

리슐랭, 진(짜)리(뷰)를 찾아서

캡스톤 프로젝트 최종 보고서

팀 이름: 현즈

20153262 이현지

20160036 조중현

20176867 전현욱

목차

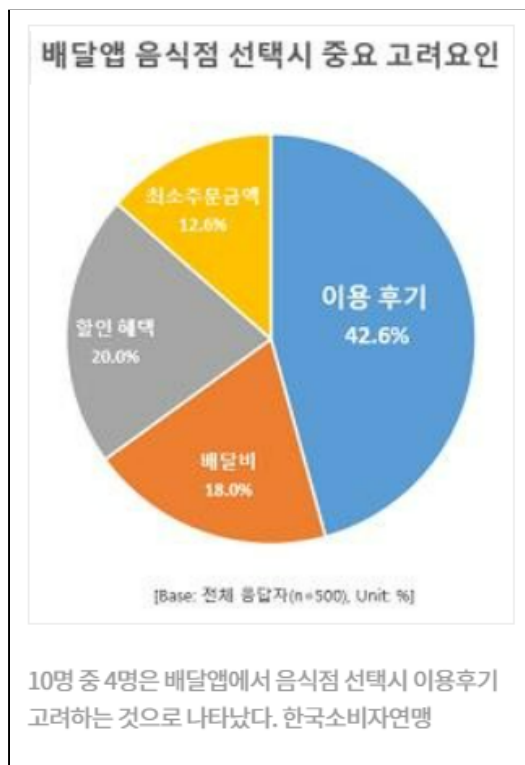
- I. Introduction
- II. Motivation
- III. Related Works
- IV. Design and implementation
- V. Evaluation
- VI. Conclusion
- VII. 팀원역할 및 일정

I. Introduction

요즘처럼 바쁜 현대 사회에서 사람들이 음식점을 선택할 때 가장 주요하게 고려하는 부분은 무엇일까요? 아래 한국 소비자 연맹 기사의 그래프에서 볼 수 있듯이 바로 경험한 소비자들 개개인이 올리는 리뷰와 별점입니다. 사람들은 배달 앱을 통해 음식을 시켜 먹을 때뿐 아니라 바깥에 나가 외식할 때에도 여러 배달어플의 소비자 리뷰, 네이버 플레이스 방문자 및 블로그 리뷰 등을 참고하여 음식점을 선택하고 결제를 고민합니다. 또한 그런 사람들의 리뷰들이 음식점의 매출과 직결되기 때문에 별점과 리뷰는 가게 점주들 역시 많은 신경을 쏟고 있는 부분입니다.

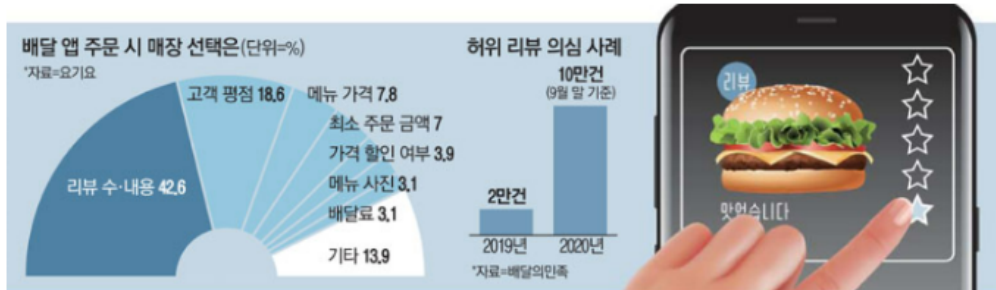
이렇게 경쟁이 치열한 배달, 외식 업계에서 과연 모든 리뷰들을 신뢰할 수 있을까요?

‘리슐랭, 진(짜)리(뷰)를 찾아서’ 프로젝트는 긍/부정 평가 및 감성 분석 모델학습을 통해 한 번 더 필터링 하여 진실한 리뷰를 찾기 위함을 목표로 하고 있습니다.



II. Motivation

월 1만건 가짜 리뷰...자영업자 "싸우기 지쳐, 나도 조작해볼까"



"믿을 건 후기뿐"... '가짜리뷰'와 전쟁 중인 유통가

배달앱 블랙 컨슈머 향한 경고 “악성 리뷰, 그거 범죄야!”

[SNS에세이]배민 별점·네이버 리뷰, 소비자 권리와 갑질 사이

실제로 이용 후기 관련 불편 사항 조사에서 48.9%가 후기를 보고 구매했으나 불만족했다고 응답했다. 이용 후기를 신뢰한다는 응답은 48.2%였다.

또한 배달의민족과 요기요에서 한식, 분식, 치킨, 중식 등 4개 카테고리당 10개 업체씩을 놓고 리뷰 이벤트 진행 여부를 조사한 결과 배달의민족에서는 60~90%, 요기요에선 20~60%에 해당하는 업체가 리뷰 이벤트를 하는 것으로 나타났다.

이렇게 많은 기사 내용에서 볼 수 있듯이 요즘 리뷰들은 음식점들의 리뷰이벤트 또는 협찬 등으로 대부분 상향평준화 되어있으며 또는 몇몇의 악질적인 블랙컨슈머들의 악성리뷰로 저평가 되어 있기도 합니다. 기사에 따르면 리뷰를 믿고 결제하였으나 불만족스러운 경험을 한 경우도 48.9퍼센트로 적지 않음을 확인 할 수 있었습니다.

저희는 이렇게 온전히 믿을 수 없는 리뷰들 속에서 진실된 리뷰만을 얻고자 ‘리술랭, 진(짜)리(뷰)를 찾아서’ 프로젝트를 시작하게 되었습니다.

III. Related Works

3.1. 기존연구/프로젝트와 유사점

소비자의 선호도를 파악하기 위한 연구 및 프로젝트는 이전에도 많이 존재합니다. 예를 들어서 식당 리뷰, 영화 리뷰등을 감정분석 하는 논문이나 자료들은 이미 많이 존재하며 연구가 활발히 진행되고 있습니다. 저희의 프로젝트 또한 별점 및 리뷰의 텍스트를 감정 분석하며, 감정 분석 결과를 소비자에게 보여주는 것을 목표로 하고 있습니다. 기존 연구들은 데이터 수집 → 데이터 전처리 → 감성 분석 → 데이터 집합 순으로 프로젝트를 진행하였고, 저희의 프로젝트 또한 그 설계 순서를 따라갑니다.

3.2. 기존 연구/프로젝트와 차이점

저희 프로젝트는 기존 프로젝트와 다르게 감정 분석에서 멈추지 않고, **새로운 리뷰 점수로** 환산하는 것에 있습니다. 기존의 방식처럼 모든 리뷰를 포함해서 분석한다면 신뢰도가 떨어지는 리뷰까지 분석에 포함되어 객관성이 떨어질 것이라고 생각했습니다. 이러한 문제점을 해결하고자 저희는 다음과 같이 변화를 주었습니다.

1. 감성사전에 신조어 및 축약어를 추가하여 요즘 시대에 맞는 감성 어휘들을 반영할 계획이며 쉽게 쓸 수 있는 ‘좋아요~’와 같은 짤막한 리뷰, 일부러 별점을 떨어뜨리기 위한 악성 리뷰, 그리고 이와 비슷한 모든 무의미한 리뷰들을 모두 배제시켰습니다. 허위 리뷰 제거에 관한 근거는 “광고 심사 지침의 권구 문구”를 바탕으로 모델을 만들어서 제거하였습니다. 노이즈를 제거함으로써 숨겨진 진실된 리뷰를 소비자들에게 제시할 수 있다고 생각합니다.
2. 기존의 감성분석에서 좀 더 발전하여 단순한 “단어” 단위의 감성분석이 아닌 “구 단위” 감성 분석을 진행하였습니다. 이에 대한 근거는 “말뭉치언어학적 관점에서 본 감성 표현 추출의 쟁점”이라는 논문을 참조하였습니다. 논문에 따르면 감성 표현의 범주는 50% 이상이 구 단위로 나오며, 이를 위해서 특정 품사 중심의 어휘 수준을 넘어 구 단위 분석이 필요하다고 합니다.

3. 새로운 리뷰점수를 도입하기 위해 구체적인 특정 키워드들에 더욱 가중치를 둘것입니다.
저희 프로젝트는 소비자들의 음식점 선택에 올바른 근거를 제시하기 위해 리뷰의 주요 키워드와 점수를 더욱 확실히 인식시키도록 하는것이 목표이기 때문입니다.

3.3. 관련 논문 및 자료

📄 음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측.pdf

📄 평점과리뷰텍스트감성분석을결합한추천시스템향상방안연구.pdf

[\[Keras\] KoNLPy를 이용한 한국어 영화 리뷰 감정 분석](#)

[리뷰 기반 맛집 추천 시스템 \[공모전 결과물\]-I am yumida](#)

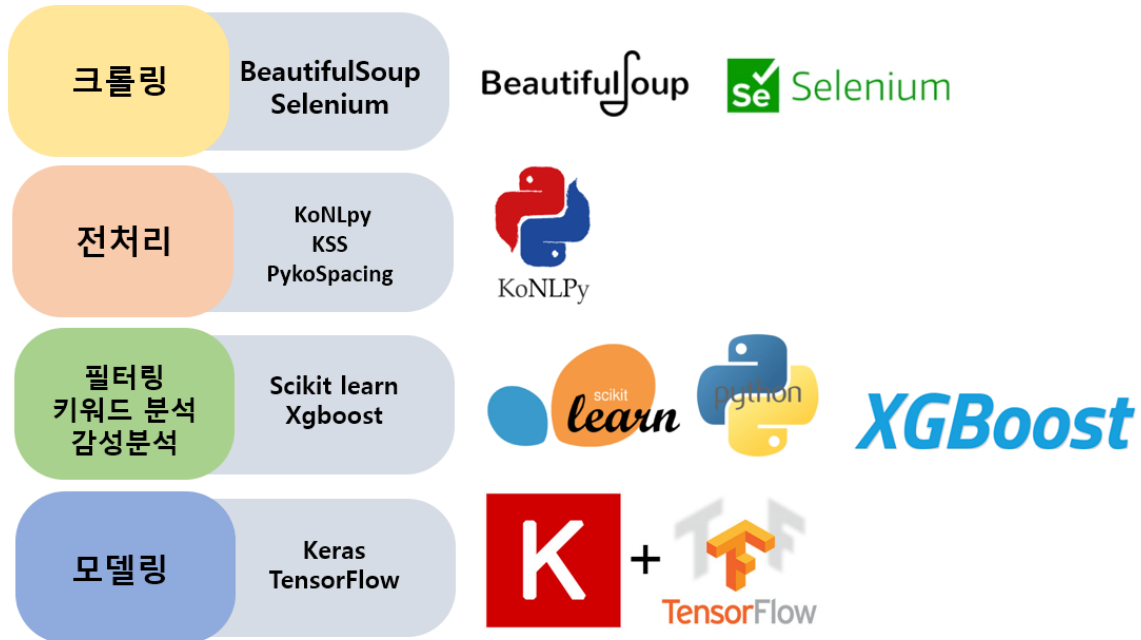
[딥러닝 - 영화리뷰 감성분석 프로그램 만들기 \(tistory.com\)](#)

3.4. 리뷰 데이터 구 단위 감성분석에 대한 근거

📄 말뭉치언어학적 관점에서 본 감성표현 추출의 쟁점 - 사용자 리뷰 말뭉치를 중심으로-.pdf

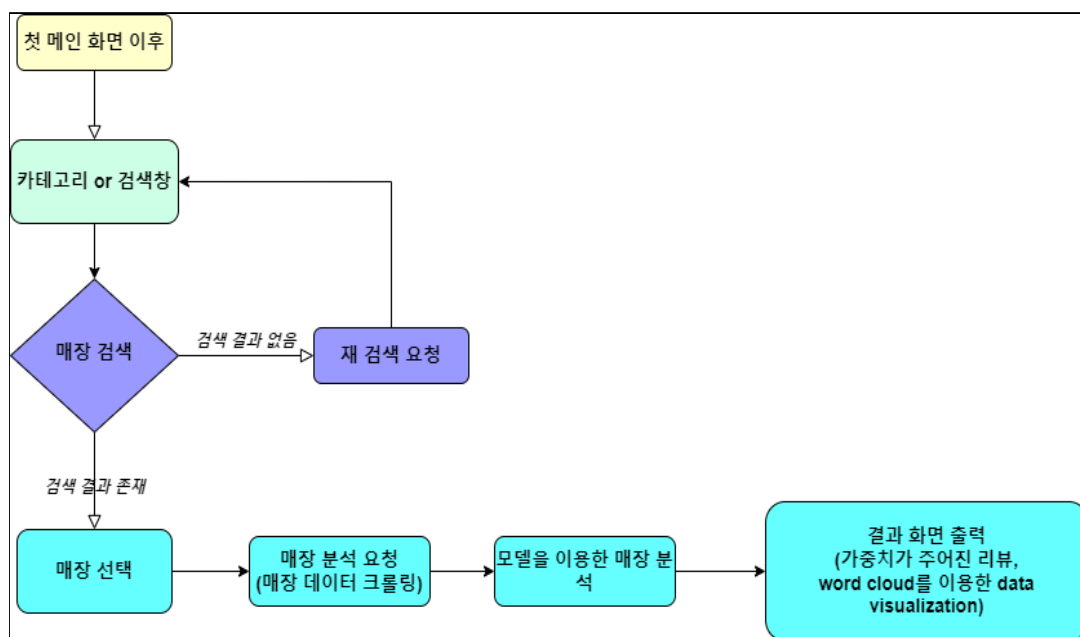
IV. Design and Implementation

4.1. 기술 현황

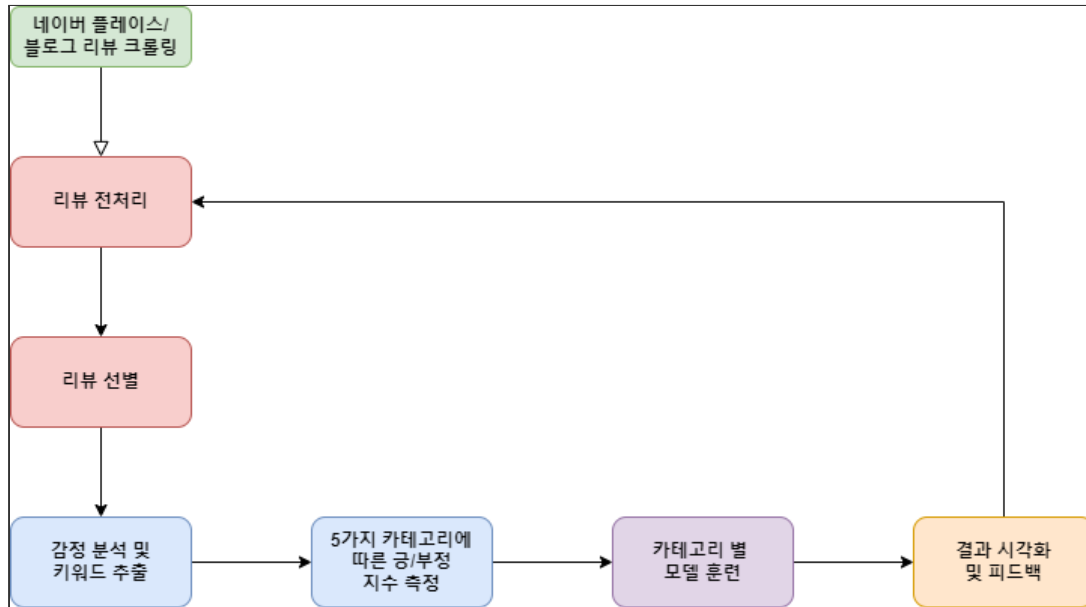


4.2. Flow Chart

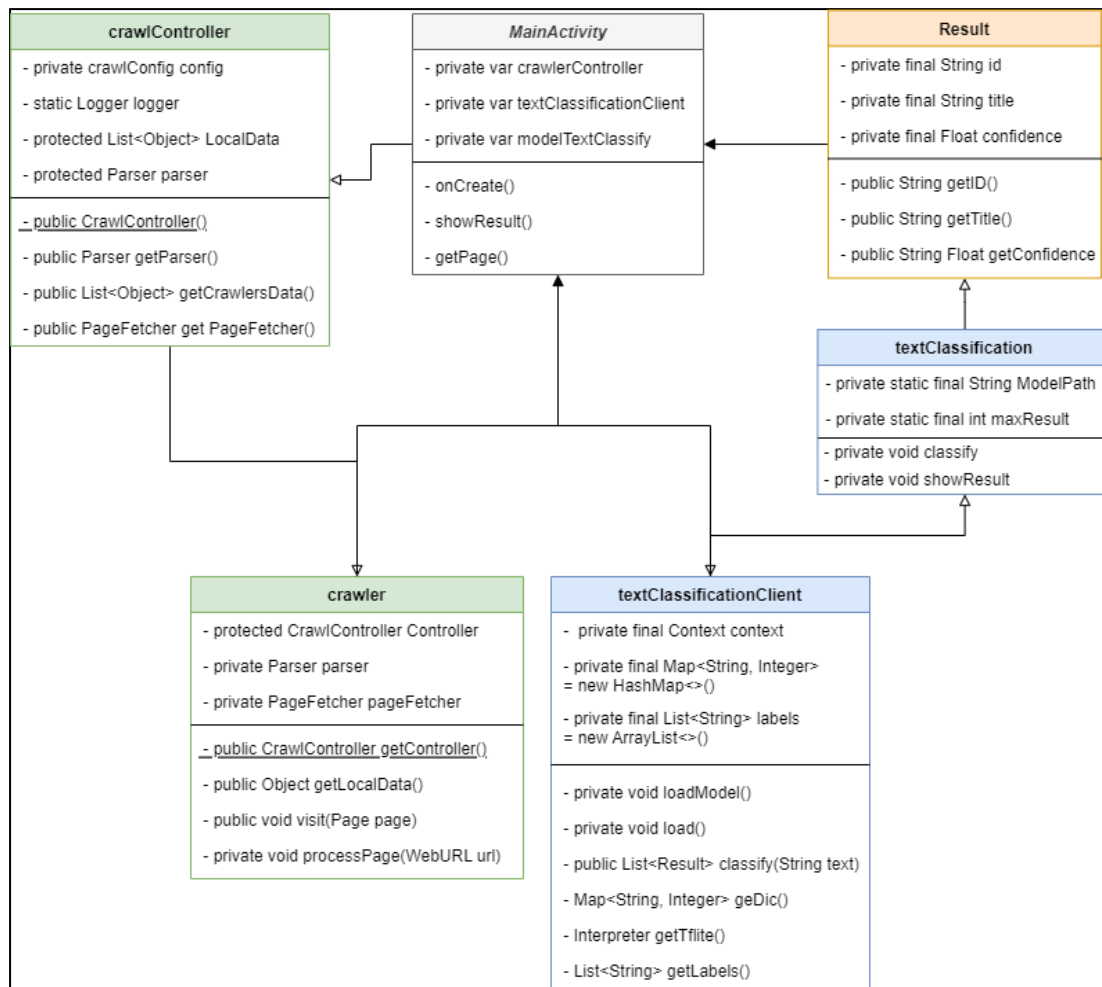
4.2.1. 초기에 기획한 Flow Chart



4.2.2 이후 변화한 Flow Chart



4.3. Class diagram



4.4. 상세 설계

4.4.1. 네이버 플레이스 방문자/블로그 리뷰 크롤링

네이버 플레이스 리뷰는 네이버에서 자동화된 크롤링을 막기 위해 주기적으로 xml파일의 Class명들을 변경하고 있습니다. 이에 프로젝트 당시에는 작동했지만, 추후 수정이 필요할 수 있습니다.

url list

	region	name	url
0	배곧	배곧 우정횃집	https://pcmap.place.naver.com/restaurant/17953...
1	배곧	배곧 소셜원 배곧본점	https://pcmap.place.naver.com/restaurant/13924...
2	배곧	배곧 등대횃집	https://pcmap.place.naver.com/restaurant/17954...
3	배곧	배곧 배곧아띠	none
4	배곧	배곧 경양카츠 배곧점	https://pcmap.place.naver.com/restaurant/10278...

네이버 플레이스 방문자 리뷰

	num	name	blog or review	visitor review content
0	1	배곧 우정횃집	방문자 리뷰	가족 모임으로 다녀왔습니다. 좋은 리뷰 평가가 많아서 일단 믿고 왔는데, 부부 사장...
1	2	배곧 우정횃집	방문자 리뷰	주말에 어디갈까 하다가 티맵에 오이도맛집 치니 딱! 하고 우정횃집이 나오더라고요. ...
2	3	배곧 우정횃집	방문자 리뷰	결혼기념일 남편이랑 다녀왔습니다.\n5월 제철 - 도다리 회. 밀반찬에 시키다시...
3	4	배곧 우정횃집	방문자 리뷰	거레저 사장님 추천으로 왔다가 완전 맛있게 잘 먹었어요. 음식 진짜 깔끔하고 맛있어...
4	5	배곧 우정횃집	방문자 리뷰	오이도에서 제일 멋진 횃집입니다.\n맛있어요.

네이버 플레이스 블로그 리뷰

	num	name	blog or review	blog name	like_count_list	command_count_list	date_list	img_count	neighbor_count	last_content_date	blog review
0	1	배곧 또 잇카페 배곧점	블로그 리뷰	TTOEATCAFE 배곧카 페 뉴맛집!	5	6	2021. 8. 21. 13:36	21	1,071	2021-11-16	배곧 한울공원이 산 책하기에 너무 좋은 곳이라 자주 가요오 후 시간에 산책하면 노을지는...
1	2	배곧 또 잇카페 배곧점	블로그 리뷰	배곧 한울공원 뷰 (또잇카페)	7	4	2021. 11. 14. 22:39	23	191	2021-11-14	원님 여러분 주말에 배곧 한울공원에 갔 다가 뷰맛집 게이쿠 맛집이라는 또잇카 페를 다녀... 주말에 다녀오기만

4.4.2. 리뷰 전처리

- 영어 및 특수문자 제거하였습니다.
- 띄어쓰기가 안 맞는 문장을 띄어쓰기 처리 하였습니다.
- 리뷰를 문장 단위로 구분하였습니다.
- 정규화 및 불용어 제거 후 형태소 분석을 수행하였습니다.

리뷰 전처리 후

	num	name	blog or review	visitor review
0	1	배곧 우정횃집	방문자 리뷰	가족 모임으로 다녀 왔습니다. 좋은 리뷰 평가가 많아서 일단 믿고 왔는데 부부 사장...
1	2	배곧 우정횃집	방문자 리뷰	주말에 어디갈까 하다가 티맵에 오이도 맛집 치니 딱 하고 우정횃집이 나오더라고요 역...
2	3	배곧 우정횃집	방문자 리뷰	결혼기념일 남편이랑 다녀왔습니다. 5월 제철 도다리 회 밑반찬에 시키다 시 매운탕가...
3	4	배곧 우정횃집	방문자 리뷰	거래처 사장님 추천으로 왔다가 완전 맛있게 잘 먹었어요. 음식 진짜 깔끔하고 맛있어...
4	5	배곧 우정횃집	방문자 리뷰	오이도에서 제일 멋진 횃집입니다. 맛있어요.
...
3226	6	배곧 그레몽디저트샵	방문자 리뷰	출산한 아내가 가장 먹고 싶다고 말 한 그레몽 에 그 타르트 맛이 정말 예술입니다....
3227	7	배곧 그레몽디저트샵	방문자 리뷰	진짜 거짓말 안 하고 동네에 친구들 놀러오면 꼭 데려가는 곳입니다. 왜냐구요. 실패...
3228	8	배곧 그레몽디저트샵	방문자 리뷰	인스타 캠성 카페예요. 디저트가 너무 맛있고 인테리어를 엄청 신경 쓰신 게 보여요...
3229	9	배곧 그레몽디저트샵	방문자 리뷰	빵 디저트 엄청 좋아해서 배곧 안 가본 곳이 없어요. 요기 휘남시에 걸은 진짜 바삭...
3230	10	배곧 그레몽디저트샵	방문자 리뷰	이런 곳 너무 좋아합니다. 디저트도 맛있네요 오자마자 바로 저장.

3231 rows x 4 columns

4.4.3. 리뷰 선별

- 허위 리뷰 키워드 사전(광고 심사 지침 권고문구)을 이용해서 spam단어가 포함된 리뷰 라벨링을 수행하였습니다.
- 허위 리뷰 제거 시 decision tree, random forest, logistic regression, xgboost 모델을 이용하여 성능 평가를 수행하였습니다.
- 여러 모델들의 성능을 비교해서 xgboost 모델을 이용한 스팸 및 악성 리뷰 제거하였습니다. 총 3231개의 리뷰 중 66개의 스팸 리뷰를 제거하였습니다.

광고 심사 지침 권고

<권고문구>

1. 추천·보증 등의 대가로 현금, 물품 등을 지급받은 경우 ‘경제적 대가’ 또는 그에 상응하는 구체적 표현(현금, 상품권, 수수료, 무료제품 등)을 사용하여 다음과 같이 표기하여야 한다.

<권고문구1> 저는 위 ○○ 상품을 추천(보증, 소개, 홍보 등)하면서 ◇◇사로부터 경제적 대가(현금, 상품권, 수수료, 포인트, 무료제품 등)를 받았음

<권고문구2> ‘유료광고’, ‘대가성 광고’ 등 경제적 이해관계가 명확히 드러나는 표현

words

독독으로
문의가
부탁이
제휴
협찬을
작성되었습니다
업체로부터
제공받아
원고료를

Spam word list		
	review	Spam or Nonspam
	식당의 지원 을 받아....	1
	고기가 맛있어...	0
	문의 를...	1
	이번에는.... 협찬 ...	1

4.4.4. 감성 분석 및 키워드 추출

- 5가지 범주에 따른 감성 사전(서비스, 맛, 가격, 분위기, 재방문의사) 하였습니다.
- 형태소 분석을 통해 토큰화 -> 구 단위 감성 분석(키워드 및 주변 조사 및 어절 포함)을 진행 하였습니다.
- 키워드와 키워드가 포함된 구에서 해당 특성의 감성사전에 포함된 단어와 비교해서 긍/부정 빈도를 평가하였습니다.

Dictionary of Sentiments



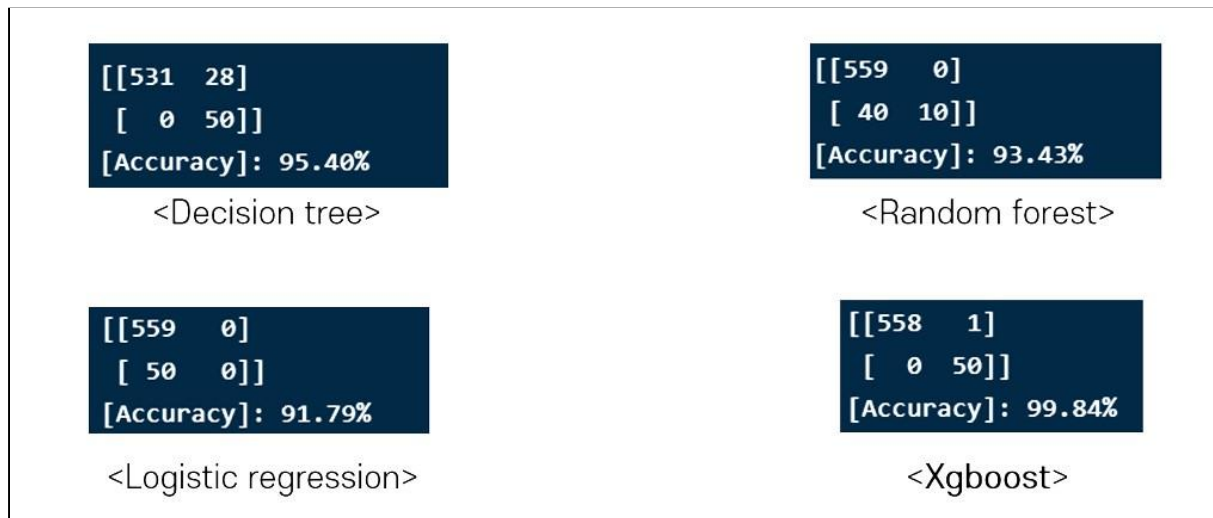
Good Features	Bad Features
<pre>service_good_feature = {'서비스': ['좋', '친절', '편찮', '최고', '빠름', '정', '훌륭', '추천', '감사', '구수', '최상', '대박', '흔흔', '특별', '개이득', '최고', '만족', '세련', '최고', '강동'], taste_good_feature = {'간': ['맛', '적절', '딱', '환상', '담백'], '음식': ['깔끔'], cost_good_feature = {'가격': ['편찮', '적당', '저렴', '적당', '싸다', '좋다', '합리적', '훌륭', '최고', '만족', '마음', '드뎌', '알맞다', '무난', '결론', '최상', '최상', '곳', '양지', '날'], atmosphere_good_feature = {'문위기': ['좋', '편찮', '조용', '깔끔', '적당', '강파', '공감', '아름', '원벽', '이끼지', '고급', '최고', '세련', '만족', '아늑', '훌륭', '예쁜', '이쁜', '편', '심쿵'], visit_good_feature = {'의사': ['있다', '충만', '백프로', '백프로', '달', '만만', '마구', '그럭', '만점', '넘칠'], '다시': ['다시', '오다', '방문', '찾다', '가보다', '한번', '갈다', '찾아가다', '가가지', '갈거다',</pre>	<pre>service_bad_feature = {'서비스': ['아쉽', '최악', '나쁜', '느려', '뻑뻑', '바주', '별로', '고양', '낙제', '쏘다쏘다', '영양', '실망', '불친절', '문제', '영플레인', '거지', '그악', '그다지', '구경', '불편', '영성', '할', '가판'], taste_bad_feature = {'간': ['세', '썰', '강하다', '별로'], '음식': ['별로', '쏘다쏘다', '최악'], cost_bad_feature = {'가격': ['비싸', '있다', '나쁜', '사악', '비호를', '높다', '무감', '아쉽', '쏘다쏘다', '별로', '그악', '그다지', '썰', 'acc', '높', '거품'], atmosphere_bad_feature = {'문위기': ['나쁘다', '비쁘다', '어수선하다', '이상하다', '혼스럽다', '별로', '무당스럽다', '시끄럼', '독잡'], visit_bad_feature = {'의사': ['글쎄'], '같이': ['다시', '말이', '여기까지', '올서서', '찾아', '시키다', '가다', '찾다', '여기', '기다리다',</pre>

4.4.5. 5가지 범주에 따른 식당의 긍/부정 지수 측정

- 각 식당별 긍/부정 지수에 대해서 계산하여 출력하였습니다.

V. Evaluation

5.1. 리뷰 필터링 과정에서 모델의 성능 비교



거짓리뷰, 악성리뷰를 제거하는 리뷰 필터링 과정에서 모델의 성능 비교해보면 다음과 같습니다. Decision tree의 정확도는 95.4 퍼센트, random forest의 정확도는 93.4 퍼센트, logistic regression의 정확도는 91.79 퍼센트, xgboost의 정확도는 99.84 퍼센트로 가장 성능이 좋은 **xgboost**로 프로젝트를 진행하였습니다.

5.2. word cloud 키워드 분석



감성 분석 및 키워드 추출 단계에서 추출한 키워드를 word cloud로 나타내면 다음과 같습니다.

5.3. 5가지 범주에 따른 음식점의 긍/부정 리뷰 분석

배곧 또잇카페 배곧점

Total review count: 44
Good service: $3/4 = 75.0$
Good atmosphere: $23/23 = 100.0$
Good cost: $1/1 = 100.0$
Good taste: $31/34 = 91.0$
Good revisit: $1/2 = 50.0$

배곧 양퐁애들갈비

Total review count: 9
Good service: $0/0 = 0$
Good atmosphere: $0/0 = 0$
Good cost: $0/1 = 0.0$
Good taste: $7/7 = 100.0$
Good revisit: $0/0 = 0$

배곧 카츠오모이

Total review count: 84
Good service: $8/9 = 89.0$
Good atmosphere: $1/1 = 100.0$
Good cost: $2/3 = 67.0$
Good taste: $62/69 = 90.0$
Good revisit: $1/2 = 50.0$

각 음식점의 총 리뷰 개수와 서비스, 맛, 가격, 분위기, 재방문의사 이렇게 5가지 범주에 따른 음식점 리뷰의 긍/ 부정 빈도를 평가하여 새로운 리뷰점수로 환산하면 다음과 같습니다.

VI. Conclusion

6.1. 기대효과

이번 ‘리슐랭, 진(짜)리(뷰)를 찾아서’ 프로젝트를 통해 배달 음식점의 리뷰를 기존에 존재하는 긍/부정 평가 및 감정분석에서 더 나아가 차별화된 리뷰 평가 방식을 만들었습니다. 리뷰들을 한번 더 필터링해서 소비자와 점주에게 도움이 되는 새로운 리뷰점수를 도입하면, 소비자와 점주 모두에게 긍정적인 효과를 가져올 것으로 생각합니다.

첫 번째 기대효과

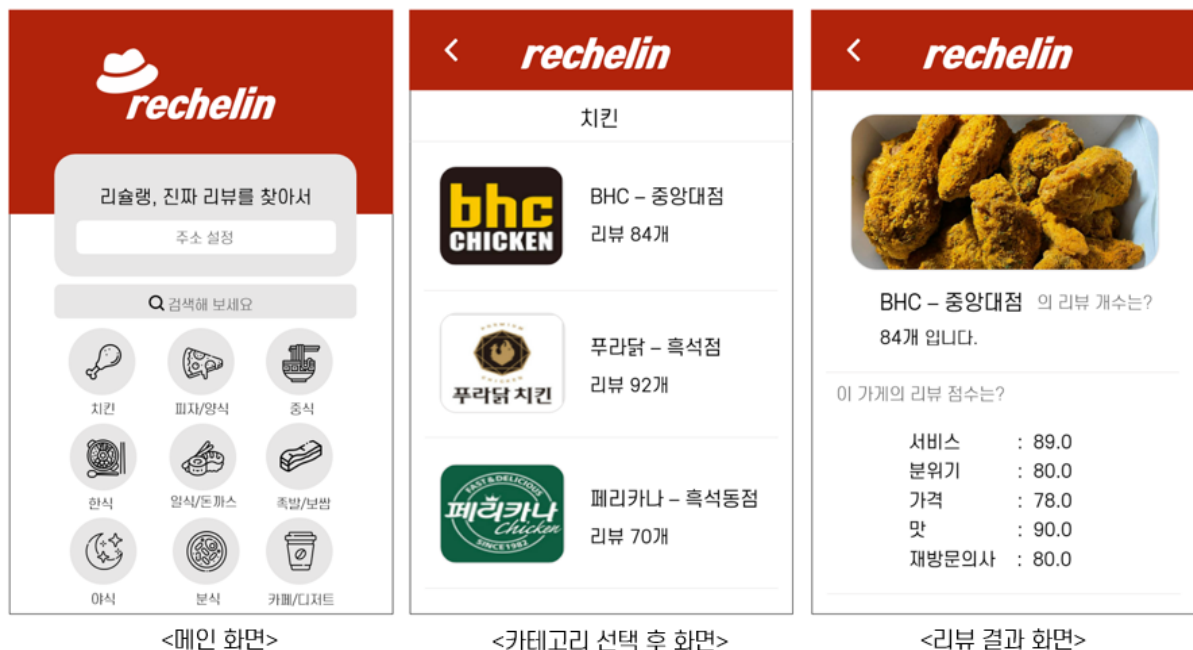
만연해 있는 리뷰 점수 불신을 신뢰로 만드는 것입니다. 새롭게 정의한 리뷰 방식으로 신뢰가 떨어지는 리뷰들을 걸러냄으로써 진실된 리뷰들만을 소비자가 볼 수 있을 것입니다. 이는 리뷰에 대한 믿음의 회복으로 이어져 소비자는 더욱 안심하고 매장 이용이 가능해질 것입니다.

두 번째 기대효과

총 5가지의 항목(서비스, 맛, 분위기, 가격, 재방문 의사)으로 세분화하여 리뷰를 분석함으로써 점주 입장에서는 자신의 서비스에 어떠한 점이 부족한지, 소비자는 어떠한 것을 원하는지를 더욱 정확하게 인지할 수 있습니다. 소비자의 항목별 평가를 통해 음식점의 부족한 부분을 찾고, 이를 개선해 나아가는것으로 더욱 경쟁력있는 음식점으로의 발전이 생길 것입니다.

소비자와 점주에게 도움이 되는 새로운 리뷰점수를 도입하는 것이 저희가 기획한 ‘리슐랭, 진(짜)리(뷰)를 찾아서’ 프로젝트의 목표입니다. 이 프로젝트는 소비자들의 음식점 선택에 올바른 근거를 제시해 줌과 동시에 음식점의 자체적인 발전에도 기여할 것으로 예상됩니다.

6.2. Future Work



‘리슐랭, 진(짜)리(뷰)를 찾아서’ 프로젝트를 적용한 앱으로의 발전을 기대해 볼 수 있겠습니다.

task	프로젝트 기간													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 선정 및 기획														
개발환경 설정														
리뷰 수집(크롤링)														
리뷰 전처리														
모델 개발														
모델 학습														
모델 테스트														
UI/UX 기획														
결과물 HTML제작														
데모 및 설명서 제출														
최종 발표														
최종보고서 제출														

개인일정

이현지

task	프로젝트 기간													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 선정 및 기획														
개발환경 설정														
UI/UX 기획														
메인 화면 뷰 개발														
검색 및 리뷰 화면 뷰 개발														
데이터 & 모델 병합 작업(필터링 및 전처리 지원)														
결과물 PPT 제작														
데모 및 설명서 제출														
최종 발표														
최종보고서 및 매뉴얼 제출														

조중현

[illegible]

모델 개발													
모델 학습													
모델 테스트													
데이터와 모델 병합													
결과물 HTML제작													
데모 및 설명서 제출													
최종 발표													
최종보고서 및 매뉴얼 제출													

전현욱

task	프로젝트 기간													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 선정 및 기획														
개발환경 설정														
리뷰 크롤링														
리뷰 데이터 전처리														
학습 데이터 및 입력 데이터 분류														
키워드 추출														
감성 사전 제작 및 감성분석														
데모 및 설명서 제출														
최종 발표														

최종보고서 및 매뉴얼 제출																
-------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--