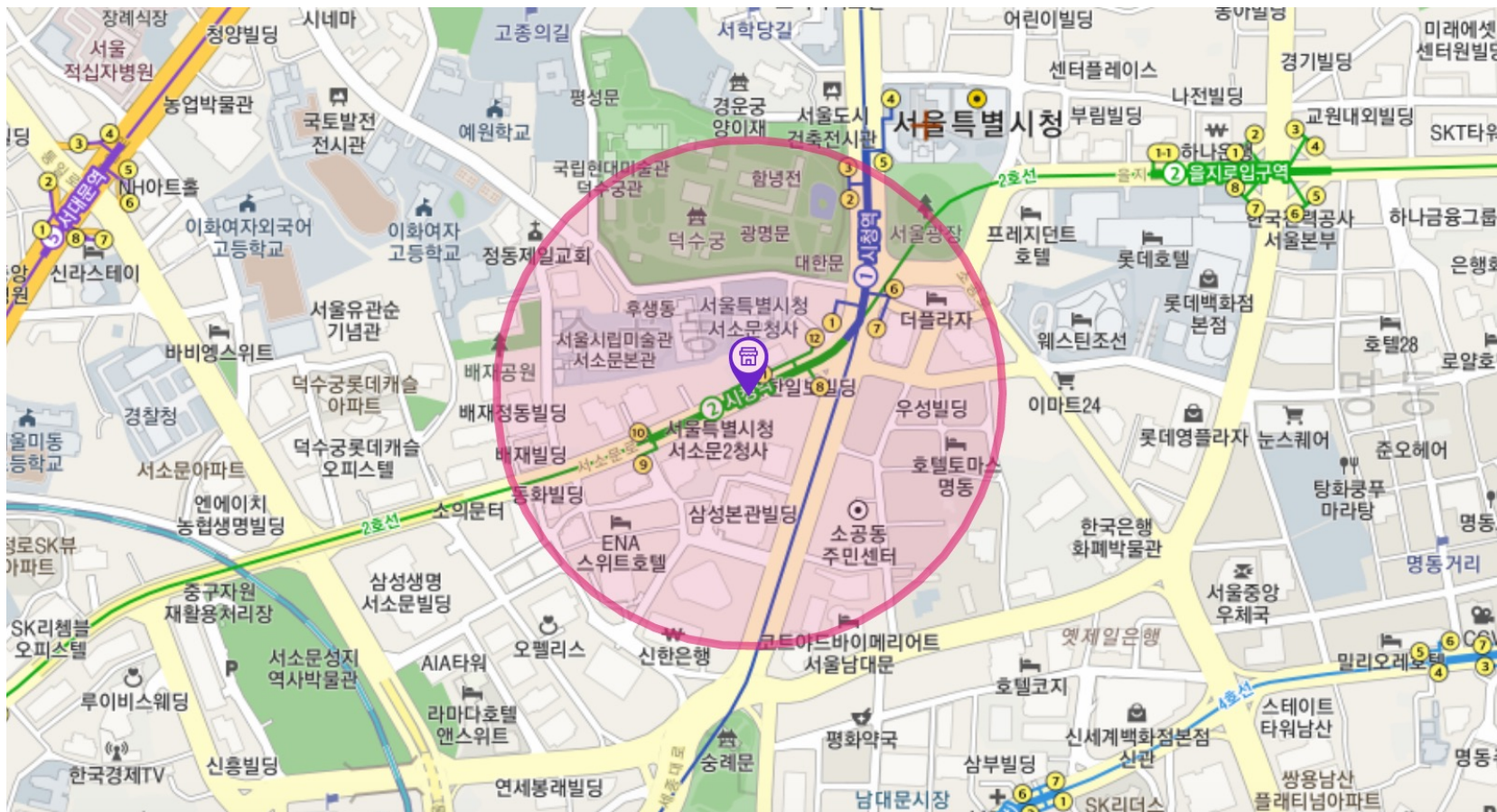
데이터 구축

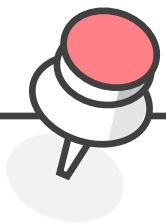




데이터 구축

2022년 2분기 기준 특정 지역 (서울시내 지하철역 + 서울시 홈페이지에 등재된 주요 상권 리스트)을 사용하여 데이터 구축

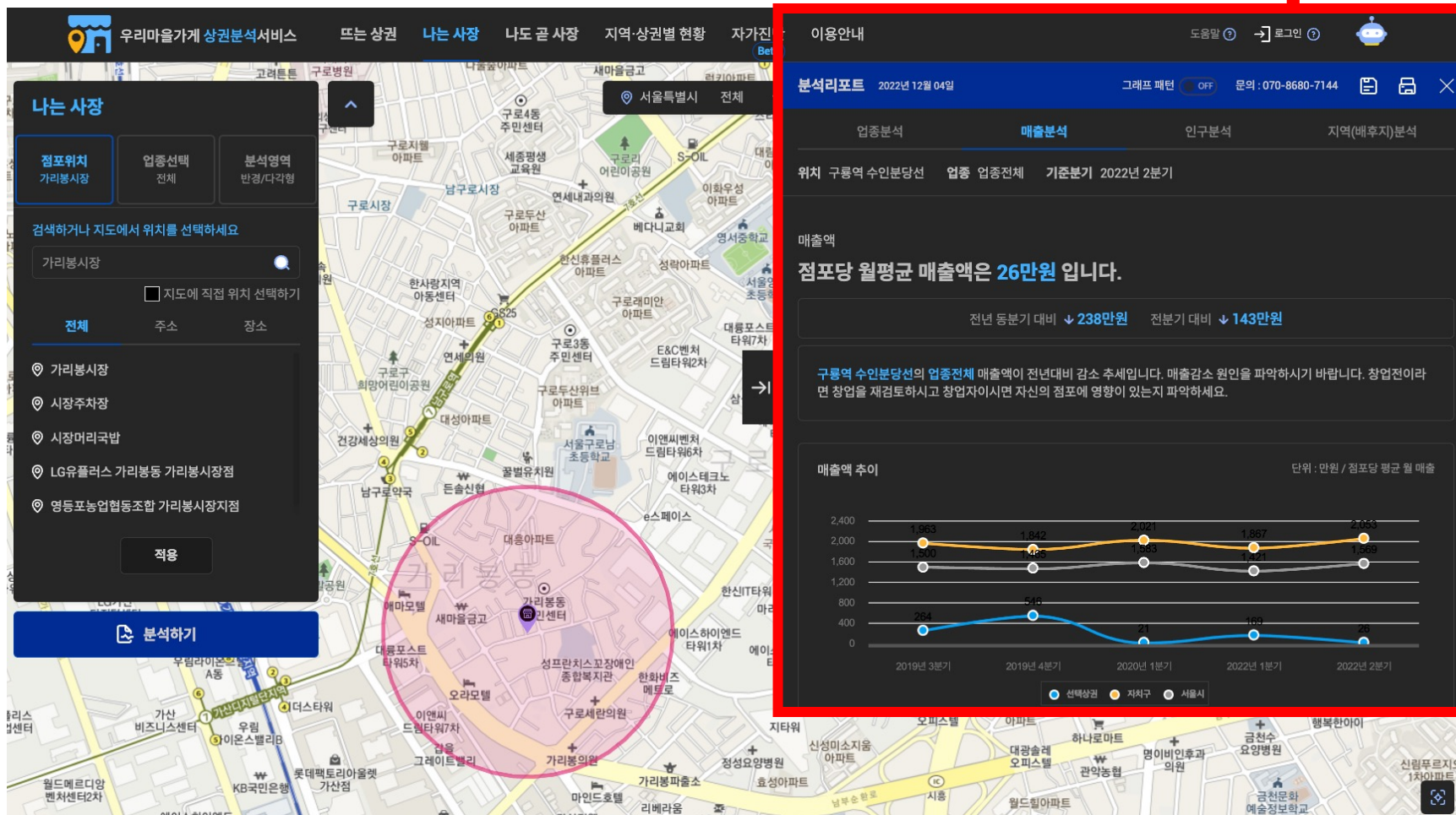


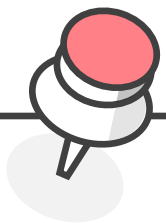


데이터 구축

해당 지역 기준 반경 300m의 데이터 크롤링(Crawling)을 통한 수집

Data crawling

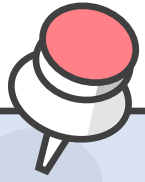




데이터 구축

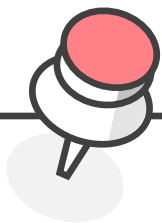
본 데이터 수집을 통해 얻은 데이터의 feature 종류는 총 36개

- 점포 수, 신생기업 생존률, 평균영업기간, 개업 수, 폐업 수, 증가하는 업, 점포 당 월 평균매출액, 월 평균 매출 건 수, 매출 많은 요일, 남자비율, 여자비율, **매출 증감, 전분기매출**, 일 평균 유동인구 수, **유동인구밀도, 유동인구 증감**, 주거인구 수, 직장인구 수, 가구세대 수, 부동산 평수 대, 부동산가격대, 관공서 수, 유통점수, 금융기관 수, 영화관 수, 병원 수, 숙박 수, 학교 수, 교통 시설 수, 소득분위 수, 행정동 소득 분위 수, 소비 트렌드, 임대료수준
- 해당 데이터를 통해 학습하고자 하는 'Label'은 매출 증가분과 유동인구 증가분을 적절히 고려하여 만들었음.
- $(10 \times \text{매출 증가 비율}) + (1 \times \text{유동인구 증가 비율})$ 을 label 값으로 선정함.



아이디어 제안





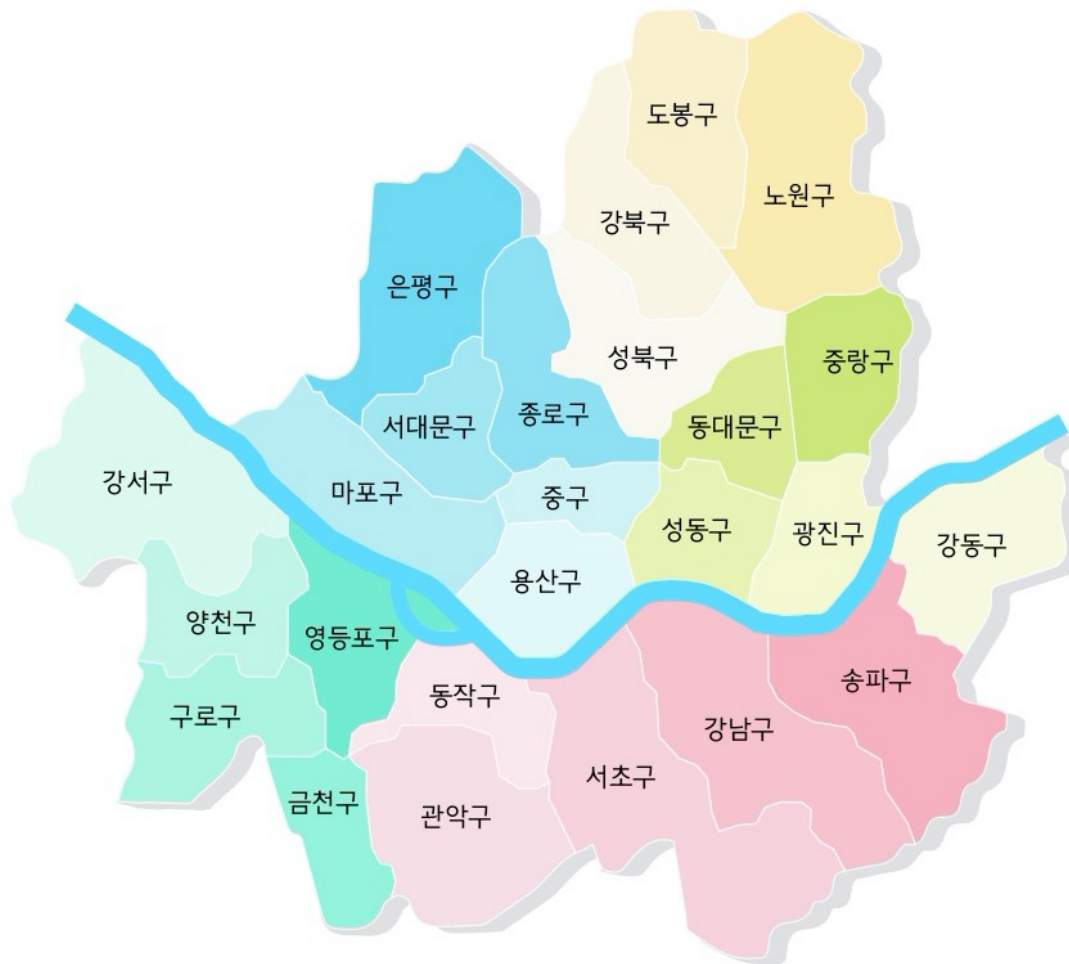
아이디어 제안

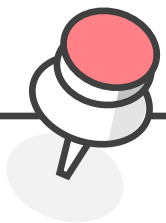
분석 대상 : 서울시 지역상권

지역상권의 사전적 정의

일정한 지역을 중심으로 재화와 용역의 유통이 이루어지는 공간적 범위.

일정한 지역을 중심으로 재화와 용역의 유통이 이루어지는 공간적 범위.
협회의 상권은 상품이 유통되는 일정한 지역을 의미하며, 상세권(商勢圈)이라고도 함.



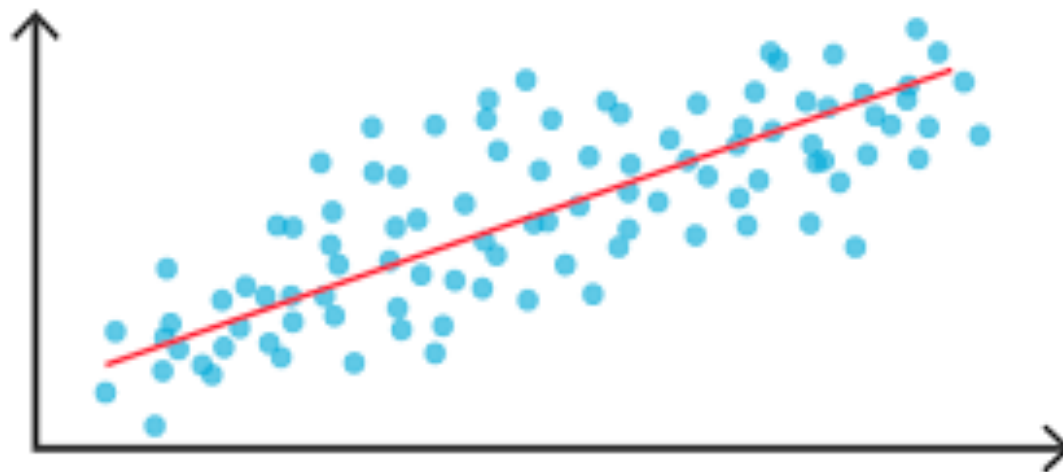


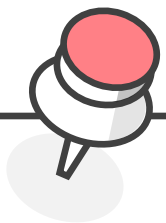
아이디어 제안

Regression analysis

회귀분석 (Regression analysis) 이란, 독립 변인이 종속 변인에 영향을 미치는지 알아보고자 할 때 실시하는 분석 방법.

단순 선형 회귀분석은 독립변수 X(설명변수)에 대하여 종속변수 Y(반응변수)들 사이의 관계를 수학적 모형을 이용하여 규명하는 것. 규명된 함수식을 이용하여 설명변수들의 변화로부터 종속변수의 변화를 예측하는 분석.





아이디어 제안

DecisionTree

여러 가지 규칙을 순차적으로 적용하면서 독립 변수 공간을 분할하는 분류 모형 알고리즘으로 내부적인 기준을 통해 참/거짓을 판별한 후 노드가 뺀어 나가는 형태

GradientBoosting

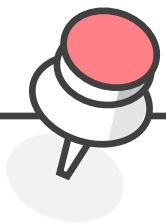
Gradient(또는 잔차(Residual))를 이용하여 이전 모형의 약점을 보완하는 새로운 모형을 순차적으로 적합한 뒤 이들을 선형 결합하여 얻어진 모형을 생성하는 지도 학습 알고리즘

XGBoost (Extreme Gradient Boosting)

Boosting 기법을 활용한 Gradient Boost 알고리즘에서 병렬 학습이 지원되도록 구현한 라이브러리

AdaBoost

분류 기반 기계학습 모형으로, 예측 성능이 조금 낮은 약한 학습기(weak classifier)를 다량 구축 및 조합하여 가중치 수정을 통해 좀 더 나은 성능을 발휘하는 하나의 강한 분류기(strong classifier)를 합성하는 방법의 알고리즘



아이디어 제안

SVR (Support Vector Regression)

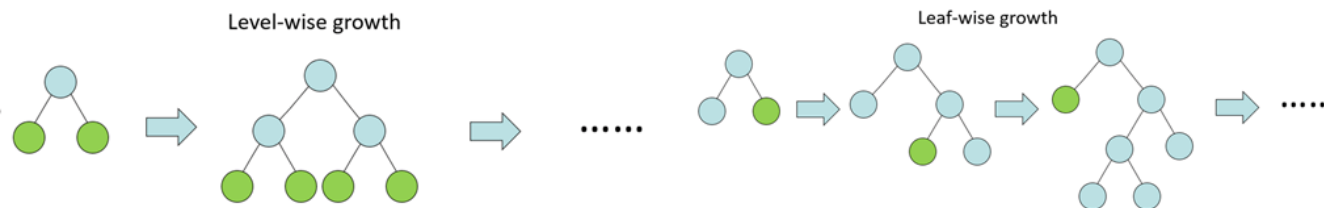
학습 데이터의 분류 예측에 사용되는 SVM을 ϵ -무감도 손실함수를 도입하여, 임의의 실수 값을 예측하도록 일반화한 알고리즘

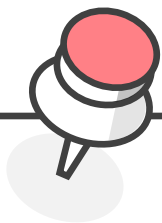
RandomForest

분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성한 다수의 결정 트리로부터 부류(분류) 또는 평균 예측치(회귀 분석)를 출력함으로써 동작하는 알고리즘

LightGBM (LGBM)

트리 기준 분할이 아닌 리프 기준 분할 방식을 사용. 트리의 균형을 맞추지 않고 최대 손실 값을 갖는 리프 노드를 지속적으로 분할하면서 깊고 비대칭적인 트리를 생성함으로써, 트리 기준 분할 방식에 비해 예측 오류 손실을 최소화하는 알고리즘

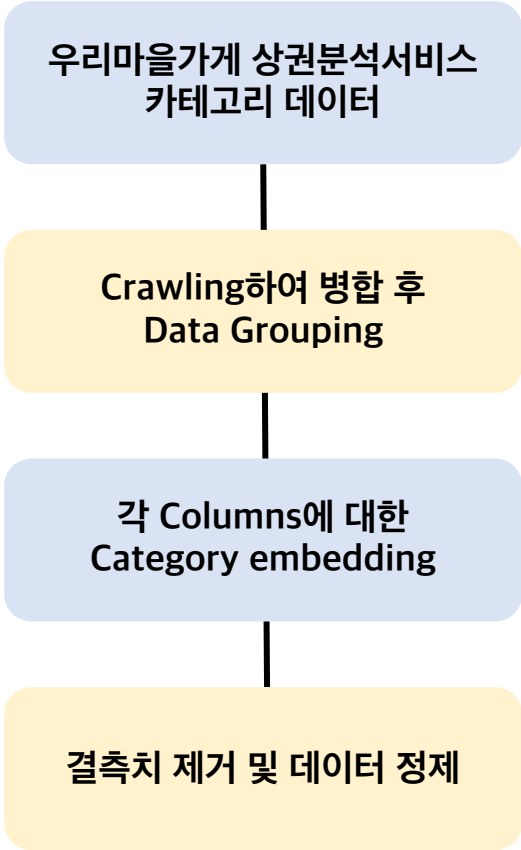




아이디어 제안

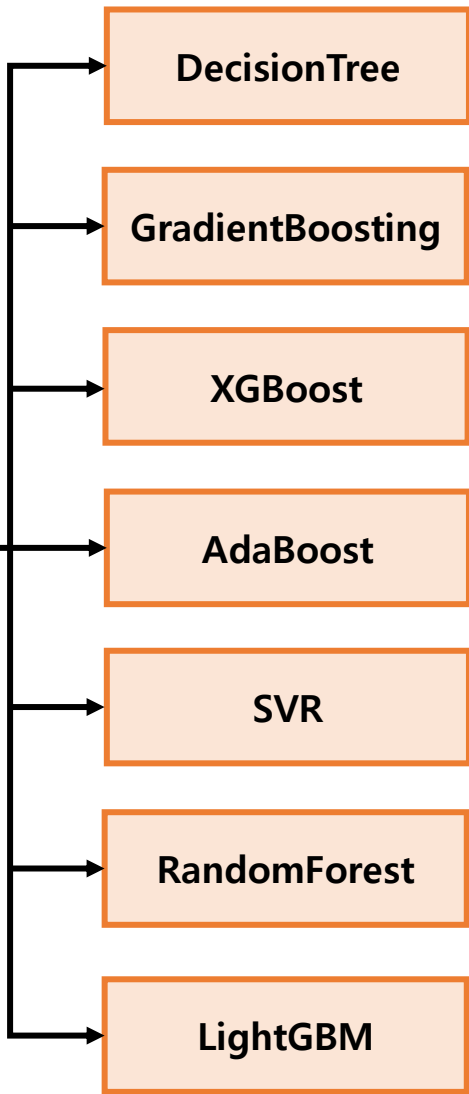
모델링 과정

Data preprocessing



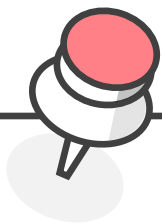
Input Data

Machine learning
regression algorithm



Model
Selection

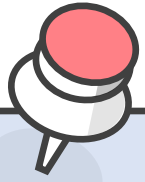
지역 상권 예측 결과
(Output)



아이디어 제안

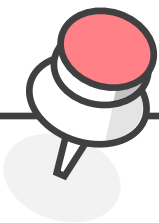
모델링 설명 및 평가

- 전체 1,371개 데이터 중 20%는 validation set으로 사용하였으며, 평가 지표는 MAE를 사용함.
- 각 모델에 따른 MAE 평가 결과, **RandomForestRegressor(0.0264)**가 가장 성능이 좋았음.
- Feature importance 추출 결과, 가장 영향력이 작은 '부동산평수대', '증가하는업', '요일', '부동산가격대'를 제거하여 성능 개선 → **(0.0253)**
- GridSearch를 사용하여 최적의 하이퍼파라미터 튜닝을 통한 성능 개선 → **(0.0245)**



결론 및 제언





결론 및 제언

- 소상공인 활성화를 위한 방안으로, 유동인구, 매출 등의 빅데이터 기반 예측 알고리즘을 활용한 AI 모델을 통해 개인이 원하는 지역을 클릭하면 해당 지역 상권이 앞으로 좋을지 나쁠지 종합적으로 판단해주는 쉽고 간편한 “로케이션 원클릭 서비스 (Location one-click service)”를 제안함.
- 로케이션 원클릭 서비스를 통해 IT 분석 플랫폼에 어려움을 겪는 소상공인 세대에게도 한번의 클릭만으로 미래 경영환경에 대비할 수 있도록 손쉬운 상권 분석이 가능해질 수 있을 것이라 생각함.
- 향후 더욱 고도화된 AI 모델을 통해 높은 예측 확률로 더 나은 원클릭 서비스를 제공하고, 더 나아가 전국의 지역구 및 업종에 따른 원클릭 서비스 확장을 기대함.

