

# 비트세이버 나이도 재분류기

과목: 기계학습의 이해 학번: 202400684 이름: 최현빈

## 1. 프로젝트 개요

기존 비트세이버 레벨에서 표기 나이도와 체감 나이도의 편차가 상당한 상황에서 여러 레벨의 데이터를 기반으로 맵의 각 위치의 초당 노트수를 분석하여 다시 새로운 나이도를 예측하는 서비스를 만들었습니다.

## 2. 진행 과정

스스로 여러번 사용하고 지속적으로 유지, 보수 작업을 하고 싶어 할 결과물을 만들고 싶었기에 자신이 즐겨했던 게임 중, 수치로서 데이터를 구하기 쉬운 리듬게임을 선정하였습니다. 그리고 배경으로 재생되는 음악과 맵 전체의 노트 숫자 외에 곡을 분류할 만한 개선된 기준이 필요하다는 것을 체감하였기에 이를 바탕으로 나이도를 재분류하는 모델을 만들기로 했습니다.

데이터는 스스로가 첫 사용자임으로 제가 이미 가지고 있던 데이터를 활용하였습니다. 추후에 포럼이나 커뮤니티에서 선정된 인기맵들도 데이터에 포함시켜 훈련할 수 있습니다. 데이터의 형식은 처음엔 모두 같은 줄 알았으나 내부 구조가 게임이 업데이트되면서 모두 바뀌어 이를 하나의 형식으로 통합한 데이터셋을 만드는데 너무 많은 시간이 소요되었습니다.

모델은 많은 시간을 쏟을 수가 없게되어 비교적 단순한 MLP 모델을 사용하게 되었습니다. 학습 결과 validation에서도 65% 정도, 실제 test에서는 44%의 정확도가 나와서 실제로 이 상태로는 서비스로 만들기 힘들다는 판단이 섰지만 동시에 현재 맵에 부여된 나이도의 기준이 상당히 모호하고 일관적이지 못하다는 것도 알 수 있었습니다.

처음에는 gradio를 통해 web ui로 접근성을 높이려고 했지만 웹 상에서 사용자의 폴더에 접근하는 문제 등이 있어, 사용자가 다운로드하여 실제 자신의 비트세이버 폴더에 추가하여 사용하는 구조로 만들었습니다

## 3. 서비스 구조

data\_analysis에서 사용한 raw data -> dataset 코드와 evaluation의 모델 로드 및 validation 코드의 일부, 그리고 간단한 CLI 형식의 코드를 더하여 서비스를 완성하였습니다.

서비스는 기본적으로 실제 CLI가 작동하는 코드와 데이터를 불러오고 추론하는 코드가 모두 각각의 파일로 구분되어 있으며 [model.py](#)의 경우 출력 형식을 유지한다면 가중치 파일과 함께 교체하여 다른 모델로도 사용이 가능합니다.

## 4. 실제 사용 결과

비트세이버 나이도 분석기

명령어: list

===== 레벨 리스트 =====

번호 이름(난이도)

0 Beat Saber(Easy)  
1 Beat Saber(Normal)  
2 Beat Saber(Hard)  
3 Beat Saber(Expert)  
4 Beat Saber(ExpertPlus)  
5 Magic(Easy)  
6 Magic(Normal)  
7 Magic(Hard)  
8 Magic(Expert)  
9 Magic(ExpertPlus)

명령어: diff 0  
난이도 재예측 결과  
표기 난이도: Easy  
재예측 난이도: Easy  
명령어: diff 1  
난이도 재예측 결과  
표기 난이도: Normal  
재예측 난이도: Easy  
명령어: diff 2  
난이도 재예측 결과  
표기 난이도: Hard  
재예측 난이도: Easy  
명령어: diff 3  
난이도 재예측 결과  
표기 난이도: Expert  
재예측 난이도: ExpertPlus  
명령어: diff 4  
난이도 재예측 결과  
표기 난이도: ExpertPlus  
재예측 난이도: ExpertPlus  
명령어: diff 5  
난이도 재예측 결과  
표기 난이도: Easy  
재예측 난이도: ExpertPlus  
명령어:

## 5. 배운 점 및 개선 방향

처음 의도했던 모델과는 많은 차이가 생겼습니다. 원 계획은 노트의 패턴을 분석하여 이를 활용하기 위해 노트를 단어처럼 사용하여 **embedding**하거나 노트 단위로 **BPE**를 시도하거나 모델 역시 데이터 자체에 집중하기 위해 **Auto encoding** 모델을 사용해볼 계획이었으나 데이터 처리 단계에서 너무 많은 시간이 소요되어 진행할 수 없었습니다. 추후에 이런 점들을 반영하여 데이터 수집 및 전처리 단계부터 모델링을 고려하여 설계할 예정입니다.