

2021 날씨 빅데이터 콘테스트 민간협력형 부문

날씨 최적화 식단 설계 및 추천 플랫폼 뭉게뭉게 냉장고

CONTENTS

1

공모배경

주제 선정 동기
분석 목표

2

활용 데이터 정의

데이터 세부내용
및 출처

3

데이터 처리 방안

데이터 전처리
설명변수 간 상관관계 분석

4

분석 기법 및 결과

예측 모델 생성
분석 및 모델링

5

활용 방안

플랫폼 개요
공공협력형 솔루션
민간협력형 솔루션

6

기대 효과

다각적 활용 및
확장 가능성

식품은 날씨와 기후의 영향을 크게 받는 상품군

날씨의 영향을 이용한 기업의 날씨경영

벌써부터 한여름 날씨... “올해 빙그레 매출 1조 넘을 것”

김광현 기자 입력 2021-06-09 14:38 수정 2021-06-09 14:56

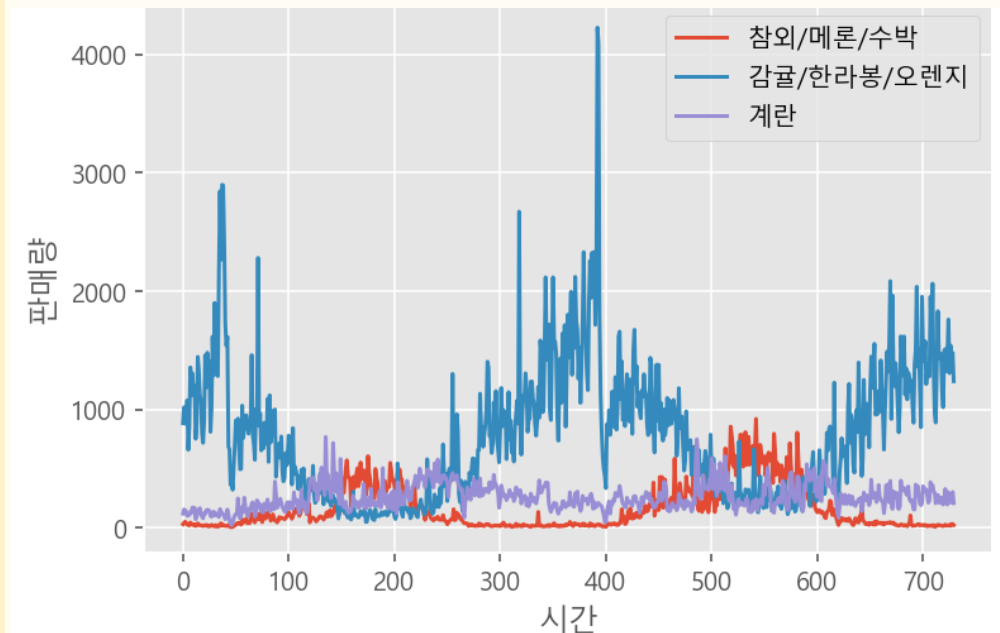
[날씨이야기]비오는 날 홈쇼핑에선 김치가 잘 팔린다

김치업체 증가집 날씨정보 활용해 수익성 개선
비오는 날에는 홈쇼핑으로 김치 판매해 매출 ↑
날씨정보 활용가치 최대 연 6조5000억원 달해

등록 2017-02-04 오전 9:00:00
수정 2017-02-04 오전 9:00:00

가 가

계절별 기후에 특화된 제철 식품



날씨변수의
소비에 대한 영향력 판별

기상요소에 따라 사람들의 소비패턴이 **유의미한 변화**를 보이는가?
날씨에 따라 사람들의 구매 경향이 **어떻게** 달라지는가?
어떤 상품군이 날씨의 영향에 대해 민감한 반응을 보이는가?

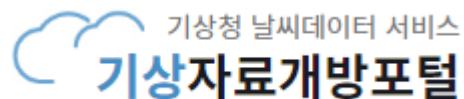
기상요소에 민감한
상품군의 선별

서로 다른 상품군을 어떻게 조합하였을 때
날씨의 영향을 긍정적으로 수용하면서
사람들의 선호도가 높도록 구성할 수 있을까?

데이터의 종합을 통한
맞춤형 서비스 제공

날씨에 민감한 식품군을 사람들의 선호도가 높은 순으로 나열하고
영양성분 데이터, 가격 데이터를 추가적으로 고려하여
가장 적합한 '**날씨 맞춤형**' 식단 조합 및 추천 **서비스**를 제공하자!

1. 기상 빅데이터



서울특별시 종관기상관측(ASOS) 데이터

	A	B	C	D	E	F	G	H
1	지점	지점명	일시	평균기온(°C)	최저기온(°C)	최저기온 시각(hhmm)	최고기온(°C)	최고기온 시각(hhmm)
2	108	서울	2018-01-01	-1.3	-5.1	739	3.8	1538
3	108	서울	2018-01-02	-1.8	-4.3	2348	1.8	1345
4	108	서울	2018-01-03	-4.7	-7.1	804	-0.4	1550
5	108	서울	2018-01-04	-4.7	-8.7	751	-0.7	1519
6	108	서울	2018-01-05	-3	-5.6	306	1.6	1530
7	108	서울	2018-01-06	-2.8	-6.8	628	2.9	1539
8	108	서울	2018-01-07	-0.8	-5.7	524	2.8	1501
9	108	서울	2018-01-08	1.3	-1.2	2352	4	1507
10	108	서울	2018-01-09	-4.2	-6.1	837	-1.2	1

Seoul2018 (2018.01.01 – 2018.12.31)

Seoul2019 (2019.01.01 – 2019.12.31)

서울특별시 황사관측(PM10) 데이터

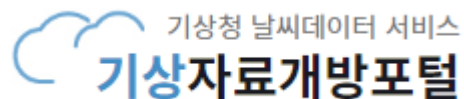
	A	B	C	D	E	F	G	H	I	J	K
1	지점	지점명	일시	일	미세먼지 농도(μg/m³)						
2	108	서울	2018-01-01	38							
3	108	서울	2018-01-02	32							
4	108	서울	2018-01-03	32							
5	108	서울	2018-01-04	40							
6	108	서울	2018-01-05	46							
7	108	서울	2018-01-06	40							
8	108	서울	2018-01-07	42							
9	108	서울	2018-01-08	41							
10	108	서울	2018-01-09	35							

dust2018 (2018.01.01 – 2018.12.31)

dust2019 (2019.01.01 – 2019.12.31)

※ 최종 예측에 사용한 데이터는 2020년 1,4,7,10월 15일 기상관측, 황사관측 데이터

1. 기상 빅데이터



전국 기상 특보 데이터

	A	B	C	D	E	F	G
1	날짜	태풍 주의보	태풍 경보	폭염 주의보	폭염 경보	한파 주의보	한파 경보
2	2018-01-01	0	0	0	0	1	0
3	2018-01-02	0	0	0	0	1	0
4	2018-01-03	0	0	0	0	1	0
5	2018-01-04	0	0	0	0	0	0
6	2018-01-05	0	0	0	0	0	0
7	2018-01-06	0	0	0	0	0	0
8	2018-01-07	0	0	0	0	0	0
9	2018-01-08	0	0	0	0	0	0
10	2018-01-09	0	0	0	0	1	0

alert (2018.01.01 – 2019.12.31)

태풍 주의보/경보, 폭염 주의보/경보, 한파 주의보/경보

기존 데이터에서 도출한 변수

체감온도 ($T = \text{기온}$, $V = \text{풍속}$)

$$T_w = 13.127 + 0.6215T - 13.947V^{0.16} + 0.486TV^{0.16}$$

불쾌지수 ($T = \text{기온}$, $RH = \text{상대습도}$)

$$THI = \frac{9}{5}T - 0.55(1 - RH) \left(\frac{5}{9}T - 26 \right) + 32$$

2. 날씨마루 제공 데이터



온라인 구매 이력 데이터

	A	B	C	D	E	F	G
1		buy2018_1.date	buy2018_1.sex	buy2018_1.age	buy2018_1.big_cat	buy2018_1.sm_cat	buy2018_1.qty
2	1	20180101	F	20	식품	가공란	37
3	2	20180101	F	30	식품	가공란	16
4	3	20180101	F	40	식품	가공란	9
5	4	20180101	F	50	식품	가공란	3
6	5	20180101	M	20	식품	가공란	13
7	6	20180101	M	30	식품	가공란	6
8	7	20180101	M	40	식품	가공란	6
9	8	20180101	F	20	냉난방가전	가열식 가습기	1
10	9	20180101	F	40	냉난방가전	가열식 가습기	3

buy2018_1 (2018.01.01 – 2018.06.30)

buy2018_2 (2018.07.01 – 2018.12.31)

buy2019_1 (2019.01.01 – 2019.06.30)

buy2019_2 (2019.07.01 – 2019.12.31)

소셜 데이터 (블로그, 커뮤니티, 인스타그램)

	A	B	C	D	E
1		sns2018_1.date	sns2018_1.big_cat	sns2018_1.sm_cat	sns2018_1.cnt
2	1	20180101	뷰티	기능성 링클케어 화장품	12.154295
3	2	20180101	뷰티	기능성 모공관리 화장품	36.000828
4	3	20180101	뷰티	기능성 아이케어 화장품	0.895782
5	4	20180101	뷰티	기능성 영양보습 화장품	14.868175
6	5	20180101	뷰티	기능성 트러블케어 화장품	48.819391
7	6	20180101	뷰티	기능성 화이트닝 화장품	23.016522

sns2018_1 (2018.01.01 – 2018.06.30)

sns2018_2 (2018.07.01 – 2018.12.31)

sns2019_1 (2019.01.01 – 2019.06.30)

sns2019_2 (2019.07.01 – 2019.12.31)

※ 음료, 건강식품을 제외한 식품 카테고리 데이터 사용 : 쌀/잡곡, 과일/견과, 채소, 축산물, 수산물, 김치

3. 영양성분 & 가격 데이터

	A	B	C	D	E	F	G	H
1	big_cat	small_cat	serving size(g)	calorie(kcal)	carbohydrate(g)	protein(g)	fat(g)	price(average)
2	쌀/잡곡	쌀	200	1.52	0.332	0.03	0.001	3.18
3	쌀/잡곡	잡곡	200	1.5	0.338	0.028	0.0038	3.73
4	과일/견과	견과류 땅콩	20	5.2	0.184	0.257	0.426	9.6
5	과일/견과	견과류 캐슈넛	20	5.76	0.267	0.198	0.476	13.5
6	과일/견과	포도/거봉/체리	100	0.67	0.1761	0.0083	0.0019	16.8
7	과일/견과	미숫가루/곡물가루	50	4.15	0.8	0.1	0.06	10
8	과일/견과	견과류 호두	20	6.88	0.07	0.155	0.7199	13.2
9	과일/견과	혼합 견과	20	1.39	0.09	0.03	0.1	32.9
10	과일/견과	견과류 피스타치오	20	5.84	0.2082	0.2599	0.489	27.1
11	과일/견과	견과류 잣/은행	20	4.165	0.3105	0.1032	0.304	134

영양성분 데이터 출처



<https://www.foodsafetykorea.go.kr/fcdb/>

가격 데이터 출처



2021 eCommerce 매출액 기준 1위
& 소비자 최대 이용 플랫폼 기준

주 이용 모바일 쇼핑 앱



남 (438)	54.8	48.6	32.0	27.9	16.9	13.5	17.6	8.9
여 (470)	54.7	56.4	24.0	24.5	22.3	20.6	7.4	14.0
20대 (301)	57.1	60.1	18.3	19.6	23.6	19.9	6.3	10.6
30대 (305)	57.0	53.8	29.5	25.6	20.3	18.0	11.8	10.8
40대 (302)	50.0	44.0	35.8	33.1	15.2	13.6	18.9	13.2

[Base: 3개월 이내 해당 보기 모바일 쇼핑 앱 이용자, N=908, 순위형 응답, %]
*연도별 음영: 전체 평균 대비 +4%p 이상인 데이터

오피서베이 모바일 쇼핑 트렌드 리포트 2020 (p.17)

3. 영양성분 & 가격 데이터

	A	B	C	D	E	F	G	H
1	big_cat	small_cat	serving size(g)	calorie(kcal)	carbohydrate(g)	protein(g)	fat(g)	price(average)
2	쌀/잡곡	쌀	200	1.52	0.332	0.03	0.001	3.18
3	쌀/잡곡	잡곡	200	1.5	0.338	0.028	0.0038	3.73
4	과일/견과	견과류 땅콩	20	5.2	0.184	0.257	0.426	9.6
5	과일/견과	견과류 캐슈넛	20	5.76	0.267	0.198	0.476	13.5
6	과일/견과	포도/거봉/체리	100	0.67	0.1761	0.0083	0.0019	16.8
7	과일/견과	미숫가루/곡물가루	50	4.15	0.8	0.1	0.06	10
8	과일/견과	견과류 호두	20	6.88	0.07	0.155	0.7199	13.2
9	과일/견과	혼합견과	20	1.39	0.09	0.03	0.1	32.9
10	과일/견과	견과류 피스타치오	20	5.84	0.2082	0.2599	0.489	27.1
11	과일/견과	견과류 잣/은행	20	4.165	0.3105	0.1032	0.304	134

칼로리(calorie), 탄수화물(carbohydrate), 단백질(protein), 지방(fat) - /g 단위

상한선 = 남성 1회 평균 식사제공량 + 3×표준오차

하한선 = 여성 1회 평균 식사제공량 - 3×표준오차

```
calorie_lo = 530
calorie_up = 800
carbo_lo = 50
carbo_up = 120
protein_lo = 20
protein_up = 50
fat_lo = 5
fat_up = 20
price_lo = 0
price_up = 7000
```


1. 변수 설정

기존 데이터에서 사용한 변수

변수	설명
temp	평균기온(°C)
temp_L	최저기온(°C)
temp_H	최고기온(°C)
rain	평균 강수량(mm)
wind_max	최대 풍속(m/s)
wind	평균 풍속(m/s)
humid	평균 상대습도(%)
dust	일평균 미세먼지 농도($\mu\text{g}/\text{m}^3$)

태풍, 폭염, 한파 각각의 경보, 주의보 데이터

: 직접 **one-hot-encoding**을 통해 6개의 변수 생성

날짜	태풍주의보	태풍경보	폭염주의보	폭염경보	한파주의보	한파경보
2018-01-01	0	0	0	0	1	0
2018-01-02	0	0	0	0	1	0
...
2019.12.31	0	0	0	0	0	1

2. NaN 처리 및 대체

- 1) 기상 데이터 : 결측치에 해당하는 측정치가 없는 것으로 간주, 결측치를 0으로 대체
- 2) 미세먼지 데이터 : 측정을 하지 못한 것으로 간주, 결측치를 전날의 값으로 대체

```
dust = []
cnt = 0
for i in range(730):
    j = i - cnt
    if (df_dust['날짜'].iloc[i] == dust20189['일시'].iloc[j]):
        dust.append(dust20189['일 미세먼지 농도( $\mu\text{g}/\text{m}^3$ )'].iloc[j])
    else:
        dust.append(np.nan)
        cnt += 1
df_dust['미세먼지'] = pd.Series(dust)
df_dust = df_dust.fillna(method='ffill')
```

미세먼지 데이터 결측치 처리

```
# 날씨 데이터 결측치 0으로
Seoul2018 = Seoul2018.fillna(0)
Seoul2019 = Seoul2019.fillna(0)
```

기상 데이터 결측치 처리

3. 새로운 변수 생성

계절 변수

봄 (3,4,5)	여름 (6,7,8)
가을 (9,10,11)	겨울 (12,1,2)

→ one-hot-encoding을 통해
봄, 여름, 가을, 겨울 indicator 변수 생성

체감온도

```
sens_temp = 13.27 + 0.6215*temp - 13.947*(wind**0.16) + 0.486*temp*(wind**0.16)
sens_temp = pd.Series(sens_temp, name='체감온도')
```

불쾌지수

```
THI = (9/5)*temp - 0.55*(1 - humid)*((9/5)*temp - 26) + 32
THI_cat = np.zeros(730)
THI_cat = pd.Series(THI_cat, name='불쾌지수')
for i in range(0, len(THI_cat)):
    if THI[i] < 68 :
        THI_cat[i] = 'THI낮음'
    elif THI[i] >= 68 and THI[i] < 75 :
        THI_cat[i] = 'THI보통'
    elif THI[i] >= 75 and THI[i] < 80 :
        THI_cat[i] = 'THI높음'
    else: THI_cat[i] = 'THI매우높음'
```

→ one-hot-encoding을 통해 불쾌지수 낮음, 보통, 높음, 매우 높음에
해당하는 indicator 변수 생성

최종 데이터 변수

설명변수			
봄	관측한 날씨가 3,4,5월인지 여부	평균 상대습도(%)	하루 평균 습도
여름	관측한 날씨가 6,7,8월인지 여부	미세먼지 농도($\mu\text{g}/\text{m}^3$)	하루 평균 미세먼지 농도
가을	관측한 날씨가 9,10,11월인지 여부	체감온도	관측한 날씨의 체감 기온
겨울	관측한 날씨가 12,1,2월인지 여부	불쾌지수(THI) 낮음	관측한 불쾌지수가 68 미만인지 여부
평균기온($^{\circ}\text{C}$)	관측한 기온의 산술평균	불쾌지수(THI) 보통	관측한 불쾌지수가 68 이상 75 미만인지 여부
최저기온($^{\circ}\text{C}$)	관측한 기온의 가장 낮은 수치	불쾌지수(THI) 높음	관측한 불쾌지수가 75 이상 80 미만인지 여부
최고기온($^{\circ}\text{C}$)	관측한 기온의 가장 높은 수치	불쾌지수(THI) 매우높음	관측한 불쾌지수가 80 이상인지 여부
일강수량(mm)	하루 동안의 강수량의 총량	태풍주의보/태풍경보	관측한 날씨의 태풍주의보/경보 발표 여부
최대 풍속(m/s)	관측한 풍속의 가장 높은 수치	폭염주의보/폭염경보	관측한 날씨의 폭염주의보/경보 발표 여부
평균 풍속(m/s)	관측한 풍속의 산술평균	한파주의보/한파경보	관측한 날씨의 한파주의보/경보 발표 여부
종속변수	온라인 구매 데이터 또는 SNS 데이터의 일별 수치 총합(성별, 나이 구분 없이)		

4. 데이터 표준화

1) 설명변수

각 변수별로 범위가 서로 달라, 모델에 적용했을 때 큰 값을 가지는 변수의 영향이 크게 작용할 수 있음

→ 각 변수의 **평균이 0, 표준편차가 1**이 되도록 표준화

2) 종속변수

각 상품별로 구매 또는 검색량의 범위가 서로 다름

→ 2018~2019년 구매 또는 검색량의 합에서 **각 일자의 구매량 또는 검색량이 차지하는 비율(%)**로 나타냄

5. Outlier 제거

종속변수의 값이 너무 크거나 작은 데이터를 제거

→ 일반적인 구매 또는 검색량에 대한 예측 성능을 높일 수 있도록 함

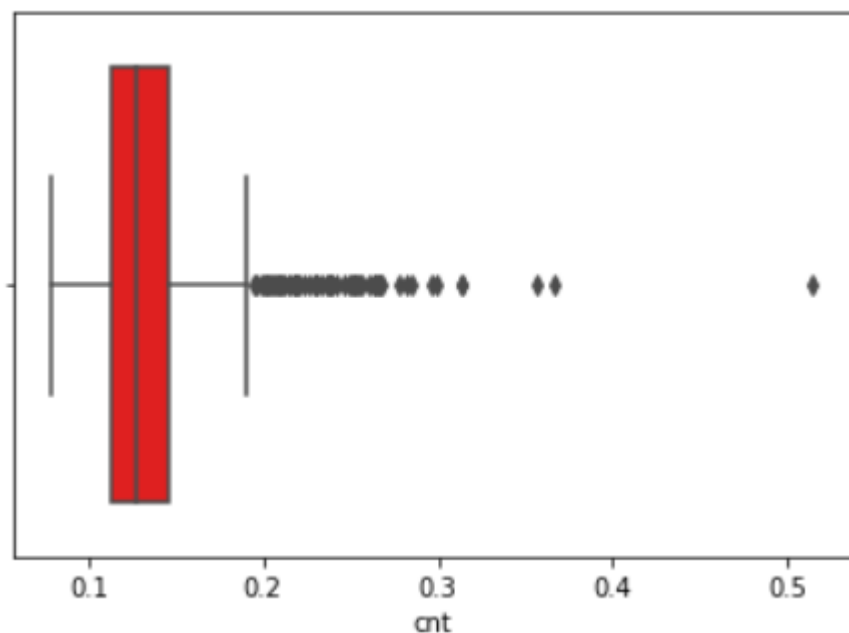
※ Outlier = 25th percentile에서 $1.5 * IQR$ (Inter Quartile Range)을 뺀 값보다 작거나
75th percentile에서 $1.5 * IQR$ 을 더한 값보다 큰 값으로 정의

1) Boxplot을 그려 outlier를 확인

2) Outlier에 해당하지 않는 index를 도출하는 함수를 통해 Outlier가 아닌 데이터를 정의

5. Outlier 제거

Box Plot



```
def outliers_iqr(data):
    q1, q3 = np.percentile(data, [25, 75])
    iqr = q3 - q1
    lower_bound = q1 - (iqr * 1.5)
    upper_bound = q3 + (iqr * 1.5)

    return np.where((data <= upper_bound) & (data >= lower_bound))

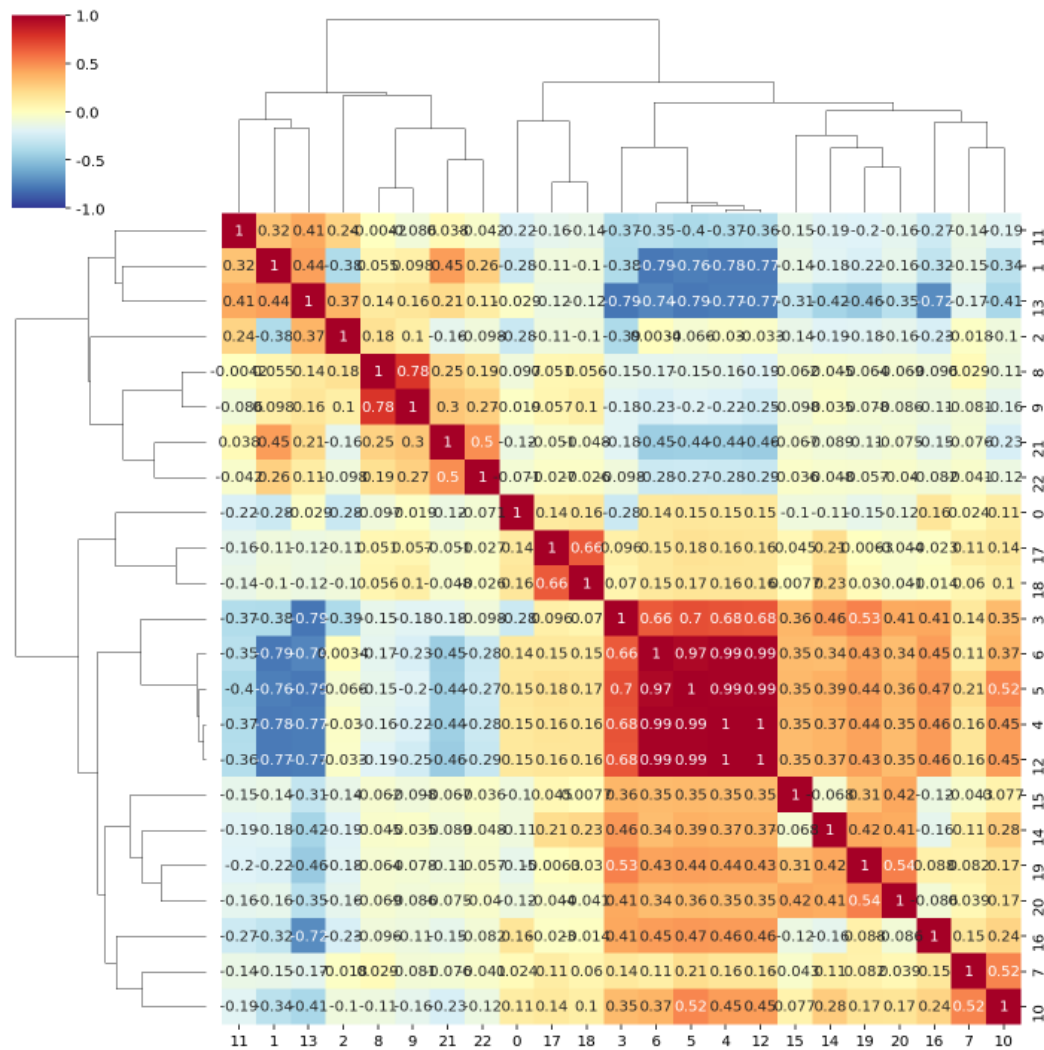
y_target_not_outlier_index = outliers_iqr(y_target)
y_target_not_outlier_index = np.array(y_target_not_outlier_index)

Y_target = []
X_data = pd.DataFrame(X_data)
new_X = pd.DataFrame(columns=range(X_data.shape[1]))
for i in y_target.index:
    if y_target.index[i] in y_target_not_outlier_index:
        Y_target.append(y_target[i])
        new_X = pd.concat([new_X, pd.DataFrame(X_data.iloc[i, :]).transpose()], axis=0)
y_target = Y_target
y_target = np.array(y_target)
X_data = new_X.reset_index(drop=True)
```

6. 설명변수 간 상관관계 분석

1) 상관계수 확인

→ 새로 생성한 변수(계절 변수, 체감온도, 불쾌지수 등)와 기존 변수(기온, 풍속, 강수 등)의 상관관계가 높게 나타남



6. 설명변수 간 상관관계 분석

2) VIF(Variation Inflation Factors) 값 확인

- 10 이상이면 다중공선성(multicollinearity)이 있다고 판단
(독립변수의 VIF값이 모두 크면 서로 상관관계가 있다고 판단)

→ 계절변수(봄,여름,가을,겨울), 평균기온, 최저기온, 최고기온, 체감온도, 불쾌지수 낮음/보통/높음/매우높음 변수는 서로 상관관계가 높음

VIF Factor features					
0	inf	0	11	1.502467	11
1	inf	1	12	1979.822615	12
2	inf	2	13	inf	13
3	inf	3	14	inf	14
4	2482.597625	4	15	inf	15
5	155.295808	5	16	inf	16
6	173.649340	6	17	1.869733	17
7	1.525512	7	18	1.879232	18
8	2.860009	8	19	1.800108	19
9	5.574139	9	20	1.824942	20
10	2.571406	10	21	1.723613	21
			22	1.450554	22

1. 교차검증 및 모델 성능 평가

1) Train set과 Test set split

- Train : Test의 비율이 8:2가 되도록 train set으로 모델을 학습시키고 test set으로 모델 성능을 평가

```
from sklearn.model_selection import train_test_split
import xgboost as xgb
from sklearn.ensemble import RandomForestRegressor
import pandas as pd
import numpy as np
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.linear_model import Ridge

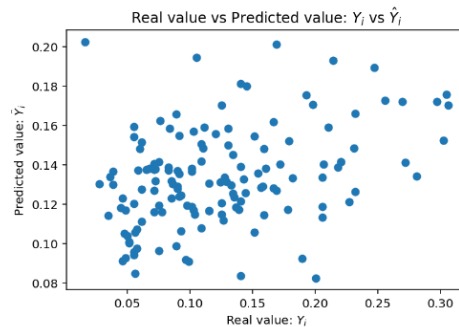
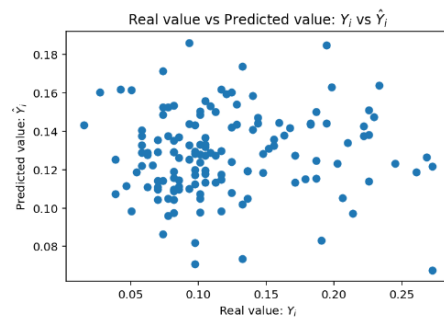
x_train, x_test, y_train, y_test = train_test_split(X_data, y_target, test_size=0.2, random_state=0)

def MAPE(y_test, y_pred):
    return np.mean(np.abs((y_test - y_pred) / y_test)) * 100
```

1. 교차검증 및 모델 성능 평가

2) K fold 교차검증

- 일반적인 성능을 구하기 위해, train set과 test set을 8:2의 비율로 선택하는 과정을 5번 반복
- 5번의 test set에 대한 R-square, MSE(Mean Squared Error), MAPE(Mean Absolute Percentage Error)를 산출하고, 평균값을 성능으로 결정



평균검증 mse: 0.0004

평균검증 r2: 0.16094

평균검증 mape: 12.961920000000001

(Ridge 모델에 대한 검증)

2. 예측 모델 선정

사용한 모델 : Xgboost, Random Forest, Lightgbm, Ridge Regression

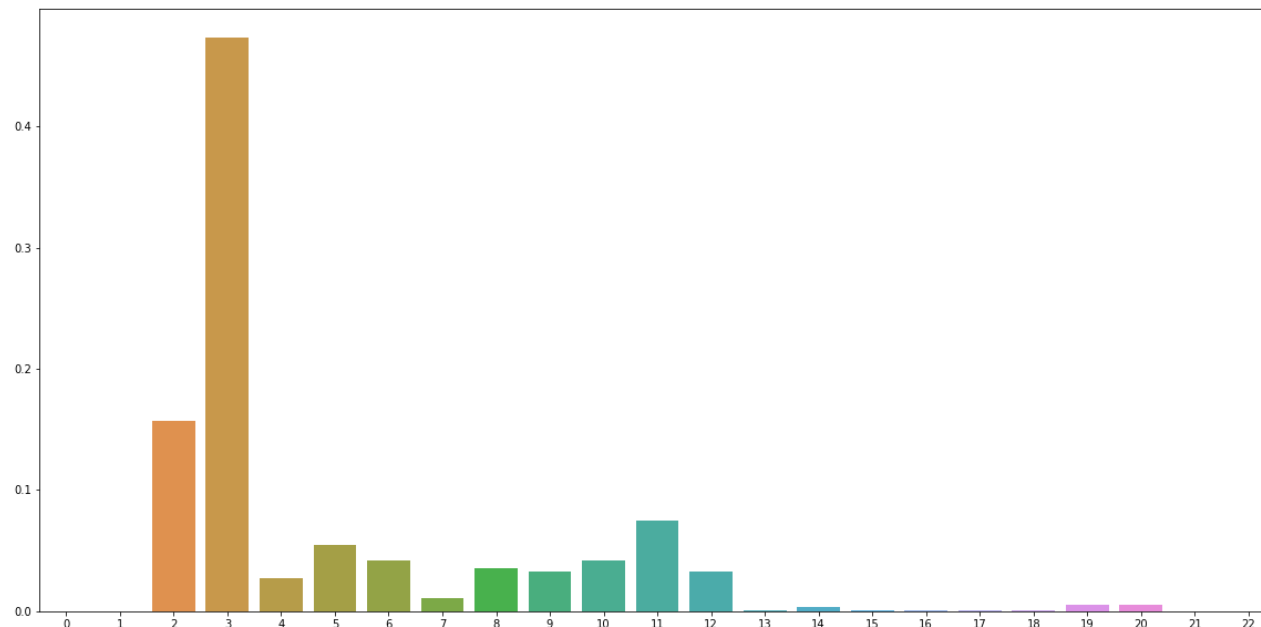
→ 앙상블 모델(Xgboost, RandomForest, Lightgbm)이 일반적으로 단일 모델에 비해 좋은 성능을 보인다고 알려져 있기에, 앙상블 모델을 다양하게 사용

→ Ridge Regression은 계수의 크기를 줄이는 효과를 가져, 일반적인 선형회귀모형에 비해 안정적인 회귀 모형을 가질 수 있도록 함

2. 예측 모델 선정

RandomForest에서 산출된 중요 변수 (split에 자주 사용된 변수) :

참외/메론/수박의 경우 여름, 봄, 평균상대습도 순으로 구매량에 영향을 많이 줌

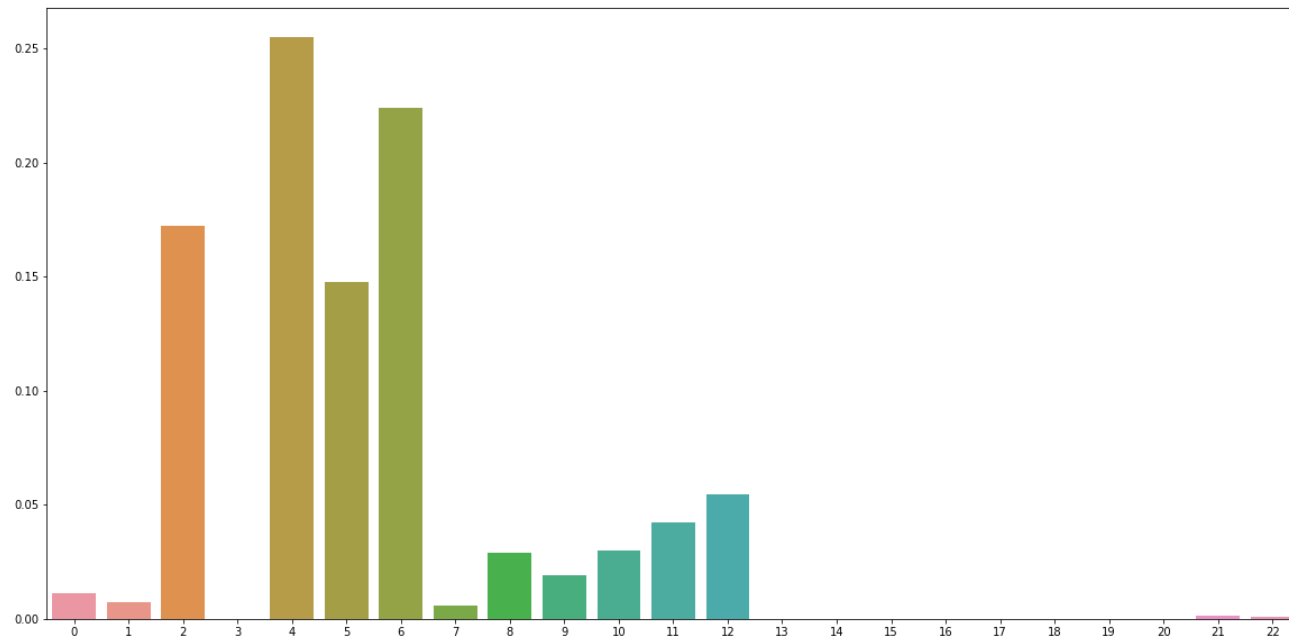


3	0.473277	20	0.005387
2	0.156802	19	0.004877
11	0.075217	14	0.003041
5	0.054290	16	0.001087
6	0.042152	17	0.001018
10	0.041946	18	0.000975
8	0.035308	15	0.000641
12	0.032543	13	0.000288
9	0.032283	1	0.000144
4	0.027310	21	0.000143
7	0.011156	0	0.000107
		22	0.000007

2. 예측 모델 선정

RandomForest에서 산출된 중요 변수 (split에 자주 사용된 변수) :

굴 생물의 경우 평균기온, 최고기온, 봄, 체감온도 순으로 소셜 데이터량에 영향을 많이 줌



4	0.254907	7	0.005803
6	0.223908	21	0.001534
2	0.172038	22	0.000864
5	0.147384	19	0.000078
12	0.054375	3	0.000074
11	0.042197	16	0.000052
10	0.030053	14	0.000042
8	0.029078	17	0.000035
9	0.018932	20	0.000033
0	0.011288	13	0.000027
1	0.007260	18	0.000024
		15	0.000014

3. 예측 변수 선정 기준

온라인 구매 데이터 카테고리 선정 기준	SNS 소셜 데이터 카테고리 선정 기준
<p>앞선 4가지 모델(Xgboost, Random Forest, Lightgbm, Ridge Regression)에서</p> <ol style="list-style-type: none"> 1) R-square 값이 음수 값으로 산출되면 제외 2) R-square 값이 0.1를 넘는 식품군을 선택 (선택되는 식품군의 수가 너무 적게 산출되지 않기 위함) 	<p>앞선 4가지 모델(Xgboost, Random Forest, Lightgbm, Ridge Regression)에서</p> <ol style="list-style-type: none"> 1) R-square 값이 음수 값으로 산출되면 제외 2) R-square 값이 0.2를 넘는 식품군을 선택 (선택되는 식품군의 수가 너무 적게 산출되지 않기 위함, 소셜 데이터가 날씨의 영향을 더 많이 받아 모델링에서 높은 성능을 보였기에 더 높은 기준으로 설정)

3. 예측 변수 선정 – 온라인 구매 데이터

big_cat	small_cat	Random Forest			XGBoost			LGBM			Ridge Regression		
		R square	MSE	MAPE	R square	MSE	MAPE	R square	MSE	MAPE	R square	MSE	MAPE
과일/견과	견과류 땅콩	0.055	0.00176	30.9239	0.082	0.00172	30.31	-0.05378	0.00198	32.28766	0.13244	0.00162	30.0439
과일/견과	견과류 캐슈넛	-0.06968	0.00324	inf	-0.0788	0.0032	inf	-0.2624	0.00378	inf	0.00029	0.00302	inf
과일/견과	포도/거봉/체리	0.26424	0.00226	54.76428	0.24394	0.00232	56.74986	0.1985	0.00245	55.77562	0.27	0.00224	55.45565
과일/견과	미숫가루/곡물가루	0.53444	0.00172	32.1744	0.52788	0.0017	31.9909	0.48488	0.0019	33.1215	0.5232	0.00178	32.60358
과일/견과	견과류 호두	0.00059	0.00176	29.63	0.00536	0.0017	29.40806	-0.1282	0.002	30.6636	0.0652	0.00164	28.8393
과일/견과	혼합견과	0.1023	0.00146	33.46134	0.0715	0.00148	33.568	-0.03614	0.00167	35.586	0.1294	0.00142	33.02662
과일/견과	견과류 피스타치오	-0.02924	0.00266	inf	-0.04152	0.00266	inf	-0.18768	0.00306	inf	-0.0163	0.0026	inf
과일/견과	견과류 잣/은행				-0.01928	0.00216	46.50765						
과일/견과	딸기/복분자/블루베리	0.55736	0.00172	30.40534	0.55324	0.00178	31.0337	0.51852	0.0019	31.22432	0.55298	0.00178	31.84092
과일/견과	감귤/한라봉/오렌지	0.61222	0.00306	45.4264	0.59636	0.0032	46.5395	0.5726	0.00335	46.74419	0.62148	0.003	46.7199
과일/견과	바나나/파인애플/망고	0.25108	0.00124	25.7565	0.19818	0.0013	26.649	0.1685	0.0013	27.2751	0.21214	0.0013	26.50876
과일/견과	견과류 밤	0.34762	0.00228	34.98186	0.29452	0.00248	36.86818	0.26964	0.00256	36.69678	0.35585	0.00226	34.8292
과일/견과	레몬/자몽	0.16314	0.0012	26.73466	0.15594	0.00122	26.95056	0.0832	0.00132	27.9898	0.17798	0.00118	26.52344
과일/견과	참외/메론/수박	0.67442	0.00452	56.8456	0.67558	0.00456	59.32142	0.65766	0.00476	64.15208	0.6237	0.00528	67.3811
과일/견과	키위/참다래	0.32148	0.00186	33.68116	0.27946	0.00196	35.39358	0.24524	0.00206	35.66854	0.28024	0.00196	35.7169
과일/견과	견과류 마카다미아	0.0596	0.00242	38.2386	0.05042	0.0024	38.519	-0.00718	0.00258	38.73674	0.1192	0.00226	37.43814
과일/견과	감/홍시	0.5379	0.0022	59.74234	0.52302	0.0023	58.5804	0.50922	0.00232	60.3635	0.42622	0.00276	68.163
채소1	쌈채소	0.07216	0.00138	25.4672	0.0334	0.00142	25.67736	-0.01816	0.0015	26.22	0.04542	0.0014	25.89938
채소1	파/양파	0.04042	0.00136	26.6845	0.02386	0.0014	26.944	-0.10534	0.0015	28.3773	0.06884	0.00134	26.25
채소1	양배추/양상추	0.0301	0.0016	29.83	0.03682	0.00166	29.619	-0.05832	0.0018	30.83548	0.02294	0.00167	30.0727

3. 예측 변수 선정 – SNS 소셜 데이터

big_cat	small_cat	Random Forest			XGBoost			LGBM			Ridge Regression		
		R square	MSE	MAPE	R square	MSE	MAPE	R square	MSE	MAPE	R square	MSE	MAPE
과일/견과	견과류 땅콩	0.07809	0.00016	6.94	0.0098	0.00017	7.162999	-0.0076	0.00016	7.29092			
과일/견과	견과류 캐슈넛							-0.17146					
과일/견과	포도/거봉/체리	0.75688	0.00022	8.334	0.75218	0.00022	8.52814	0.748	0.0002	8.55406	0.71	0.00026	9.48108
과일/견과	미숫가루/곡물가루	0.8091	0.00043	13.4566	0.80726	0.00048	13.2651	0.79244	0.00052	13.85888	0.7922	0.00052	14.0746
과일/견과	견과류 호두	0.32248	0.0001	6.00256	0.31338	0.0001	6.00978	0.2882	0.0001	6.1866	0.31928	0.0001	6.06388
과일/견과	혼합견과							-0.27244					
과일/견과	견과류 피스타치오	0.11432	0.00076	17.581	0.0898	0.00076	17.65178	0.02052	0.0008	18.31224	0.0619	0.00076	18.2439
과일/견과	견과류 잣/은행							-0.01758					
과일/견과	딸기/복분자/블루베리	0.8705	0.0003	8.9218	0.8676	0.0003	9.531	0.8675	0.0003	9.24874	0.8327	0.0004	11.2798
과일/견과	감귤/한라봉/오렌지	0.3789	0.00022	8.19918	0.3605	0.0002	8.3122	0.3251	0.00022	8.6096	0.3658	0.0002	8.4521
과일/견과	바나나/파인애플/망고	0.63192	0.0001	5.72438	0.6289	0.0001	5.8209	0.6083	0.0001	5.91894	0.6138	0.0001	6.1235
과일/견과	견과류 밤							-0.09978					
과일/견과	레몬/자몽							-0.0337					
과일/견과	참외/메론/수박	0.8796	0.00124	17.02772	0.87419	0.00128	18.48338	0.87938	0.00122	17.7194	0.84328	0.00162	26.22006
과일/견과	키위/참다래	0.5587	0.0003	9.7378	0.5539	0.00028	9.8605	0.5117	0.0003	10.31108	0.4844	0.0003	10.65846
과일/견과	견과류 마카다미아	0.26036	0.0008	18.49374	0.222	0.00086	18.9783	0.17134	0.00088	19.5655	0.2023	0.0008	19.3318
과일/견과	감/홍시	0.39938	0.0001	4.7	0.3958	0.0001	4.77468	0.29076	0.0001	5.0246	0.3144	0.0001	5.0735
채소1	쌈채소	0.3368	0.00056	14.5962	0.31656	0.00056	14.97336	0.28148	0.0006	15.1568	0.29864	0.0006	15.15396
채소1	파/양파	0.38736	0.00014	7.1375	0.38978	0.00012	7.16919	0.37532	0.00014	7.2138	0.35848	0.00016	7.39942
채소1	양배추/양상추	0.31476	0.0003	10.43242	0.28608	0.0003	10.67114	0.24488	0.0003	10.9621	0.27238	0.00031	10.8186
채소1	미나리	0.6761	0.0006	13.9131	0.6885	0.0005	13.7867	0.64142	0.0006	14.863	0.6462	0.0006	15.16578
채소1	브로콜리/셀러리	0.25612	0.0003	11.4459	0.2156	0.0003	11.80772	0.1945	0.0004	11.93728	0.25128	0.0003	11.77718
채소1	오이/가지							-0.0344					
채소1	시금치	0.68408	0.00044	11.8601	0.66544	0.00046	12.17382	0.65088	0.0005	12.45488	0.6913	0.0004	11.70244
채소1	무/배추							-0.07512					
채소1	콩나물/숙주							-0.0131					
채소1	부추	0.4963	0.0003	10.1281	0.5134	0.0003	10.0199	0.4795	0.0003	10.2359	0.4848	0.0003	10.37716
채소2	새송이버섯							-0.14594					
채소2	표고버섯	0.31262	0.00031	10.77704	0.33226	0.0003	10.7191	0.24542	0.0003	11.29058	0.32308	0.0003	10.8052
채소2	도라지/더덕	0.2163	0.0004	12.3413	0.17528	0.00044	12.7812	0.0933	0.00052	13.39662	0.21944	0.0004	12.5366

4. 분석 및 모델링

온라인 구매 데이터

	item	g	calorie	carbo	protein	fat	price	pred1	pred2	pred3	pred4
0	포도/거봉/체리	100	0.67	0.1761	0.0083	0.0019	16.80	0.426712	0.813871	0.454498	0.482078
1	미숫가루/곡물가루	50	4.15	0.8000	0.1000	0.0600	10.00	0.127101	0.208000	0.226498	0.193195
2	딸기/복분자/블루베리	100	0.55	0.1257	0.0100	0.0080	42.00	0.241218	0.117907	0.116887	0.086758
3	감귤/한라봉/오렌지	100	0.50	0.1280	0.0100	0.0010	11.80	1.416313	0.973386	-0.263388	0.447830
4	바나나/파인애플/망고	100	0.84	0.2190	0.0110	0.0010	3.83	0.137823	0.157678	0.151188	0.143918
5	견과류 밤	50	1.51	0.3340	0.0330	0.0050	20.70	0.618773	0.486843	0.260381	0.397196
6	레몬/자몽	100	0.28	0.0930	0.0071	0.0008	6.95	0.047517	0.326221	0.388889	0.260059
7	참외/메론/수박	100	0.39	0.0940	0.0120	0.0018	7.90	0.079475	0.207373	0.323638	0.134406
8	키위/참다래	100	0.64	0.1480	0.0080	0.0100	11.27	0.104552	0.124291	0.119538	0.122077
9	감/홍시	100	0.58	0.1575	0.0040	0.0004	11.94	0.125146	0.131956	0.139542	0.100104

SNS 소셜 데이터

	item	g	calorie	carbo	protein	fat	price	pred1	pred2	pred3	pred4
0	포도/거봉/체리	100	0.67	0.1761	0.0083	0.0019	16.80	0.112691	0.136948	0.166813	0.178377
1	미숫가루/곡물가루	50	4.15	0.8000	0.1000	0.0600	10.00	0.108416	0.214852	0.214075	0.215105
2	견과류 호두	20	6.88	0.0700	0.1550	0.7199	13.20	0.136192	0.126098	0.115369	0.139469
3	딸기/복분자/블루베리	100	0.55	0.1257	0.0100	0.0080	42.00	0.170853	0.135265	0.098110	0.080592
4	감귤/한라봉/오렌지	100	0.50	0.1280	0.0100	0.0010	11.80	0.154215	0.130752	0.105990	0.107275
5	바나나/파인애플/망고	100	0.84	0.2190	0.0110	0.0010	3.83	0.125138	0.151829	0.148085	0.126960
6	참외/메론/수박	100	0.39	0.0940	0.0120	0.0018	7.90	0.074717	0.197835	0.303753	0.114353
7	키위/참다래	100	0.64	0.1480	0.0080	0.0100	11.27	0.130875	0.144404	0.150869	0.139160
8	견과류 마카다미아	20	7.20	0.1220	0.0830	0.7670	38.50	0.148354	0.104170	0.104538	0.180990
9	감/홍시	100	0.58	0.1575	0.0040	0.0004	11.94	0.129073	0.120915	0.121023	0.142583

-> 각 상품별 가장 높은 성능을 보인 모델을 이용하여 2020년 날씨에 따른 예측값 산출 (ex. 과일/견과)

5. 모델링에 따른 예측 결과

칼로리, 탄수화물, 단백질, 지방, 가격의 상한, 하한 내에 들어오는 카테고리별 상품 조합 중 상품 구매량에 대한 예측값의 합이 가장 높은 상위 5개의 조합

2020년 1월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred	
587	쌀	돼지	곱창	조개	미나리	표고버섯	물김치	감귤/한라봉/오렌지	2.532469
679	쌀	돼지	곱창	조개	미나리	고추/피망/파프리카	물김치	감귤/한라봉/오렌지	2.526738
608	쌀	돼지	곱창	조개	미나리	도라지/더덕	물김치	감귤/한라봉/오렌지	2.515951
632	쌀	돼지	곱창	조개	미나리	토마토	물김치	감귤/한라봉/오렌지	2.492982
655	쌀	돼지	곱창	조개	미나리	감자	물김치	감귤/한라봉/오렌지	2.490320

2020년 4월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred	
671	쌀	돼지	곱창	조개	미나리	고추/피망/파프리카	무김치	감귤/한라봉/오렌지	2.203417
679	쌀	돼지	곱창	조개	미나리	고추/피망/파프리카	물김치	감귤/한라봉/오렌지	2.193603
512	쌀		계란	꽃게	오이/가지	토마토	무김치	감귤/한라봉/오렌지	2.187481
520	쌀		계란	꽃게	오이/가지	토마토	물김치	감귤/한라봉/오렌지	2.177667
534	쌀		계란	꽃게	오이/가지	감자	무김치	감귤/한라봉/오렌지	2.165308

2020년 7월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
555	쌀	계란	꽃게	오이/가지	고추/피망/파프리카	무김치	레몬/자몽	2.422541
498	쌀	계란	꽃게	오이/가지	표고버섯	무김치	레몬/자몽	2.415475
537	쌀	계란	꽃게	오이/가지	감자	무김치	레몬/자몽	2.405378
515	쌀	계란	꽃게	오이/가지	토마토	무김치	레몬/자몽	2.393716
559	쌀	계란	꽃게	오이/가지	고추/피망/파프리카	물김치	레몬/자몽	2.370215

2020년 10월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
534	쌀	계란	꽃게	오이/가지	감자	무김치	감귤/한라봉/오렌지	1.744499
512	쌀	계란	꽃게	오이/가지	토마토	무김치	감귤/한라봉/오렌지	1.727530
541	쌀	계란	꽃게	오이/가지	감자	물김치	감귤/한라봉/오렌지	1.723692
520	쌀	계란	꽃게	오이/가지	토마토	물김치	감귤/한라봉/오렌지	1.706723
536	쌀	계란	꽃게	오이/가지	감자	무김치	견과류 밤	1.693865

5. 모델링에 따른 예측 결과

칼로리, 탄수화물, 단백질, 지방, 가격의 상한, 하한 내에 들어오는 카테고리별 상품 조합 중 SNS 소셜 데이터에 대한 예측값의 합이 가장 높은 상위 5개의 조합

2020년 1월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
2889	쌀	닭가슴살	굴 생물	시금치	우영/연근	물김치	견과류 마카다미아	1.310543
2887	쌀	닭가슴살	굴 생물	시금치	우영/연근	무김치	견과류 마카다미아	1.310027
2888	쌀	닭가슴살	굴 생물	시금치	우영/연근	물김치	견과류 호두	1.298380
2886	쌀	닭가슴살	굴 생물	시금치	우영/연근	무김치	견과류 호두	1.297864
2923	쌀	닭가슴살	굴 생물	시금치	마늘/생강	물김치	견과류 마카다미아	1.296890

2020년 4월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
4490	쌀	오리고기/훈제오리	고등어	시금치	도라지/더덕	물김치	참외/메론/수박	1.276180
4488	쌀	오리고기/훈제오리	고등어	시금치	도라지/더덕	무김치	참외/메론/수박	1.270051
4492	쌀	오리고기/훈제오리	고등어	시금치	도라지/더덕	파김치	참외/메론/수박	1.266271
3669	쌀	오리고기/훈제오리	오징어	시금치	토마토	물김치	미숫가루/곡물가루	1.245197
7036	쌀	오리고기/훈제오리	갈치	부추	도라지/더덕	물김치	미숫가루/곡물가루	1.245074

2020년 7월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
2443	쌀	닭가슴살	갈치	부추	도라지/더덕	물김치	참외/메론/수박	1.466224
932	쌀	닭가슴살	고등어	부추	도라지/더덕	물김치	참외/메론/수박	1.454718
1978	쌀	닭가슴살	갈치	파/양파	도라지/더덕	물김치	참외/메론/수박	1.448813
2129	쌀	닭가슴살	갈치	양배추/양상추	도라지/더덕	물김치	참외/메론/수박	1.444043
566	쌀	닭가슴살	고등어	파/양파	도라지/더덕	물김치	참외/메론/수박	1.437307

2020년 10월 15일

	rice	livestock	fish	veg1	veg2	kimchi	fruit	pred
935	쌀	닭가슴살	고등어	부추	도라지/더덕	파김치	포도/거봉/체리	1.161905
485	쌀	닭가슴살	오징어	부추	도라지/더덕	파김치	견과류 마카다미아	1.155758
1529	쌀	닭가슴살	조개	부추	도라지/더덕	파김치	견과류 마카다미아	1.152520
929	쌀	닭가슴살	고등어	부추	도라지/더덕	물김치	포도/거봉/체리	1.145935
483	쌀	닭가슴살	오징어	부추	도라지/더덕	물김치	견과류 마카다미아	1.139787

이러한 예측 결과를 토대로 기획한
날씨 최적화 식단 설계 및 추천 플랫폼

뭉게뭉게 냉장고



모바일 플랫폼 예시 UI

1 날씨에 민감한 식품군 선별

- 기상 요소에 따른 대중적인 식품군별 선호도를 파악하고, 이를 기반으로 식단 조합 추천

2 영양성분을 고려한 건강한 식단

- 열량 및 탄수화물, 단백질, 지방 함량을 1회 식사 권장 제공량에 맞추어 적절히 구성

3 예산에 맞춘 적절한 단가의 식단

- 가격 데이터의 고려를 통해 용도에 맞추어 적절한 단가의 식단 설계 및 조합 가능

뭉게뭉게 냉장고
추천 식단



메인 디시 1(육류)

메인 디시 2(수산물)

사이드 디시 1(녹황채소)

사이드 디시 2(근채류/담자균류/기타)

김치류

디저트(과일/견과)

뭉게뭉게 냉장고의 추천 식단을 바탕으로 공공기관 및 단체 배식이 필요한 시설의 식단 메뉴 설계

코로나 격리시설

코로나19로 격리된 사람들이 날씨에 따라 선호할 만한 식단을 제공하여
코로나 블루를 완화할 수 있는 솔루션

노인 복지시설

연령별 수요 예측을 바탕으로 기상 요소에 따라 노년층이 선호할 만한
메뉴를 분석하여 이를 토대로 식사를 제공하는 솔루션

기타 식사 배급 단체

공공기관 구내식당, 배식 봉사 단체 등의 메뉴 설계에 있어 영양성분 및
단가, 기상요소에 따른 사람들의 보편적 선호도를 고려하는 솔루션

날씨에 따른 고객들의 수요 예측을 통해 기상 요소에 따라 인기 있을 식사 메뉴 기획 및 판매

도시락 메뉴 기획

기상 요소에 따라 수요가 많을 것으로 예상되는 식단 구성을 도시락 메뉴로 조합하여 판매하는 솔루션

시즌 메뉴 프로모션

특정 기상기후가 지속적으로 나타나는 계절에 맞춘 시즌 메뉴를 설계, 예보 데이터를 고려해 적절한 프로모션 시기를 결정하는 솔루션

오프라인 음식점

그날그날 다른 메뉴를 제공하는 오프라인 음식점에서 날씨에 따라 많은 사람들이 선호할 것으로 예측되는 동시에 적절한 단가를 갖춘 식사 메뉴를 결정할 수 있는 솔루션

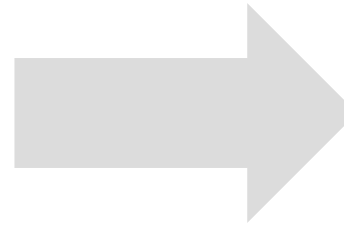


6.기대 효과

매일매일의 날씨에 따른 요리와 음식점을 추천

요리에 포함된
재료군 데이터

각 음식점의
요리 데이터



추후 다양한 데이터의 도입으로 플랫폼 기능의 혁신 및 확장 가능성에 따른 기대 효과

1. 연령별, 성별 소비패턴 및 선호도 분석

- 온라인 구매 이력 및 SNS 데이터의 세부 분석을 바탕으로 각 성별 및 연령대가 선호할 만한 식단 구성

2. 알레르기 및 비선호 재료/식품의 배제

- 플랫폼 상에서 식단 구성 시 제외할 재료 및 식품 종류를 선택하여 맞춤형 영양성분을 갖춘 식단 설계

3. 예보 데이터와의 결합을 통한 체계적 식단 구성

- 기상 예보 데이터를 함께 고려하여 날씨에 따른 사람들의 수요를 더욱 정확하게 예측 가능

- (1) 기상자료개방포털, 종관기상관측(ASOS) 데이터
<https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>
- (2) 기상자료개방포털, 황사관측(PM10) 데이터
<https://data.kma.go.kr/data/climate/selectDustRltmList.do?pgmNo=68>
- (3) 기상청, "불쾌지수"
https://www.kma.go.kr/HELP/basic/help_01_05.jsp
- (4) 기상자료개방포털, "체감온도"
<https://data.kma.go.kr/climate/windChill/selectWindChillChart.do?pgmNo=111>
- (5) 기상자료개방포털, 기상특보
<https://data.kma.go.kr/data/weatherReport/wsrList.do?pgmNo=647>
- (6) 벌써부터 한여름 날씨… "올해 빙그레 매출 1조 넘을 것", 동아일보
<https://www.donga.com/news/Economy/article/all/20210609/107348324/1>
- (7) [날씨이야기]비오는 날 홈쇼핑에선 김치가 잘 팔린다, edaily,
<https://www.edaily.co.kr/news/read?newsId=01390726615827240&mediaCodeNo=257>

