

카카오맵 맛집 평점의 긍정 편향 분석

Section 4 Project

AI 14기 주민선

Contents

1. 프로젝트 목적

2. 개발 프로세스

3. 결론 및 개선점

프로젝트 목적

제주도로 여행을 간다고 생각해 봅시다
여러분은 어떤 것부터 찾아보시나요?

예쁜 풍경



맛있는 음식



호캉스



프로젝트 목적

맛있는 음식을 원한다면 그 지역의 맛집을 우선 검색해야겠죠
어느 사이트에서 검색하든 평점과 후기가 가장 중요한 기준일텐데요
모든 식당의 평점이 너무 높게 책정되어 있는 것 같지 않나요?
평점이 4.9라서 굉장히 기대했는데 맛은 3점 정도인 경험, 한번쯤은 있을 겁니다

kakao

제주 서귀포 맛집

검색

길찾기

버스

장소 6,331

A 연돈 돈까스우동

3.9 ★★★★★ 507건 | 리뷰 1,028

제주특별자치도 서귀포시 일주서로 968-10 (지번) 색달동 2156-2

0507-1386-7060 · 상세보기 · 홈페이지

B 원앤온리 카페

4.0 ★★★★★ 538건 | 리뷰 2018

제주특별자치도 서귀포시 안덕면 산방로 14 (지번) 안덕면 사계리 86

영업중 | 매일 09:00 ~ 20:00

064-794-0117 · 상세보기 · 홈페이지

C 색달식당 중문갈치조림구이

4.6 ★★★★★ 1,353건 | 리뷰 4,898

제주특별자치도 서귀포시 예래로 255-18 1층 (지번) 하예동 350

제주도 서귀포 맛집

제주 성산 해마루

해물, 생선 요리

영업중 ★ 4.28 | 리뷰 999+

중문수원음식점

영업중 ★ 4.54

성산 타쿠마시

생선회

영업전 ★ 4.43 | 리뷰 999+

원앤온리

카페

영업중 ★ 4.34

제주도 서귀포 맛집

천짓골식당

4.3 ★★★★★ (765) · WWW

한식 · 천지동 294-10

영업 종료 · 오후 5:00에 영업 시작

뽕살집

4.6 ★★★★★ (1,028) · WW

한국식 BBQ · 중앙동 중정로91번

영업 종료 · 오후 3:00에 영업 시작

서귀포동경우동본점

4.1 ★★★★★ (125) · WW

우동 · 중앙로89번길 4

영업중 · 오후 8:00에 영업 종료

제주도 서귀포 맛집 베스트 95곳

1. 미영이네식당 4.7

제주특별자치도 서귀포시 대정읍 대정로 42

소위 제주도 갈매기다 병원에서... 일반에 방문한지도 모를것은 미영이네, 근대 찜갈매기다 맛있디... ㅠ 친구랑 내가 고향에 와서 너무 좋아해서 자주... 더보기

2. 남양수산 4.6

제주특별자치도 서귀포시 성산읍 고성동서로56번길 11

ysom 지못난 고향어회와 찜닭이 기대되는 다른 곳보다는 있는 편이긴 한데 포장합니다! 고등어와 내우 산란하고 대패한 허문골에 꼭꼭오니 정말 맛있었어요... 더보기

3. 중문수두리보말갈국수 4.4

제주특별자치도 서귀포시 장례로 192

한뼘 왕복 보고 제주도에서 인생 갈국수집 찾았네요... 왜이렇게 시간이 전혀 아깝지 않은 맛과 친절함이있어요 깔끔한 국물과 깨끗하고 고소한 것... 더보기

프로젝트 목적

그런데 카카오맵 평점은 신뢰할만 하다는 얘기가 있습니다
사실일까요?

역시 카카오맵은 실패를 안함...

다른 집 먼저 알아봤었는데

카카오맵 리뷰점수가 훨씬 더 높은데로 갔는데 대박이
야...

저는 전혀 모를 때는 카카오맵 평점 참고 합니다.

네이버지도 같은 경우에는 소위 '작업' 쳐놓은 리뷰가 너무 많은 반면에,

상대적으로 카카오맵은 훨씬 적어요.

사용자가 적어서 그럴 수도 있지만 카카오맵은 리뷰가 충분히 있으면서 평점을 보면 어느정도 맞다

네이버지도에서 평점은 4점대 후반 안되는 집이 거의 없는 반면에 상대적으로 카카오맵은 편차가 크

물론 카카오맵 평점 좋은 곳도 가보면 실망하는 경우가 있긴해서 절대적이진 않았고요.

망고플레이트는 맛이 좋으면 대체로 평가가 후하고

카카오맵 리뷰는 맛이 좋아도 서비스, 위생, 분위기,
하다못해 셰프의 인성이나 도덕성 등에 논란이 조금
만 있어도 점수를 확 깎음

구글,네이버 리뷰는 거르면 됨. 예시로
4.9다. 무슨뜻인지 알지?

지금은 없어지긴 했지만 예전 네이버 식당 리뷰는
점수가 네이버 웹툰 급으로 높아, 4점 못 넘기면 이
상한 거 아니냐는 느낌이 들었는데
반대로 카카오맵 식당 점수(아직 살아있음)는 대체
어디가 4점을 넘길 수 있나 싶을 정도로 점수가 째.
똑같은 맵 어플인데 이렇게 점수 기준이 극단적일
일인가

프로젝트 목적

카카오맵 평점이 신뢰할 만한지 어떻게 평가할 수 있을까요?

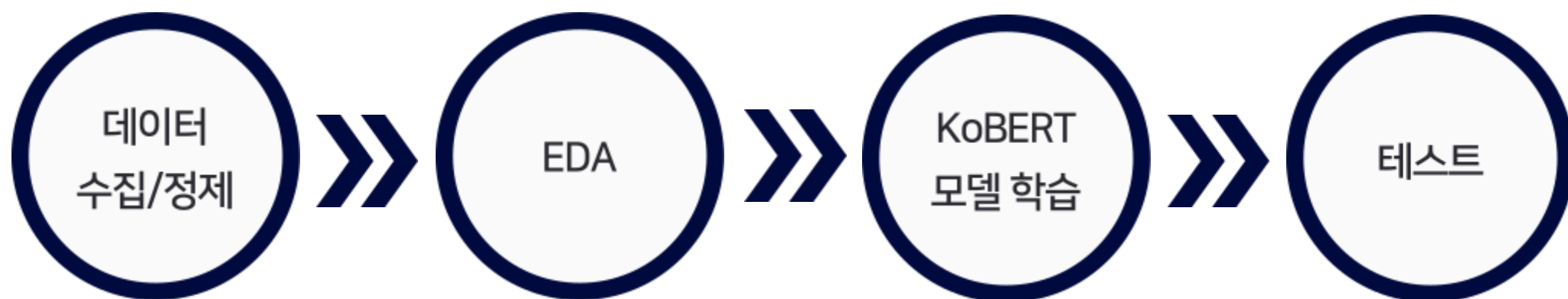
-> 좋거나 나쁜 극단적인 경우가 아닌, 무난한 후기의 평점을 통해 확인



카카오맵 이용 후기의 실제 중립 감정은 몇 점일까?

가설: 긍정 편향 되어있으므로 4점 이상일 것이다

개발 프로세스



개발 프로세스



데이터 수집

- 제주도 음식점 리스트.csv의 식당 이름을 카카오맵에 검색한 후 후기가 20건 이상이면 해당 후기들을 모두 수집하는 커스텀 크롤러 제작

출처: 공공데이터포털 제주관광공사_제주관광정보시스템(VISITJEJU)_음식점콘텐츠

<https://www.data.go.kr/data/15041984/fileData.do>

콘텐츠명	노키즈존	예약가능여부	동보유여부	유아서비스가능여부	평일오	평일클로징시	주말오	주말클로징시	휴식시	휴식종료시간	대표메뉴기타
0 탐라독배기	NaN	NaN	NaN	NaN							
1 토담골	NaN	NaN	NaN	NaN							
2 우유부단	NaN	NaN	NaN	NaN							
3 명당양과	NaN	NaN	NaN	NaN							
4 땅고분식뷔페	NaN	NaN	NaN	NaN							
...							
1931 카페아오오	y	y	NaN	NaN							
1932 샤오츠	y	n	NaN	NaN							
1933 읍씨빵집	y	y	NaN	NaN							
1934 바렘목장&카페	y	n	NaN	NaN							
	y	y	NaN	NaN							

KakaoMap Search Results for '제주 우유부단':

- 장소 1: 우유부단 (2.9 stars, 83 reviews, 187 likes)
- 주소: 제주특별자치도 제주시 한림읍 금악동길 38-1 (지번) 한림읍 금악리 142-2
- 영업중: 매일 10:00 ~ 18:00

Reviews:

- 남남남남: 후기 35, 별점평균 4.4 (2022.08.29), 5 stars, 맛, 분위기
- L: 후기 370, 별점평균 4.4 (2022.08.22), 4.5 stars, 좋아요
- 나: 후기 20, 별점평균 4.5 (2022.05.11), 5 stars, 맛있었어요, 우리 아기가 양을 좋아해서 아이스크림

Tools/Formats: CSV, KakaoMap, Selenium

개발 프로세스



데이터 저장



	ABC 상호명	ABC 리뷰어	ABC 리뷰날짜	123 별점	ABC 코멘트
1	우유부단	L	2022.08.22	80	
2	우유부단	나	2022.05.11	100	맛있었어요. 우리 아기가 영
3	우유부단	ㅎ	2022.04.16	60	우유아이스크림과 밀크쉐이
4	우유부단	beaucoupderiz	2022.01.12	40	
5	우유부단	석원	2022.01.11	100	
6	우유부단	BEARDOOR	2022.01.05	60	이시돌목장 우유를 좋아해
7	우유부단	타	2021.12.05	20	맛은 잘모르겠고 가격이 너
8	우유부단	DH. KIM	2021.11.21	40	밀크쉐이크. 밀크티 먹기
9	우유부단	지연	2021.11.05	40	
10	우유부단	유 동호	2021.10.31	60	우유의 고소함이 진해서 맛
11	우유부단	베토벤	2021.09.21	100	맛있어요!!!! 1달백한 밀크쉐
12	우유부단	ㅂㅂ	2021.09.03	20	맛은 그런대로 괜찮은데 가
13	우유부단	내돈내먹솔직후기	2021.09.01	40	아이스크림 별로. 상하목장
14	우유부단	우당탕	2021.08.26	100	우유를 컨셉으로 한 카페. 소
15	우유부단	우가차차	2021.08.19	20	분위기는 모르겠고 일단 아
16	우유부단	양술랭	2021.07.27	80	1. 맛1- 우유맛이 진한 우유
17	우유부단	배고폰데	2021.07.24	40	맛 3점 서비스 2점 파리 짱!
18	우유부단	ㅇㅇ	2021.07.18	80	완전 맛있어요! 근데 화장실
19	우유부단	호호호	2021.07.16	80	목장구경후 시원한 밀크쉐이
20	우유부단	~	2021.07.14	20	맛은 있는데 직원들이 하나

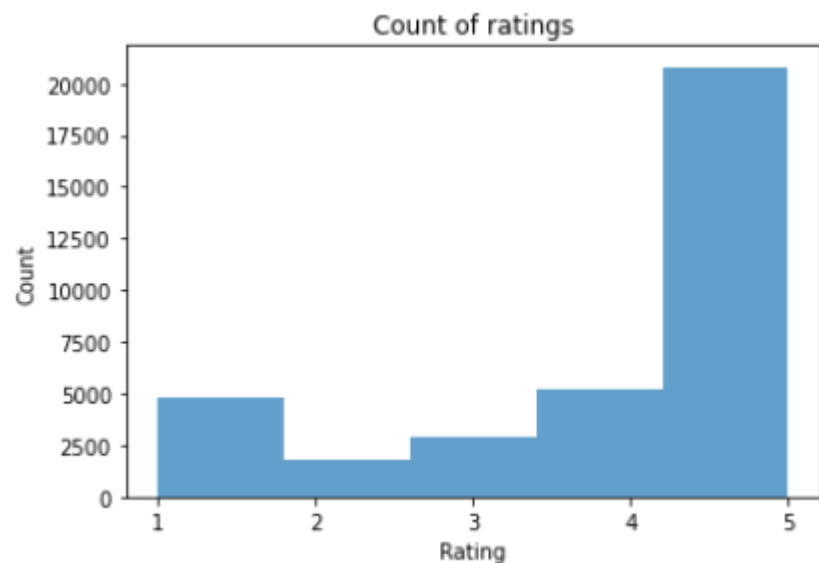
```

graph TD
    A[데이터 수집/정제] --> B[EDA]
    B --> C[KoBERT 모델 학습]
    C --> D[테스트]
  
```

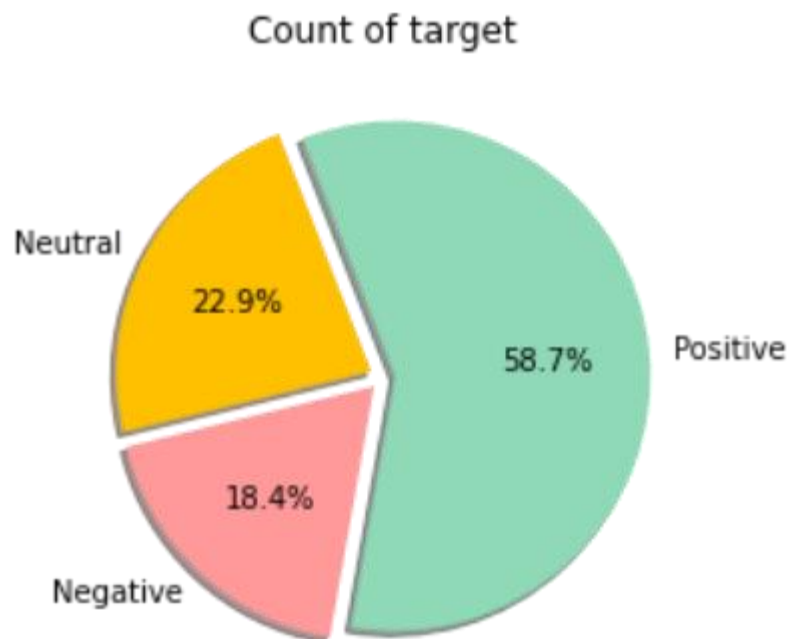
- 리뷰 코멘트=NaN 삭제
- 별점 1~5점으로 변경
- 불필요한 컬럼 삭제
- 컬럼명 및 순서 정리

```
df.shape >>> (35441, 3)
```

개발 프로세스



별점이 5점인 후기가 월등히 많음
→ label 추가 시 고려



Target labeling

별점1~2 : 부정(0) = Negative

별점3~4 : nan = Neutral

별점5 : 긍정(1) = Positive

개발 프로세스

형태소 분석을 통해 명사 추출 후 워드클라우드 생성

- 부정적인 단어: 불친절, 별로, 최악 등



긍정적인 단어가 더 많이 나타남

개발 프로세스



- label이 0 또는 1인 데이터(별점: 1~2, 5)로 학습 진행(train/val)

- KoBERT 모델 사용

구글 BERT의 한국어 성능 한계
한국어 버전 자연어 처리 모델

<https://github.com/SKTBrain/KoBERT>



- 토큰화, 패딩 처리

```
(array([ 2, 4155, 6925, 7295, 6896, 1801, 6133, 5451, 7120, 5152, 5330,
        517, 6750, 6750, 7788, 1967, 7141, 6885, 6855, 6916, 517, 5330,
        5400, 5859, 4483, 5415, 7788, 4204, 5702, 6916, 3, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
        1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=int32),
array(31, dtype=int32),
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        dtype=int32),
```

1)

개발 프로세스



- 모델 학습

train 정확도: 0.9974 / validation 정확도: 0.9651

epoch 3 batch id 1 loss 0.10994336754083633 train acc 0.96875

epoch 3 batch id 201 loss 0.04294877126812935 train acc 0.9777674129353234

epoch 3 train acc 0.9801401869158879

100%  107/107 [00:26<00:00, 4.43it/s]

epoch 3 validation acc 0.9467971183800623

100%  321/321 [03:52<00:00, 1.76it/s]

epoch 4 batch id 1 loss 0.1256030648946762 train acc 0.9375

epoch 4 batch id 201 loss 0.003911632113158703 train acc 0.9897388059701493

epoch 4 train acc 0.9916764018691588

100%  107/107 [00:26<00:00, 4.44it/s]

epoch 4 validation acc 0.9650992990654206

100%  321/321 [03:52<00:00, 1.75it/s]

epoch 5 batch id 1 loss 0.018370486795902252 train acc 0.984375

epoch 5 batch id 201 loss 0.0016452973941341043 train acc 0.9968905472636815

epoch 5 train acc 0.9974201713395638

100%  107/107 [00:26<00:00, 4.42it/s]

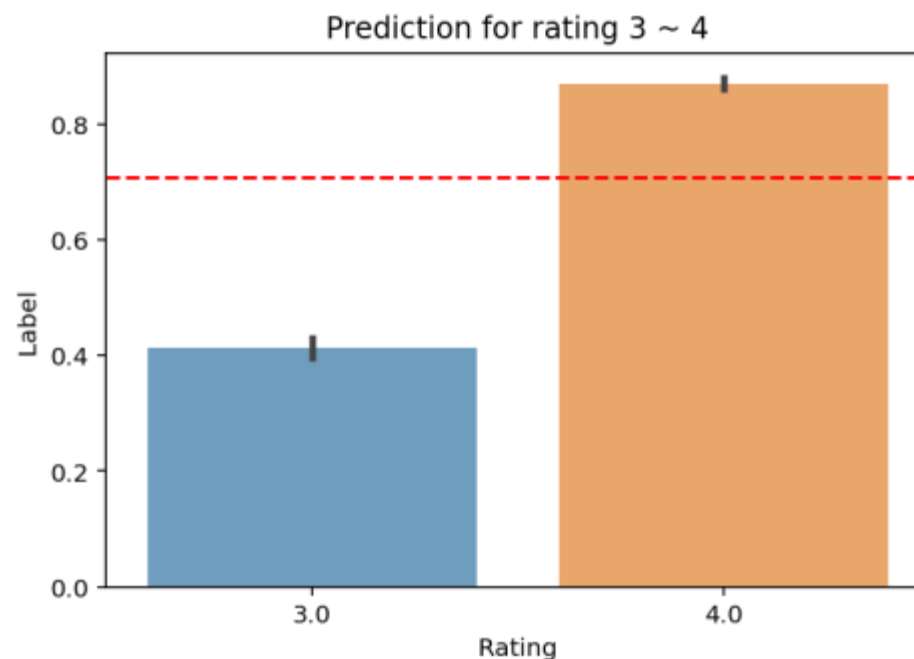
epoch 5 validation acc 0.9650992990654206

개발 프로세스



- 별점이 3~4인 중립 감정 데이터의 label 예측(test)

rating	reviews	label
3.0	우유아이스크림과 밀크셰이크, 우도땅콩밀크티는 먹을만한데요 아메리카노,라떼, 카카오밀...	0.0
3.0	이시돌목장 우유를 좋아해서 밀크티도 기대했는데 조금 아쉬웠네요...그래도 목장 구경...	0.0
3.0	우유의 고소함이 진해서 맛있지만 너무 비싸요. 주변 볼거리도 딱히 없어요. 풀바셋이...	0.0
4.0	1. 맛₩n- 우유맛이 진한 우유와 아이스크림₩n2. 가격₩n- 저렴하지는 않음₩n...	0.0
4.0	완전 맛있어요! 근데 화장실이 협소하고 더러워요ㅜ	1.0



test셋의 평균 label: 0.7079

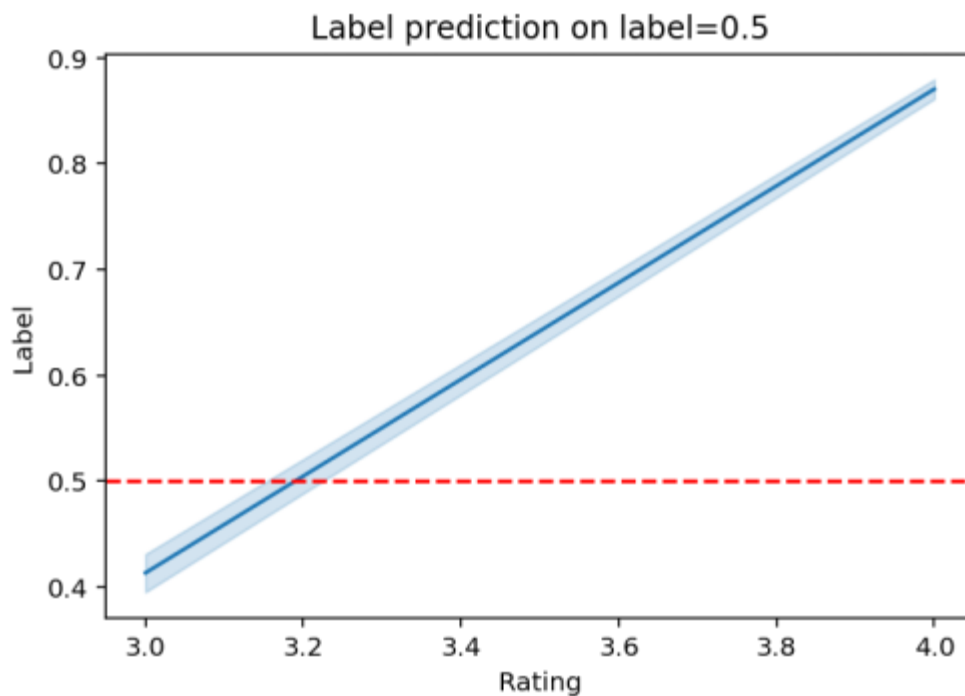
별점 3점 평균 label: 0.4133

별점 4점 평균 label: 0.8704

개발 프로세스



- 중립(label=0.5) 감정의 별점 → 3.2 정도로 예측



결론 및 개선점

카카오맵 이용 후기의 실제 중립 감정은 몇 점일까?

~~가설: 긍정 편향 되어있으므로 4점 이상일 것이다~~

결론

- 카카오맵 후기의 중립 감정은 3.2 정도로 예측되므로,
약간의 긍정 편향은 있을 수 있어도 3에서 크게 벗어나지 않음
- 따라서 카카오맵의 후기는 신뢰할만 하다고 여겨짐

아쉬운 점 및 추후 개선점

- 학습 시 긍정(58.7%)과 부정(18.4%) 데이터의 비율이 불공평했으므로
예측에 오류가 있을 수 있음. 다른 데이터 라벨링 방식 시도
- KoBERT를 이용하여 다중분류 시도
- 기타 feature를 이용하여 비즈니스 인사이트 도출

THANK YOU