

프로젝트 기획서

OTT서비스 데이터 분석 14팀 일사천리

타이틀

1. 프로젝트 주제

- “콘텐츠들의 특징을 기반으로 본인에게 알맞은 OTT서비스를 추천”

2. 엔드유저에게 보이는 웹서비스 타이틀

- Which OTT

3. 팀 구성 및 역할

- 김선교 : 팀장, 데이터분석 및 추천 모델 개발
- 박나은 : 데이터분석 및 문서 작성/정리
- 최성호: Backend
- 양덕규 : Frontend

데이터세트

1. 사용할 데이터셋

- 미디어 통계포털 미디어 이용행태 조사(OTT서비스)
(<https://stat.kisdi.re.kr/main.html>)
- 넷플릭스 데이터
(<https://www.kaggle.com/akashguna/netflix-prize-shows-information>)
- 디즈니 데이터
(<https://www.kaggle.com/unanimad/disney-plus-show>)
- 키노라이츠 크롤링 데이터

2. 사용할 데이터 세트를 통해 얻고자 하는 인사이트

- 각 플랫폼별 콘텐츠 특징 및 장단점
- OTT서비스 이용도 및 관심도

3. 분석을 위해 데이터를 준비하는데 예상되는 어려움

- 크롤링 시 알 수 없는 오류로 데이터 수집이 늦어질 수 있음
- 데이터에 부족 및 탐색 오류로 편향되거나 잘못된 분석을 하는 것
- 플랫폼별 콘텐츠 특징을 찾을 수 없거나 그 차이가 크지 않은 경우

기획 방법

1. 문제 정의와 가설 설정 방법

- 1) 문제 정의: 코로나 이후 ott 서비스 이용량이 급증함과 동시에 여러 서비스들이 런칭함에 따라 사용자들이 자신에게 맞는 서비스를 선택하는데 어려움을 겪거나 다수의 서비스를 구독하는 것에 대해 부담을 느끼고 있다.
- 2) 가설 설정:
 - ① 각 OTT서비스가 보유한 콘텐츠의 장르/국가/관람가 등의 데이터와 플랫폼의 오리지널/독점 콘텐츠 보유 비중을 시각화하면 각 서비스의 콘텐츠 특성을 파악할 수 있다.
 - ② 분석으로 발견한 특성을 바탕으로 콘텐츠 기반 추천(CBF) 모델 적용 시 사용자가 원하는 콘텐츠를 많이 보유한 서비스를 추천받을 수 있을 것이다.
 - ③ 사용자의 OTT 서비스 이용 시간/빈도 데이터를 바탕으로 사용자가 어느 정도로 OTT를 사용하고 있는지 파악할 수 있도록 돕고 이를 바탕으로 OTT 서비스 추천에 대해 홍보한다면, 서비스에 대한 흥미를 높일 수 있다.

2. 기대 효과

- 타겟층의 트래픽 유도를 통해 광고 수익 창출
- OTT 서비스 별 강점 및 단점을 분석하는 서비스를 제공하여 솔루션 제공 가능

4. 엘리스에서 배운 데이터분석 기술, 알고리즘 및 라이브러리 사용 계획

0.분석 툴: Colab

1. 시각화 :

- 코랩: Plotly
- 웹: Recharts

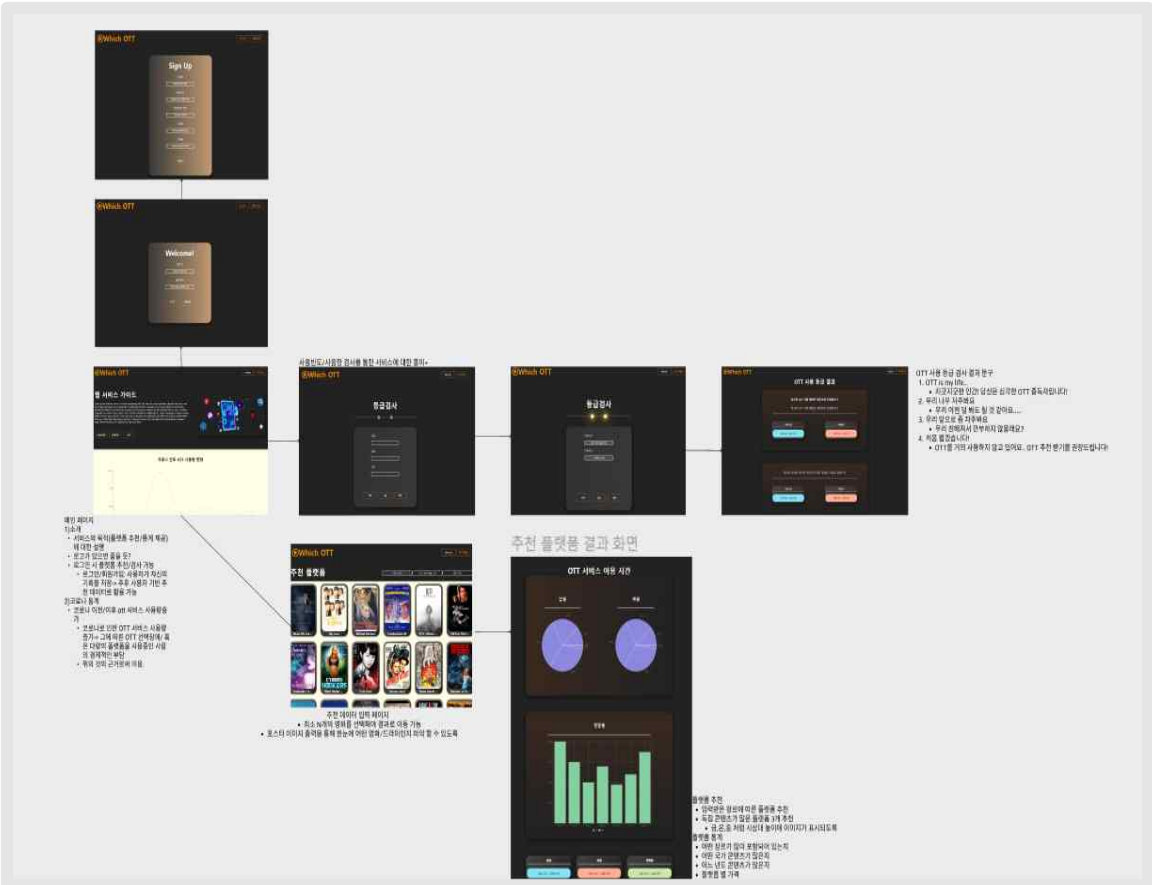
2.데이터 처리: Numpy/pandas

3.back-end: Python, Flask, MariaDB

4.front-end: HTML5, CSS3, JavaScript, React

5. 데이터 분석 종류 (시각화, 통계 대시보드)

- 추천 OTT서비스 표시: 시각화
- 독점/오리지널 콘텐츠 비중이 높은 서비스: 시각화
- 플랫폼별로 장르 포함 비중: 통계 대시보드
- 콘텐츠 제작 국가별 비중: 통계 대시보드
- 콘텐츠 제작 시기별 비중: 통계 대시보드
- 성별 OTT 사용시간/사용빈도: 통계 대시보드
- 연령별 OTT 사용시간/사용빈도: 통계 대시보드



2. 시나리오

1-1. 메인 페이지에서 추천 버튼을 클릭한다.

- 로그인을 하지 않았을 경우 경고창을 띄운다
- 로그인한 상태일 경우 사용자의 선호 콘텐츠를 받는 페이지로 넘어간다

1-1-1. 사용자의 선호 콘텐츠를 받는다

- 포스터 이미지 출력을 통해 어떤 드라마/ 영화인지 한눈에 파악하도록 돕는다.
- 선택된 콘텐츠는 연보라로 하이라이트 효과
- 10개를 선택하지 않았을 경우 결과 페이지로 넘어가지 않도록 한다.
- 콘텐츠 10개를 선택할 경우 결과 페이지로 넘어간다

1-1-2 추천 결과와 OTT서비스 분석 결과를 출력한다

- 독점 콘텐츠가 많은 콘텐츠 플랫폼 3개를 보여준다(금/은/동처럼 시상대 형식)
- 취향에 맞는 콘텐츠가 많은 플랫폼 3개를 띄운다(금/은/동처럼 시상대 형식)
- 서비스별 콘텐츠 분석 자료를 출력한다.
 - 플랫폼별로 어떤 장르가 많이 포함되어 있는지
 - 어떤 국가 콘텐츠가 많은지
 - 어느 년도 콘텐츠가 많은지
 - 플랫폼별 가격

1-2 검사 버튼을 클릭한다

- 로그인을 하지 않았을 경우 경고창을 띄운다
- 로그인한 상태일 경우 성별/사용시간/빈도 등을 받는 페이지로 넘어간다

1-2-1 성별/사용시간/빈도 등을 받는다

- 단계별로 진행하여 다음 페이지 이동시 하이라이트 효과

1-2-2 사용 등급 검사 결과를 출력한다

- 개인의 사용 등급과 문구를 나타낸다
- 성별/나이 정보와 이용 시간/빈도 데이터를 사용해 평균을 통계로 나타낸다.

1-3 메인페이지에서 로그인 버튼 클릭 시 로그인 창으로 이동

- 없을 경우 경고 메시지를 출력한다

1-3-2 회원가입 버튼 클릭 시 회원가입을 진행한다

- 비밀번호의 경우 영어/특수문자/숫자 포함
- id의 경우 영어로 작성