

---

# 개봉 예정 영화의 관객 수 예측 모델 사업 계획서

---

팀 식스 (6조)

2017.07.30.



# 목차

1. 개요 및 프로젝트 필요성 .....	3
1-1 분석 개요	
1-2 배경 및 동향	
1-3 제안의 효과	
2. 목표 .....	6
2-1 목표	
2-2 목표 달성의 평가지표	
3. 분석 및 개발방법 .....	8
3-1 추진일정	
3-2 인력의 주요 이력	
3-3 수행분야별 업무 분장	

# 개요 및 프로젝트 필요성

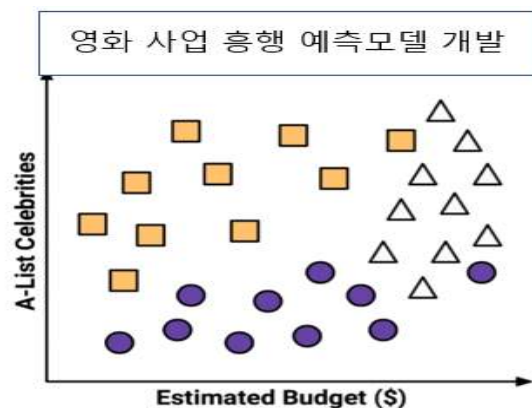
## 1-1 분석개요

1) 명칭 : 개봉 예정 영화의 관객 수 예측 모델 프로젝트

2) 기대 효과 :

- 통계적 이론을 통한 접근, 다양한 변수간 상관관계 분석 등 정밀한 통계 분석을 토대로 높은 수준의 예측 지표를 획득할 수 있음
- 특정 조건에 따른 영화 관객수를 예측함으로써 영화관의 수익 개선, 소비자의 행동 예측에 도움이 될 수 있음
- 8월에 있는 실제 빅데이터 관련 공모전 접수에 모델 출품

3) 예상 결과물 :



4) 과정 :

- 데이터 전처리 및 변수 선택 (관련 논문 참조, 영화사이트 크롤링)
- 기초 통계 분석 및 모델링
- 최적의 모델 도출
- 공모전 제시 데이터 실제 대입 및 검증 작업
- 결과 도출 및 시각화

## 1-2 배경 및 동향

### 1) 배경

- 한국 문화산업의 큰 부분을 맡고 있는 한국 영화 산업의 최근 2017년도의 수익률은 15.9%로 지속적인 성장세를 보이고 있음. 하지만 개별적인 영화 산업의 수익률은 흥행에 성공, 실패가 쉽게 예측되지 않으며, 이는 시장 매출 극대화에 크나큰 손실로 나타나고 있음
- 영화 산업은 다양한 기업이 연관되어져 있으며 또한 기존에 방대한 산업 자료를 가진 산업임. 이는 고객의 수요를 미리 예측하고 그에 따른 기업 전략을 활용할 수 있어야함을 보여줌
- 따라서 영화 산업에 있어서 기업에서 실제 보유하고 있는 데이터를 자유롭게 활용하여 새로운 비즈니스 모델 및 참신한 아이디어를 적용할 수 있는 모델 개발이 요구되어 짐.

### 2) 동향

- [시행 2018. 7. 1.] [법률 제15513호] 근로기준법의 시행이후 한국 문화산업의 수요가 급증하고 있다. 해당 산업의 큰 비중 차지하는 한국 영화산업의 경우 그 시장 규모가 16억 달러로 세계6위가 되는 상황이며 점점 그 규모는 증가하는 추세이다.
- 추가로 세계적인 흐름의 AI, 빅데이터 기술의 도입은 다양한 분야에 접목되어지고 있음. 정부, 민간 기업 등 다양한 산업체가 여러 데이터를 공공 데이터화에 공개하며 여러 부분에서의 활동 및 평가가 진행되고 있음.

### 3) 선행연구

- [영화 흥행 결정 요인과 흥행 성과 예측 연구, 김연형 외, 11년 저] 본 연구는 선행연구로는 영화의 흥행 요소에 관한 분석, 다양한 통계적인 값을 사용해 영화의 흥행을 예측한 과거의 연구가 존재
- [영화 흥행 실적 예측을 위한 빅데이터, 전희국 외, 14년 저] 허나 해당 연구들은 과거의 자료들만 사용한 연구이거나, 최근 영화 흥행에 있어서 주요 변수로 지적되는 소셜 미디어가 주요 변수로 포함되어 있지 않는 한계를 지님 따라서 특정 변수 뿐만 아닌 다양한 변수를 추가적으로 사용하였고 최근 날짜인 ,2018 8월 개봉 예정 작들에 관한 관객수를 예측해보는 모델은 이전에 없던 새로운 모델이 될 것이다.

### 1-3 제안의 효과

- 본 연구의 장점으로서는 빅데이터 기술을 활용하여, 기업 데이터를 활용, 다양하고 폭넓은 여러 변수들을 고려해 여러 모델을 만들고 각각에 대한 통계적인 분석 기법이 가해졌다는 것. 또한 과거의 연구들에 여러 수정 사항들이 가미됨.
- data web crawling, 회귀분석, p-value, 엑셀 데이터 전처리 기법 등을 이용하여, 실무적이며 논증적인 방법의 수치적 해석 (통계분석 i.e. 다중회귀분석, 로지스틱 회귀, 퍼셉트론)과 같은 분석들을 통해 각 변수들 간 각각의 다양한 관계성 설명할 수 있다는 장점이 존재.
- 본 연구는 추가적으로 감독명 배우에 국한되어있는 기존 연구에서 벗어나, 권위적인 평론가의 평가, 일반 대중들의 다양한 평가 지수 등을 고안, 활용함으로써 기존까지 진행되었던 연구를 획기적으로 발전, 개선함.

## 목표

### 2-1 목표

- 본 분석의 최종 목표는 다양한 빅데이터 분석방법을 활용하여 영화 흥행도(관객수) 예측 모델을 만들고 2018년 8월 개봉예정작 중 하나를 임의로 선정하여, 흥행도를 실제로 예측해보고 실제 해당 영화의 흥행결과와 비교 분석하는 것에 의의를 둠.
- 본 분석은 사업적 이익을 추구로 하지 않고, 기존에 주류를 이뤘던 감독-배우 및 스크린 수에 국한되어 있던 영화 흥행 모델에 다양한 변수를 추가함으로써 독창적 모델을 만들어 최종적으로 다양한 분야에서의 연구자들이 빅데이터 전문가로서의 활약을 펼칠 수 있게 도모하는 것에 그 목표를 둠.

### 2-2 목표 달성의 평가지표

[데이터 분석을 활용한 한국 영화 흥행 예측, 2018.6. 김세윤]

- 예측 모델 분석 결과 다양한 모델 중 배우, 배급사, 장르를 포함하는 모델이 회귀계수의 설명력이 높게 나타났다.
- 향후 연구에서는 TV드라마나 예능 등에서 인지도를 쌓고 영화에 출연하는 배우에 대한 영향력 분석도 필요하다. 또한 SNS상에서의 배우, 감독, 배급사 등의 평판 및 영향력 분석도 필요하다.
- 위 연구 논문에서 나타나있듯, 기존 분석은 배우, 배급사, 장르를 포함하는 모델이 주를 이룸. 또한 최근 사회에 강력한 영향력을 행사하고 있는 SNS에 대한 변수에 대한 필요성 언급.

□ 이에 따라, 본 분석에서 활용한 변수

종속변수	총 관람객 수
독립변수	상영 스크린 수
	개봉전, 개봉 1~3주 평점
	개봉전 미디어 노출 횟수
	개봉 1~3주차 관객 수
	제작비
	국적
	배우 및 감독
	배급사
	제작사
	등급
	장르
	온라인 버즈
	관객수 증감분

## 분석 및 개발방법

### 3-1 추진 일정

구분	세 부 개 발 내용	개발 기간																	비 고
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
데 이 터 전처리	영 화 데 이 터 수 집																		
	평가 데 이 터 크 롤링																		
	다 양 한 변수 데 이 터 수 집																		
통 계 분석	변수 별 모델링																		
	회 귀 분 석																		
	통 계 량 계산																		
	의사결정 분 류 기 제작																		
머 신 러 닝, 데 이 터 분석	다 양 한 모델 학 습																		
	실제 데 이 터 대 입 및 검 증																		
	최 적 의 모델 선 정																		
결 과 물 시 각 화 및 발표																			

### 3-2 인력의 주요 이력

성명	전공	소속
김운식	정치외교학과	단국대학교
김익환	소프트웨어학과	경희대학교
박창영	응용수학과	강남대학교
송준영	응용통계학과	단국대학교
이종범	컴퓨터공학과	가천대학교
장재원	경영학과	아주대학교



### 3-2 수행분야별 업무분장

수행분야	성명	역할	분석비중(5)
데이터 전처리	박창영	분석 데이터 전처리	30%
	송준영		
	이종범		
	장재원		
통계분석	김운식	모델 구성	25%
	김익환		
	박창영		
머신러닝 데이터 분석	송준영	최적 모델 선택	25%
	이종범		
	장재원		
결과시각화 및 발표	김운식	모델 예측 시각화	20%
	김익환		
	박창영		