

LIS4813 딥러닝개론 Final Project

텍스트 리뷰 감성 분석을 통한 서대문구 내 음식점 별점 리뉴얼

2020122006 권수현
2020122037 김채성
2021122070 박지은

INDEX

01 Introduction

02 Data & Model

03 Process

04 Conclusion

01

Introduction

01 Introduction

오늘은
어디 가서
먹지?



< 신촌맛집

인증

인기메뉴 ▾

영업중

쿠폰

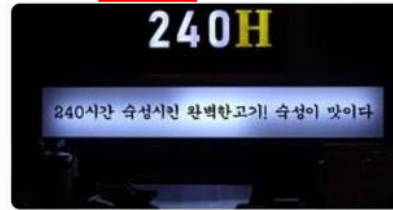
N 포장주문

N 예약

광고 240H N 예약 육류, 고기요리

숙성 고깃집 명품 이베리코

영업 종료 ★ 4.95 리뷰 231



광고 우드문 특 특 카페, 디저트

크로플이 맛있는 카페

영업 종료 · 리뷰 105 · 서울 서대문구 창천동 ▾



한림돈가 본점 돼지고기구이

영업 종료 ★ 4.38 리뷰 999+ TV 생방송투데이



“ 꼬들목살. 대존맛. 진짜. 감격적. 갓김치. 대존맛. 고사리. 하.

별점 리뷰 종료 안내

방문자 리뷰의 별점 선택 기능이 종료되어
현재 표시되는 **평균 별점은 2021년 10월**
까지 참여한 결과입니다.

평균 별점 노출은 **2022년 내에 종료**될 예
정입니다.

확인

01 Introduction

음식점의 평판을 한 눈에 파악할 수 있는 지표의 부재

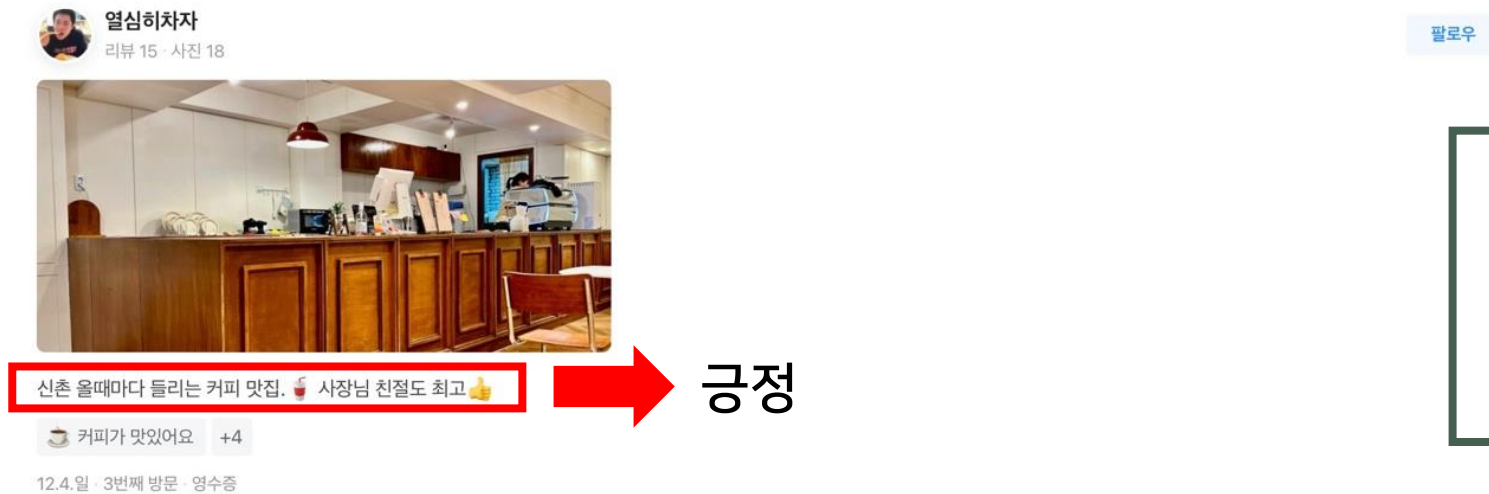
오늘은
어디 가서
먹지?

텍스트 리뷰 감성 분석을 통한 별점 리뉴얼

01 Introduction

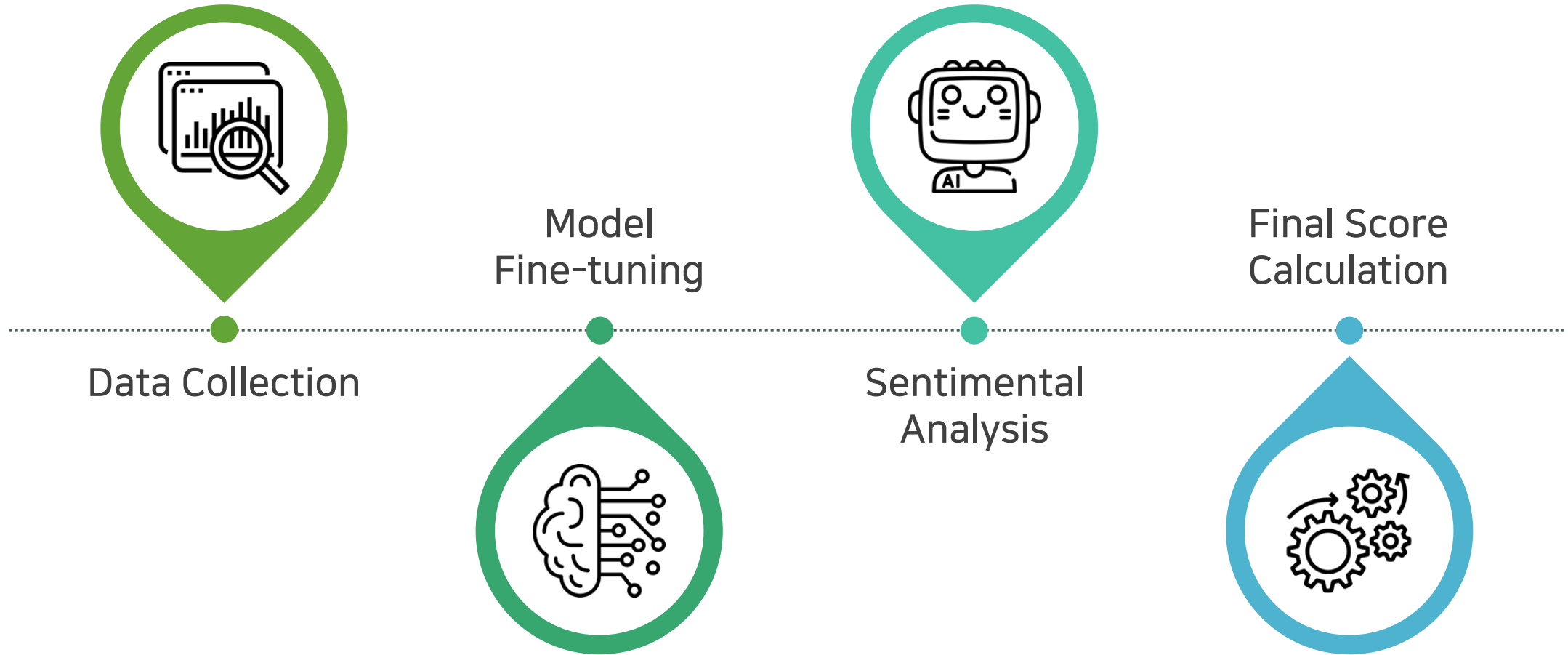


→ 긍정



텍스트 리뷰를
별점의 근거로 이용

01 Introduction



02

Data & Model

02.1 Data

네이버 지도 리뷰 데이터

소상공인시장진흥공단_상가(상권)정보

영업 중인 전국 상가업소 데이터를 제공합니다.
(상호명, 업종코드, 업종명, 지번주소, 도로명주소, 경도, 위도 등)

👍 38

👎 0

🔖 관심

크롤링 데이터
가이드라인으로 이용

	이름	상호명	지점명	상권업종대 분류명	상권업종중 분류명	시도명	시군구명	행정동명	법정동명	지번주소	도로명	도로명주소
0	손가네빈대떡	손가네빈대떡	NaN	음식	한식	서울특별시	서대문구	홍은2동	홍은동	서울특별시 서대문구 홍은동 409-51	서울특별시 서대문구 증가로10길	서울특별시 서대문구 증가로10길 63
1	ZINO본점본점	ZINO본점	본점	음식	양식	서울특별시	서대문구	신촌동	대신동	서울특별시 서대문구 대신동 84-1	서울특별시 서대문구 연대동문길	서울특별시 서대문구 연대동문길 45
2	플라토어학학원	플라토어학학원	NaN	음식	닭/오리요리	서울특별시	서대문구	홍은2동	홍은동	서울특별시 서대문구 홍은동 426	서울특별시 서대문구 연희로	서울특별시 서대문구 연희로 286
3	에브리키친	에브리키친	NaN	음식	한식	서울특별시	서대문구	홍제1동	홍제동	서울특별시 서대문구 홍제동 82	서울특별시 서대문구 통일로25길	서울특별시 서대문구 통일로25길 30
4	참숯돼지천국	참숯돼지천국	NaN	음식	한식	서울특별시	서대문구	홍은1동	홍은동	서울특별시 서대문구 홍은동 9-43	서울특별시 서대문구 홍은중앙로	서울특별시 서대문구 홍은중앙로 118

02.1 Data

네이버 지도 리뷰 데이터

Q 손가네빈대떡

×

손가네빈대떡 전,빈대떡
영업 종료 · 16:00에 영업 시작
서울 서대문구 남가좌동 ▾

출발 | 도착

손가네지짐이 전,빈대떡
영업 종료 · 16:00에 영업 시작
원주 단계동 ▾

출발 | 도착

손가네빈대떡 전,빈대떡
★ 4.46/5 방문자리뷰 72 · 블로그리뷰 11

출발 | 도착

저장하기 | 거리뷰 | 공유

홈 | 소식 | 메뉴 | 리뷰 | 사진

이벤트, 혜택 알림을 받아보세요! 알림받기

서울 서대문구 거북골로 33-10 1층 ▾

명지대학교 정문 이디야커피 골목

영업 종료 · 16:00에 영업 시작 ▾

02-308-3250 복사

제로페이

손가네빈대떡

홈 | 소식 | 메뉴 | 리뷰 | 사진

리뷰 31 사진 리뷰만

전 4 닭발 2 막걸리 2 빈대떡 2 참소라 1

만족도 7 청결도 2 가격 2 음식량 2 배달

추천순 Beta 최신순

672G
리뷰 57 · 사진 142 · 팔로워 2

팔로우

갑자기 집에 들어가는 길에 전이 먹고 싶어서 포장 주문 했습니다! 대중 시간 얘기하고 마트에서 장보고 갔는데 미리 준비해주셔서 바로 집에와서 너무 ... ▾

친절해요 +4

11.7.월 · 1번째 방문 · 영수증

희망50
리뷰 266 · 사진 143 · 팔로워 5

팔로우

일반지도 | 위성지도

대극장
홍은동두산위브 아파트
홍연아파트
산도시연공원

© NAVER 10

Selenium & BeautifulSoup을
이용한 크롤링

02.1 Data

기타 음식점 및 쇼핑 리뷰 데이터

음식점 리뷰 데이터

		review	score
0		친절하시고 깔끔하고 좋았습니다	5
1		조용하고 고기도 굿	5
2		갈비탕과 냉면, 육회비빔밥이 맛있습니다.	4
3		대체적으로 만족하나wn와인의 구성이 살짝 아쉬움	4
4		고기도 맛있고 서비스는 더 최고입니다~	5
...	
8009		가짜원조... 줄서는사람 뭐지???	1
8010	짜통 여긴 원조랑 아무 관계 직원 연관되게 없는 그냥 이름만 .. 가짜입니다 속지 마세요		1
8011		가성비 별로 비싸기만함	1
8012	별 한개주기도 아까움. 별을 줘야 등록돼서 평가함! 나 같은 사람 또 안생겼으면 좋...		1
8013	자장면은 맛있는데 나머지는 동네 중국집 수준 회식으로 갔어서 가격은 잘모르겠네요		4

쇼핑 리뷰 데이터

		review	score
0		배송빠르고 굿	5
1		택배가 엉망이네용 저희집 밑에층에 말도없이 놔두고가고	2
2		아주좋아요 바지 정말 좋아서2개 더 구매했어요 이가격에 대박입니다. 바느질이 조금 ...	5
3		선물용으로 빨리 받아서 전달했어야 하는 상품이었는데 머그컵만 와서 당황했습니다. 전...	2
4		민트색상 예뻐요. 옆 손잡이는 거는 용도로도 사용되네요 ㅎㅎ	5
...	
9995		기대를 너무 많이했나봐요 일단 머릿결 상하신 분은 고민좀 해보세요 ...	1
9996		너무 귀엽고 폭신평신했고 예뻐요	5
9997		눈뒤에 실거는부분이커서 다뜨고 눈을볼일수가없네요 ㅏ	1
9998		짜증 옷이 뒤틀려서 왔음. 짜증 ㄷ신ㅅㅇㄷㄹ즌ㅅㄴㅅㄷㅅ가ㅡㄱㅅㄱㄴㅅ튼ㅅㄱㄷ쓰는할ㅎ...	2
9999		유통기한이 이번달이랑 다음달까지네요 너무 짧아요	2



Fine-tuning시 이용

02.2 Model

KoBERT

SKTBrain/**KoBERT**

Korean BERT pre-trained cased (KoBERT)



5

Contributors

3

Issues

627

Stars

159

Forks



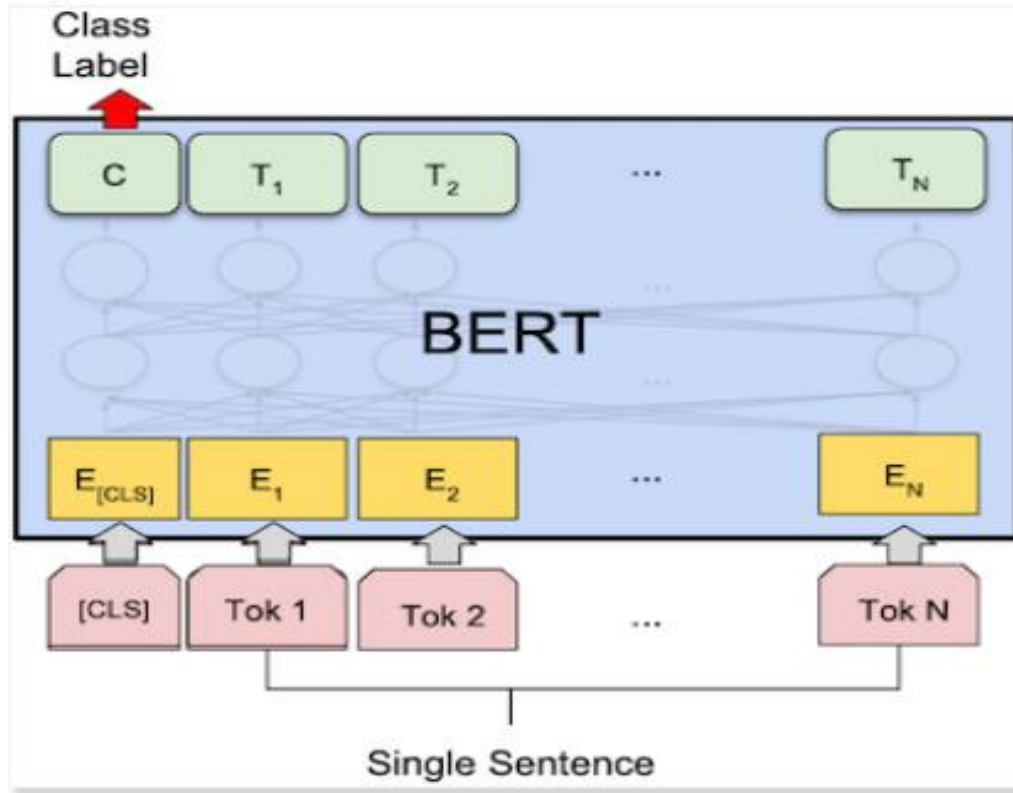
KoBERT

KoBERT

- SKTBrain에서 공개한 한국어 버전의 BERT 모델
- 한국어 위키 5백만 문장 & 뉴스 2천만 문장으로 pretrained

02.2 Model

KoBERT



BERT

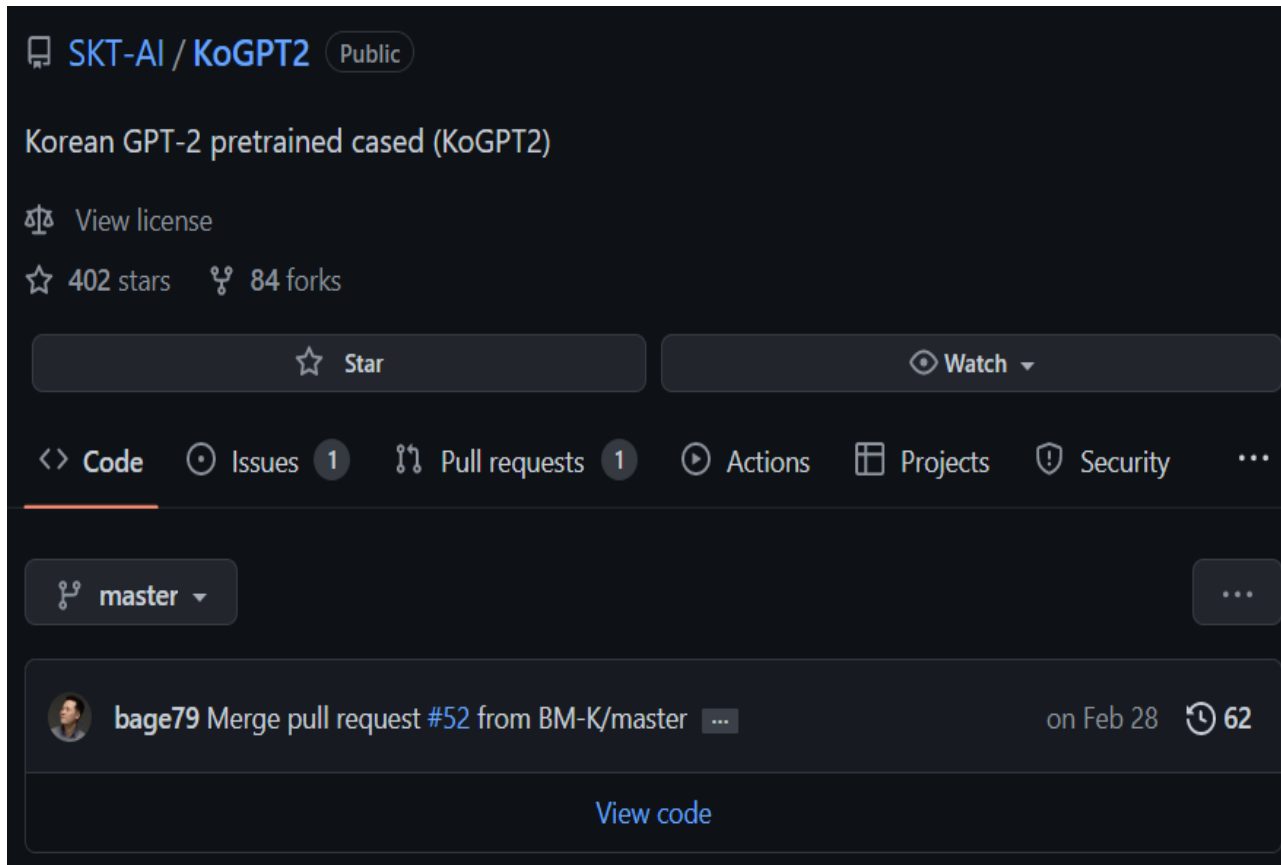


BERT

- Bidirectional Encoder Representation from Transformers
- Transformer의 encoder 구조를 이용
- 문맥 이해도 ↑
- 수행하려는 task에 맞게 fine-tuning 진행

02.2 Model

KoGPT-2



KoGPT-2

- SKT-AI에서 GPT-2의 한국어 성능 개선을 위해 개발한 모델
- 한국어 위키 & 청와대 국민청원 등을 포함한 데이터셋으로 pretrained
- CBPE tokenizer 사용

02.2 Model

KoGPT-2



GPT-2

- Generative Pretrained Transformer
- Transformer의 decoder 구조를 이용
- Auto-regressive
- Zero-shot learning => fine-tuning 필수 X

03

Process

03.1 Data Preprocessing

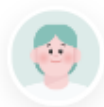
	review	score	label
0	친절하시고 깔끔하고 좋았습니다	5	1
1	조용하고 고기도 굿	5	1
2	갈비탕과 냉면, 육회비빔밥이 맛있습니다.	4	1
3	대체적으로 만족하나wn와인의 구성이 살짝 아쉬움	4	1
4	고기도 맛있고 서비스는 더 최고입니다~	5	1
...
8009	가짜원조... 줄서는사람 뭐지???	1	0
8010	짹통 여긴 원조랑 아무 관계 직원 연관되게 없는 그냥 이름만 .. 가짜입니다 속지 마세요	1	0
8011	가성비 별로 비싸기만함	1	0
8012	별 한개주기도 아까움. 별을 줘야 등록돼서 평가함! 나 같은 사람 또 안생겼으면 좋...	1	0
8013	자장면은 맛있는데 나머지는 동네 중국집 수준 회식으로 갠어서 가격은 잘모르겠네요	4	1

	review	score	label
0	배공빠르고 굿	5	1
1	택배가 엉망이네용 저희집 밑에층에 말도없이 놔두고가고	2	0
2	아주좋아요 바지 정말 좋아서2개 더 구매했어요 이가격에 대박입니다. 바느질이 조금 ...	5	1
3	선물용으로 빨리 받아서 전달했어야 하는 상품이었는데 머그컵만 와서 당황했습니다. 전...	2	0
4	민트색상 예뻐요. 옆 손잡이는 거는 용도로도 사용되네요 ㅎㅎ	5	1
...
9995	기대를 너무 많이했나봐요 일단 머릿결 상하신 분은 고민좀 해보세요 ...	4	0
9996	너무 귀엽고 폭신평신하고 예뻐요	5	1
9997	눈뒤에 실거는부분이커서 다뜨고 눈을볼일수가없네요 T_T	4	0
9998	짜증 웃이 뒤틀려서 왔음. 짜증 드신ㅅㅇㄷㄹ즌ㅅㄴㅅㅌㅅ가ㅡㄱㄱㄴㅅ튼ㅅㄱㅇ드쓰는할ㅎ...	2	0
9999	유통기한이 이번달이랑 다음달까지네요 너무 짧아요	2	0

labeling

Score ≤ 3 : negative(0)
Score > 3 : positive(1)

03.1 Data Preprocessing



Popop

리뷰 102 · 사진 3 · 팔로워 2

팔로우

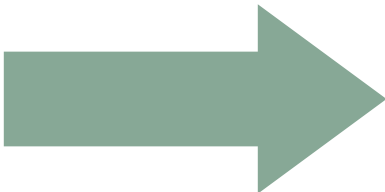
모듬전 가성비 좋은듯해요 좀 멍멍한 하얀 전이 있던 데(녹두가 조금 들어간걸까) 그것만 남겼네용 다른 선택지도 있다면 좋겠다 싶었어요



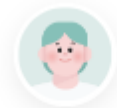
가성비가 좋아요

+1

7.16.토 · 1번째 방문 · 영수증



문장 분리 기준?



정의구현사회

리뷰 5

팔로우

이전 후 오랜만에 방문했네요. 가게는 더 깔끔해졌고. 맛과 사장님 친절함은 여전했습니다. 힘든 때인디 다들 힘내요~

★5 · 21.1.5.화 · 1번째 방문 · 영수증

03.1 Data Preprocessing

KSS 한국어 문장 분리 알고리즘

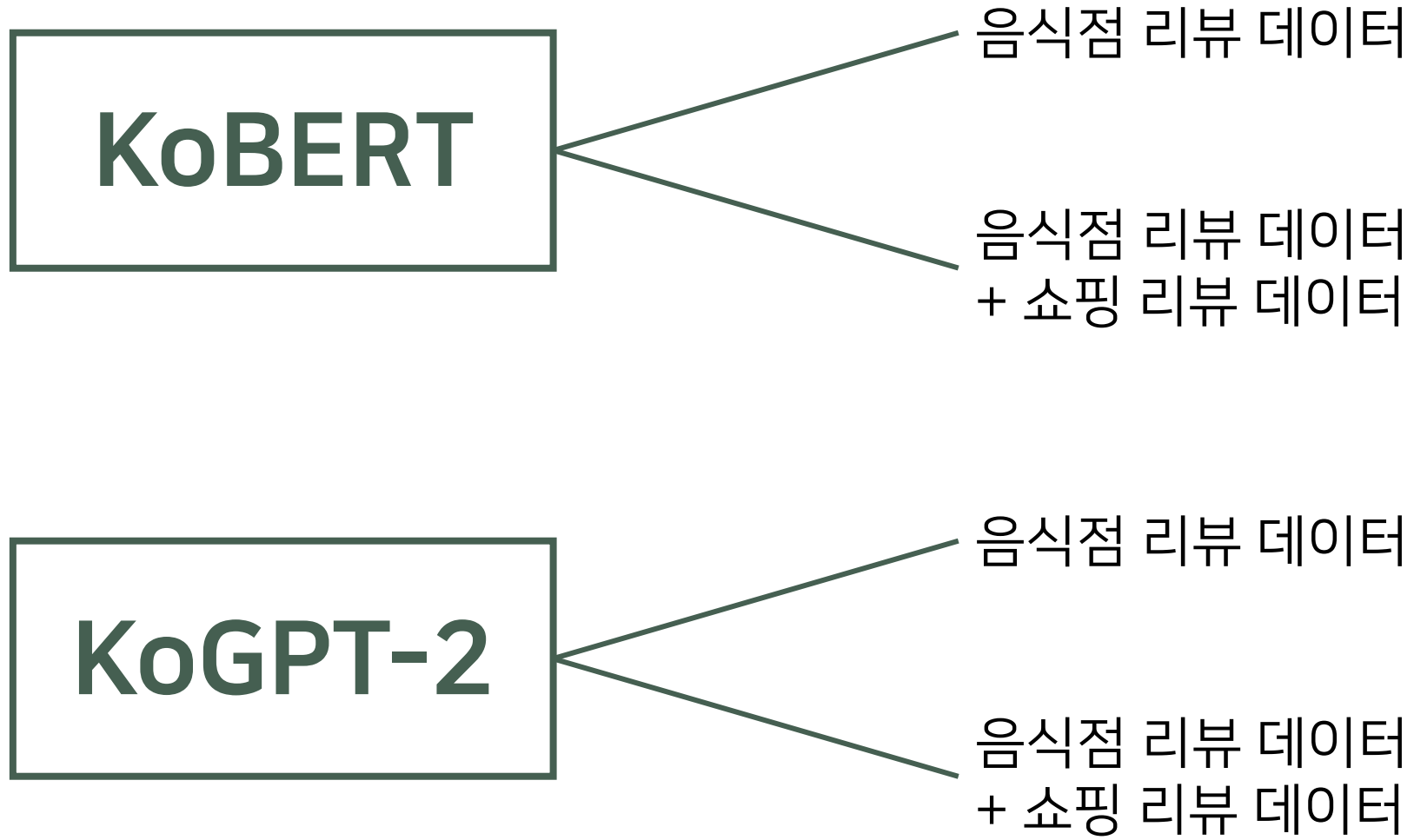
Kss: A Toolkit for Korean sentence segmentation

release v3.7.3 issues 5 open

This repository contains the source code of Kss, a representative Korean sentence segmentation toolkit. I also conduct ongoing research about Korean sentence segmentation algorithms and report the results to this repository. If you have a good idea about Korean sentence segmentation, please feel free to talk through the [issue](#).

준?

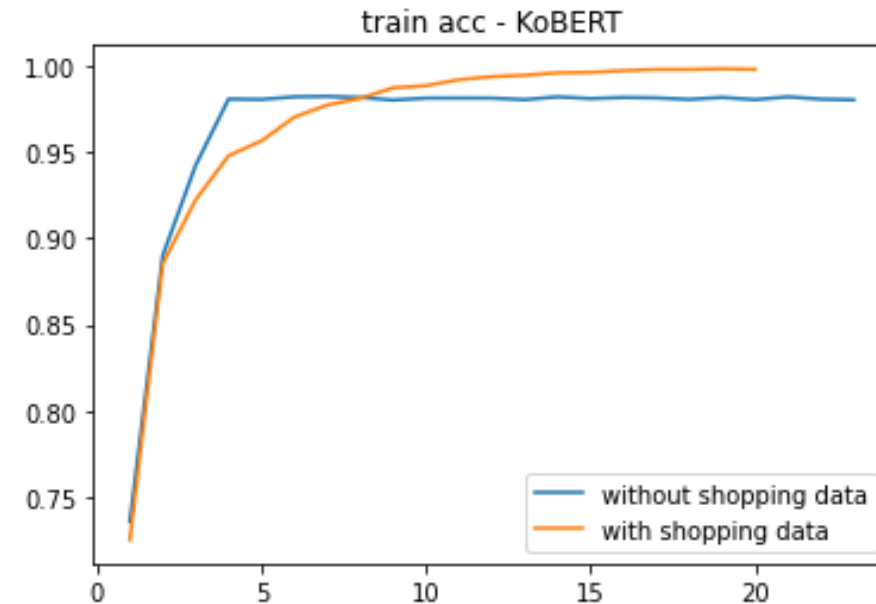
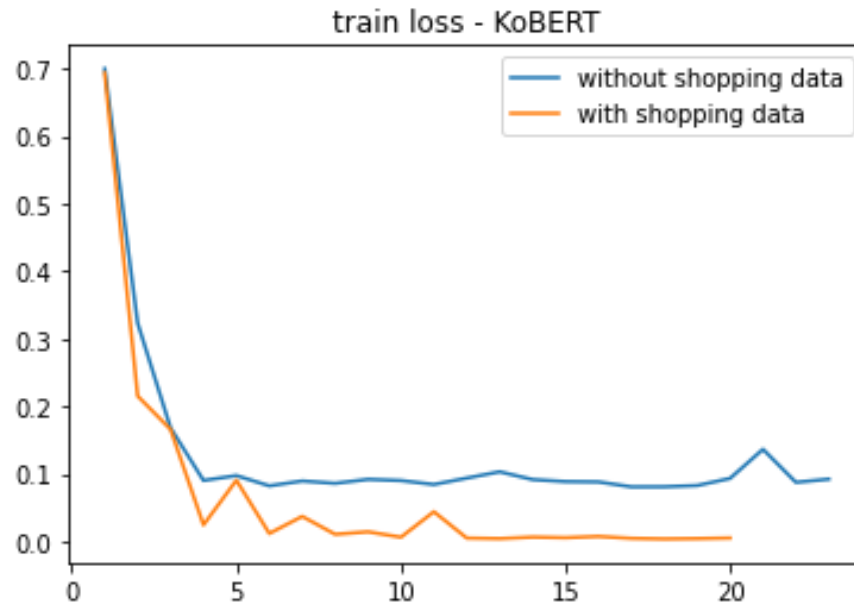
03.2 Model Fine-tuning



03.2 Model Fine-tuning

KoBERT

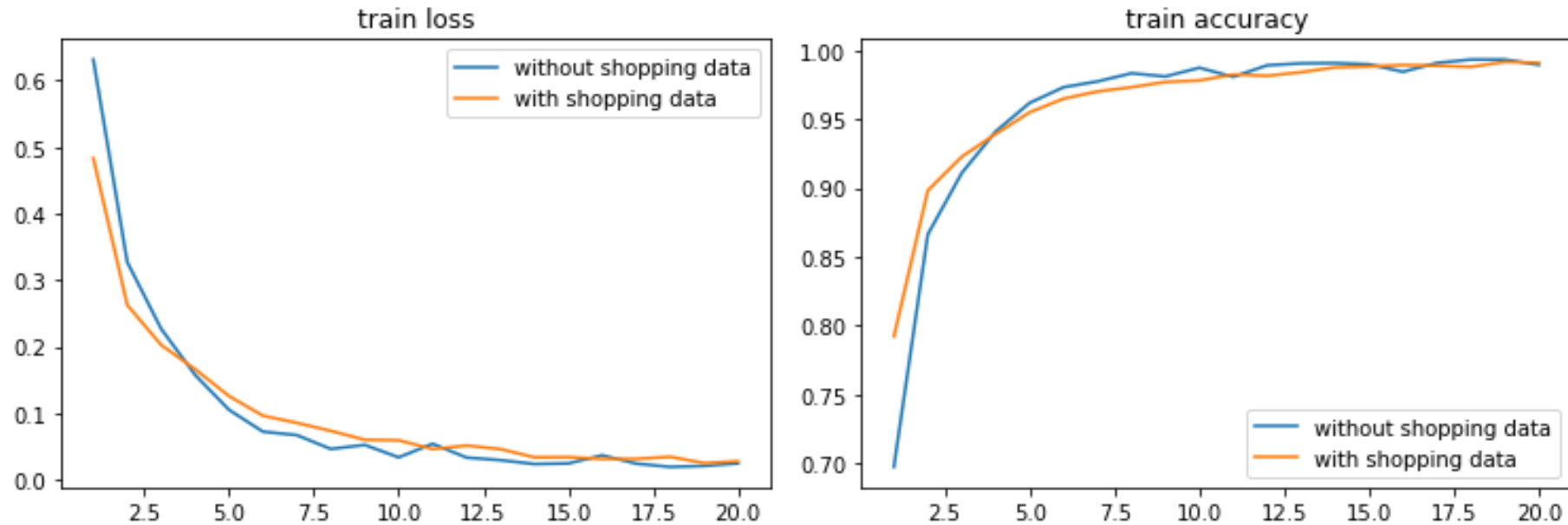
Data	Train_loss	Train_acc	Test_loss	Test_acc
음식점	0.092	0.980		0.9180
음식점+쇼핑	0.004	0.998		0.9294



03.2 Model Fine-tuning

KoGPT-2

Data	Train_loss	Train_acc	Test_loss	Test_acc
음식점	0.0247	0.9893	0.8350	0.8674
음식점+쇼핑	0.0279	0.9908	0.6677	0.8767



03.2 Model Fine-tuning

KoBERT

92.94%

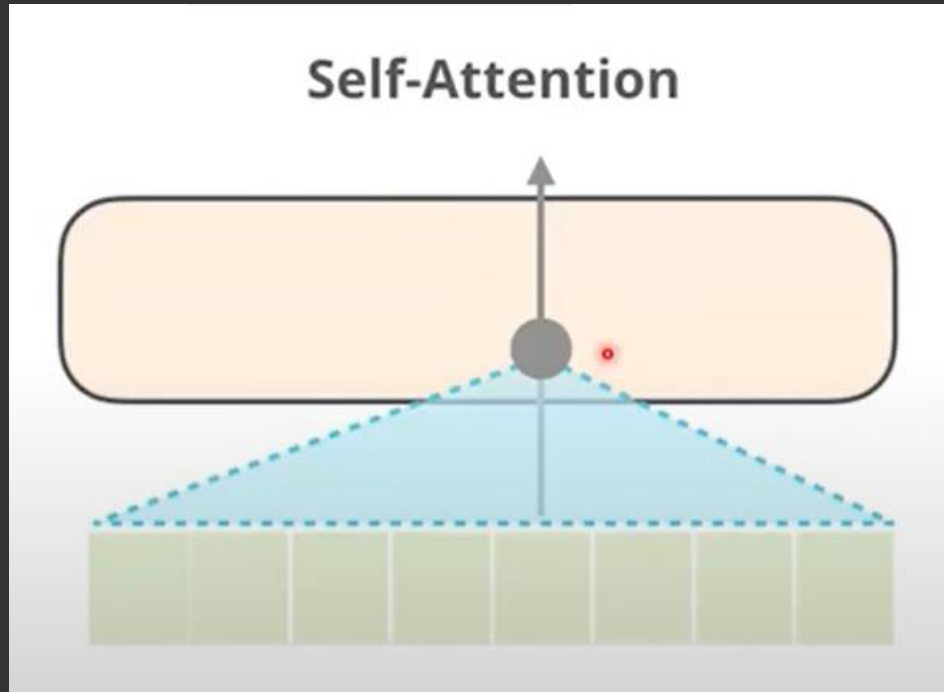
VS

KoGPT-2

87.67%

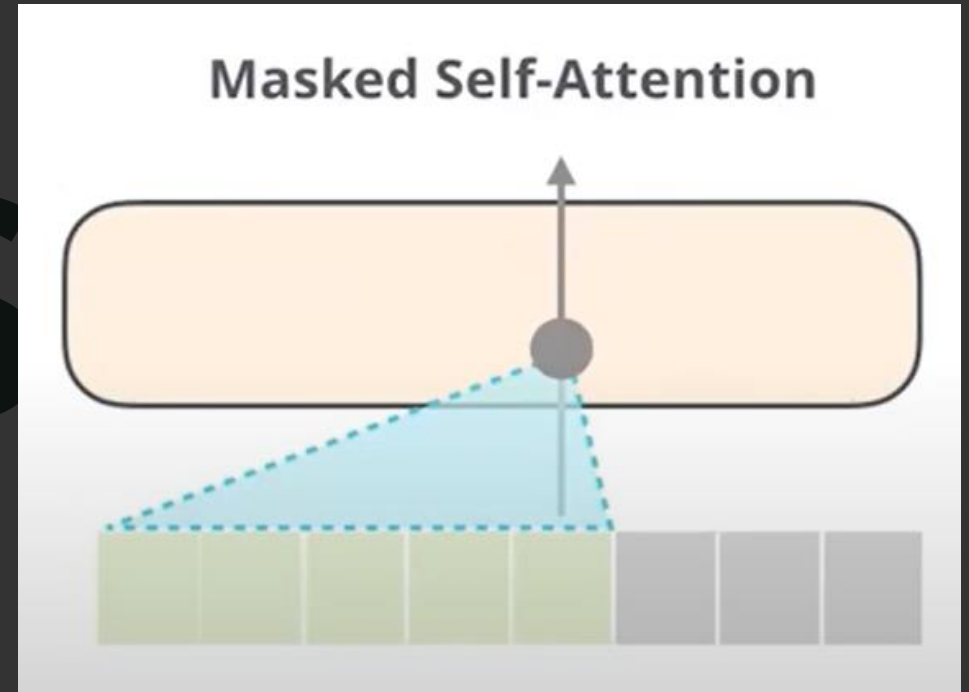
03.2 Model Fine-tuning

KoBERT



맥락 파악 good

KoGPT-2



다음 단어 예측 good

03.3 Sentimental Analysis

```
[51] sentiment_predict('저렴한 전집이지만 바로 부쳐주시니 맛은 보장됩니다.')
```

```
1/1 [=====] - 0s 43ms/step  
99.01% 확률로 긍정 리뷰입니다.
```

```
[52] sentiment_predict('조금씩 시켜먹기 좋게 메뉴판 구성이 되어있고 사이드도 다채롭습니다.')
```

```
1/1 [=====] - 0s 34ms/step  
99.99% 확률로 긍정 리뷰입니다.
```

```
[39] sentiment_predict('정문 근처 카페 찾다가 발견했는데 진짜 올해 간 카페 중에 제일 좋았어요...♥')
```

```
1/1 [=====] - 0s 40ms/step  
100.00% 확률로 긍정 리뷰입니다.
```

```
[42] sentiment_predict('크림치즈크로플 너무 맛있었습니다..!')
```

```
1/1 [=====] - 0s 38ms/step  
100.00% 확률로 긍정 리뷰입니다.
```

```
[43] sentiment_predict('카야크로플은 호불호 갈릴거 같은데 저는 편향았습니다.')
```

```
1/1 [=====] - 0s 72ms/step  
99.97% 확률로 긍정 리뷰입니다.
```

```
[44] sentiment_predict('콘센트도 많고 테이블 크기도 적당해서 노트북 작업하기 좋은데 다만 테이블 높이가 좀 낮아서 목이 아팠어요ㅠㅠ')
```

```
1/1 [=====] - 0s 40ms/step  
53.74% 확률로 긍정 리뷰입니다.
```

03.4 Final Score Calculation

1. 한 리뷰 내에 긍정&부정 문장이 모두 존재할 경우

“음식 맛은 너무 좋았어요. 하지만 직원들이 불친절했습니다.” 긍정? 부정?



각 리뷰 내 모든 문장을 분리하여 각 문장별로 감성 분석 진행

2. 한 문장 내에 긍정&부정 문장이 모두 존재할 경우

“맛있는데 직원들은 불친절해요.” 긍정? 부정?

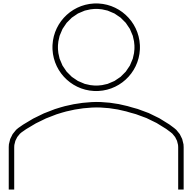


‘긍정으로 판단할 확률’을 이용

03.4 Final Score Calculation

1. 방문자별 리뷰 score 계산 : 각 리뷰 내 모든 문장이 '긍정으로 판단될 확률'의 평균

음식점1



방문자1

정문 근처 카페 찾다가 발견했는데 진짜 올해 간 카페 중에 제일 좋았어요...♥
할로윈 분위기도 물씬 나고 제가 평소에 너무 좋아하는 우드톤과 아늑한 분위기라 너무 맘에 들었어요ㅠㅠ
그리고 사장님께서 너무 친절하고 귀여우시고 선곡 센스도 최고 🎧

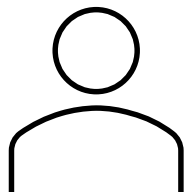
1.000

1.000

1.000



1.000



방문자2

크림치즈크로플 너무 맛있었습니다...!
카야크로플은 호불호 갈릴거 같은데 저는 괜찮았습니다.
콘센트도 많고 테이블 크기도 적당해서 노트북 작업하기 좋은데 다만 테이블 높이가 좀 낮아서 목이
아팠어요ㅠ

1.000

0.997

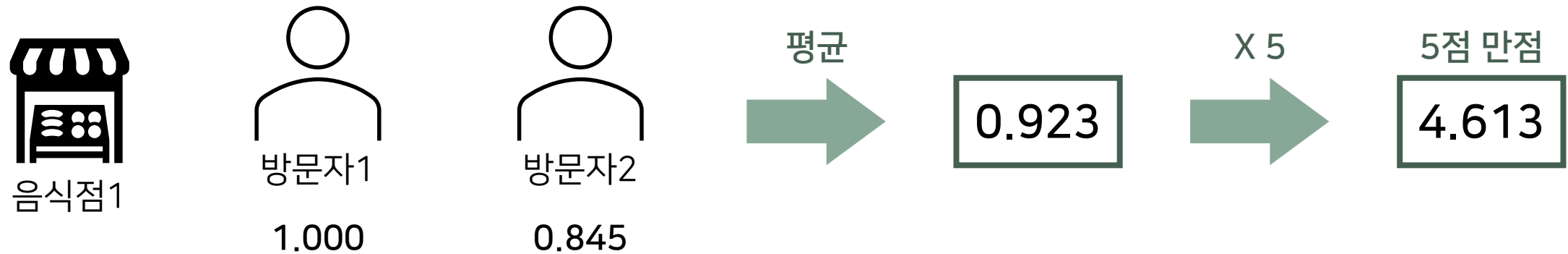
0.537



0.845

03.4 Final Score Calculation

2. 음식점별 최종 별점 계산 : 음식점의 모든 방문자별 리뷰 score의 평균



03.4 Final Score Calculation

최종 별점 계산 결과

index	식당명	기존 별점	KoBERT
1	느티나무	4.83	4.31
2	김덕후의차돌조	4.51	4.27
3	본가생고기	4.46	4.60
4	여우골초밥앤참치	4.43	3.93
5	낙원식당	4.42	4.39
6	레몬	4.39	4.12
7	코쿤캡	4.34	3.98
8	본우리반상	4.29	3.85
9	핑크버터	4.27	3.98
10	거북이집한정식	4.20	3.52
11	명물갈비	4.19	4.11
12	초장쌈밥	4.03	3.56
...			

04

Conclusion

04.1 기대 효과 및 활용 방안

1. 네이버 플레이스 별점 서비스 대체 가능

- 일관적인 기준으로 별점 계산 가능
- 리뷰에 대한 양적 수치화를 통해 음식점의 평판에 대한 대략적인 파악 가능

2. 리뷰의 질적 향상 기대

- 소상공인과 소비자 간의 긍정적인 순환 관계 성립

3. 연세대 학생 및 서대문구 거주민의 음식점 선택에 도움

4. 별점 계산 알고리즘의 다양한 지역과 분야로의 확장 가능성 존재

04.2 한계점

1. 최종적으로 계산한 별점이 얼마나 정확한가에 대한 성능 평가 진행 X
2. '긍정으로 분류할 확률'에 대한 성능 평가 진행 X
3. 단순한 '사실 묘사' 문장 제거 불가
Ex. "엄마와 함께 방문했어요."



THANK YOU

