

영화의 흥행 요인 분석 및 개봉 예정 영화의 관객수 예측

- 훈련 과정명 : 파이썬과 R을 활용한 빅데이터 분석(머신러닝,딥러닝)전문가 양성과정
- 훈련 기간 : 2019-01-11 ~ 2019-07-12
- 팀명 : NA
- 팀장 : 정호진 | 팀원 : 김지수 박건우 이형욱

목차

- ✓ 주제 및 목적
- ✓ 개발 환경 (개발 리소스)
- ✓ 요구 사항 분석
- ✓ 팀원 별 업무 분담
- ✓ 프로젝트 일정 계획

주제

영화의 흥행 요인 분석 및 개봉 예정 영화의 관객수 예측

목적

최근 봉준호 감독이 영화 <기생충>으로 2019년 칸 영화제 황금종려상을 수상함에 따라 한국 영화계에 한 획을 그었다는 평을 받고 있다. 한국 영화 투자에 대한 긍정적 기류가 흐르는 가운데 영화 흥행에 어떤 요인이 영향을 미치는지 분석하고, 이를 기반으로 개봉 예정작의 관객수를 예측하여 투자 위험을 줄이고자 한다. 또한, 더욱 다양한 영화가 투자 받을 기회를 늘려 영화 산업의 성장과 발전의 기반을 마련한다.

개발 환경 (개발 리소스)

OS

MicroSoft Windows 10

Python

Python 3.5.2

DB

Oracle

R

R 3.6.0

기타 Tool

Edit Plus

요구 사항 분석

1

영화 관객수와 관련된 변수(장르, 제작사, 국가, 인물 등)를 선정한다.

2

오픈API(영화진흥위원회, 네이버 영화 등), BeautifulSoup, selenium 등을 이용하여 영화 관객수와 관련된 데이터를 수집한다.

3

OracleDB를 이용하여 데이터베이스를 설계하고,
수집한 데이터를 저장한다.

4

R을 이용하여 영화 데이터의 통계분석(선형 회귀 분석, 상관관계 분석 등)을 시행하고 그래프를 통해서 시각화(산점도, 상자 수염 그래프 등)한다.

5

Tensorflow를 이용하여 기계학습(선형 회귀를 통한 지도학습)을 수행하고 관객수를 예측한다.

팀원 별 역할 분담

* 추후 변동될 수 있음

팀장 : 정호진

■ 담당

- 팀장

■ 수행 역할

- 역할 분담 및 의견 수렴
- 공용 저장소(Github) 구축
- 문서 작업

팀원 : 김지수

■ 담당

- 팀원

■ 수행 역할

- 데이터 수집(SNS 데이터)
- 데이터 전처리(결측치 및 이상치 제거)
- 데이터 마이닝
- 예측 모델 생성

팀원 : 박건우

■ 담당

- 팀원

■ 수행 역할

- 데이터 수집(정형 데이터)
- 데이터 전처리(코딩변경)
- 데이터 마이닝
- 예측 모델 생성

팀원 : 이형욱

■ 담당

- 팀원

■ 수행 역할

- 데이터 수집(비정형 데이터)
- 데이터 전처리(파생변수 생성)
- 데이터 마이닝
- 예측 모델 생성

프로젝트 일정 계획

