

# # 둥꿁둥꿁 # 음악추천

카카오 아레나 데이터를 활용한 음악 플레이리스트 추천 서비스

Melón kakao multicompus

# 팀 구성 및 역할



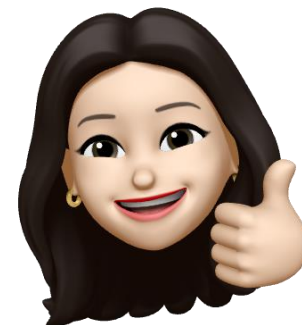
황인범

- 웹 프론트 개발
- 웹, DB 기획
- 웹 백엔드 개발
- DB 구현
- 알고리즘 분석



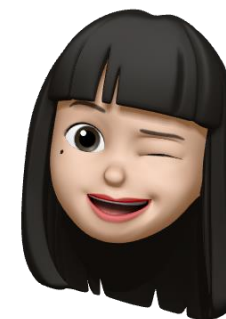
정한솔

- 웹 프론트 개발
- 웹 화면 기획서  
작성
- 알고리즘 분석
- PPT 제작



오혜수

- 웹, DB 기획
- 웹 백엔드 개발
- DB 구현
- 알고리즘 분석
- PPT 제작



박세진

- 웹 프론트 개발
- 웹 백엔드 개발
- 알고리즘 분석
- 데이터 크롤링
- PPT 제작

# Contents

1. 프로젝트 배경

---

2. 프로젝트 준비과정

---

3. 프로젝트 수행 결과

---

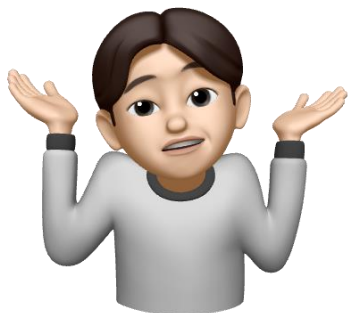
4. 활용 및 발전 방안

---

01

## 프로젝트 배경

---



음원 사이트의 곡들을 모두 들으려면 얼마나 오래 걸릴까?

음원 사이트 곡이 44,629,583곡이니까  
곡 당 4분으로 계산하면 339년이나 걸리네? Oh My God !!!!!



11,335

Today Album

110,314

Today Track

4,427,186

Total Album

44,629,583

Total Track

## Needs



## Solution

너무 많은 음원  
매일 쏟아지는 수많은 신곡  
차트와는 일치하지 않는 내 취향

“ 추천 시스템을 통해  
취향에 맞는 음악을 추천하여  
플레이리스트를 만들어보자 !”

## 기존 주제

사용자 데이터와 플레이리스트 데이터를 통해  
하이브리드형 추천 시스템을 구현





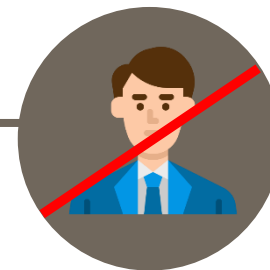
## 문제점

알고리즘의 성능 평가 불가



## 해결책

카카오 아레나 대회에 결과물  
제출을 통해 점수 확인



## 문제점

유저 기반 데이터가 없음



## 해결책

컨텐츠 기반의 추천시스템 구현

## 최종 주제

카카오 아레나에서 제공하는 **플레이리스트 데이터**를  
사용하여 **컨텐츠 기반 추천 시스템**을 구현

## AS-IS

## TO-BE

01

사용자가 청취한 노래 기반이나  
음원 사이트에서 제공하는  
플레이리스트만 청취 가능



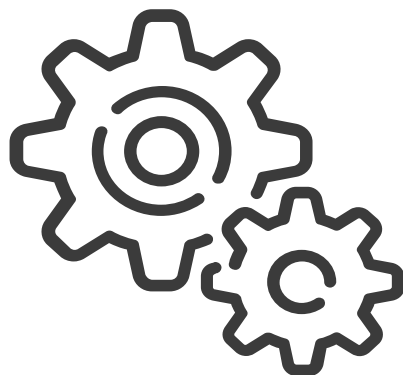
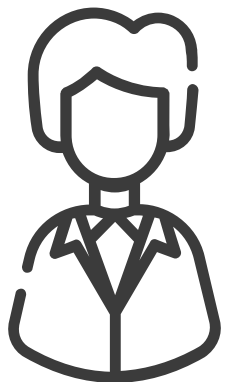
자신의 기분이나 상황에 따라  
노래 / 태그 선택하여  
플레이리스트 추천 가능

02

노래를 하나씩 공유함



추천 받은 나만의 플레이리스트를  
공유 가능하게 함으로써  
좀 더 폭 넓은 음악을 공유 가능



노래 / 태그 선택

모델링에 적용

비슷한 노래 /  
플레이리스트를  
생성

## 데이터베이스



## 개발 환경



## 프레임 워크

django

## 라이브러리



NumPy



## 협업



# 프로젝트 준비과정

---

1) Model

2) Web

02



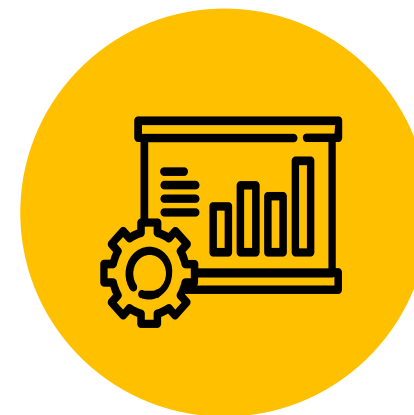
## 데이터 수집

- 카카오 아레나에서 제공된 30GB 멜론 데이터
- 웹 구현을 위한 크롤링 데이터



## 모델 선택 과정

- Word2Vec, FastText, Neighbor, KNN, Khaiii 등 다양한 알고리즘



## 모델 분석 결과

- 실제 멜론 대회에 참가해 정확도 확인

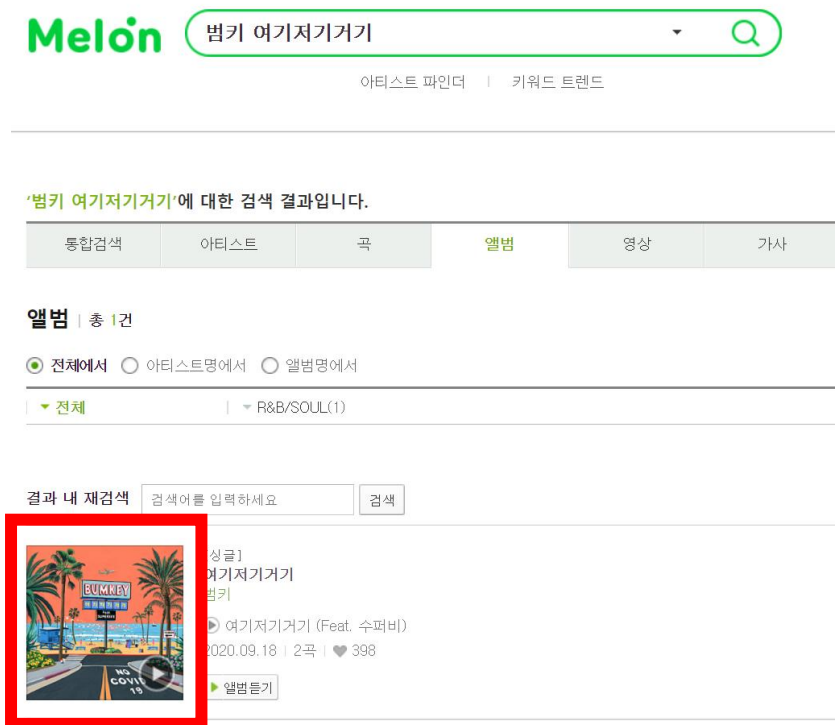
# 데이터 수집



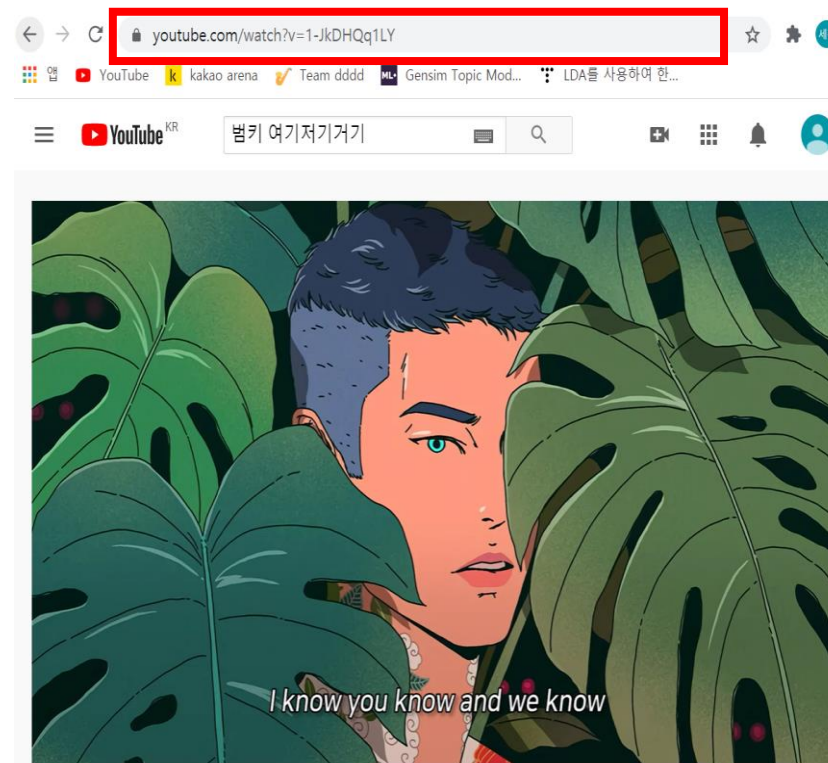


## “ 웹 구현을 위해 필요한 데이터 ”

### ① 멜론 앨범 이미지 크롤링

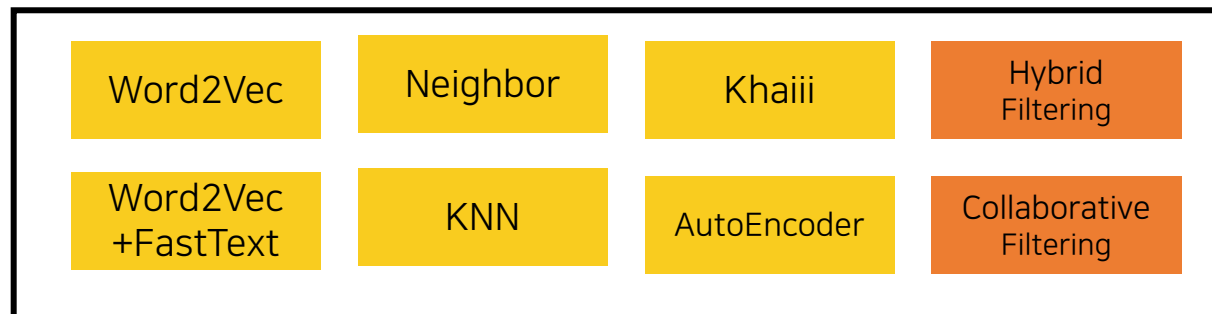


### ② 유튜브 URL 크롤링



# 모델 선택 과정

## 다양한 알고리즘



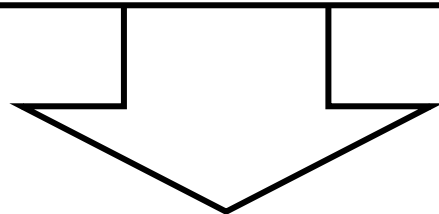
### 1차 모델 분류

- 유저 데이터 부재
- Hybrid & collaborative filtering 분석 불가

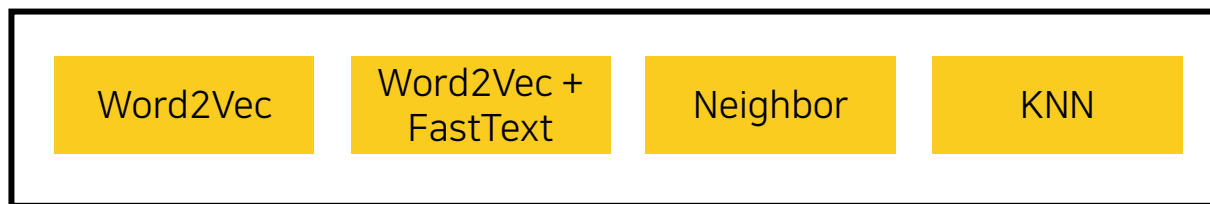


### 2차 모델 분류

- 컴퓨팅 사양 부족
- Kahiii & autoencoder 분석 불가



## 모델 선택 과정



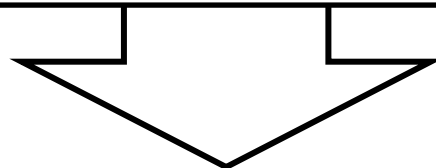
데이터 양이 많아  
분석에 많은 시간 소요



Data 개수	Song	Tag	Total
130,000	0.250919	0.471058	0.28394
50,000	0.170272	0.389103	0.203097
40,000	0.170215	0.390988	0.203331
30,000	0.186839	0.414329	0.220963
20,000	0.184976	0.417139	0.2198
10,000	0.142083	0.345522	0.172599



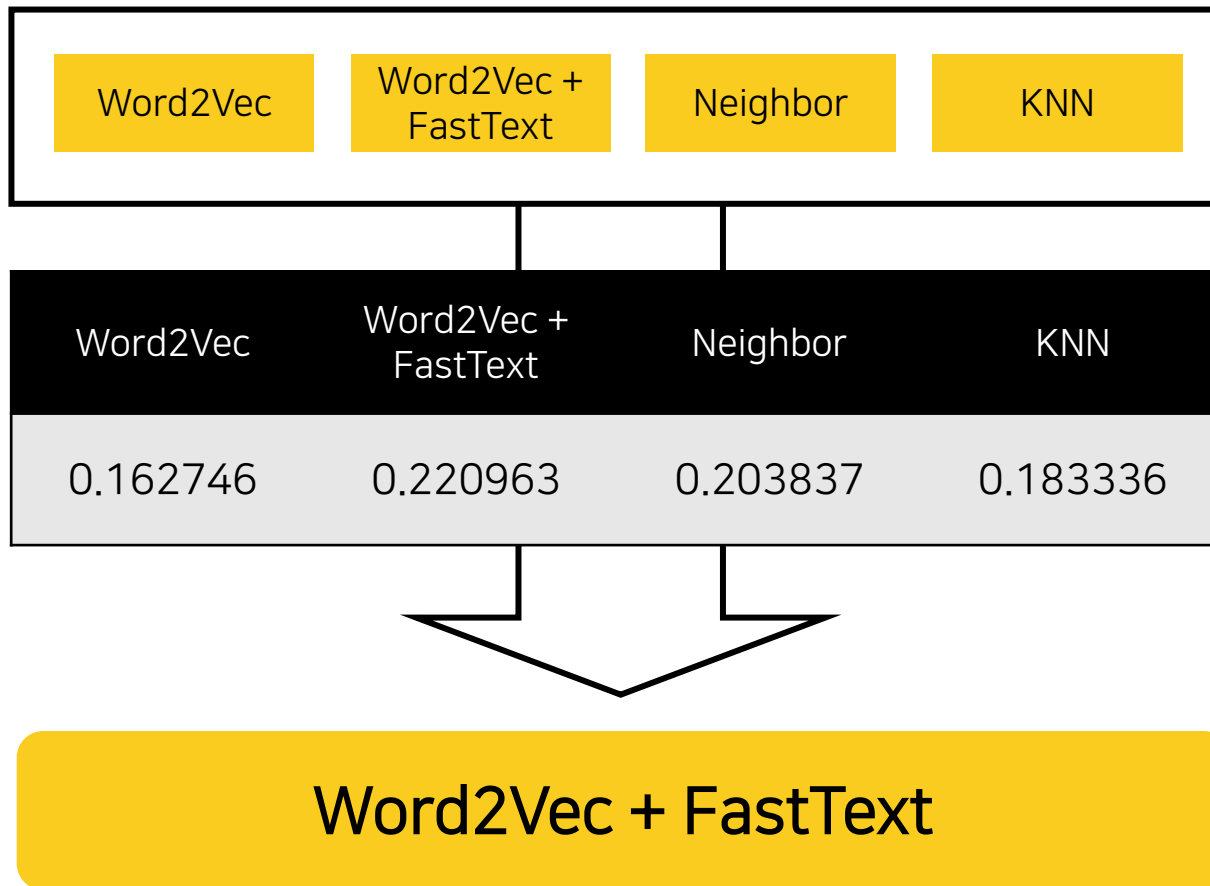
“ Word2Vec + FastText ”  
알고리즘으로 데이터 개수를 줄여서  
분석의 정확도를 비교해보자



## 2. 프로젝트 준비과정

### 1) Model

# 모델 선택 과정



3만개의 데이터로 알고리즘 구현

전체 데이터로 정확도를 비교 후  
최종 알고리즘 선택

전체 데이터로 학습한 결과

Song : 0.251

태그 : 0.471

Score : 0.2841

분석 걸리는 시간 : 6시간

# 모델 결과

“ Play list 예측 전 후 ”

	id	plylst_title	songs	tags
0	118598		[373313, 151080, 275346, 696876, 165237, 52593...	[]
1	131447	앨리스테이블	[]	[]
8	80810	리듬타면서 뽀시게 운동하자!!! (스트레스 날리자):}	[]	[스트레스]
11	136650		[174132, 257367, 539802, 568691, 296891, 38471...	[아침, 눈, 크리스마스]



	id	plylst_title	songs	tags
0	118598		[373313, 151080, 275346, 696876, 165237, 52593...	[OST, 디즈니, 애니메이션, 픽사, 기분 전환, disney, 디즈니OST, 영원...
1	131447	앨리스테이블	[549178, 348200, 140867, 298056, 394622, 42982...	[팝, 기분전환, 드라이브, 감성, 락, 힙합, 레이블, 힐링, 알앤비, 아이돌]
8	80810	리듬타면서 뽀시게 운동 하자!!! (스트레스 날리자):}	[643498, 420396, 700850, 355939, 285157, 70475...	[스트레스, 기분전환, 운동, 댄스, 드라이브, 힙합, 클럽, 헬스, 여름, 일렉,...
11	136650		[174132, 257367, 539802, 568691, 296891, 38471...	[아침, 눈, 크리스마스, 캐롤, 성탄절, 겨울, Christmas, 연말, Car...

## “ 카카오 아레나 리더보드에 도전한 결과 ”

kakaoarena

대회 목록   데이터셋   포럼   공지/문의

로그인

플레이그라운드

Melon Playlist Continuation

플레이리스트에 있는 곡들과 어울리는 곡들을 찾아주세요

3달 전 시작

참여 팀 수 60

총 상금 ₩0

메뉴

개요   데이터   리더보드   포럼

다른 대회 선택

리더보드

공개 리더보드

#	△1d	팀명	멤버	스코어	곡 nDCG	태그 nDCG	제출 횟수	마지막 제출
1	▲ 4	동공동공		0.284174	0.251201 (2)	0.471018 (2)	8	9시간 전
2	▼ 1	세진이와아이들		0.28394	0.250919 (3)	0.471058 (1)	2	10일 전
3	▼ 1	hh		0.276863	0.244249 (4)	0.461677 (3)	1	15일 전
4	▼ 1	js		0.273414	0.255058 (1)	0.377428 (6)	2	21일 전
5	▼ 1	귀여운 회귀		0.267297	0.234293 (5)	0.454323 (4)	4	한 달 전
6	-	멜론파이브		0.225228	0.186099 (6)	0.446957 (5)	36	한 달 전
7	-	MTG		0.028538	0.005149 (8)	0.161075 (8)	7	25일 전
8	-	음잘알		0.025581	0.000129 (9)	0.169805 (7)	2	10일 전
9	-	TASP		0.010762	0.008083 (7)	0.025944 (9)	8	한 달 전

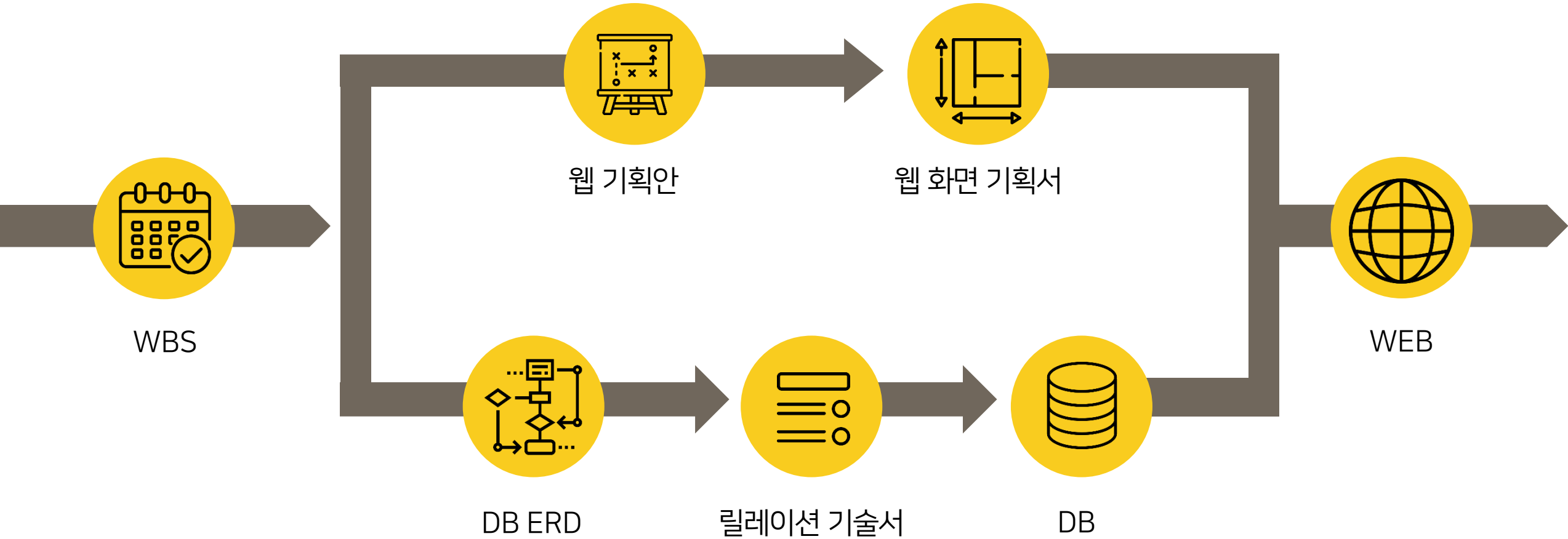
# 프로젝트 준비과정

---

- 1) Model
- 2) Web

02

# Preview





## 2) Web

# WBS



(

Project 명		(프로젝트 명 작화)				전체진행 :			현재날짜:		2020-11-16																										
구분	공정	완료율	상호율	기간	시작 날짜	완료 날짜	담당자	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	13월	14월	15월	16월	17월	18월	19월	20월	21월	22월	23월	24월	25월	26월	27월	28월	29월	30월
프로젝트 관리	문제 정의		50%	문제정의서	2일	2020-11-16	2020-11-17																														
	현장성 검토	100%	현장성검토	1일	2020-11-16	2020-11-16																															
	문제 분석	0%		2일	2020-11-17	2020-11-18																															
	1차 Photo Type	0%		2일	2020-11-19	2020-11-20																															
	원격	0%		4일	2020-11-21	2020-11-24																															
	2차 Photo Type	0%		5일	2020-11-25	2020-11-29																															
	구현	0%		20일	2020-11-30	2020-12-19																															
	Alpha Test	0%		3일	2020-12-13	2020-12-15																															
	Beta Test	0%		4일	2020-12-16	2020-12-19																															
	완료 Test	0%		3일	2020-12-20	2020-12-22																															
결과	0%		2일	2020-12-23	2020-12-24																																
프로젝트 수행	시스템조사	100%		1일	2020-11-16	2020-11-16																															
	주요시스템에 대한 사전정보 이해	17%		6일	2020-11-16	2020-11-21																															
	필요한 리소스까지 리서치	0%		6일	2020-11-19	2020-11-24																															
	데이터 수집	0%		8일	2020-11-22	2020-11-29																															
	데이터지 정제하기	0%		4일	2020-11-23	2020-11-26																															
	데이터지 구조 파악	0%		3일	2020-11-24	2020-11-26																															
	데이터지 이상치 파악	0%		2일	2020-11-25	2020-11-26																															
	데이터지 결측치 제거	0%		2일	2020-11-25	2020-11-26																															
	데이터지 간헐 분할	0%		2일	2020-11-27	2020-11-28																															
	데이터지 결합	0%		3일	2020-11-28	2020-11-30																															
	특수 로컬 학습 및 리서치	0%		6일	2020-11-30	2020-12-05																															
	현실한 예측 로컬 시험	0%		4일	2020-12-02	2020-12-05																															
	특수 로컬 구현	0%		5일	2020-12-03	2020-12-07																															
	시뮬레이션	0%		4일	2020-12-04	2020-12-07																															
	시뮬레이션 도구 학습 및 리서치	0%		3일	2020-12-05	2020-12-07																															
	현실한 시뮬레이션 도구 시험	0%		3일	2020-12-05	2020-12-07																															
	특수 로컬 시작점	0%		2일	2020-12-06	2020-12-07																															
	모델 분석	0%		3일	2020-12-03	2020-12-05																															
	각 예측 모델의 결과값 비교 및 분석	0%		3일	2020-12-03	2020-12-05																															
	시뮬레이션 구현			9일	2020-12-05	2020-12-13																															
DB 운영 / 데이터 관리	0%		3일	2020-12-05	2020-12-09																																
여름리제이션 구현	0%		16일	2020-12-08	2020-12-23																																
기초 UI 구현	0%		11일	2020-12-13	2020-12-23																																
서비스용 톨인 / 디버깅스 제어	0%		16일	2020-12-05	2020-12-20																																

## 2. 프로젝트 준비과정

### 2) Web

# 웹 기획



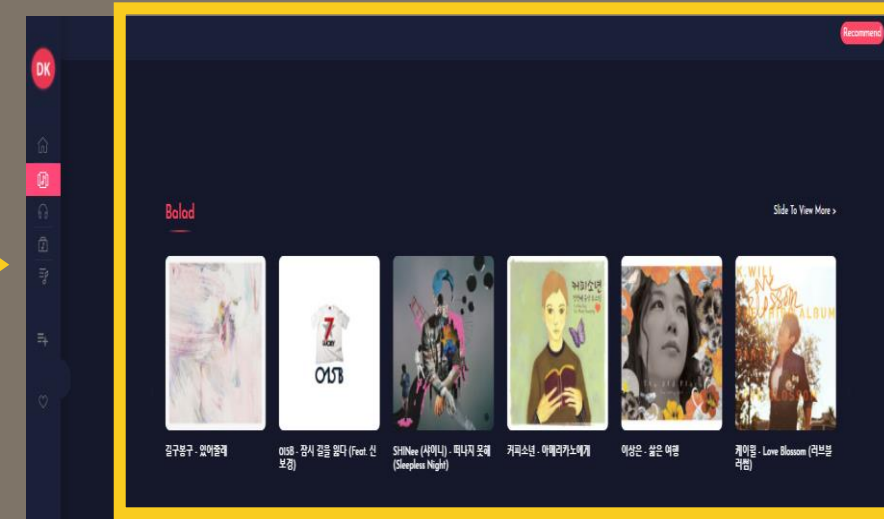
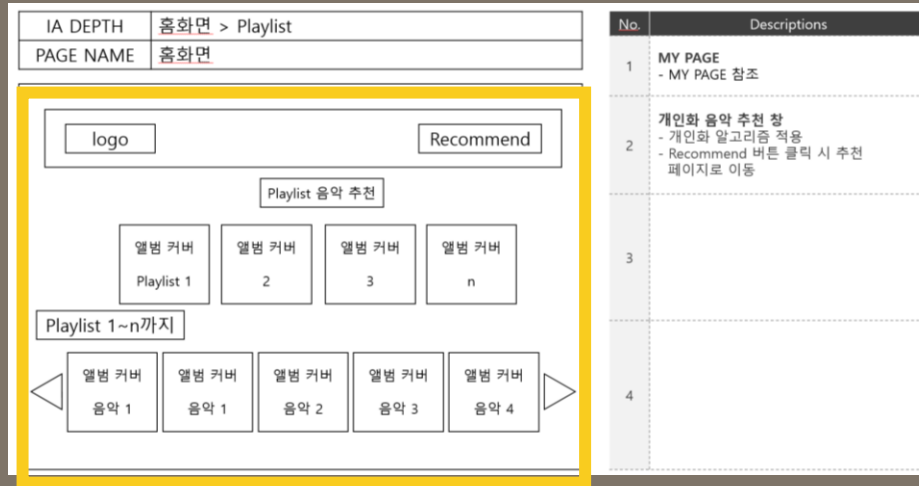
### 개발 요구사항 명세서

팀명 : 동공동공  
팀원 : 황인범, 오혜수, 박세진, 정한솔  
2020.11.12 요구사항 조안 작성 v1.0

목표 : 좋은 음악들을 널리 알리는 것이 목표 (shout out 정신) + 개인화 추천!

#### < 홈 화면 >

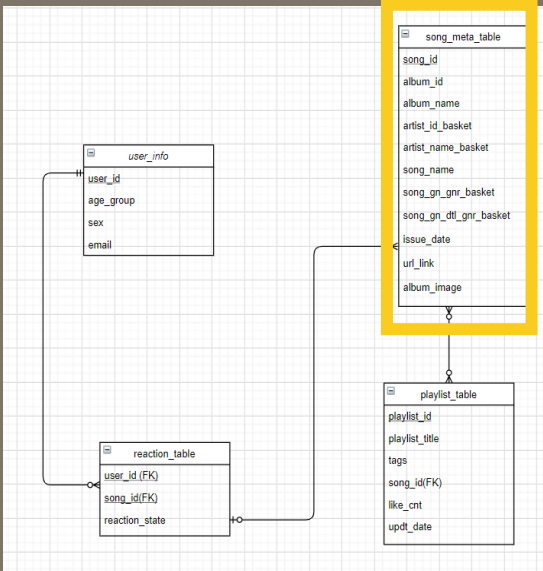
1. 음악 추천 창
  - 버튼 클릭 > 추천 창으로 이동
2. 탑15 노래
  - 좋아요 수가 많은 노래 15개 보여줌
3. 탑 장르 3개
  - 인기있는 장르 3개 보여줌
  - 앨범 3개 이미지로 보여줌
  - 아래에 장르에 따른 곡 리스트로 하나씩 보여줌
4. 탑 태그
  - 여러가지 태그들 표지를 보여줌
  - 이미지 클릭하면 태그 창으로 넘어감



[웹 기획안]

[웹 화면 기획서]

[실제 웹 화면]



개체명	노래 정보	릴레이션 기술서	작성일		2020.11.16	Page	
	song_meta_table		작성자		오혜수	2/4	
개체 설명	사용자의 정보를 저장하는 테이블						
Attribute		Data Type	Key	NOT NULL	UNIQUE	Domain	Default
노래 id	Song_id	INT(50)	PK	√	√		
앨범 id	Album_id	INT(50)					
앨범 이름	Album_name	Varchar2(200)					
가수 id 리스트	Artist_id_basket	VARCHAR2(200)					
가수 이름 리스트	Artist_name_basket	Varchar2(200)					
곡명	Song_name	Varchar2(200)					
곡 장르 리스트	song_gn_gnr_basket	Varchar2(200)					
곡 세부 장르 리스트	song_gn_dtl_gnr_basket	Varchar2(200)					
발매일	issue_date	date					

이름: song\_meta

코멘트:

열: + 추가 - 제거 ▲ 위로 ▼ 아래로

#	이름	데이터 유...	길이...	...	NULL...	0으...	기본값
1	song_id	INT	50				기본값 없...
2	album_id	INT	50		☑		NULL
3	album_...	VARCHAR	200		☑		NULL
4	artist_id...	VARCHAR	200		☑		NULL
5	artist_n...	VARCHAR	200		☑		NULL
6	song_n...	VARCHAR	200		☑		NULL
7	song_g...	VARCHAR	200		☑		NULL
8	song_g...	VARCHAR	200		☑		NULL
9	issue_d...	DATE			☑		NULL
10	url_link	VARCHAR	300		☑		NULL

[ERD]

[릴레이션 기술서]

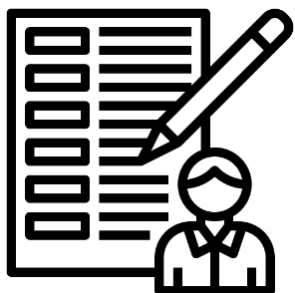
[DB]

03

# 프로젝트 수행 결과

---

# Service Flow



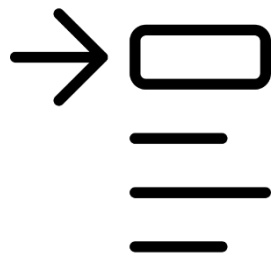
좋아하는 곡 / 태그 선택

Recommend 30%



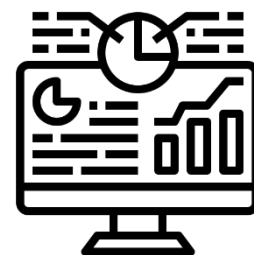
백엔드에서 모델링 적용

Recommend 50%



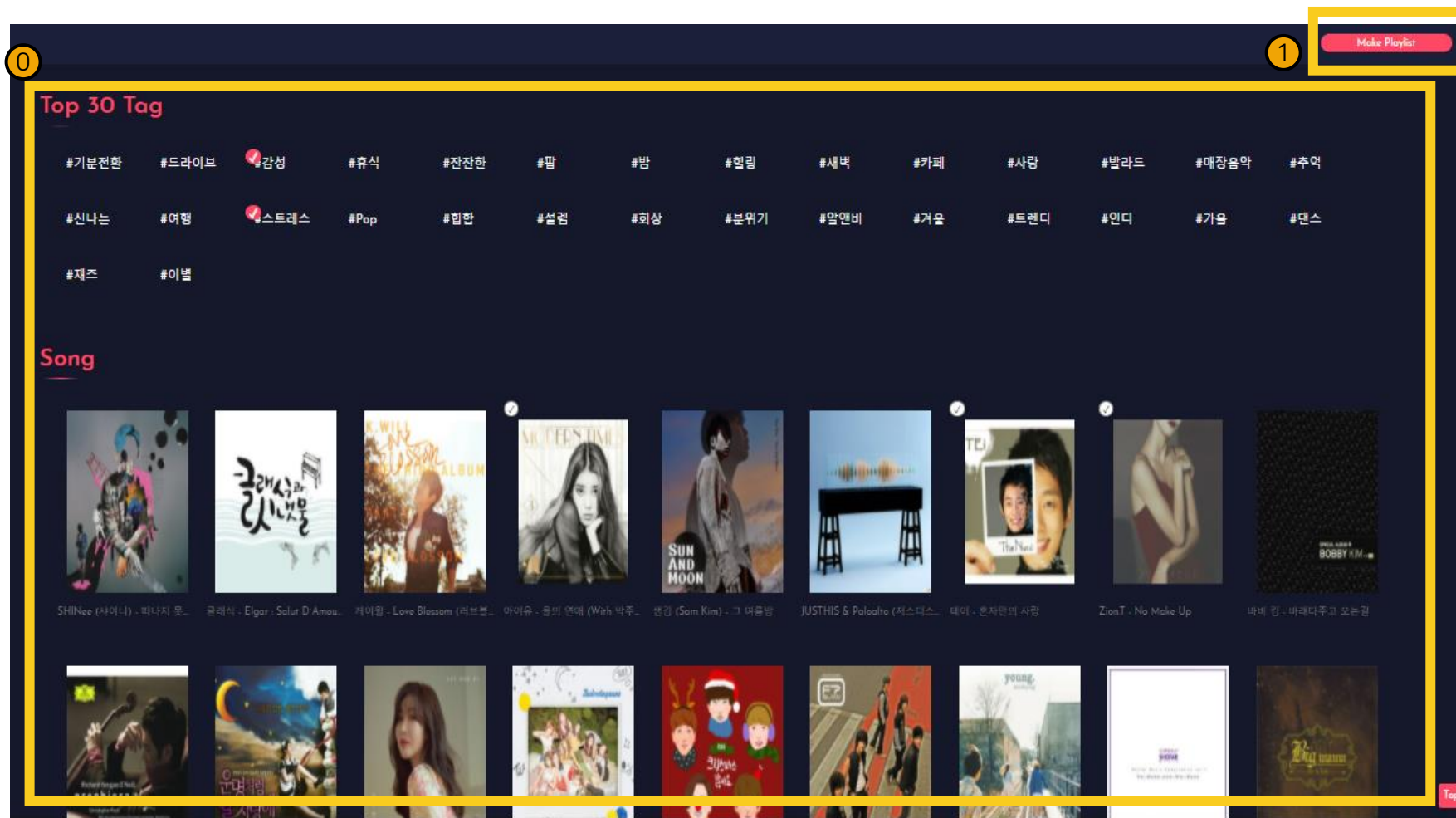
맞춤 플레이리스트 생성

Recommend 70%



웹 페이지에 곡 리스트 출력

Recommend 100%

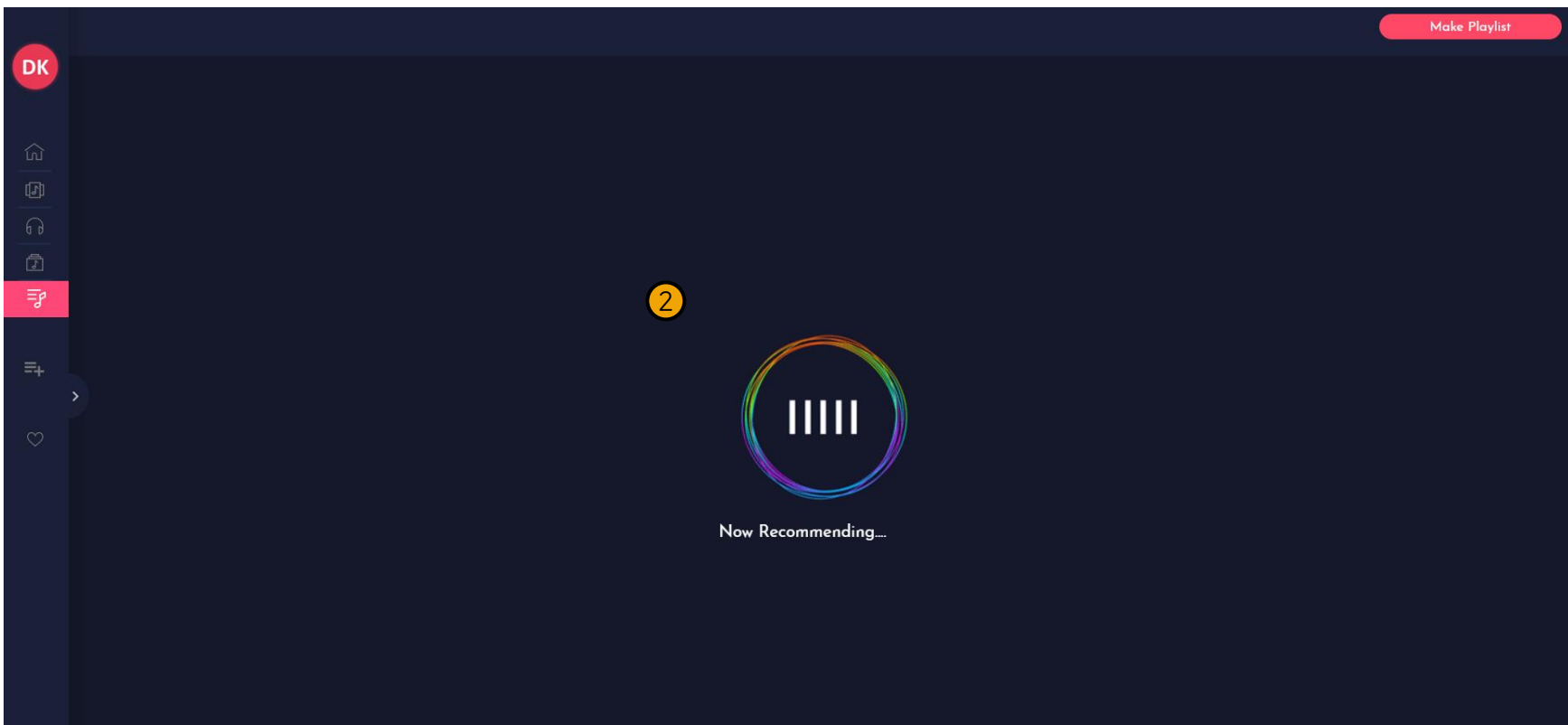


STEP 0.

선호하는 노래/태그 선택

STEP 1.

플레이 리스트 생성 버튼 클릭



STEP 2.  
모델링

3

Recommend Playlist					
#	Song Title	Singer	Release Date	Play	Share
1	연능연능	홍아	June 12, 2018	4	5
2	걱정말아요 그대	이적	Nov. 7, 2015		
3	꿀맛	정미애	Aug. 15, 2019		
4	소란했던 시절에	벌리어코스티	April 22, 2014		
5	Rain Drop	아이유	June 3, 2010		
6	홀로 (Feat. 김나영)	정키	Aug. 31, 2012		
7	너로 가득해	혜지 (Hyeji)	Dec. 4, 2017		
8	MOM (겨울나무)	홍아	Dec. 27, 2017		
9	고마운 사람 (Vocal By 이소진)	업라이트 (Upright)	April 18, 2019		
10	너 사용법	에디킴	April 11, 2014		
				5	All Share

## STEP 3.

웹에서 보여지는 추천 플레이리스트

## STEP 4.


Play 버튼 클릭 시 유튜브 url로 연결

## STEP 5.

Share 버튼 클릭 시 카카오톡으로 공유



Recommend









## Dung KKung

### Recommend Playlists ! :D

We analyze with melon data provided by Kakao Arena  
Choose your taste and get recommended playlists

Recommend

### Liked Song Top 15

01		VVS (Feat. JUSTHIS) 미란이, 먼치맨, Khundi Panda, 머쉬베놈	▶
02		내일이 오면 릴보이 (LILBOI)	▶
06		What Do I Call You 태연 (TAEYEON)	▶
07		헛된 건 사랑이 아니다 임창정	▶
11		Achoo (Feat. pH-1, HAON) 미란이 (Mirani)	▶
12		Life Goes On 방탄소년단	▶

# 활용 및 발전 방안

---

04

## Solution

태그와 곡을 선택하여  
나만의 맞춤 플레이리스트 생성



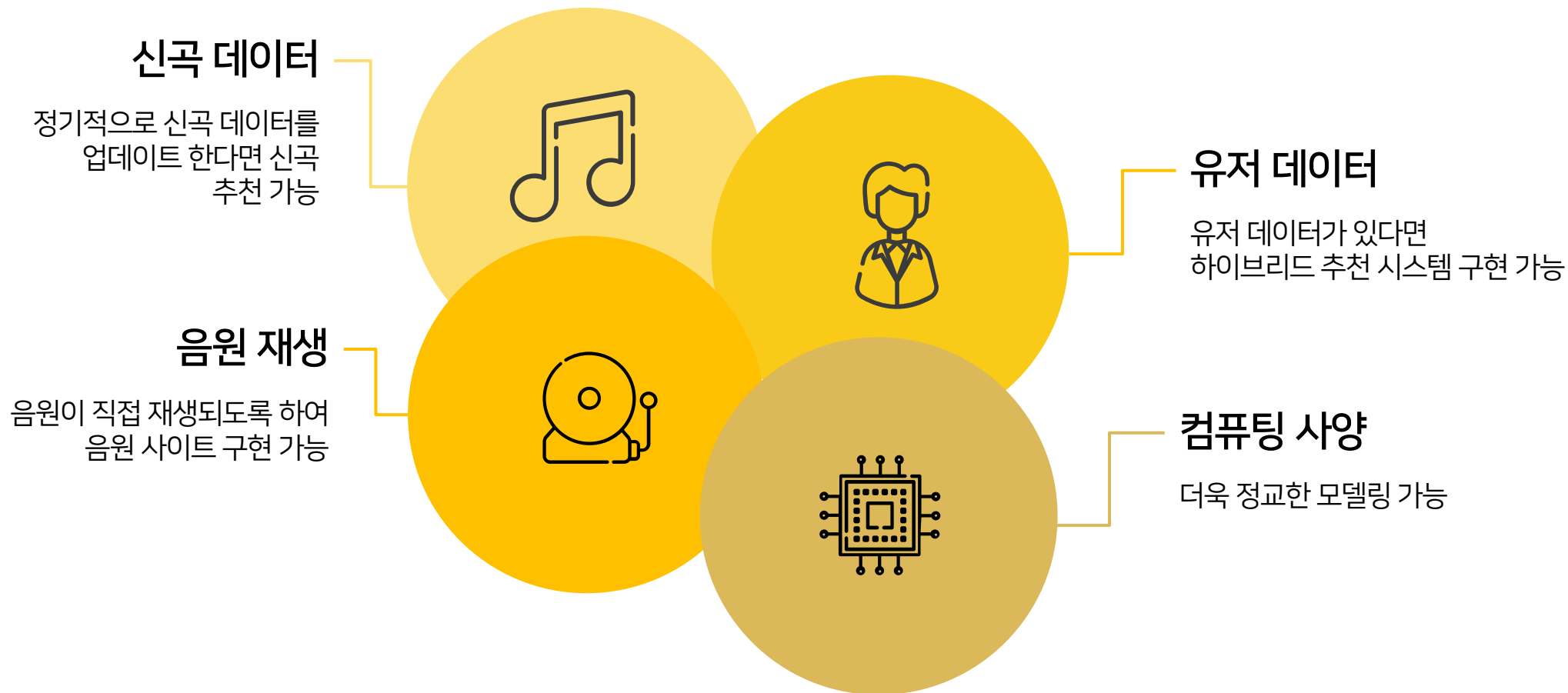
## Improvement

### 1. 매장음악 서비스

태그를 통해 매장에서 트는 음악 플레이리스트를 제공

### 2. 킬러 콘텐츠로서의 역할

반복되는 노래 추천이 아닌 내가 선택한 것에 따라 차별  
화 된 음악 추천을 받을 수 있음



느낌점

---

# 느낀점

구분	내용
프로젝트 수행상 어려움 극복 사례	카카오 아레나에서 제공 받은 데이터가 용량이 너무 커서 여러 알고리즘으로 분석하는데 시간과 메모리에 대한 어려움을 겪었다. 그렇지만 강사님과 멘토님들의 피드백과 팀원들의 원활한 커뮤니케이션을 토대로 데이터를 줄여서 구현해보자는 결론이 났고 그 결과 프로젝트를 무사히 마칠 수 있었다.
프로젝트에서 잘한 부분	추천 서비스부터 웹 페이지 구현까지 프론트와 백엔드를 도전하는 첫 프로젝트였다. 모두에게 생소한 프로젝트였지만 누구 하나 포기하지 않고 각자의 상황과 능력에 맞게 맡은 임무를 끝까지 해내려고 노력한 과정 속에서 시너지가 발휘했다.
프로젝트에서 아쉬운 부분	사용자 데이터가 없어서 처음 생각한 주제인 하이브리드 추천 시스템을 구현하지 못한 점이 아쉬웠다. 또한, 음성 분석 데이터인 Mel-spectrogram 데이터는 약 238GB로 분석을 시도조차 해보지 못했다. Mel 데이터에 있는 주파수 정보를 활용하여 다양한 분석을 하고 싶었지만, 해보지 못한 점이 가장 아쉬웠다.
프로젝트를 통한 진로설계, 취업분야 탐색 및 결정 등 도움	팀원들과의 협업을 통해 업무 분담 및 소통 능력을 길렀다. 직접 웹 페이지 구현을 해보면서 프론트, 백엔드 개발을 경험했고 DB와 웹을 연결하여 데이터베이스를 실제로 활용해보았다. 이번 프로젝트를 통해 기본적인 웹 개발 지식을 쌓을 수 있었고, 데이터 분석 뿐만 아니라 웹 개발이라는 분야에도 관심을 가지게 되었다.

# 느낀점



컴퓨터 사양이 부족해서 못돌려본  
알고리즘이 있어서 아쉬웠다. 또한 한번  
학습하는데 시간이 오래 걸려서 더 많은  
알고리즘을 못한게 아쉬웠다.

황인범



개인화를 통해 발전시키지 못한 점은  
아쉽지만 결과물을 만들어낸 뛰어난  
조원들과 함께할 수 있어 영광이었습니다.  
많은 점을 배워갑니다.

정한솔



프로젝트 진행하면서 약 5개월간  
배운 부분 외 생소한 부분도 많았지만  
팀원들 덕분에 잘 마무리할 수 있었습니당  
다들 고생했어유 ㅎㅎ

오혜수



멜 데이터를 사용 할 수 있었으면  
좋았을텐데 ... 용량이 너무 커서  
시도조차 못한게 아쉬웠다  
좋은 언니 오빠들 만나서 양 행복띠

박세진

**Q & A**



**감사합니다!**