SWE3028-Fall-2024
Capstone Design Project

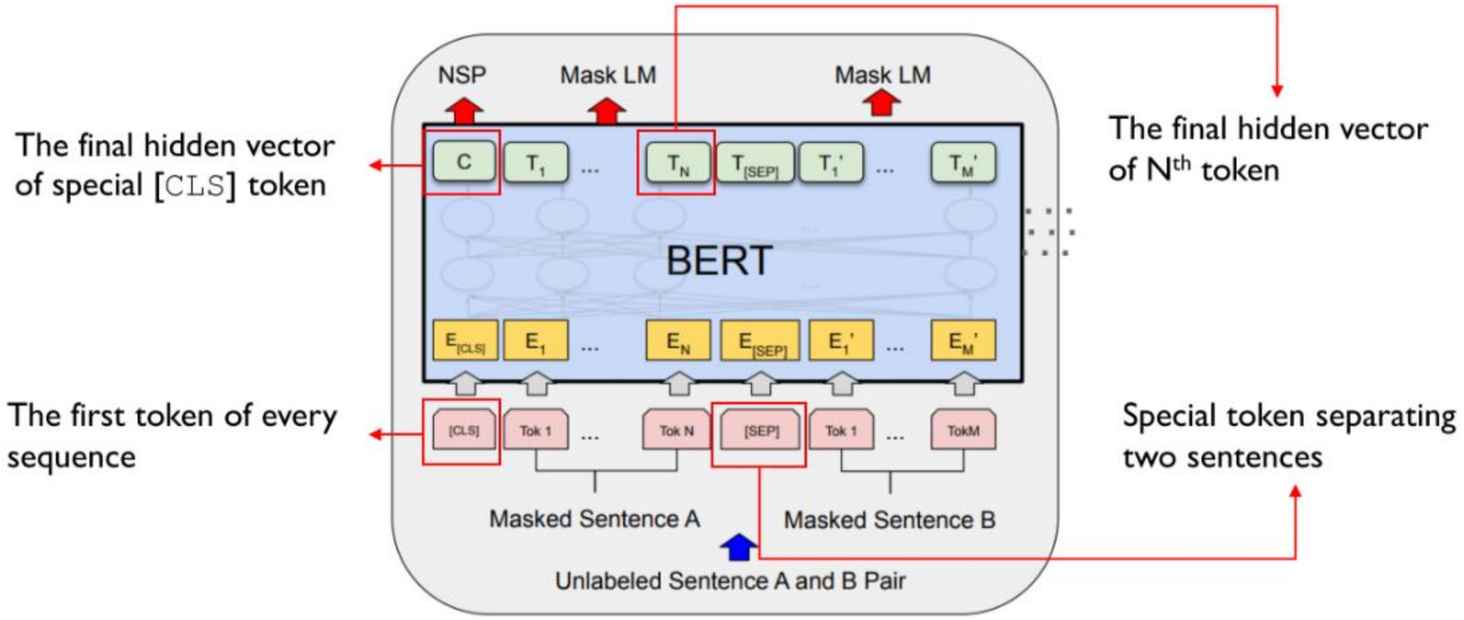
NLP-based Psychological Diagnostic Diary

Weekly Progress Meeting (1)

런닝머신 팀(TEAM C)
김나영 김예담 박준현 방기호

CONTENTS

01. Technical Background - BERT



A stack of Transformer Encoders

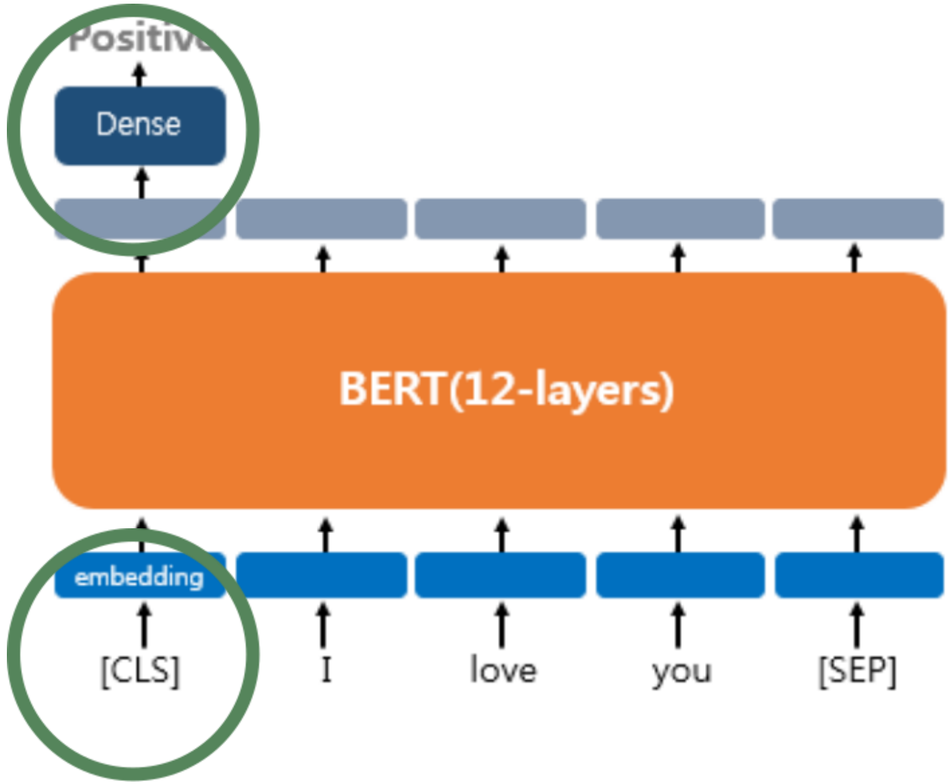
Working with fine-tuned BERT

01. Technical Background - BERT

Fine-Tuning

Improve performance
in diverse downstream tasks

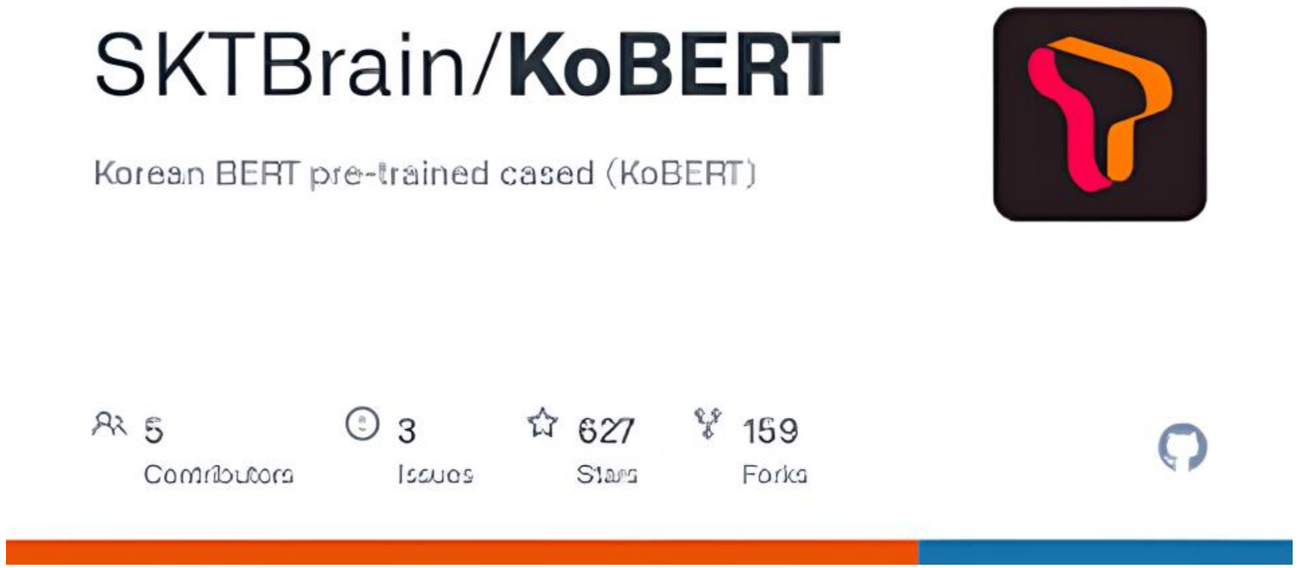
Single Text Classification



01. Technical Background - BERT

<limitation>

01. Technical Background - KoBERT



A version of BERT fine-tuned for Korean

01. Technical Background - KoBERT

input: 오늘 하루는 정말 힘들었지만 그래도 기분이 나쁘지 않았다

[CLS] 오늘 하루는 정말 힘들었지만 그래도 기분이 나쁘지 않았다 **[SEP]**

1. Wordpiece Embedding: The input sentences are **tokenized** and divided into **subwords**.

[CLS] 오늘 하루 ##는 정말 힘 들 ##었 ##지만 그래도 기분 ##이 나쁘 ##지 않 ##았 ##다 **[SEP]**

→ This allows the model to partially handle not only frequently appearing words

but also first-time complex words.

01. Technical Background - KoBERT

[CLS] 오늘 하루 ##는 정말 힘 들 ##었 ##지만 그래도 기분 ##이 나쁘 ##지 않 ##았 ##다 [SEP]

2. Position Embedding: Information about how the tokens are **arranged** within the sentence is added.
→ It plays an important role in **grasping the context**.

3. Segment Embedding: **Different sentences** are distinguished.
→ It recognizes the differences between sentences by distinguishing several sentences.

01. Technical Background - KoBERT

4. KoBERT - 12 Transformer layers

Each layer learns how a particular word relates to another word in a sentence
through a **self-attention**.
→ The [CLS] token is transformed into a vector
that reflects the **full context** of multiple input sentences.

01. Technical Background - KoBERT

[CLS]

5. Dense layer: It selects the final emotion by transforming the vector from the **[CLS]** token.

① **Linear Transformation**: transforms dimensions by applying weights and biases according to the number of emotion labels.

② **Softmax**: The probability for each emotion class is calculated and predicted as the emotion with the highest probability value.

01. Technical Background

Why?

BERT

Bidirectional to understand context

back and forth

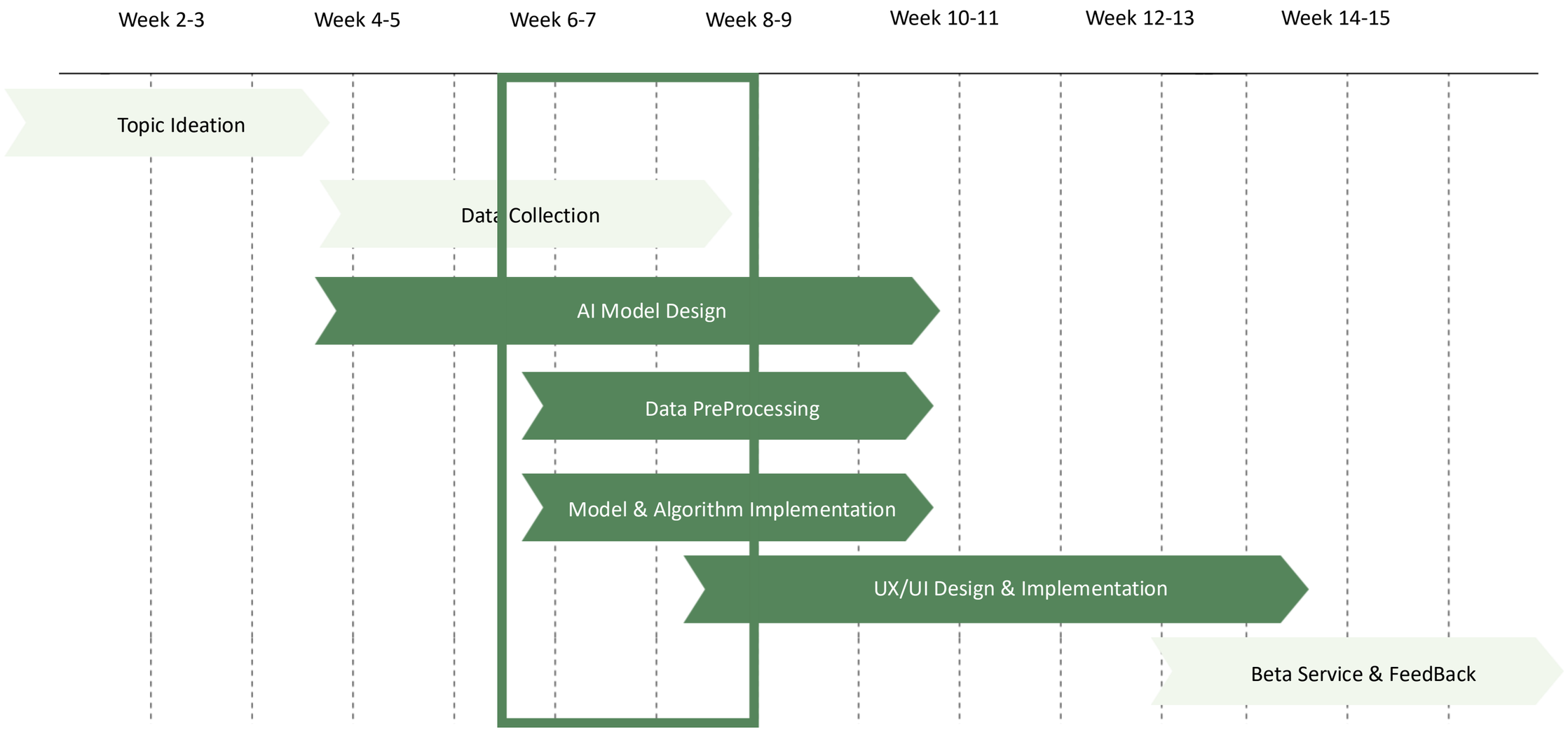
+

KoBERT

Optimized for Korean

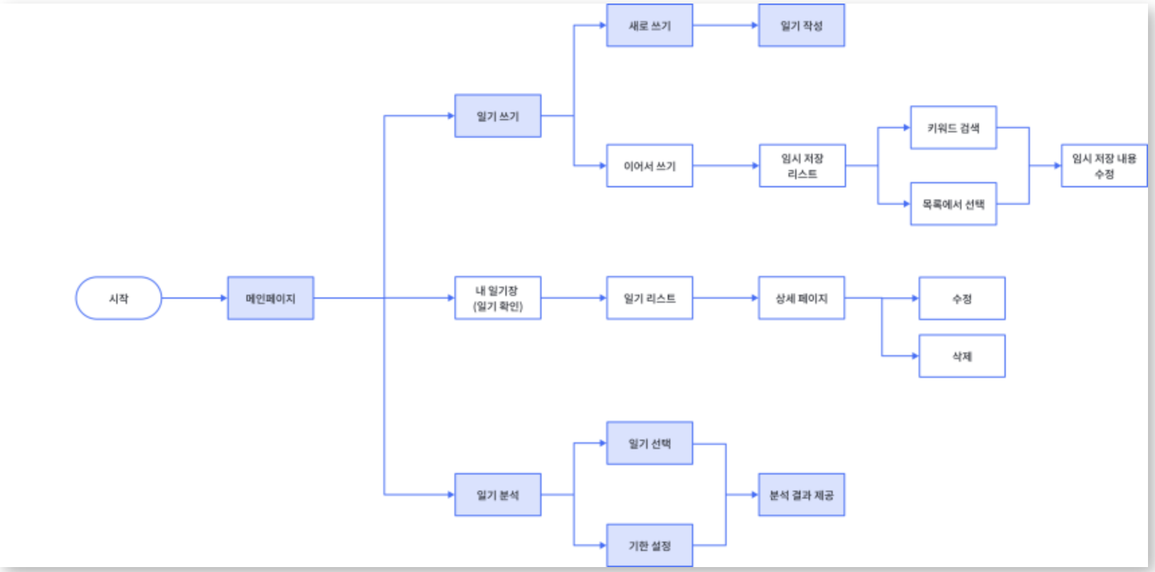
reflecting irregular language changes in Korean

02. Progess - Expected Progess



02. Progress - UX / UI Design

Flow Chart

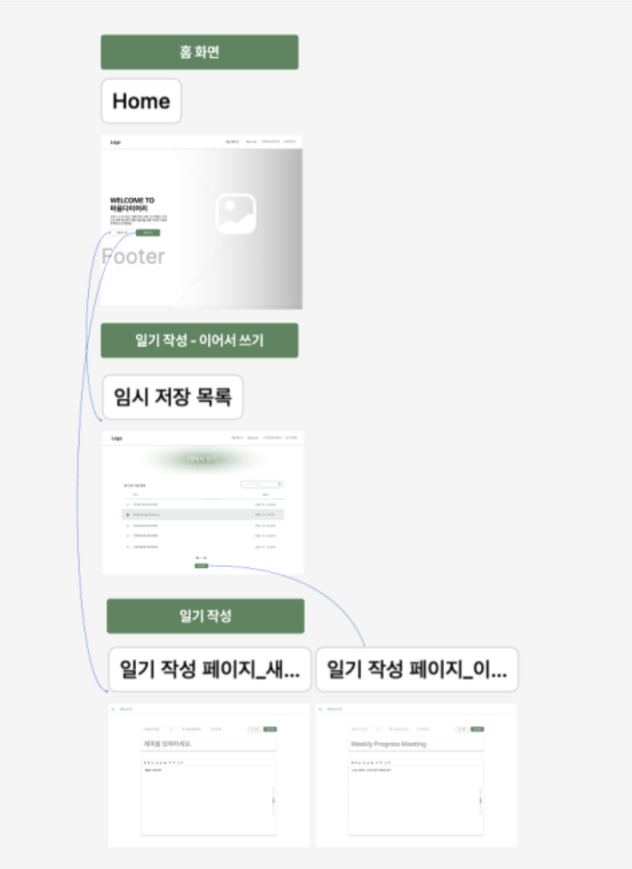


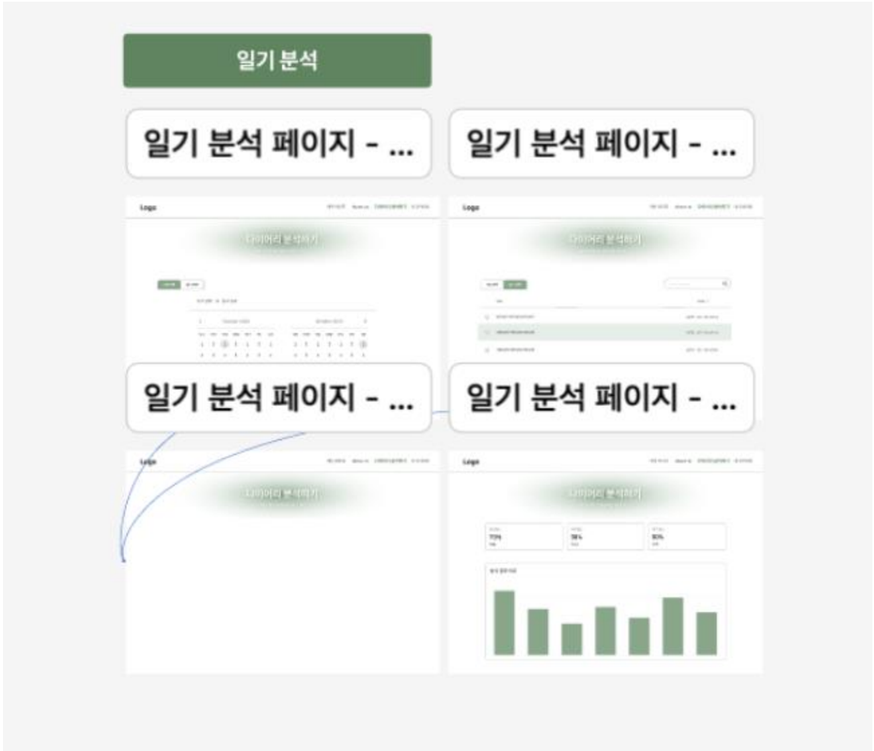
IA

1 Depth	2 Depth	3 Depth	4 Depth	Contents
메인 페이지				간단한 서비스 소개, 메인 기능
	일기 작성			
		새로 작성		날짜, 날씨, 내용 작성 +@(추후 구현)
		이어서 작성		
			일기 리스트	검색, 날짜 선택
	내 일기장			
		일기 리스트		
			일기 상세	작성 내용, 작성 시점, 분석 내용(분석했던 일기의 경우), 수정 및 삭제 버튼
	일기 분석			
		일기 선택		
			결과 제공	분석 근거, 분석 결과

02. Progress - UX / UI Design

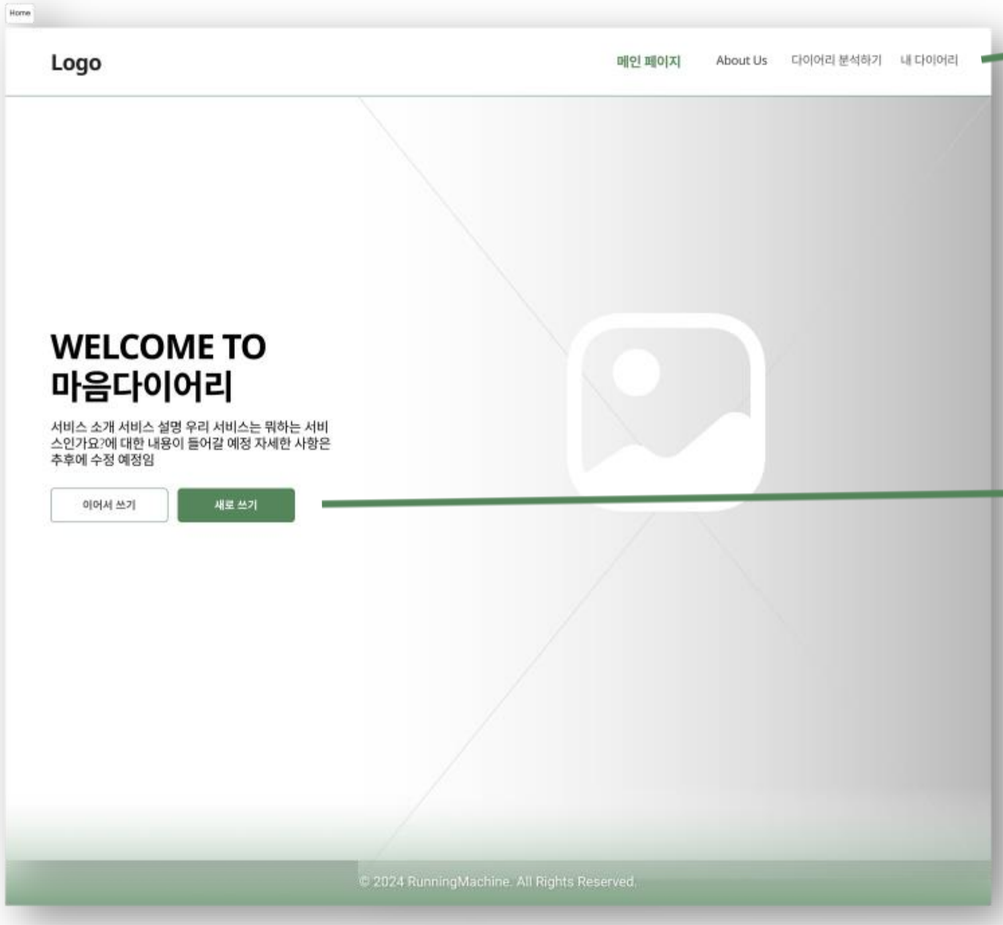
Wireframe





02. Progress - UX / UI Design

<Main Page>



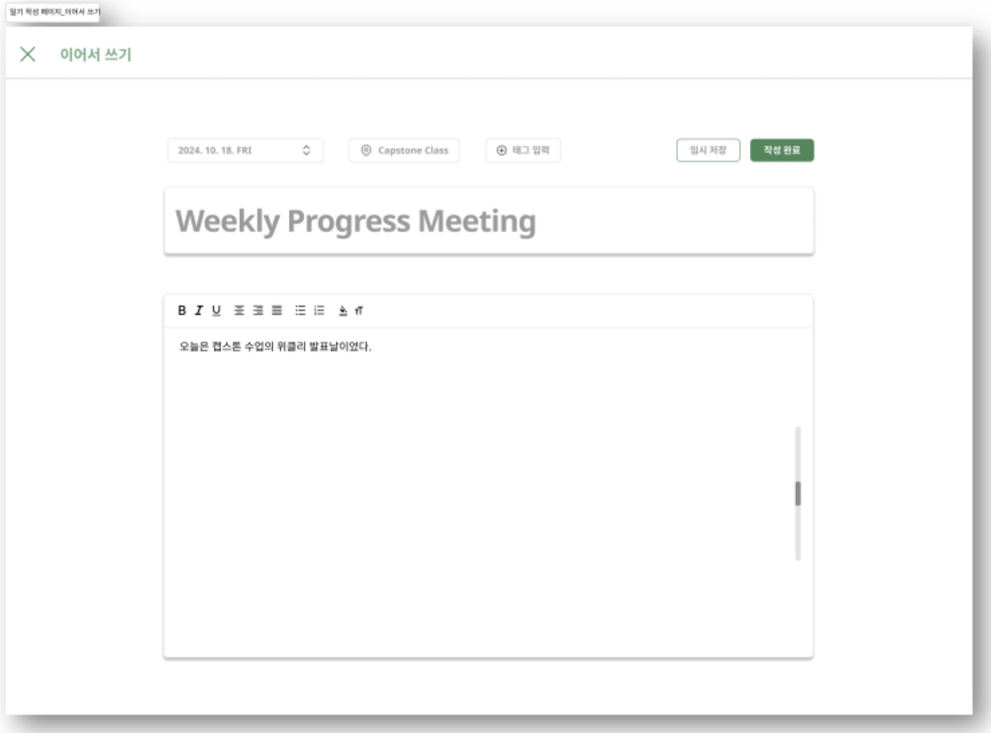
Navigation Bar

Button to Writing Page

02. Progress - UX / UI Design

<Writing Page>

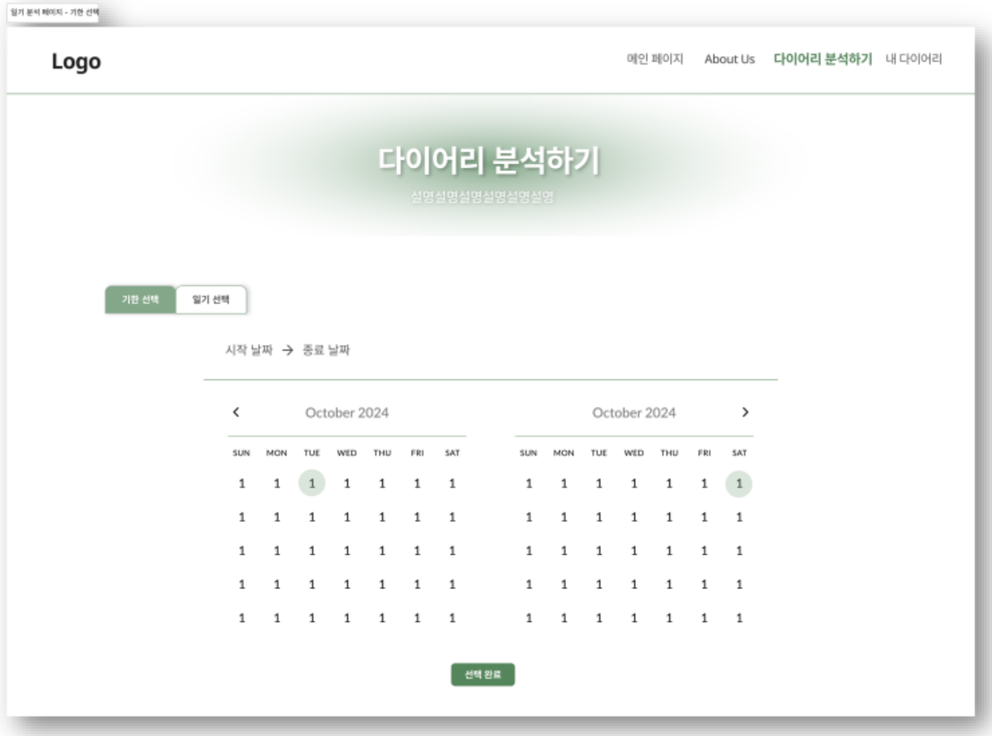
Create New Journal



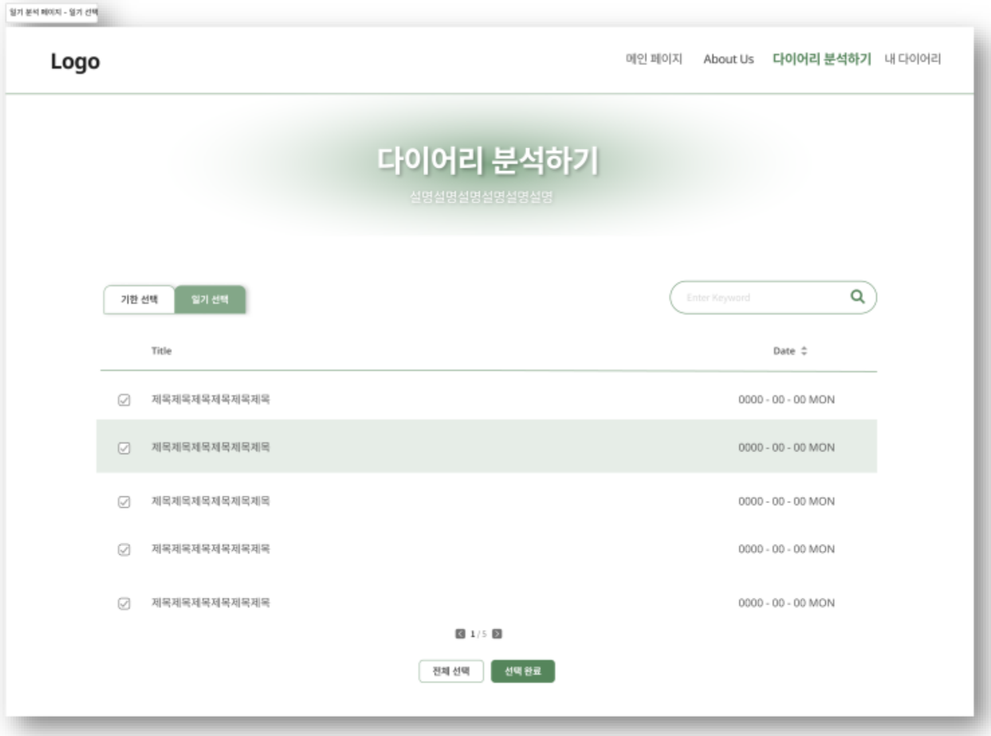
Use Drafted Journal

02. Progress - UX / UI Design

<Analyzing Page>



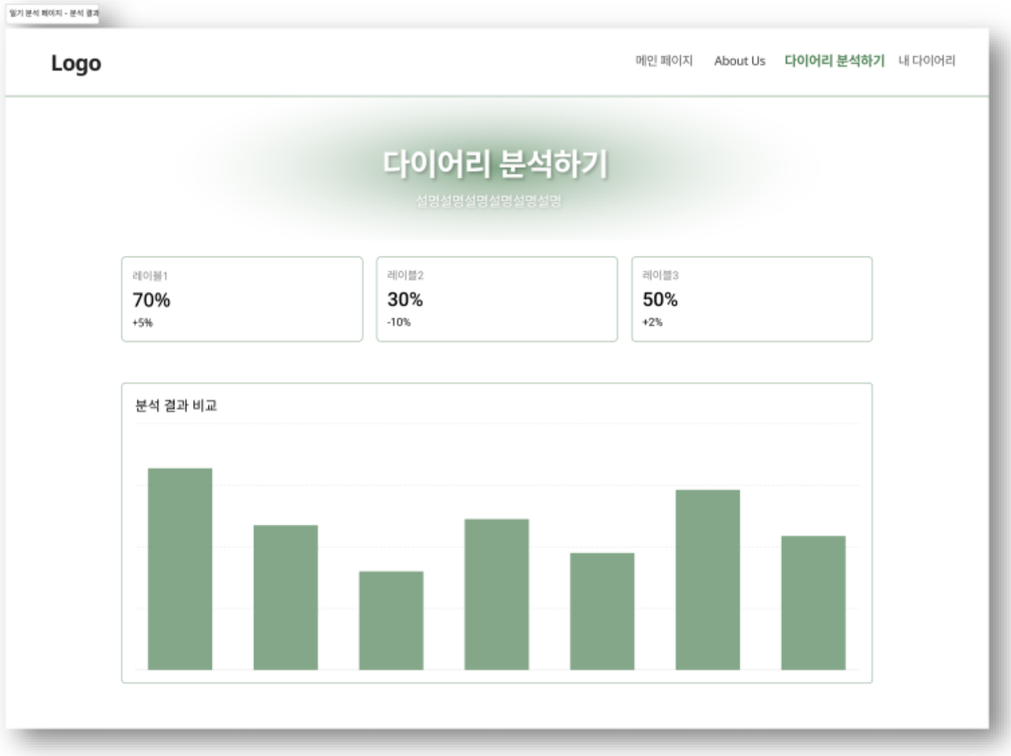
Choosing by Date



Choosing Manually

02. Progress - UX / UI Design

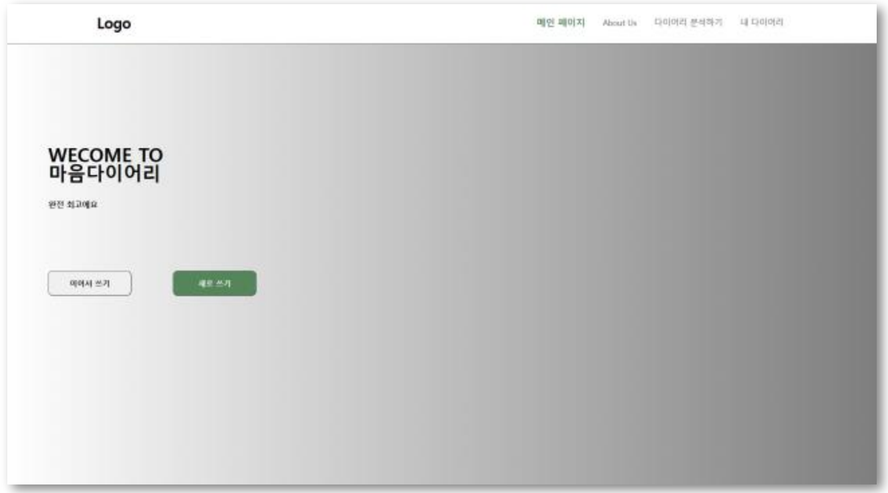
<Analyzing Page>



02. Progress - Front-end

Prepare a Framework

Implement Several Pages



02. Progress - AI

Data Collection

Wellness Conversation Script Dataset

From 16,000 **counsel data** received from Gangnam Severance,
5,232 user utterances were extracted
when the patient himself/herself visited the hospital.

Depression / Sadness / Loneliness / Anger / Lethargy / Abnormal Emotional Control
Sense of Loss / Loss of Appetite / Increase in Appetite / Insomnia
Nervousness / Fatigue / Feeling of Guilt / Loss of Concentration
Decrease in Confidence / Decline in Self-esteem
Feeling of Despair / Suicidal Impulse / Feeling of Insecurity

19 emotional labels in total

02. Progress - AI

Data Preprocessing

Before

무기력	우울하고 무기력해.
	우울해서 움직일 힘이 하나도 없어.
	너무 우울해서 움직일 기운도 없다.
	너무 우울해서 움직일 힘도 없어요.
	너무 우울하고 무기력해요.
	우울하고 기력도 없네요.
	우울해서 움직이고 싶은 마음도 없어요.
	우울하고 무기력감이 느껴져.
	마음이 우울하니까 힘도 안 나요.
	마음이 우울하니까 뭔가를 할 수 있는 힘이 안 생겨.
	우울해서 움직일 힘도 없다.

Utterances separated
by clinical keywords

After

utterance	label
우울해	0
너무 우울	0
우울해죽겠	0
기분이 우	0
우울모드임	0
우울함	0
우울해 미	0
나 우울함.	0

그래서 퇴	18
나이프의	18
이런 제가	18
엄청 숨이	18
사무실에	18
혼자있을	18
원래 안 그	18
근데 일을	18
불안하고,	18

Numbered emotion label
tagged for each utterance

02. Progress - AI

Pipeline Constructing

```
Input your sentence (or type 'exit' to stop): 오늘 정말 힘든 일이 많았다. 지갑도 잃어버리고, 시험에서 50점밖에 받지 못했다.
Predicted Emotion: nervousness, Confidence: 0.1605
----------------------------------------------------------

Input your sentence (or type 'exit' to stop): 종료
```

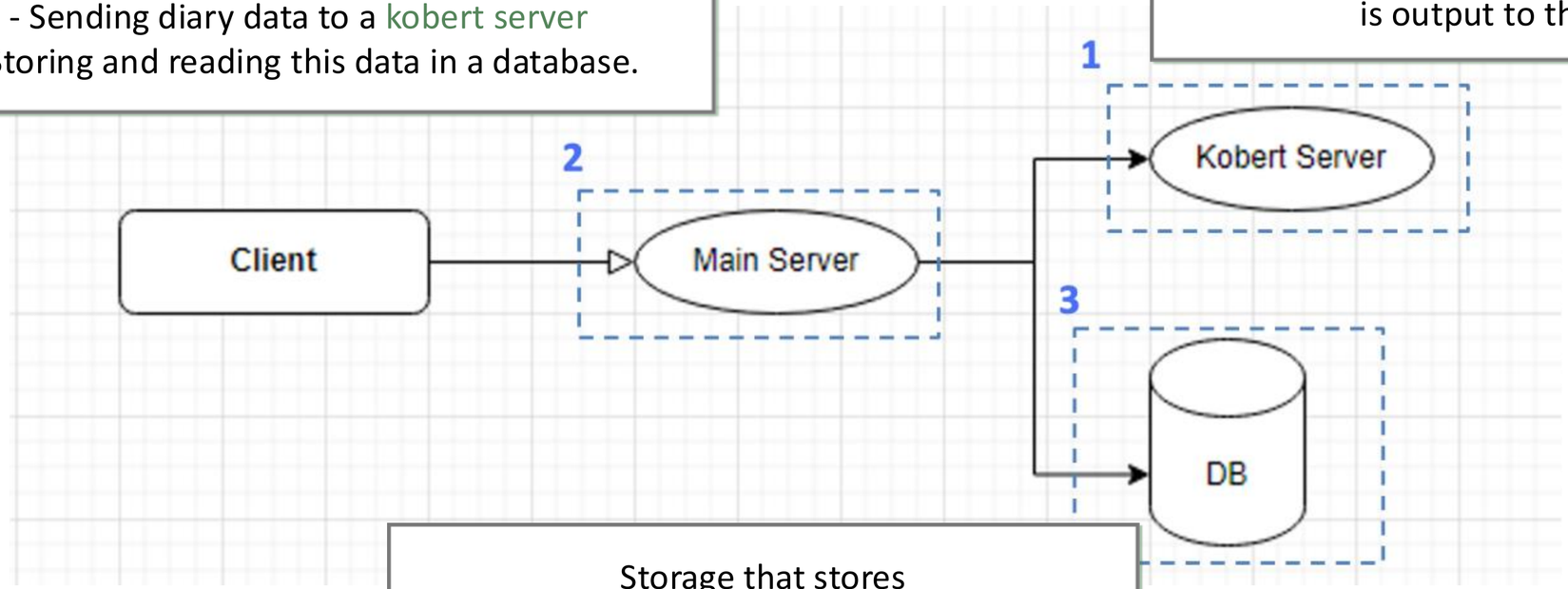
A baseline pipeline was constructed
using pre-trained KoBERT models provided by SKT Brain.
The above is the result of testing with the built simple emotion classifier.
(https://github.com/SKTBrain/KoBERT)

02. Progress - Back-end

Architecture

A server that performs major functions
- Sending diary data to a kobert server
- Storing and reading this data in a database.

Input diary data using KoBERT,
emotional data extracted from the text
is output to the server.



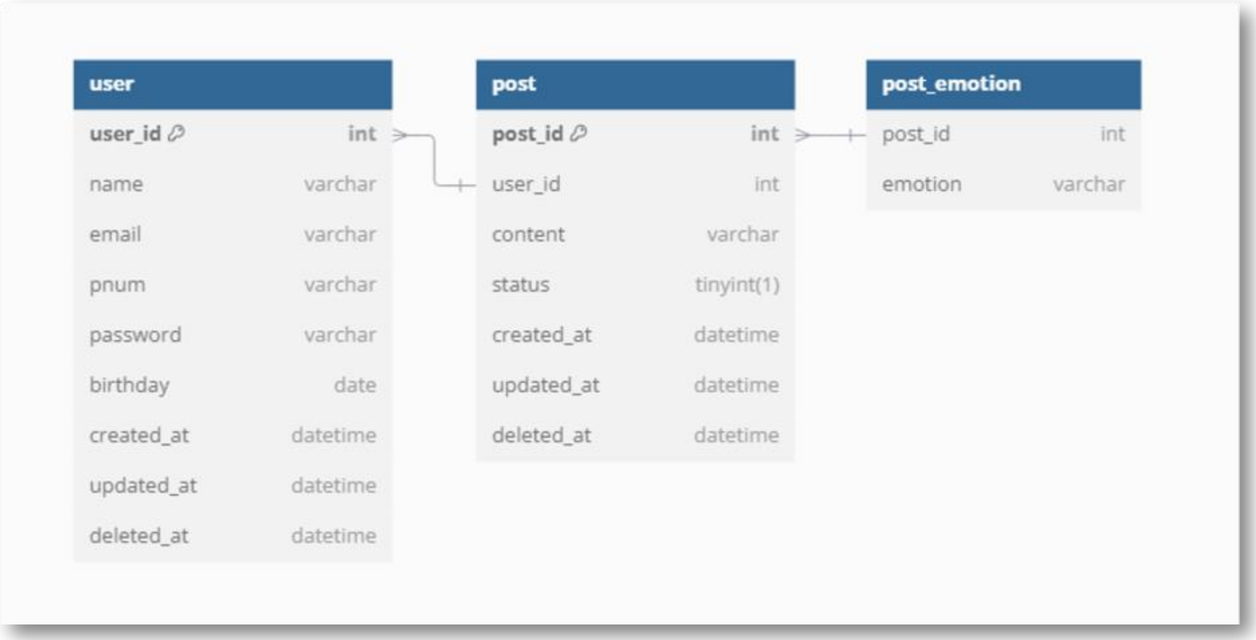
Storage that stores
user information, diary data, etc.

02. Progress - Back-end

API List

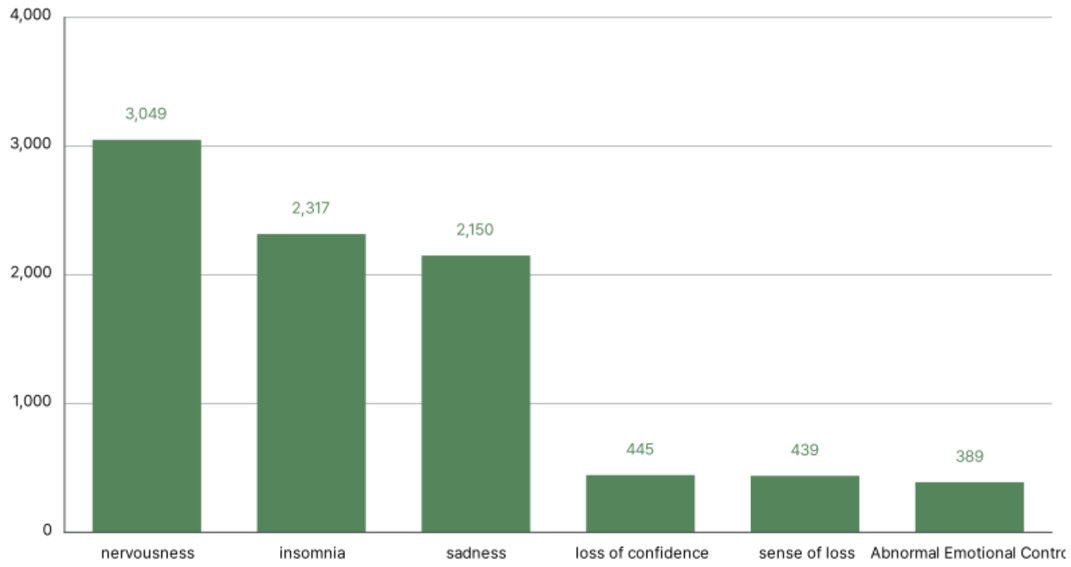
- user
  - GET get user
  - POST create user
  - POST login
- main
  - GET getPostList
  - GET getPost
  - POST createPost
  - PUT editPost
  - DEL deletePost
  - GET getPostEmotions
  - GET getPostListEmotions
  - POST createPostEmotions

ERD



03. Challenging Problem

Optimize Performance



Imbalance in the number
of utterances by emotion label

Data Augmentation

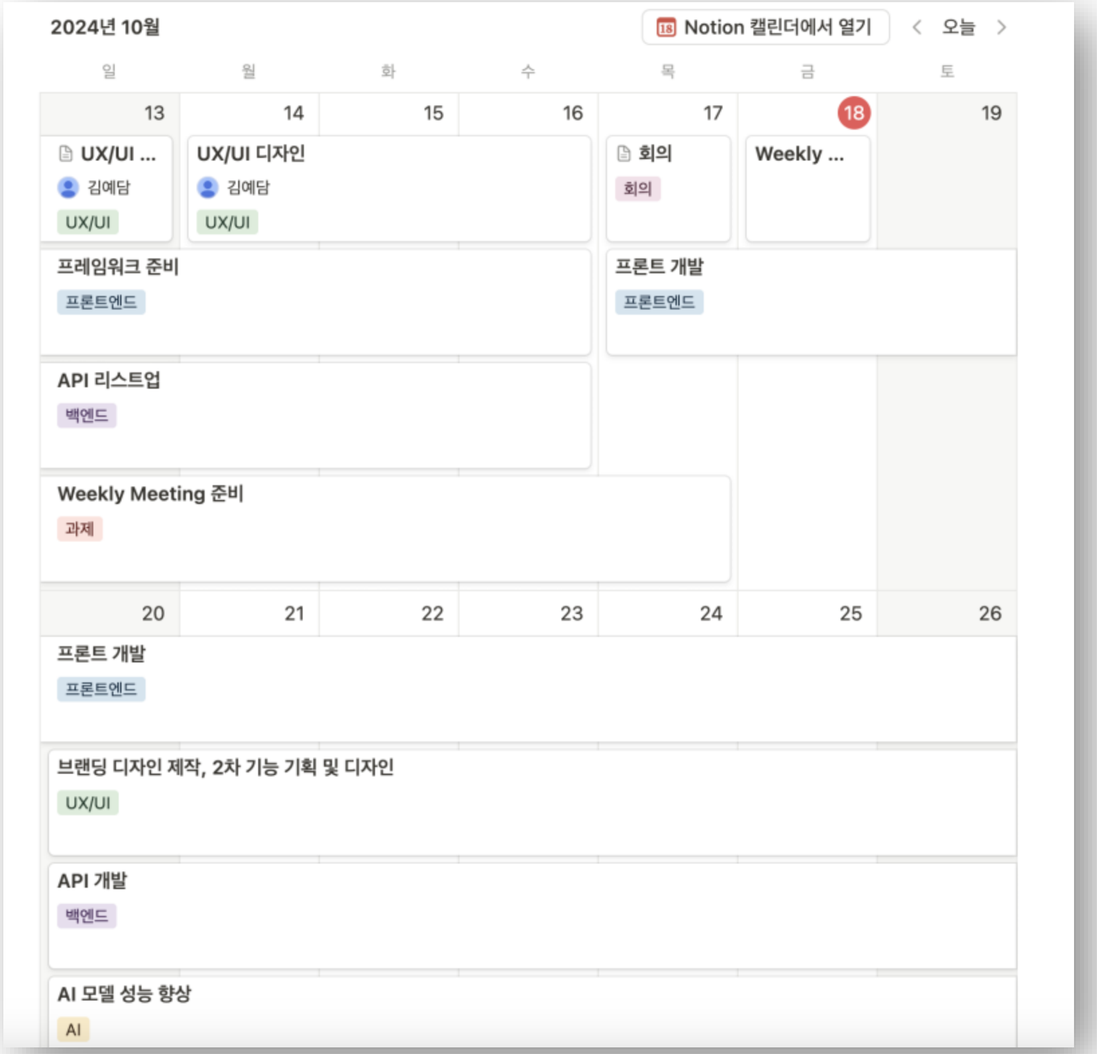
03. Challenging Problem

Hyperparameter Tuning

```
kobert_config = {
    'attention_probs_dropout_prob': 0.1,
    'hidden_act': 'gelu',
    'hidden_dropout_prob': 0.1,
    'hidden_size': 768,
    'initializer_range': 0.02,
    'intermediate_size': 3072,
    'max_position_embeddings': 512,
    'num_attention_heads': 12,
    'num_hidden_layers': 12,
    'type_vocab_size': 2,
    'vocab_size': 8002
}
```

Plan to use various deep learning techniques to find the best combination of parameters

04. Plan


2024년 10월
Notion 캘린더에서 열기
오늘
일
월
화
수
목
금
토
13
14
15
16
17
18
19
UX/UI ...
김예담
UX/UI
UX/UI 디자인
김예담
UX/UI
회의
회의
Weekly ...
프레임워크 준비
프론트엔드
프론트 개발
프론트엔드
API 리스트업
백엔드
Weekly Meeting 준비
과제
20
21
22
23
24
25
26
프론트 개발
프론트엔드
브랜딩 디자인 제작, 2차 기능 기획 및 디자인
UX/UI
API 개발
백엔드
AI 모델 성능 향상
AI

04. Plan

UX/UI Design

- Specify
  Detailed Features and Policy
- Create a Branding Design
  Logo CI, Color Chip, Typography
- Planning Additional Features and
  Building Wireframes

Front-end

- Fetch to the Back-end
  of the entered parameters
- Implementation
  of an Unimplemented Pages
- Implementation
  of Detailed Features

AI

- Structural modification and
  Performance improvement

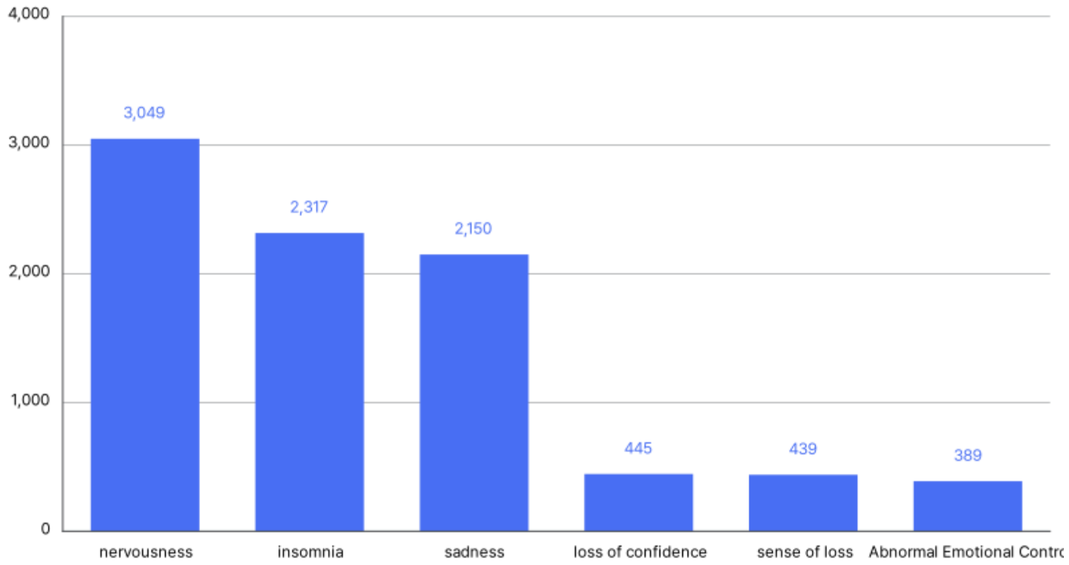
Back-end

- Developing API

Q & A

NLP-based Psychological Diagnostic Diary

04. AI / Back-end Progress - AI: performance




```
kobert_config = {
    'attention_probs_dropout_prob': 0.1,
    'hidden_act': 'gelu',
    'hidden_dropout_prob': 0.1,
    'hidden_size': 768,
    'initializer_range': 0.02,
    'intermediate_size': 3072,
    'max_position_embeddings': 512,
    'num_attention_heads': 12,
    'num_hidden_layers': 12,
    'type_vocab_size': 2,
    'vocab_size': 8002
}
```

1. Data Augment

There is an imbalance in the number of utterances by emotion label. Since this can harm the accuracy of the model, we will try the data augmentation.

2. Hyper-parameter Tuning

We will use various deep learning techniques to find the best combination of parameters.