

텍스트 리뷰를 활용한 평가 항목 점수 예측

호텔관광경영학과 17010187 장준원
데이터사이언스학과 19011715 임다영

Contents

1. 문제상황 정의
2. 관련 내용
3. 데이터셋 구축
4. 전처리
5. 모델링
6. 결과
7. 한계 및 개선방향

1. 문제상황 정의

- 긴 텍스트보다 별점이 리뷰에 대한 정보를 집약적으로 나타냄

ke**님 2021년 11월 11일

신고

★★★★★ | 맛 ★ 5 양 ★ 5 배달 ★ 5



기네스머쉬룸와퍼 세트/1(음료변경(코카-콜라R (변경 없음)),사이드변경(프렌치프라이 (R))),21치즈스틱/2

늘, 버거엔 진심인 버거킹셰프님들께서 세심하게 만들어주심에 버거는only 갓거킹에서만 주문합
지요:D 기사님께서서는 눈썹을 휘날리시며 날아와주시어 벨요청에도 불구하고 문자까지 세심하게 날
려주신 덕에, 이 쌀쌀한 추위속에서도 마음만은 따스히 불타올랐답니다:) 항상 감사드리고 오늘도 든
든한 힘이 됩니다! #식후감 #치유는 사람으로써 #버거킹은 못참치

so**님 2021년 11월 11일

신고

★★★★★ | 맛 ★ 5 양 ★ 4 배달 ★ 5

더블화이트갈릭팩1/1(음료변경(코카-콜라R (변경 없음)),음료변경2(코카-콜라R (변경 없음))),코코넛쉬림프...

콜라도 흐르지 않았고 빠르게 잘 도착했습니다 햄버거가 납작(?)하게 도착해서 그렇지 맛은 보장되
잖아요 ^^ 배달상태로 보면 햄버거의 납작함이 이해되진 않아요 주방에서? 아님 보기보다 내용물
이 알차지 않은건가?

> 별점 확인이 직관적임!

1. 문제상황 정의

- 프로젝트의 활용 방향

up**님 2021년 9월 25일
★★★★★ | 맛 ★ 3 양 ★ 3 배달 ★ 3

후라이드치킨/1,참이슬 후레쉬/2
오랜만에 생각나서 시켜봤는데 맛있습니다

신고



ch**님 2021년 11월 4일
★★★★★ | 맛 ★ 3 양 ★ 4 배달 ★ 3

소시지 레볼루션 치즈를 크러스트 L/1(추가 선택(치즈 스틱))
깔끔하고 맛있고 다른 파자 브랜드에 비해 심플해서 좋다

신고

me**님 2021년 1월 21일
★★★★★ | 맛 ★ 4 양 ★ 4 배달 ★ 4

페퍼로니 피자 (오리지널 도우 L) /1,코카콜라 1.25L/1
도미노 페페로니 피자는 항상 맛있네요

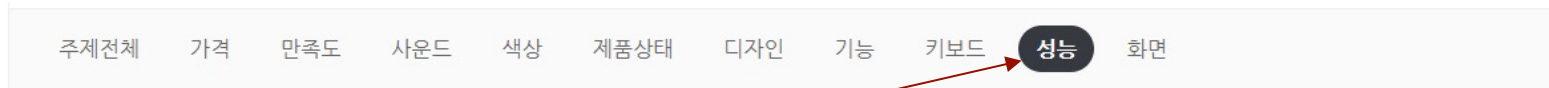
신고



**텍스트와 별점의 평가가 상이한 데이터의 경우,
모델의 결과값 바탕으로 향후 대처(ex. 필터링) 가능하다..!**

2. 관련 내용 - 네이버쇼핑 상품평

- 네이버쇼핑 > 상품별 리뷰



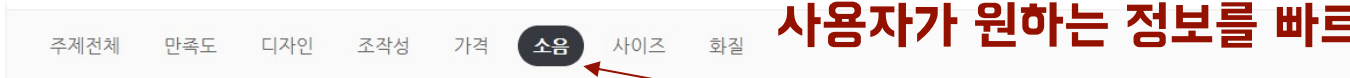
★★★★★ 5 할리스애플공인판매점 · heej**** · 20.07.07.

화면이 크니 시원시원하고 성능도 정말 좋아요. 다만 발열이.. 지금도 맥북으로 리뷰쓰는 중인데 손에 땀나네요... ㅠ

화면이 크니 시원시원하고 성능도 정말 좋아요.

다만 발열이.. 지금도 맥북으로 리뷰쓰는 중인데 손에 땀나네요... ㅠ

**리뷰에서 해당 키워드와 관련된 상품의 평가를 추출해
사용자가 원하는 정보를 빠르게 전달할 수 있다!**



★★★★★ 5 11번가 · ad***** · 21.05.08.

최고예요

원래 모니터랑 본체로 사려고 했었는데 코로나의 영향으로 보류중이었다가 선이 없이 쓰는 것이 깔끔해 보여서 일체형으로 사게 되었습니다선없는것도 좋았지만 팬돌아가는 소리가 거의 안나서 그것도 마음에 듭니다다만 전원잭이 받침대 뒷부분 경가운데 있어서 책상이 벽에 붙어있으면 모니터 받침대가 많이 밀려나와서 공간을 많이 차지하게 되네요일체형pc도 많이 좋아져서 업그레이드 되고 성능도 좋고 저는 만족합니다

리뷰펼치기 ▾

★★★★★ 4 11번가 · so***** · 21.08.28.

좋아요

밤에 주문해서 다음날 바로 발송해주었습니다. 일체형PC는 처음 써보는데 깔끔하고 좋을것 같습니다 음악을 듣기에 사운드가 그리 좋지는 않지만 소리가 크긴해서 들을만 합니다 원도가 깔려있어서 참 좋고, 원도를 재설치하는 것과 같은 리셋기능이 있다고 해 더 좋을것 같습니다 시작시간은 약간 걸리지만 종료시간은 순식간에 꺼집니다 화면 넓어서 좋습니다 히든 카메라 있어 좋습니다

3. 데이터셋 구축 - 요기요 리뷰 데이터

- 요기요 > 가게별 리뷰
 - 맛, 양, 배달 3가지 카테고리로 별점이 나뉘어져 존재

kg**님 2일 전

신고

★★★★★ | 맛 ★ 4 양 ★ 3 배달 ★ 5

육회혼술세트/1(메뉴 선택(육회90g + 고추장육회90g))

비록 혼술세트지만 양이 조금 아쉬워요

eb**님 2021년 10월 12일

신고

★★★★★ | 맛 ★ 5 양 ★ 5 배달 ★ 5

연어회빔밥/1,생연어초밥 (6pcs) /2,연어사시미 18pcs/1

연어 메뉴를 싹 주문해 본 것 같은데 사시미, 초밥, 회덮밥 세가지 다 너무 만족스러웠습니다. 비린내나 잡내에 민감한 편인데도 그런걸 못느끼고 맛있게 먹은걸 보면 기본적으로 좋은 연어를 쓰시고 식자재 관리에도 많은 노력을 하고 계신게 느껴졌어요. 서비스 김치전도 넘 맛있어서 메뉴를 다시 살펴봤는데 사이드로 판매하는 메뉴는 아니군요ㅠㅠ 주문하시는 분들 꼭! 리뷰 약속 남기시고 김치전도 같이 맛보세요 ㅎㅎ

3. 데이터셋 구축 - 크롤링

- Selenium, BeautifulSoup 사용 > 리뷰 데이터 CSV로 저장 자동화
- Target : 세종대학교 주변 음식점 (69개)

```
# 요기요 크롤링
from selenium import webdriver
import time
from bs4 import BeautifulSoup
import re
import pandas as pd

# url 입력
driver = webdriver.Chrome( )
url = "https://www.yogiyo.co.kr/" # 사이트 입력
driver.get(url) # 사이트 오픈
driver.maximize_window() # 전체창
time.sleep(2) # 2초 지연

# 검색창 선택
xpath = '//*[@id="search"]/div/form/input' # 검색창
element = driver.find_element_by_xpath(xpath)
element.clear()
time.sleep(2)

# 검색창 입력
value = input("지역을 입력하세요")
element.send_keys(value)
time.sleep(2)

# 조회버튼 클릭
search_xpath = '//*[@id="button_search_address"]/button[2]'
driver.find_element_by_xpath(search_xpath).click()
```

3. 데이터셋 구축 - 크롤링

- CSV로 저장된 데이터 (총 172,615개 크롤링)

- Name(음식점 이름), taste(맛 점수), quantity(양 점수), delivery(배달 점수), text(리뷰)로 구성

	name	taste	quantity	delivery	text
0	귀한족발-직영점	5	5	5	족발만 먹어봤는데 보쌈도 맛있어요!!!
1	귀한족발-직영점	4	5	5	고기가 부드럽고 맛나요. 찬들도 맛이 좋아요.
2	귀한족발-직영점	5	5	5	보쌈 너무 맛있어요
3	귀한족발-직영점	5	5	5	보쌈냄새하나도안나고 부드럽고 너무맛있어요~
4	귀한족발-직영점	5	5	5	오 여기 괜찮네요\n\n배달음식에 만족 한거 오랜만..
5	귀한족발-직영점	5	5	5	배달도 빨리왔고 서비스는 생각도 안했는데 챙겨주셔서 감사하 잘 먹었습니다. 맛도 있...
6	귀한족발-직영점	5	5	5	역시 양 많고 맛있어요! \n여기서만 주문해서 먹어요~ 고기 상태도 정말 좋아요~ ...
7	귀한족발-직영점	5	5	5	오영 여기 진짜 맛있어요...인생 족발 찾았어열..
8	귀한족발-직영점	5	5	5	맛있게 잘 먹었어요~
9	귀한족발-직영점	5	5	5	맛있게 잘 먹었습니다
10	귀한족발-직영점	5	5	5	맛있습니다!!!!
11	귀한족발-직영점	5	5	-1	양도 많고 너무 맛있어요 ㅎㅎㅎ 짬짬
12	귀한족발-직영점	5	5	5	맛있게 먹었습니다.
13	귀한족발-직영점	5	5	5	두번째 주문인데 앞으로 귀한족발에서만 주문하려고요!! 여기 부추 존맛인데 사장님께 ...
14	귀한족발-직영점	5	5	5	너무맛있고 양도많고 행복하게 먹었어요★
15	귀한족발-직영점	5	5	5	배달이 엄청 빨리 와서 놀랬어요 양도 많구 따뜻하고 맛있게 잘 먹었습니다 감사해요!
16	귀한족발-직영점	5	5	5	처음 먹어봤는데...\n족발 간이 세지 않고 살코기가 부드러워 맛있게 먹었습니다.\n...
17	귀한족발-직영점	5	5	5	이 지점은 처음 주문해봤는데 맛있었어요 또 시킬게요!
18	귀한족발-직영점	5	5	5	가끔 포장해서 먹다가 이번에 배달을 이용했는데 역시 맛있네요.\n잘 먹고 있습니다
19	귀한족발-직영점	5	5	5	족발이 짜지 않고 아주 부드럽고 맛있었어요~\n미국수도 너무 자극적이지 않으면서 적...
20	귀한족발-직영점	4	4	5	맛있게잘먹었습니다~~~
21	귀한족발-직영점	5	5	5	맛있게 잘 먹었습니다~
22	귀한족발-직영점	5	5	5	보통 진짜맛있거나 문제있을때만 리뷰쓰는데 진짜 맛있어요. 위생걱정때문에 배달 못시켜...
23	귀한족발-직영점	5	5	5	배달이 진짜 빠르고 진짜 맛있었어요! 요청사항도 잘 들어주시고 ㅎㅎ푸짐하게 맛있게 ...
24	귀한족발-직영점	5	5	5	맛도 맛있고 무엇보다 배달속도가...15분도 안되서 오네요 ㅎㅎ 맛있게잘먹었습니다

4. 전처리

- 정규표현식을 활용해 특수기호 및 이모티콘 제거

* 족발인데 보쌈으로 잘못 배달되었어요~ 그치만 전화하자마자 바로 족발배달해주셨고 덕분에 두 메뉴를 맛보게 되었네요ㅎㅎㅎㅎㅎ
둘 다 정말정말 맛있었는데 의외로 잘못배달된 보쌈이 엄청 맛있어서 또 생각나요😋 조만간 또 시켜먹을게요~

* 너무 맛있어요. 🥰🥰🥰🥰🥰🥰

* 너무맛있고 양도많고 행복하게 먹었어요★



* 족발인데 보쌈으로 잘못 배달되었어요~ 그치만 전화하자마자 바로 족발배달해주셨고 덕분에 두 메뉴를 맛보게 되었네요ㅎㅎㅎㅎㅎ
둘 다 정말정말 맛있었는데 의외로 잘못배달된 보쌈이 엄청 맛있어서 또 생각나요 조만간 또 시켜먹을게요~

* 너무 맛있어요.

* 너무맛있고 양도많고 행복하게 먹었어요

- 네이버 맞춤법 검사기 바탕의 py-hanspell을 활용해 맞춤법 검사를 진행

* 족발인데 보쌈으로 잘못 배달되었어요~ 그렇지만 전화하자마자 바로 족발 배달해주셨고 덕분에 두 메뉴를 맛보게 되었네요ㅎㅎㅎㅎㅎ
둘 다 정말 정말 맛있었는데 의외로 잘못 배달된 보쌈이 엄청 맛있어서 또 생각나요 조만간 또 시켜 먹을게요~

*너무 맛있어요.

*너무 맛있고 양도 많고 행복하게 먹었어요

4. 전처리

- 아래와 같이 '맛', '배달', '양' 과 관련된 형태소가 없는 경우, 텍스트 기반 별점 추론이 매우 어려움

맛	배달	양	리뷰
3	3	2	주꾸미가 몇 개 없는데 다 아기들이네요. ㅠㅠ
3	3	3	그냥 그래요....
5	5	5	앞으로 여기서 자주 시켜 먹을 듯 너무 좋았다 진짜로
3	2	4	정식이 두 개인데 국물은 하나만 와서 전화 여러 번 했는데 안 받으시네요 앞으로 꼼꼼히 좀 챙겨주세요.
1	5	1	전에도 그러시더니 또 비빔장에 고추장을 안 주셨어요 ㅠ.ㅠ 이제 여기서 안 시켜 먹을래요.
4	3	4	솔직한 리뷰 달면 영업에 반영될 것 같아 별점으로 만 남기겠습니다
5	5	5	좋아요 가끔 시켜 먹는데 항상 기대 이상은 하는 거 같아요^^
3	5	4	주인 대응 불친절합니다

4. 전처리

- Rule-based 방법을 도입해 '맛', '배달', '양'과 관련된 형태소가 존재하는 리뷰들만 필터링해 모델링을하기로 결정
- Konlpy의 Mecab 형태소 분석기를 활용해 빈도수 기준 상위 400개의 형태소에 대해서 '맛', '배달', '양'과 관련된 형태소를 추출

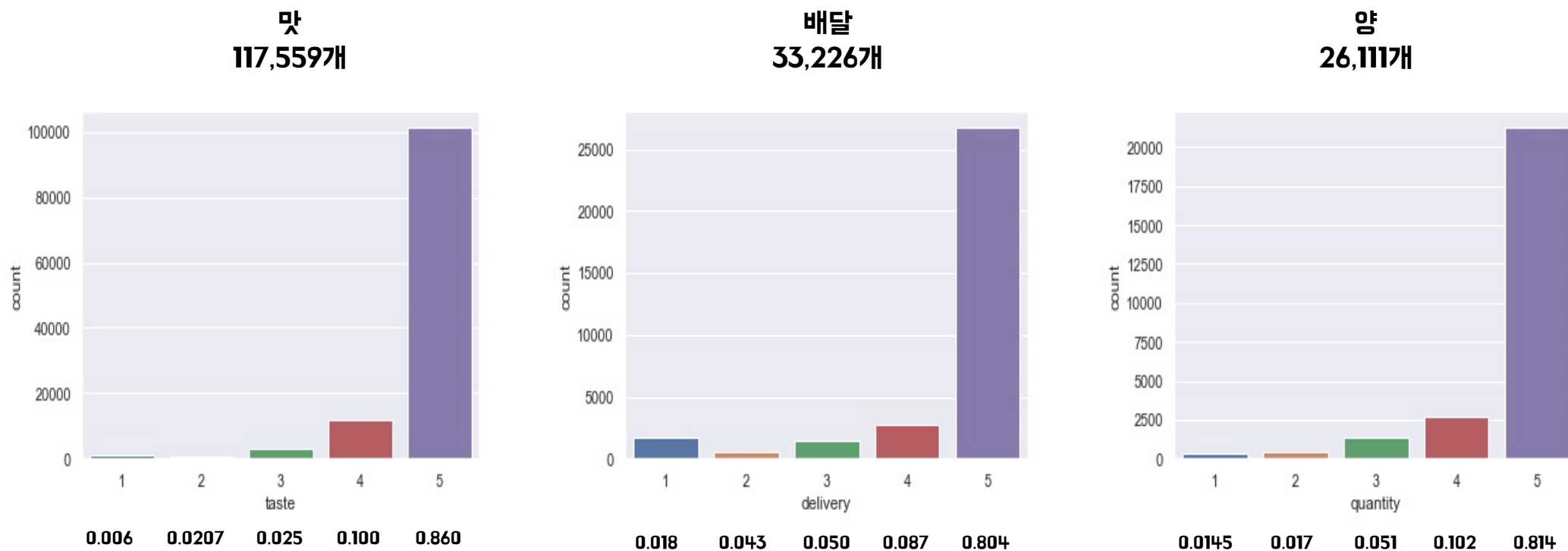
항목	추출한 형태소
맛	'맛' or '느끼' or '짜' or '달' or '부드' or '눅눅' or '매콤' or '맵' or '매운'
배달	'배달' or '시간' or '배송' or '빠' or '빨' or '시간' or '늦' or '느' or '도착'
양	'양' or '많' or '푸짐' or '적' or '조금' or '적당'

- 각 항목별로 데이터셋 및 모델을 독립적인 파이프로라인으로 구축해 실험을 진행함



4. 전처리

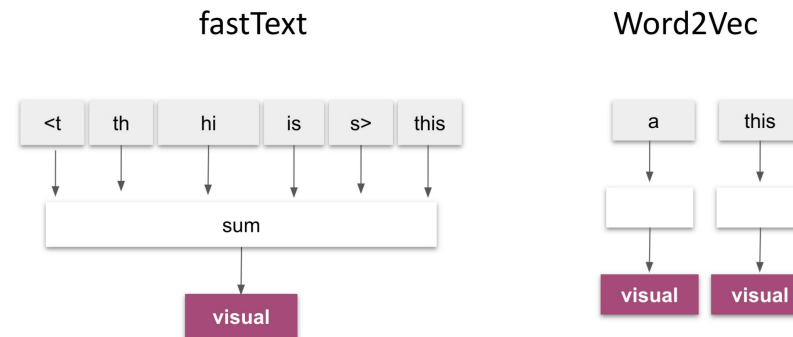
- 데이터셋 통계량 및 분포



- 3개의 데이터셋 분포가 모두 5점에 치우쳐져 있음
- 계층적 샘플링을 활용해 훈련데이터와 테스트데이터를 0.9 : 0.1의 비율로 나누어서 분포가 유지되도록 함

5. 모델

- Machine Learning - Embedding
 - 자연어를 컴퓨터가 이해하도록 하기 위해서는 수치형 데이터 (ex. 벡터)로 변환해주어야 함
 - 리뷰 텍스트를 ML 모델의 입력 Feature로 활용하기 위해 Fasttext Embedding 기법을 활용
- Fasttext Embedding
 - Facebook에서 개발한 Embedding 방법론으로, Subword를 활용해 Word2vec의 OOV문제를 해결한 Embedding
 - 각 단어를 자기 자신과 N-gram의 조합으로 표현함
Ex) apple = <ap + app + ppl + ppl + le> + <app + appl + pple + ple> + <appl + pple> + , ..., + <apple> (n=3~6)
 - 한국어의 형태론적 특징을 반영하기 쉬워, 다른 Embedding에 비해 성능이 우수함

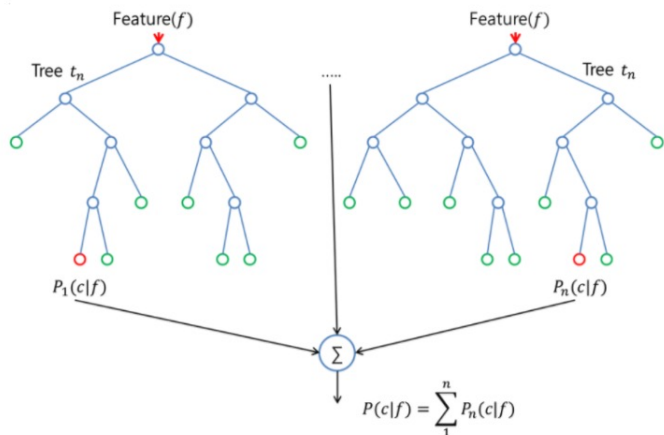


[Fasttext vs Word2Vec]

5. 모델

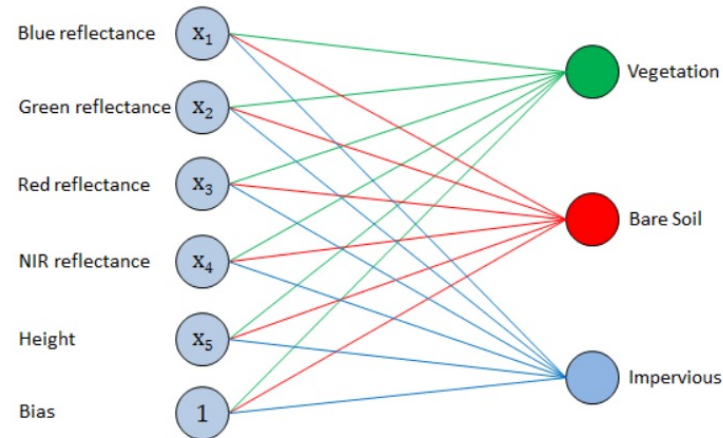
- Machine Learning Model - Random Forest, SGD Classifier, XGBoost

Random Forest



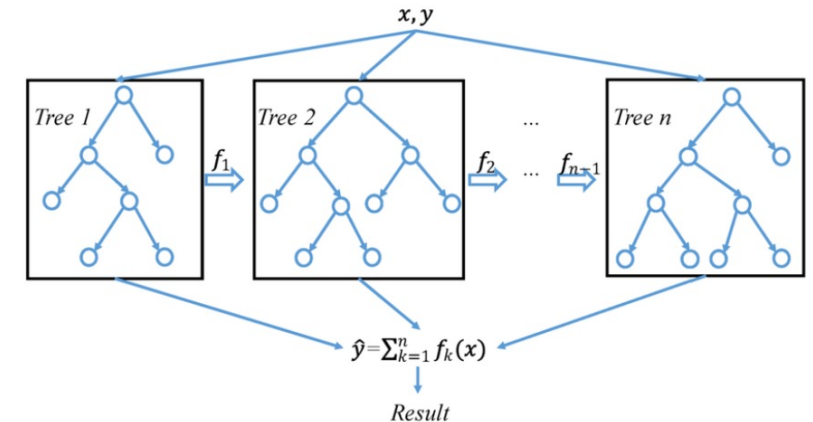
여러 개의 Decision Tree를 형성하고 새로운 데이터 포인트를 각 트리에 동시에 통과시키며, 각 트리가 분류한 결과에서 투표를 실시하여 가장 많이 득표한 결과를 최종 분류 결과로 선택하는 Ensemble 모델

SGD Classifier



Gradient Descent를 전체 데이터(Batch)가 아닌 일부 데이터의 모음(Mini-Batch)를 대상으로 진행해 학습을 진행하는 모델

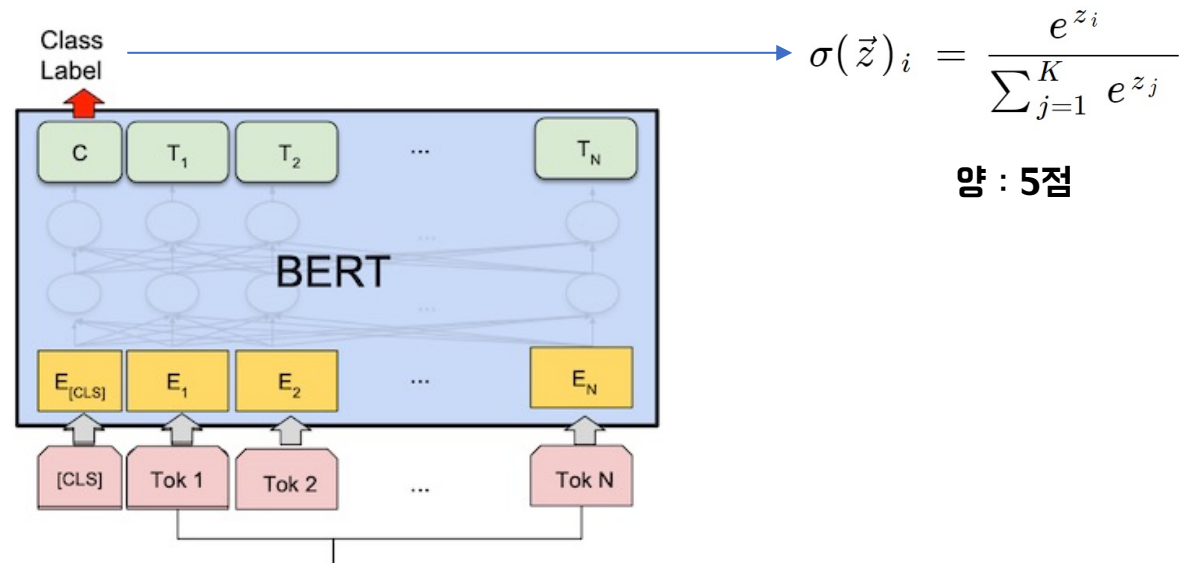
XGBoost (Extreme Gradient Boosting)



Ensemble Boosting 모델의 한 종류로, 이전 모델에서의 예측값과 실제값의 오차(Loss)를 훈련데이터에 반영하는 Gradient를 이용하여 오류를 보완하는 모델

5. 모델

- Deep Learning - BERT
 - BERT(Bidirectional Encoder Representations from Transformers)는 RNN계열의 장기 의존성 및 병렬처리 문제를 해결한 Transformer의 Encoder를 활용해 만든 딥러닝 모델
 - 2018년에 발표됨과 함께 대부분의 자연어처리 task에서 SOTA를 달성했으며, 현재까지 자연어처리 분야의 주류가 되는 모델
 - SKT-Brain에서 공개한 한국어 pretrained BERT인 KoBERT를 각 항목별 데이터셋에 fine-tuning해 실험을 진행



양 : 5점

앞으로 파스타는 이곳에서만 시켜 먹을 예정입니다. 다음에는 로제나 토마토소스로 주문해보겠습니다. 스테이크 등은 처음 시켜봤는데 가격 대비 양과 질이 뛰어나서 무척 만족스럽습니다. 오래오래 좋은 음식 부탁드립니다. 감사합니다

6. 결과 - 정량적

- ML - Random Forest
 - Precision, Recall, F1, Accuracy

맛

	Precision	Recall	F1-Score
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.86	1.00	0.93
accuracy			0.86
macro avg	0.17	0.20	0.19
Weighted avg	0.74	0.86	0.80

배달

	Precision	Recall	F1-Score
1	1.00	0.03	0.06
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.81	1.00	0.89
accuracy			0.81
macro avg	0.36	0.21	0.19
Weighted avg	0.70	0.81	0.72

양

	Precision	Recall	F1-Score
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	1.00	0.01	0.03
4	0.00	0.00	0.00
5	0.81	1.00	0.90
accuracy			0.81
macro avg	0.36	0.20	0.19
Weighted avg	0.71	0.81	0.73

- Random Forest는 '맛' 을 전부 5점으로 예측함
- '배달'과 '양'도 대부분 5점으로 예측해 불균형한 레이블에 대처를 하지 못함

6. 결과 - 정량적

- ML - SGD Classifier
 - Precision, Recall, F1, Accuracy

맛

	Precision	Recall	F1-Score
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.86	1.00	0.93
accuracy			0.86
macro avg	0.17	0.20	0.19
Weighted avg	0.74	0.86	0.80

배달

	Precision	Recall	F1-Score
1	0.62	0.47	0.53
2	0.00	0.00	0.00
3	0.42	0.08	0.13
4	0.00	0.00	0.00
5	0.84	1.00	0.91
accuracy			0.83
macro avg	0.38	0.31	0.32
Weighted avg	0.72	0.83	0.77

양

	Precision	Recall	F1-Score
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.35	0.18	0.24
4	1.00	0.00	0.01
5	0.83	0.99	0.91
accuracy			0.82
macro avg	0.44	0.24	0.23
Weighted avg	0.80	0.82	0.75

- Random Forest와 마찬가지로 SGD Classifier 역시 '맛' 을 전부 5점으로 예측함
- Random Forest에 비해 '배달'과 '양'에 대해서는 다른 레이블로도 예측을 진행하나 불균형한 역시 레이블에 대처를 하지 못함

6. 결과 - 정량적

- ML - Xgboost
 - Precision, Recall, F1, Accuracy

맛

	Precision	Recall	F1-Score
1	0.42	0.12	0.19
2	0.25	0.01	0.03
3	0.47	0.05	0.09
4	0.46	0.02	0.04
5	0.87	1.00	0.93
accuracy			0.86
macro avg	0.49	0.24	0.25
Weighted avg	0.81	0.86	0.81

배달

	Precision	Recall	F1-Score
1	0.64	0.57	0.60
2	0.31	0.08	0.13
3	0.45	0.17	0.25
4	0.31	0.06	0.10
5	0.86	0.98	0.92
accuracy			0.83
macro avg	0.52	0.38	0.40
Weighted avg	0.78	0.83	0.79

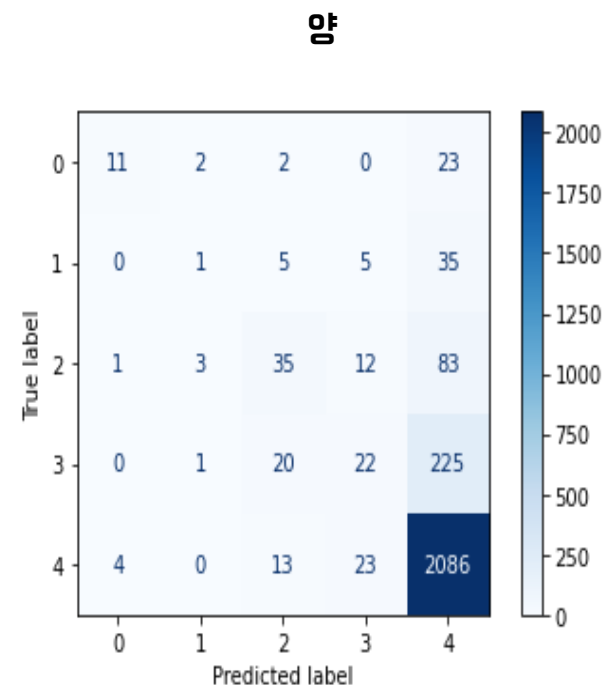
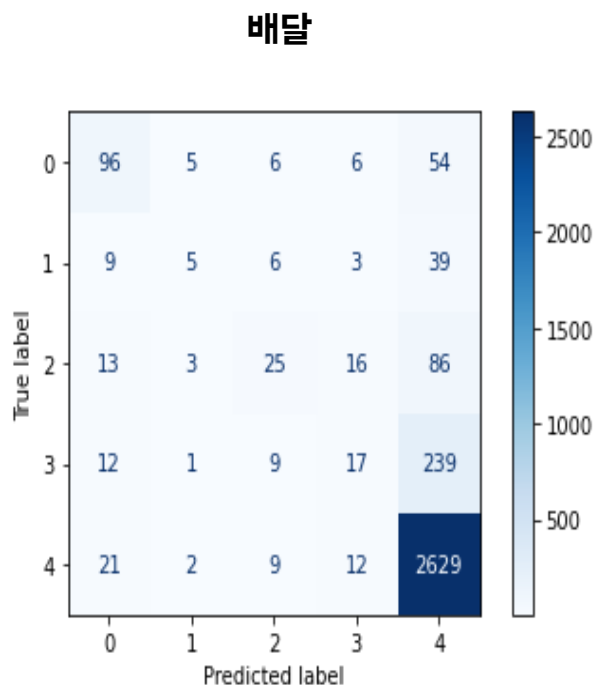
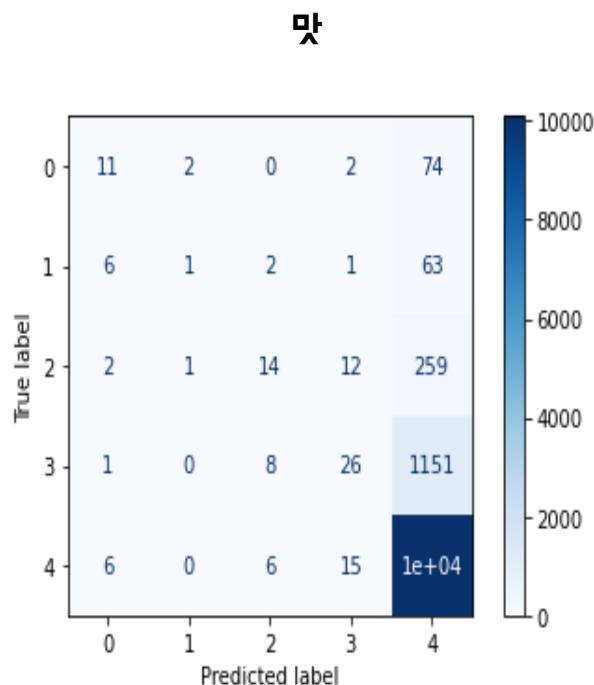
양

	Precision	Recall	F1-Score
1	0.69	0.29	0.41
2	0.14	0.02	0.04
3	0.47	0.26	0.33
4	0.35	0.08	0.13
5	0.85	0.98	0.91
accuracy			0.83
macro avg	0.50	0.33	0.36
Weighted avg	0.77	0.83	0.78

- 전체 sample을 5로 예측했을 때보다 정확도가 소폭으로 증가함
- 다른 ML 모델들에 비해서 5로 치우친 예측을 하는 경향이 떨어짐

6. 결과 - 정량적

- Xgboost - Result
 - Confusion Matrix



- 대부분 5로 예측하지만 다른 모델에 비해 5 이외의 점수도 예측하는 것을 확인할 수 있었음
- 데이터에서 5가 많아 5로 예측하는 경우가 여전히 많음

6. 결과 - 정량적

- BERT Result
 - Precision, Recall, F1, Accuracy

맛

	Precision	Recall	F1-Score
1	0.52	0.48	0.50
2	0.10	0.27	0.14
3	0.25	0.45	0.32
4	0.16	0.42	0.23
5	0.98	0.90	0.94
accuracy			0.87
macro avg	0.40	0.51	0.43
Weighted avg	0.93	0.87	0.90

배달

	Precision	Recall	F1-Score
1	0.83	0.60	0.70
2	0.00	0.00	0.00
3	0.35	0.35	0.35
4	0.16	0.40	0.23
5	0.97	0.92	0.94
accuracy			0.85
macro avg	0.46	0.45	0.44
Weighted avg	0.91	0.85	0.88

양

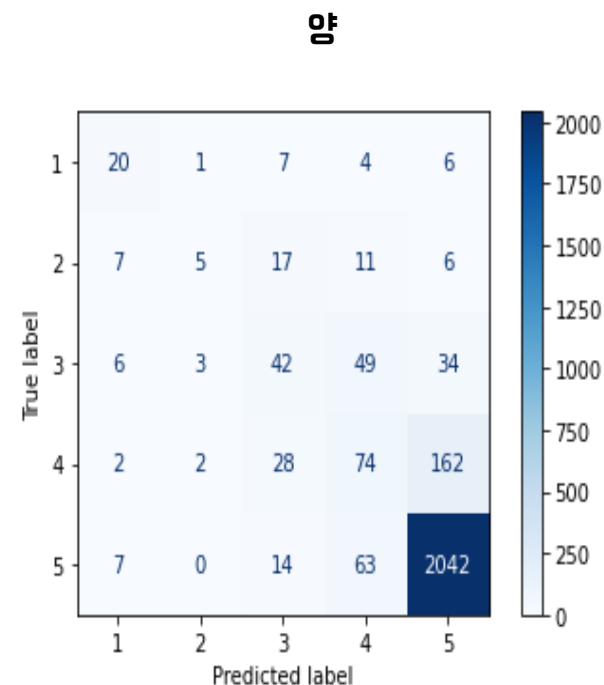
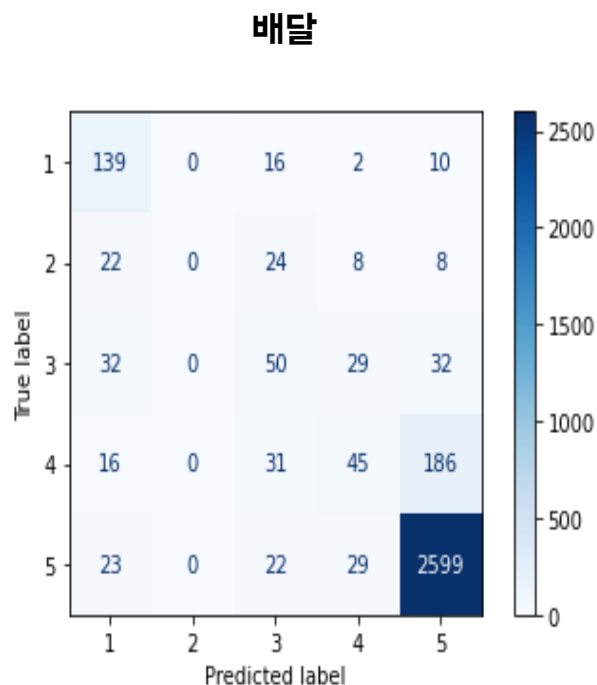
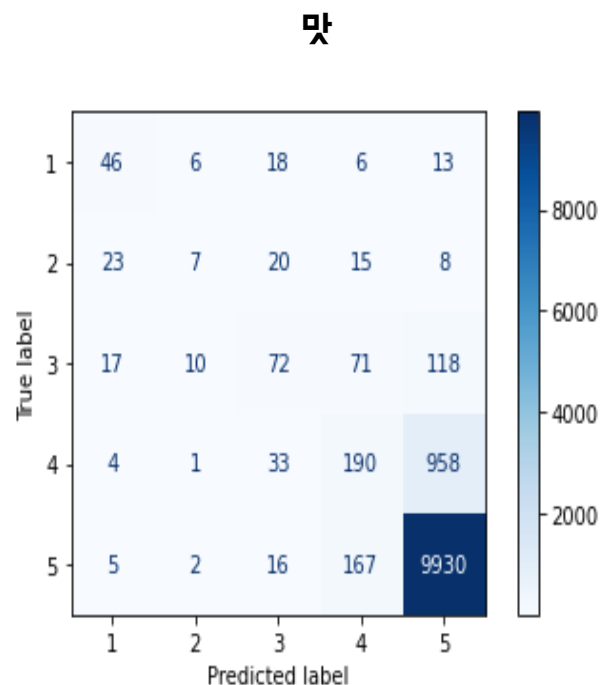
	Precision	Recall	F1-Score
1	0.53	0.48	0.50
2	0.11	0.45	0.18
3	0.31	0.39	0.35
4	0.28	0.37	0.32
5	0.96	0.91	0.93
accuracy			0.84
macro avg	0.44	0.52	0.45
Weighted avg	0.87	0.84	0.85

- 전체 sample을 5로 예측했을 때보다 정확도가 소폭으로 증가함

6. 결과 - 정량적

- BERT Result

- Confusion Matrix



- 대부분 5로 예측을 하나, '맛'과 '배달'의 경우에는 1로 예측을 하는 경우도 존재함
- '배달'의 경우, 2로는 전혀 예측을 하지 않음

6. 결과 - 정성적

- Xgboost Result
 - 잘못 예측했지만 결과가 어느정도 합당한 예측들

항목	실제	예측	리뷰
맛	3	1	햄버거가 .. 점점 맛이 없어져요.
	4	5	처음 먹어봤는데 맛있어요!
	4	3	너무 식어서 왔어요. 맛은 있네요
배달	3	1	배달시간 오래 걸리는 거 빼곤 좋음
	2	3	맛있게 잘 먹었습니다. 근데 좀 늦네용
	4	1	배달,, 너무 오래 걸리네요 ㅠㅠ
양	3	4	배달 잘 안 시켜 먹는데 맛있어서 또 시켜요 족발 양이 ... ㅠㅎ
	2	3	맛은 그럭저럭. 양은 흐 으음..... 배달은 빠름.
	4	3	프라이드는 역시 이집인 듯 요ㅎ양이 살짝 적은 거 같지만 잘 먹겠습니다

6. 결과 - 정성적

- BERT Result
 - 잘못 예측했지만 결과가 어느정도 합당한 예측들

항목	실제	예측	리뷰
맛	5	3	이게 무슨 피자죠. 맛이나 양은 그렇다 치고 배달이 문젠지 잘못 만든 건지...
	3	4	맛있어요. 배달도 빠르고 좋아요. 근데 짜장면 면은 별로입니다. 땡글땡글한 맛은 없어요.
	5	3	맛있는데 매운맛 시켰더니 너무 맵더라고요 ㅠㅠ담엔 보통맛 시켜야겠어요!!
배달	2	3	맛있게 잘 먹었습니다. 배달이 잘못 오고 또 늦게 와서.... 좀 아쉬웠지만 치킨은 맛있네요 ㅎㅎ
	5	3	배달이 좀 걸리네요.. 잘 먹었습니다
	3	1	맛은 좋은데 배달이 너무 오래 걸리네요ㅠ주문이 많아 그럴것죠?
양	2	3	대짜리 시켰는데.. 양도 너무 적고 젓가락도 2개밖에 안 주시고....
	3	4	오토케 너무 마식자남 그래서 양이 부족해쌔ㅠ
	2	1	맛은 그럭저럭. 양은 흐 으음..... 배달은 빠름.

7. 한계점 및 향후 개선 방향

- 한계점
 - ‘맛’, ‘배달’, ‘양’과 관련된 형태소가 존재하지 않으면 예측을 어려워 Rule-based 기반의 전처리를 활용함
(상위 빈도수 기반으로 관련 형태소 추출 시, ‘맛’ 대비 ‘배달’ 과 ‘양’ 관련 리뷰가 직관적이어서 성능이 좋은 것으로 판단됨)
 - 데이터셋의 부족으로 동일한 리뷰에 대해서 ‘맛’, ‘배달’, ‘양’을 한번에 예측하는 파이프라인을 구축하지 못함
- 개선 방향
 - Rule-based 기반의 전처리의 고도화
 - 동일한 리뷰에 대해서 ‘맛’, ‘배달’, ‘양’을 한번에 예측하는 파이프라인을 구축
 - Over Sampling (SMOTE)나 Under Sampling (Edited Nearest Neighbors)등을 활용해 레이블 불균형 문제 해결
(관련 방법론들이 수치형 데이터에는 잘 작동하나, 고차원 자연어 임베딩에서 잘 작동하지는 않아 추가적인 데이터 수집도 고려)
 - BERT 이외의 다른 Deep Learning 모델 활용
 - 이미지 정보도 활용하는 Multi-Modal Model 제안

감사합니다

Appendix - 형태소 기반으로 필터링된 리뷰

- '맛' 항목의 리뷰 (Test Data)

네 여러분들이 아는 그 맛이네요

맛있게 잘 먹었습니다 맛도 좋았고 가격 대비 양도 많았습니다

우선 양이 너무 많네요~! 맵기도 적당하고 튀김 맛집 인정ㅋ 많이 못 시킨 게 너무 아쉬울정도ㅠㅠ 돈으로 혼내주고 싶네요 맛있게 잘 먹었습니다 사장님ㅋㅋ

- '배달' 항목의 리뷰 (Test Data)

양도 괜찮고 배달도 생각보다 안 늦게 왔습니다 잘 먹었습니다

배달은 한 시간 정도 걸렸고 철판에 양이 적을 거 같다 생각했는데 고기가 큼지막해서 전혀 안 모자랐고요 마신했습니다!

배달 빠르고 프렌치프라이는 맛있게 됐는데 해시브라운이 덜 익어서 왔네요ㅠ

- '양' 항목의 리뷰 (Test Data)

양도 많고 부드러네요~또 시켜 먹을게요~ 정말 맛나요

양도 듬뿍 들어 먹기도 버거움 ㅋㅋㅋ

맛있네요. 양이 조금 적은듯하나 맛있습니다.