

# Mbtify

Spotify Track을 활용한 MBTI 분류 서비스

AI07

강동우

# CONTENTS

## 엠비티파이 소개

Mbtify

Track

MBTI

Spotify

## 제작 과정

Flask

MongoDB

API

ML

Heroku

Sqlite

## 데이터 시각화 및 DB 스키마

EDA

XGboost

Metabase

Scikit-Learn

Schema

## 에필로그

느낀점

한계

# 서비스 소개

## “Find MBTI using track”

Spotify에서 제공한 MBTI 별 Playlist를 통해  
추출한 트랙들을 기반으로 노래에 대한 MBTI 성향 예측

엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

# 서비스 시연

<https://mbtify.herokuapp.com>

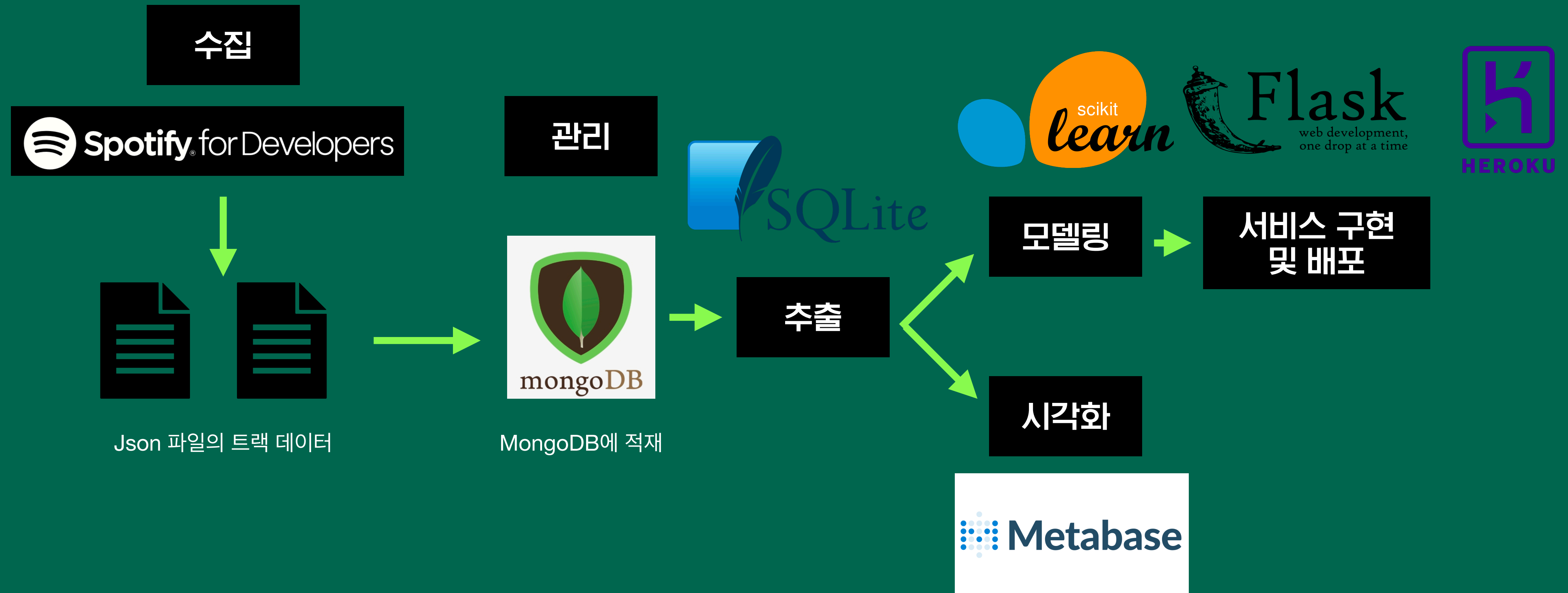
엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

# 파이프 라인 소개

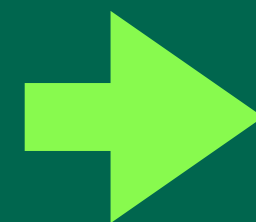


# 수집에서 저장까지

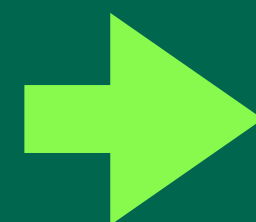
API

JSON

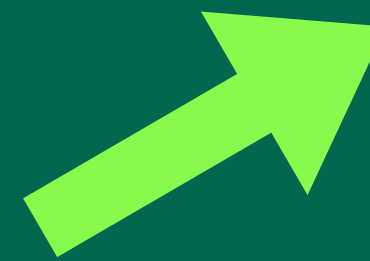
데이터 베이스에 저장



MBTI 별  
플레이리스트



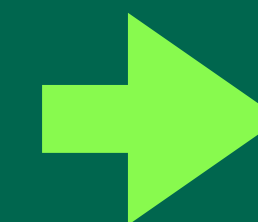
플레이리스트의  
트랙 정보



트랙의 기본 정보  
(가수 이름, 트랙 명칭 등)



트랙의 음악적 특징 정보  
(어쿠스틱함, 에너지틱함 BPM 등)



MongoDB에 적재

myFirstDatabase

tracks\_features

tracks\_info

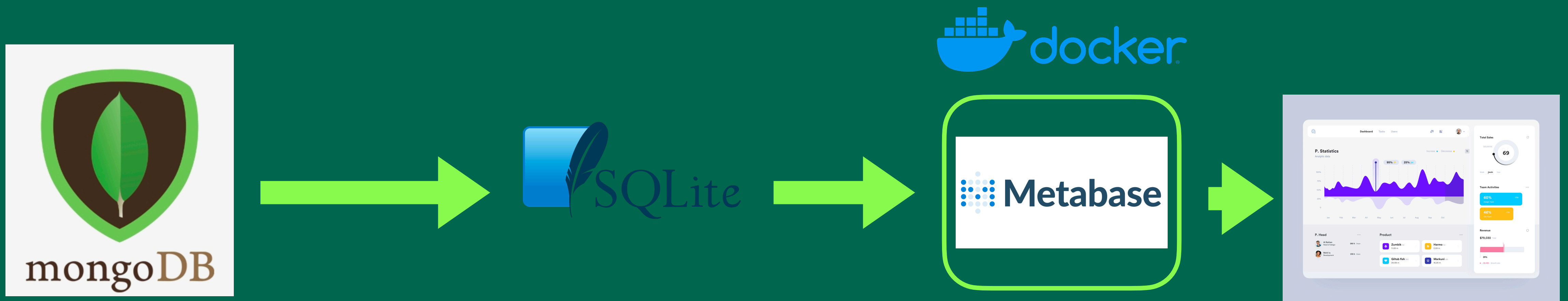
엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

# 추출에서 시각화



MongoDB 데이터를 Pymongo를 통해 LocalDB로 옮김.

Docker container로 db 파일을 가상환경에 copy

Dashboard로 데이터 분석

엠비티파이 소개

제작과정

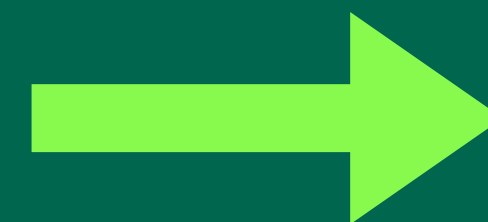
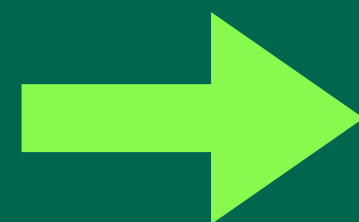
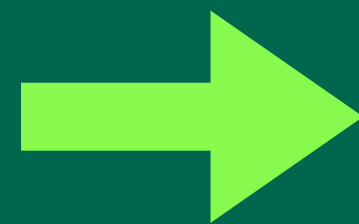
시각화 및 스키마

에필로그

# 모델링에서 API 서비스 배포



MongoDB 데이터를 Pymongo를 통해 LocalDB로 옮김.



모델링, API 파이썬 파일



서버 연결 WSGI 파이썬 파일



웹 페이지 구현 HTML Template  
(Send, recv, 404, 500)



엠비티파이 소개

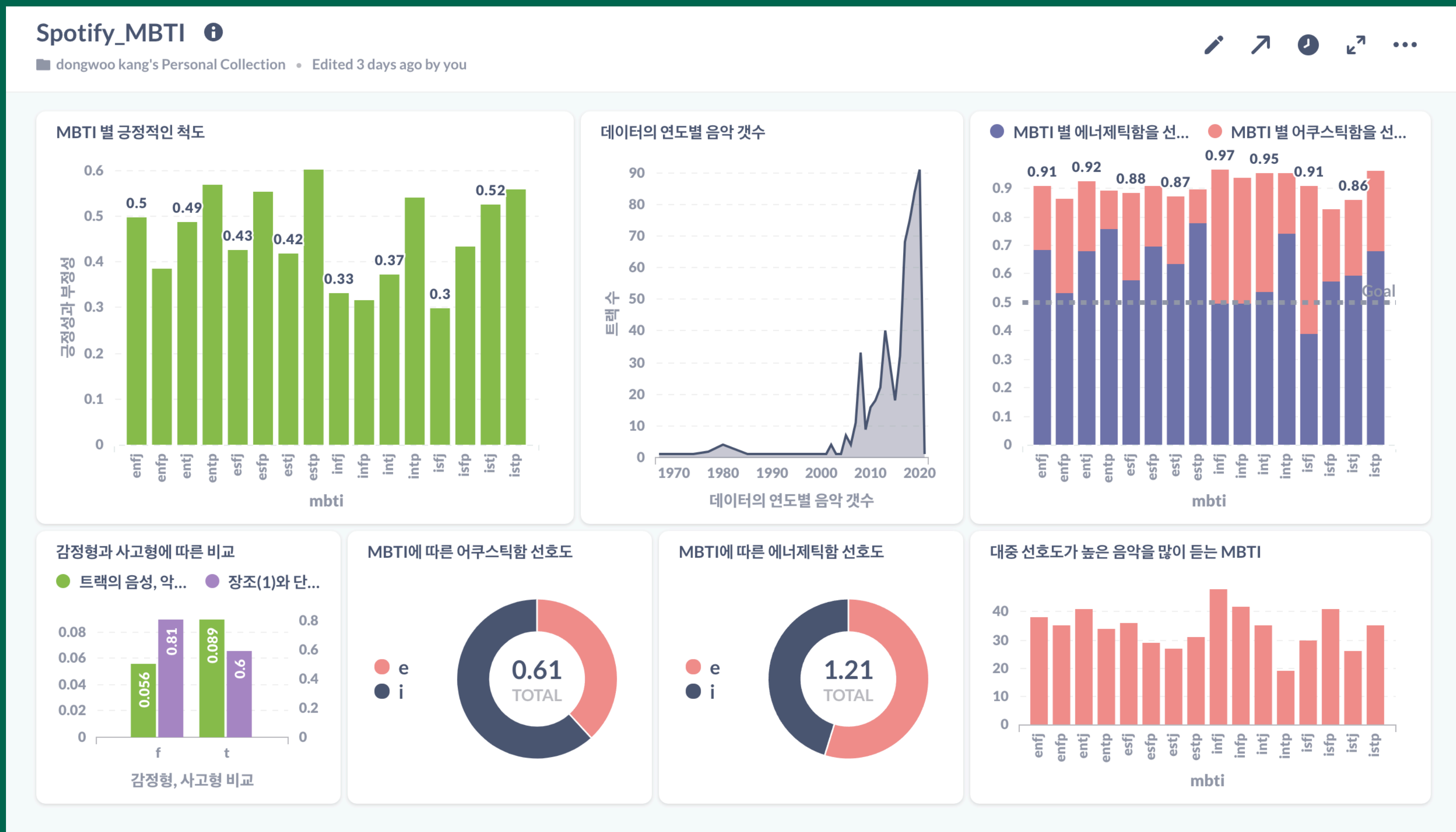
제작과정

시각화 및 스키마

에필로그



# 대시보드 시각화



엠비티파이 소개

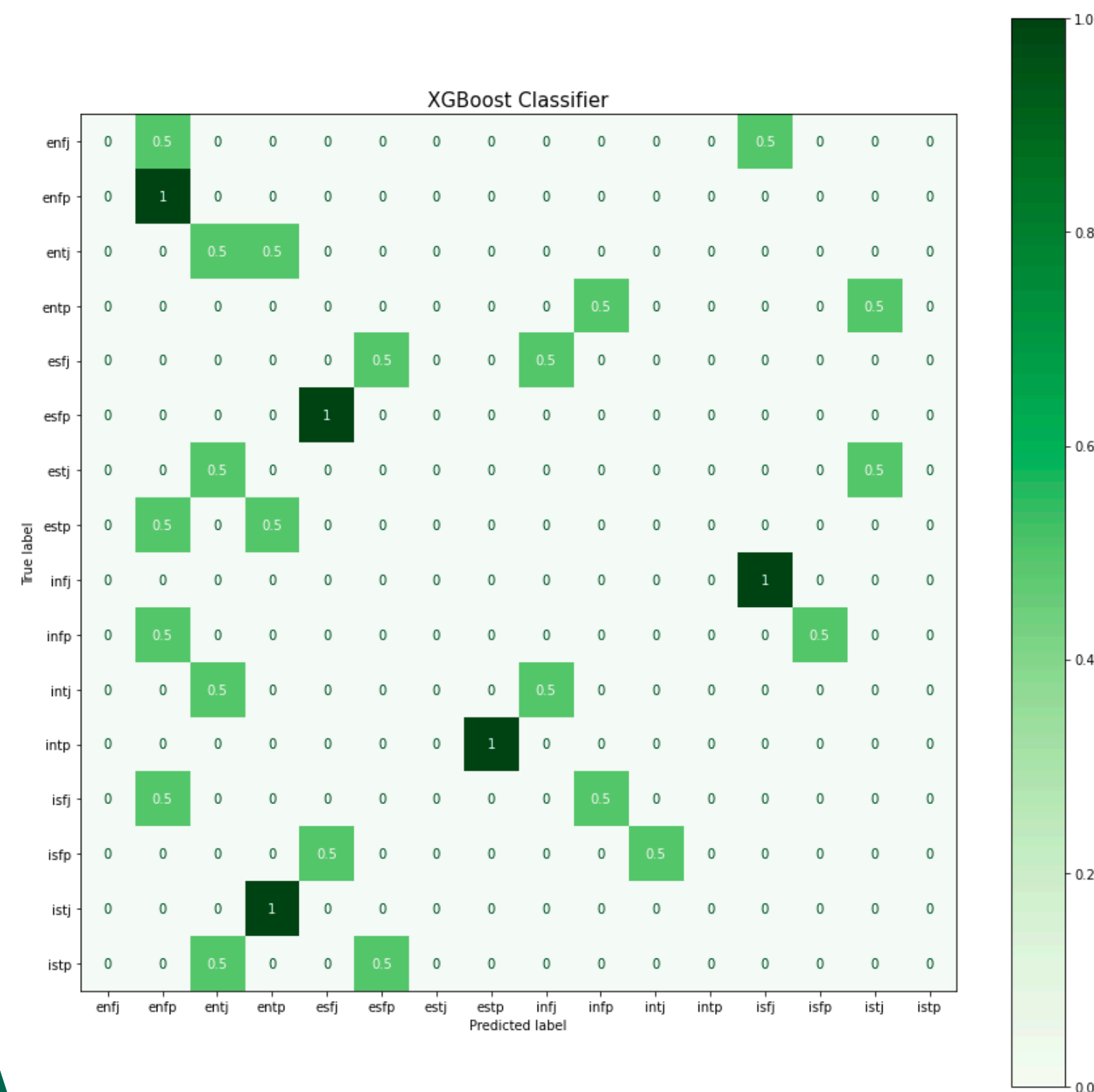
제작과정

시각화 및 스키마

에필로그

# 머신러닝 모델

## “XGboost Classification”



I랑 E[외향성(E)과 내향성(I)],  
F랑 T[감정형(F), 사고형(T)]는  
그래도 어느정도 분류가 된것 같지만

N(직관)이나 S(감각) 아니면  
J(판단)나 P(인식)가 잘 분류가 되지 않았습니다.

엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

# 데이터베이스 스키마

## Info

**\_id** VARCHAR(32) NOT NULL PK  
**Id** VARCHAR(32) NOT NULL  
**Artist** VARCHAR(32)  
**image\_url** VARCHAR(32)  
**release\_date** VARCHAR(32)  
**Name** VARCHAR(32)  
**Popularity** INT  
**MBTI** VARCHAR(32)

## Features

**\_id** VARCHAR(32) NOT NULL PK  
**Id** VARCHAR(32) NOT NULL FK  
**Acousticness** FLOAT  
**energy** FLOAT  
**Instrumentalness** FLOAT  
**key** INT  
**liveness** FLOAT  
**Loudness** FLOAT  
**Mode** INT  
**Speechiness** FLOAT  
**tempo** FLOAT  
**Valence** FLOAT  
**dancability** FLOAT



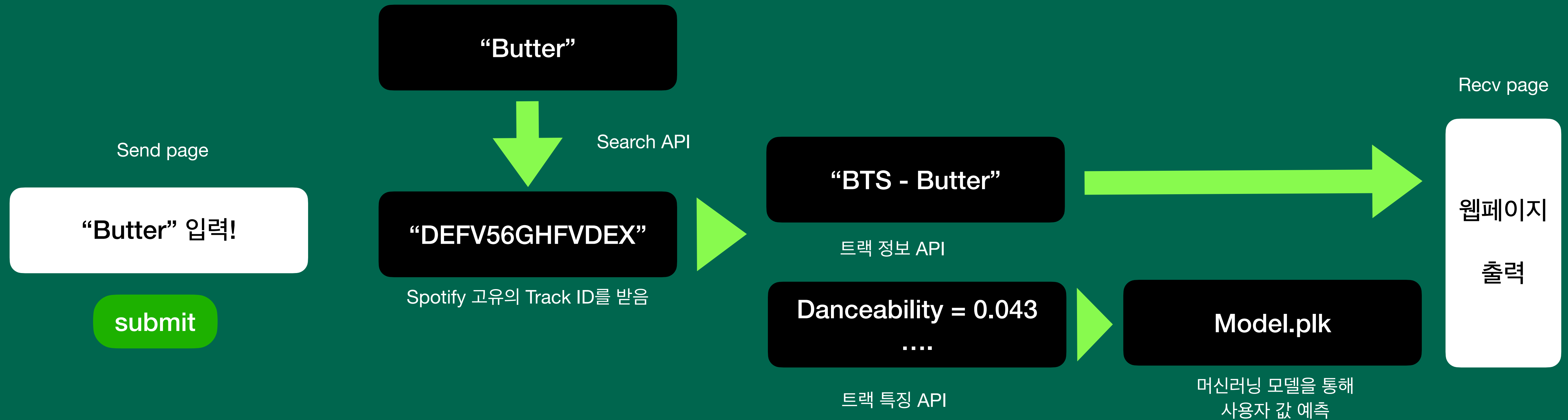
엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

# 유저 서비스 사용 프로세스



# 한계점

## “데이터 부족과 다중분류의 한계”

생각보다 Playlist를 다 모은다 해도 API에서 주어지는 제한이 있을 뿐 더러 생각보다 많은 데이터가 아니었다.

그리고 Xgboost를 이용해 다중분류를 진행했는데, 분류 대상이 많아지다보니 정확도가 10~20%정도라는 결과가 나왔다.  
군집화를 이용해 분류를 해보려고 했으나 아직 더 공부가 필요해 보인다.

# 느낀점

## “Data 분야 모든 프로세스 경험”

데이터 수집부터 저장 그리고 분석 서비스 개발 배포 까지  
데이터 분석가, 사이언티스트, 엔지니어 그리고 웹 개발자의 단계까지  
경험을 해보았다. 전체적인 흐름을 파악할 수 있었던 프로젝트 였다.

음원 플랫폼에서 이 서비스를 일회성 마케팅 용도로 사용하면  
많은 유입이 있지 않을까 싶다.

엠비티파이 소개

제작과정

시각화 및 스키마

에필로그

감사합니다.

