

머신러닝 프로젝트 : 영화 관람객 수 예측

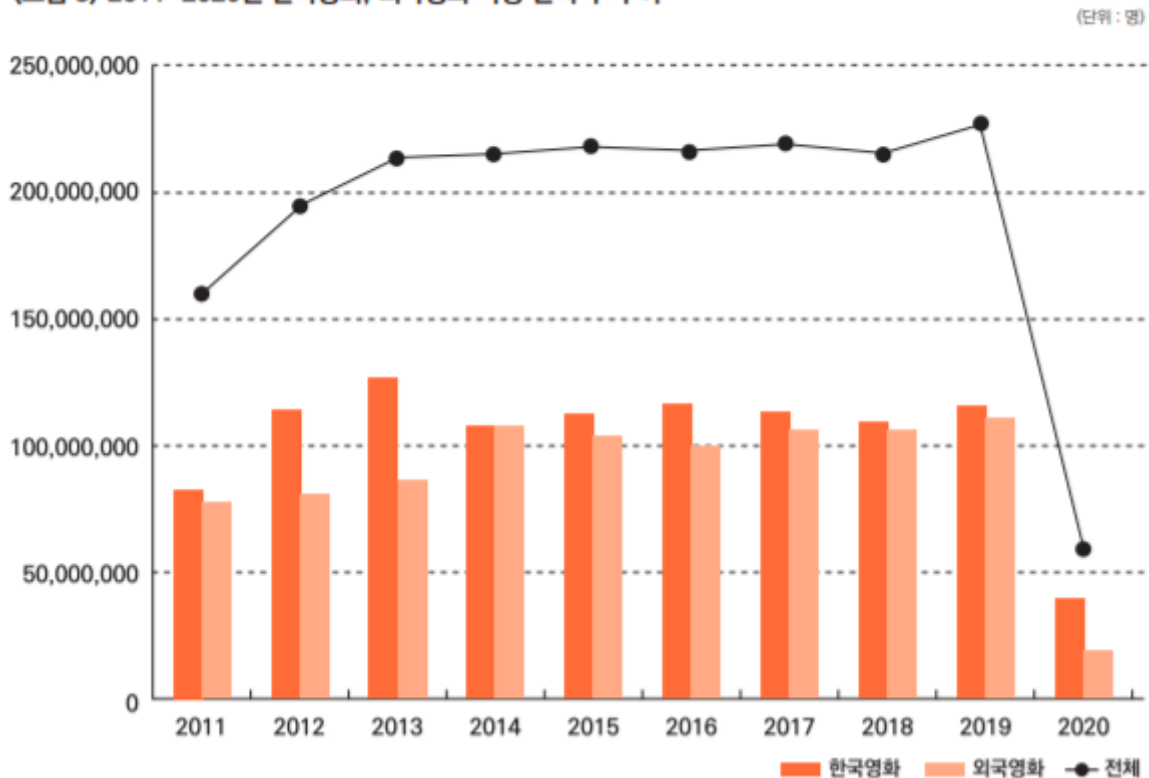
1. 주제

- 영화의 흥행에 영향을 주는 변수 데이터를 수집하여 머신러닝 학습을 통해 관람객 수 예측 모델 구현(~ 2019년 데이터 활용)
- 2020년의 영화들이 코로나19의 영향이 없었을 때 관객의 수 예측치를 확인

2. 배경

- 영화 산업은 지속적인 성장을 거듭하고 있었다.
- 2019년 말 코로나19가 유행하며 2020년 관객수는 70% 가량 감소하였다.
- 코로나19 이전의 데이터로 관객수에 영향을 미치는 요소를 분석하여 코로나19의 영향을 2020년 영화 관객수 예측으로 파악해보려한다.

〈그림 3〉 2011-2020년 한국영화, 외국영화 극장 관객 수 추이



[영화진흥위원회]

3. 데이터 소개

■ 데이터 : KOBIS 오픈데이터, 네이버 영화 API, 구글 트렌드

- KOBIS 오픈 데이터 : KOBIS 공식통계(연도 별) 데이터
: 영화 전반적인 정보(영화명, 감독, 스크린수, 장르, 매출액, 관객수 등)
- 네이버 영화 API : 네이버 API 이용한 데이터 추출
: 영화의 평점 정보, 배우 정보 등

4. 머신러닝을 통해 얻고자 하는 결과

- 영화의 여러 변수 데이터를 이용해 관람객 수를 예측하는 모델의 구현하여 전반적인 데이터 분석에 익숙해지는 것을 목표로 한다.
- 영화의 흥행에 영향을 미치는 요인을 분석한다.
- 2020년 영화 실제 관객수와 예측치로 코로나19의 영향을 확인한다.

5. 타임테이블

날짜	일정
11/25 – 26 (목 – 금)	주제 선정 및 데이터 수집
11/27 - 11/28 (토 - 일)	1차 모델 구현 및 결과 토론
11/29 – 12/1 (월 - 수)	모델 개선 및 분석
12/1 (수)	머신러닝 과제, 분석
12/2 (목)	팀1 : 결과물 분석 + 시사점 파악 팀2 : 발표자료 제작

6. 팀원 각자의 목표한 역할

- 스토리별 역할 분배
- 구글 드라이브 : <https://drive.google.com/drive/folders/1KfbVX6P0QhPFNE1A--cmuceCTLGKelgU?usp=sharing>
- 김남규
- 노연우
- 이은지
- 이태기

7. 참고 자료

- <https://www.koreascience.or.kr/article/CFKO202015463051319.pdf>
- http://www.bigdata-map.kr/datastory/new/story_24
- <http://koreascience.or.kr/article/JAKO201834663385693.pdf>
- <https://rpubs.com/leesohee/689993>
- <https://smecsm.tistory.com/69>