



AI기반 식수예측 서비스 개발

키워드

ML

NLP

Python

Python Flask

Web



주요 내용

- 분석 배경
 - 미배식 되어서 나오는 음식물 쓰레기를 줄이는 방안 마련
 - 집단 급식소에서의 식수 예측은 영양사의 경험과 직관에 의한 예측이 주를 이루고 있음
- 분석 목적
 - 데이터 기반 식수 예측 모델링을 통해 예측의 정확성과 객관성을 보충하고 실무에서 바로 활용할 수 있도록 모델 배포 및 서비스 웹사이트 구현
- 분석 내용
 - 지자체에서 제공한 약 6년간의 식단표, 카드 데이터와 공공데이터 포털의 데이터를 활용하여 AI 기반의 집단급식소 식수 식수 인수 인원 예측
 - 비정형 데이터(식단표) 전처리를 통한 모델 성능 개선
 - Gradient Boosting, Random Forest, CatBoost 모델을 블렌딩하여 최종 앙상블 모델 도출
 - 웹 서비스 구현

분석 개요

분석 목적

- 미배식 되어서 나오는 음식물 쓰레기를 줄이는 방안 마련

분석 배경

- 집단 급식소에서의 식수 예측은 영양사의 경험과 직관에 의한 예측이 주를 이루고 있음
- 데이터 기반 식수 예측 모델링을 통해 예측의 정확성과 객관성, 보편타당성을 보충

분석 목표

- 지자체에서 제공한 식단표, 카드 데이터와 공공데이터 포털의 데이터를 활용하여 AI 기반의 집단급식소 식수 인수 인원을 예측
- 실무에서 바로 활용할 수 있도록 모델 배포 및 서비스 웹사이트 구현

분석 데이터

분석 데이터 목록 및 설명

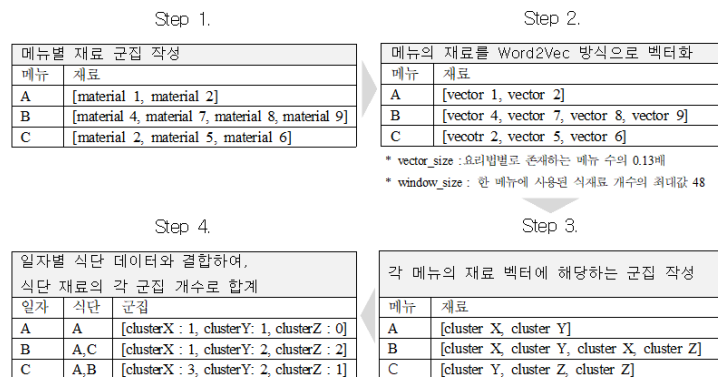
- 대구광역시 O구청으로부터 약 6년 간의 일일 식단표 및 구내식당 카드 결제 데이터 내역을 제공받음
- 선행 연구를 기반으로 제공받은 데이터 외에 추가적으로 식수 인원 예측에 활용한 데이터는 공공데이터 포털에서 수집

카테고리	활용데이터 목록	구성 내용	데이터 규모	출처
지자체 데이터	구내식당 일일 식단표	일자별 식단	1,448건	대구시 O구청
	메뉴별 식재료	메뉴, 메뉴별 식재료	1,065건	

	메뉴별 조리법	메뉴, 메뉴별 조리법	917건	
	일자별 구내식당 카드결제 내역	날짜, 성명, 수량, 단가	385,482건	
기상청 데이터	대구광역시 기상 데이터	일 최고/최저기온, 일 강수량 등	-	공공 데이터 포털

데이터 정제 방안

- 메뉴 특성에 따른 파생 변수 생성
 - 메뉴는 고객의 관점에서 식사 여부를 결정하는 중요한 기준
 - 동일한 메뉴임에도 다르게 표기되어있는 문제로 인해 범주형 변수로 그대로 활용할 수 없다고 판단
 - ex) 돈육김치찌개, 돼지김치찌개 → 동일한 메뉴지만 다르게 표기
 - 메뉴의 이름은 달라도 식재료, 조리법과 같은 메뉴의 본질적인 요소는 유사하므로 식재료와 조리법 정보를 활용한 파생변수를 통해 메뉴의 특성을 예측에 반영함
 - 조리법 정보 (국, 김밥/주먹밥, 김치, 무침/샐러드) 활용
 - 조리법을 기준으로 메뉴를 분류
 - 재료 정보 활용
 - 재료 종류가 약 1,200개 존재하여 워드임베딩과 군집화를 통해 유사한 재료들끼리 묶어주는 전처리 작업 수행
 - Word2vec에서 착안하여 메뉴의 재료들을 하나의 문장으로, 재료 각각을 단어로 취급하여 재료를 벡터공간에 임베딩
 - 임베딩된 데이터들의 유사도 계산은 통상적으로 텍스트마이닝에서 주로 활용되는 Spherical-Kmeans를 활용, 군집의 개수는 실루엣 계수 적용
 - 최종적으로는 재료별로 분류된 군집 정보와 조리법 정보를 파생변수로 생성하여 활용
 - <재료 데이터 활용 프로세스>

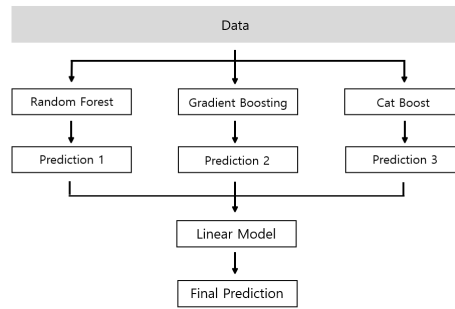


- 기타 외생 변수 생성
 - 선행 연구와 EDA를 기반으로 생성
 - 체감온도, 폭염 여부, 강우 여부, 적설 여부와 같은 기상 변수 생성
 - 요일, 월, 연도, 직전일 식수 인원, 연휴 전일 여부와 같은 시계열 변수도 생성
- 앞서 만든 변수들을 일자 기준으로 결합하여 최종 데이터셋을 생성

분석 내용

- 분석 방법
 - 여러 예측 모형(Gradient Boost, XGB, LGBM, CatBoost) 중 MSE와 MAE를 기준으로 가장 유의한 모델 3가지 (Gradient Boosting, Random Forest, CatBoost)를 블렌딩하여 최종 앙상블 모델 도출

<스태킹 앙상블 모형 구조>



- 스택킹 앙상블을 적용한 결과, MSE가 865.28로 기존 단일 모델들에 비해 약 8~12% 개선되었음.

Model	MSE	MAE	MAPE
Random Forest	1,056.61	24.85	0.083
GradientBoosting	972.41	24.12	0.081
CatBoost	943.59	23.88	0.079
Ensemble	865.28	22.26	0.074

분석 결과

• 메뉴명을 그대로 사용

Model	MSE
Random Forest	1,077.29
GradientBoosting	987.68
CatBoost	983.91
Ensemble	967.83

• 조리법, 요리법 정보를 활용

Model	MSE
Random Forest	1,056.61
GradientBoosting	972.41
CatBoost	943.59
Ensemble	865.28

한식 특성 상 메뉴의 종류가 많고 본질적으로 같은 메뉴라도 다른 이름으로 표기되는 경우가 있으므로 메뉴를 그대로 사용하였을 때 보다 조리법, 요리법 정보를 활용하여 메뉴 특성을 반영하였을 때 성능이 향상된 것을 확인할 수 있음

웹 서비스

- python flask를 사용하여 웹 서비스 구현
- 날짜, 메뉴를 입력하면 예측값을 도출