수고했어, 오늘도.

자연어 처리 기반 감정 기록 서비스

[감정분석단] 김현주 / 박준혁 / 박호준 / 조영훈

서비스 산업 데이터를 활용한 빅데이터 분석 8회차

Index

1 . 프로젝트 개요

2 . 프로젝트 설계

3. 프로젝트 진행과정

4. 프로젝트 결과물

일기

감정일기 앱

청사진

[건강플러스] 현대인의 15% 경험하는 우울증

한윤조 기자 hanyunjo@imaeil.com

매일신문 입력 2021-11-17 15:30:00 수정 2021-11-17 08:16:19

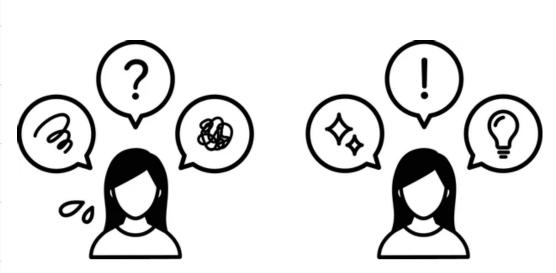
현대인의 질병, 미소 우울증 "내가 웃을 땐, 내가 울고 있다"

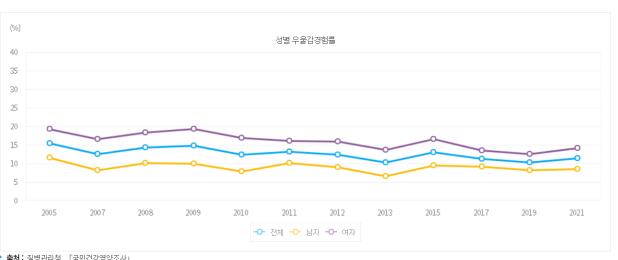
등록 2021-05-27 17:05 | 수정 2021-07-21 10:30 성진규 | 하이닥 건강의학기자

[출근길 인터뷰] '코로나 블루' 덮친 10·20대...우울증·공 황장애 급증

코로나로 5명중 1명 우울증..코로나 발생 초기 대비 자살 생각 40% 증가

우울증 80% 혼자 해결하려다 상황 악화..자가진단 5점 이상 꼭 진료봐야





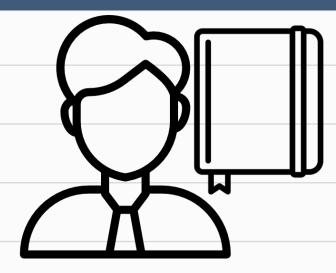
- 출처: 질병관리청, 「국민건강영양조사」
 - * 자료: 질병관리청, 「2021 국민건강통계-국민건강영양조사 제8기 3차년도(2021)」
- ▶ 주석: 1) 우울감경험률은 최근 1년 동안 연속 2주 이상 일상생활에 지장이 있을 정도로 슬프거나 절망감 등을 느낀 적이 있는 사람들의 비율임. 2) 전체 및 성별 자료는 2005년 추계인구로 연령표준화한 수치임

프로젝트 진행과정

프로젝트 결과물

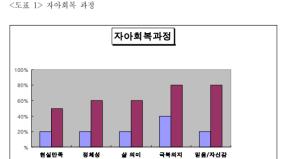
40대 직장인 남성 대상

저널 치료 프로그램 3개월간 진행 결과 분석



<도표 3> 자아실현 과정

자아 회복↑ 자아 발견↑ 자아 실현↑



■저널쓰기 전 ■저널쓰기 후



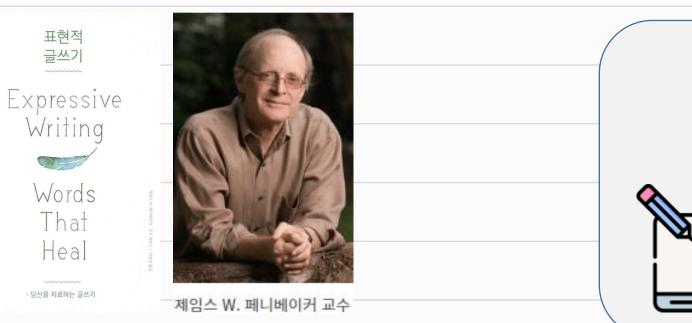
차미숙. "저널치료를 통한 자아발견." 국내석사학위논문 경기대학교 문화예술대학원, 2010. 경기도

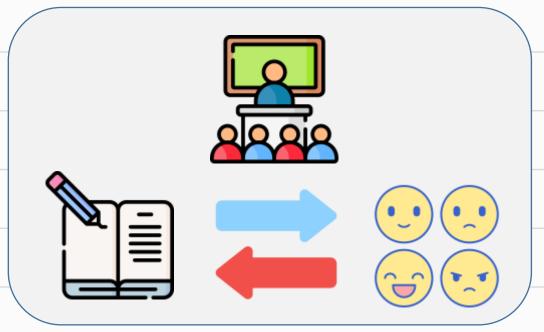
[논문]이봉희. (2007). 저널치료: 새로운 일기쓰기. 새국어교육, 77, 235-264.

일기

감정일기 앱

청사진





- · '80년대 중반, 50여명의 학생을 대상으로 실험을 진쟁.
- · 실험주제는 '글쓰기와 면역체계의 영향을 미치는가'
- · 실험 결과 글로 감정을 털은 이들은 그렇지 않은 이들보다 진료율이 43%↓

일기

감정일기 앱

청사진



MOODA 4+

내 안의 나와 마주하는 시간

OLIVESTONE Lab.

라이프 스타일 앱 1위

★★★★ 4.4 + 111개의 평가

₩1,500

감정 기록 #지난 날짜 감정기록하기 #월별 감정 보기



[영화 'HER'] 인공지능 운영체제 '사만다'와의 대화를 통해 위로를 받는 모습



일상 - 다이어리, 일기 4+

당신의 소중한 일상을 간직하세요.

DONGWOOK KIM

iPad용으로 디자인됨

★★★★ 4.8 • 738개의 평가

무료ㆍ앱 내 구입 제공











일기

감정일기 앱

청사진



MOODA 4+

내 안의 나와 마주하는 시간 OLIVESTONE Lab.

라이프 스타일 앱 1위

★★★★ 4.4 • 111개의 평가

₩1,500

감정 기록 #지난 날짜 감정기록하기 #월별 감정 보기



일상 - 다이어리, 일기 4+

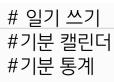
당신의 소중한 일상을 간직하세요.

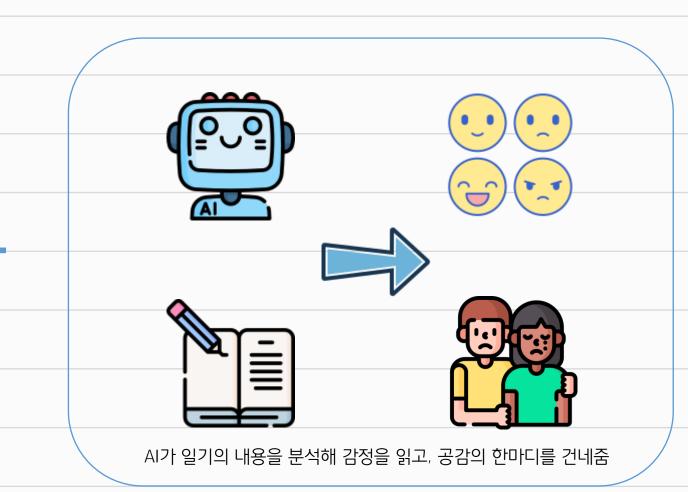
DONGWOOK KIM

iPad용으로 디자인됨

★★★★★ 4.8 • 738개의 평가

무료ㆍ앱 내 구입 제공



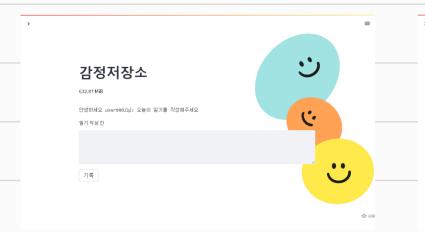


일기

감정일기 앱

청사진

- page







일기쓰기 PAGE

- 일기 쓰기/저장 기능
- 감정 분류(KoBERT)
- 문장 생성(BERT)
- 일기 작성 TIP



지난 기록 보기 PAGE

- 달력에 기록된 감정스탬프
- 지난 일기 보기



감정 그래프 PAGE

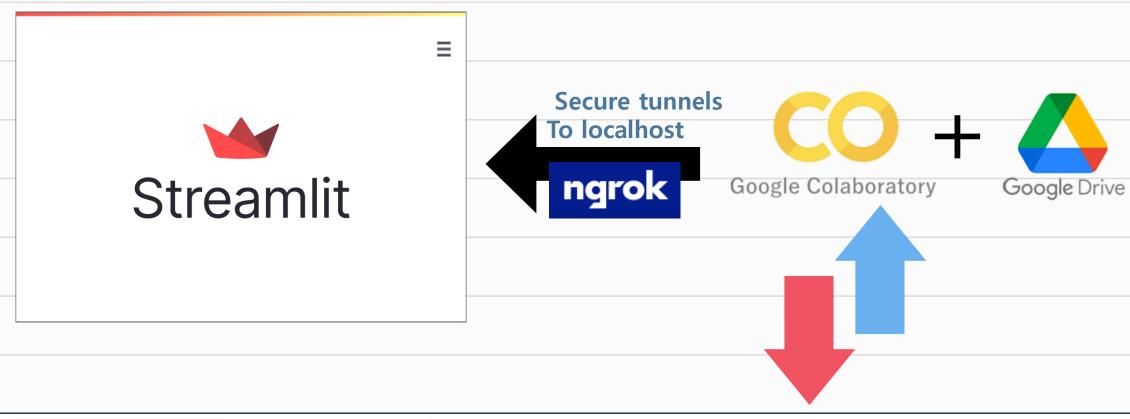
- -월별 감정 그래프 출력
- 현재 상태에 대한 문구 출력

일기

감정일기 앱

청사진

- 대시보드



user**0001**

2023-01-31

2023-02-02

2023-02-03

daylog: "슬픈 하루야"

emotion: "sad"



운영방법

컨셉회의	데이터 전처리	l	기능구현		마무리	
WEEK 1 WEEK 2	WEEK 3	WEEK 4	WEEK 5	WEEK 6	WEEK 7	
아이디어 회의						
	데이터 수집 및 정제	탐색적 자료 분석				
			모델	구축	대시 보드 구축	

플랫폼 // 회의 및 자료 관리



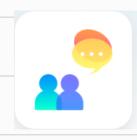
데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

대시보드



#코퍼스 # 감성대화 # 감성 챗봇 # 우울증 예방 *****

감성 대화 말뭉치

분야 한국어 유형 오디오,텍스트

갱신년월: 2022-12 구축년도: 2020 조회수: 12,490 다운로드: 5,865 용량: 20,35 MB

질문과 답변이 있는 대화 데이터셋 # 6가지 감정



#텍스트 감정 인식 # 다분류 감정 # 텍스트 데이터 레이블링

한국어 감정 정보가 포함된 단발성 대화 데이터셋

분야 기타 **유형** 텍스트

갱신년월: 2022-05 조회수: 866 다운로드: 223 용량: 1,57 MB



#일상 대화문 # 감정 레이블링 # 챗봇

한국어 감정 정보가 포함된 연속적 대화 데이터셋

분야 기타 유형 텍스트

갱신년월: 2022-05 조회수: 622 다운로드: 136 용량: 2,44 MB

문장과 감정 # 7가지 감정 [감성 대화 말뭉치]

감정대분류, 감정 소분류 // 상황, 문장, ... # 데이터 수 58270 x 13

KoBERT // "사람문장1" 과 "감정_대분류(6가지 감정)" # BERT // 질문과 답변

[한국어 감정 단발성]

감정대분류, 감정 소분류 // 상황, 문장, ... # 데이터 수 38595 x 2

KoBERT // "Sentence" 와 "Emotion(7가지 감정)"

프로젝트 개요	프로젝트 설계	프로젝트 진행과정

프로젝트 결과물

데이터 전처리	감정 분류 모델	문장 생성 모델	대시보드

감정분류 데이터셋

데이터 수집 및 분석

감성대화말뭉치	단발성, 연속적		감정통합	20000 -	12000 -			
기쁨	행복			기쁨	기쁨	15000 -	8000 -	
슬픔	슬픔		슬픔	5000 - - - - - - - - - - - - - - - - - -	2000 -			
상처	20		20	0	Sentence Emotion 일은 왜 해도 해도 끝이 없을까? 화가 난다. 짜증			
	분노	>>>	짜증	짜증	짜증		1 2	이번 달에 또 급여가 깎였어! 물가는 오르는데 월급만 자꾸 깎이니까 너무 화가 나. 짜증 회사에 신입이 들어왔는데 말투가 거슬려. 그런 애를 매일 봐야 한다고 생각하니까 스 짜증
분노	혐오						M[o	3 4
불안	공포		불안	 79153				
당황	놀람		당황	79154 79155 79156	이럴 땐 진짜 질색하는 표정이 된다니깐. 감탄스럽기까지 해. 기쁨			
-	중립		중립	79157	고맙긴요. 제가 할 수 있는 건 이 정도가 전부인걸요. 선배님, 부디 힘내세요. 기쁨 ows × 2 columns			

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

대시보드

문장생성 데이터셋

연령	성별	•••	사람문장1	시스템문장1	사람문장2	시스템문장2	사람문장3	시스템문장3
청년	남성		일은 왜 해도	많이 힘드시겠어요 …				•••
노년	여성	***	이번달에 또 급여가…	급여가 줄어 속상 …				

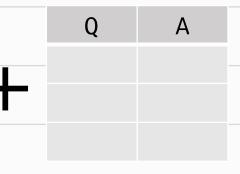
대화형 데이터

사람문장 1,2,3 → 'Q' 시스템 문장 1,2,3 → 'A'

감성	대화	데이	터셋
----	----	----	----

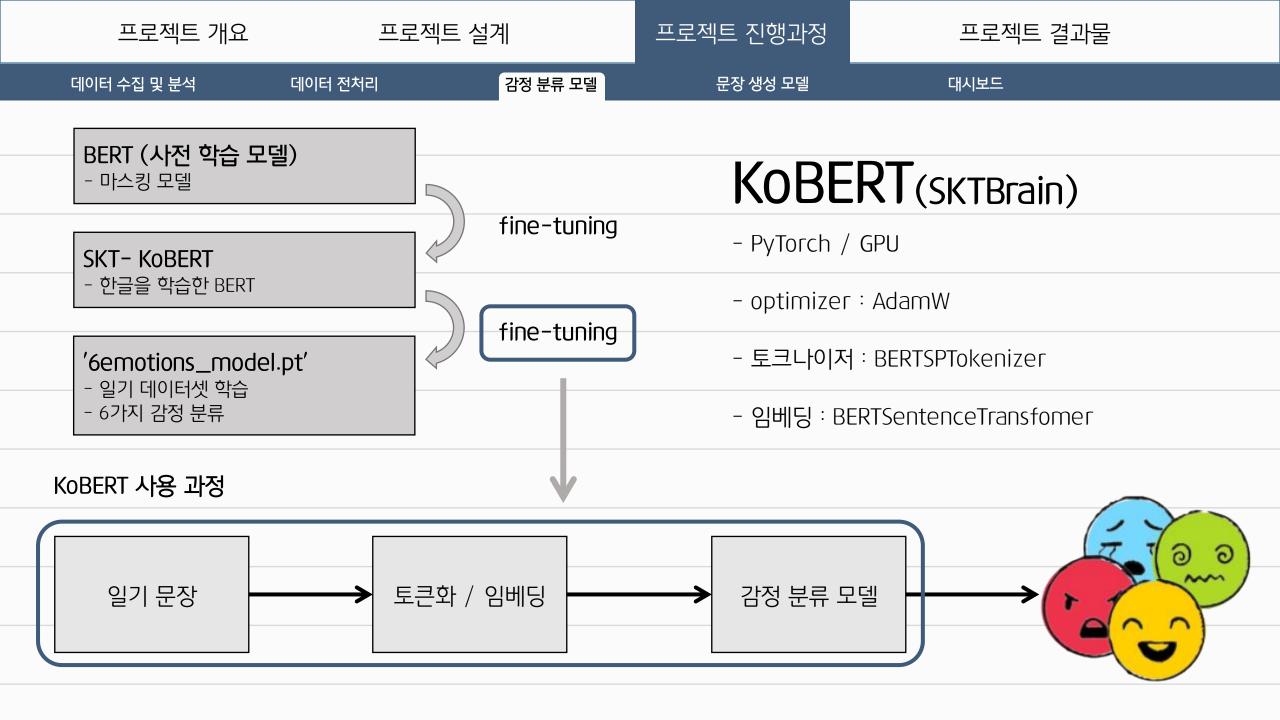
Q A +

Chatbot data

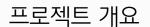




	Unnamed: 0	Q	A
0	0	일은 왜 해도 해도 끝이 없을까? 화가 난다.	많이 힘드시겠어요. 주위에 의논할 상대가 있 나요?
1	1	이번 달에 또 급여가 깎였어! 물가는 오르는데 월급만 자꾸 깎이니까 너무 화가 나.	급여가 줄어 속상하시겠어요. 월급이 줄어든 것을 어떻게 보완하실 건가요?
2	2	회사에 신입이 들어왔는데 말투가 거슬려. 그 런 애를 매일 봐야 한다고 생각하니까 스	회사 동료 때문에 스트레스를 많이 받는 것 같 아요. 문제 해결을 위해 어떤 노력을
3	3	직장에서 막내라는 이유로 나에게만 온갖 심부 름을 시켜. 일도 많은 데 정말 분하고	관련 없는 심부름을 모두 하게 되어서 노여우 시군요. 어떤 것이 상황을 나아질 수 있
4	4	얼마 전 입사한 신입사원이 나를 무시하는 것 같아서 너무 화가 나.	무시하는 것 같은 태도에 화가 나셨군요. 상대 방의 어떤 행동이 그런 감정을 유발하는
175740	175740	훔쳐보는 것도 눈치 보임.	티가 나니까 눈치가 보이는 거죠!
175741	175741	훔쳐보는 것도 눈치 보임.	훔쳐보는 거 티나나봐요.
175742	175742	흑기사 해주는 짝남.	설렜겠어요.
175743	175743	힘든 연애 좋은 연애라는게 무슨 차이일까?	잘 헤어질 수 있는 사이 여부인 거 같아요.
175744	175744	힘들어서 결혼할까봐	도피성 결혼은 하지 않길 바라요.
175745 rov	ws × 3 colun	nns	



	<u>-</u> 3	프로젝트	개요		프로젝트	트 설계		프로젝.	트 진행괴	정	<u>.</u>	프로젝트 결과물	
	데이터 수	-집 및 분석		데이터 전처리	:	감정 분	류 모델	- -	문장 생성 모덜	<u> </u>	대기	시보드	
	1. 감정 -	분류 데이터	l셋 (문장 +	라벨)									
		1.	감정(str)	-> 숫자	(int) 변환	<u> </u>							
		Sentence	<u> </u>	Emotion	변환	숫자							
	일은 왜	해도 끝이 없을	까? 화가 난다.	짜증		0				3. 임베딩	당 문장과	라벨링을 통해 학습	=
Ī	오	늘 너무 행복한	하루였어	기쁨	>>>	1							
		-							4. 1	ine-tu	ning mod	lel	
		-										_model.pt'	
		•	2. [레이터 전	처리(토	크화, 정수	인코딩, 피	대딩)					
	일은	왜	해도	해도	끝이	없을까	?	화가	난다		4		
	2	4049	2459	517	5859	2822	6168	4049	2459	1083	0		
	3	1	1										
								1	1	1			



프로젝트 설계

프로젝트 진행과정

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

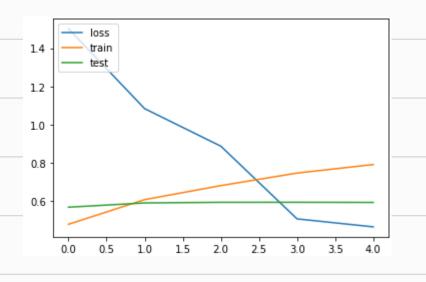
대시보드

fine-tuning 결과 (epoch=5)

감정 분류: 7, 데이터 분포 불균형

train_accuracy = 0.79 / test_accuracy = 0.59

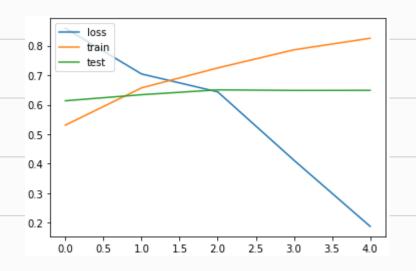
- 학습시킨 모델 결과는 7emotions_model.pt로 저장



감정 분류 : 6, 데이터 분포 균형

train_accuracy = 0.83 / test_accuracy = 0.65

- 학습시킨 모델 결과는 6emotions_model.pt로 저장



프로젝트 진행과정

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

대시보드

하고싶은 말을 입력해주세요 : 행복한 하루였어 행복이 느껴집니다.

하고싶은 말을 입력해주세요 : 출근하기 싫어 분노가 느껴집니다.

하고싶은 말을 입력해주세요 : 나는 불행해 슬픔이 느껴집니다.

하고싶은 말을 입력해주세요 :

▶ 하고싶은 말을 입력해수제요 :

어른들의 말이 문득 떠오른다...크면 다일게될꺼라는 말들이제야 문득문득 어릴때를 떠올려 보면 아! 맞구나!란 생각이 든다. 어렸을때는 잔소리로만 들었던 말들이 이제야 느껴진다하는 순간 내나이 어느덧 40줄 믿기지 않지만 세월이 이렇게 빨리 간다니. 내소중한 사람과 같이 있을 시간이 이제 생각만큼 길지 않게 됐다는 것을 느낀 나이가되어버렸다. 좀더 빨리 알아으면 좋았을것을 10년뒤 이글을 다시보게 된다면 그때는 어떤 기분이 들까?후회없이 살아 왔다고 느낄 수 있게 소중히 보내자. 10년뒤 20년뒤 메도후회하지 않도록 나와 가족들 친구들 그리고내 이웃 그리고 나를 아는 모든사람들과 평범하게 소박하게 살고 싶다.

슬픔이 느껴집니다

프로젝트 개요	프로젝트 설계	프로젝트 진행과정	프로젝트 결과물
데이터 수집 및 분석 데이터 전기	더리 감정 분류 모델	문장 생성 모델	대시보드
Transformer		SBERT	
Q: `3박 4일 정도 놀러가고	. 싶다'	Q: 데이터셋 전부	
인코더		임베딩	
Q: `3박 4일 정도 놀러가고	. 싶다'	임베딩 결과 데이터	선셋에 추가 문장 입력
디코더		임베딩 /	사전 임베딩
			 임베딩 결과 비교
	Q	А	→ 텍스트 유사도 score 저장
[예측] + [예측] +	+ [예측] _{일은 왜 해도}	해도 많이 힘드시겠	기정 표는 SCOIE의
	이번달에 또 급	급여가 급여가 줄어 =	속상 응답 출력
AI			

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

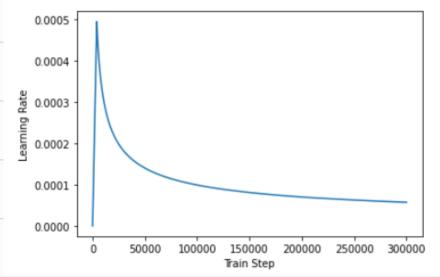
대시보드

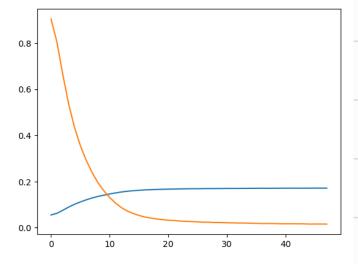
Transformer

	N	d_{model}	$d_{ m ff}$	$_{\text{I}}$ h	d_k	d_v	P_{drop}	ϵ_{ls}	train steps
base	6	512	2048	8	64	64	0.1	0.1	100K

Transformer - base

learning_rate : 1e-4





프로젝트 설계

<u>프로젝트 진행과정</u>

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

대시보드

BERT (응답 임베딩 사전)

embedding 결과를 저장해 책 처럼 사용

장점

Transformer와 비교하였을 때 똑같은 말을 따라하기 보다, 저장된 문장을 출력하기 때문에 자연스러운 느낌을 준다.

단점

일기내용과 비슷한 embedding 결과가 없다면 엉뚱한 문장을 출력한다.

📭 '종료' 입력시 프로그램 종료

Q : 오늘 안풀리던 일들이 조금씩 풀려서 안심이 됐다. 내일도 아무일 없이 지나갔으면 좋겠다

A: 하루속히 괴롭힘이 해결돼서 편안한 생활을 보낼 수 있기를 바랄게요

Q : 시험이 한달 남았다. 시험이 끝나자마자 하고 싶던거 다해야지

A: 응원해요. 시험 잘 봤으면 좋겠네요

Q : 지긋지긋한 장마 얼른 여름이 끝났으면 좋겠다

A : 즐거운 휴가를 보내면서 기분이 나아지길 바라요.

Q : 이번달 급여가 또 깎였어!

A : 좋은 소식이네요. 축하드려요.

Q: 내가 잘못한거 아닌데. 엄마는 언니편이야

A : 언니로서 해야 할 역할을 못 한다는 생각에 힘드시겠어요.

Q : 종료

프로젝트 진행과정

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델 `

대시보드

결과 분석 및 비교 (샘플링 문장)

장문 일기 내용 요약	Transformer	SBERT	
일하러 가기 싫다. 직장인의 일상적 고충	생각만해도 시원하네요 .	지금 상황이 어떻게 바뀌기를 바라시나요?	
과거에 대한 반성, 미애에 대한 다짐, 기대	천천히 지워 보세요	해왔던 것처럼 잘하시리라 믿어요	
아내의 생일날 아침을 준비하는 남편의 모습	새로운 시작 응원해요 !	정말 행복하실 것 같아요. 남편 분과 함께 행복한 노후가 되시면 좋겠어요.	

프로젝트 개요

프로젝트 설계

프로젝트 진행과정

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

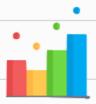
문장 생성 모델

대시보드



스트림릿(streamlit)이란?







Data ──►Insight ──► Deploy

상용툴: 어려운 커스텀

전문툴: 어려운 개발

장점

- 1. 간단한 사용법
- 백엔드를 작성하고, 경로를 정의하고,

HTTP 요청을 처리하고, 프론트엔드를 연결하고,

HTML, CSS, JavaScript 등을 작성할 필요가 없음

2. 다양한 API 제공

- streamlit에서 제공하는 API를 사용하여 몇 줄의 코드 만으로 손쉽게 앱 구현 가능

3. 즉시 배포

4

Streamlit: 둘다 해결

프로젝트 개요

프로젝트 설계

프로젝트 진행과정

프로젝트 결과물

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

Copy

문장 생성 모델

대시보드

%%writefile streamlit.py



ngrok Your Authtoken

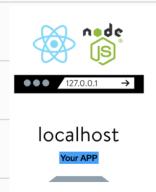
This is your personal Authtoken. Use this to authenticate the ngrok agent that you downloaded.



nohup: appending output to 'nohup.out'
<NgrokTunnel: "http://4bf0-34-87-103-57.ngrok.io" -> "http://localhost:80">



1 # ## streamlit 종료 후 초기화 해주기!! 2 ngrok.kill() # url 초기화









- 1. %%writefile로 py파일 생성
- 2. ngrok 개인 토큰번호 발급
- 3. ngrok.connect로 url연결
- 4. streamlit 종료 후 ngrok.kill()로 url 초기화

데이터 수집 및 분석

데이터 전처리

감정 분류 모델

문장 생성 모델

대시보드

Colab - Streamlit 연동 과정에서 발생한 문제

	문제점	방안	진행	결과
	colab 메모리 초과		1) csv 파일 -> Q&A(csv)	1) 메모리가 감소하여 초기 연결 시 connecting error는 해결
1	로 인한 connect error	csv 파일 용량 줄이기	-> 임베딩(npy) 2) csv, npy 파일 불러온 뒤 병합	2) 문장을 여러번 입력하거나 페이 지를 이동하면 결국 메모리가 쌓여 connecting error 발생!
2	페이지 로딩시간	1. model, data를 사용할 때만 호출 2. 감정 분류, 문장 생성이 끝난 뒤 메모리 삭제	1)일기 입력 함수안에 model, data load 코드를 작성 2) del 함수로 메모리 초기화	1. 페이지 로딩시간 단축 2. 메모리 감소

데이터 전처리 문장 생성 모델 대시보드 데이터 수집 및 분석 감정 분류 모델 Firebase Cloud Firestore I■ calendar **=** : **3** 02 🛜 finalpjt-14eb3 기쁨 + 문서 추가 + 컬렉션시작 + 컬렉션시작 20% 중립 calendar February + 필드추가 불안 91 user0001 w0: "[__ 당황 기쁨 기쁨 기쁨 기쁨]" 53,3% 13.3% 02 w1: "[기쁨 당황 짜증 기쁨 슬픔 당황 중립]" January w2: "[기쁨 기쁨 슬픔 16 17 18 19]" 6.67% w3: "[20 21 22 23 24 25 26]" 6.67% user0001 2023-02-15 finalpjt-14eb3 + 문서 추가 + 컬렉션시작 + 컬렉션시작 15 calendar 과거의 감정 + 필드추가 • • , user0001 2023-01-30 answer: "기다리던 월급을 받아서 정말 기쁘시겠어 2023-01-31 2023-02-02 daylog: "오늘은 월급날이야 신난다" · 2023-02-03 emotion: "기쁨" 2023-02-04 2023-02-05 2023-02-06 지난 일기를 보려면 날짜를 선택하세요 2023-02-07 2023-02-15

프로젝트 결과물

결과물

기대효과 및 전망

결과물

기대효과 및 전망



소감

• 김현주

3개월간의 배운내용을 바탕으로 프로젝트를 진행하면서, 기초부터 자세히 배우는 과정과 실제로 프로젝트를 하는 과정 모두가 중요하다는 것을 알게되는 시간이었습니다.

쉽진 않았지만 많은 보람을 얻는 시간이었습니다. 수강 기간 동안 많은 도움을 주신 강사님과, FT님, 그리고 팀원덕에 프로젝트를 잘 마무리 하게 된 것 같습니다.

• 박호준

이번 프로젝트를 통해 KoBERT, Streamlit을 처음 써봤는데, 어려운 부분이 많았지만 새로운 것들을 알아가는 재미가 더 컸던 것 같습니다.

부족한 실력이지만 팀원들과 단합하여 프로젝트를 끝마쳤다는 것이 자랑스럽고 앞으로의 일이나 프로 젝트에대한 자신감을 얻어가는 시간이었습니다.

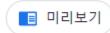
• 박준혁

인간 실격

다자이 오사무의 책 :

책 미리보기

161페이지 중 33페이지 보기 가능





>

• 조영훈

프로젝트를 진행하면서 많은 우여곡절이 있었지만,

팀원들이 마지막까지 힘을 내줘서 무사히 프로젝트 를 마칠 수 있었습니다.

각자 맡은 역할에 최선을 다해준 팀원들에게 감사 합니다!





THANK YOU

서비스 산업 데이터를 활용한 빅데이터 분석 8회차