

Read your mind

“ KoBERT 모델 기반 일기에 어울리는 노래 추천 프로젝트 ”



Phoenix
김이정(PM)
김나린
박지우
최형규

Contents

1. Introduce
2. Detail
3. Process
4. Demonstration
5. Q&A

프로젝트 소개
상세 설명
진행과정
프로젝트 시연
질의 응답



INTRODUCE

1. 기획 의도

일기 기록하기

기분전환이 필요할 때 주로 음악 감상의 필요를 느꼈다.
89.2%가 노래 감상 시 가사를 고려하는 것으로 나타났으며,
가사의 비중은 5점 척도에서 평균 4.09점으로 높게 측정되었다. (중략)...
대부분은 가사를 통해 공감과 위로를 얻는다고 보고되었다.

〈노래 감상 시 대학생이 인식하는 가사의 비중과 역할〉
이화여자대학교 교육대학원

내 마음과 딱 맞는 가사가 있는
노래가 듣고 싶어.



BUT,

기존의 노래 플랫폼(멜론, 지니, 유튜브 등)에서 추천하는 노래들은
사용자의 선호 알고리즘에 의한 추천일 뿐,
사용자의 현재 감정에 초점을 맞추는 형태는 아님.



Read your mind with music



사용자가 일기 입력한다면?

작성한 일기 속 감성을 분석



사용자에게 적합한 노래 추천 가능!

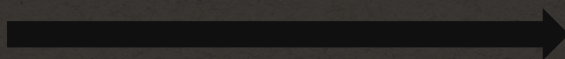


Read your mind with music

일상의 감정과 어울리는 노래를 추천해줄 수는 없을까?



작성한 일기 속 감정을 분석



사용자가 일기 입력한다면?

사용자에게 적합한 노래 추천 가능!

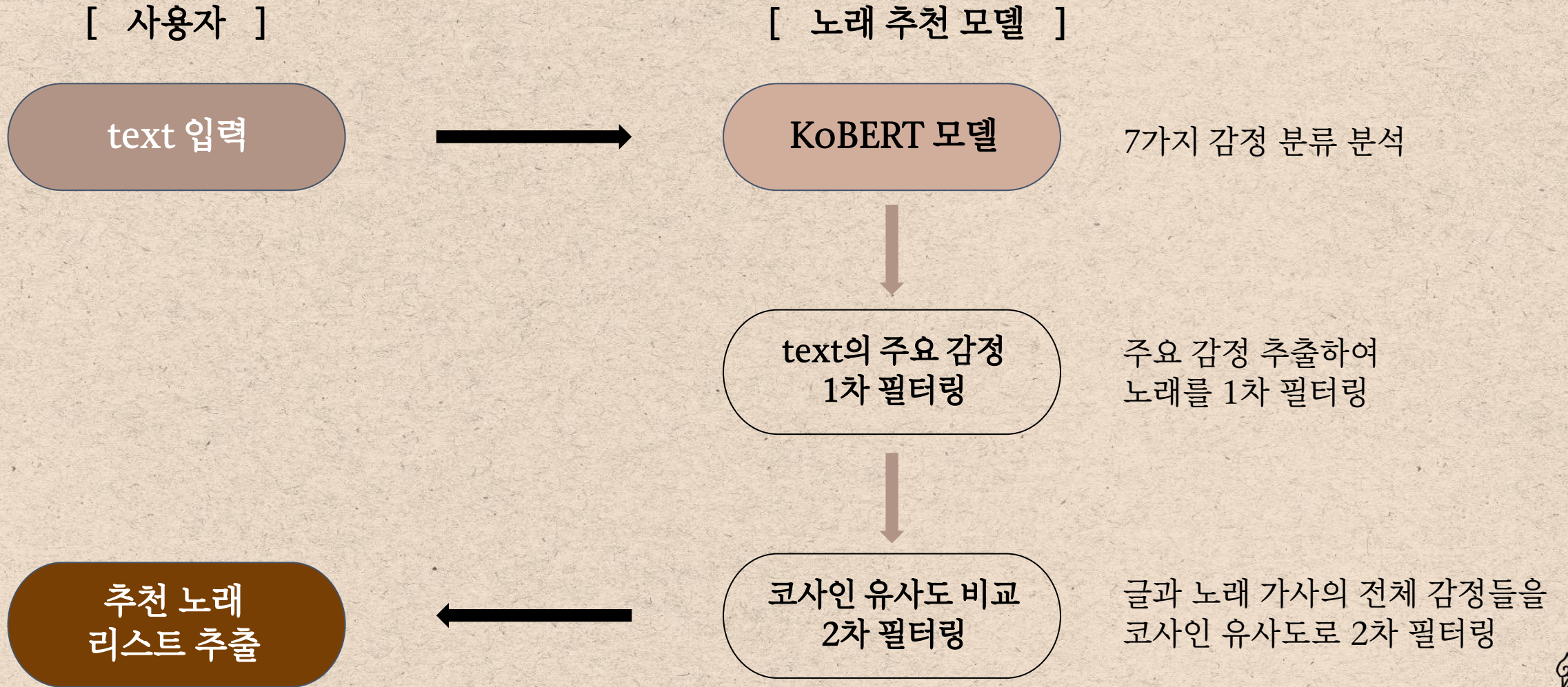




Detail

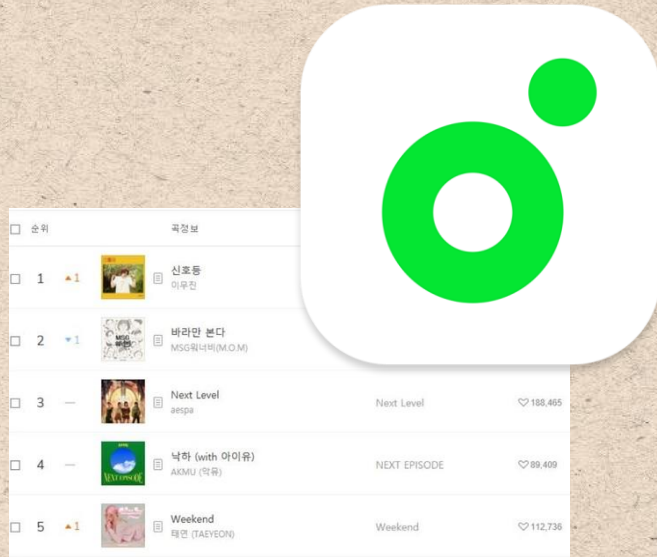
1. 구조도
2. Data
3. Modeling

일기 기록하기





Melon에서 사용자에게 추천해주기 위한 장르별 음악 크롤링



Melon

스테디셀러 50개,
인기곡 50개 크롤링

장르별
노래 제목, 가수, 가사,
이미지, 좋아요

*장르: 발라드, 댄스, 인디, 랩, 락, 소울, 트로트

Selenium을 활용한 크롤링 코드



웹에서 원하는 정보를 얻기 위한 패키지
사람이 버튼을 누르는 등 행동을 하는 대신
컴퓨터가 할 수 있도록 지원

```
# 크롤링할 페이지가 있는 주소값을 넣기
url = "https://www.melon.com/genre/song_list.htm?gnrCode=GN0200&steadyYn=Y"

driver.get(url)
wait = WebDriverWait(driver, 10)

# popular_button # song_info = song_info.split('\n')
# popular_button. # 추출된 노래 정보 출력
song_list = []
# # 클릭 후에 페이지 title = []
# wait.until(EC.singer = []
txt = []
# 테이블의 각 행을 가 img = []
song_info = [] like = []
table_rows = driver
for song in song_info:
    song_l = song['text'].split('\n')
    title.append(song_l[1])
    singer.append(song_l[2])
    txt.append(song['id'])
    img.append(song['img'])
    like.append(song_l[5])

for row in table_
checkbox_valu
image_link =

song_info.append
'id': che
'img': in
'text': r
})

# 노래명, 가수명, 노래 가사, 앨범사진, 좋아요 수

# 브라우저 닫기
driver.quit()
```


Model 학습을 위한 Data Sets 구축



- 감성 대화 말뭉치 58271개
- 한국어 연속성 대화 데이터 셋 55630개
- 감정 분류를 위한 음성 데이터 셋 43991개

행복, 슬픔, 분노, 공포, 혐오, 놀람 + 중립

감정 분류 by Susan David, 하버드 심리학자

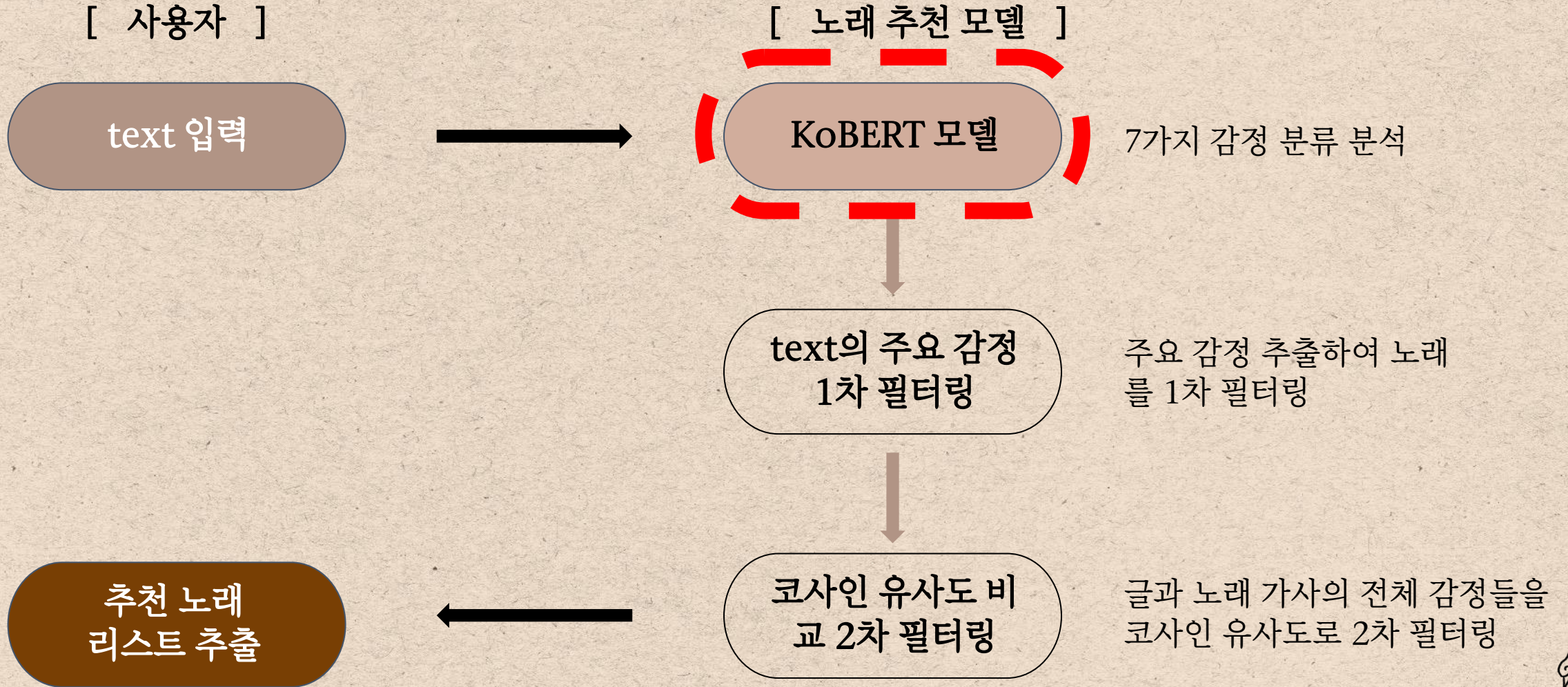
감정 분포의 비율을 맞추기 위한

Data Augmentation 진행

가장 일기의 Text와 노래 가사와의 일치도가 높은 DataSets 선택

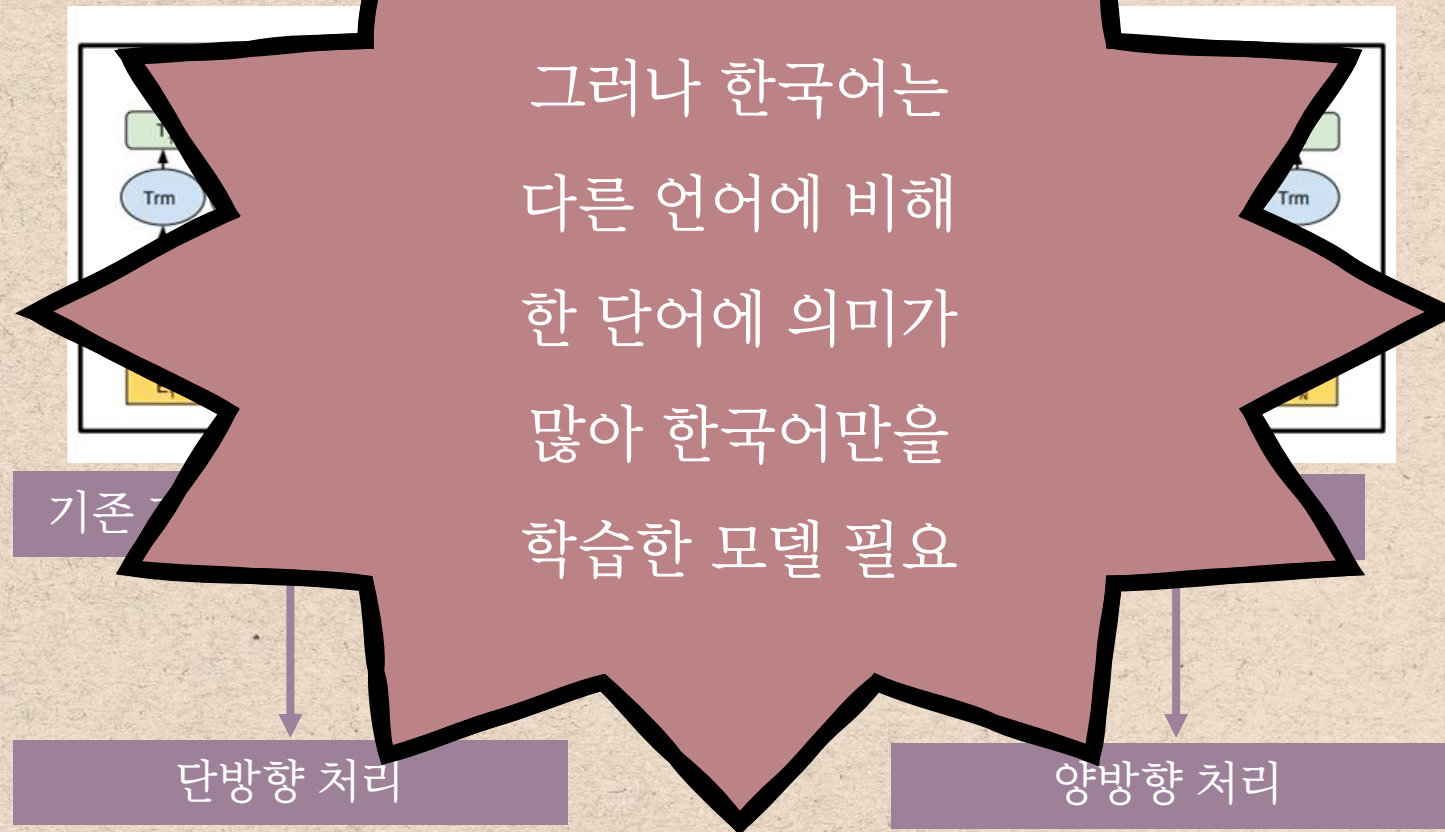
최종 데이터셋 총 개수 157,862개

Emotion	
중립	47048
슬픔	36250
분노	25680
놀람	16425
공포	14662
행복	12917
혐오	4880



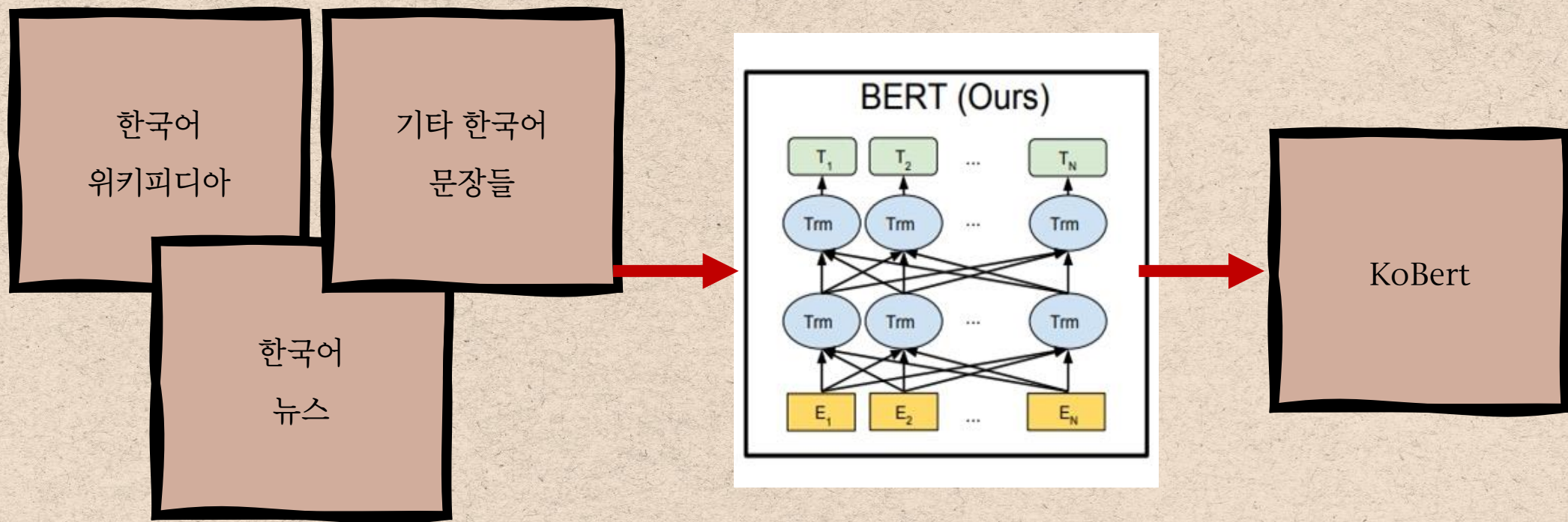
BERT model

현존하는 가장 강력한 NLP 언어모델. 언어모델 전이학습 컨셉을 적용.

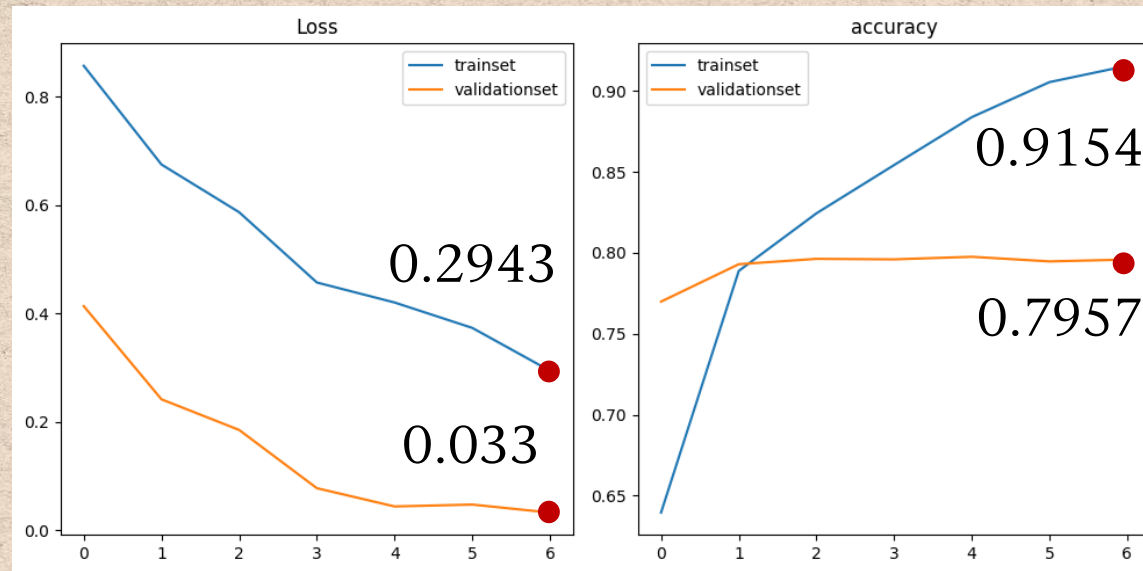


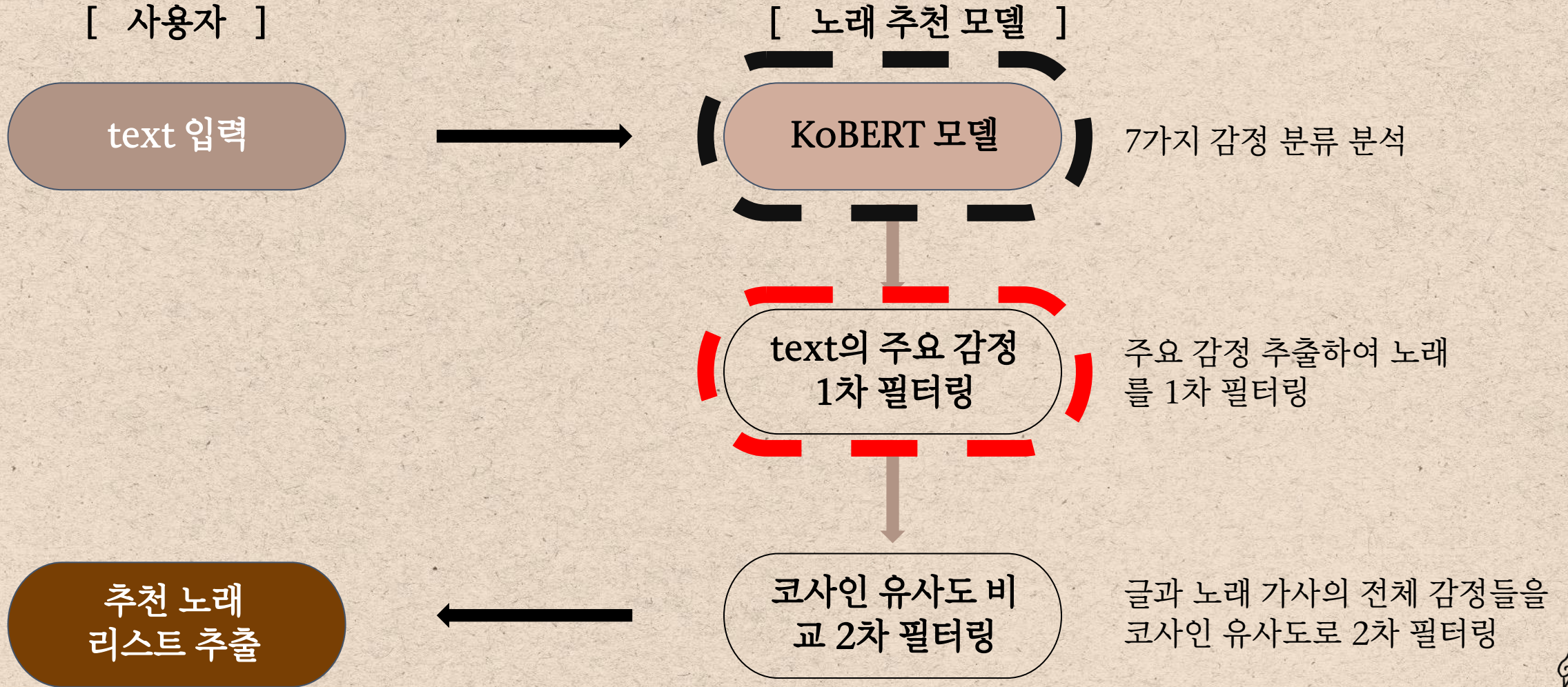
KoBERT model

SKT-Brain에서 개발한 한국어 기반 BERT 모델.



03. Modeling / ReUmi model





KoBERT 모델

7가지 감정 분류 분석 후 최종 Datasets

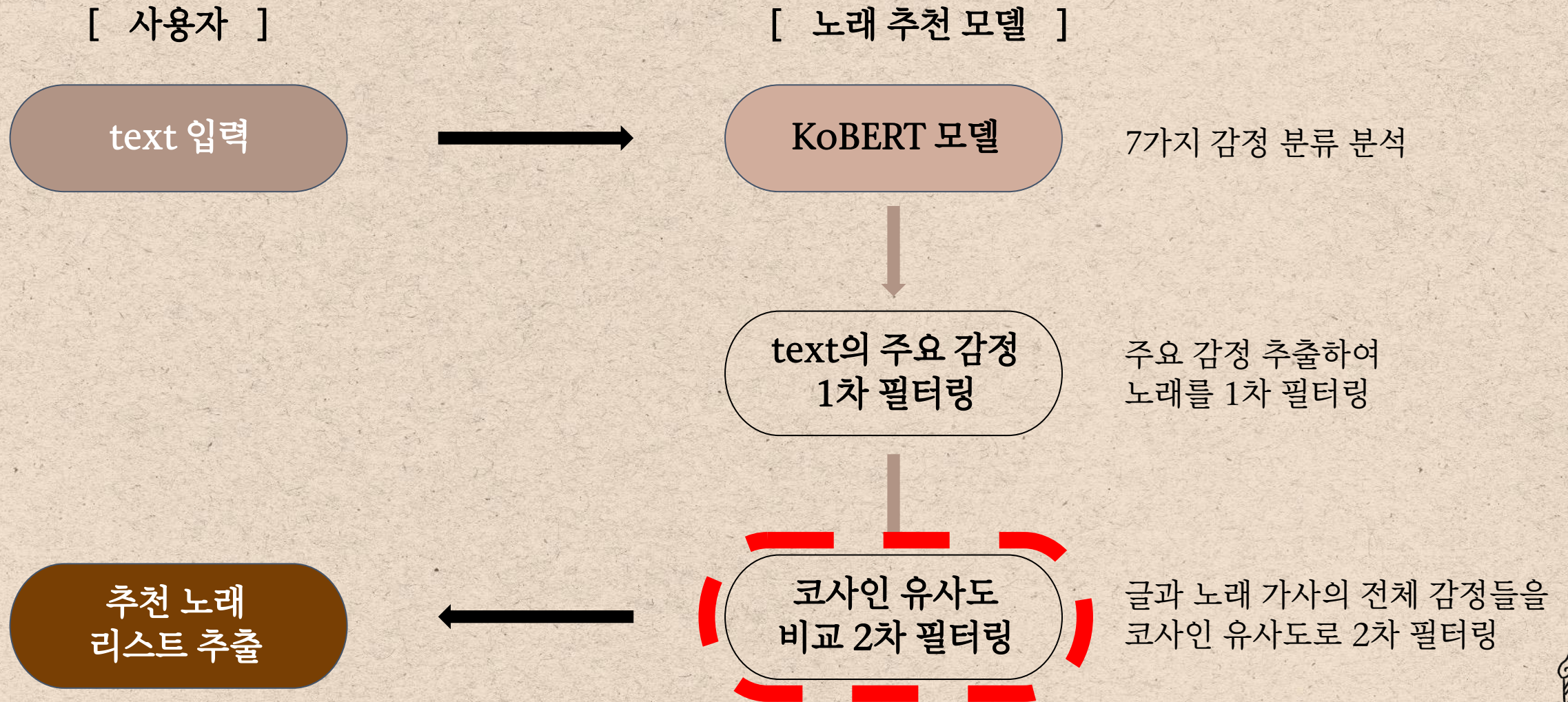
[행복, 슬픔, 분노, 공포, 혐오, 놀람, 중립]

title	singer	lyrics	img	like	emotions	category
헤어지자 말해요	박재정	헤어지자고 말하려 오늘 너에게 가다가 우리 추억 생각해 봤어 처음 본 네 얼굴 마주친 눈동자 가까스로 본 너의 그 미소들 ...	https://cdnimg.n	143,076	[5, 26, 0, 1, 1, 1, 1]	슬픔
사막에서 꽃을 피우듯	우디 (Woody)	아침에 눈 뜨는 게 너무 행복해졌어 널 사랑한다 말할 수 있어서 하늘에 감사해 ...	https://cdnimg.n	68,945	[7, 22, 11, 0, 0, 1, 0]	슬픔
인사	범진	돌아서는 너를 보며 난 아무 말도 할 수 없었고 슬퍼하기엔 짧았던 나의 해는 저물어 갔네 ...	https://cdnimg.n	45,377	[0, 16, 3, 2, 2, 0, 2]	슬픔

text의 주요 감정
1차 필터링

사용자의 TEXT가 '슬픔'으로 인식 → '슬픔'에 해당하는 노래만 불러옴





코사인 유사도 Cosine Similarity

| 두 벡터 사이의 각도를 계산하여 두 벡터가 얼마나 유사한지 측정하는 척도

TF-IDF 벡터화

단어의 빈도를 수치화하여 표현



구분	문장						
문서1	나는 아침보다 저녁이 좋다						
문서2	사과는 아침보다 저녁이 좋다						
문서3	사과는 점심 간식으로 좋다						
	나는	아침보다	저녁이	좋다	사과는	점심	간식으로
문서1	1	1	1	1	0	0	0
문서2	0	1	1	1	1	0	0
문서3	0	0	0	1	1	1	1

TF-IDF는 단어의 중요도를 고려하여 가중치를 주기 때문에 문서에서 유의미한 정보를 얻어내는 데 더욱 효과적

코사인 유사도 비교
2차 필터링

글과 노래 가사의 전체 감정들을 코사인 유사도로 2차 필터링

사용자 입력text



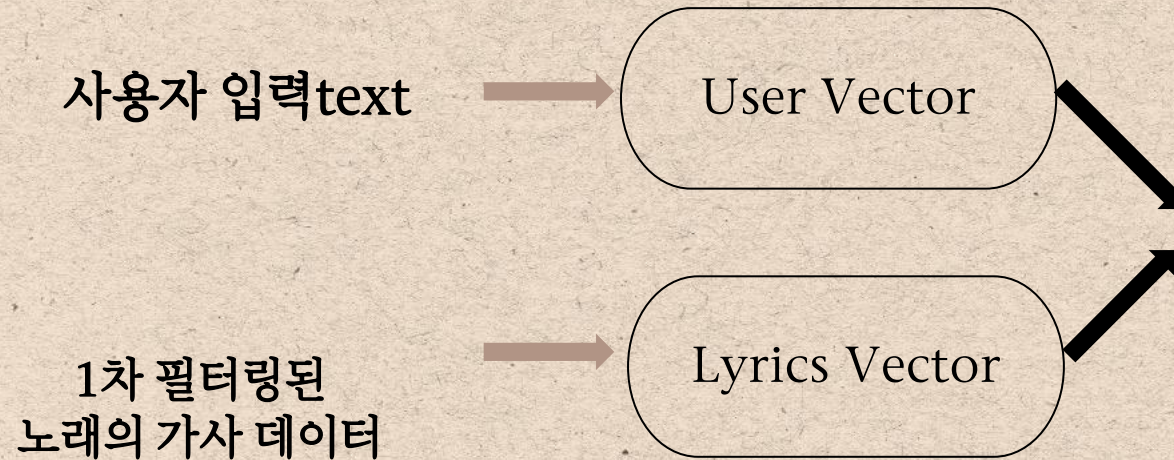
1차 필터링된
노래의 가사 데이터



코사인 유사도 비교
2차 필터링

글과 노래 가사의 전체 감정들을 코사인 유사도로 2차 필터링

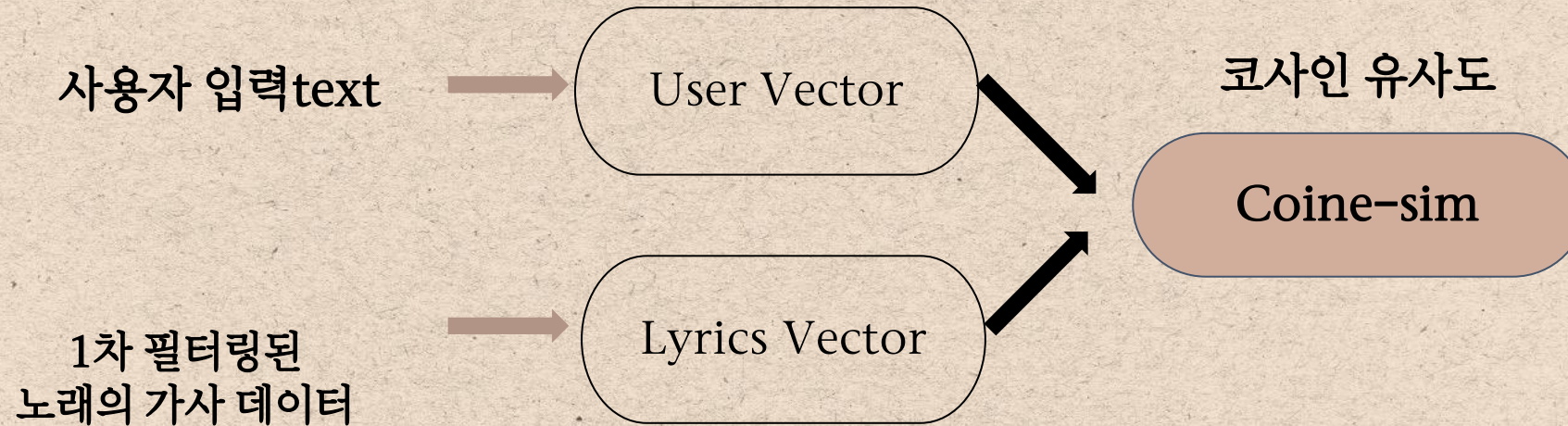
TF-IDF 벡터화



코사인 유사도 비교
2차 필터링

글과 노래 가사의 전체 감정들을 코사인 유사도로 2차 필터링

TF-IDF 벡터화



코사인 유사도 비교
2차 필터링

글과 노래 가사의 전체 감정들을 코사인 유사도로 2차 필터링

TF-IDF 벡터화

사용자 입력text

User Vector

1차 필터링된
노래의 가사 데이터

Lyrics Vector

코사인 유사도

Cosine-sim

상위 3개의
곡 리스트 추출

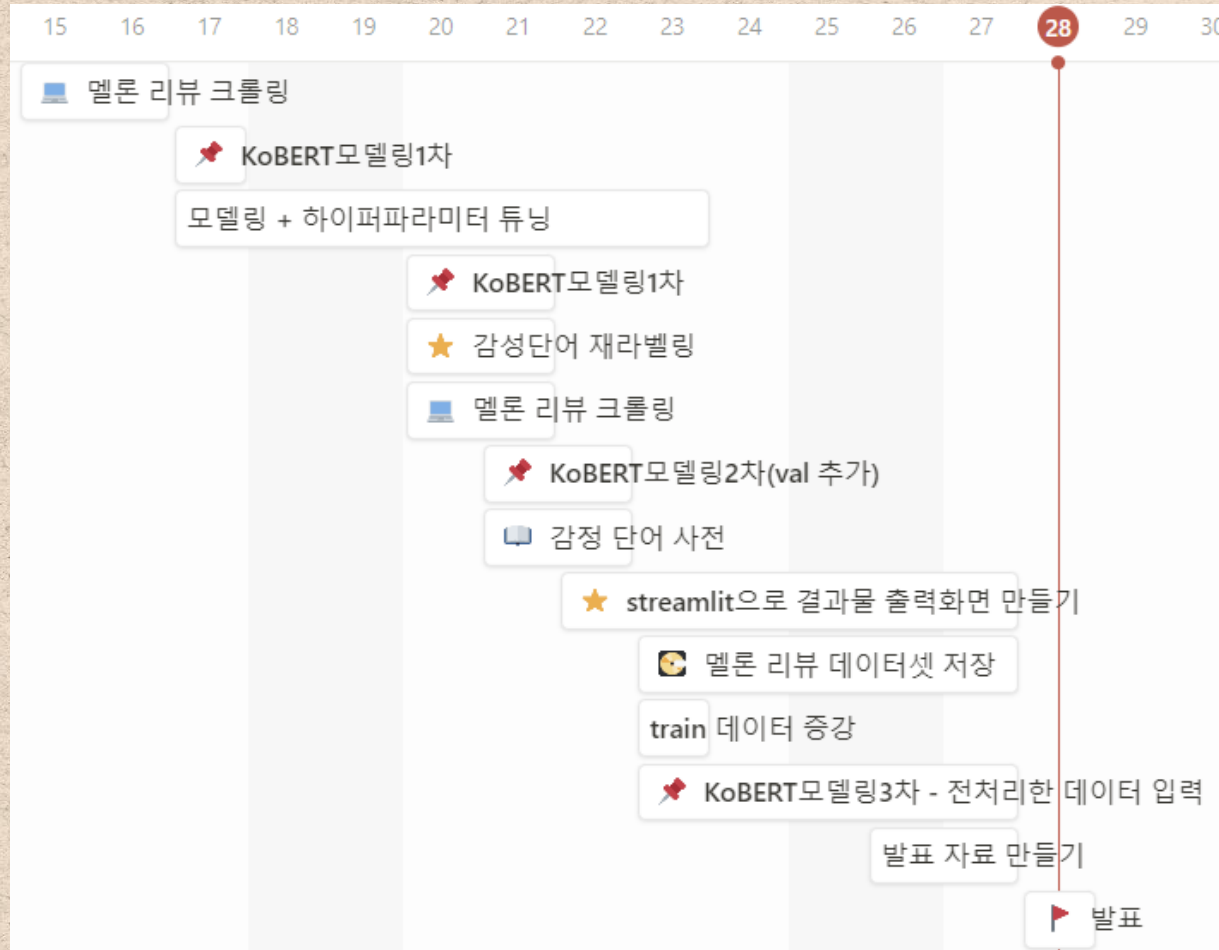


Process



1. 개발 일정
2. 프로젝트 이슈

일기 기록하기



프로젝트 기간: 11/13(금)~11/28(화)



[KoBert import]

```
from kobert import get_tokenizer
from kobert import get_pytorch_kobert_model
```

ModuleNotFoundError Traceback (most recent call last)

<ipython-input-9-9f2efe92f183> in <cell line: 1>()

----> 1 from kobert import get_tokenizer

2 from kobert import get_pytorch_kobert_model

ModuleNotFoundError: No module named 'kobert'

NOTE: If your import is failing due to a missing package, you can manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the "Open Examples" button below.

OPEN EXAMPLES SEARCH STACK OVERFLOW



2023년 8월 부터 colab에서 KoBert가 실행이 안되는 문제 발생.



[KoBert import]

```
#KoBERT
from kobert.utils import get_tokenizer
from kobert.pytorch_kobert import
get_pytorch_kobert_model
#transformer
from transformers import AdamW
from transformers.optimization import
get_cosine_schedule_with_warmup

# ★ Hugging Face를 통한 모델 및 토큰라이저 Import
from kobert_tokenizer import KoBERTTokenizer
from transformers import BertModel

from transformers import AdamW
from transformers.optimization import
get_cosine_schedule_with_warmup
```

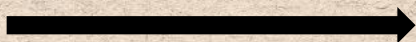
KoBert를 import 하는 대신
Hugging Face를 활용해
BERTDataset과
BERTSentenceTransform의 class 수정.



[데이터의 품질]



데이터 셋을 shuffle 할 때마다
모델의 train acc 및 valid acc가 달라짐.



데이터의 감정별 비율이 불균형, 품질저하 데이터셋 포함되어 있음.
일기에 맞는 데이터셋이 아닌 것이 원인



- 데이터 셋 shuffle 후, 가장 성능이 좋은 데이터 셋 선정.
- 데이터 품질에 영향을 끼치는(비속어, 정치적 언어셋) 데이터 삭제.
- 가사와 일기 간 유사도를 추가하여 정확도 향상.



[colab과 streamlit]



BERT 모델은 각 토큰의 인덱스와 타입, 유효길이를 필요로 한다.

gluonnlp 라이브러리는 이를 매우 편하게 해주는 함수를 제공하므로 반드시 필요.

However...

gluonnlp가 로컬에서 실행되지 않고 colab에서만 실행되는 이슈 발견.



예상 가능한 해결 방법은 아래 두 개였다.

방법1. 로컬에서 gluonnlp 실행 가능하도록 하기

방법2. Streamlit과 colab 연결하기



[colab과 streamlit]

결국 해결된 방법은 Colab에서 streamlit을 실행시키는 것이었으나,
streamlit은 코드를 이용해 새 창을 띄우는 것이기 때문에 기본적인 Google Colab에서는 실행불가능하였다.
그래서 ngrok를 이용하여 streamlit에서 실행되는 주소를 우회시켜 브라우저를 띄울 수 있게 하였다.



```
!pip install pyngrok==4.1.1
from pyngrok import ngrok
!ngrok authtoken <PASTE YOUR API KEY>
public_url = ngrok.connect(port='8501')
print(public_url)
```





Demonstration

1. 프로젝트 시연
2. 팀원 소개

일기 기록하기

일기 텍스트 입력

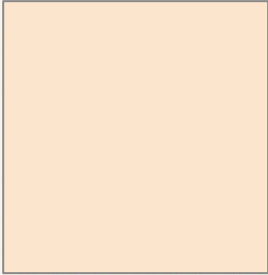
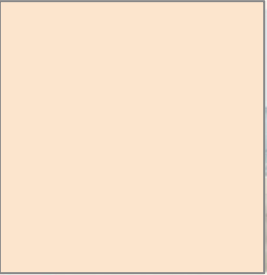

죽은 것은 이미 정해진 일이기에
명랑하게 살아라
언젠가는 끝날 것이기에 온 힘을 다해 맞서자.
시간은 한정되어 있기에 기회는 늘 지금이다
울부짖는 일 따윈 오페라 가수에게나 맞겨라

니체의 말 中 / 프리드리히 니체

일기 기록하기

텍스트 감정 추출을 통해
사용자를 위한 노래 추천

나만의 감정일기 노래리스트

		
<small>포크/블루스</small> THE BOYZ 2ND ALBUM [PHANTASY] Pt.2 Sixth Sens [PHANTASY] Pt.2 ... AKMU (악뮤)	<small>발라드</small> 미리 미리 크리스마스 (Feat. 천동 Of MBLAQ) AKMU (악뮤)	<small>댄스</small> THE BOYZ 2ND ALBUM [PHANTASY] Pt.2 Sixth Sens [PHANTASY] Pt.2 ... AKMU (악뮤)

02. Phoenix team 팀원 소개



김나린

Front-End
Development



김이정

PM
Development



최형규

Development



박지우

Development





감사합니다!

fin

Q&A

일기 기록하기

Read Your Mind