MISO 보고서

1. 프로젝트 배경 및 목표

프로젝트 배경

- **현대 사회에서의 음악 역할**
  - 음악은 단순한 오락을 넘어 일상 속 어디든 존재하는 필수 요소
  - 현대인은 음악으로 지친 감정을 위로 받기도 하고 삶의 원동력이 되어줌

음악 시장 소비 트렌드

- **음악 소비 트렌드 변화**
  - 음악 감상 트렌드는 다양한 주제, 감성을 담은 플레이리스트를 통해 소비가 이루어지고 있음
  - 최근에는 자신만의 색깔, 개성을 드러내고 찾고 싶은 욕구가 높음
  - 따라서 개인이 플레이리스트를 만들어 본인의 개성 표현 및 공유
- 플레이리스트 활용 마케팅 및 이용 증가
  - 플레이리스트 음악과 섬네일로 소비자들의 감성 자극
  - 유튜브 플레이리스트 채널 구독자 수 (100만 이상 : `essential`, `떼껄룩`)
  - 플레이리스트 조회수
  - 브랜드와 플레이리스트 협업

- AI 기반 음악 추천 서비스 증가

    - AI 기술이 발전하고 관심도가 증가하면서 AI를 음악에 활용하는 플랫폼 증가
    - 기존 AI를 활용한 음악 추천 시스템은 보통 자연어 인풋 기반의 추천 시스템이 대다수
    - 주변 분위기와 상호작용하는 비정형 데이터 활용 AI 기반 음악 추천 서비스 많지 않음

2. 문제정의

- **기존의 음악 추천의 한계**
    - 다른 이용자나 플랫폼에서 만든 플레이리스트에서 음악을 선택하는 시스템이 일반적임
    - 어떤 곡을 선택해야 유사한 노래들이 재생됨
    - 그날 기분을 입력해 감정 분석을 통해 음악을 추천하는 방식도 일부 존재함 → 번거로움
    - 신곡이나 곡에 대한 정보 업데이트가 느림 → 사람이 직접 구성하는 플레이리스트가 나오는데 시간이 걸림

음악 시장 소비 트렌드의 문제점

- 플레이리스트
    - 내 상황, 기분을 직접 검색해 플레이리스트를 찾아야함
    - 플레이리스트를 사람이 직접 만들고 구성하기 때문에 업데이트 주기가 김
    - 최신곡이 플레이리스트에 반영되려면 시간이 걸림

- AI 기반 음악 추천
    - 자연어 기반이므로 플레이리스트와 마찬가지로 단어를 입력하거나 챗봇과 대화를 진행함 → 번거로움 존재
    - 음성을 이용한 추천도 mood를 직접 얘기 해야함(사용자 무드 선택 → 자동화 X)

3. 프로젝트 목표

프로젝트 목표

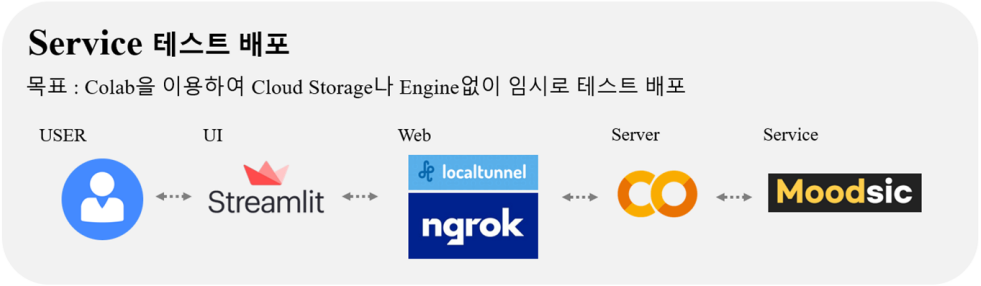
- 기존의 시스템과 다른 방향으로 접근하여 사진 기반의 새로운 추천 시스템 장르 개발
- 사용자의 이미지(일상사진, 추억, 현재 표정)를 통해 분위기에 잘 맞는 음악 추천 서비스 구현

- 팀 목표
  - 이미지 기반 음악 추천 서비스
    - 직접 감정을 얘기하거나 입력을 통한 방식이 아닌 이미지를 기반해 분위기 분석
    - 분위기와 맞는 음악 여러개 추천

  - 웹을 통한 서비스 구현

4. 프로젝트 수행 과정

서비스 전체 구조



BLIP 모델

- image를 input으로 받아 text로 상황을 묘사해주는 모델
- 멀티모달 모델
  - 파인튜닝 데이터
    - 데이터 원본: https://huggingface.co/datasets/visheratin/laion-coco-nllb
    - 데이터 크기: (893884, 5)
    - 해당 데이터를 통해 한국어로 학습을 진행하려했으나 환경 및 시간 문제로 학습된 모델 이용(데이터의 1% 학습시키는데 15시간 소요)

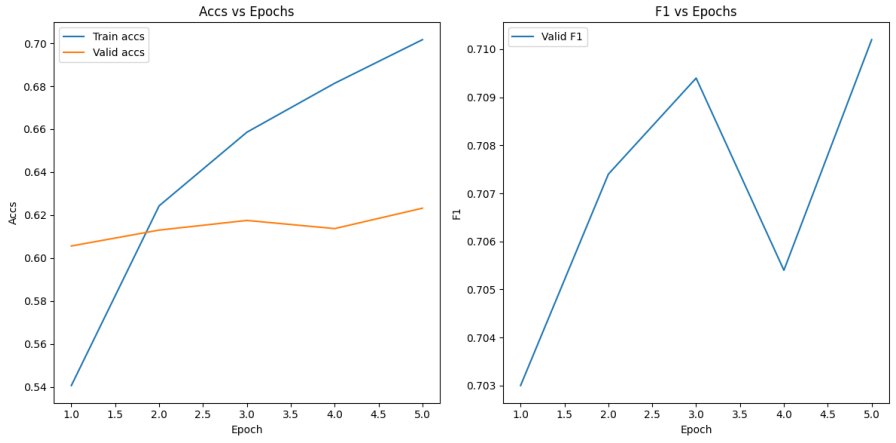
Bert 모델

- Bert 모델을 이용해 BLIP 모델이 출력한 text의 분위기(mood)를 파악
- 자연어 모델
  - 파인튜닝 데이터

        - 데이터 원본: https://research.google/blog/goemotions-a-dataset-for-fine-grained-emotion-classification/
        - 데이터 크기: (265488, 3)
        - target 감정 라벨 → 28가지('admiration', 'amusement', 'anger' 등등..)
        - Reddit(소셜, 콘텐츠, 토론 웹사이트) 사이트에서 수집한 댓글 데이터

- 구조 설계
    - 28가지의 감정을 ['happy', 'romantic', 'sad', 'aggressive', 'dramatic'] → 5가지 감정으로 축약
        - ex) df['romantic'] = df[['love', 'caring']].max(axis=1)
    - bert-base 모델 이용
    - 각 라벨(5가지의 감정) sigmoid 함수 이용한 예측 확률 출력
    - Optimizer = AdamW
    - loss 함수 = BCEWithLogitLoss

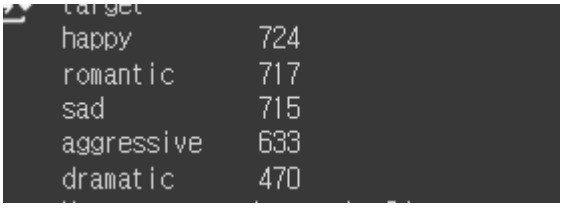
- 성능
    - 평가 그래프



    - F1_score = 0.712

음원분류 모델

- 음원에서 5가지 무드(aggressive, dramatic, happy, romantic, sad)를 분류하는 다중분류 모델
  - 음악의 감정태그를 확인할 수 있는 last.fm을 활용하여 메타데이터 수집
  - 이후 메타데이터를 활용하여 youtube에서 wav파일 크롤링 진행
  - 총 3259개의 학습 데이터 수집 완료



- 각 데이터의 mel spectogram 수집
  - 주파수의 단위를 공식에 따라 멜 단위(Mel unit)로 바꾼 스펙트럼
- wav파일을 활용하여 총 3가지 카테고리의 피처 수집

▼ 스펙트럴 특징

Spectral Centroid	스펙트럼의 무게 중심으로, 음의 밝기와 관련됨
Spectral Bandwidth	스펙트럼의 폭, 주파수의 분포 범위를 나타냄
Spectral Flatness	스펙트럼이 얼마나 평탄한지를 나타내는 지표
Spectral Rolloff	스펙트럼 에너지의 특정 비율을 포함하는 주파수 지점
Spectral Contrast	주파수 대역 간의 에너지 차이
MFCC (Mel-Frequency Cepstral Coefficients)	멜 주파수 대역에서의 스펙트럼 정보를 요약한 계수들
Chroma Feature	스펙트럼을 12개의 반음 계열로 요약한 특징
Spectral Flux	스펙트럼의 시간적 변화율

▼ 음향 특징

Tempo	곡의 속도 (BPM)
RMS (Root Mean Square) Energy	신호의 에너지 수준을 나타내는 RMS 값
ZCR (Zero Crossing Rate)	신호의 0 교차 비율

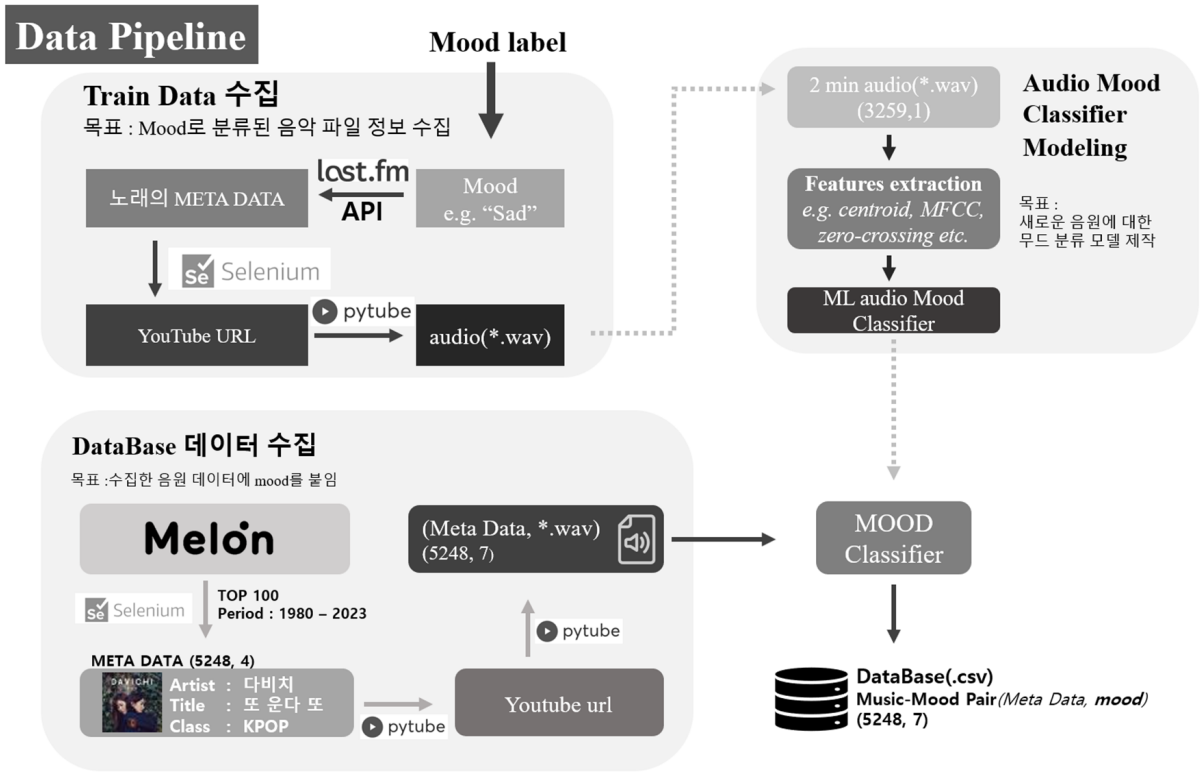
▼ 구조적 특징

Beat	박자
Harmonic	조화적 성분 분

Percussive Components	타악적 성분 분리

- 수집 데이터 셋 : 스펙트럴 특징(`3248,54`), 음향 특징(`3248,6`) , 구조적 특징 (`3248,7`)
- 각각의 데이터 셋을 활용한 모델 구축
- 최종 데이터 셋 : 스펙트럴 특징(`3248,54`)
- 모델 선정
  - mel sepctogram을 활용한 cnn 이미지 분류 모델
    - 0.43
  - 피처 데이터를 활용한 이미지 분류 모델
    - Gradient Boosting Classifier : 0.539
    - Extra Trees Classifier : 0.534
    - Random Forest Classifier : 0.527
  - 성능이 높은 3개 모델 선정 (Gradient Boosting Classifier, Random Forest Classifier, Extra Trees Classifier) → 0.561

데이터 수집 및 데이터 베이스 구축



- 멜론 뮤직에서 월간 TOP 100 목록 수집 (1980 ~ 2023)
  - TOP 노래에 대한 meta-data 획득 후 `qurey="제목 + 아티스트명"` 로 검색 → 유튜브 링크 및 음원 데이터 확보

image_url	artist	title	top_year	class	youtube_top_url	yt_id	mood
str	str	str	i64	str	str	str	str
"https://cdnimg...	"윤시내"	"고목"	1980	"kpop"	"https://youtub...	"C-ZWHdSgrdA"	"romantic"
"https://cdnimg...	"희자매"	"그 사람 바보"	1980	"kpop"	"https://youtub...	"4lEhHkdLl-0"	"romantic"
"https://cdnimg...	"계은숙"	"기다리는 여심"	1980	"kpop"	"https://youtub...	"hpWvqLQSflY"	"romantic"
"https://cdnimg...	"윤항기"	"나는 행복합니다"	1980	"kpop"	"https://youtub...	"-6Zd1mhlHRA"	"happy"
"https://cdnimg...	"김학래"	"내가 (대상)"	1980	"kpop"	"https://youtub...	"ivkRj7NRz8g"	"romantic"

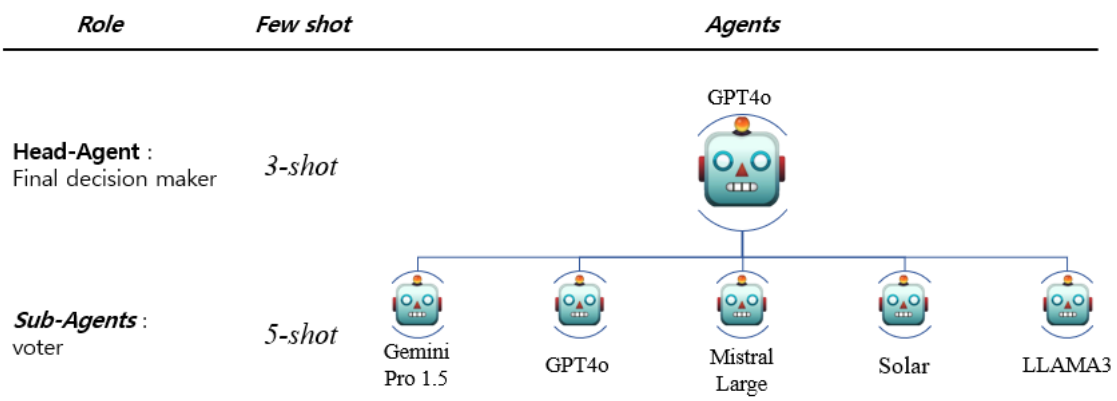
- 움원 감정 분류 모델을 이용하여 각 음원에 mood tag를 붙여줌

LLM API를 활용한 감정 분류기

- LLM API를 활용하여 BLIP 모델이 반환한 텍스트를 무드로 분류하는 Multi-Agent 구현
- 작동 방식
  - sub-agnets가 각자 자신이 생각하는 답과 그에 대한 판단 근거를 의견으로 제시

- head-agent가 5개의 의견을 종합 후 의견 타당성을 분석 및 다수결 원칙에 의거해서 최종 판단을 진행

- Multi-Agent 구조



- 사용 API
  - Head-Agent : `GPT4o`
  - Sub-Agents : `LLAMA3`, `Solar`, `GPT4o`, `Gemini 1.5 pro`, `Mistral Large`
- 모델 파인튜닝 방법
  - 퓨샷 러닝
    - Head-Agent : `3-shot learning`
    - Sub-Agents : `5-shot learning`
  - Prompt

▼ Head-Agent

```
You are a bot responsible for aggregating sentiment 

happy
romantic
sad
aggressive
dramatic

Analyze the provided opinions, consider the rationale
```

```
###
Here are some examples:

Situation: 'a street in the city of Havana, Cuba'
Opinions:
- The category is "happy" and the reason is "Havana
- The category is "happy" and the reason is "Streets
- The category is "dramatic" and the reason is "Hava
- The category is "happy" and the reason is "Havana,
- The category is "happy" and the reason is "Havana,
Category: happy
Explanation: Most opinions lean towards 'happy' with

Situation: 'a street with a tall building in the bac
Opinions:
- The category is "dramatic" and the reason is "A st
- The category is "dramatic" and the reason is "The
- The category is "None" and the reason is "The give
- The category is "None" and the reason is "The situ
- The category is "None" and the reason is "The give
Category: dramatic
Explanation: While the majority of opinions categori

Situation: 'two trees in the field'
Opinions:
- The category is "romantic" and the reason is "It e
- The category is "dramatic" and the reason is "The
- The category is "None" and the reason is "The sigh
- The category is "None" and the reason is "The situ
- The category is "happy" and the reason is "The pre
Category: happy
Explanation: While the opinions are varied and many
###

<<<
Situation : {insert inquiry context here}
Opinions : {insert inquiry opinion here}
>>>
```

▼ Sub-Agnets

```
You are a situational sentiment analysis bot. Your t

happy
romantic
sad
aggressive
dramatic

You must classify the given situation into the mood

###
Here are some examples:

Situation: A child is hiding behind a small tree to
Category: dramatic
Explanation: The child might be caught by the thief,
Situation: A man and a woman holding hands under a c
Category: romantic
Explanation: The cherry blossom setting creates a ro
Situation: Two gladiators are fighting with swords.
Category: aggressive
Explanation: Fighting with swords is a threatening a
Situation: A little child running with a basket full
Category: happy
Explanation: The child feels happy with a basket ful
Situation: A child is lying on the roadside after fa
Category: sad
Explanation: The child might be crying or in pain af
###

<<<
Situation: {insert inquiry text here}
>>>
```

추천 시스템

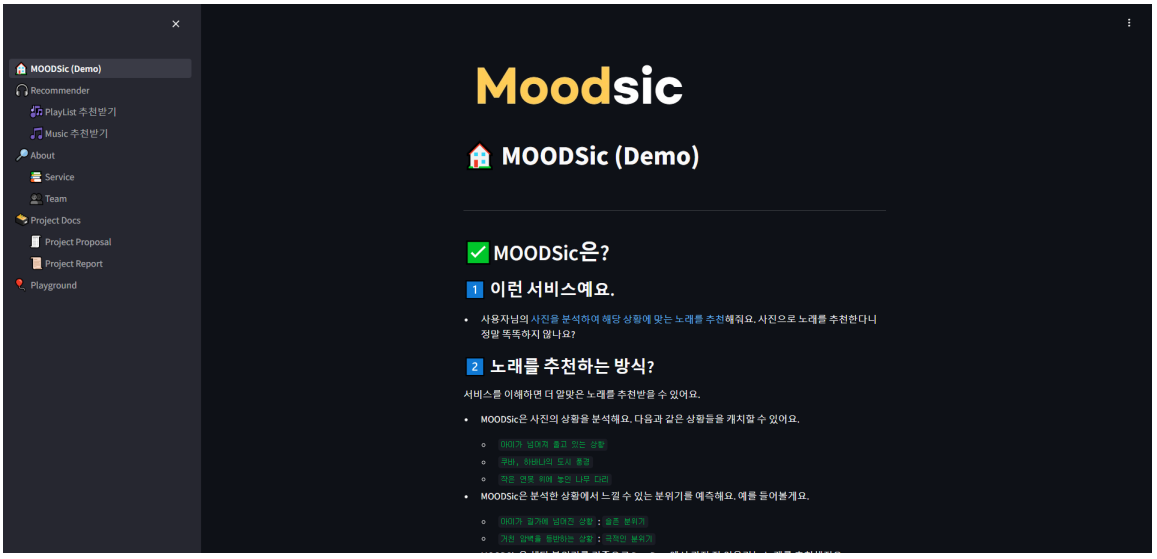
- 주어진 조건에 맞는 음악 데이터 중 Random Sampling으로 추천하도록 단순하게 구현
- 사용자가 장르(`kpop` , `pop` , `both`) 선택 가능
- 사용자의 사진에서 읽어낸 무드에 맞는 노래를 랜덤으로 추천하는 방식
- 데이터 베이스는 csv 파일 형태를 dataframe으로 읽는 방식으로 임시 구현

웹 설계

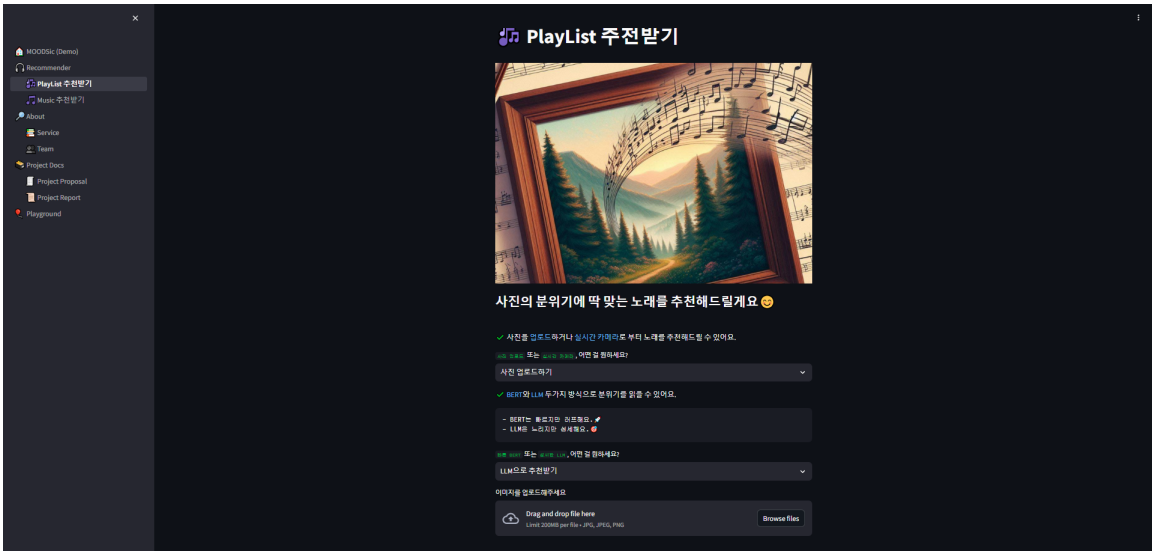
- Streamlit을 이용하여 구현
- colab 환경에서 돌아가도록 구현 (colab 노트북)

- 구현 서비스 화면

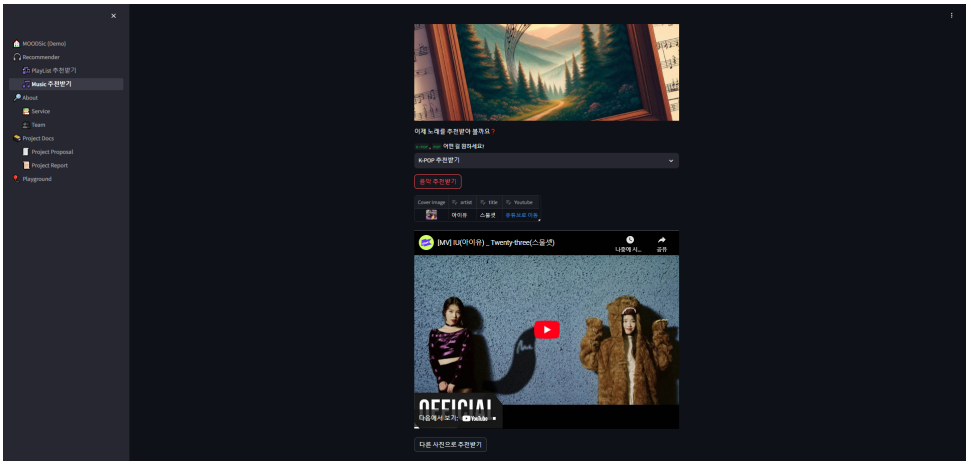
▼ `Home`



▼ `PlayList 추천 페이지`



▼ 실제 음악 추천 화면



5. 프로젝트 결과

기대효과 및 발전 방향

- 기존의 플레이리스트 채널처럼 상황과 감정을 검색해 입력할 필요가 없음(자동화 O)
- 당장 생각나는 노래가 없을때, 해당 기술을 통해 분위기에 맞는 음악을 들을 수 있음
- 특정 이미지에 대한 개인의 감정이나 추억을 음악으로 연결해줌으로써 더 깊은 정서적 연결을 만들어내고 그 노래들이 자기만의 플레이리스트가 될 수 있음
- 소셜 미디어에 공유하는 사진에 맞는 음악을 추천해주는 기능으로 서비스 활용도 가능

한계 및 개선점

- 더 좋은 개발 환경과 질 좋은 데이터를 확보했다면 여러 실험을 통해 더 좋은 성능이 나올 수 있을 것이라고 생각함
- 데이터를 좀 더 확보했다면 감정을 5가지로 한정하는 것이 아닌 더 많은 감정으로 나눠 분류 가능
- 사진으로부터 무드를 분석하는 것은 만족할 성능이 나왔지만, 실제 음원을 감정에 분류한다는 것이 생각했던 것보다 어려운 문제였고, 문제 정의가 잘 되지 않아서 모델링 성능을 개선하지 못함