```
from google.colab import drive
import pandas as pd
import json


# 구글 드라이브 마운트
drive.mount('/content/drive')


```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
six_label_all_text = pd.read_csv('/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/six_label_all_text')
#six_label_first_text = pd.read_csv('/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/six_label_first_text')
four_label_all_text = pd.read_csv('/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/four_label_all_text')
#four_label_first_text = pd.read_csv('/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/four_label_first_text')
```

```
raw_data = pd.read_excel('/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/Training_221115_add/원천데이터/감성대화말뭉치(최종데이터)_Training.xlsx')
raw_data

```

	Unnamed: 0	연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	사람문장1	시스템문장1	사람문장2	시스템문장2	사람문장3	시스템문장3
0	1	청년	여성	진로,취업,직장	해당없음	분노	노여워하는	일은 왜 해도 해도 끝이 없을까? 화가 난다.	많이 힘드시겠어요. 주위에 의논할 상대가 있나요?	그냥 내가 해결하는 게 나아. 남들한테 부담 주고 싶지도 않고.	혼자 해결하기로 했군요. 혼자서 해결하기 힘들면 주위에 의논할 사람을 찾아보세요.	NaN	NaN
1	2	청년	여성	진로,취업,직장	해당없음	분노	노여워하는	이번 달에 또 급여가 깎였어! 물가는 오르는데 월급만 자꾸 깎이니까 너무 화가 나.	급여가 줄어 속상하시겠어요. 월급이 줄어든 것을 어떻게 보완하실 건가요?	최대한 지출을 억제해야겠어. 월급이 줄어들었으니 고정지출을 줄일 수밖에 없을 것 같아.	월급이 줄어든 만큼 소비를 줄일 계획이군요.	NaN	NaN
2	3	청년	여성	진로,취업,직장	해당없음	분노	노여워하는	회사에 신입이 들어왔는데 말투가 거슬려. 그런 애를 매일 봐야 한다고 생각하니까 스...	회사 동료 때문에 스트레스를 많이 받는 것 같아요. 문제 해결을 위해 어떤 노력을 ...	잘 안 맞는 사람이랑 억지로 잘 지내는 것보단 조금은 거리를 두고 예의를 갖춰서 대...	스트레스받지 않기 위해선 인간관계에 있어 약간의 거리를 두는 게 좋겠군요.	NaN	NaN
3	4	청년	여성	진로,취업,직장	해당없음	분노	노여워하는	직장에서 막내라는 이유로 나에게만 온갖 심부름을 시켜. 일도 많은 데 정말 분하고 ...	관련 없는 심부름을 모두 하게 되어서 노여우시군요. 어떤 것이 상황을 나아질 수 있...	직장 사람들과 솔직하게 이야기해보고 싶어. 일하는 데에 방해된다고.	직장 사람들과 이야기를 해 보겠다고 결심하셨군요.	NaN	NaN
4	5	청년	여성	진로,취업,직장	해당없음	분노	노여워하는	얼마 전 입사한 신입사원이 나를 무시하는 것 같아서 너무 화가 나.	무시하는 것 같은 태도에 화가 나셨군요. 상대방의 어떤 행동이 그런 감정을 유발하는...	상사인 나에게 먼저 인사하지 않아서 매일 내가 먼저 인사한다고!	항상 먼저 인사하게 되어 화가 나셨군요. 어떻게 하면 신입사원에게 화났음을 표현할 ...	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
51625	51626	노년	남성	재정	만성질환 무	분노	성가신	나이가 먹고 이제 돈도 못 벌어 오니까 어떻게 살아가야 할지 막막해. 능력도 없고.	경제적인 문제 때문에 막막하시군요. 마음이 편치 않으시겠어요.	아무것도 할 수 없는 내가 무가치하게 느껴지고 실망스러워.	지금 할 수 있는 가장 합리적인 행동은 무엇인가요?	노년층을 위한 경제적 지원이나 부업 같은 것도 알아보아야겠어.	좋은 결과 얻으시길 바랄게요.
51626	51627	노년	여성	재정	만성질환 유	불안	초조한	몸이 많이 약해졌나 봐. 이제 전과 같이 일하지 못할 것 같아 너무 짜증 나.	건강에 대한 어려움 때문에 기분이 좋지 않으시군요. 속상하시겠어요.	마음 같아서는 다 할 수 있는 일인데 이젠 몸이 안 따라와 주니 화만 나.	어떻게 하면 지금의 기분을 나아지게 할 수 있을까요?	남편과 함께 게이트볼이나 치러 가야겠어. 그럼 기분이 나아질 것 같아.	남편과 함께하는 좋은 외출 시간 되시길 바랄게요.

다음 단계: raw_data 변수로 코드 생성 New interactive sheet

```
SAVE_PATH= '/content/drive/MyDrive/프로젝트/2025_과외_프로젝트1/data/'
```

```
raw_data=raw_data.drop('Unnamed: 0', axis=1)
```

```
teen_data=raw_data[raw_data['연령']=='청소년']
```

```
teen_data.head()
```

	연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	사람문장1	시스템문장1	사람문장2	시스템문장2	사람문장3	시스템문장3
346	청소년	남성	학업 및 진로	해당없음	분노	좌절한	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다.	음악 실기 시험을 망쳐서 기분이 안 좋으시겠어요.	보통 공부를 못하면 노래라도 잘 부르는데 나는 노래도 못 불러.	공부도 못하고 노래도 못 부르시는군요. 앞으로 음악 실기를 망치지 않으려면 어떻게 ...	노래를 연습하면 실기 성적이 좋아질까?	노래를 연습해 보려고 하시는군요.
347	청소년	남성	학업 및 진로	해당없음	분노	좌절한	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네.	친구가 전시 그림을 이틀 만에 끝냈군요. 어떤 기분이 드셨나요?	난 왜 이렇게 느린지 좌절스러워.	느려서 좌절을 느끼셨군요. 이 기분을 어떻게 낫게 할 수 있을까요?	나는 작업을 꼼꼼하게 한다고 생각할래.	꼼꼼하게 작업하는 거라고 생각하려 하시는군요.
348	청소년	남성	학업 및 진로	해당없음	분노	좌절한	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아.	피아노를 아무리 연습해도 실력이 늘지 않으시는군요. 마음이 어떠신가요?	속상하고 좌절스러워. 난 아무래도 피아노에 재능이 없나 봐.	피아노에 재능이 없는 것 같아 속상하시군요. 도움을 줄 수 있는 사람이 없을까요?	피아노 선생님한테 도움을 청해 보고 싶어.	피아노 선생님께 도움을 청하려 하시는군요.
349	청소년	여성	학교폭력/따돌림	해당없음	분노	노여워하는	반 애들이 나만 따돌려서 너무 화가 나.	친구들이 따돌려서 화가 나시는군요.	도대체 나한테 왜 이러는 걸까?	나에게 이러는 이유를 알고 싶으시군요. 이유를 알 수 있는 방법이 있을까요?	글쎄. 처음부터 날 따돌린 건 아니었는데 무슨 오해를 하고 있나?	친구들과 오해가 빨리 풀렸으면 좋겠어요.
350	청소년	여성	학교폭력/따돌림	해당없음	분노	노여워하는	친구들이 내 말을 무시하는 게 화가 나.	친구들이 말을 무시해서 정말 화가 나겠어요.	왜 내 말을 무시한 거지?	친구들이 말을 무시하는 이유를 알고 싶으시군요. 이유를 알아낼 방법이 있을까요?	또 그러면 다음에는 직접 물어봐야겠어.	다음에는 이유를 직접 물어보려 하시는군요.

다음 단계: teen_data 변수로 코드 생성 New interactive sheet

```
teen_data=teen_data.drop(columns=['연령','성별','신체질환','시스템문장1','시스템문장2','시스템문장3'])
```

```
len(teen_data['상황키워드'])
```

```
10582
```

```
keywords_series = teen_data['상황키워드'].reset_index(drop=True)
```

데이터

```
six_label_all_text_phq9=pd.concat([six_label_all_text, keywords_series], axis=1)
```

```
six_label_all_text_phq9=six_label_all_text_phq9.drop(columns='Unnamed: 0')
```

```
six_label_all_text_phq9.head()
```

	text	감정_대분류	상황키워드
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	분노	학업 및 진로
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	분노	학업 및 진로
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	분노	학업 및 진로
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	분노	학교폭력/따돌림
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	분노	학교폭력/따돌림

다음 단계: six_label_all_text_phq9 변수로 코드 생성 New interactive sheet

```
#six_label_first_text_phq9=pd.concat([six_label_first_text, keywords_series], axis=1)
```

```
#six_label_first_text_phq9=six_label_first_text_phq9.drop(columns='Unnamed: 0')
```

```
#six_label_first_text_phq9.head()
```

```
four_label_all_text_phq9=pd.concat([four_label_all_text, keywords_series], axis=1)
```

```
four_label_all_text_phq9=four_label_all_text_phq9.drop(columns='Unnamed: 0')
```

```
four_label_all_text_phq9.head()
```

	text	감정 분류	상황키워드
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	우울 신호	학업 및 진로
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	우울 신호	학업 및 진로
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	우울 신호	학업 및 진로
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	일반 부정	학교폭력/따돌림
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	일반 부정	학교폭력/따돌림

다음 단계: four_label_all_text_phq9 변수로 코드 생성 New interactive sheet

```
#four_label_first_text_phq9=pd.concat([four_label_first_text, keywords_series], axis=1)
```

```
#four_label_first_text_phq9=four_label_first_text_phq9.drop(columns='Unnamed: 0')
```

```
#four_label_first_text_phq9.head()
```

PHQ-9

```
direct_core = [
    "자살 생각이 든다", "죽고싶다", "죽는게 낫다",
    "자해 생각이 든다", "자해할 생각이다"
]
```

```
core_list = [
    "우울하다", "희망이 없다", "즐겁지 않다",
    "행동이 느려졌다", "말이 느려졌다",
    "일상에 집중을 못한다", "실패했다",
    "가족을 실망 시켰다", "기운이 없다", "피곤하다"
]

```

```
indirect_list = [
    "흥미가 없다", "흥미가 떨어지다",
    "식욕이 줄다", "입맛이 없다", "많이 먹다",
    "잠들기 어렵다", "잠을 너무 많이 잔다",
    "가만히 있질 못한다", "안절부절하다", "잘못하고있다"
]
```

```
phq9_symptoms={'direct_core':direct_core,'core_list':core_list,'indirect_list':indirect_list }
```

```
phq9_keys = list(phq9_symptoms.keys())
```

```
phq9_keys
```

```
['direct_core', 'core_list', 'indirect_list']
```

키워드 점수 가중치

- 진행안함 유사도 점수만 돌렸을때가 오히려 최종 점수로 돌렸을때 보다 결과가 더 좋음

형태소 분석

- phq9_symptoms 단어들과 정제된 텍스트 데이터를 각각 형태소 분석해준다

```
# #Mecab 설치 (Colab)
# !pip install konlpy
# !pip install mecab-python3
# !bash <(curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/scripts/mecab.sh)
```

```
# from konlpy.tag import Mecab

# # Mecab 객체 생성
# mecab = Mecab()
```

```
# pos_phq9={}

# for i in phq9_keys:
#     #print(phq9_symptoms[i])

#     steam_list=[]

#     for j in phq9_symptoms[i]:
#         #print(j)
#         pos=mecab.pos(j)
#         #print(pos)

#         steam=[]
#         for word, tag in pos:

#             #print(word, tag)

#             if tag.startswith('VV') or tag.startswith('VA'): # VV(동사), VA(형용사)의 경우 어간만 사용
#                 steam.append(word)
#             elif tag.startswith('NN'):  # NNG(일반명사), NNP(고유명사) 등 명사의 경우 단어 자체를 사용
#                 steam.append(word)

#             #print(steam)

#         steam_list.append(' '.join(steam))
#         #print(steam_list)

#     pos_phq9[i]=steam_list


# print(pos_phq9)
```

```
# pos_phq9
```

```
# def pos(i):

#     steam_list=[]

#     pos=mecab.pos(i)
#     #print(pos)

#     steam=[]
#     for word, tag in pos:

#         #print(word, tag)

#         if tag.startswith('VV') or tag.startswith('VA'): # VV(동사), VA(형용사)의 경우 어간만 사용
#             steam.append(word)
#         elif tag.startswith('NN'):  # NNG(일반명사), NNP(고유명사) 등 명사의 경우 단어 자체를 사용
#             steam.append(word)

#         #print(steam)

#     return ' '.join(steam)

```

```
# six_label_all_text_phq9['text_pos']=six_label_all_text_phq9['text'].apply(lambda x: pos(x))
```

```
# six_label_all_text_phq9
```

phq9_score 생성

```
# phq9_keys
```

```python
# import numpy as np

# def create_phq9_vector(text):
#     phq9_secore=0

#     for i,v in enumerate(phq9_keys):

#         if v =='direct_core':

#             for j in pos_phq9[v]:

#                 if j in text:
#                     phq9_secore+=3.0


#         elif v =='core_list':

#             for j in pos_phq9[v]:

#                 if j in text:
#                     phq9_secore+=2.0


#         else:

#             for j in pos_phq9[v]:

#                 if j in text:
#                     phq9_secore+=1.0

#     return  phq9_secore

```

```python
# six_label_all_text_phq9['phq9_score']=six_label_all_text_phq9['text_pos'].apply(lambda x: create_phq9_vector(x))
```

```python
# six_label_all_text_phq9
```

의미 유사도 가중치

텍스트 데이터 벡터화

```python
# max_length 정하기

all_text_max_length=161
first_text_max_length =62
```

```python
batch_size=32
max_length=all_text_max_length
```

```python
import torch
import torch.nn.functional as F
from transformers import AutoTokenizer, AutoModel
from tqdm import tqdm
import numpy as np
from tqdm import tqdm


device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print(f"Using device: {device}")

MODEL_NAME = "klue/roberta-base"
tokenizer = AutoTokenizer.from_pretrained(MODEL_NAME)
klue = AutoModel.from_pretrained(MODEL_NAME).to(device)
```

```
Using device: cuda
Some weights of RobertaModel were not initialized from the model checkpoint at klue/roberta-base and are newly initialized: ['pooler.dense.bias', 'pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```python
def encode_texts(texts, max_length, batch_size):

    if hasattr(texts, 'tolist'):
        texts = texts.tolist()

    with torch.no_grad():

        text_vecs=[]
        for i in tqdm(range(0, len(texts), batch_size)):

            batch=texts[i:i+batch_size]

            token=tokenizer(batch,
                    padding=True,  # 짧은 문장은 패딩 추가
                    truncation=True,  # 긴 문장은 잘라냄
                    max_length=max_length,  # 최대 256토큰
                    return_tensors="pt"  # 파이토치 텐서로 반환
                )


            input_ids=token['input_ids'].to(device)
            attention_mask=token['attention_mask'].to(device)

            output=klue.forward(input_ids, attention_mask )
            last_hidden_state=output.last_hidden_state


            mask=attention_mask.unsqueeze(-1).float() # 패딩 제외하기
            mean=(last_hidden_state*mask).sum(dim=1)/mask.sum(dim=1).clamp(min=1e-9) # 평균 계산
            mean=F.normalize(mean, p=2, dim=1) # 단위 벡터로 만들기

            text_vecs.append(mean.detach().cpu())

        text_vecs=torch.cat(text_vecs, dim=0) # 리스트=> 텐서화


    return text_vecs
```

```python
encode_texts(six_label_all_text_phq9['text'], max_length, batch_size)
```

```
100%|██████████| 331/331 [00:40<00:00,  8.11it/s]
tensor([[ 0.0118, -0.0319, -0.0625,  ...,  0.0341, -0.0270, -0.0193],
        [-0.0230, -0.0296, -0.0496,  ...,  0.0127, -0.0217, -0.0005],
        [-0.0071, -0.0654, -0.0354,  ..., -0.0049, -0.0245,  0.0103],
        ...,
        [-0.0006, -0.0424, -0.0302,  ..., -0.0108, -0.0255, -0.0236],
        [-0.0056, -0.0479, -0.0203,  ..., -0.0009, -0.0633, -0.0080],
        [-0.0008, -0.0683, -0.0631,  ...,  0.0032, -0.0299,  0.0045]])
```

phq-9 단어 벡터화

```python
phq9_words=[]

for i in phq9_symptoms:
    for v in phq9_symptoms[i]:
        phq9_words.append(v)
```

```python
phq9_emb=encode_texts(phq9_words, len(phq9_words), batch_size)
```

```
100%|██████████| 1/1 [00:00<00:00, 26.81it/s]
```

```python
text_emb = encode_texts(six_label_all_text_phq9['text'], max_length, batch_size)
```

```
100%|██████████| 331/331 [00:42<00:00,  7.77it/s]
```

내적

```python
cos_sim=(text_emb @ phq9_emb.T).numpy()

max_sim=cos_sim.max(axis=1)
```

```python
six_label_all_text_phq9['similarity']=max_sim
```

```python
six_label_all_text_phq9.head()
```

	text	감정_대분류	상황키워드	similarity
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	분노	학업 및 진로	0.606270
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	분노	학업 및 진로	0.625640
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	분노	학업 및 진로	0.626699
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	분노	학교폭력/따돌림	0.580797
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	분노	학교폭력/따돌림	0.652328

다음 단계: six_label_all_text_phq9 변수로 코드 생성 | New interactive sheet

최종 점수 계산

- total_all = {(np.array(kw_scores_all) * W_KW) + (max_sim_all * W_SEM)} * TARGET_CATS
    - W_KW = 1.0 # 키워드 점수 가중치 자살
    - W_SEM = 0.6 # 의미 유사도 가중치
    - TARGET_CATS = ["학업 및 진로", "학교폭력/따돌림"] CATEGORY_BOOST = 1.5
-
    - 진행안함 유사도 점수만 돌렸을때가 오히려 최종 점수로 돌렸을때 보다 결과가 더 좋음

```python
# W_KW=1.0
# W_SEM= 0.6
```

```python
# six_label_all_text_phq9['TARGET_CATS']=six_label_all_text_phq9['상황키워드'].apply(lambda x: 1.5 if x in ["학업 및 진로", "학교폭력/따돌림"] else 1.0)
```

```python
# six_label_all_text_phq9['phq9_total_score']=((six_label_all_text_phq9['phq9_score']*W_KW)+(six_label_all_text_phq9['similarity']*W_SEM))*six_label_all_text_phq9['TARGET_CATS']
```

```python
# six_label_all_text_phq9.columns
```

```python
# six_label_all_text_phq9=six_label_all_text_phq9[['text', '감정_대분류','similarity', 'phq9_score',
#        'phq9_total_score']]
```

```python
# six_label_all_text_phq9.head()
```

모델링

Fine_Tuning

데이터로드 함수

```python
# max_length 설정

all_text_max_length, first_text_max_length
```

```
(161, 62)
```

```python
from transformers import AutoModel, AutoTokenizer
from sklearn.preprocessing import LabelEncoder
from torch.utils.data import Dataset, DataLoader
from sklearn.model_selection import train_test_split


batch_size=32
```

```python

class TextDataset(Dataset):
    def __init__(self, texts, phq9_features, labels, tokenizer, max_length):
        self.texts = texts.reset_index(drop=True)   # 확실하게 리셋
        self.labels = labels.reset_index(drop=True) # 확실하게 리셋
        self.tokenizer=tokenizer
        self.max_length=int(max_length)
        self.phq9_features = phq9_features.reset_index(drop=True) # Series 형태

    def __len__(self):
        return len(self.texts)

    def __getitem__(self, idx):
        text = str(self.texts.iloc[idx])
        label = int(self.labels.iloc[idx])

        feature_val = self.phq9_features.iloc[idx]


        encoding=self.tokenizer.encode_plus(
            text,
            add_special_tokens=True,
            max_length=self.max_length,
            padding='max_length',
            truncation=True,
            return_attention_mask=True,
            return_tensors='pt'
             )
        return (
                encoding['input_ids'].flatten(),
                encoding['attention_mask'].flatten(),
                torch.tensor([feature_val], dtype=torch.float32),
                torch.tensor(label, dtype=torch.long)
        )
```

```python
def FineTuning_data_load(text, max_length):

    tokenizer=AutoTokenizer.from_pretrained('klue/roberta-base')


    label_encoder=LabelEncoder()
    label_encoded=label_encoder.fit_transform(text.iloc[:,1])
    label_count=text.iloc[:,1].nunique()

    print(f"클래스 수: {label_count}")
    for idx, label_name in enumerate(label_encoder.classes_):
        print(f"  {idx}: {label_name}")
```

```

test['label_encoded']=label_encoded

X_train, X_test, X_phq9_train, X_phq9_test, y_train, y_test = train_test_split(
    test['text'],
    test['similarity'],
    test['label_encoded'],
    test_size=0.3,
    stratify=test['label_encoded'],
    random_state=42
)

X_train=X_train.reset_index(drop=True)
X_test=X_test.reset_index(drop=True)

X_phq9_train=X_phq9_train.reset_index(drop=True)
X_phq9_test=X_phq9_test.reset_index(drop=True)

y_train=y_train.reset_index(drop=True)
y_test=y_test.reset_index(drop=True)


train_dataset = TextDataset(
texts=X_train,
phq9_features=X_phq9_train,
labels=y_train,
tokenizer=tokenizer,
max_length=max_length
)

test_dataset = TextDataset(
    texts=X_test,
    phq9_features=X_phq9_test,
    labels=y_test,
    tokenizer=tokenizer,
    max_length=max_length
)

train_loader=DataLoader(train_dataset, batch_size=batch_size, shuffle=True)
test_loader=DataLoader(test_dataset, batch_size=batch_size, shuffle=False)

print(f"✅ 훈련 데이터: {len(train_dataset)}개")
print(f"✅ 테스트 데이터: {len(test_dataset)}개")

return train_loader, test_loader, label_count
```

모델 함수

```
from transformers import AutoModel,AutoConfig
import torch.nn as nn
from tqdm import tqdm

epoch=1
```

```
class klueClassifier(nn.Module):
    def __init__(self, num_class, num_layers_to_train=3, dropout=0.3):
        super(klueClassifier,self).__init__()

        self.klue=AutoModel.from_pretrained('klue/roberta-base')
        config= AutoConfig.from_pretrained('klue/roberta-base')
        hidden_size=config.hidden_size

        # ========================================
        # 부분 파인튜닝
        total_layers = 12 # klue는 12개 레이어
        layers_to_freeze = total_layers - num_layers_to_train

        for param in self.klue.embeddings.parameters():
            param.requires_grad = False

        # 2) 하위 N개 레이어 freeze
        for i in range(layers_to_freeze):
            for param in self.klue.encoder.layer[i].parameters():
                param.requires_grad = False
        # ========================================


        self.phq9_feature_layer=nn.Linear(1,16)

        self.dropout=nn.Dropout(dropout)
        self.classifier= nn.Linear(hidden_size*16, num_class)


    def forward(self, input_ids, attention_mask, phq9_features):

        output=self.klue(input_ids=input_ids, attention_mask=attention_mask)
        cls_token=output.pooler_output

        phq9=self.phq9_feature_layer(phq9_features)
        phq9=torch.relu(phq9)
        combined =torch.cat([cls_token, phq9], dim=1)

        x=self.dropout(combined)
        y_pred=self.classifier(x)

        return y_pred
```

```

def fineTuning_train(train_loader, model, criterion, optimizer, device=device, epochs=5):

    history = {'epoch': [], 'loss': []}

    for i in range(epochs):
        model.train()
        total_loss=0

        for x, attention_mask, phq9, y in tqdm(train_loader):

            x= x.to(device)
            attention_mask=attention_mask.to(device)
            phq9=phq9.to(device)
            y= y.to(device)

            y_pred=model(x, attention_mask, phq9)
            loss=criterion(y_pred, y)

            optimizer.zero_grad()
            loss.backward()

            optimizer.step()

            total_loss+=loss.item()

        history['loss'].append(total_loss / len(train_loader))
        history['Epoch'].append(i)

        if i % 2 == 0:
            print (f'Epoch {i}: Loss={total_loss/len(train_loader)}')

    return history
```

모델평가 함수

```
from sklearn.metrics import(
    accuracy_score,
    f1_score,
    precision_score,
    recall_score,
    classification_report,
    confusion_matrix
)
import numpy as np

```

```
def fineTuning_eval(test_loader, model, depression_classes, device=device, label_names=None ):
    model.eval()

    all_preds=[]
    all_labels=[]

    with torch.no_grad():
        for x, attention_mask, phq9, y in test_loader:  # ✅ 3개!
            x = x.to(device)
            attention_mask = attention_mask.to(device)
            phq9=phq9.to(device)

            y_pred = model(x, attention_mask, phq9)  # ✅ attention_mask 포함
            preds = torch.argmax(y_pred, dim=1)

            all_preds.extend(preds.cpu().numpy())
            all_labels.extend(y.cpu().numpy())


    accuracy = accuracy_score(all_labels, all_preds)
    f1_macro = f1_score(all_labels, all_preds, average='macro')
    f1_weighted = f1_score(all_labels, all_preds, average='weighted')
    cm = confusion_matrix(all_labels, all_preds)

    print(f"\n✅ Accuracy: {accuracy:.4f}")
    print(f"✅ Macro F1-Score: {f1_macro:.4f}")
    print(f"✅ Weighted F1-Score: {f1_weighted:.4f}")
    print(f"\n✅ Confusion Matrix:")
    print(cm)


    # ========================================
    # 3. 우울 신호 집중 분석
    # ========================================

    print("\n" + "="*60)
    print(f"🎯 우울 신호 집중 분석 ({depression_classes})")
    print("="*60)

    # 이진 분류로 변환
    all_labels_np = np.array(all_labels)
    all_preds_np = np.array(all_preds)

    y_true_binary = np.isin(all_labels_np, depression_classes).astype(int)
    y_pred_binary = np.isin(all_preds_np, depression_classes).astype(int)

    # Precision & Recall
    precision = precision_score(y_true_binary, y_pred_binary, zero_division=0)
    recall = recall_score(y_true_binary, y_pred_binary, zero_division=0)
    f1 = f1_score(y_true_binary, y_pred_binary, zero_division=0)


    # 실제 우울 신호인데 기타로 예측
    false_negatives = np.sum((y_true_binary == 1) & (y_pred_binary == 0))
    # True Positive
    true_positives = np.sum((y_true_binary == 1) & (y_pred_binary == 1))
    # 전체 우울 신호
    total_depression = np.sum(y_true_binary == 1)


    print(f"\n{'='*60}")
    print(f"🎯 우울 신호 집중 분석 (불안 + 상처 + 슬픔)")
    print(f"{'='*60}")
    print(f"✅ Precision: {precision:.4f}")
    print(f"✅ Recall: {recall:.4f}")
    print(f"✅ F1-Score: {f1:.4f}")
    print(f"\n⚠️ False Negative: {false_negatives}건 / {total_depression}건")
    print(f"✅ True Positive: {true_positives}건 / {total_depression}건")


    # Classification Report
    if label_names: # e.g. ['기쁨', '당황', '분노', '불안', '상처', '슬픔']
        print(f"\n📋 Classification Report:")
        print(classification_report(all_labels, all_preds, target_names=label_names))


    return {
        'accuracy': accuracy,
        'f1_macro': f1_macro,
        'f1_weighted': f1_weighted,
        'depression_precision': precision,
        'depression_recall': recall,
        'depression_f1': f1,
        'false_negatives': false_negatives,
        'true_positives': true_positives,
        'total_depression': total_depression,
    }
```

두개 선별 데이터

- six_label_all_text + finetuning
- four_label_all_text + finetuning
- six_label_all_text_phq9 + finetuning
- four_label_all_text_phq9" + finetuning

```
import torch
import torch.nn.functional as F
from transformers import AutoTokenizer, AutoModel
from tqdm import tqdm
import numpy as np
from tqdm import tqdm


device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
print(f"Using device: {device}")

MODEL_NAME = "klue/roberta-base"
tokenizer = AutoTokenizer.from_pretrained(MODEL_NAME)
klue = AutoModel.from_pretrained(MODEL_NAME).to(device)
```

```
Using device: cuda
Some weights of RobertaModel were not initialized from the model checkpoint at klue/roberta-base and are newly initialized: ['pooler.dense.bias', 'pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

four_label_all_text_phq9

```
four_label_all_text_phq9
```

	text	감정 분류	상황키워드
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	우울 신호	학업 및 진로
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	우울 신호	학업 및 진로
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	우울 신호	학업 및 진로
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	일반 부정	학교폭력/따돌림
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	일반 부정	학교폭력/따돌림
...	...	...	...
10577	엄마가 내 학교생활까지 간섭하고 있어서 매번 싸워. 내가 지금보다 더 좋은 모습을 ...	일반 부정	가족관계
10578	부모님은 내 성적으로 매번 뭐라고 하셔. 학교에서는 공부 잘하는 편이라고 하는데 말...	일반 부정	학업 및 진로
10579	친구랑 싸웠어. 너무 속상해. 친구가 나랑 놀면 우리 부모님이 하지 말라는 것 때문...	일반 부정	가족관계
10580	부모님이 내가 밖에 나가려고만 하면 간섭을 하셔. 속상해. 내가 너무 험하게 놀고 ...	일반 부정	가족관계
10581	부모님은 나에게 관심도 없어. 내가 무슨 말을 하든 다 잊어버리는 것 같아서 짜증 ...	일반 부정	가족관계

10582 rows × 3 columns

다음 단계: four_label_all_text_phq9 변수로 코드 생성 | New interactive sheet

```
# max_length 정하기

all_text_max_length=71
first_text_max_length =28
```

```
#four_label_all_text_phq9['text_pos']=four_label_all_text['text'].apply(lambda x: pos(x))
```

```
#four_label_all_text_phq9['phq9_score']=four_label_all_text['text_pos'].apply(lambda x: create_phq9_vector(x))
```

```
phq9_emb=encode_texts(phq9_words, len(phq9_words), batch_size)
```

```
100%|██████████| 1/1 [00:00<00:00, 79.32it/s]
```

```
text_emb = encode_texts(four_label_all_text_phq9['text'], max_length, batch_size)
```

```
100%|██████████| 331/331 [00:42<00:00,  7.83it/s]
```

```
cos_sim=(text_emb @ phq9_emb.T).numpy()

max_sim=cos_sim.max(axis=1)
```

```
four_label_all_text_phq9['similarity']=max_sim
```

```
# W_KW=1.0
# W_SEM= 0.6
```

```
# four_label_all_text_phq9['TARGET_CATS']=four_label_all_text_phq9['상황키워드'].apply(lambda x: 1.5 if x in ["학업 및 진로", "학교폭력/따돌림"] else 1.0)
```

```
# four_label_all_text_phq9['phq9_total_score']=((four_label_all_text_phq9['phq9_score']*W_KW)+(four_label_all_text_phq9['similarity']*W_SEM))*four_label_all_text_phq9['TARGET_CATS']
```

```
four_label_all_text_phq9
```

	text	감정 분류	상황키워드	similarity
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	우울 신호	학업 및 진로	0.606270
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	우울 신호	학업 및 진로	0.625640
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	우울 신호	학업 및 진로	0.626699
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	일반 부정	학교폭력/따돌림	0.580797
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	일반 부정	학교폭력/따돌림	0.652328
...	...	...	...	...
10577	엄마가 내 학교생활까지 간섭하고 있어서 매번 싸워. 내가 지금보다 더 좋은 모습을 ...	일반 부정	가족관계	0.604876
10578	부모님은 내 성적으로 매번 뭐라고 하셔. 학교에서는 공부 잘하는 편이라고 하는데 말...	일반 부정	학업 및 진로	0.594684
10579	친구랑 싸웠어. 너무 속상해. 친구가 나랑 놀면 우리 부모님이 하지 말라는 것 때문...	일반 부정	가족관계	0.597930
10580	부모님이 내가 밖에 나가려고만 하면 간섭을 하셔. 속상해. 내가 너무 험하게 놀고 ...	일반 부정	가족관계	0.617870
10581	부모님은 나에게 관심도 없어. 내가 무슨 말을 하든 다 잊어버리는 것 같아서 짜증 ...	일반 부정	가족관계	0.587160

10582 rows × 4 columns

다음 단계: four_label_all_text_phq9 변수로 코드 생성 | New interactive sheet

```
# four_label_all_text_phq9=four_label_all_text_phq9[['text', '감정 분류','similarity', 'phq9_score',
#        'phq9_total_score']]
```

```
phq9_total_result=[]

results={'task':'four_label_all_text_phq9',
        'text': 'all',
        'label':'four',
        'model':'finetuning',
        'phq9': 'yes'
                }
```

```
max_length=all_text_max_length

batch_size=8
epochs=6
```

```
train_loader, test_loader, label_count=FineTuning_data_load(four_label_all_text_phq9, max_length)
```

```
클래스 수: 4
  0: 긍정
  1: 우울 신호
  2: 일반 부정
  3: 중립
✅ 훈련 데이터: 7407개
✅ 테스트 데이터: 3175개
```

```
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
finetuning_model = klueClassifier(num_class=label_count, num_layers_to_train=3, dropout=0.3).to(device)


FineTuning_criterion=nn.CrossEntropyLoss()
FineTuning_optimizer=torch.optim.AdamW([{'params':finetuning_model.klue.parameters(), 'lr': 1e-4},
                            {'params':finetuning_model.classifier.parameters(), 'lr': 1e-4}
                            ], weight_decay=0.01)
```

```
Some weights of RobertaModel were not initialized from the model checkpoint at klue/roberta-base and are newly initialized: ['pooler.dense.bias', 'pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```
FineTuning_train(train_loader, finetuning_model, FineTuning_criterion, FineTuning_optimizer, device, epochs=6)
```

```
100%|██████████| 926/926 [00:57<00:00, 16.14it/s]
Epoch 0: Loss=0.78063283300683
100%|██████████| 926/926 [00:52<00:00, 17.77it/s]
100%|██████████| 926/926 [00:53<00:00, 17.27it/s]
Epoch 2: Loss=0.5974393697942307
100%|██████████| 926/926 [00:52<00:00, 17.80it/s]
100%|██████████| 926/926 [00:53<00:00, 17.38it/s]
Epoch 4: Loss=0.5091130429794348
100%|██████████| 926/926 [00:54<00:00, 17.05it/s]
{'Epoch': [0, 1, 2, 3, 4, 5],
 'loss': [0.78063283300683,
  0.6474522269626609,
  0.5974393697942307,
  0.5468721830851316,
  0.5091130429794348,
  0.4532699808719735]}
```

```
torch.save(finetuning_model.state_dict(), SAVE_PATH + 'phase2_four_label_all_text_phq9.pt')
print("모델 저장 완료")
```

```
모델 저장 완료
```

```
six_label_depression_classes = [3, 4, 5]
four_label_depression_classes = [1]
```

```
result=FineTuning_eval(test_loader, finetuning_model, four_label_depression_classes, device=device)
```

```
✅ Accuracy: 0.7276
✅ Macro F1-Score: 0.6583
✅ Weighted F1-Score: 0.7278

🔢 Confusion Matrix:
[[ 347    8   17    1]
 [  10  678  268    7]
 [  17  426 1259   31]
 [   1   31   48   26]]

============================================================
🎯 우울 신호 집중 분석 ([1])
============================================================

============================================================
🎯 우울 신호 집중 분석 (불안 + 상처 + 슬픔)
============================================================
✅ Precision: 0.5932
✅ Recall: 0.7040
✅ F1-Score: 0.6439

⚠️ False Negative: 285건 / 963건
✅ True Positive: 678건 / 963건
```

```
results.update(result)
```

```
phq9_total_result.append(results)
```

```
import gc
del finetuning_model
del FineTuning_optimizer
torch.cuda.empty_cache()
gc.collect()
```

```
409
```

six_label_all_text_phq9

```
six_label_all_text_phq9
```

	text	감정_대분류	상황키워드	similarity
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	분노	학업 및 진로	0.606270
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	분노	학업 및 진로	0.625640
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	분노	학업 및 진로	0.626699
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	분노	학교폭력/따돌림	0.580797
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	분노	학교폭력/따돌림	0.652328
...	...	...	...	...
10577	엄마가 내 학교생활까지 간섭하고 있어서 매번 싸워. 내가 지금보다 더 좋은 모습을 ...	당황	가족관계	0.604876
10578	부모님은 내 성적으로 매번 뭐라고 하셔. 학교에서는 공부 잘하는 편이라고 하는데 말...	분노	학업 및 진로	0.594684
10579	친구랑 싸웠어. 너무 속상해. 친구가 나랑 놀면 우리 부모님이 하지 말라는 것 때문...	불안	가족관계	0.597930
10580	부모님이 내가 밖에 나가려고만 하면 간섭을 하셔. 속상해. 내가 너무 험하게 놀고 ...	당황	가족관계	0.617870
10581	부모님은 나에게 관심도 없어. 내가 무슨 말을 하든 다 잊어버리는 것 같아서 짜증 ...	불안	가족관계	0.587160

10582 rows × 4 columns

다음 단계: six_label_all_text_phq9 변수로 코드 생성 | New interactive sheet

```
# max_length 정하기

all_text_max_length=71
first_text_max_length =28
```

```
# six_label_all_text_phq9['text_pos']=six_label_all_text_phq9['text'].apply(lambda x: pos(x))
```

```
# six_label_all_text_phq9['phq9_score']=six_label_all_text_phq9['text_pos'].apply(lambda x: create_phq9_vector(x))
```

```
phq9_emb=encode_texts(phq9_words, len(phq9_words), batch_size)
```

```
100%|██████████| 4/4 [00:00<00:00, 69.26it/s]
```

```
text_emb = encode_texts(six_label_all_text_phq9['text'], max_length, batch_size)
```

```
100%|██████████| 1323/1323 [00:40<00:00, 32.90it/s]
```

```
cos_sim=(text_emb @ phq9_emb.T).numpy()

max_sim=cos_sim.max(axis=1)
```

```
six_label_all_text_phq9['similarity']=max_sim
```

```
# W_KW=1.0
# W_SEM= 0.6
```

```
# six_label_all_text_phq9['TARGET_CATS']=six_label_all_text_phq9['상황키워드'].apply(lambda x: 1.5 if x in ["학업 및 진로", "학교폭력/따돌림"] else 1.0)
```

```
# six_label_all_text_phq9['phq9_total_score']=((six_label_all_text_phq9['phq9_score']*W_KW)+(six_label_all_text_phq9['similarity']*W_SEM))*six_label_all_text_phq9['TARGET_CATS']
```

```
six_label_all_text_phq9
```

	text	감정_대분류	상황키워드	similarity
0	음악 실기 완전 망쳤어. 나는 왜 잘하는 게 하나도 없는지 모르겠다. 보통 공부를 ...	분노	학업 및 진로	0.606271
1	축제 때 전시할 그림을 나는 한 달 내내 준비했는데 친구는 겨우 이틀 만에 끝냈다네...	분노	학업 및 진로	0.625640
2	내가 어제 종일 피아노 쳤는데도 아무리 연습해도 실력이 늘지 않아. 속상하고 좌절스...	분노	학업 및 진로	0.626699
3	반 애들이 나만 따돌려서 너무 화가 나. 도대체 나한테 왜 이러는 걸까? 글쎄. 처...	분노	학교폭력/따돌림	0.580797
4	친구들이 내 말을 무시하는 게 화가 나. 왜 내 말을 무시한 거지? 또 그러면 다음...	분노	학교폭력/따돌림	0.652328
...	...	...	...	...
10577	엄마가 내 학교생활까지 간섭하고 있어서 매번 싸워. 내가 지금보다 더 좋은 모습을 ...	당황	가족관계	0.604876
10578	부모님은 내 성적으로 매번 뭐라고 하셔. 학교에서는 공부 잘하는 편이라고 하는데 말...	분노	학업 및 진로	0.594684
10579	친구랑 싸웠어. 너무 속상해. 친구가 나랑 놀면 우리 부모님이 하지 말라는 것 때문...	불안	가족관계	0.597930
10580	부모님이 내가 밖에 나가려고만 하면 간섭을 하셔. 속상해. 내가 너무 험하게 놀고 ...	당황	가족관계	0.617870
10581	부모님은 나에게 관심도 없어. 내가 무슨 말을 하든 다 잊어버리는 것 같아서 짜증 ...	불안	가족관계	0.587160

10582 rows × 4 columns

다음 단계: six_label_all_text_phq9 변수로 코드 생성 New interactive sheet

```
# six_label_all_text_phq9=six_label_all_text_phq9[['text', '감정_대분류','similarity', 'phq9_score',
#         'phq9_total_score']]
```

```

results={'task':'six_label_all_text_phq9',
        'text': 'all',
        'label':'six',
        'model':'finetuning',
        'phq9': 'yes'
                }
```

```
max_length=all_text_max_length

batch_size=32
epochs=6
```

```
train_loader, test_loader, label_count=FineTuning_data_load(six_label_all_text_phq9, max_length)
```

```
클래스 수: 6
  0: 기쁨
  1: 당황
  2: 분노
  3: 불안
  4: 상처
  5: 슬픔
✅ 훈련 데이터: 7407개
✅ 테스트 데이터: 3175개
```

```
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
finetuning_model = klueClassifier(num_class=label_count, num_layers_to_train=3, dropout=0.3).to(device)


FineTuning_criterion=nn.CrossEntropyLoss()
FineTuning_optimizer=torch.optim.AdamW([{'params':finetuning_model.klue.parameters(), 'lr': 1e-4},
                           {'params':finetuning_model.classifier.parameters(), 'lr': 1e-4}
                           ], weight_decay=0.01)
```

```
Some weights of RobertaModel were not initialized from the model checkpoint at klue/roberta-base and are newly initialized: ['pooler.dense.bias', 'pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```
FineTuning_train(train_loader, finetuning_model, FineTuning_criterion, FineTuning_optimizer, device, epochs=6)
```

```
100%|██████████| 232/232 [00:43<00:00,  5.30it/s]
Epoch 0: Loss=1.2709504556039284
100%|██████████| 232/232 [00:43<00:00,  5.31it/s]
100%|██████████| 232/232 [00:43<00:00,  5.31it/s]
Epoch 2: Loss=0.8366146749206658
100%|██████████| 232/232 [00:43<00:00,  5.30it/s]
100%|██████████| 232/232 [00:43<00:00,  5.31it/s]
Epoch 4: Loss=0.6219044909009646
100%|██████████| 232/232 [00:43<00:00,  5.30it/s]
{'Epoch': [0, 1, 2, 3, 4, 5],
 'loss': [1.2709504556039284,
  0.9458683048342836,
  0.8366146749206658,
  0.7166754432536405,
  0.6219044909009646,
  0.504972870365299]}
```

```
torch.save(finetuning_model.state_dict(), SAVE_PATH + 'phase2_six_label_all_text_phq9.pt')
print("모델 저장 완료")
```

```
모델 저장 완료
```

```
six_label_depression_classes = [3, 4, 5]
four_label_depression_classes = [1]
```

```
#six_label_all_text_phq9

result=FineTuning_eval(test_loader, finetuning_model, six_label_depression_classes,device=device)

```

```
✅ Accuracy: 0.6923
✅ Macro F1-Score: 0.7057
✅ Weighted F1-Score: 0.6928

🔢 Confusion Matrix:
[[343   6  10   1   5   8]
 [  3 394  30  39  44  43]
 [  3  39 347  43  50  56]
 [  4  57  52 370  66  51]
 [  9  37  52  28 351  68]
 [  4  32  36  32  69 393]]

============================================================
🎯 우울 신호 집중 분석 ([3, 4, 5])
============================================================

============================================================
🎯 우울 신호 집중 분석 (불안 + 상처 + 슬픔)
============================================================
✅ Precision: 0.8317
✅ Recall: 0.8346
✅ F1-Score: 0.8331

⚠️ False Negative: 283건 / 1711건
✅ True Positive: 1428건 / 1711건
```

```
results.update(result)
```

```
phq9_total_result.append(results)
```

```
phq9_total_result
```

```
[{'task': 'four_label_all_text_phq9',
  'text': 'all',
  'label': 'four',
  'model': 'finetuning',
  'phq9': 'yes',
  'accuracy': 0.7275590551181103,
  'f1_macro': 0.6582672327099572,
  'f1_weighted': 0.7277941938978384,
  'depression_precision': 0.5931758530183727,
  'depression_recall': 0.7040498442367601,
  'depression_f1': 0.6438746438746439,
  'false_negatives': np.int64(285),
  'true_positives': np.int64(678),
  'total_depression': np.int64(963)},
 {'task': 'six_label_all_text_phq9',
  'text': 'all',
  'label': 'six',
  'model': 'finetuning',
  'phq9': 'yes',
  'accuracy': 0.6922834645669291,
  'f1_macro': 0.705692417726166,
  'f1_weighted': 0.6927634271763351,
  'depression_precision': 0.8316831683168316,
  'depression_recall': 0.8345996493278784,
  'depression_f1': 0.8331388564760793,
  'false_negatives': np.int64(283),
  'true_positives': np.int64(1428),
  'total_depression': np.int64(1711)}]
```

```
import gc
del finetuning_model
del FineTuning_optimizer
torch.cuda.empty_cache()
gc.collect()
```

```
218
```

비교 및 시각화

```
phq9_total_result=pd.DataFrame(phq9_total_result)
```

```

# 기존 결과 불러오기
total_result = pd.read_csv(SAVE_PATH + 'phase1_total_result.csv') # Phase 1에서 저장한 이름 확인

total_result['phq9'] = 'no' # 기존 결과에는 phq9 없음 표시

# 병합 (concat 리스트로 묶기!)
concat_total_result = pd.concat([total_result, phq9_total_result], ignore_index=True)

# # 최종 저장
# concat_total_result.to_csv(SAVE_PATH + 'phase2_total_result.csv', index=False)
```

```
concat_total_result
```

	Unnamed: 0	task	text	label	model	accuracy	f1_macro	f1_weighted	depression_precision	depression_recall	depression_f1	false_negatives	true_positives	total_depression	phq9
0	0.0	six_label_all_text_mlp	all	six	mlp	0.493228	0.494794	0.481526	0.810362	0.484512	0.606437	882	829	1711	no
1	1.0	six_label_all_text_finetuning	all	six	finetuning	0.690709	0.702094	0.689634	0.835608	0.822911	0.829211	303	1408	1711	no
2	2.0	four_label_all_text_mlp	all	four	mlp	0.635591	0.464339	0.604500	0.552265	0.329180	0.412492	646	317	963	no
3	3.0	four_label_all_text_finetuning	all	four	finetuning	0.728819	0.663972	0.729996	0.584020	0.736241	0.651355	254	709	963	no
4	4.0	six_label_first_text_mlp	first	six	mlp	0.479685	0.483210	0.474238	0.721618	0.698422	0.709831	516	1195	1711	no
5	5.0	six_label_first_text_finetuning	first	six	finetuning	0.590551	0.603807	0.590082	0.764738	0.811222	0.787294	323	1388	1711	no
6	6.0	four_label_first_text_mlp	first	four	mlp	0.627717	0.443433	0.590027	0.539499	0.290758	0.377868	683	280	963	no
7	7.0	four_label_first_text_finetuning	first	four	finetuning	0.685984	0.587333	0.682067	0.541133	0.614746	0.575596	371	592	963	no
8	NaN	four_label_all_text_phq9	all	four	finetuning	0.727559	0.658267	0.727794	0.593176	0.704050	0.643875	285	678	963	yes
9	NaN	six_label_all_text_phq9	all	six	finetuning	0.692283	0.705692	0.692763	0.831683	0.834600	0.833139	283	1428	1711	yes

다음 단계: final_combined_result 변수로 코드 생성 New interactive sheet

```
ph2_total_result=concat_total_result[(concat_total_result['text']=='all')&(concat_total_result['model']=='finetuning')]
```

```
ph2_total_result=ph2_total_result.drop(['Unnamed: 0'], axis=1)
```

```
ph2_total_result
```

	task	text	label	model	accuracy	f1_macro	f1_weighted	depression_precision	depression_recall	depression_f1	false_negatives	true_positives	total_depression	phq9
1	six_label_all_text_finetuning	all	six	finetuning	0.690709	0.702094	0.689634	0.835608	0.822911	0.829211	303	1408	1711	no
3	four_label_all_text_finetuning	all	four	finetuning	0.728819	0.663972	0.729996	0.584020	0.736241	0.651355	254	709	963	no
8	four_label_all_text_phq9	all	four	finetuning	0.727559	0.658267	0.727794	0.593176	0.704050	0.643875	285	678	963	yes
9	six_label_all_text_phq9	all	six	finetuning	0.692283	0.705692	0.692763	0.831683	0.834600	0.833139	283	1428	1711	yes

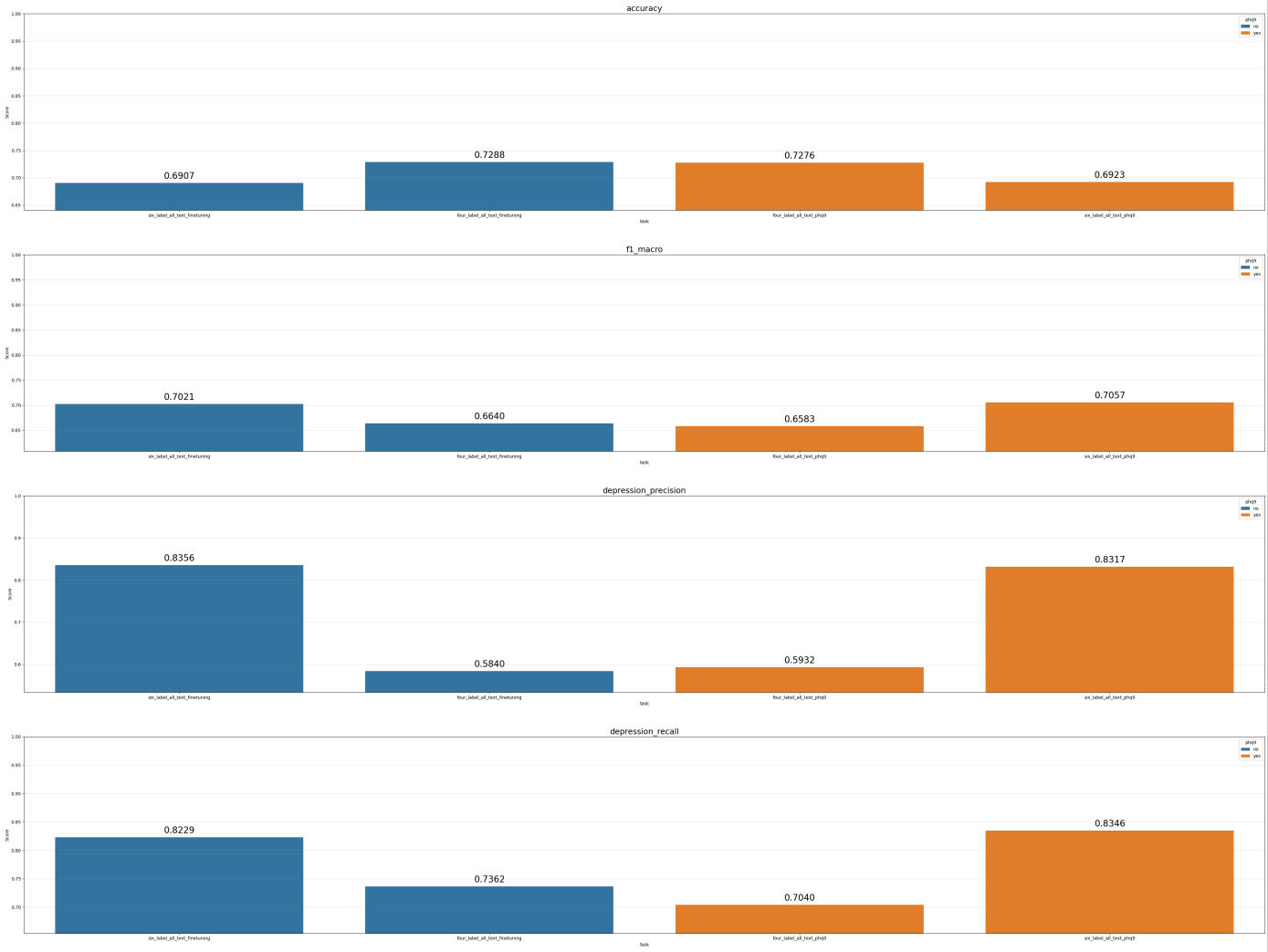
다음 단계: ph2_total_result 변수로 코드 생성 New interactive sheet

```
ph2_total_result.to_csv(SAVE_PATH + 'phase2_total_result.csv', index=False)
```

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
figure, axs = plt.subplots(4,1,figsize=(40,30))

metrics=['accuracy', 'f1_macro', 'depression_precision', 'depression_recall']


for i,v in enumerate(metrics):
    #axs[i].bar(total_result['task'], total_result[v])
    sns.barplot(data=ph2_total_result, x='task', y=v, hue='phq9', ax= axs[i])
    axs[i].set_title(v, fontsize=18)
    axs[i].set_ylabel('Score')
    axs[i].grid(axis='y', alpha=0.3)


    # ========================================
```

```
    # 🔥 [여기만 추가] y축을 확대해서 차이를 잘 보이게 함
    # 최소값에서 0.05 뺀 지점부터 1.0(또는 최댓값)까지 보여줌
    min_score = phq9_total_result[col].min() - 0.05
    axs[i].set_ylim(bottom=min_score if min_score > 0 else 0, top=1.0) # top은 필요하면 조절

    for container in axs[i].containers:
        # fmt='%.4f' : 소수점 4자리까지 표시
        # fontsize=20 : 그림이 크니까 글자도 크게
        axs[i].bar_label(container, fmt='%.4f', padding=5, fontsize=20)
    # ▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲▲

plt.tight_layout(h_pad=5.0)

plt.show()
```



1 코딩을 시작하거나 AI로 코드를 생성하세요.

1 코딩을 시작하거나 AI로 코드를 생성하세요.

1 코딩을 시작하거나 AI로 코드를 생성하세요.