



오늘 뭐 먹지?

식사 메뉴 추천 시스템 구현을 위한
데이터 분석 프로젝트

8조 이선정

CONTENTS

01 서론

- 문제 현황 및 연구 배경
- 주제선정

02 본론

- 데이터 수집 및 전처리 과정
- 데이터 분석

03 결론

- 기대효과
- 활용방안

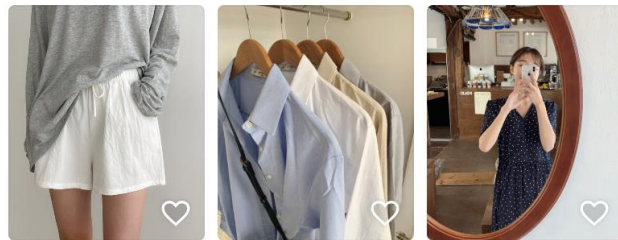
서론 연구배경 및 문제현황

1) 데이터 기반 식사 메뉴 추천 시스템이 필요한 이유



▲ 넷플릭스

당신을 위한 추천 아이템



▲ 지그재그

왓차 영화 순위



넷플릭스 영화 순위

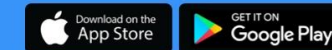


▲ 왓차피디아

게임 평가 및 추천 서비스

미니맵

- Steam, PS, Xbox 연동으로 게임 관리를 한 번에!
- 340만 개의 메타데이터 분석을 통한 크로스플랫폼 게임 추천



minimap

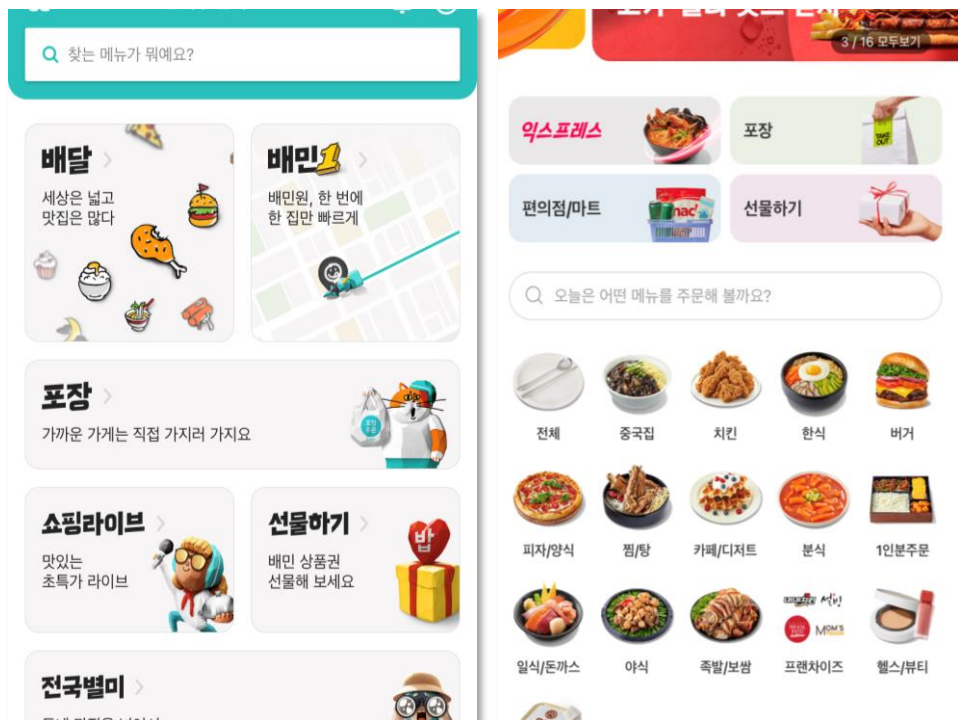
▲ 미니맵

현재 다양한 기업들이 현대인들의 라이프스타일에 맞게 추천 서비스를 제공하고 있으며, 추천 시스템에 대한 니즈 꾸준히 증가하고 있음

서론

연구배경 및 문제현황

1) 데이터 기반 식사 메뉴 추천 시스템이 필요한 이유



▲ 배달의 민족

▲ 요기요

그러나 현재 음식 어플 중
가장 많이 사용하는 배달 어플의 경우,
추천 서비스를 따로 제공하지 않고 있음

YTN | 2021.08.17. | 네이버뉴스

10명 중 4명, 배달 음식점 '후기' 보고 선택...절반 불만족

소비자 10명 가운데 4명이 배달 앱 음식점을 선택할 때 별점 리뷰 등 이용 후기를
최우선으로 고려한다고... 조사했더니, 절반에 가까운 소비자 48.9%가 후기를 보고



또한, 많은 소비자들이 식당의 리뷰를 보고
메뉴를 선택하지만
이에 대한 불만을 가지고 있음

서론 연구배경 및 문제현황

2) 현재의 식사 메뉴 추천 서비스의 한계점



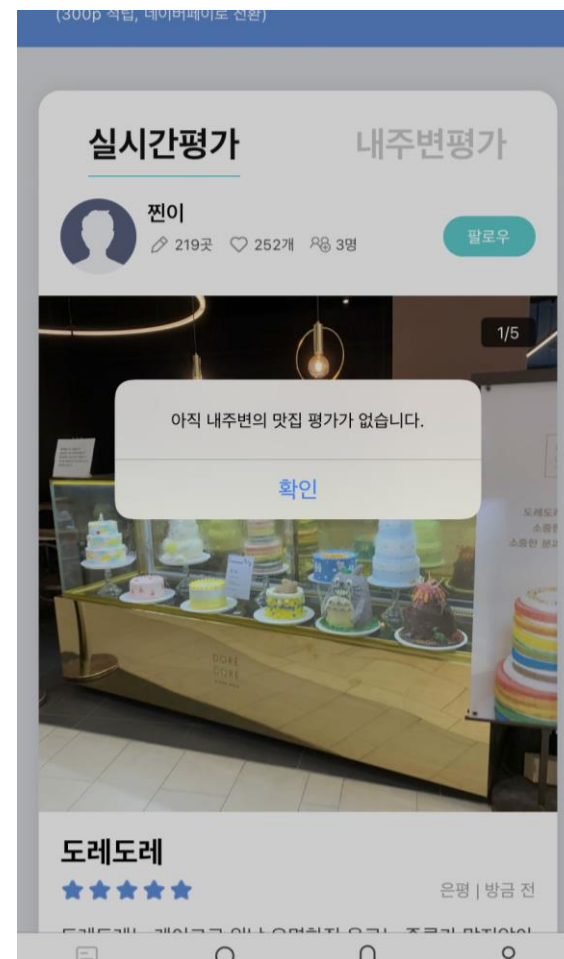
▲ 어플-뭐 먹을까?

랜덤 메뉴 추천 서비스

랜덤 메뉴 추천서비스로,
사용자가 원하는 자신의 데이터 기반
추천을 받을 수 없음

사용자 평가 기반 추천 서비스

이용활성화 부족으로 인해
원하는 정보를 추천 받을 수 없음



▲ 어플-다이닝코드

① 배경

현대인의
추천시스템에 대한
니즈



② 배경

그러나, 음식어플과
현재의 식사 메뉴 추천
서비스는 이용자들이
원하는 추천 정보를
받아볼 수 없음

개인의 상황과 특성을
반영해 맞춤형 식사메뉴를
추천해줄 방법은 없을까?



머신러닝 기법과 통계적
분석 기법을 사용해
고객 맞춤 식사 메뉴를
추천해주는 서비스를
구축하고자 함



데이터 수집 및 전처리 개요

Framework

'지역명(서초구) + 점심 메뉴 키워드로 포털사이트에 검색 후 반환되는 블로그 게시물 제목&본문 크롤링



메뉴 추출 작업 후 KoNLPy 이용해 어떤 메뉴들이 많이 나오는지 분석

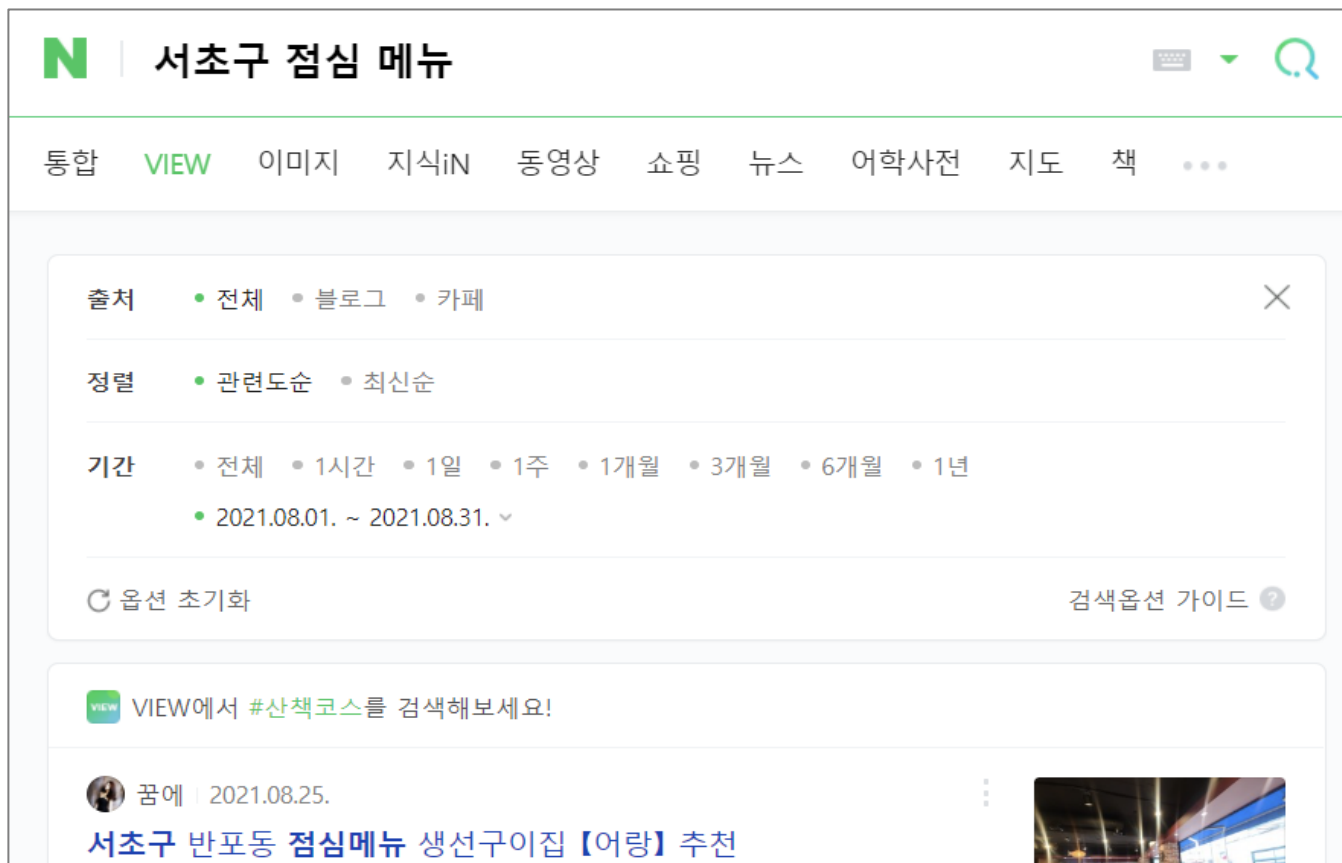


해당 데이터를 분류별로 묶어 식사메뉴(종속변수)로 사용할 변수 확정



설문조사를 통해 응답자들의 최근 식사메뉴와 이외 식사를 할 때 고려되는 요소들을 파악

1) 식사 메뉴 데이터 수집



▲네이버 블로그

- 데이터 수집 방법

: 네이버 블로그에서 “서초구 점심 메뉴” 키워드로 Jupyter Notebook 및 Python을 이용해 크롤링

- 데이터 수집 기간

: 2021.08.01~2021.08.31

- 데이터

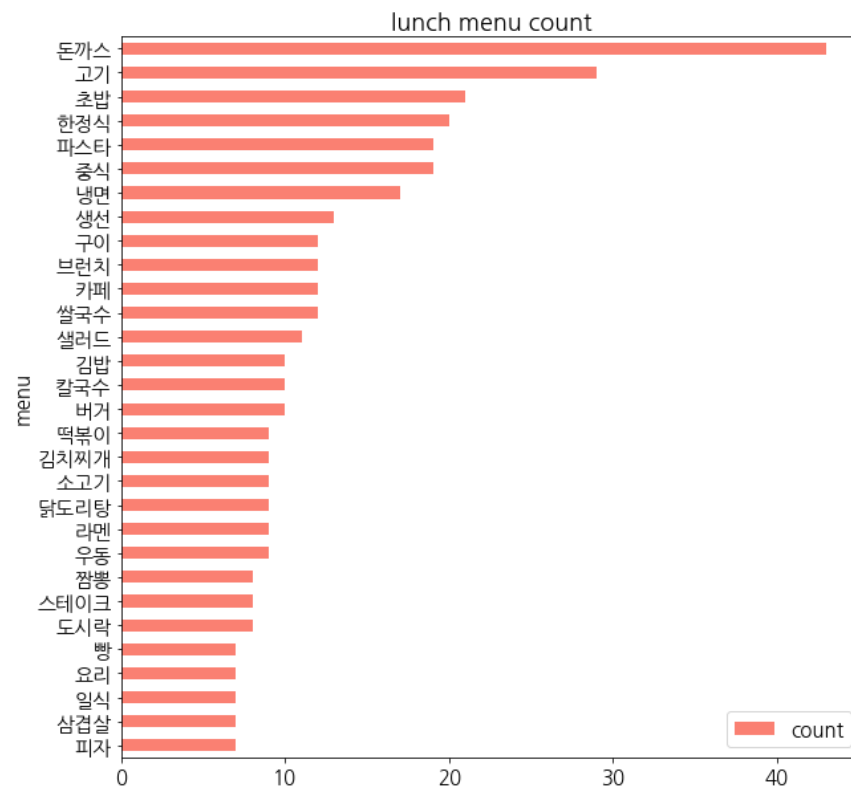
: “서초구 점심 메뉴“ 키워드를 가진 670개의 데이터를 대상으로 함

1) 식사 메뉴 데이터 수집 - Crawling(크롤링)

```
search_keyword = '서초구 점심 메뉴'
URL = 'https://search.naver.com/search.naver?where=blog&query={}&sm=tab_opt&nso=so%3Ar%2Op%3Afrom{}to{}'
#url = URL.format(quote(search_keyword), date, date)

#무한스크롤 & 크롤링코드
date_index = pd.date_range(start='20210801', end='20210831')
date_list = date_index.strftime("%Y%m%d").tolist()
```

제목	본문	날짜
[서울 강남] 조양관 _ 강남한정식 식당에서 즐기는 강남...	강남한정식 전문점 조양관 포스팅입니다.강남 보리굴비 맛집! 강남한정식 조...	2021.08.01.
2021.07.31~08.06	7/31 같이 일하는 동료가 빵 줬다 학학 서초 루엘드파리 썩 좋아하는 거알고 썩...	2021.08.01.
[일상일기] 즐거운기만나의 Happy_birthmonth, 21년 7월...	즐거운기만나가 제일 좋아하는 7월 일상일기오늘 안에 다 쓸 수 있을까? 사진 ...	2021.08.01.
7월의 일상	2021. 07. 03 토 점심에 사장님이랑 지코바 시켜먹엇당예전에 부산에만 지코...	2021.08.01.
내방역 중국집 '만다린' / 16년 전통 방배동 중식 맛집 추천	[내방역] 만다린- 02-596-6767- 서울 서초구 서초대로25길 17- 매...	2021.08.02.



2021.08.01~2021.08.31을 기간으로 설정해 '서초구 점심 메뉴' 키워드를 가진 블로그
게시글들을 크롤링 한 후, 전처리 작업을 수행해 식사 메뉴 단어들을 카운트 함

본론 데이터 수집 및 전처리 과정

1) 식사 메뉴 데이터 수집 - 크롤링 후 분류

	menu	count
0	돈까스	43
1	고기	29
2	초밥	21
3	한정식	20
4	파스타	19
5	중식	19
6	냉면	17
7	생선	13



범주형 분류	메뉴 분류
0	고기류(삼겹살, 소고기, 불고기 등)
1	한식(생선구이, 비빔밥, 제육 등)
2	일식(돈까스, 초밥, 우동, 라멘 등)
3	중식(짜장면, 짬뽕, 탕수육, 마라탕 등)
4	분식(떡볶이, 김밥, 튀김, 볶음밥 등)
...	...
11	아시안요리(쌀국수, 팟타이, 뽕양꿍 등)

카운트한 상위 100개의 메뉴 중 종속변수로 사용할 식사 메뉴들을
비슷한 음식군끼리 묶어 분류함

2) 식사 메뉴 추천 모델 구현을 위한 데이터 수집

점심메뉴 추천 서비스 구현을 위한 설문조사

안녕하세요. (주)코드프레소의 활동의 일환으로 데이터분석 프로젝트를 진행하고 있는 숙명여자대학교 학생입니다. 프로젝트 준비과정 중 '점심메뉴 추천 서비스' 구현을 위한 데이터가 필요해 다음과 같이 설문을 만들었습니다. 해당 설문조사는 5분정도 소요될 것으로 예상됩니다.

귀하의 귀중한 시간을 해당 연구에 참여해주셔서 감사드립니다.

해당 설문지는 점심메뉴 추천 서비스 알고리즘 구현을 위해 진행되는 것이며, 연구 목적으로만 활용될 예정입니다. 설문조사의 결과는 익명으로 처리되며 비밀이 보장됩니다. 또한, 조사 결과는 프로젝트 종료시 즉시 파기됨을 알려드립니다. 모든 문항에 솔직하게 답변해주시길 바랍니다.

추첨을 통해 기프티콘을 증정할 예정이니 많은 참여 부탁드립니다.

최근 외식 혹은 배달을 통해 점심에 먹은 음식은 무엇입니까?

▲ Google 설문지 사용

• 데이터 수집 방법

: 설문조사를 통해 직접 응답을 받아 데이터를 수집함

• 데이터 수집 기간

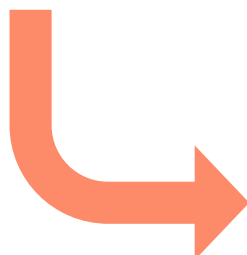
: 2021.09.07~2021.09.07

• 데이터 표본

: 전국 성인남녀 50명 대상

2) 식사 메뉴 추천 모델 구현을 위한 데이터 수집 - 전처리

타임프	최근 외식 혹은 배달을 통해 메인요리의 명칭은 무엇입	식사는 외식이었습니까, 배	귀하의 연령대는 어떻게	도 귀하의 성별은 어떻게 되	귀하의 직업상태는 어떻게	식사를 하실때 일행은 어떻	식사를 하시는 목적이 있	드	식당 혹은 배달의 대기시간	식당 선택 기준은 어
9-7-2021 0:34:33	고기류(삼겹살, 소고기, 불 석쇠불고기	배달	20대	남성	학생	친구	지인과의 만남	20분-30분	지인 추천	
9-7-2021 0:36:13	패스트푸드(치킨, 햄버거, 치킨	배달	20대	남성	무직	혼자 먹었다	혼밥	20분-30분	기존에 아는 가게이	
9-7-2021 0:37:34	분식(떡볶이, 김밥, 튀김, 떡볶이	배달	20대	여성	학생	가족	가족과의 외식 혹은 식사	20분-30분	기존에 아는 가게이	
9-7-2021 0:59:27	찜,탕(찜닭, 백숙, 삼계탕, 찜닭	배달	20대	여성	학생	가족	가족과의 외식 혹은 식사	40분-50분	쿠폰 또는 할인 프로	
9-7-2021 1:04:11	고기류(삼겹살, 소고기, 불 삼겹살, 소고기 부위별	외식	20대	여성	학생	가족	가족 생일 기념	10분-20분	기존에 아는 가게이	
9-7-2021 1:23:10	중식(짜장면, 짬뽕, 탕수육 짜장면	배달	20대	여성	학생	가족	가족과의 외식 혹은 식사	10분-20분	기존에 아는 가게이	
9-7-2021 1:34:25	중식(짜장면, 짬뽕, 탕수육 마라탕	배달	20대	여성	학생	혼자 먹었다	혼리	50분-60분	기존에 아는 가게이	



인코딩

식사메뉴	식사타입	성별	직업상태	일행	식사목적	대기시간	식당선택기	가격대	날씨	식당타입	식사가치	권만족여부
0	0	0	1	1	1	1	1	1	1	0	0	1
5	0	0	2	4	3	1	0	0	2	0	0	0
4	0	1	1	0	0	1	0	1	0	0	0	1
9	0	1	1	0	0	3	3	3	1	0	0	1
0	1	1	1	0	0	0	0	5	0	1	0	1
3	0	1	1	0	0	0	0	0	0	1	0	1
3	0	1	1	4	3	4	0	1	0	1	1	0

기존 20개 질문 중 식사메뉴(종속변수)에 영향이 있을 독립변수들만 간추려 범주형 데이터로 변환했으며, 결측치를 제거하는 데이터 전처리 과정을 이행했다.

또한, 새로운 데이터에 대한 예측을 위한 테스트 데이터 셋을 따로 만들었다.

2) 식사 메뉴 추천 모델 구현을 위한 데이터 수집 - 변수설명

Decision Tree 모델의 종속변수

* 각 변수들은 모두 범주형으로 분류함

식사메뉴	고기류/한식/일식/중식/분식/패스트푸드/양식/국밥 및 찌개/국수/짬.탕/카페/아시안요리				
식사타입	배달 / 외식	성별	남성 / 여성	직업상태	직장인 / 학생 / 무직
일행	가족 / 친구 / 애인 / 직장동료 / 혼자 먹었다				
식사목적	가족과의 외식 혹은 식사 / 지인과의 만남 / 데이트 / 배고픔을 해결하기 위함				
대기시간	10-20분 / 20-30분 / 30-40분 / 40-50분 / 50-60분 / 1시간 초과				
가격대	1,000-10,000원/10,000원-20,000원/20,000원-30,000원/30,000-40,000원/40,000원-50,000원/50,000원 초과				

종속변수

독립변수

본론 데이터 수집 및 전처리 과정

2) 식사 메뉴 추천 모델 구현을 위한 데이터 수집 - 변수설명

* 각 변수들은 모두 범주형으로 분류함

선택기준	기준에 아는 가게 / 리뷰를 통해 선택 / 프로모션이 있는 가게 / 가격이 합리적인 가게
날씨	맑음 / 흐림 / 비
식당타입	프랜차이즈 / 개인운영식당
식사가치관	음식의 맛이 중요하다 / 음식의 가격이 중요하다
만족여부	불만족 혹은 보통 / 만족

로지스틱 회귀 분석을 할 때 종속변수로 사용

종속변수

독립변수

식사 메뉴 추천 모델 구현 개요

Framework

ML 알고리즘 중 하나인 Decision Tree를 이용해 각 변수들을 Input해 식사메뉴를 예측할 수 있는 모델을 구축함



테스트 데이터셋의 식사 메뉴를 예측함



이후 로지스틱 회귀분석 기법을 이용해 모델을 학습시켜 만족/(불만족or보통) output 예측할 수 있는 모델 구축함



위에서 만든 식사 메뉴 예측 모델의 테스트 데이터셋에 로지스틱 회귀분석을 통해 학습시킨 만족여부 예측 모델을 적용시켜 만족/(불만족or보통) 예측

사용 모델 : Decision Tree Learning

결정 트리 학습법(decision tree learning)은 어떤 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델로서 결정 트리를 사용한다. 결정 트리 중 분류 트리를 활용해 트리를 구축할 예정이다.

```
train.head()
```

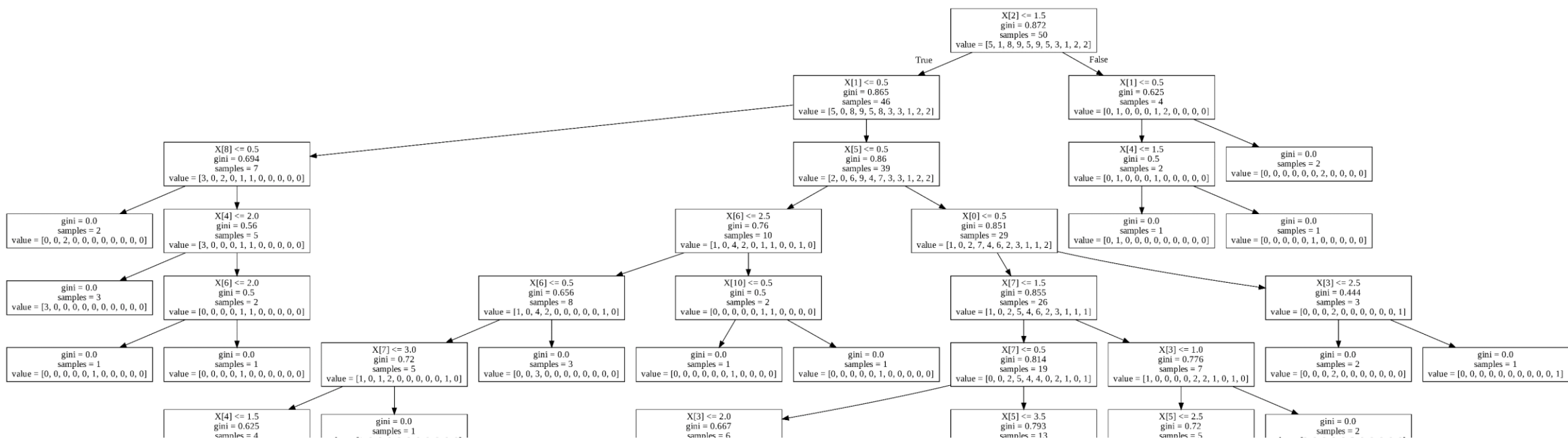
	식사메뉴	식사타입	성별	직업상태	일행	식사목적	대기시간	식당선택기준	가격대	날씨	식당타입	식사가치관
0	0	0	0	1	1	1	1	1	1	1	0	0
1	5	0	0	2	4	3	1	0	0	2	0	0
2	4	0	1	1	0	0	1	0	1	0	0	0
3	9	0	1	1	0	0	3	3	3	1	0	0
4	0	1	1	1	0	0	0	0	5	0	1	0

종속변수(Yi)

독립변수(Xi)

사용 모델 : Decision Tree Learning

Train 데이터를 사용해 모델을 학습시켜 의사결정트리를 완성했다.



* 의사결정트리 시각화 - 일부 생략

사용 모델 : Decision Tree Learning

테스트 데이터의 독립변수들에 따른 식사메뉴를 예측할 수 있었으며,
사용자의 특성 및 상황에 따른 식사메뉴를 예측할 수 있는 모델을 완성했다.

```
test['예측메뉴'] = y_pred
```

```
test.head()
```

	식사타입	성별	직업상태	일행	식사목적	대기시간	식당선택기준	가격대	날씨	식당타입	식사가치관	예측메뉴
0	0	1	1	1	1	2	4	0	0	0	0	3
1	1	1	0	2	2	0	2	3	1	1	0	2
2	1	1	1	1	1	1	0	1	0	1	0	3
3	0	0	1	4	3	0	4	0	0	1	1	2
4	1	0	0	0	0	2	2	2	2	0	0	0

메뉴예측완료

사용 모델 : Logistic Regression

로지스틱 회귀(Logistic Regression)는 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법이다. 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에 일종의 분류 기법으로도 볼 수 있다. 로지스틱 회귀 분석 기법 중 종속변수가 이항형인 로지스틱 회귀 분석을 진행할 예정이다.

1) 1차 로지스틱 회귀 분석 - 만족여부 예측을 위해 식사메뉴 변수 포함

train.head()

만족여부 - 보통or불만족(0), 만족(1)

	식사메뉴	식사타입	성별	직업상태	일행	식사목적	대기시간	식당선택기준	가격대	날씨	식당타입	식사가치관	만족여부
0	0	0	0	1	1	1	1	1	1	1	0	0	1
1	5	0	0	2	4	3	1	0	0	2	0	0	0
2	4	0	1	1	0	0	1	0	1	0	0	0	1
3	9	0	1	1	0	0	3	3	3	1	0	0	1
4	0	1	1	1	0	0	0	0	5	0	1	0	1

독립변수(Xi)

종속변수(Yi)

사용 모델 : Logistic Regression

1) 1차 로지스틱 회귀 분석 - 결과

Logit Regression Results

Dep. Variable:	만족여부	No. Observations:	50
Model:	Logit	Df Residuals:	37
Method:	MLE	Df Model:	12
Date:	Wed, 08 Sep 2021	Pseudo R-squ.:	0.6266
Time:	04:48:36	Log-Likelihood:	-8.8002
converged:	True	LL-Null:	-23.570
Covariance Type:	nonrobust	LLR p-value:	0.003275

	coef	std err	z	P> z	[0.025	0.975]
const	6.1268	4.819	1.271	0.204	-3.318	15.572
식사메뉴	-0.9809	0.630	-1.557	0.120	-2.216	0.254
식사타입	2.3101	3.634	0.636	0.525	-4.812	9.433
성별	5.4492	4.673	1.166	0.244	-3.709	14.608
직업상태	1.5497	1.829	0.847	0.397	-2.034	5.134
일행	-0.0921	0.842	-0.109	0.913	-1.742	1.558
식사목적	1.8300	1.374	1.331	0.183	-0.864	4.524
대기시간	-1.5609	1.154	-1.353	0.176	-3.823	0.701
식당선택기준	2.8271	1.384	2.043	0.041	0.114	5.540
가격대	-1.0395	1.035	-1.005	0.315	-3.068	0.988
날씨	-4.5732	2.417	-1.892	0.058	-9.311	0.164
식당타입	-7.0804	4.186	-1.692	0.091	-15.284	1.124
식사가치관	-1.8174	2.643	-0.688	0.492	-6.997	3.362

Pseudo R-square 값 0.6266으로
모델이 유의하며,
분석가능성이 있다고 판단했다.

유의수준(α)를 0.1로 검정했을 때,
유의한 p-value를 가지는 변수로는
식사메뉴, 대기시간, 식당선택기준, 날씨,
식당타입이 있었다.
(식사메뉴, 대기시간의 경우 p-value가
 α 보다 작다고 할 수 없지만
유의성이 있다고 판단)
따라서, 해당 변수들만을 가지고
2차로 로지스틱 회귀 분석을 진행한다.

사용 모델 : Logistic Regression

2) 2차 로지스틱 회귀 분석 - 결과

Logit Regression Results

```

=====
Dep. Variable:          만족여부   No. Observations:          50
Model:                  Logit      Df Residuals:                44
Method:                 MLE        Df Model:                   5
Date:                  Wed, 08 Sep 2021   Pseudo R-squ.:          0.4439
Time:                  04:54:41    Log-Likelihood:         -13.106
converged:              True      LL-Null:                 -23.570
Covariance Type:        nonrobust    LLR p-value:            0.0008362
=====

```

	coef	std err	z	P> z	[0.025	0.975]
const	6.9167	2.390	2.894	0.004	2.232	11.601
식사메뉴	-0.4311	0.256	-1.687	0.092	-0.932	0.070
대기시간	-0.9435	0.503	-1.877	0.061	-1.929	0.042
식당선택기준	1.0882	0.437	2.492	0.013	0.232	1.944
날씨	-2.1510	0.846	-2.544	0.011	-3.808	-0.494
식당타입	-3.0491	1.351	-2.256	0.024	-5.698	-0.400

Pseudo R-square 값이 1차 분석 때보다 낮아졌지만 0.4439로 유의하다고 판단. 유의한 p-value값이 많아지면서 모델의 설명력이 더 좋아졌다.

유의수준(α)을 0.1로 가정했을 때, 식사메뉴, 대기시간, 식당선택기준, 날씨, 식당타입 모두 유의한 변수로 판단됐다. 따라서 해당 변수들은 식사만족여부에 영향을 미친다고 할 수 있다.

사용 모델 : Logistic Regression

3) 테스트 데이터에 로지스틱 회귀 모델 적용

앞에서 구축한 로짓 회귀 모델을 테스트 데이터셋에 적용해 만족여부를 예측했으며, 최종적으로 모델을 완성했다.

```
Y_pred=logit.predict(X_test)
```

```
test['만족여부예측'] = Y_pred
```

```
test.head()
```

만족여부 예측 완료

	식사타입	성별	직업상태	일행	식사목적	대기시간	식당선택기준	가격대	날씨	식당타입	식사가치관	예측메뉴	만족여부예측
0	0	1	1	1	1	2	4	0	0	0	0	3	0.999693
1	1	1	0	2	2	0	2	3	1	1	0	2	0.953949
2	1	1	1	1	1	1	0	1	0	1	0	3	0.836295
3	0	0	1	4	3	0	4	0	0	1	1	2	0.999363
4	1	0	0	0	0	2	2	2	2	0	0	0	0.948043

테스트 데이터셋에 모델을 적용한 결과



A씨

식사타입	성별	일행	날씨	...	예측메뉴	만족여부예측
배달	여성	친구	맑음	...	중식	만족



B씨

식사타입	성별	일행	날씨	...	예측메뉴	만족여부예측
배달	남성	혼밥	맑음		일식	만족



C씨

식사타입	성별	일행	날씨	...	예측메뉴	만족여부예측
외식	여성	연인	흐림	...	일식	만족



위 결과와 같이, 모델을 적용해 새로운 데이터에 대한
식사메뉴와 만족여부를 예측할 수 있음

기대효과 ①

사용자의 특성과 상황에 맞는
식사 메뉴 추천 가능



현대인의 추천 서비스에 대한
니즈를 충족



만족도 극대화

기대효과 ②

다양한 데이터들을 함께 활용해
응용할 수 있음

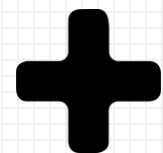


다른 서비스로의 확장가능성이
존재함

① 기존 서비스에 기능 추가



기존의 룰렛, 위치기반 랜덤 메뉴 추천 서비스



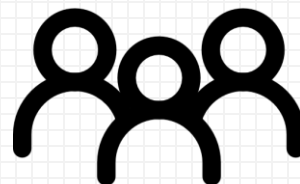
개인 맞춤형 식사 메뉴 추천 서비스 기능

② 맛집, 사용자 평점 데이터 수집해 새로운 시스템 개발

개인 맞춤 메뉴 추천 시스템

맛집 텍스트 리뷰

사용자 평점



이용자 커뮤니티 기반 종합 추천 시스템 개발



감사합니다

