

개인화된 OTT 추천

딥러닝 기반 OTT 플랫폼 선택 시스템

사용자 맞춤형 콘텐츠 추천 시스템 구현 보고서

1조 : 김명재, 박찬, 유승환, 정의철

목차

프로젝트 개요

3-5



- 선택의 피로 문제 정의
- 목적 및 기획 의도
- 핵심 차별점

시스템 구조

6-8



- 전체 시스템 아키텍처
- 기술 스택 구성
- 데이터 모델링 및 관리

핵심 알고리즘

9-10



- 추천 알고리즘 워크플로우
- 로직 파이프라인

주요 파일 구성

11-12



- 핵심 파일 역할 및 기능
- 시스템 구조 요약
- 모듈 간 상호작용

기대효과 및 향후 계획

13-15



- 현재 한계점
- 향후 개선 방향
- 비즈니스 가치 및 결론

프로젝트 배경: 현대인의 콘텐츠 선택 고민



선택의 피로 (Paradox of Choice)

현대 OTT 및 영화 콘텐츠의 폭발적 증가로 인해 사용자는

🕒 원하는 콘텐츠를 찾는 데 많은 시간과 노력을 소모

🗣️ 결국 만족스럽지 못한 선택으로 이어지는 경우가 많음



기존 추천 시스템의 한계

현재 대부분의 추천 시스템이 가진 문제점:



콜드 스타트 문제

신규 사용자가 가입하면 이전에 본 콘텐츠가 없기 때문에 추천이 어렵다는 문제.



시청 이력 기반의 단순 분석

단순한 시청 기록만으로 사용자의 취향 패턴 파악에 한계



플랫폼 독립성 부족

OTT 플랫폼 단위로 추천되므로, 사용자가 모든 플랫폼을 넘나드는 통합 추천을 받기 어려움

💡 결과적으로 사용자들은 콘텐츠 선택에 더 많은 시간을 소비하면서도 낮은 만족도를 경험하게 됨

프로젝트 목적 및 기획 의도

◎ 핵심 목적

🧩 사용자의 영화 취향과 시청 성향을 **정밀하게 분석**하여
개인에게 최적화된 영화와 OTT 플랫폼 추천

🕒 콘텐츠 **탐색 시간 단축**으로 사용자의 미디어 소비
경험의 질 극대화

👥 '선택의 피로' 문제 해결로 사용자 만족도 향상

💡 기획 의도



딥러닝 기반 정교한 추천

심층 신경망으로 복잡한 사용자-영화
상호작용 학습



인구통계학적 데이터 활용

연령, 성별 등 인구통계 정보를 통한 추천
정확도 향상



과학적 검증

A/B 테스트를 통한 추천
시스템 효과 측정



OTT 플랫폼 정보 결합

추천 영화의 시청 가능 플랫폼 정보 제공

“ 데이터 기반의 정교한 개인화 추천을 통해 콘텐츠 탐색 시간을 단축하고, 미디어 소비 경험의 질을 극대화합니다. ”

핵심 차별점



딥러닝 기반 정교한 평점 예측

- ✓ 심층 신경망을 활용해 사용자-영화 간 복잡한 상호작용 학습
- ✓ 미묘한 취향 패턴까지 포착하는 정밀한 분석
- ✓ TensorFlow/Keras 기반 협업 필터링으로 평점 예측 정확도 향상

기존 추천 시스템 대비 예측 정확도 향상



인구통계학적 데이터 활용

- ✓ 연령, 성별 등 인구통계 정보 반영한 정교한 추천
- ✓ A/B 테스트를 통한 인구통계 정보의 영향력 과학적 검증
- ✓ 유사 사용자 그룹 탐색 시 인구통계적 특성 고려



OTT 플랫폼 가용성 정보 결합

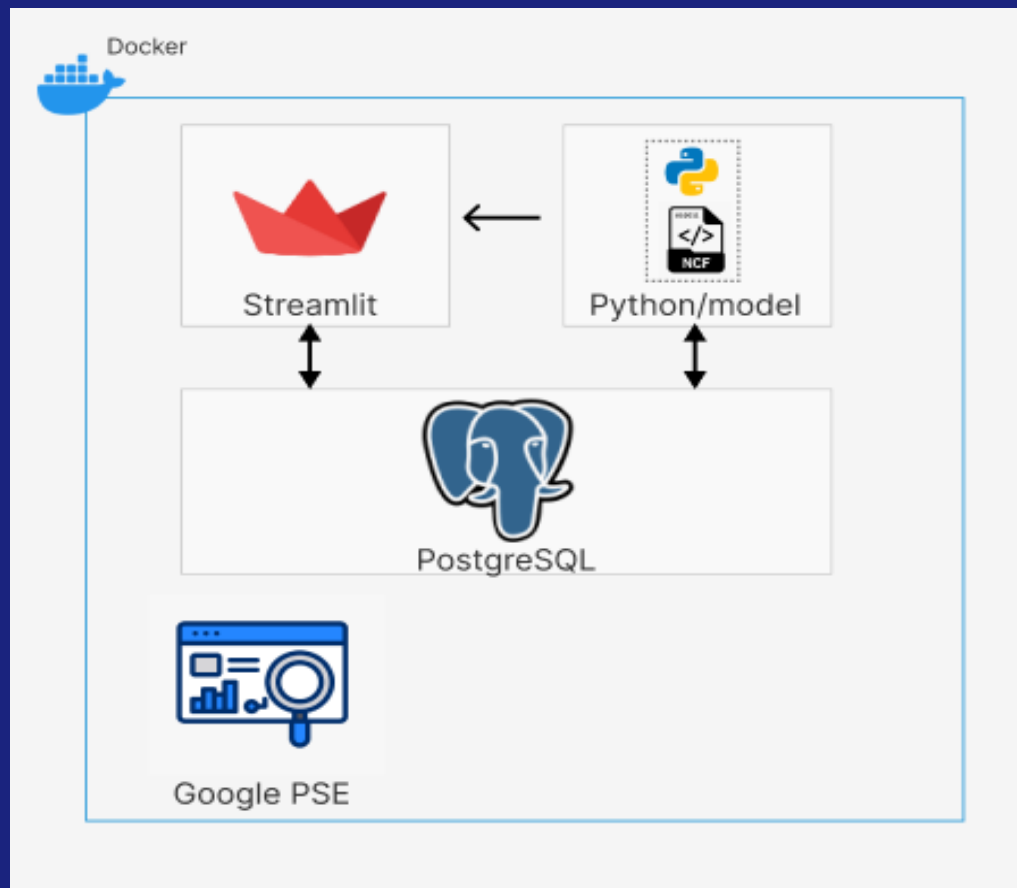
- ✓ 추천 영화가 시청 가능한 OTT 플랫폼 정보 제공
- ✓ 사용자 구독 현황과 연계한 최적 플랫폼 추천
- ✓ 콘텐츠 접근성 향상으로 사용자 만족도 증대

추천 영화의 실제 시청 가능성 극대화

기존 추천 시스템의 한계를 극복하는 혁신적 접근

전체 시스템 아키텍처

데이터 수집부터 추천 결과 제공까지의 전체 시스템 흐름



기술 스택 구성



프론트엔드/UI

 Streamlit

- 설문 및 대시보드 개발
- 사용자 인터페이스 구현
- 결과 시각화 및 표현



데이터베이스

 PostgreSQL

- 사용자/영화/평점 데이터 저장
- 관계형 데이터 구조 관리
- JSONB 타입으로 유연한 데이터 처리
- PGvector Extension 사용을 위해 MYSQL 사용 지양



추천 모델

TensorFlow/Keras

- 딥러닝 기반 평점 예측
- 사용자-영화 임베딩 학습
- 다층 퍼셉트론(MLP) 구현



데이터 처리

 Pandas, NumPy

- 데이터 전처리 및 분석
- 행렬 연산 및 통계 처리
- 데이터셋 변환 및 정제



인프라

 Docker

- 개발/테스트/배포 환경 일관성 유지
- 컨테이너화된 서비스 관리
- 시스템 확장성 제공



데이터 수집

 Google Cloud Platform (GCP)

- 콘텐츠 정보에 대해 CSE(Custom Search Engine) 이용
- 각 콘텐츠가 어느 플랫폼에서 방영되는지 확인

데이터 모델링 및 관리

데이터 소스



MovieLens 데이터셋

- 대규모 사용자-영화 평점 데이터
- 영화 메타데이터 정보
- 사용자 인구통계 정보



신규 사용자 설문 데이터

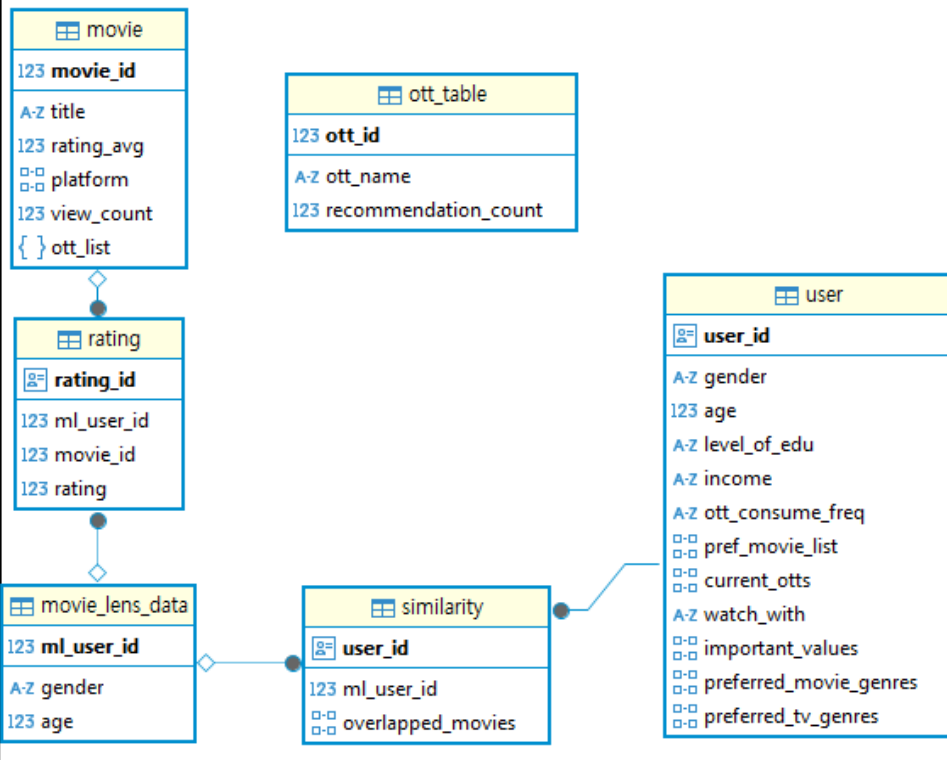
- Streamlit 설문을 통한 1차 데이터 수집
- 나이, 성별, 선호 영화 등 정보
- 최소 10편의 선호 영화 입력 필요



공공 마이데이터 API

- OTT 이용 현황 데이터
- 인구통계 기반 외부 데이터
- 향후 확장성을 위한 데이터 소스

데이터베이스 스키마



데이터베이스 정보

PostgreSQL 기반 관계형 데이터베이스 사용
사용자/영화/평점 데이터의 효율적 저장 및 관리

핵심 추천 알고리즘 워크플로우



사용자 설문 데이터 수집

Streamlit 설문 인터페이스를
통해 인구통계 및 선호 영화
최소 10편 입력



유사 사용자 그룹 탐색

MovieLens 데이터셋 내
Jaccard 유사도 기반으로
상위 N명 유사 사용자 그룹 식별



딥러닝 평점 예측

유사 사용자 그룹의 평점 패턴
패턴을 활용, 미시청 영화에
대한 예측 평점 산출



영화 추천 리스트 생성

예측 평점 상위 Top-K 영화
선정, 장르 다양성 고려



OTT 플랫폼 추천

추천 영화의 OTT 제공 현황
및 사용자의 구독 현황을
반영해 최적 플랫폼 추천

핵심 로직 요약

- **콜드 스타트 해결**: 최소 10편의 선호 영화 입력으로 초기 추천 정확도 확보
- **Jaccard 유사도**: 두 사용자 선호 영화 집합의 교집합/합집합 비율로 유사성 측정
- **딥러닝 모델**: 사용자/영화 ID 임베딩 → 다층 퍼셉트론(MLP) → 평점 예측(1~5점)
- **OTT 매핑**: 사용자 구독 정보와 영화 제공 플랫폼 정보 결합으로 최적 시청 경로 제안

로직 파이프라인



주요 파일 구성 및 역할

● UI/사용자 인터페이스 ● 데이터 처리 ● 모델 학습/예측 ● 데이터베이스 작업

🏠 models.py

sqlAlchemy사용해서 DB 내용을 자유롭게 사용, 연동

📁 01_Survey.py

사용자 설문 페이지

사용자 인구통계 및 선호 영화 데이터 수집 인터페이스

👤 user_matching.py

유사 사용자 탐색

Jaccard 유사도 기반 유사 사용자 그룹 식별

📄 load_data.py

데이터 전처리 및 DB 업로드

원시 데이터 정제 및 데이터베이스 적재

✅ check_data.py

users.csv 중복 검사

사용자 데이터 중복 제거 및 정합성 검증

🎬 mysqlmovie.py

NCF 기반 딥러닝 모델 학습

신경망 협업 필터링 모델 학습 및 파라미터 최적화

🔧 modeltest.py

모델 테스트 및 예측값 확인

학습된 모델 성능 평가 및 추론 결과 검증

🗄 database.py

DB 연결

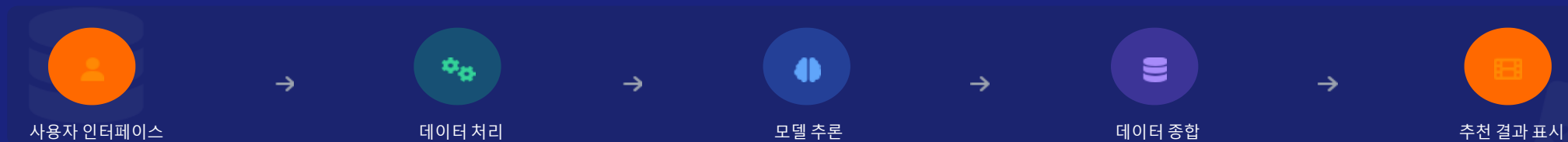
PostgreSQL 데이터베이스 연결 및 세션 관리

🔍 ott_serach.py

웹 api 이용

CSE(custom search engine)로 영화 검색

파일 간 데이터 흐름



시스템 구조 요약

모듈화된 시스템 구성

각 기능별 파일로 분리되어 독립적 개발 및 유지보수가 용이한 구조



데이터 파이프라인 특징

🟡 모듈화: 각 기능별 파일로 분리, 독립적 개발 및 유지보수 용이

🟡 데이터 파이프라인: 설문 → DB 저장 → 유사도 매칭 → 딥러닝 예측 → 추천 결과

현재 한계점



콜드 스타트 문제

신규 사용자가 선호 영화 10편 미만 입력 시 추천 정확도 저하 현상

❗ 충분한 초기 데이터 확보의 어려움으로 인한 개인화 추천 품질 저하



데이터 희소성

마이너 영화 및 비활성 사용자에게 대한 추천 성능 한계 발생

❗ 인기도가 낮은 영화나 활동이 적은 사용자에게 대한 정확한 예측 어려움



데이터의 구식화

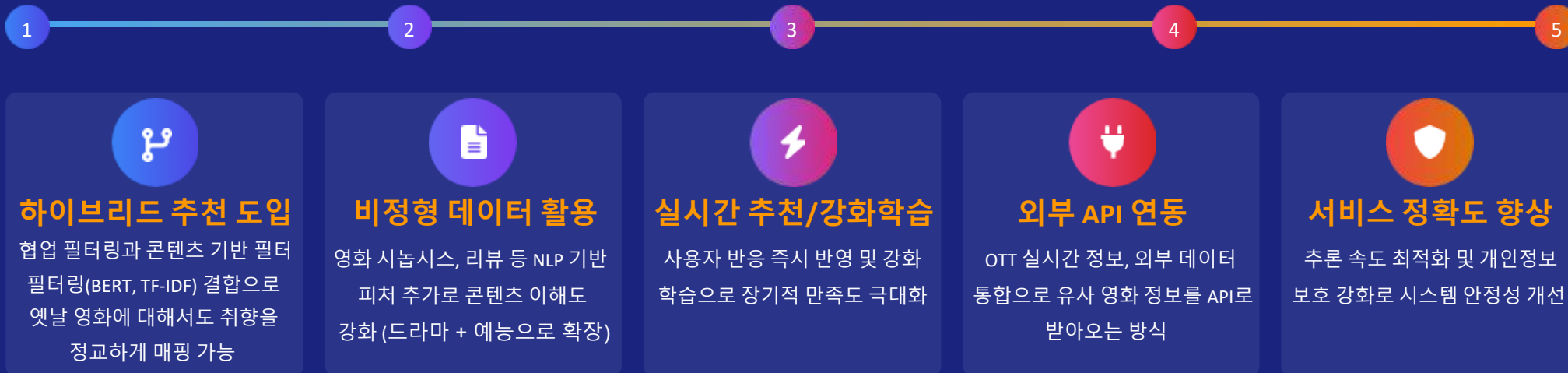
최신 사용자 시청 기록 데이터 부재로 최신 사용자 시청 기록 리스트 질문에 대해 유사도 매칭이 어려움

❗ 데이터만 있으면 해결되고 시스템적으로는 실시간 처리가 될 수 있도록 구축함

이러한 한계점들은 향후 개선 계획을 통해 단계적으로 해결해 나갈 예정

향후 개선 방향

시스템 발전을 위한 로드맵



기대 효과

👉 사용자간 유사도, 정확도 향상

👉 사용자 만족도 20% 이상 증가

👉 OTT 구독 전환율 15% 개선

기대 효과 및 비즈니스 가치



사용자 경험 혁신

- ✓ 콘텐츠 탐색 시간 단축으로 사용자 피로감 해소
- ✓ 개인 취향에 최적화된 만족도 높은 추천 제공
- ✓ 데이터 기반 투명한 추천으로 신뢰도 향상
- ✓ OTT 플랫폼 선택 효율성 증대

◎ 미디어 소비 경험의 질적 극대화



비즈니스 측면 가치

- ✓ 검증된 PoC 모델 확보로 상용화 기반 마련
- ✓ OTT 구독/구매 전환율 15% 이상 증대 예상
- ✓ 콘텐츠 제작사와 OTT 플랫폼 간 시너지 창출
- ✓ 데이터 축적을 통한 지속적 비즈니스 모델 개선

◎ 사용자와 비즈니스 모두에 실질적 가치 제공



기술적 우위 확보

- ✓ 딥러닝 기반 복합적 취향 패턴 학습 역량
- ✓ A/B 테스트 등 과학적 검증 방법론 적용
- ✓ 데이터 기반 혁신으로 지속적 모델 개선 가능
- ✓ 확장 가능한 모듈식 아키텍처 구축

◎ 최신 기술과 실증적 검증의 균형적 결합

💡 본 프로젝트는 **사용자 중심 혁신**과 **비즈니스 가치 창출**을 균형있게 달성하며, **기술적 우수성**을 바탕으로 지속가능한 발전 가능성을 제시합니다.

결론 및 제안

프로젝트 요약

본 프로젝트는 데이터 기반의 정교한 개인화 추천, 딥러닝 및 인구통계 정보 융합, OTT 플랫폼 정보 결합 등 최신 기술과 실증적 검증을 결합한 혁신적 추천 시스템의 프로토타입을 제시합니다.

사용자 가치

- 콘텐츠 탐색 시간 단축
- 만족도 높은 맞춤형 추천
- 미디어 소비 경험의 질 향상

비즈니스 가치

- OTT 구독/구매 전환율 증대
- PoC 모델 확보로 상용화 가능성
- 데이터 기반 혁신으로 시장 우위 확보

기술적 실현 가능성

- 체계적 아키텍처 구축 완료
- 과학적 실험 통한 유효성 검증
- 명확한 개선 로드맵 수립

최종 제안

본 프로젝트는 사용자와 비즈니스 모두에게 **실질적 가치**를 제공할 수 있는 혁신적인 추천 시스템으로, 지속적인 개발과 투자를 통해 **시장 경쟁력**을 갖춘 서비스로 발전할 수 있습니다.

Q & A



감사합니다.

