

548

FASTCAMPUS  
DATA SCIENCE SCHOOL  
PROJECT

# 나만을 위한 영화 별점 예측 시스템

취향통계분석

보고싶어요 4

봤어요 544

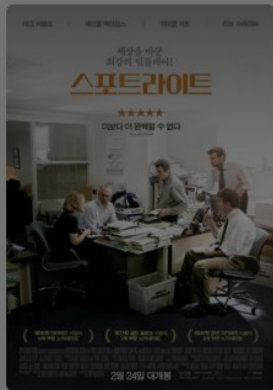
코멘트 0

팔로잉 2

팔로워 6

공유하기

장르 | 국가 | 담은 순



★★★★☆



★★★★☆



★★★★☆



★★★★☆



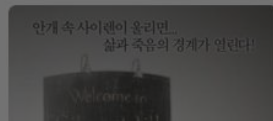
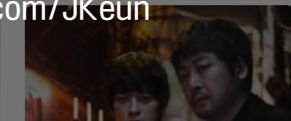
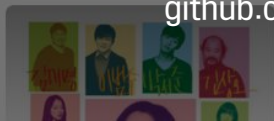
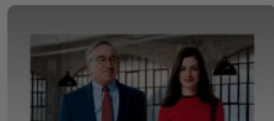
★★★★☆



★★★★☆

2016. 4. 15

PARK JAE KEUN  
github.com/JKeun



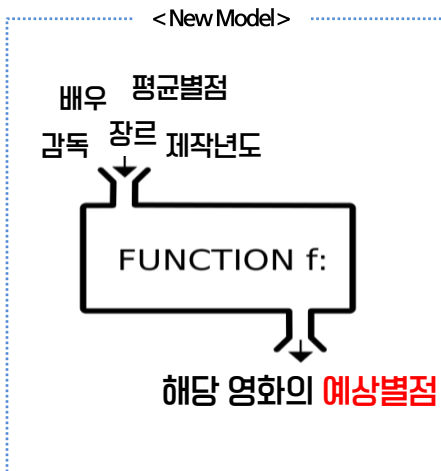
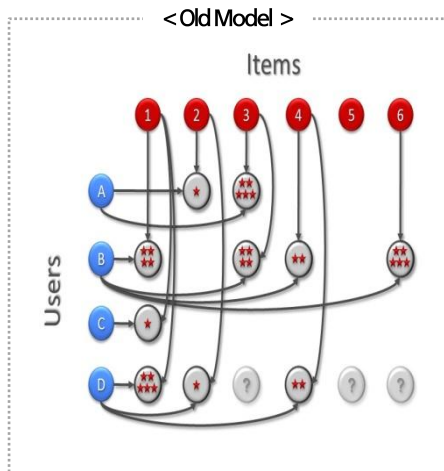
## Why ?

왜 시작하게 되었는가?

- 왓차의 예상 별점이 생각보다 나에게 잘 맞지 않는다?
- 영화를 보기 전 찾아보는 정보들(감독, 배우, 평균별점 등)은 곧 내가 영화를 볼 때 중요시하는 요인들

프로젝트의 목표는 무엇인가?

- 나만을 위한 개인화된 모델
- 기존의 유사도 Matrix를 통한 모델이 아닌, 위의 요인들을 바탕으로 한 예측 모델





## How ?

어떤 방법으로 사용하였는가?

- Supervised learning: Naïve Bayesian Classifier

$$\hat{y} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k).$$

- Data Frame

|        | 평균별점   | 좋아요수  | ... | 감독    | 배우       | ... |
|--------|--------|-------|-----|-------|----------|-----|
| 영화 1   | 4.2268 | 9691  | ... | 봉준호   | 송강호, 박해일 | ... |
| 영화 2   | 2.9962 | 3166  | ... | 김광식   | 김강우, 정진영 | ... |
| 영화 3   | 3.9011 | 33565 | ... | 데이빗핀처 | 브래드피트    | ... |
| :      | :      | :     | :   | :     | :        | :   |
| 영화 544 | 4.1277 | 51829 | ... | 박찬욱   | 최민식, 유지태 | ... |

GaussianNB

MultinomialNB

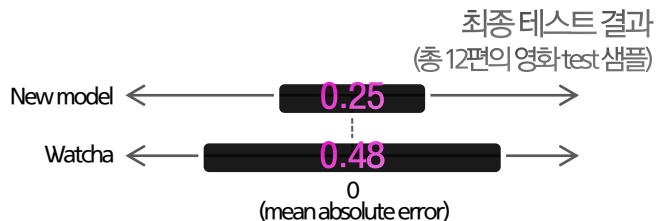
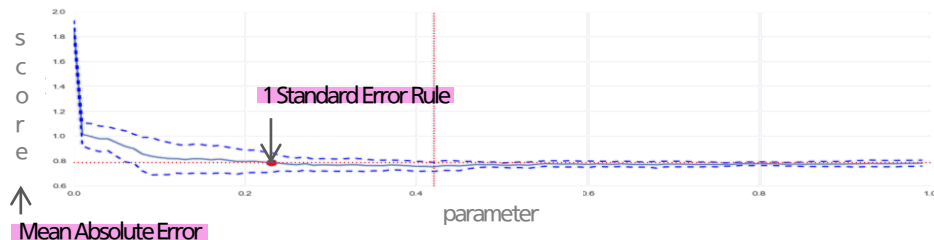
New model with conditional probability



## What ?

프로젝트를 통해 배운 것은 무엇인가? (무엇을 중점적으로 다루었나?)

- Parameters Tuning & Improving Performance

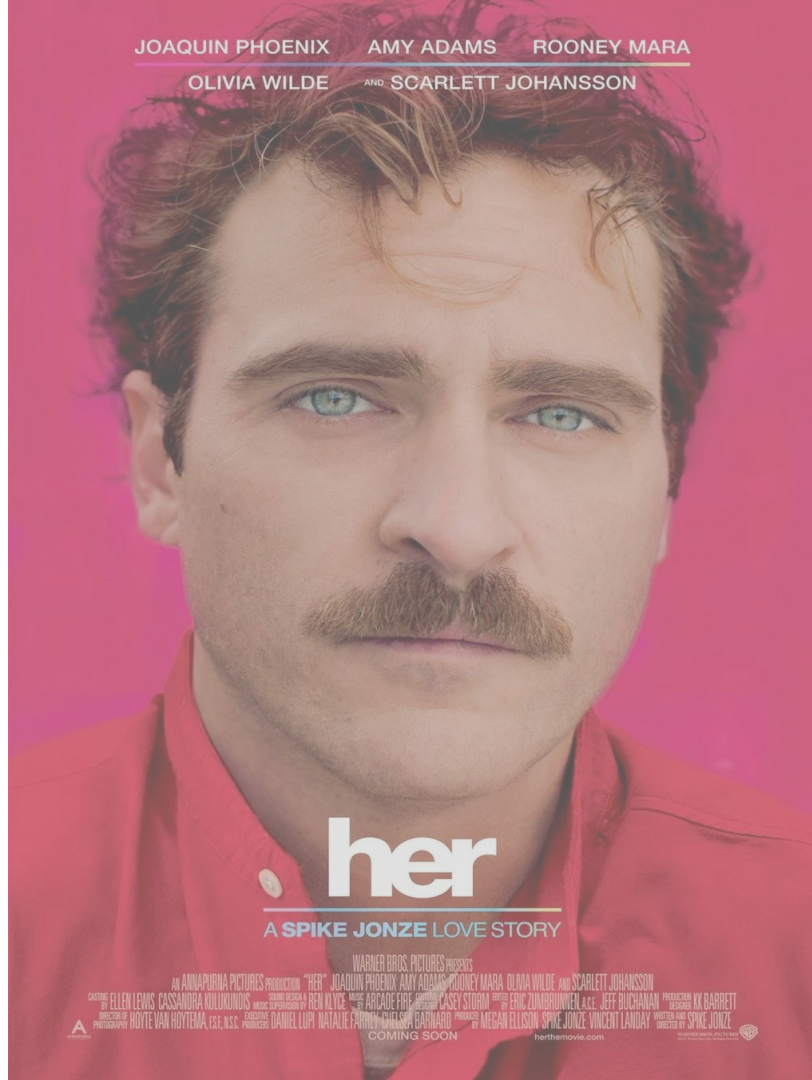


한계점은 무엇인가?

(어떤 Resource가 주어진다면 어떤 Performance를 이룰 수 있는가?)

- 다른 User의 별점 예측에 적용하기 어려운 부분 존재
- 다른 User의 영화 평가 정보  $\Rightarrow$  User마다의 개인화된 별점 예측모델  
 $\Rightarrow$  유사도 Matrix와의 결합 모델

※ 그러나 User별 각기 다른 모델을 만들어야하므로 Cost 발생이 큼







[ 감사 합니다 ]

2016. 4. 15  
PARK JAE KEUN  
[github.com/JKeun](https://github.com/JKeun)