

알고리즘 구축 계획서

과제명 : 노래 추천 알고리즘

2022. 05. 19

목 차

I. 개 요	
1. 문서의 목적	
2. 문제 정의 및 문서의 범위	
II. 문제 정의	
1. 골든 서클	
2. 데이터 명세	
III. 알고리즘 파이프라인	

I. 개 요

1. 문서의 목적

본 문서는 “노래 추천 알고리즘” 개발의 계획서이다.

본 계획서는 “노래 추천 알고리즘” 설계 및 구축의 기초가 되며, 알고리즘의 목적 및 파이프라인에 대하여 정의하고 서술한다.

본 계획서는 가능한 구체적이고 간결하게 서술되어야 하고, 추후 프로토타입을 통해 전체 알고리즘을 시연할 수 있어야 한다. 본 계획서를 사용하는 대상은 본 과제를 기획하는 기획자, 알고리즘을 개발하는 개발자 등이며, 본 과제의 설계 및 개발 과정에서 본 계획서를 사용할 수 있도록 한다.

2. 문제 정의 및 문서의 범위

본 계획서는 “노래 추천 알고리즘”의 문제 정의 및 데이터 명세, 파이프라인의 기술을 그 범위로 한다.

II. 문제 정의

1. 골든 서클

Why: 작곡, 음반 참여 등에 대한 진입 장벽이 낮아지면서 수많은 음악이 쏟아져 나오는 오늘날, 처음 보는 아티스트의 처음 보는 제목만 가지고는 어떤 음악이 나의 취향에 맞을지 판단하기 어렵다.

How: 노래를 직접 듣지 않고도, AI가 노래의 내용을 기반으로 나에게 맞는 노래를 추천해준다.

What: 내가 최근에 들었던 노래를 입력하면, 그 노래들을 기반으로 사용자가 좋아할만한 노래를 추천해주는 알고리즘

2. 데이터 명세

특성 번호	특성 설명	수집 방법	인코딩 형태
Feature 1	노래 제목	멜론 크롤링	명목 척도
Feature 2	노래 가사	네이버 뮤직 크롤링 “{ 가사”.format(노래제목) 으로 쿼리를 만든다.	Word2Vec
Feature 3	노래 장르	멜론 크롤링	명목 척도 (원-핫 인코딩)
Feature 4	아티스트	멜론 크롤링	유사도 기반 임베딩
Feature 5	작곡가	멜론 크롤링	유사도 기반 임베딩
Feature 6	작사가	멜론 크롤링	유사도 기반 임베딩
Feature 7	리메이크 여부	멜론 크롤링	명목 척도
Feature 8	원곡 아티스트	멜론 크롤링	리메이크를 안했다면 영행렬, 했다면 유사도 기반 임베딩
Feature 9	경연곡 여부	네이버 뮤직 크롤링	명목 척도
Feature 10	발매 날짜	멜론 크롤링	시계열 인코딩
Feature 11	곡의 길이	멜론 크롤링	비율 척도
Feature 12	게이름

III. 알고리즘 파이프라인

데이터 수집	주로 멜론 및 네이버 뮤직 크롤링
전처리	아티스트, 작곡가, 작사가 등 유사도 기반 임베딩
탐색적 분석	노래 제목이 비슷하면 가사도 비슷한지 확인
특성 공학	작사가와 가사 간의 유사성, 작곡가와 곡의 길이 간의 유사성 확인 필요
모델 선택	레이블이 없으므로 비지도학습 방식이 필요. 유사도 기반의 인코딩이 추가 되므로, 계층적 군집 분석이 가장 적절할 것으로 판단
모델 학습	한번만에 맞출 필요는 없으므로, 10개의 예상 레이블을 출력하도록 학습
모델 평가	실루엣 지수
최적화	향후 서비스를 만든 이후, 10개의 예상 레이블 중 고객이 원하는 노래가 있을 확률이 높아지도록 설계
적용 계획	노래 특성을 기반으로 추천시스템을 먼저 설계한 후, 상용화 이후에는 고객이 노래를 들었던 이력을 바탕으로 추천시스템을 개선