

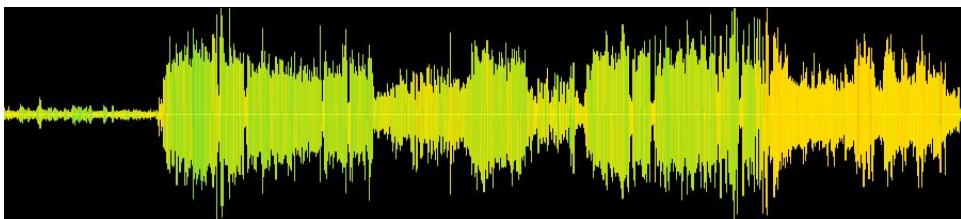


# AI Rush Tasks 4-6: Music Classification

Adrian Kim, NSMLxAutoML, Naver Corp.

## Task Description (rush4, rush5, rush6)

- 공통 목표
  - 변환된 음악파일을 input으로, 올바른 classification 결과를 output 해서 목표 metric을 올리자
  - 필요시 메타 정보 이용 가능



외로울 때

파티할 때

집중할 때

휴식할 때

## Task Description (rush4, rush5, rush6)

- Q1 (rush4): Station Classification
  - Single class classification
  - 4 classes: 외로울때, 파티할때, 집중할때, 휴식할때

top-1 accuracy

PM: 이창신

- Q2 (rush5): Multi-class Mood Classification
  - Multi class classification
  - 100 classes for mood
  - 1~5 tags for each track

micro-averaging f1-score

PM: 정영규

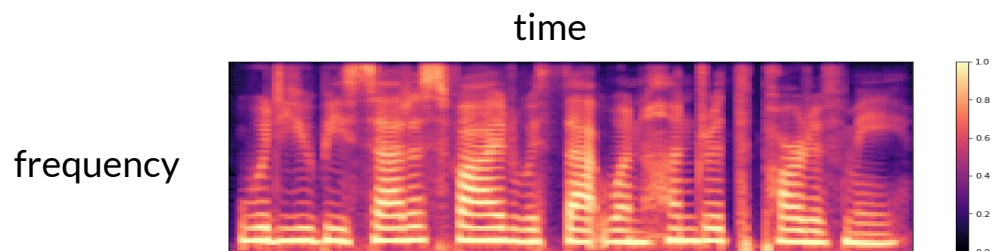
- Q3 (rush6): Japanese Music Genre Classification
  - Single class classification
  - 4 classes: 가요곡/엔카, R&B, 레게, J-ROCK
  - 음악적 특징을 반영한 장르 선정

top-1 accuracy

PM: 정영규

## Data Description (Input)

- Input: Mel-spectrogram (train\_data/mel\_spectrogram/\*.npy)
  - 음원 파일(.mp3, .wav)은 저작권상 이유로 제공 불가
  - 음원파일(time series)을 변환하여 얻을 수 있는 spectrogram 데이터가 주어짐
    - [1 x 주파수 대역(frequency bin) x 시간(time)]
    - number of mel bins = 128
    - sampling rate = 16000
    - length of each time frame = 32ms
  - 각 음원마다 길이가 다르기 때문에 시간축이 고정되어있지 않음
- Note: 어떻게 쪼개어서 사용할지, 특정 부분을 공략을 할지 말지 등에 대한 결정이 관건



## Data Description (Meta Data)

- Input: Meta data
  - Q1, Q2
    - train\_data/meta/track\_meta.json
    - train\_data/meta/artist\_meta.json
    - train\_data/meta/lyric\_meta.json
    - train\_data/meta/album\_meta.json
  - Q3
    - train\_label
      - track\_title
      - artist\_name\_list
      - album\_name
  - Note: 텍스트 데이터를 적절하게 사용해 보는것도 도움이 되는 **task**가 있을 수 있습니다

## Data Description (Labels)

- nsmi.DATASET\_PATH/train/train\_label
  - Q1: {'track\_index': {data\_idx: track\_id},  
      'station\_name': {data\_idx: label}}
  - Q2: {'track\_index': {data\_idx: track\_id},  
      'mood\_tag': {data\_idx: list of labels}}
  - Q3: {'track\_index': {data\_idx: track\_id},  
      'genre': {data\_idx: label},  
      ...(기타 메타 데이터)}

## Baseline model

- 5 layer 2D CNN
  - 1 -> 64 -> 128 -> 256 -> 512
- mel-spectrogram의 첫 32초를 큼지막하게 chunk를 뽑아 input으로 넣음
- meta data는 사용 안함
  
- Note: 2D CNN이 과연 정답일까?
- Note: Data input을 어떻게 설계해서 task를 수행하기 위한 데이터를 효율적으로 사용할 것인가?
- Note: meta data는 사용 어떻게 할까?

# Evaluation

- 1주차: test data의 25%, 2주차: test data의 50%, 3주차: test data의 100% 사용
- Q1, Q3 - Easy
  - top 1 accuracy
  - Q1: 200곡에 대해 평가
  - Q3: 400곡에 대해 평가
- Q2 - Hard
  - micro f1 score (threshold = 0.5)
  - 98곡에 대해 평가



Q/A