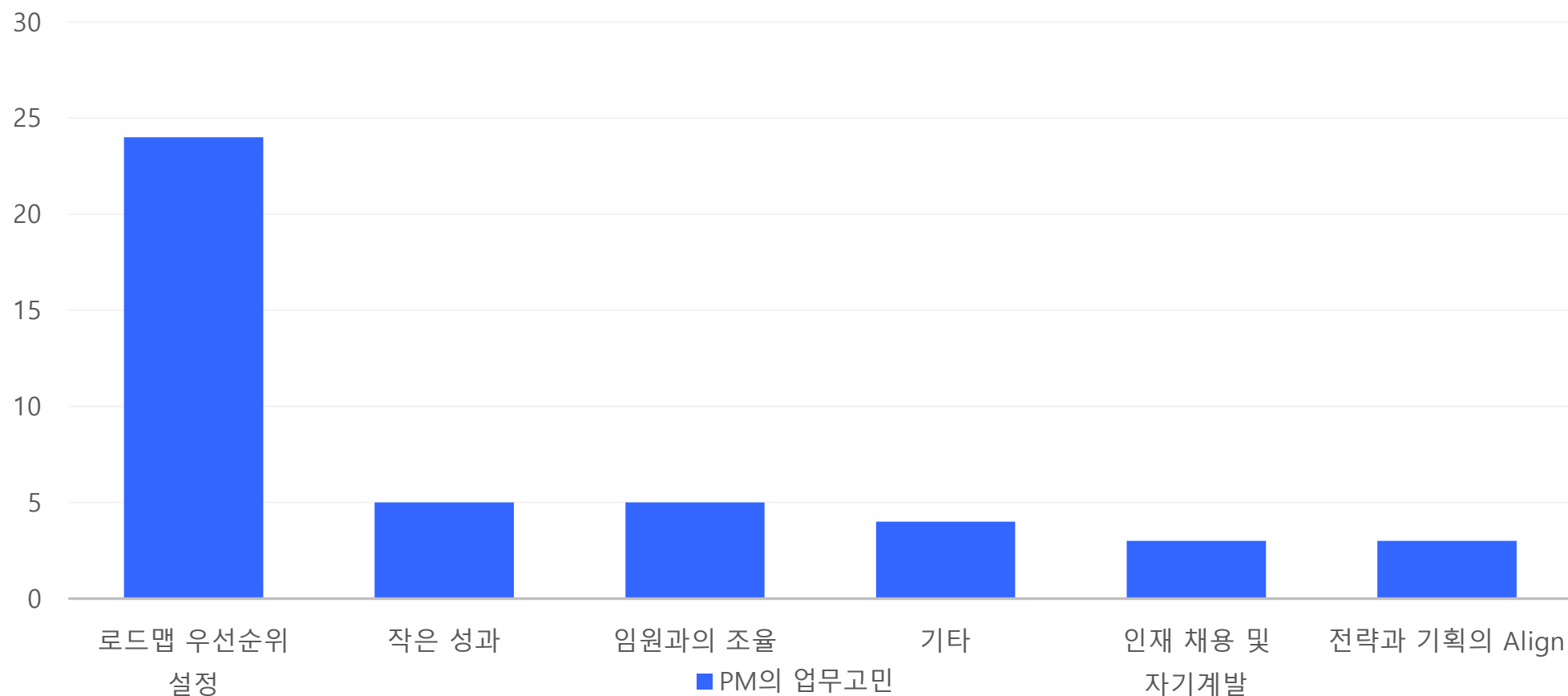


Review

생생한 유저들의 이야기

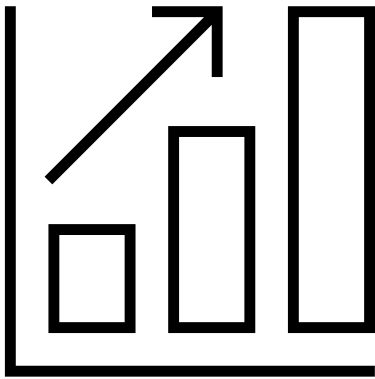
우선 순위가 문제다

Background

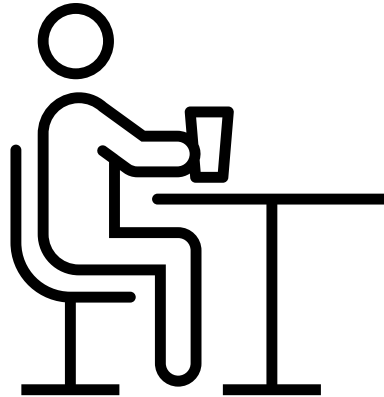


로드맵 설정 방식은?

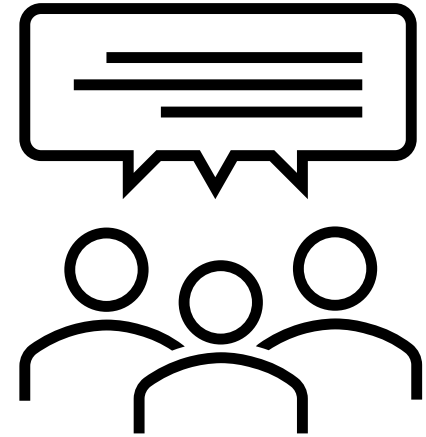
Background



데이터 기반
인사이트



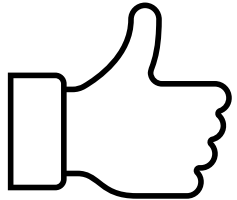
사용자 인터뷰
인사이트



리뷰 기반
인사이트

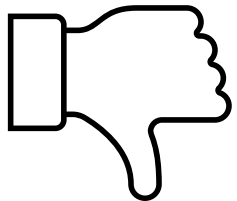
유저들의 의견을 들어보자

Background



씩씩한 너구리

“꼭잠 잘 수 있는 짧은 시간도 생기면 좋을거 같아요 ㅎㅎ (10분, 30분, 1시간)”



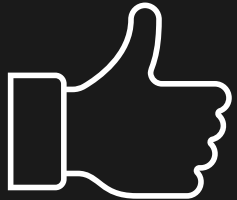
불편한 다람쥐

“저번에도 이런 적 있긴 한데 매트리스에 올려놓으라는 설명 이후 다음을 누르니까
그냥 포인트만 날아가고 재시작되네요. 벌써 세개 날려서 기분이 안좋습니다.”



유저들의 의견을 들어보자

Background



씩씩한 너구리

“꼭잠 잘 수 있는 짧은 시간도 생기면 좋을거 같아요 ㅎㅎ (10분, 30분, 1시간)”



불편한 다람쥐



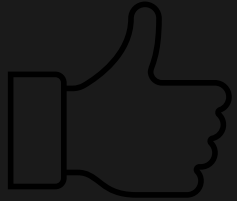
“개미도 이런 점 인기 하데 매트리스에 올려놓으라는 설명 이후 다음을 누르니까 그냥 포인트만 날아가고 재시작되네요. 벌써 세개 날려서 기분이 안좋습니다.”

“10분 단위 서비스를 새로 추가해볼까?”



유저들의 의견을 들어보자

Background



씩씩한 너구리



잠 잘 수 있는 짧은 시간 동안 “오류 발견! 오류발견!!!” (10분, 30분, 1시간)



불편한 다람쥐

“저번에도 이런 적 있긴 한데 매트리스에 올려놓으라는 설명 이후 다음을 누르니까
그냥 포인트만 날아가고 재시작되네요. 벌써 세개 날려서 기분이 안좋습니다.”

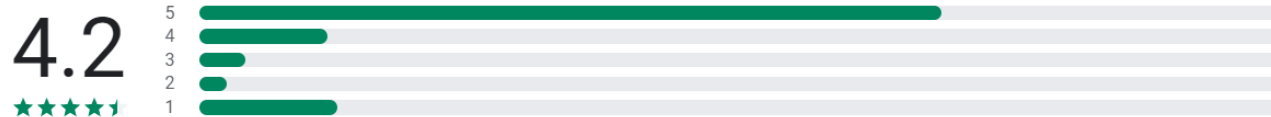


하지만.. 실제 리뷰 데이터는?

Background

평가 및 리뷰 →

검증된 평점과 리뷰입니다. ⓘ



리뷰 3.38만개

T TORA TOKI

★★★★☆ 2022년 8월 6일

컬리앱 너무 잘 쓰고 있었는데.. 얼마전 업데이트 후 재고가 있는 것으로 나오는 제품도 바로 장바구니 들어가보면 품절인 걸로 나오네요.. 한두 번이면 아 사람들이 많이 찾는 제품이네 할텐데 이런 경우가 수도 없이 많습니다.. 품절인데도 품절 표시가 미리 안 되는 이유가 뭔지.. 장보면서 무지 스트레스네요.. 덧붙여서 재입고 알림도 거의 안 옵니다.. 딱 한번 받아봤고요. 뷰티섹션을 늘릴 게 아니라 기존 것을 잘 돌아 가게 해주셔야 될 거 같습니다.. (폰기종은 노트9)

사용자 83명이 이 리뷰가 유용하다고 평가함

이 리뷰가 유용했나요?

예

아니요

(주)컬리

2022년 8월 9일

Love food, Love life! 고객님 안녕하세요. 마켓컬리입니다. 이용에 불편을 드려 대단히 죄송합니다. (_ _) 당시 일시적으로 오류가 확인되어 조치완료하였습니다. 혹여 지속적으로 동일한 이슈가 발생될 경우 어플을 최신버전으로 업데이트하시거나 재설치 부탁드립니다. 해소되지 않을 경우 1:1게시판 혹은 카카오톡으로 문의주시면 담당자 최대한 신속하게 도움드릴 수 있도록 하겠습니다. 꾸준히 모니터링 및 테스트하며 더욱 편리한 환경에서 컬리를 이용하실 수 있도록 최선을 다하겠습니다. 감사합니다. 마켓컬리 드림.

3.3만개?

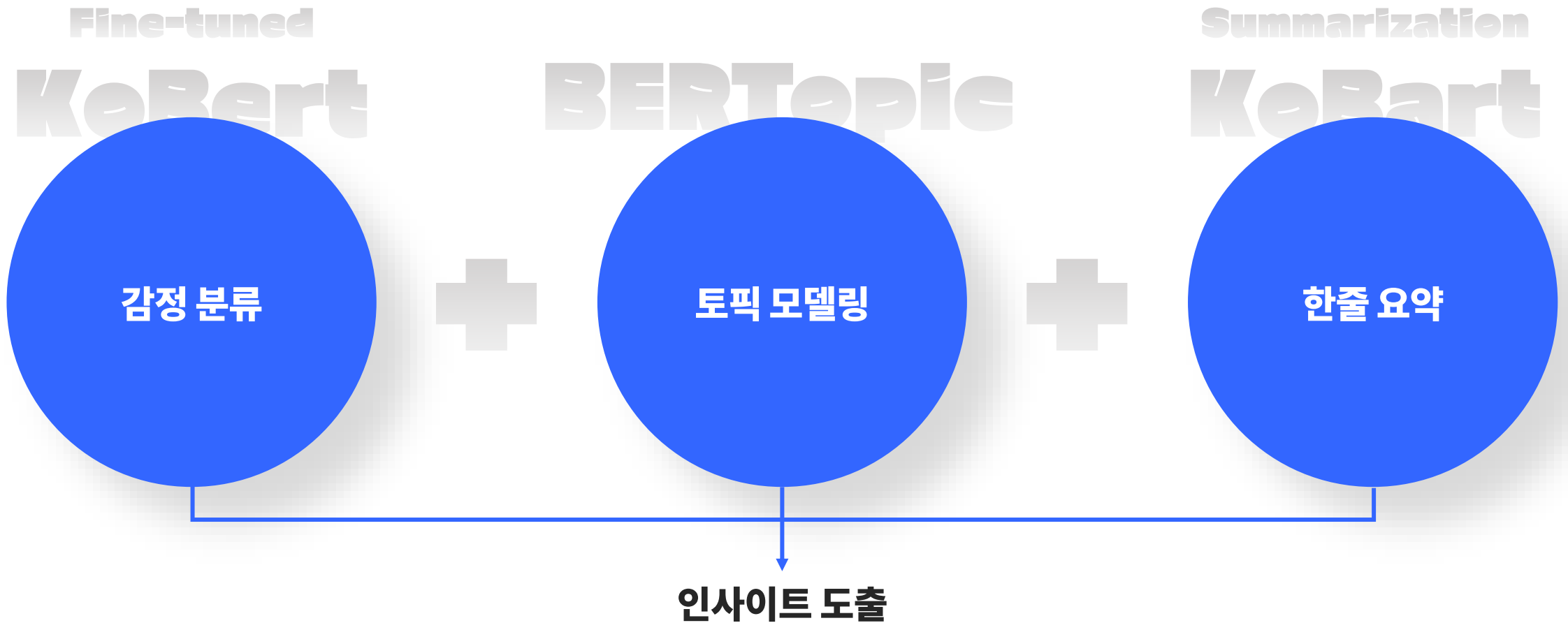
→ 인사이트를 어떻게 뽑으면 될까?

리뷰를 카테고리별로 요약하자

Background



Review - 모델링 파이프라인



Positive

2.1만개

- ✓ **명상_도움** : 불면증이 있었는데 명상이 굉장히 도움되네요
- ✓ **알람_깔끔** : 다른 앱에 비해서 알람 소리가 깔끔해요
- ✓ **자동_결제** : 한달마다 자동으로 결제돼서 너무 편해요!
- ✓ **수면_추적** : 잘 때 수면을 추적해서 보여주니까 신기해요

Negative

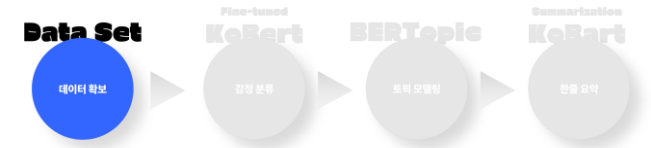
1.2만개

- ✓ **자동_결제** : 자꾸만 자동으로 결제가 돼서 짜증납니다
- ✓ **앱_팅김** : 아니 왜 자꾸 장바구니만 들어가면 앱 튕김?
- ✓ **광고_사기** : 광고가 너무 많아요...이거 사기 앱인가요?
- ✓ **개인_정보** : 왜 이렇게 개인 정보 요구가 많아요? 불안합니다

리뷰 데이터 확보

1. Crawling

1-1 플레이스토어 리뷰



Google Play

수면 트래커
수면, 오토슬립, 코골이
LEAP FITNESS

수면분석 - 수면, 오토슬립, 코골이
Leap Fitness Group
4.9 ★

Sleep Cycle

수면 사이클: 수면 분석 및 스마트 알람 시계
Sleep Cycle AB
4.4 ★

당신의 수면을 추적

당신의 수면을 추적
SM Health Team
4.8 ★

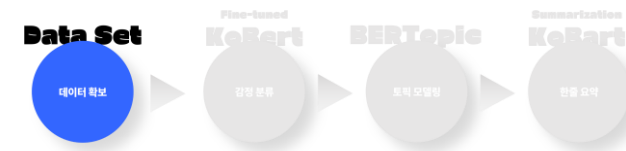
better sleep

BetterSleep: 수면 추적, 고요함 & 명상
Ipnos Software
4.7 ★

수면 앱 이름	리뷰 수
0 미라클나잇	21
1 코끼리	1274
2 수면소리	1097
3 Calm	1682
4 슬립사이클	1734
5 수면분석	383
6 스노어랩	920
7 슬립모니터	323
8 BetterSleep	4600
9 슬립 시어리	353
10 Sleepo	1239
11 Sleepzy	249
12 빗소리	2954
13 SleepTown	593
14 마보	892
15 Sleep Tracker	6
16 꿀밤	21
17 편안한 음악	136
18 마인드위즈	1331
19 굿슬립	25

- ✓ 플레이스토어의 '수면' 키워드 앱 기준
- ✓ 중복된 어플을 제외한 20개 어플
- ✓ 총 19,833개의 리뷰 수집

1-2 확보 데이터 목록



수면분석 - 수면, 오토슬립, 코골이
평가 및 리뷰

Hyun yee

리뷰를 작성한 고객의 ID

★★★★★ 2022년 2월 2일

리뷰를 쓴 날짜와 별점

잠자는데 시간이 많이 걸리는 편인데 앱 내에 잠 잘하는데 좋은 음악들도 다양하고 심지어 명상을 할 수 있는 음악도 있어서 따로 다른 앱을 깔지 않아도 된다는 점이 좋았습니다. 수면 측정도 정확하게 되고 자는 도중에 내가 얼마나 깊게 잤는지 체크해주고 클립으로 잘 때 내가 냈던 소리들이 자동으로 녹음되어 알 수 있어서 정말 좋은 것 같습니다. 특히 이 앱이 무료라는 점이 놀라운데, 만약 유료라고 하더라도 구매해서 사용하고 싶을 정도로 유용하고 쉽게 사용할 수 있도록 만든 앱 같아서 좋았습니다!!!

사용자 37명이 이 리뷰가 유용하다고 평가함

이 리뷰가 유용했나요? ☐ 예 ☐ 아니요

0416

★★★★★ 2022년 5월 21일

넘 잘 쓰고 있는데.. 아침에 일어나면 핸드폰이 너무 뜨겁고 cpu가 과부화되어있다고 하는데 이건 어쩔 수 없는 건가요??ㅜㅜ 그리고 알람 설정할 때는 분명 크게 설정해놨는데 막상 올릴 때는 너무 작아요... 핸드폰 자체의 볼륨이 작으면 그렇게 되는 건가요? 이 두개 외에는 다 너무 좋은데ㅜㅜ

사용자 33명이 이 리뷰가 유용하다고 평가함

이 리뷰가 유용했나요? ☐ 예 ☐ 아니요

	id	date	rating	review
5808	Hyun yee	2022년 2월 2일	5	잠자는데 시간이 많이 걸리는 편인데 앱 내에 잠 잘하는데 좋은 음악들도 다양하고 심...
5809	0416	2022년 5월 21일	5	넘 잘 쓰고 있는데.. 아침에 일어나면 핸드폰이 너무 뜨겁고 cpu가 과부화되어있다...

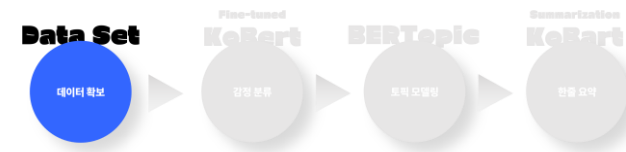
✓ 확보 데이터 목록 (앱 이름별)

1. User ID
2. Rating
3. Date
4. Review

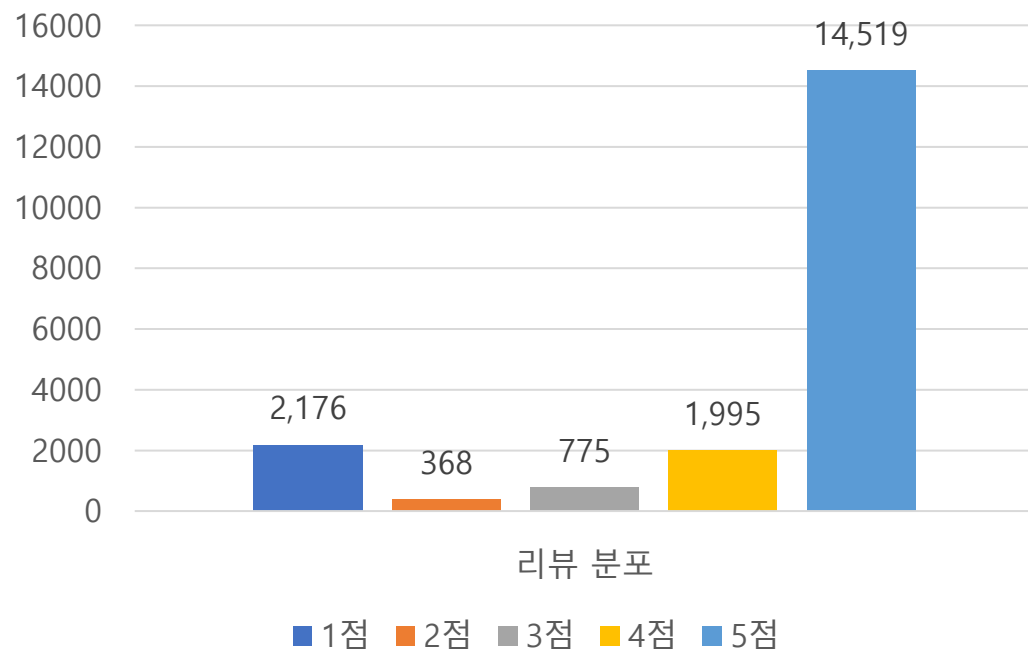
데이터 전처리

2. Preprocessing

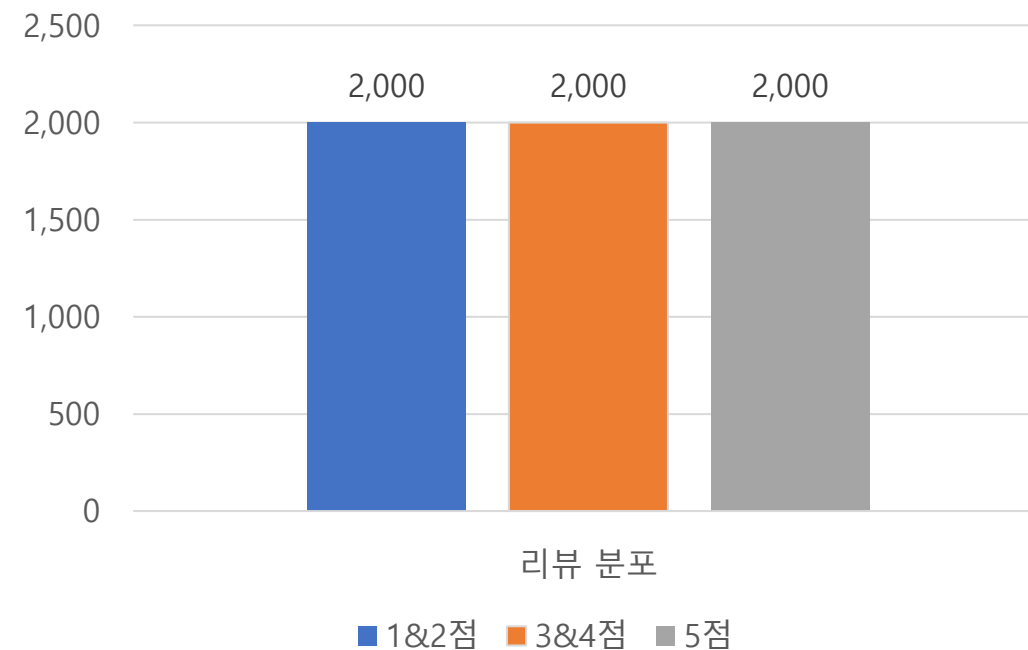
2-1 데이터 불균형 해소



기존 리뷰 데이터 분포

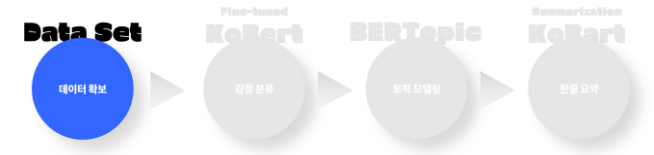


리뷰 데이터 분포 조절



- ✓ Labeling을 위한 데이터 불균형 해소
- ✓ 감정 분류 성능 향상 목적 (긍정/부정)

2-2 문장 단위 분리



Google 사용자

★★★★★ 2019년 9월 4일

pdf파일 보는데 최적화되어 있습니다 하지만 pdf파일이 많을 경우 찾는데 어려움이 있는데요 폴더별로 볼 수는 있으나 폴더안에 문서도 보여주어서 찾기가 좀 힘들습니다. 폴더별로 볼때 처음에 폴더만 보여주고 그 폴더로 들어가면 pdf파일을 볼 수 있는 형식으로 만들어주실 수 있나요?

장점

단점

개선점

리뷰 데이터 특징

1. 문장 구분이 '로 정확히 구분되지 않음
2. 정보가 부족한 문장이 있음
3. 한 문장/리뷰 내에 긍정/부정 감정이 함께 들어 있음
4. 평점이 높더라도 부정 리뷰가 있음

전처리 방식

1. 리뷰를 문장 단위로 분리
2. 품사 수가 부족한 문장을 제거 ($Pos > 10$)

→ 각 문장에 긍/부정 태깅이 가능하도록 분리

2-3 리뷰 전처리 결과



* Konlpy 내의 kkma와 Okt 분석기 이용

1. kkma.sentences 를 통해 문장별로 리스트 생성
2. Okt를 통해 조사, 접미사, 구두점, 외국어, 숫자를 제외
3. 품사가 5개 미만인 경우 문장 결합
4. 합쳐서 정리된 문장을 구두점과 외국어를 제거하여 리스트로 문장 저장

예시 문장

“넌 좋아요..... ♡ 코끼리 앱 주변에도 막 엄청 추천하고 다녀요 ㅎㅎ 명상, 수면, 심리 등등 도움이 엄청 됩니다.”

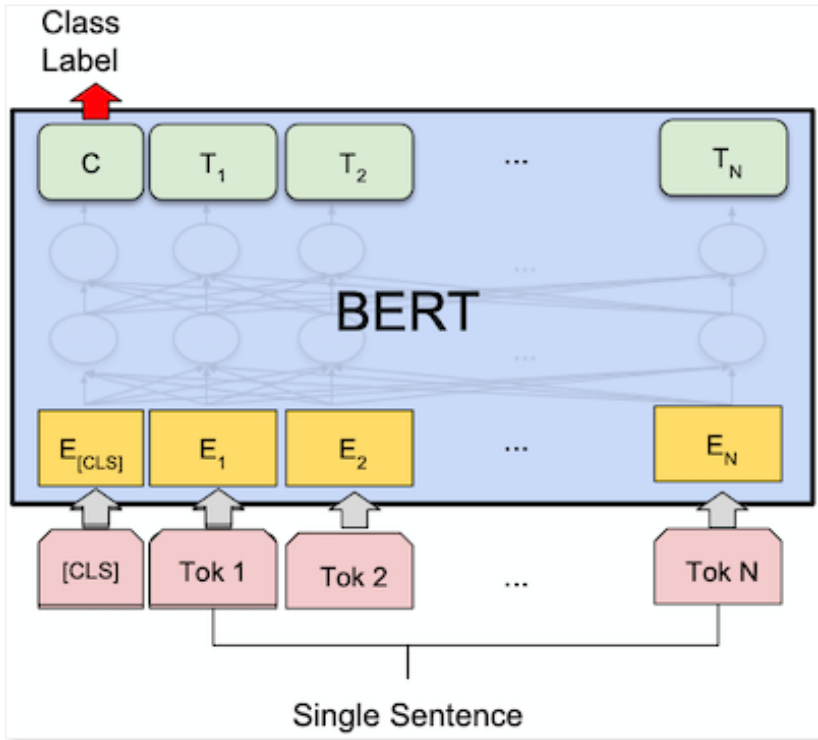
전처리 문장

[‘코끼리 앱 주변에도 막 엄청 추천하고 다녀요’, ‘ㅎㅎ 명상 수면 심리 등등 도움이 엄청 됩니다’]

리뷰 내 긍정/부정 감정 분류

3. Sentiments

3-1 BERT



BERT



내 이름이 BERT란 걸 알고 있었나?

✓ BERT란?

1. Bidirectional Encoder Representation from Transformers
2. Transformer의 Encoder 구조를 차용

✓ Pre-Train model

1. 방대한 Corpus들로 Masked-LM과 같은 Pre-train방식을 거침

✓ Transfer Learning

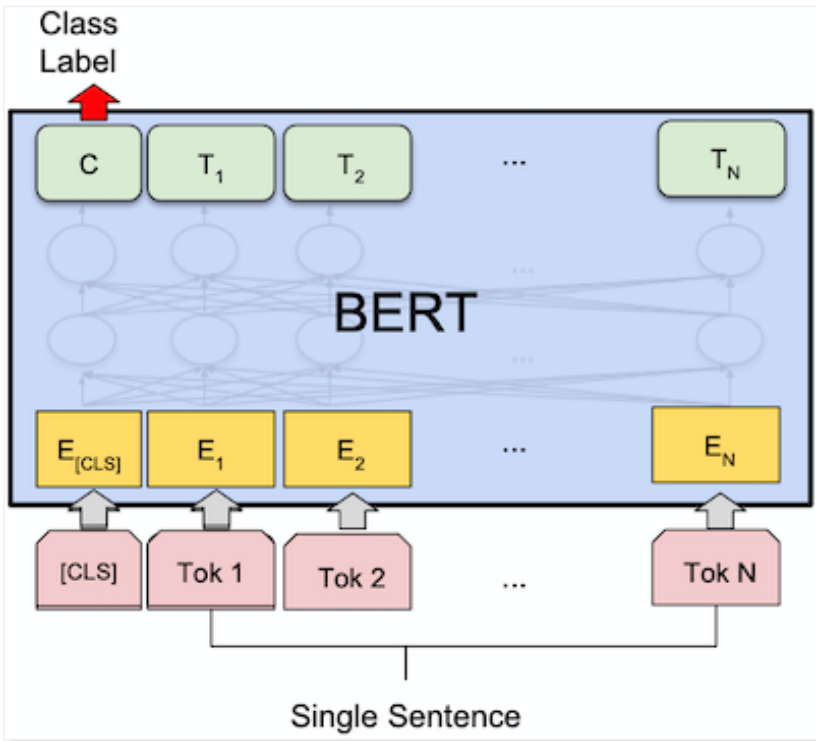
1. 수행하려는 Task를 설정하여 fine-tuning
2. 우리 모델의 경우 ‘앱 리뷰 문장 감성분류’에 맞게 fine-tuning

→ 즉, 이미 잘 학습된 pre-trained된 BERT 모델을 ‘리뷰 감성 분류’에 맞게 조율한 것

3-1 BERT



S | W



BERT



이제 BERT를 까먹지 않겠지?

✓ Fast-Train

: Pre-train 모델이기 때문에
Fine-tuning만 하면 됨!

✓ Small Data

: 적은 데이터셋으로 좋은 성능을
낼 수 있음

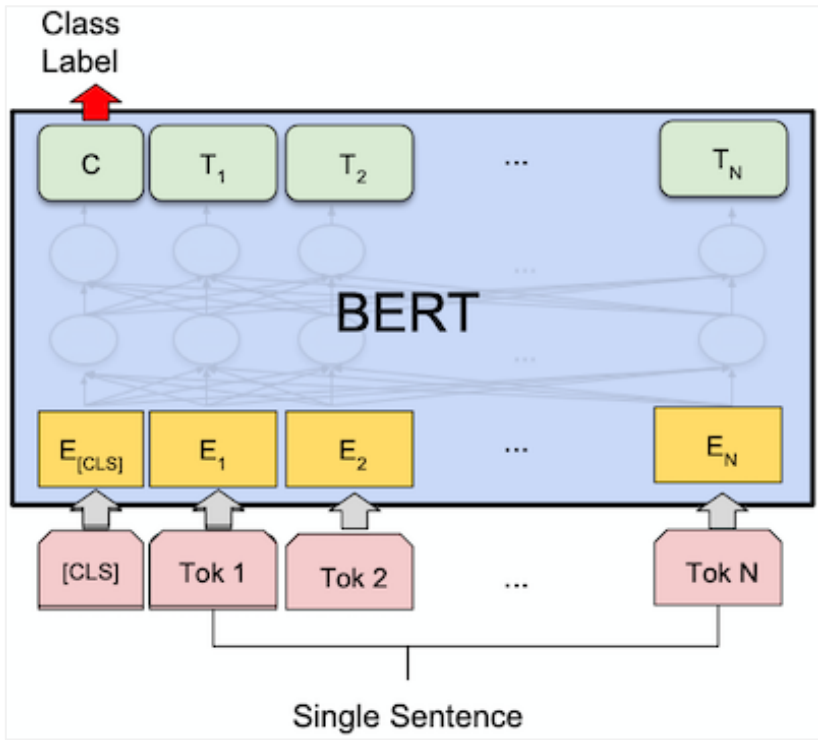
✓ 한국어 특성 반영 한계

: 영어 기반으로 학습된 모델
→ 때문에 한국어 정확도 ↓

3-1 BERT



S | W



BERT



이제 BERT를 까먹지 않겠지?

✓ Fast-Train

: Pre-train 모델이기 때문에
Fine-tuning만 하면 됨!

✓ Small Data

: 적은 데이터셋으로 좋은 성능을
낼 수 있음

✓ 한국어 특성 반영 한계

: 영어 기반으로 학습된 모델
→ 때문에 한국어 정확도 ↓



KoBERT

✓ 그러면 한국어 학습 모델 쓰자!

3-2 KoBERT



SKTBrain/KoBERT

Korean BERT pre-trained cased (KoBERT)



5

Contributors

3

Issues

627

Stars

159

Forks



KoBERT

S

✓ Korean-focused Model

: 한국어 위키의 500만 문장과 뉴스 2,000만 문장을 pre-train한 한국어 전용 BERT 모델

✓ Multi-sentiment Classification

: 긍정/부정을 포함한 다중분류가 가능한 모델

DATA Sets



단발성 대화 DataSet

각 문장에 대해 7개 감정(기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립)으로 Labeling된 데이터 셋

NAVER 영화 리뷰 DataSet

한국어 자연어 처리에 많이 이용되는 감성어 corpus로 긍/부정 Labeling의 정확도가 높은 리뷰 단위 데이터 셋

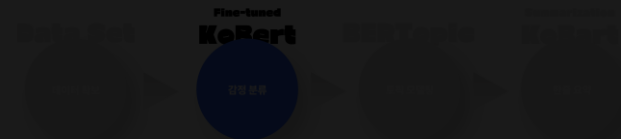
KNU(군산대학교) 한국어 감성사전

표준어국어대사전을 구성하는 형용사, 부사, 동사, 명사의 모든 뜻풀이에 대한 긍정, 중립, 부정 Labeling이 된 데이터 셋

연어툰 Labeling DataSet

팀 내부에서 긍정/부정/개선/일반문장으로 Labeling한 플레이스토어 수면 어플 리뷰 데이터 셋

DATA Sets



단발성 대화 DataSet

각 문장에 대해 7개 감정(기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립)으로 Labeling된 데이터 셋

NAVER 영화 리뷰 DataSet

한국어 자연어 처리에 많이 이용되는 감성어 corpus로 긍/부정 Labeling의 정확도가 높은 리뷰 단위 데이터 셋

KNU(건국대학교) 한국어 감성사전

표준어국어대사전을 구성하는 형용사, 부사, 동사, 명사의 모든 뜻풀이에 대한 긍정, 중립, 부정 Labeling이 된 데이터 셋

연어툰 Labeling DataSet

팀 내부에서 긍정/부정/개선/일반문장으로 Labeling한 플레이스토어 수면 어플 리뷰 데이터 셋

3-3 4가지 감정으로 분류



첨엔 접속이 잘 되더니 요즘 들어 부쩍 접속도 잘 안되고 눌러도 들어가지지 않을 때가 많아요

부정



정말 지치고 힘든 사람 들이라면 지출이 된다 해도 분명 불만 없이 결제를 할 것이라고 압니다

긍정



제가 휴대폰이 없어서 공기계로 듣는데 와이파이 없는 밖에서도 명상음을 듣고 싶어요

개선



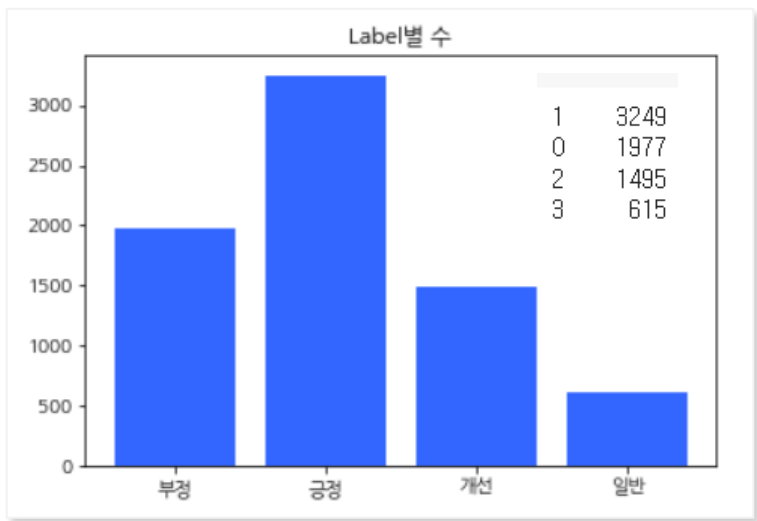
항상 4시 후에 왔던 한 초등학생이에요

일반



우리의 데이터를 4가지 감정으로 분류해보자!

3-3 4가지 감정으로 분류



Our Data

Fine-tuned
KoBERT

Our Model

부정



긍정



개선

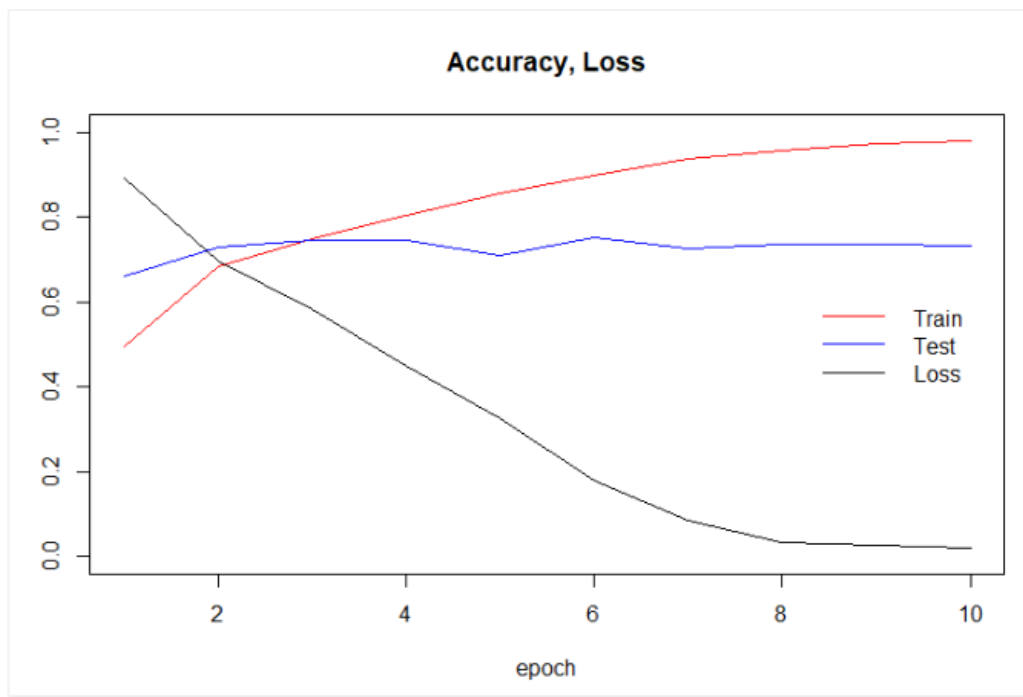


일반



우리의 데이터를 4가지 감정으로 분류해보자!

3-3 4가지 감정으로 분류 - 결과



Accuracy

Train : 98%

Test : 73%

Why..?

3-3 4가지 감정으로 분류- 결과



개선?: 완성도도 좋고 내용도 좋은데 1년 사용료가 너무 비싸다

Accuracy

Train : 98%

Test : 73%

‘개선’ 라벨링 기준 모호

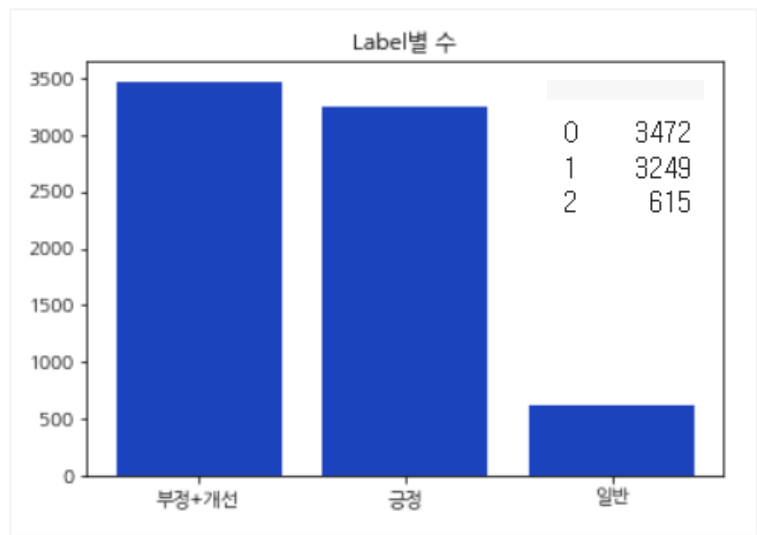
Overfitting 의심

부정?: 59000원 저도 모르게 자동 결제 돼버렸네요 바로 환불해주세요



부정, 개선 감정 합쳐보자

3-4 3가지 감정으로 분류



부정+개선 통합 데이터

KoBERT
Ver.1

Max Length: 40
Epoch: 10

KoBERT
Ver.2

Max Length: 64
Epoch: 5

Our Model

**부정
+
개선**



긍정

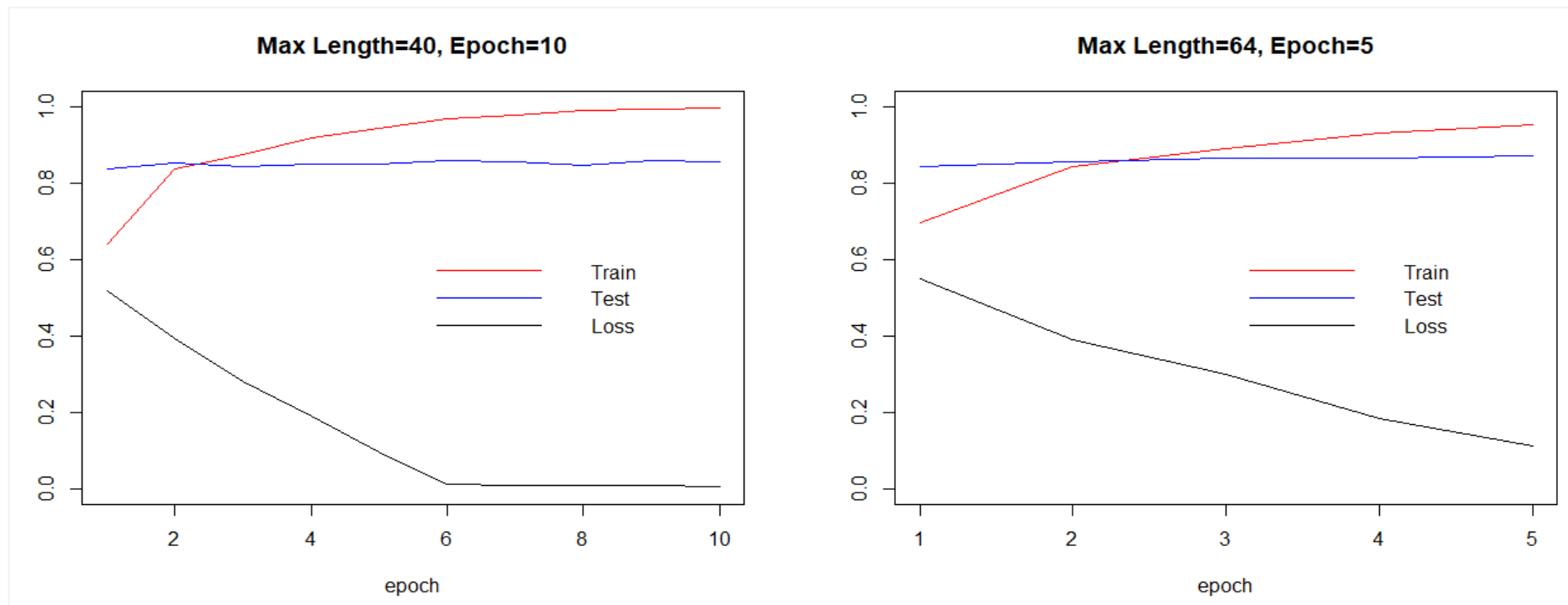


일반



우리의 데이터를 3가지 감정으로 분류해보자! (w/ Hyper Parameter 조정)

3-4 3가지 감정으로 분류- 결과



Accuracy

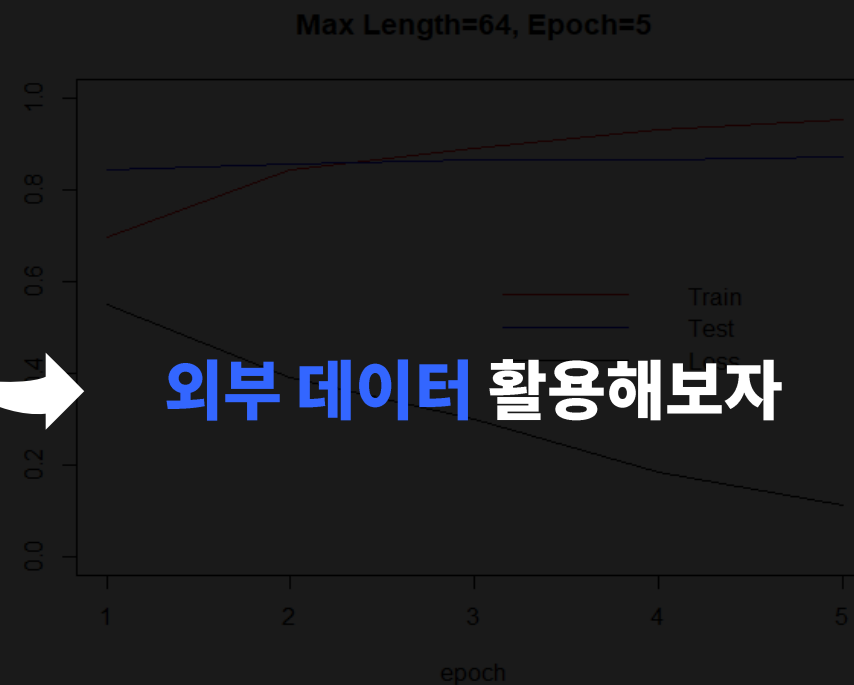
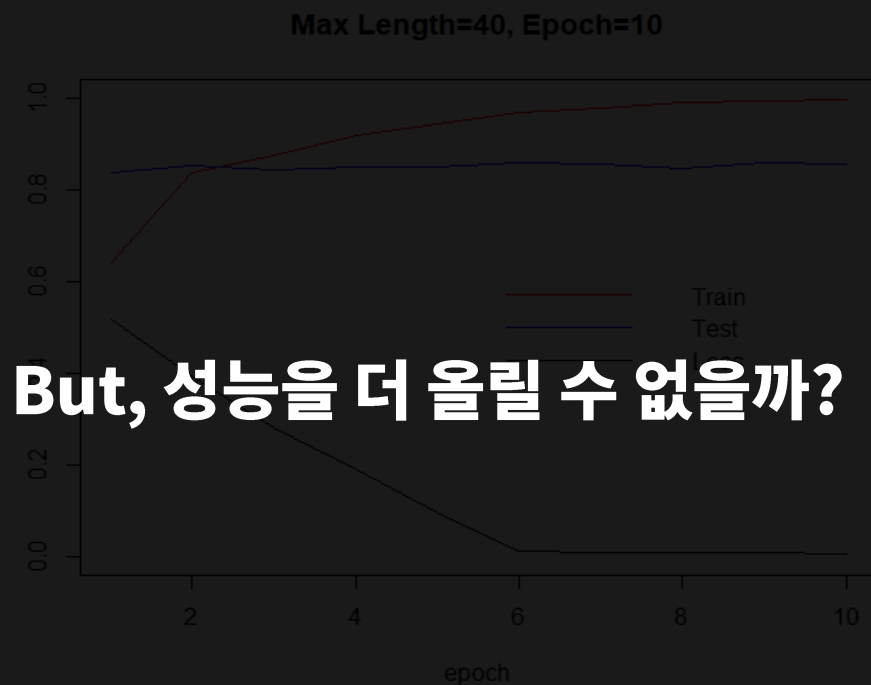
Train : 99%

Test : 85%

Train : 95%

Test : 87%

3-4 3가지 감정으로 분류- 결과



Accuracy

Train : 99%

Test : 85%

Train : 95%

Test : 87%

3-5 외부 데이터 활용 (네이버 영화 리뷰)



	Review	Label
0	굳	1
2	뭐야 이 평점들은 나쁘진 않지만 점 짜리는 더더욱 아니잖아	0
3	지루하지는 않은데 완전 막장임 돈주고 보기에	0
4	만 아니었어도 별 다섯 개 줬을텐데 왜 로 나와서 제 심기를 불편하게 하죠	0
5	음악이 주가 된 최고의 음악영화	1
...
199987	인간이 문제지 소는 뭘죄인가	0
199988	평점이 너무 낮아서	1
199989	이게 뭐요 한국인은 거들먹거리고 필리핀 혼혈은 착하다	0
199990	청춘 영화의 최고봉방황과 우울했던 날들의 자화상	1
199991	한국 영화 최초로 수간하는 내용이 담긴 영화	0

198436 rows × 2 columns

Naver
영화 리뷰 라벨링 데이터

3-5 외부 데이터 활용 (네이버 영화 리뷰)



	Review	Label
0	굳	1
2	뭐야 이 평점들은 나쁘진 않지만 점 따리는 더더욱 아니잖아	0
3	지루하지는 않은데 완전 막장임 돈주고 보기에	0
4	만 아니었어도 별 다섯 개 줬을텐데 왜 로 나와서 제 심기를 불편하게 하죠	0
5	음악이 주가 된 최고의 음악영화	1
...
199987	인간이 문제지 소는 뭔진인가	0
199988	평점이 너무 낮아서	1
199989	이게 뭐요 한국인은 거들먹거리고 필리핀 혼혈은 착하다	0
199990	청춘 영화의 최고봉방황과 우울했던 날들의 자화상	1
199991	한국 영화 최초로 수간하는 내용이 담긴 영화	0

198436 rows × 2 columns

약 20만 개의 데이터 → 7000개 Sampling



	Review	Label
0	매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 놔둘 때 딱...	1
2	그리고 가끔 아침에 일어났을 때 머리가 뭉통 한 데 그 땐 아침밥을 먹거나 커피 ...	1
3	단점 가끔 휴대폰 소프트웨어가 밤에 자기 마음대로 업데이트 되는데 그때 폰이 꺼졌...	0
7	뭐 부족한 부분이 있다는 후기가 있어서 한번 경험하고 결제 진행을 선택 하긴 어렵네요	0
8	개인적인 문 제일 수도 있지만 미라클 나 잇을 사용 할 때마다 휴대폰 진동에 머리...	0

긍, 부정 리뷰만 남긴 우리의 데이터

KoBERT

Max Length: 64
Epoch: 10

부정



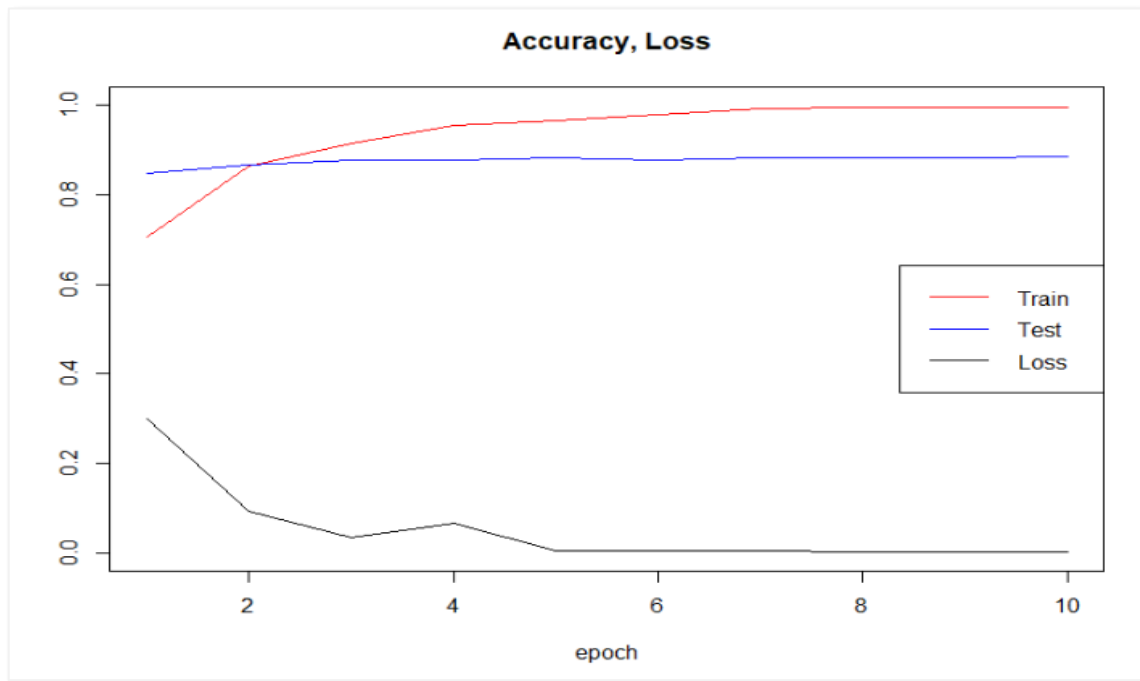
긍정



총 13,721개의 데이터 (부정 6,961, 긍정 6,760)

긍정, 부정의 두 가지 감정으로 분류

3-5 외부 데이터 활용 - 결과



Accuracy

Train : 99%

Test : 89%

성능 소폭 향상

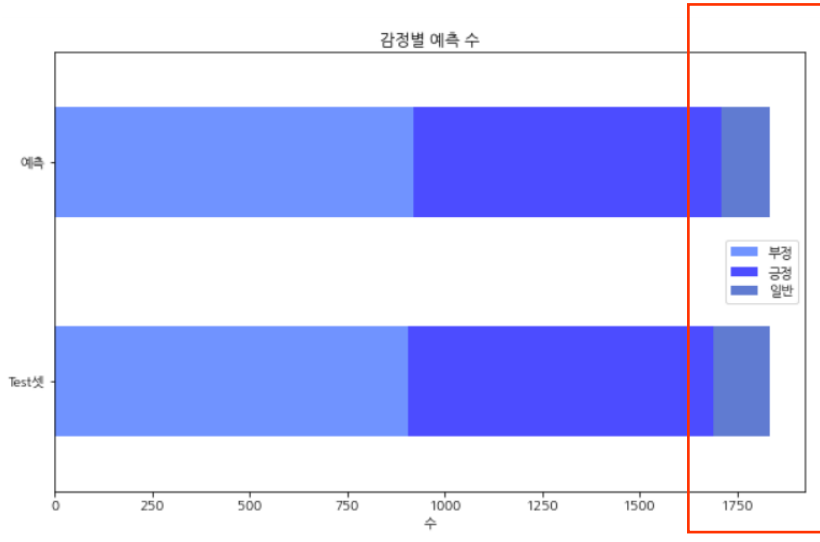


“일반 문장 예측 성능을 더 올릴 수 있지 않을까?”

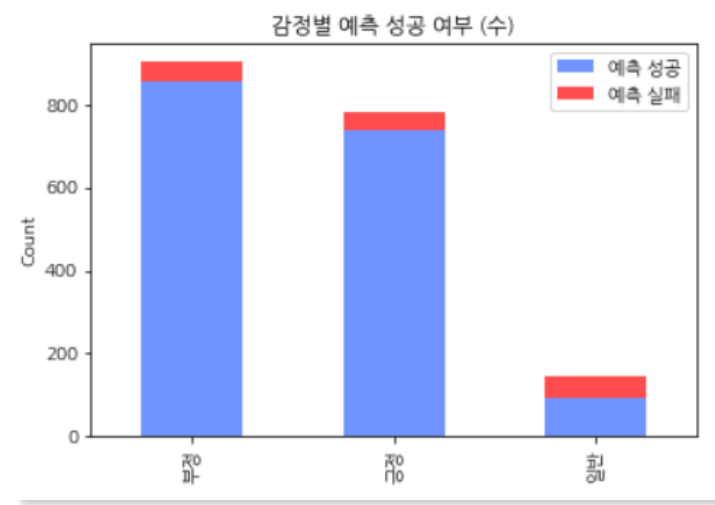
3-6 감정별 예측 정확도 분석



	부정	긍정	일반
Test셋	906	783	145
예측	919	791	124

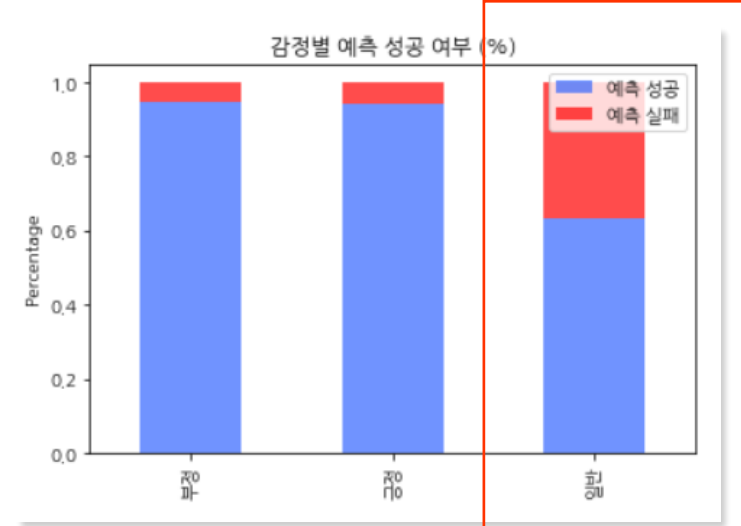


Test 데이터 내 감정별 개수 및 예측한 감정별 개수



	예측 성공	예측 실패
부정	858	48
긍정	738	45
일반	92	53

감정별 예측 개수



	예측 성공	예측 실패
부정	0.947020	0.052980
긍정	0.942529	0.057471
일반	0.634483	0.365517

감정별 예측 비율

리뷰 데이터 증강

4. Augmentation

4-1 EDA (Easy Data Augmentation)

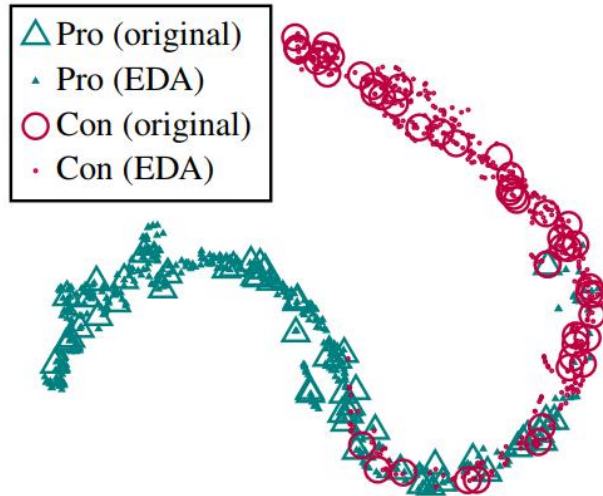


Figure 2: Latent space visualization of original and augmented sentences in the Pro-Con dataset.

Operation	Sentence
None	A sad, superior human comedy played out on the back roads of life.
SR	A <i>lamentable</i> , superior human comedy played out on the <i>backward</i> road of life.
RI	A sad, superior human comedy played out on <i>funniness</i> the back roads of life.
RS	A sad, superior human comedy played out on <i>roads</i> back <i>the</i> of life.
RD	A sad, superior human out on the roads of life.

Table 1: Sentences generated using EDA. SR: synonym replacement. RI: random insertion. RS: random swap. RD: random deletion.

Original Sentence

저기 맛있게 생긴 빨간 사과가 책상 위에 있다.

SR (Synonym Replacement)

저기 맛있게 생긴 **붉으스름한** 사과가 책상 위에 있다.

RI (Random Insertion)

저기 맛있게 생긴 **파란** 빨간 사과가 책상 위에 있다.

RD (Random Deletion)

저기 맛있게 “ 빨간 사과가 책상 위에 있다.

RS (Random Swap)

저기 **생긴 맛있게** 빨간 사과가 책상 위에 있다.

4-2 EDA (Easy Data Augmentation)



```
[ ] EDA(review['Review'][0],  
      alpha_sr=0.5,  
      alpha_ri=0.2,  
      alpha_rs=0.2,  
      p_rd=0.3,  
      num_aug=9)
```

'매번 9시간 넘게 좋아하십니다 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 자고 4시간 딱 반 좋았습니다 알람을 맞추지 않아도 제때 일어날 수 잘 때 면서 있어서 않아도 깨우지 일어난다며 참 자거나',
'매번 9시간 넘게 잠들지 못하고 깨는데 거 딱 반 알람을 않아도 제때 있어서 좋았습니다 면서 않아도 일어난다며 좋아하십니다',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 시 딱 4시간 반 자 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'매번 때 넘게 일어난다며 깊게 잠들지 깨는데 면서 못하고 이 거 켜 났을 9시간 딱 4시간 반 자고 맞추지 알람을 않아도 제때 일어날 수 있어서 좋았습니다 자거나 깨우지 않아도 잘 수시로 참 좋아하십니다',
'매번 9시간 일어난다며 자거나 깊게 잠들지 좋아하십니다 수시로 못하고 이 거 켜 일어날 때 딱 4시간 반 자고 알람을 맞추지 않아도 제때 수 났을 면서 좋았습니다 있어서 깨우지 않아도 잘 넘게 참 깨는데',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 자 거 켜 났을 때 딱 자 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 수 있어서 좋았습니다 자 면서 깨우지 않아도 잘 자 자 일어난다며 참 좋아하십니다',
'매번 9시간 넘게 이 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 이 켜 났을 이 때 이 딱 이 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'9시간 깊게 잠들지 수시로 깨는데 이 났을 때 딱 않아도 제때 수 좋았습니다 깨우지 않아도 잘 일어난다며',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 형 딱 4시간 반 고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 때 딱 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다']

✓ KorEDA / TextAttack 라이브러리를 통해 구현

: SR/RI/RD/RS를 Stochastic하게 적용하여 Text 데이터 증강

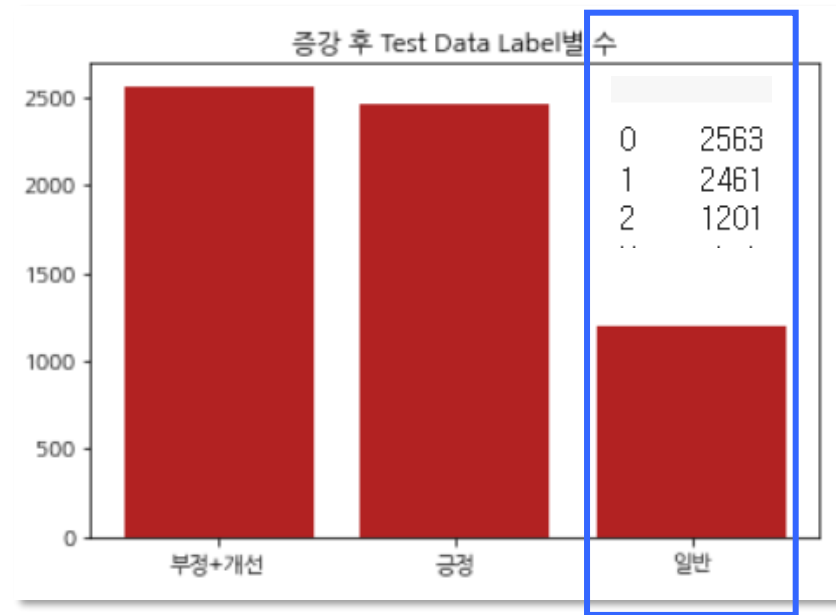
4-2 EDA 적용



Train Data



Augmentation

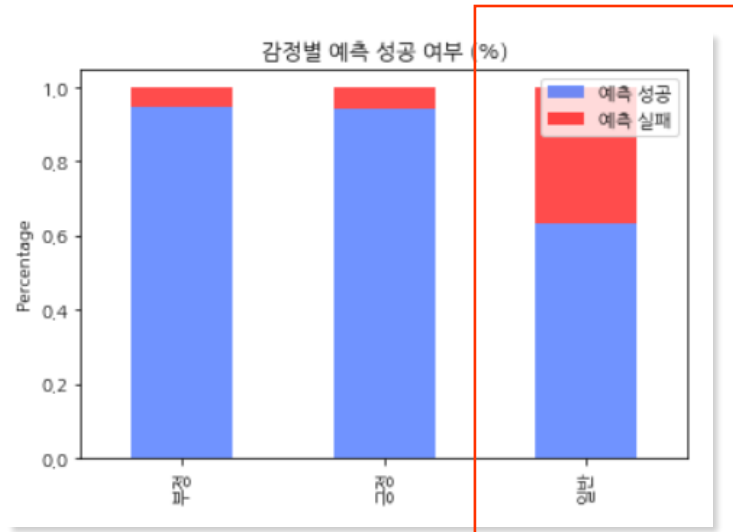


증강 결과 1201개의 일반 문장 확보

4-3 증강 전후 결과 비교



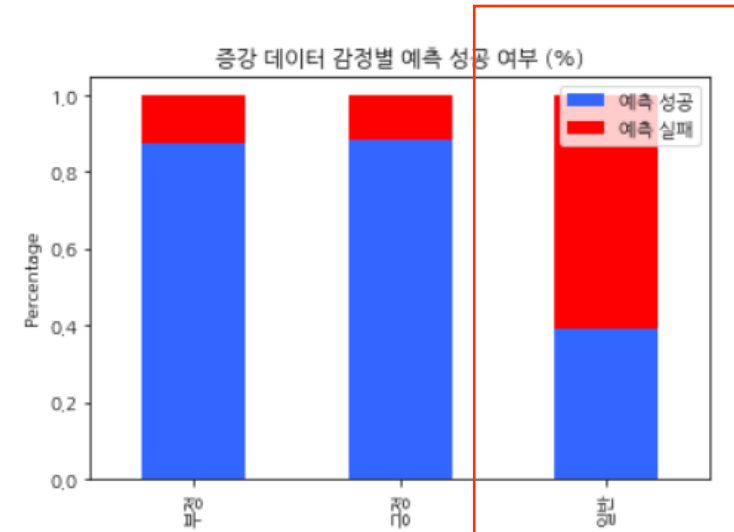
Data 증강 전



	예측 성공	예측 실패
부정	0.947020	0.052980
긍정	0.942529	0.057471
일반	0.634483	0.365517

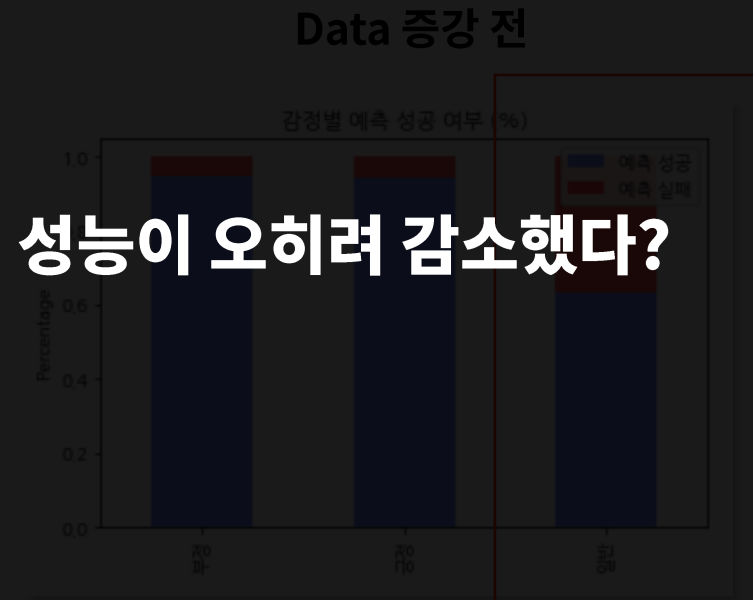
VS

Data 증강 후



	예측 성공	예측 실패
부정	0.873068	0.126932
긍정	0.885057	0.114943
일반	0.393103	0.606897

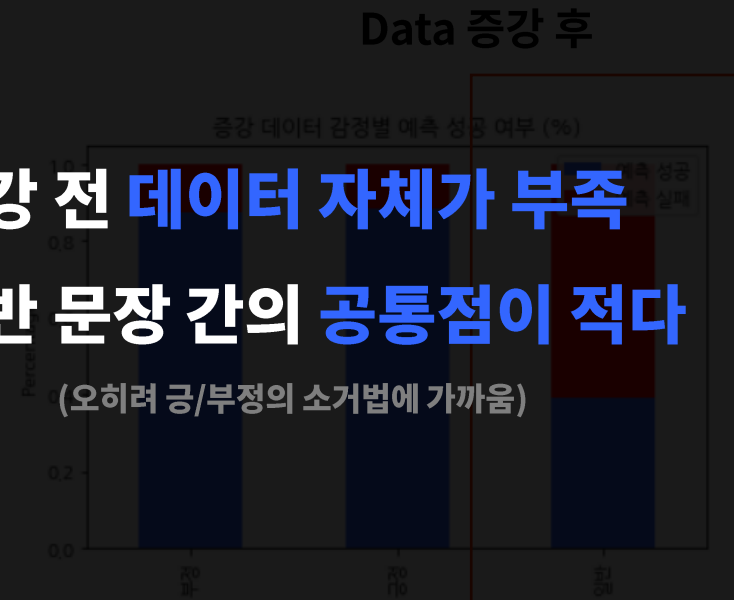
4-3 증강 모델 감정별 예측 정확도 분석



	예측 성공	예측 실패
부정	0.947020	0.052980
긍정	0.942529	0.057471
일반	0.634483	0.365517



1. 증강 전 데이터 자체가 부족
 2. 일반 문장 간의 공통점이 적다
- (오히려 긍/부정의 소거법에 가까움)



	예측 성공	예측 실패
부정	0.873068	0.126932
긍정	0.885057	0.114943
일반	0.393103	0.606897

4-4 EDA (Easy Data Augmentation) 돌아보기



```
[ ] EDA(review['Review'][0],  
      alpha_sr=0.5,  
      alpha_ri=0.2,  
      alpha_rs=0.2,  
      p_rd=0.3,  
      num_aug=9)
```

'매번 9시간 넘게 좋아하십니다 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 자고 4시간 딱 반 좋았습니다 알람을 맞추지 않아도 제때 일어날 수 잘 때 면서 있어서 않아도 깨우지 일어난다며 참 자거나',
'매번 9시간 넘게 잠들지 못하고 깨는데 거 딱 반 알람을 알아도 제때 있어서 좋았습니다 면서 않아도 일어난다며 좋아하십니다',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 시 딱 4시간 반 자 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'매번 때 넘게 일어난다며 깊게 잠들지 깨는데 면서 못하고 이 거 켜 났을 9시간 딱 4시간 반 자고 맞추지 알람을 알아도 제때 일어날 수 있어서 좋았습니다 자거나 깨우지 않아도 잘 수시로 참 좋아하십니다',
'매번 9시간 일어난다며 자거나 깊게 잠들지 좋아하십니다 수시로 못하고 이 거 켜 일어날 때 딱 4시간 반 자고 알람을 맞추지 않아도 제때 수 났을 면서 좋았습니다 있어서 깨우지 않아도 잘 넘게 참 깨는데',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 자 거 켜 났을 때 딱 자 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 수 있어서 좋았습니다 자 면서 깨우지 않아도 잘 자 자 일어난다며 참 좋아하십니다',
'매번 9시간 넘게 이 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 이 켜 났을 이 때 이 딱 이 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'9시간 깊게 잠들지 수시로 깨는데 이 났을 때 딱 알아도 제때 수 좋았습니다 깨우지 않아도 잘 일어난다며',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 형 딱 4시간 반 고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다',
'매번 9시간 넘게 자거나 깊게 잠들지 못하고 수시로 깨는데 이 거 켜 났을 때 딱 4시간 반 자고 알람을 맞추지 않아도 제때 일어날 수 있어서 좋았습니다 면서 깨우지 않아도 잘 일어난다며 참 좋아하십니다']

✓ **문장의 변화가 적음**

✓ **유의어 사전이 부족함**

✓ **데이터가 적음**

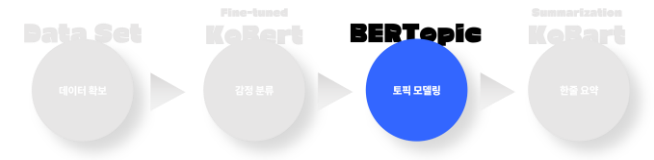
고려한 해결법

1. word2vec similarity 기반으로 유의어 사전을 만들어 보자!
2. word2vec → vector SMOTE/ADASYN(증강) → vec2word

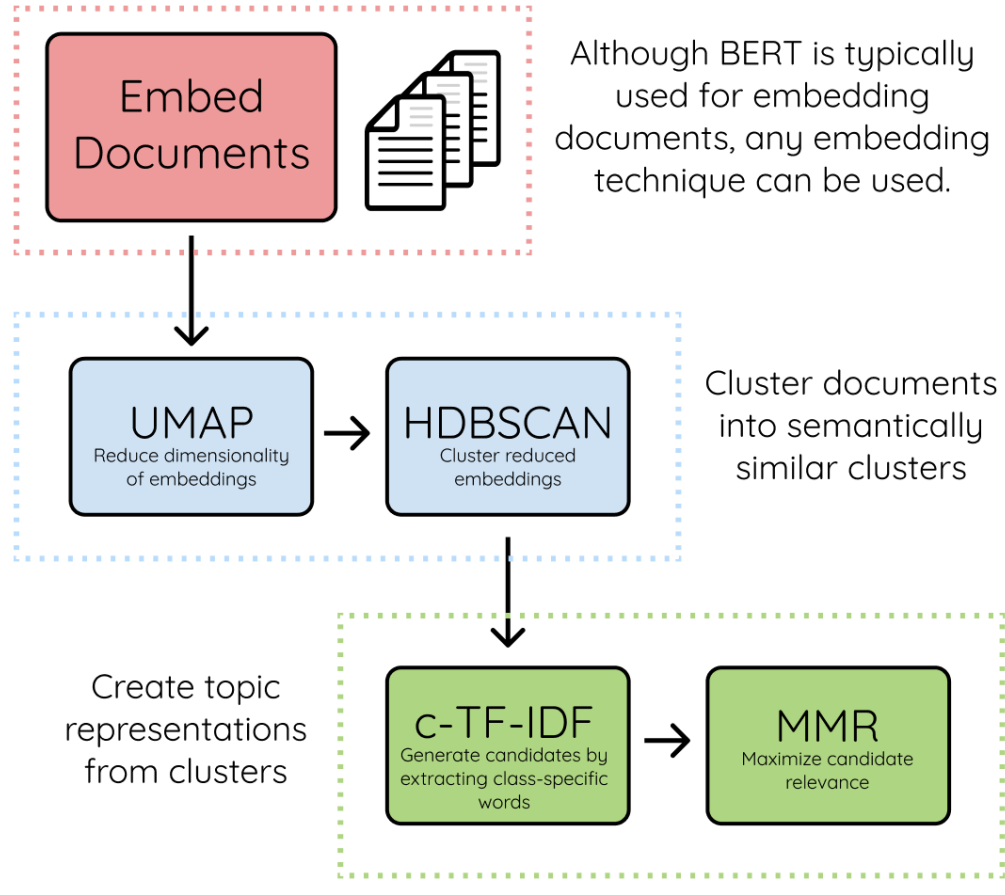
토픽 모델링

5. Topics

5-1 BERTopic 사용 이유



Topic Modeling의 수많은 기법 중에 왜 BERTopic?



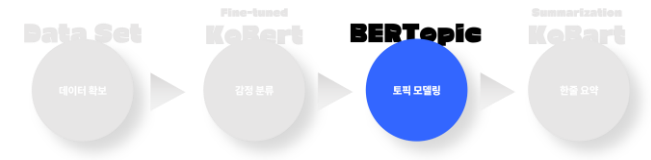
✓ BERTopic 이란?

1. Topic Modeling 기법
2. BERT 기반의 Embedding + Class-based TF-IDF 사용

✓ BERTopic 구조

1. Bert를 이용하여 Document Embedding
2. UMAP을 이용하여 Document Vector 차원 축소
3. HDBSCAN을 이용하여 Clustering (유사 Document 묶음)
4. C-TF-IDF를 통해서 묶여진 그룹의 Topic을 표현하는 단어 탐색
5. Maximize Candidate Relevance Algorithm을 이용해서 Topic을 설명하는 Keyword들이 다양하게 선별되도록 Keyword List를 Return한다.
6. 유사도 기준으로는 Cosine Similarity를 사용한다.

5-2 감정분 문장 전체기



#전처리 함수

```

def to_words(corpus):
    korean_only = re.sub('[가-힣]', '', corpus)
    words = korean_only.split()
    stops = set(stopwords_list)
    meaningful_words = [w for w in words if not w in stops]
    mystring = ''.join(map(str, meaningful_words))
    nouns = okt.nouns(mystring)
    nounstring = ','.join(map(str, nouns))
    return nounstring
  
```



```

1      처음, 휴대폰, 소리, 적응, 머리, 이틀, 적응, 아집
4      앱, 표시, 광고, 대한, 평가, 란, 신경안, 때, 만사, 앱, 광고, 평가, 란, 보고, 처음
5      위치, 번, 점수, 기록
6      꿈결, 집, 진동, 알람, 땀, 앱, 소리, 뭐, 묘아, 직, 베타, 단계
11     오류, 버그, 중지, 이용권, 다시, 드릴, 필요, 이용자, 입장, 버그, 문제, 이용권, 문제, 월간, 플...
      ...
7203      기와, 달, 무료, 체험, 후연
7206      거미, 리, 결제, 체험, 시작, 결제, 임
7229      전자파, 몸, 머리, 직빵, 생각
7255      무료, 체험, 안, 건너, 기하, 알람, 기능, 사용
7314      구매, 프로모션, 코드, 프리미엄, 등록
  
```

Noun만 추출한 Topic

1. 한국어가 아닌 단어 제거
2. Stopwords 제거
3. Okt를 이용하여 Noun만 남기기
4. 전체 pos 활용 Topic과 비교

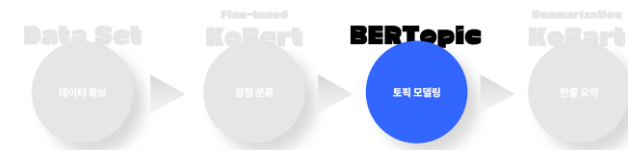


```

-1_너무_자동결제_결제_계속
0_소리가_소리_소리를_음악이
1_결제가_환불_환불해주세요_refund
2_좋아요_좋다_so_잡생각음이
3_마음이_명상이_마보를_있습니다
4_ㅋㅋ_ㅋㅋ_실행이_볼륨이랑
5_잘쓰고있어요_사랑_좋네요_잘자고있어요
6_안돼요_안되네요_끊어지네요_줄수
7_좋은_app_앱을_앱입니다
8_재올때_아기_좋아요_좋은거같아요
9_앱입니다_앱이에요_이용하고있습니다_명상앱중
  
```

모든 POS를 활용한 Topic

5-3 BERTopic 돌아보기



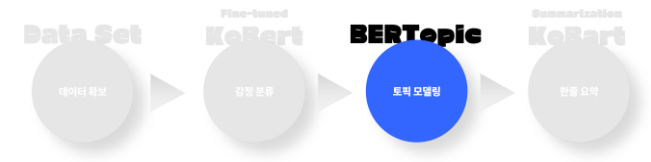
1. Topic 추출 방식 시각화

	Topic	Count	Name
0	-1	1782	-1_결제_취소_유료_명상
1	0	274	0_무료_체험_결제_기간
2	1	189	1_설치_삭제_사용_유료
3	2	158	2_가입_
4	3	136	3_환불_결제_요청_신청
5	4	127	4_자동_결제_취소_최악
6	5	96	5_안드로이드_핸드폰_갤럭시_휴대폰
7	6	70	6_음악_배경음악_소리_유투브
8	7	69	7_내야_내지_돈돈_비용
9	8	61	8_회원_탈퇴_가입_요청
10	9	61	9_탈퇴_처리_방법_자동
11	10	61	10_구독_취소_결제_신청
12	11	57	11_녹음_소리_코고_파일
13	12	47	12_정기_방법_결제_취소
14	13	45	13_계정_삭제_탈퇴_해주시
15	14	43	14_버튼_탈퇴_계정_가기
16	15	42	15_이용권_무료_사용_달이
17	16	41	16_기능_측정_호흡_코골
18	17	40	17_진짜_로딩_건가_가요
19	18	37	18_네이버_어디가_프리미엄_보이지
20	19	36	19_사용_유료_사용자_후주

Topic Word Scores



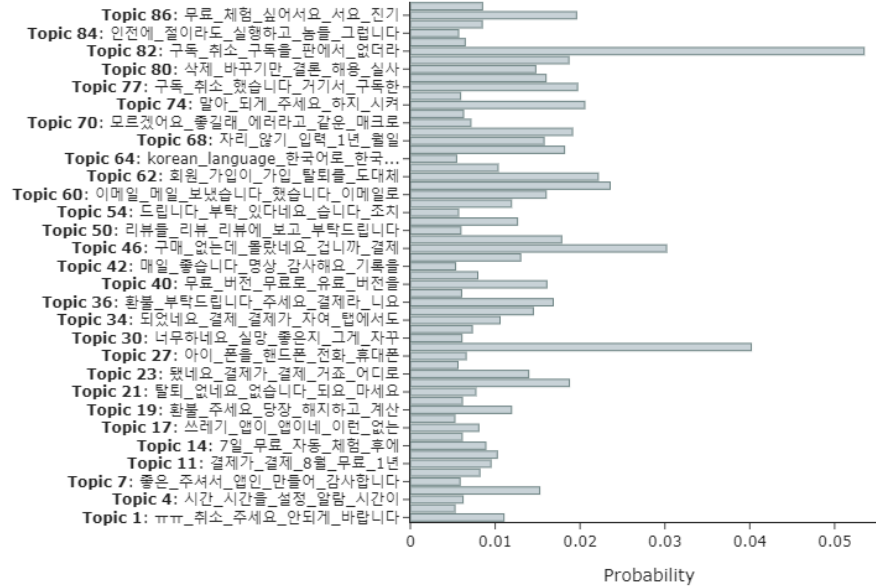
5-3 BERTopic 돌아보기



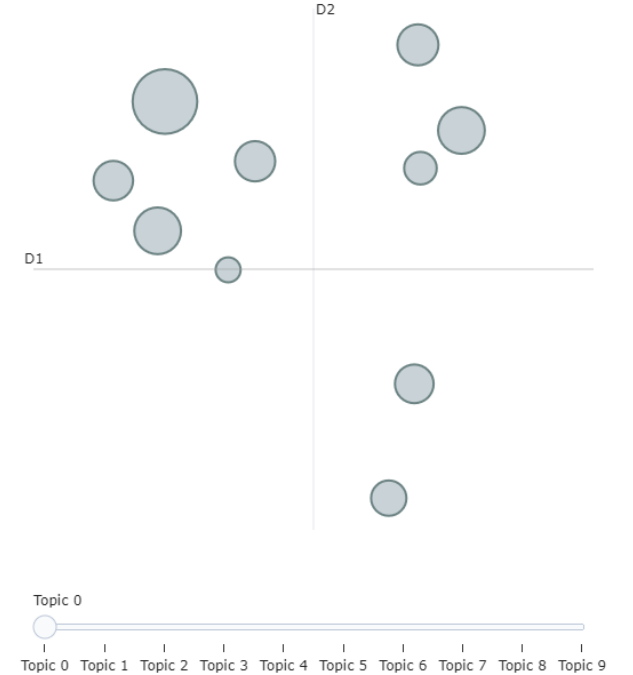
2. Topic별 분포 시각화

	Topic	Count	Name
0	-1	1782	-1_결제_취소_유료_명상
1	0	274	0_무료_체험_결제_기간
2	1	189	1_설치_삭제_사용_유료
3	2	158	2_가입_
4	3	136	3_환불_결제_요청_신청
5	4	127	4_자동_결제_취소_최악
6	5	96	5_안드로이드_핸드폰_갤럭시_휴대폰
7	6	70	6_음악_배경음악_소리_유투브
8	7	69	7_내야_내지_돈돈_비용
9	8	61	8_회원_탈퇴_가입_요청
10	9	61	9_탈퇴_처리_방법_자동
11	10	61	10_구독_취소_결제_신청
12	11	57	11_녹음_소리_코고_파일
13	12	47	12_정기_방법_결제_취소
14	13	45	13_계정_삭제_탈퇴_해주시
15	14	43	14_버튼_탈퇴_계정_가기
16	15	42	15_이용권_무료_사용_달이
17	16	41	16_기능_측정_호흡_코골
18	17	40	17_진짜_로딩_건가_가요
19	18	37	18_네이버_어디가_프리미엄_보이지
20	19	36	19_사용_유료_사용자_후주

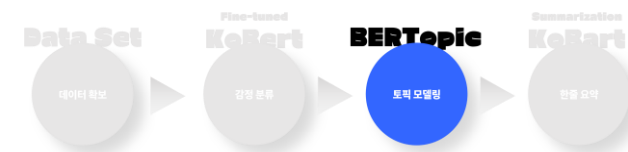
Topic Probability Distribution



Intertopic Distance Map



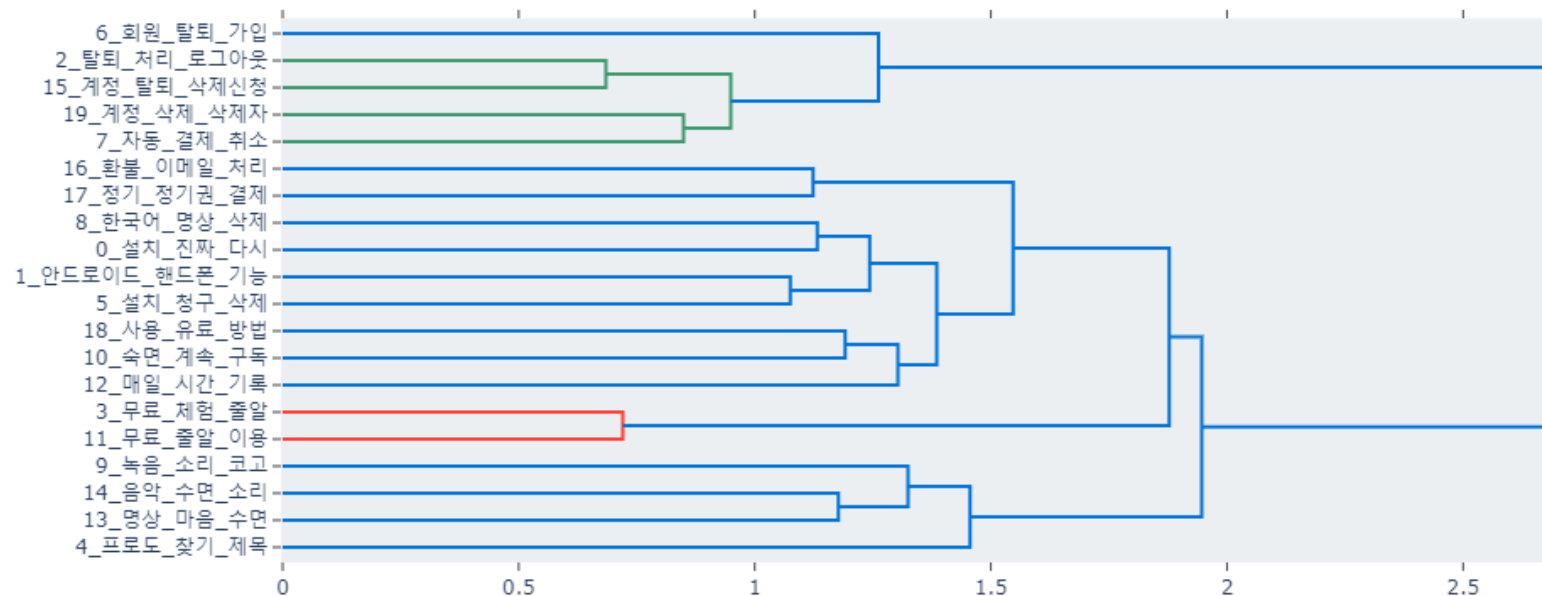
5-3 BERTopic 돌아보기



3. Topic 개수 Cutting

	Topic	Count	Name
0	-1	1782	-1_결제_취소_유료_명상
1	0	274	0_무료_체험_결제_기간
2	1	189	1_설치_삭제_사용_유료
3	2	158	2_가입_
4	3	136	3_환불_결제_요청_신청
5	4	127	4_자동_결제_취소_최악
6	5	96	5_안드로이드_핸드폰_갤럭시_휴대폰
7	6	70	6_음악_배경음악_소리_유투브
8	7	69	7_내야_내지_돈돈_비용
9	8	61	8_회원_탈퇴_가입_요청
10	9	61	9_탈퇴_처리_방법_자동
11	10	61	10_구독_취소_결제_신청
12	11	57	11_녹음_소리_코고_파일
13	12	47	12_정기_방법_결제_취소
14	13	45	13_계정_삭제_탈퇴_해주시
15	14	43	14_버튼_탈퇴_계정_가기
16	15	42	15_이용권_무료_사용_달이
17	16	41	16_기능_측정_호흡_코골
18	17	40	17_진짜_로딩_건가_가요
19	18	37	18_네이버_어디가_프리미엄_보이지
20	19	36	19_사용_유료_사용자_후주

Hierarchical Clustering



Positive 2.1만개

	Topic	Count	Name
0	-1	1776	-1_소리_마음_명상_집중
1	0	303	0_광고_최고_정말_사용
2	1	130	1_무료_결제_유료_버전
3	2	100	2_소리_녹음_여러가지_파도
4	3	99	3_잠도_수면_꿀잠_평소
5	4	73	4_
6	5	72	5_도움_마음_안정_프로
7	6	67	6_음악_자연_기분_기능
8	7	65	7_아침_시작_하루_알람
9	8	63	8_아기_아들_평소_금방
10	9	58	9_제거_모습_레전드_수치화
11	10	53	10_마음_안정_명상_심리
12	11	49	11_원래_효과_정말_한국어
13	12	47	12_사용_안드로이드_핸드폰_다가
14	13	46	13_매일_하루_하루하루_오늘
15	14	45	14_도움_수면_불면증_향상
16	15	45	15_코끼리_명상_마음_하루
17	16	41	16_수면_상태_불면증_확인가능
18	17	39	17_불면증_요즘_캐치_수면장애
19	18	39	18_시간_타이머_개월_지금
20	19	39	19_가격_구매_서비스_결제

- ✓ 소리 만족
- ✓ 마음이 안정되는 효과
- ✓ 아기가 금방 잠듦
- ✓ 불면증 개선에 도움
- ✓ 시간 타이머 기능 좋음
- ✓ 가격 합리적

Negative 1.2만개

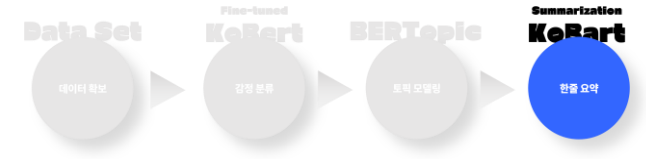
	Topic	Count	Name
0	-1	1782	-1_결제_취소_유료_명상
1	0	274	0_무료_체험_결제_기간
2	1	189	1_설치_삭제_사용_유료
3	2	158	2_가입_
4	3	136	3_환불_결제_요청_신청
5	4	127	4_자동_결제_취소_최악
6	5	96	5_안드로이드_핸드폰_갤럭시_휴대폰
7	6	70	6_음악_배경음악_소리_유튜브
8	7	69	7_내야_내지_돈돈_비용
9	8	61	8_회원_탈퇴_가입_요청
10	9	61	9_탈퇴_처리_방법_자동
11	10	61	10_구독_취소_결제_신청
12	11	57	11_녹음_소리_코고_파일
13	12	47	12_정기_방법_결제_취소
14	13	45	13_계정_삭제_탈퇴_해주시
15	14	43	14_버튼_탈퇴_계정_가기
16	15	42	15_이용권_무료_사용_달이
17	16	41	16_기능_측정_호흡_코골
18	17	40	17_진짜_로딩_건가_가요
19	18	37	18_네이버_어디가_프리미엄_보이지
20	19	36	19_사용_유료_사용자_후주

- ✓ 환불 요청
- ✓ 자동 결제 취소
- ✓ 음악 소리 관련 불만
- ✓ 긴 로딩 문제

리뷰 한 줄 요약

6. Summary

6-1 Text Summarization



✓ **F(text) = Comprehensive Information**

✓ **Extractive Summarization**

1. 문장 단위로 중요도를 Scoring
2. Score를 기반으로 문장을 선택하여 조합하여 요약
≒ N번째 문장 추출

✓ **Abstractive Summarization**

1. NLG (Natural Language Generation)
2. 문장 내 토큰을 추출하여 새로운 요약 문장을 생성하는 것

✓ **BART**

1. 입력 텍스트 일부에 노이즈를 추가
2. 이를 다시 원문으로 복구하는 AutoEncoder 형태로 학습



6-2 BART



✓ BART란?

1. Bidirectional Auto-Regressive Transformer
2. 인코더/디코더를 모두 학습시킨 모델
3. 인코더/디코더 중 일부만 학습하는 BERT/GPT와 구별됨

✓ Token Masking

1. 랜덤한 토큰을 [MASK]토큰으로 치환

✓ Token Deletion

1. 랜덤한 토큰을 제거

✓ Text Infilling

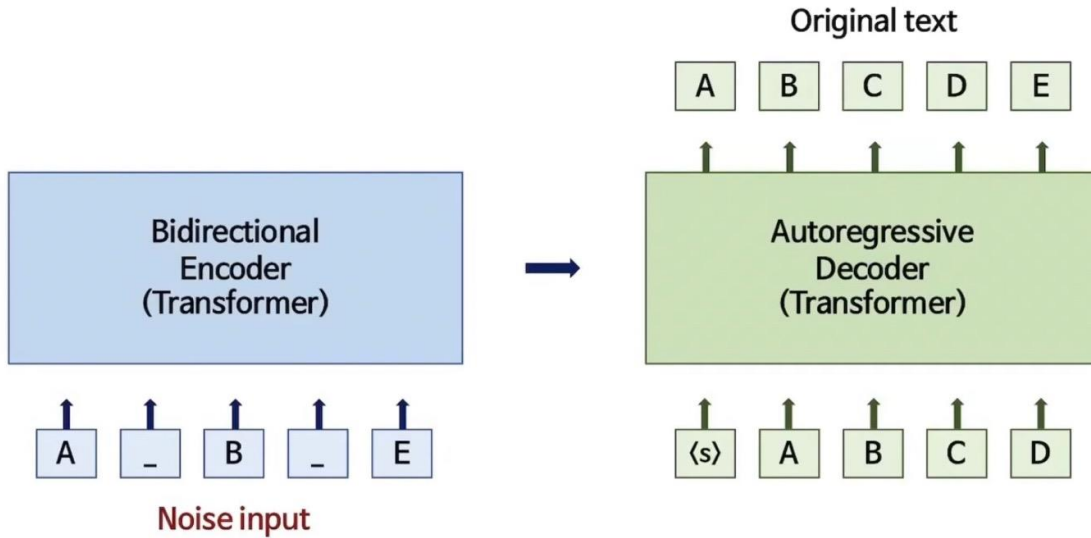
1. 랜덤한 Text Span을 [MASK] Span으로 치환

✓ Sentence Permutation

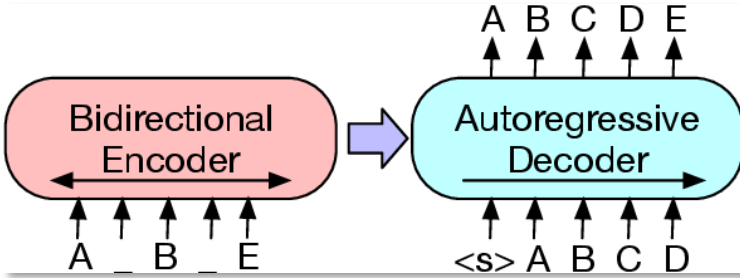
1. 문장별로 분리한 후 랜덤으로 순서를 Shuffle

✓ Document Rotation

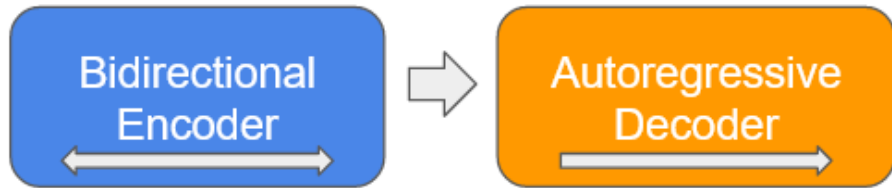
1. 랜덤하게 선택된 토큰을 기준으로 문서를 시작함 (Rotating)



6-3 BART



가 나 다. 라 마.



가 _ 다. _ 마.

<s>가 나 다. 라 마.

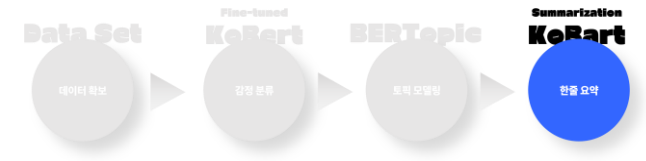


KoBART

✓ KoBART란?

1. 영어로 학습된 BART모델을 한국어에서도 잘 작동할 수 있게 한국어 Corpus를 기반으로 학습시킨 모델

6-4 키워드별 리뷰 분류



1. 토픽별로 문서 할당

	Topic	Count	Name
0	-1	1782	-1_결제_취소_유료_명상
1	0	274	0_무료_체험_결제_기간
2	1	189	1_설치_삭제_사용_유료
3	2	158	2_가입_
4	3	136	3_환불_결제_요청_신청
5	4	127	4_자동_결제_취소_최악
6	5	96	5_안드로이드_핸드폰_갤럭시_휴대폰
7	6	70	6_음악_배경음악_소리_유튜브
8	7	69	7_내야_내지_돈돈_비용
9	8	61	8_회원_탈퇴_가입_요청
10	9	61	9_탈퇴_처리_방법_자동
11	10	61	10_구독_취소_결제_신청
12	11	57	11_녹음_소리_코고_파일
13	12	47	12_정기_방법_결제_취소
14	13	45	13_계정_삭제_탈퇴_해주시
15	14	43	14_버튼_탈퇴_계정_가기
16	15	42	15_이용권_무료_사용_달이
17	16	41	16_기능_측정_호흡_코골
18	17	40	17_진짜_로딩_건가_가요
19	18	37	18_네이버_어디가_프리미엄_보이지
20	19	36	19_사용_유료_사용자_후주

✓ 무료_체험_결제_기간

리뷰_묶음_1

✓ 설치_삭제_사용_유료

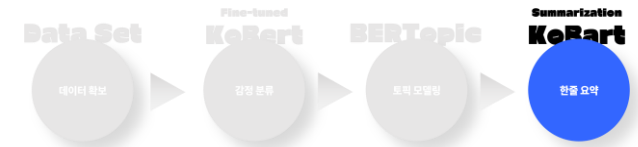
리뷰_묶음_1

✓ 환불_결제_요청_신청

리뷰_묶음_1

... ..

6-5 리뷰Text 한 문단 묶기



2. 한 문단으로 문서 묶음

✓ 무료_체험_결제_기간

```
[ ] topic0_fullreview_im
```

```
27         근데 무료 이용권 총 3회 있었는데 2회 뭐 때문인지 그냥 날 아가 있더라 구 용
106     마음이 편해 지내요 그리고 대부분 무료로 들을 수 있게 해 주셨으면 좋겠네요 이 ...
107     어른만 쓴다는 보장 규칙이 없으니 전체는 아니 여도 거의 다 무료 가입 안하고 들...
351         유료 콘텐츠를 조금 무료로 조금만 바꿀 순 없을까요
545     유료에서 몇 가지를 무료로 만들든지 무료로 늘리든지 해 주시면 좋을 것 같아요

...
5774         좀더 무료 화가 되었으면 좋겠어요
5850         그냥 공짜로 다해 주시면 좋은데
5873         무료로 바탕 화면 나 가도 소리가 재생 되게 해 주세요
6223         무료 버전이 많아 졌음 좋겠 네 여
6714         무료 콘텐츠와 유료 콘텐츠의 경계가 애매한 데 공지를 더 명확히 해 주세요
Name: Review, Length: 122, dtype: object
```



```
[ ] topic0_sen[:1000]
```

'근데 무료 이용권 총 3회 있었는데 2회 뭐 때문인지 그냥 날 아가 있더라 구 용.마음이 편해 지내요 그리고 대부분 무료로 들을 수 있게 해 주셨으면 좋겠네요 이 점 고 쳐 주세요. 어른만 쓴다는 보장 규칙이 없으니 전체는 아니 여도 거의 다 무료 가입 안하고 들을 수 있게 해 주셨으면 좋겠네요. 유료 콘텐츠를 조금 무료로 조금만 바꿀 순 없을까요. 유료에서 몇 가지를 무료로 만들든지 무료로 늘리든지 해 주시면 좋을 것 같아요. ㄴ 다른 거 무료 말로 이 것만 무료로 다시 바꿔 주면 너무 좋겠어요. 근데 무료 이용권 총 3회 있었는데 2회 뭐 때문인지 그냥 날 아가 있더라 구 용. 마음이 편해 지내요 그리고 대부분 무료로 들을 수 있게 해 주셨으면 좋겠네요 이 점 고 쳐 주세요. 어른만 쓴다는 보장 규칙이 없으니 전체는 아니 여도 거의 다 무료 가입 안하고 들을 수 있게 해 주셨으면 좋겠네요. 유료 콘텐츠를 조금 무료로 조금만 바꿀 순 없을까요. 유료에서 몇 가지를 무료로 만들든지 무료로 늘리든지 해 주시면 좋을 것 같아요. ㄴ 다른 거 무료 말로 이 것만 무료로 다시 바꿔 주면 너무 좋겠어요. 실제로 마음이 편해 지는 느낌이 에요 무료로 바뀐다면 모든 사람이 좋아할 거 같네요. 이어 플 설치하고 무료 체험 할 때 신중히 하세요 ㅡ 무료 체험 끝나면 바로 일 년 결 재가 되 어서 옴머 거자 먹기로 일 년 사용 하게 되었어요. 무료 체험 드길래 뒤로 가기 했는데 자동 가입이 되네요. 무료 체험 후 오늘 결제 됐길래 앱스토 어에서 정기 구독 취소 했는데 환불 처리 해 주세요 어디에다 요청 해야 하는지 어플에도 안 나와 있네요. 계정 탈퇴 앱에서 불가능 메일 로만 가능 처리 되었는지 확인 불가능 무료 체험을 하면 7일 후 이 자동 결제 되니 주의 하세요. 아니 무료 기능이 있어서 체험을 해보고 하는 것도 아니고 일단 결제 해 라면 띠 바 하. 오늘 무료 체험 하기 눌렀는데 무료 체험 끝나면 5만 9 천원 자동 결제 되는 거네요. 무료 체험 '

6-6 리뷰 요약



3. KoBart-Summarization을 이용하여 한 문장 요약 추출

✓ Topic1 문단

✓ 무료_체험_결제_기간

1. 35000원 자동 결제 됨요. 2년 가량 사용하고 있습니다
2. 코 골이나 기침 등 녹음이 안되어 있습니다 무엇이 문제인가요.
3. 35000 원씩 3년 해도 10만원 적은 돈이 아님을 알려주셨으면 합니다.
4. 잘 일어나 진다고 평점 좋길래 깔아 봤는데 솔직히 잘 모르겠어요.
5. 35000원 자동 결제 되었습니다. 9/17 일까지 무료 이용이라고 표시 되어 있는데
6. 오늘 9월 16일 자동 결제가 되었네요 환불 요청 합니다 구독도 취소 했습니다
7. 무슨 상황인가요. 10/30 일까지 무료 이용이라고 되어 있는데
8. 오늘 10월 29일 자동 결제 되었습니다 결재 취소 또는 환불 요청 합니다

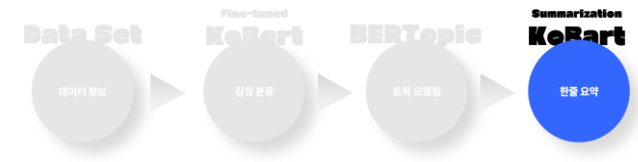
**“결제가 되는 줄 몰랐는데
환불 요청합니다.”**

KoBart-Summarization

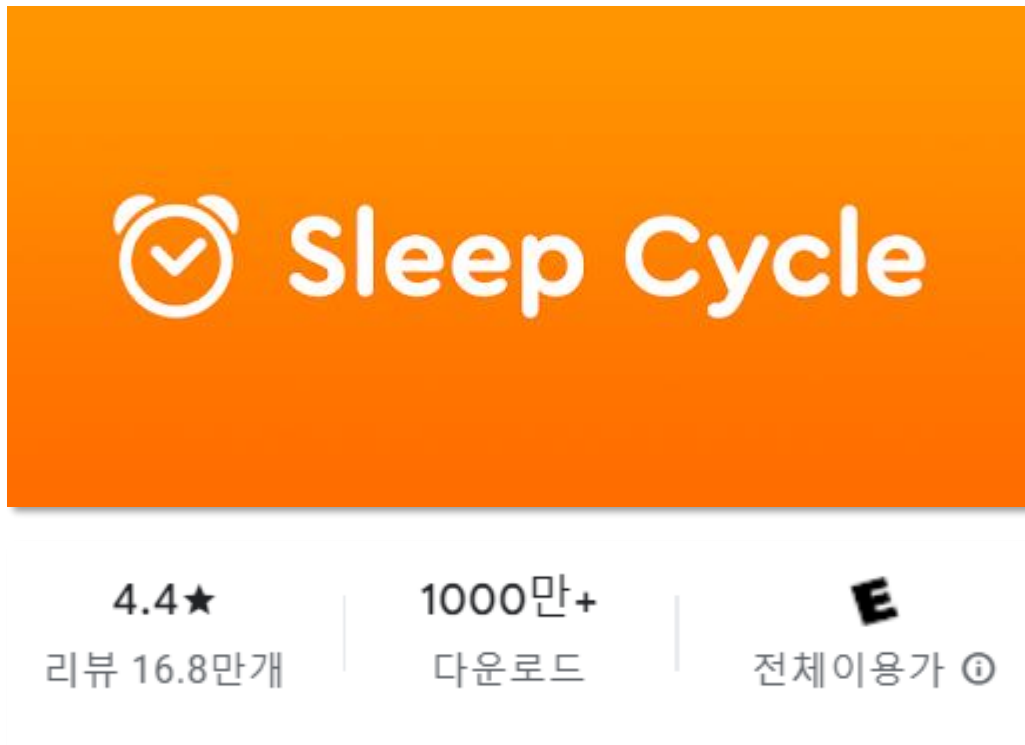
모델 결과 해석

Result

R-1 실제 적용 (Sleep Cycle)



✓ Sleep Cycle 앱 기준으로 결과 확인



✓ Sleep Cycle이란?

1. 현재 1위 수면 어플리케이션
2. 수면 추적 기능 / 지능형 알람 시계 / 수면 추적기 등

Positive

36개

✓ 알람_수면_진짜_아침

: 수면 알람 키고 자면 무조건 일어나게 되고 앱 중에 유로로 결제 하는 것 들 중에는 이게 최고인 것 같고 수면 알람 기능도 좋으며 잠도 잘 자고 일어나면 알람으로 깨워주는 앱도 있습니다

Negative

371개

✓ 무료_체험

: 무료 사용 한 달 끝나는 시점부터 자동 결제 된다고 미리 고지 되어 있는 스마트 알람을 개인 인증으로 개인정보를 내놓거나 프리미엄 결제 하라고 하네요

✓ 안드로이드_핸드폰

: 안드로이드 폰으로 바꾸면서 1년 사용 요금이네요 그 외의 기능 자체는 괜찮지만 삼성 헬스랑 연동 안됨 추가 지원 바랍니다.

✓ 소리_알람

: 알람 끄는 것과 알람 소리처럼 백색 소음 설정 하는 것도 홈에 같이 빼주시면 편할 것 같아요

✓ 노트_안내

: 수면 노트 오류로 인해 사용하던 수면 노트 다수가 삭제 되었는데 취소 환불은 어디서 하는 겁니까.

Positive

2.1만개

✓ 알람_수면_진짜_아침

: 수면 알람 키고 자면 무조건 일어나게 되고 앱 중에 유로로 결제 하는 것 들 중에는 이게 최고인 것 같고 수면 알람 기능도 좋으며 잠도 잘 자고 일어나면 알람으로 깨워주는 앱도 있습니다

이용자 충성도별로

우선 순위를 보다 세분화 할 수 있을까?

1점 리뷰 : 이탈 예정 유저 / 5점 리뷰 : 충성 유저

Negative

1.2만개

✓ 무료_체험

: 무료 사용 한 달 끝나는 시점부터 자동 결제 된다고 미리 고지 되어 있는 스마트 알람을 개인 인증으로 개인정보를 내놓거나 프리미엄 결제 하라고 하네요

✓ 안드로이드_핸드폰

: 안드로이드 폰으로 바꾸면서 1년 사용 요금이네요 그 외의 기능 자체는 괜찮지만 삼성 헬스랑 연동 안됨 추가 지원 바랍니다.

✓ 소리_알람

: 알람 끄는 것과 알람 소리처럼 백색 소음 설정 하는 것도 홈에 같이 빼주시면 좋을것 같습니다

✓ 노트_안내

: 수면 노트 오류로 인해 사용하던 수면 노트 다수가 삭제 되었는데 취소 환불은 어디서 하는 겁니까.

Positive

36개

✓ 알람_수면_진짜_아침

: 수면 알람 키고 자면 무조건 일어나게 되고 앱 중에 유료로 결제 하는 것 들 중에는 이게 최고인 것 같다고 생각합니다.

5점 리뷰

✓ Null

1점 리뷰

Negative

371개

✓ 결제_자동_사용_무료

: 작년 수험생 때 사용했으나 6개월 이상 사용하지 않아 정기 결제 해 놓은 걸 깜빡 했습니다. 유료 구매 하고 싶을 정도로 좋지만 1년 이용료 3만 4천 무료 버전이라도 잘 써서 리뷰 남겨드릴게요.

✓ 스토어_재구매

: 앱스토어에서 구매 했었는데 구글에서 재구매 해야 하는 건가요 ㅠ.

✓ 무료_체험_결제_자동

: 무료 사용 한 달 끝나는 시점부터 자동 결제 된다고 미리 고지 되어 있는 스마트 알람을 개인 인증으로 개인정보를 내놓거나 프리미엄 결제 하라고 하네요

✓ 무료_체험_기간_결제

: 1 달 무료 체험 후 자동 결제 되버림 정기 결제 취소조차도 까다롭게 만들어 놓음 사 기 어플이네요 무료라고 표시 하지 마세요

Positive

36개

✓ 알람_수면_진짜_아침

: 수면 알람 키고 **이용료를 낮춰볼까?** 되고 앱 중에 유료로 결제 하는 것 중에는 이게 최고인 것 같다고 생각합니다.
✓ **IOS와 안드로이드 서버를 통합할까?**

5점 리뷰

✓ Null

Negative

371개

✓ 결제_자동_사용_무료

: 작년 **수험생 때 사용했으나** 6개월 이상 사용하지 않아 정기 결제 해 놓은 걸 깜빡 했습니다. **유료 구매 하고 싶을 정도로 좋지만 1년 이용료 3만 4천 무료 버전이라도 잘 써서** 리뷰 남겨드릴게요.

✓ 스토어_재구매

: 앱스토어에서 구매 했었는데 **구글에서 재구매** 해야 하는 건가요 ㅠ.

✓ 무료_체험_결제_자동

: **무료 사용 한 달** 끝나는 시점부터 **자동 결제** 된다고 미리 고지 되어 있는 스마트 알람을 개인 인증으로 개인정보를 내놓거나 프리미엄 결제 하라고 하네요

✓ 무료_체험_기간_결제

: 1 달 무료 체험 후 자동 결제 되버림 정기 결제 취소조차도 까다롭게 만들어 놓음 사 기 어플이네 요 무료라고 표시 하지 마세요

Positive

36개

✓ 알람_수면_진짜_아침

: 수면 알람 키고 자면 무조건 일어나게 되고 앱 중에 유료로 결제 하는 것 들 중에는 이게 최고인 것 같다고 생각합니다.

✓ Null



- ✓ **이용자들에게 안내를 더 잘해주자!**
- ✓ **혹시 오류가 있는 것은 아닌지 확인하자!**
- ✓ **결제 플로우 자체를 개선하자!**

1점 리뷰

Negative

371개

✓ 결제_자동_사용_무료

: 작년 **수험생** 때 **사용했으나** 6개월 이상 사용하지 않아 정기 결제 해 놓은 걸 깜빡 했습니다. **유료 구매 하고 싶을 정도로 좋지만 1년 이용료 3만 4천 무료 버전이라도 잘 써서** 리뷰 남겨드릴게요.

✓ 스토어_재구매

: 앱스토어에서 구매 했었는데 **구글에서 재구매** 해야 하는 건가요 ㅠ.

✓ 무료_체험_결제_자동

: **무료 사용 한 달** 끝나는 시점부터 **자동 결제** 된다고 미리 고지 되어 있는 스마트 알람을 개인 인증으로 개인정보를 내놓거나 프리미엄 결제 하라고 하네요

✓ 무료_체험_기간_결제

: 1 달 무료 체험 후 **자동 결제 되버림** 정기 결제 취소조차도 까다롭게 만들어 놓음 사 기 어플이네요 무료라고 표시 하지 마세요



감사합니다!!

프로젝트 회고

개선점

✓ Data Crawling

1. 도메인별 리뷰 내용이 너무 상이하여 ‘수면’도메인에 한정하여 모델을 만든 것이 아쉬웠다.

✓ Pre-Processing

1. 문장 구분을 보다 세밀하게 할 수 있는 기준을 마련해보면 좋을 것 같다.

✓ Labeling

1. Labeling 수가 부족했다.
2. Labeling 기준을 보다 정교하게 세우지 못한 게 아쉽다.
3. 결국 ‘개선’분류를 못했다.
4. 어플리케이션 도메인을 아우르는 좋은 감성사전이 있었다면 성능이 더 좋았을 것 같다.

✓ KoBART

1. KoBART-summarization 모델을 사용할 경우 최대 토큰 갯수 제한이 있어서 상위 N개 토큰을 잘라서 요약을 진행
2. 토픽의 개수를 임의로 20개로 정함
3. Fine-Tuning을 적용 못해봤다.

✓ Text-Augmentation

1. 일반 문장의 수가 너무 작았던 것이 아쉽다.
2. Word2vec기반으로 고안했던 모델들을 구현하지 못한 것이 아쉽다.

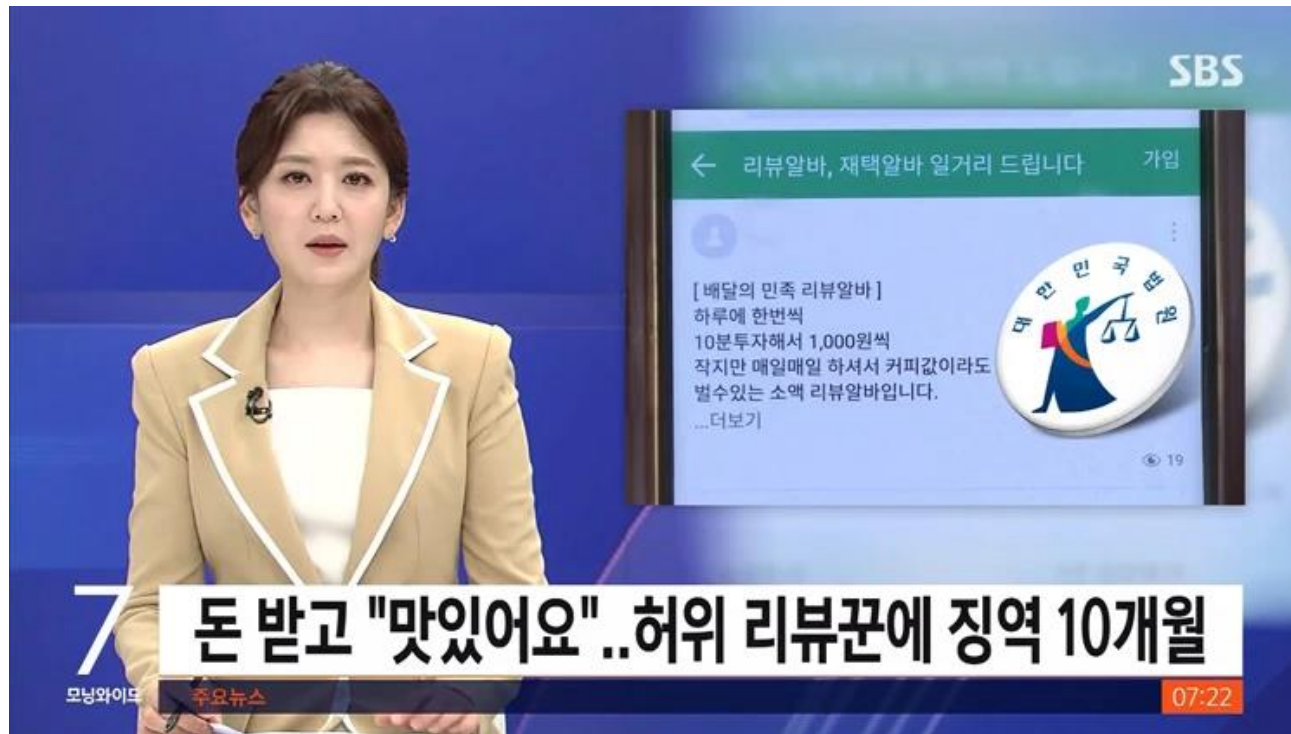
✓ Similarity

1. 자카드 유사도와 별도로 코사인 유사도도 돌려봤다면 좋았을 것 같다.
2. 내용 기반 유사도를 넘어서 말투 기반 유사도도 확인할 수 있었다면 좋았을 것 같다.

Appendix

4.4 Similarity

Appendix - 배경



✓ 구매 리뷰가 너무 많아요!

1. 리뷰 분석을 통해 경쟁사 분석을 실행할 때 구매 리뷰를 제거하면 실제 이용자들의 의견을 들을 수 있을 것!

Appendix - Similarity

✓ 기본 전처리

1. 정규 표현식 사용 후 삭제된 리뷰 제거

✓ Tokenizer

1. Okt(Open Korean Text)형태소 분석기 사용
: '좋아요' 등의 무성의 리뷰 제거
2. Stem = False, Norm = False 로 설정
: 실제 리뷰의 말투를 고려하기 위해 노력

✓ Jaccard 유사도 계산

1. 순서를 고려하지 않은 계산법 필요
2. 유사도를 계산할 때 리뷰들의 Token 개수가 달라도 됨

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Result

1. 전체 리뷰 중 유사 리뷰 비율

감정 Label	유사 리뷰 비율
긍정	4%
부정	2.4%
개선	4.6%
일반	1.1%

2. 유사 리뷰 중 감정별 비율

감정 Label	유사 리뷰 비율
긍정	44.2%
부정	30%
개선	20.4%
일반	8.4%