

대한민국 ICT 인재를 키우는 힘,

2023년 한이음 ICT 멘토링 프로젝트 참여기 #2

빅데이터, AI를 활용한 감정 다이어리 앱 개발

강자원 컴퓨터시스템응용기술사 / KBS MNC(Media Network Center)팀
이단비·김가영·윤예진·이지은 부산대학교 IT응용공학과

이번 호에서는 2023년 4월부터 11월까지 대학생들과 함께 진행한 빅데이터와 인공지능 기술을 활용한 감정 분석 앱 개발 프로젝트에 대해 소개하는 시간을 가져보려 한다.



빅데이터, AI를 활용한 감정분석 다이어리란?

개인의 일상을 기록 및 공유하고 기록을 바탕으로 감정을 분석해 결과를 다양한 통계로 제공하며, 감정 개선을 위한 활동까지 추천하는 다이어리 애플리케이션이다. MZ세대 사이에서 ‘블로그’에 대한 인기가 높아지고 있다. 자신의 솔직한 생각을 ‘기록’하고 특정 사람들과 ‘공유’할 수 있기 때문이다. 자신만의 그룹을 만들어 ‘나’의 감정을 기록하고 공유할 수 있는 아늑한 공간을 만들고자 했다. 최근 정신건강의 어려움을 겪고 있는 청년의 수가 증가하는 상황이다. 이에 따라, 감정 일기와 노래 및 명언 추천(솔루션 제공)을 통해 스스로 감정을 인식하고 관리할 수 있게 도와주고자 한다.

프로젝트 성과

‘빅데이터, AI를 활용한 감정분석 다이어리’란 주제로 ‘2023년 한이음 ICT 멘토링 공모전’에 참가하여 입선작으로 선정되었다. 또한 한국정보처리학회의 ACK 2023 논문집(제30권 제2호)에 논문 게재의 성과도 얻을 수 있었다.

- 한이음 공모전 - 입선
- 한국정보처리학회 논문 게재

감정 일기를 통한 6가지 감정 분류 : LSTM모델 연구
(2023년 한국정보처리학회 학술대회논문집, 제 30권 2호)



팀원	이름	소속	부서/학과	역할	직위/학년
멘토	강자원	KBS	MNC	프로젝트 지도	감독
멘티	팀장	부산대학교	IT응용공학과	PM, Flask 서버 구축, 추천 시스템 개발	4학년
	팀원1			Android Studio (앱 제작), UI/UX 디자인	
	팀원2			AI 구현, UI/UX 디자인	3학년
	팀원3			Spring 서버, 데이터베이스 관리, 추천시스템 개발	

표 1. 프로젝트 수행 구성원

최종구현 화면

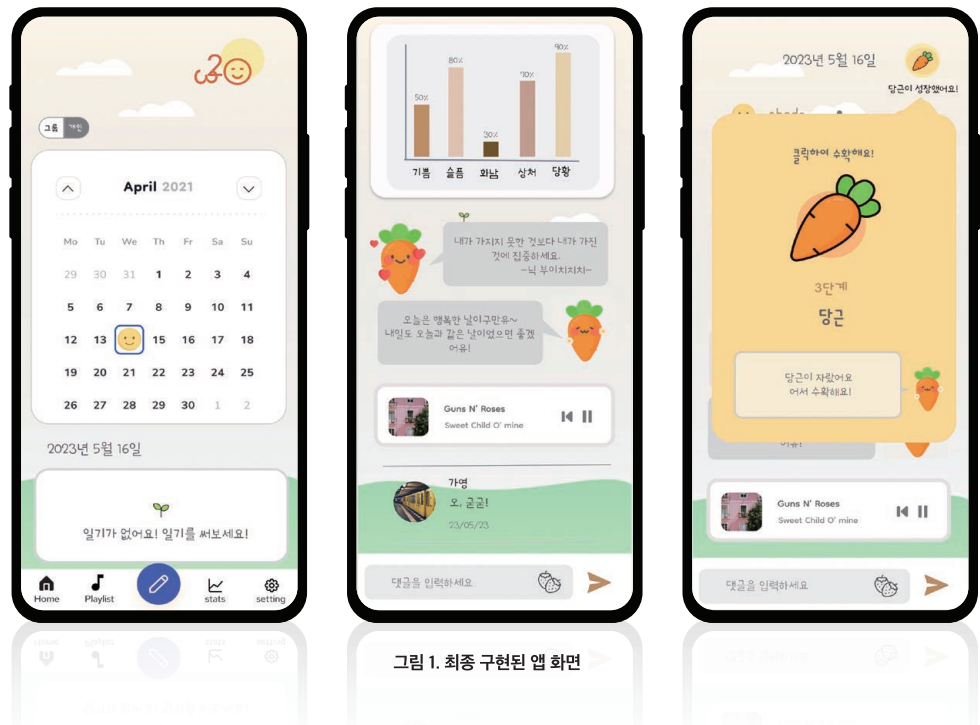


그림 1. 최종 구현된 앱 화면

애플리케이션 대표기능



- 기록 : 사용자가 양식대로 '감정 일기'를 작성한다.
- 공유 : 나의 일상을 친구들과 '공유'한다
- 진단 : 일기를 바탕으로 'AI가 감정을 분석'해 공감의 말을 전한다.
- 보상 : 일기를 기록하면 '농작물'이 성장하고, 수확한 농작물로 상대방의 일기에 '좋아요'를 누르는 등 참여할 기회가 많아진다.
- 분석 : 개인별, 그룹별 분석된 감정의 변화 추이 '시각화 그래프'를 제시한다.
- 공감 : 그룹원들끼리 '댓글'을 남기거나 '좋아요'를 누를 수 있다.
- 음악 추천 : 분석한 감정을 바탕으로 적합한 '노래'를 추천한다.
- 명언 추천 : 분석한 감정을 바탕으로 적합한 '명언'을 추천한다.

프로젝트 개요 및 핵심 기술

시스템 흐름도

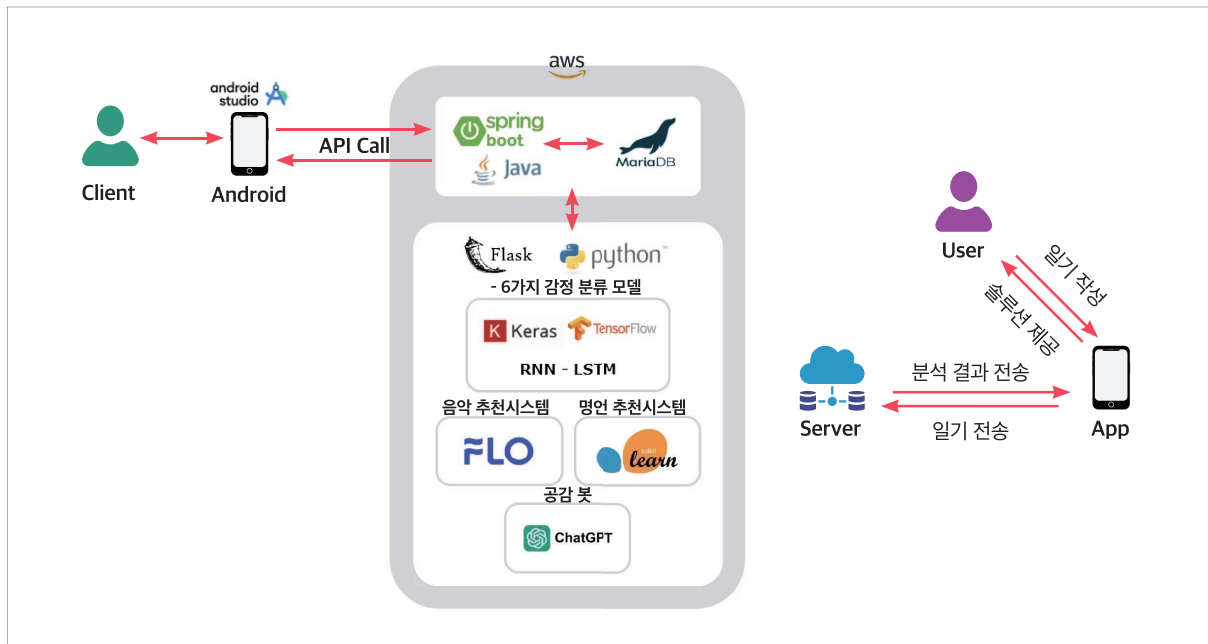


그림 2. 전체 시스템 흐름도 소개

- 클라이언트는 회원, 친구목록, 캘린더, 타임라인 API를 요청한다.
- Spring Boot 서버에서 회원, 친구목록, 캘린더, 타임라인 API를 응답한다.
- 클라이언트가 작성한 일기를 Flask 서버로 전송한다.
- Flask 서버에서 텍스트 마이닝을 통해 감정을 분석하고, 분석 결과를 바탕으로 사용자에게 적절한 음악과 명언을 추천한다.
- 이용자의 Web Browser(W/B)에서 Web Server(W/S)로 요청을 보내고, 감정분석 등 AI 기술이 필요한 요청일 경우 Web Server for AI(AIS)로 전달한다.
- AI는 사용자의 감정 다이어리 내용을 기반으로 개인화된 학습 및 자기 개선 계획을 제안할 수 있다. 이를 통해 사용자는 자신의 감정을 관리하고, 새로운 습관을 형성하며, 지속적인 성장을 이루어낼 수 있다.

앱 개발에 적용된 알고리즘과 핵심 기술

☑ 머신러닝 인공지능 알고리즘 - NLP(Natural Language Processing)

컴퓨터에 인간의 언어를 이해시키기 위해 자연어처리(NLP) 기술을 이용한다. 각 단계에서 이전 단계의 출력값을 다음 단계의 input으로 사용하며 순환적인 연산을 수행하는 순환 신경망 모델을 사용한다. 감정 분석 훈련 데이터는 문장과 그 문장이 지닌 감정(예 : 긍정적, 부정적 또는 중립적 감정)으로 구성된다. 머신러닝 알고리즘은 이러한 데이터 세트를 읽고 문장이 입력되면 그 문장의 감정을 반환하는 모델을 생성한다. 이렇게 문장이나 문서가 입력되면 해당 입력값 대한 레이블을 반환하는 종류의 모델로 자연어처리 기술을 사용한다.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	연령	성별	상황키워드	신체질환	감정	대문(환)명	소문명	사람문장1	시스템문장1	사람문장2	시스템문장2	사람문장3	시스템문장3		
2	1	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
3	2	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
4	3	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
5	4	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
6	5	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
7	6	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
8	7	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
9	8	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
10	9	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
11	10	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
12	11	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
13	12	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
14	13	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
15	14	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
16	15	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
17	16	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
18	17	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
19	18	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
20	19	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							
21	20	여성	진로 취업 1	학당입금	분노	노약자라는 말을 듣게 해 줘서	행복스런 내가	혼자 해결하기 힘들면 주위에 의논할 사람							

그림 3. 전처리를 통해 분류된 데이터셋

☑ 생성형 인공지능 엔진 - KoGPT2 (Korean GPT-2)

사용자가 작성한 일기 데이터를 전처리하고 토큰화하여 특징값을 추출하고 Konlpy 한국어 형태소 분석기를 이용하여 중요한 단어, keywords를 선별한다. 선별된 단어를 크롤링하여 적절한 명언을 추천한다. 또한, KoGPT2 모델을 이용하여 사용자의 감정에 공감하는 자연스러운 문장(공감봇)을 생성하도록 한다.

개념 정리

※ KoGPT2 : 2020년에 SK텔레콤에서 GPT-2를 파인 튜닝하여 개발한 모델

※ 파인튜닝(Fine Tuning) : 기존 모델을 변형하여 새로운 모델을 만드는 행위

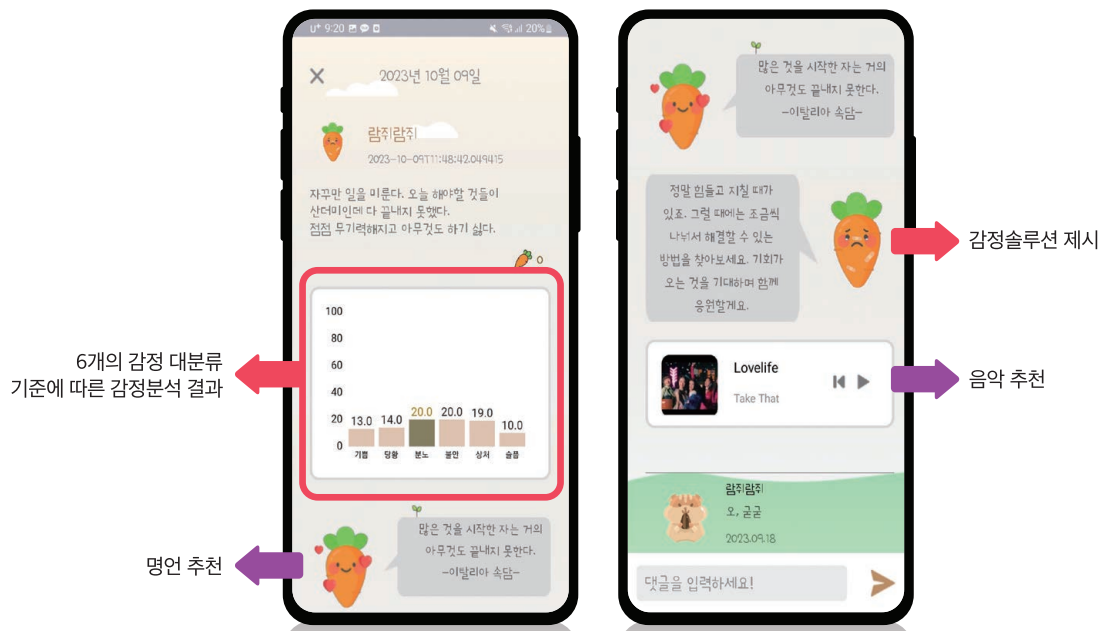


그림 4. 인공지능 기술이 적용된 화면 예시

☑ 감정분석을 위한 딥러닝 학습 모델 - LSTM(Long Short Term Memory)

‘string punctuation’를 이용하여 물음표, 온점 등 불필요한 문장 부호를 사용하지 않도록 사용자가 입력한 데이터를 전처리한다. 학습용 데이터와 테스트용 데이터를 분류하는 과정에서 ‘stratify’ 파라미터를 활용하여 층화추출을 통해 감정별 데이터가 골고루 포함될 수 있도록 한다. 모델 학습에는 시계열 데이터를 학습하는 방법 중 RNN의 문제점을 보완한 LSTM 모델과 GRU 모델을 비교군으로 두어 더 높은 정확도를 가진 LSTM을 활용하여 결과가 도출될 수 있도록 하였다.

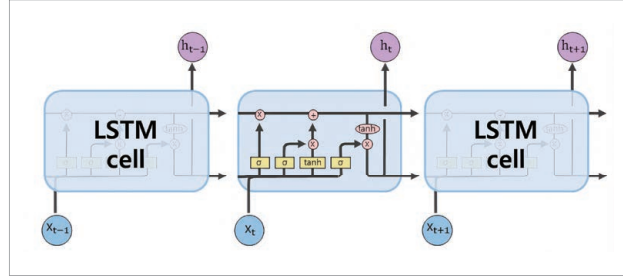


그림 5. LSTM 모델의 기본 구조

☑ LSTM 모델을 선택을 위한 GRU 모델과 비교 실험과정

AI hub에서 제공하는 데이터 중 일반인 1,500명 대상의 인터뷰 및 크라우드 소싱을 수행한 ‘감성 대화 말뭉치’ 데이터를 이용하였다. 총 51,630개의 데이터로 이루어져 있으며 ‘불안’, ‘분노’, ‘상처’, ‘슬픔’, ‘당황’, ‘기쁨’의 6개의 감정 대분류와 각각 대분류에 속해있는 여러 개의 감성 소분류로 라벨링되어 있다.



그림 6. 감정 분류에 따른 이미지

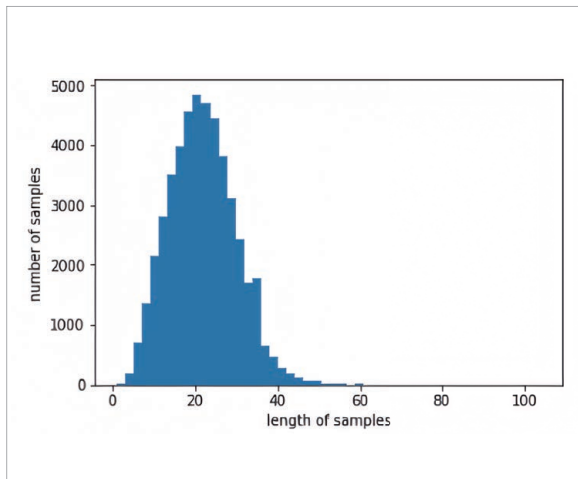
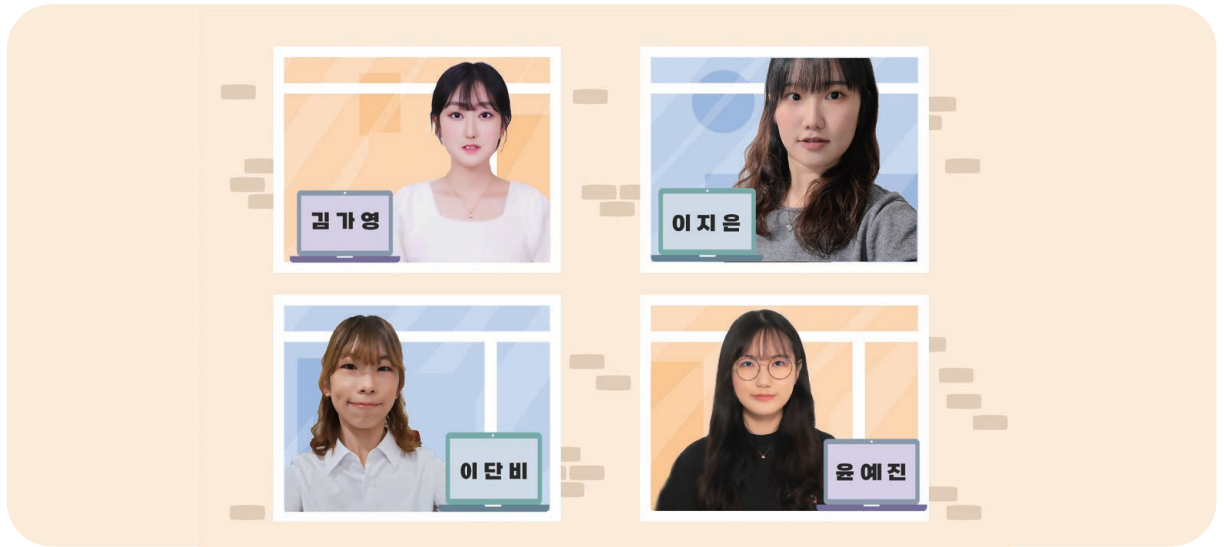


그림 7. Okt 형태소 분석기를 활용한 학습용 데이터의 문장 길이 시각화 결과

타깃값인 ‘불안’, ‘분노’ 등 각 감정을 0부터 6까지의 숫자로 Label Encoding 하고, 단어들을 수치화하기 위해 형태소 분석기 Okt(Open-source Korean Text Processor)를 사용하여 토큰나이징 작업을 시행하였다. 학습을 위해 문장의 길이를 동일하게 맞추기 위해 문장의 최대 길이를 확인하고 패딩을 통해 공간을 채운다. 전처리한 데이터로 감정분석을 위한 모델링을 진행한다. 한글 데이터를 정제하고 각 단어를 벡터화한 후 파이썬의 넘파이 배열 데이터를 외부 파일로 저장하였다.

빠른 학습을 위해 LSTM과 GRU 모델을 간단하게 구축하여 텍스트 분류를 진행한 결과 정확도가 높지 않았다. 두 모델 중 테스트 데이터에 대한 정확도가 더 높았던 LSTM 모델을 기준으로 모델의 은닉층을 추가하고, 모델의 optimizer를 Adam(Adaptive Moment Estimation)으로 수정하고 dropout을 0.3으로 수정한 결과 기존 코드 대비 7% 이상의 정확도를 향상할 수 있었다.

프로젝트 완료 후기



김가영 이 프로젝트를 통해 다양한 문제를 경험하면서 팀원들과의 소통과 위기 극복에 대한 중요성을 깨달았습니다. 이전에는 작은 프로젝트를 주로 수행하고 있었고, 애자일 방법론을 선호하여 설계를 간소화하고 개발에 중점을 두었습니다. 그러나 이 프로젝트를 통해 워터폴 방법론에 대한 이해와 경험을 얻을 수 있었습니다. 기획 단계에서 여러 문서들을 작성하면서 탄탄하고 명확한 기획이 완성도 있는 프로그램을 만들어내는 중요성을 다시 깨닫게 되었습니다. 초반에는 서툰 부분도 많았지만, 멘토님의 지원과 팀원들 간의 노력 덕분에 프로젝트를 성공적으로 마무리할 수 있었습니다. 이번 프로젝트 경험은 그 어떤 프로젝트 보다 잊지 못할 가치 있는 경험이 될 것 같습니다.

윤예진 이 프로젝트를 통해 요구사항 정의서, 유스케이스 다이어그램, 서비스 흐름도 등의 문서를 직접 작성하며 개발뿐만 아니라 소프트웨어 설계 과정의 중요성을 알게 되었고, 소프트웨어 공학적 지식을 적용할 기회가 되었습니다. 기술 구현에 크고 작은 어려움들을 겪으며 다른 팀원들보다 개발 속도가 늦어지기도 하였으나 여러 시행착오 끝에 감정 분류 모델을 성공적으로 구현할 수 있었고 논문 제출, 공모전 입선을 통해 프로젝트를 잘 마무리할 수 있었습니다. 좋은 팀원들과 함께한 덕분에 프로젝트가 보다 풍성하고 효과적으로 진행된 것 같아 기쁘고 도움 주신 멘토님께도 감사드립니다.

이단비 PM이라는 생소한 역할을 맡아 팀장 역할을 제대로 해내지 못한 것 같아 아쉬운 마음이 들었습니다. 백엔드를 제대로 다뤄본 경험이 처음이었지만 API를 직접 만들며 성취감을 많이 느꼈던 시간이었습니다. 팀원들 덕분에 같이 의기투합하여 결과물을 만들어내서 팀원들에게 매우 감사합니다. 또, PM이라는 역할의 중요성과 문서 작업의 중요성을 알려주신 멘토님께도 진심으로 감사 인사드립니다.

이지은 평소 감정 관리에 관심이 있었지만, 우울이나 화남 등 사람의 감정을 분석해서 솔루션을 준다는 것이 처음엔 모호했습니다. 하지만 팀원들과 주제를 구체화하고 실질적인 개발을 진행하면서 점점 재미를 느꼈습니다. 또, 디자인과 개발, 다양한 문서 작업을 통해 실무적인 개발 과정을 경험할 수 있어 유익했습니다. 첫 협업이라 막막함과 우여곡절이 있었지만 프로젝트가 올바른 방향으로 갈 수 있게 도와주신 멘토님과 끝까지 함께한 당다리 팀원들에게 감사합니다! 