

# 삼성청년 SW·AI아카데미

관통프로젝트:  
맞춤형 큐레이션을 위한  
추천 시스템 구현

# 맞춤형 큐레이션을 위한 추천 시스템 구현

## 챕터의 포인트

- 관통 프로젝트 안내

## 관통 프로젝트 안내

## 목표

- 최종 목표는 사용자에게 맞춰 콘텐츠를 선별해주는 AI 기반의 큐레이션 시스템을 구축하는 것입니다. Django Rest Framework와 PostgreSQL을 활용하여 백엔드 환경을 구성하고, OpenAI API와 벡터 서치(Vector Search) 기법을 통해 사용자의 선호도에 기반한 맞춤형 추천 시스템을 구현합니다.
- 이를 통해 콘텐츠의 코사인 유사도 등을 활용한 개인화 추천 로직을 제공하고, 향후 다양한 추천 알고리즘과 상호작용 이벤트(좋아요, 조회수 등)를 반영하여 큐레이션 품질을 높이는 데 목적을 둡니다.
- 임베딩 및 벡터 서치 기법을 이해하고 적용할 수 있다.
- Django Rest Framework를 통해 RESTful API 서버를 구축하고 관리할 수 있다.
- PostgreSQL에서 pgvector를 이용한 코사인 유사도 검색을 구현할 수 있다.
- 사용자 행동 데이터를 수집·분석하여 추천 시스템에 반영할 수 있다.



## 준비사항

- 사용 데이터
  - 과정을 통해 수집되는 데이터
  - 데이터 콘텐츠에 대한 임베딩(embedding) 정보  
(Embedding 관련 API기반 서비스 등 또는 사전 학습된 오픈소스 모델 학습 활용)
  - 사용자 인터랙션(좋아요, 조회, 댓글 등) 데이터
- 개발언어/프로그램
  - Python & Django Rest Framework: 백엔드 API 서버 구현
  - PostgreSQL: pgvector 확장 기능을 사용해 벡터와 유사도 검색 구현
  - OpenAI API: 내용 임베딩을 위한 모델 활용
  - pgvector: 코사인 유사도(CosineDistance) 기반의 벡터 검색 가능

## 구현 방법

### 1) 벡터 임베딩 생성

- OpenAI Embeddings API 등의 임베딩 모델 기반의 처리를 활용하여 콘텐츠 벡터화(Embedding)
- Django 모델에 임베딩 정보를 저장할 필드를 추가하여 벡터 데이터 관리

## 구현 방법

### 2) 사용자 인터랙션 관리

- 사용자의 좋아요(Like), 조회수, 스크랩 등의 컴포넌트를 대상으로 한 상호작용을 별도로 저장
- 상호작용 종류에 따라 가중치를 다르게 부여하여 추후 분석에 활용



## 구현 방법

### 3) 추천 알고리즘

- 사용자 별 임베딩 정보를 분석하고, 코사인 유사도 등 관련성을 평가하여 수치가 높은 기준으로 콘텐츠를 추천
- Django ORM + pgvector를 사용하여 효율적으로 거리 계산
- 이미 좋아요 등 사용자가 소비한 콘텐츠를 제외하고, 노출되는 콘텐츠는 무작위로 섞는 등의 다양성 확보 방안 고민

## 구현 방법

### 4) 연관 콘텐츠(관련 콘텐츠) 노출

- 사용자가 확인 중인 콘텐츠와 유사한 콘텐츠를 추출
  - 예시)
  - 사용자의 정보를 기반으로 해당 콘텐츠 위주로 제공하는 추천 기반의 콘텐츠들을 노출
  - 코사인 유사도 등 관련도가 높은 상위 n개(예: 5개)를 함께 노출

## 구현 방법

### 5) API 설계 및 로직 구현

- Django Rest Framework로 추천 콘텐츠 조회, 관련 콘텐츠 조회 등 API Endpoint 구성
- RESTful API 형태로 데이터를 제공하여 다양한 클라이언트(모바일, 웹 등)에서 활용 가능
- PostgreSQL + pgvector 환경 설정

## 관통 프로젝트 가이드

### PJT 내용

- 데이터 수집
  - 크롤링 등을 통해 텍스트 및 이미지 데이터 확보
- AI 임베딩 생성
  - OpenAI API 등의 임베딩 모델을 활용해 콘텐츠를 벡터화
- 데이터베이스 저장
  - 서비스를 위한 OLTP 형의 데이터베이스인 PostgreSQL에 데이터와 임베딩을 벡터 형태로 저장 (pgvector 사용)
  - 이외의 벡터 DB를 사용하는 것은 자유
- 추천 시스템
  - Django Rest Framework에서 사용자별로 추천하는 콘텐츠 제공 로직 구현
- 관련 콘텐츠 표시
  - 각 콘텐츠에 대해 유사도가 높은 콘텐츠 목록도 함께 출력

## 관통 프로젝트 가이드

**SSAFYNEWS**

검색어를 입력하세요

나만의 뉴스 큐레이팅

대시보드

로그아웃

🔍 AI 맞춤 추천 뉴스

당신이 원하는 뉴스, 이제 AI가 직접 추천해드립니다.  
나만의 취향을 기반으로, 맞춤형 뉴스만 쏙쏙 골라주는  
뉴스 큐레이팅 서비스 SSAFYNEWS에 빠져보세요.  
AI 맞춤형 기사에 대해 대화하며 궁금한 점을 물어보고,  
원하는 보기 쉬운 대시보드를 통해 나의 뉴스 소비 패턴도 확인할 수 있습니다.

데이터사이언티스트 관련 정보를 보려면 IT/과학 카테고리를 클릭하세요!

전체

연예

경제

교육

국제

산업

정치

지역

건강

문화

취미

스포츠

사건사고

사회일반

IT/과학

여성특집

여행레저

라이프스타일

testuser님을 위한 뉴스 피드

최신순

추천순

사생활만

이예슬 기자 brightpearl@kyunghyang.com · 2025.03.19.

**"현재 앞 식당 불매하겠다" 탄핵반대 시위대...집주들 "가게 지키기도 바쁜데" 당혹**

국무 세력을 중심으로 서울 종로구 헌법재판소 인근 식당에 대한 '불매 리스트'가 확산하면서 집주들이 당혹해하고 있다. 이들은 "일부 인근 집주들이 탄핵 반대 시위대에 자살을 했다"며 온라인상에서 불매를 요청하고 나섰다. 윤석열 대통령 탄핵심판으로 시위대가 인제로 몰려 매출이 줄어든 데 이어 국무 세력의 불매 운동까지 일면서 집주들은 두배로 고통을 받고 있다. 해당 집주들은 "먹을거리다"며 "상황이 빨리 끝나기만 바랄 뿐"이라고 밝혔다. 18일 온라인 커뮤니티와 사회관계망서비스(SNS) 등에는 '탄핵에 ...

0 0 0

#국무 세력 #불매 리스트 #탄핵 심판 #집주 고통 #헌법재판소

IT\_과학

배문규 기자 sobbe@kyunghyang.com · 2025.03.20.

**지난해 개인정보 유출 사고 307건...원인 1위는 해킹**

지난해 개인정보 유출 원인도 해킹이 절반 이상을 차지했다. 공공기관 유출 신고 건수는 전년도에 비해 두 배 이상 늘어난 것으로 집계됐다. 개인정보보호위원회는 20일 지난해 개인정보 유출 사고를 분석한 '2024년 개인정보 유출 신고 통량 및 예방 방법' 보고

**SSAFYNEWS**

검색어를 입력하세요

나만의 뉴스 큐레이팅

대시보드

로그아웃

←

기사

**"우연히 듣고 빠졌다"던 SNS, 광고였다...카카오엔터 기만광고 제재**

김세훈 기자 ksh3712@kyunghyang.com · 2025.03.24.

유명 사회관계망서비스(SNS) 채널을 운영하며 일반 소비자인 것처럼 자사 음원을 광고한 카카오엔터테인먼트가 공정거래위원회로부터 과징금 제재를 받게 됐다. 공정위는 카카오엔터테인먼트의 표시광고법 위반행위에 대해 시장조치 및 과징금 3억9000만원을 부과했다고 24일 밝혔다. 자사가 기획·유동하는 음원 음반을 SNS와 온라인 커뮤니티에 기만적으로 광고한 혐의다. 카카오엔터는 음반·음반 유통시장 1위 사업자다. 공정위 조사에 따르면 카카오엔터는 유명 SNS 채널을 인수하거나, SNS 채널을 개설해 홍보글을 올리면서 해당 SNS 채널이 자사 소유·운영 채널임을 알리지 않았다. 카카오엔터는 해당 채널에 "우연히 듣고 빠졌다던 아이티스트" "오늘 내 알고리즘에 온 노래" "들자마자 반한 노래" 등의 문구를 사용해 마치 일반인이 추천하는 것처럼 했다. 카카오엔터는 이런 방식으로 2016년 10월부터 지난해 2월까지 아이돌연구소(메이스트), 노래는듣고다녀-!(메이스트-인스타그램 등) 등 15개 채널에서 총 2353개의 홍보 게시물을 올렸다. 또 카카오엔터 소속 직원들은 2021년부터 2023년 12월까지 다수, 뽕뽕 등 주요 온라인 커뮤니티 11곳에 자신의 신분을 밝히지 않고 37개 홍보 게시글을 작성했다. 카카오엔터는 광고대행사 8억 6000만원을 주고 광고를 하면서도 광고에 경제적 이해관계를 밝히지 않았다고 했다. 이처럼 상업적 광고라는 점을 알리지 않은 광고는 427건에 달했다. 카카오엔터가 활용한 SNS 채널은 팔로워 수가 411만명, 온라인 커뮤니티 가입자 수도 최대 150만명 수준이었다. 공정위는 이런 문제·누락으로 소비자가 해당 광고를 기획된 광고물이 아닌 일반인에 의한 추천글로 받아들이기 가능성이 높다고 봤다. 공정위는 게시물 작성자가 일반 소비자인지 경제적 이해관계가 있는 광고주인지가 소비자 선택에 상당한 영향을 미치는 데도 이를 알리지 않은 것은 기만광고에 해당한다고 판단했다. 카카오엔터는 내부 법무 검토를 통해 경제적 이해관계를 표시하지 않은 것이 부당한 광고에 해당할 가능성이 높다는 것을 일고도 위반행위를 계속한 것으로 조사됐다. 공정위가 대중음악 분야에서 기만적 광고행위를 제재한 것은 이번이 처음이다. 공정위는 "대중음악과 같이 타인의 선호·추천이 영향을 많이 받는 분야에서는 SNS를 통한 홍보·시 사업자가 경제적 이해관계를 명확히 밝히야 한다"면서 "문화산업 분야의 부당한 표시·광고들 지속적으로 감시할 것"이라고 했다.

카카오엔터테인먼트 · 공정거래위원회 · 과징금 · 기만광고 · SNS 채널

1 1 0

좋아요

관련 기사

**"내돈내산" 아니었어? 알고보니 원상...SNS '뒷광고' 2만건 적발**

김세훈 기자 ksh3712@kyunghyang.com · 2025.03.16.

0 0

**카카오 김범수, 경영 일선서 물러난다... "건강상 이유"**

최민지 기자 ming@kyunghyang.com · 2025.03.13.

0 0

**검찰, '핵심 부패'가 의혹 카카오엔터에 또 압수수색**

박재민 기자 applaudi@kyunghyang.com · 2025.03.20.

0 0

**"국단적일수록 돈 본다"...도 넘는 국무 유희버 제지할 길이 없다**

배문규 기자 sleumb@kyunghyang.com · 2025.03.12.

0 0

**성 실종파환 '역설방송' BJ-사이버 레거 유희버, 세무조사받는 다**

김종나영 기자 nayoung@kyunghyang.com · 2025.03.06.

0 0

## 요구사항 1

- 기본 기능
  - 유저 행동 기반으로 추천 시스템을 적용해서 Django Rest Framework 백엔드에 반영
- 데이터 분석 및 추천
  - 콘텐츠 임베딩 생성 및 데이터베이스 저장
  - 사용자 상호작용 정보를 활용한 추천 로직 구현
  - pgvector의 Cosine Similarity를 이용한 유사도 계산 (이외의 고도화된 방법을 활용할 수도 있음)
- 요청 조건
  - OpenAI API(또는 유사 AI 서비스)를 통해 임베딩 생성
  - Django Rest Framework로 API 구현
  - 벡터 검색 및 유사도 랭킹 활용



## 요구사항

- 결과
  - 사용자 선호(좋아요, 조회 등)에 따라 추천된 콘텐츠 리스트
  - 해당 콘텐츠 목록을 반환하는 백엔드 엔드포인트(API)

## 요구사항 2

추천된 콘텐츠와 유사도가 높은 관련 콘텐츠 목록 제공

- 요청 조건
  - 메인 추천 콘텐츠와 함께 상위 5개 유사 콘텐츠 조회 등 다양한 방식을 활용
  - 콘텐츠 간 유사도 계산 시 코사인 거리(Cosine Distance) 등 콘텐츠 유사도 판단 기준 활용
- 결과
  - 연관 콘텐츠를 JSON 형태(혹은 지정 형식)로 함께 전달
  - 사용자에게 콘텐츠 간 맥락을 쉽게 파악할 수 있도록 도움

## 추가기능

### 추가 피쳐 엔지니어링 및 고급 알고리즘 적용

- 요청 조건
  - 기존 임베딩 외의 카테고리, 출처, 시간대 등을 고려한 멀티모달 추천 모델 적용 가능
  - 사용자 개인화(사용자 그룹화, 협업 필터링, 시계열 분석 등) 기법 적용 가능
- 결과
  - 향상된 추천 정확도 및 사용자별 차별화된 콘텐츠 큐레이션
  - 추천 성능 비교(기본 vs 고도화) 분석과 보고서 작성

## 산출물

- Gitlab에 올라온 de-project 코드 기반의 레포지토리 지속적 커밋
- 이후 PJT도 해당 레포지토리에 이어 나가면서 개발

## 정리

구현 기능	체크 포인트
데이터 임베딩 및 DB 저장	벡터화(임베딩) 정확도 pgvector 설정 여부
Django Rest API 구현	추천 리스트 반환 및 API 정상 동작
관련 콘텐츠(연관 콘텐츠) 관련 추천 구현	유사 콘텐츠 목록, pgvector 등 vector 활용 코사인 유사도 등 유사도 기준 활용
추가 피쳐 엔지니어링 및 고급 알고리즘	콘텐츠 카테고리, 시간대, 사용자 프로필 등 추가 고려

# 내일 방송에서 만나요!

삼성청년SW·AI아카데미