

특성값, 기준값 확정

📅 날짜

2022년 12월 2일

☰ 미팅멤버

구현모 **최은진** 장동근 이종은

☰ 비고

홍행도(이분, 다중), 관중수 예측

© 댓글 추가

▼ 영화 흥행도, 관객수 예측

1. 기간 :

- a. 2015-2019 년 르네상스
- b. 2020-2021 영화 암흑기
- c. 2022년 부활기

2. 전처리

- a. 대표국적과 국적 중 대표국적을 선택해서 사용한다.
- b. 기간매출액, 매출액 점유율은 분석에서 제외
- c. 연말이랑 년초에 겹치는 기간에 대해서는 최근 누적 관객수로 업데이트 한다
- d. 개봉일 기준:
 - i. 2020년 이후 데이터는 사용하지 않는다.
 - ii. 재개봉 데이터는 사용하지 않는다.
- 1. 기간
- e. 배급사 기준: 4대 메이저 배급사 / 그 외
- f. 장르 기준(장르가 여러 개인 경우 어떻게 나눌 것인지)
- g. 누적관객수 5,000명 이상 또는 스크린수 5개 이상 데이터 사용(성인물 제외됨)
- h. 누적관객수가

3. 예측 시나리오

[흥행도 분류]

- 흥행기준
 - 제작비가 있는 경우 제작비 대비 매출이 나오는 경우 흥행
 - 최근 10년 박스 오피스 200 위 내 관람객수가 평균 이상인 경우 흥행으로 판단
- 수익에 대한 부분은 제작비가 나오지 않기 때문에 검토 제외, 관객수로 계산한다.
- (예시) 아바타

- 장르: SF
- 감독 : 제이슨카메론
- 배우: 주연배우
- 배급사(매출) : 주식회사 해리슨앤컴퍼니, 이십세기폭스코리아(주)
- 제작사(X)
- 러닝타임
- 등급기준
 - 2진 분류
 - 미흥행(0), 흥행(1)
 - 다항 분류

관람객수	흥행 등급
100만 미만	미흥행
100만 - 500만	흥행
500만 - 700만	중박
700만 - 1000만	대박
1000만 이상	초대박

[관객수 예측]

- 감독, 배우, 장르, 상영시간

[왓차피디아 지수]

- <https://pedia.watcha.com/>
- 영화제목, 상영년도, 장르, 상영국가, 평점, 평점참여수, 코멘트

[네이버 지수]

- 리스트

4. 배급사 - CJ, 롯데시네마 쇼박스, NEW

5. 역할배분

공통: 분류, 회귀모델 학습

구현모: 감독 등급, 배우등급

장동근: 왓차피디아, 분류모델/회귀모델

최은진: 관객수에 영향을 줄 수 있는 요인 (스토리라인에 사용할 것)

이종은: SNS 지수 / 지수들을 나누는 기준에 대한 근거 찾기 + 다른 사람들이 어떻게 지수화했는지 찾아보기

▼ 프로젝트 계획안

프로젝트명: 무비퍼스(무비 예측 문어-옥토퍼스)

주제:

영화 정보에 따른 영화 흥행도, 관객수 예측 분석

선정 배경:

코로나 19로 인해 영화 산업이 타격을 받게 되었고, 점차 거리두기가 해제되면서 영화에 대한 수요가 증가하고 있다. 이를 통해 앞으로 영화 산업의 미래가 어떻게 변화될지 확인해보고자 한다.

진행 방향:

영화진흥위원회 관람객수, 왓차피디아 평점, 배우, 감독 등 영화에 미치는 특징(요인)을 찾아 흥행하는 영화, 관객수를 예측하는 모델을 만들고자 합니다.

의의:

분류 머신러닝, 회귀분석 학습을 통해서 영화 흥행이 미치는 요인과 관객수를 찾아보고자 한다.

담당역할:

장동근: 웹크롤링, API, 모델링

최은진, 이종은, 구현모 : 영화진흥위원회 데이터 전처리, 모델링

