

# 영화 흥행 예측 모델링

.Directed by.  
빛나구형

최형구  
김빛나  
서승표  
이제훈  
강윤영  
신우진

홍희지  
한정현  
신재욱  
백용수  
강유희

11

---





오늘 영화 뭐보지?

# NAVER

영화



검색

영화

영화순위

집에서 볼만한 영화추천

영화개봉작

영화 무료보기 사이트

상영중인 영화

영화추천

영화 다운로드 사이트

영화관

영화예매순위

영화의전당

개봉예정 영화

영화 아가씨

영화 사냥

영화 개봉예정작

도움말 | 신고

자동완성 끄기

영화총

상영작·예정작

· 현재 상영작  
· 개봉 예정작  
· TV/DVD 영화  
· 예고편

영화평점

매거진

예매

시사

평점

다운로드

인디극장

오늘도 새로운 재미! 네이버 영



현재상영영화

상영작 영화



조선마술사 The Magician, 2015

관람객 [?] ★★★★★ 6.91 | 기자·평론가 ★★★★★ 4.63 | 네티즌 [?] ★★★★★ 6.64 | 내 평점 ★★★★★ 등록>

드라마 | 한국 | 122분 | 2015.12.30 개봉 | [국내] 12세 관람가

감독 김대승 출연 유승호(한희), 고아라(한영) 더보기>

다운로드 원작 도서 1,779

주요정보 배우/제작진 포토 동영상 평점/리뷰 명대사/연관영화

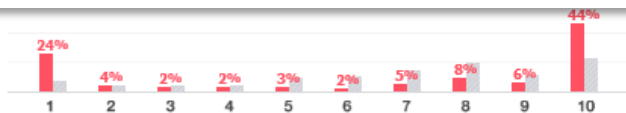
개봉 전 · 개봉 후 ·

네티즌 평점 [?] ★★★★★ 6.64 4,921명

관람객 평점 [?] ★★★★★ 6.64

점수 분포

남녀별 연령별



한줄분석이 영화를 가장 좋아하는 그룹은 10대 여자입니다.

그래프는 6분 주기로 업데이트 됩니다.

베스트

결말 얘기는 아무도 안하는건가요? ㅋㅋㅋㅋ 할 말을 잃었습니다. 마지막 무대는 무슨, 조선마술사가 마술은 안하고 연애나 하답니다.



유승호의 차기작 기대됩니다!

가능하여, 스포일러는 식별할 수 있습니다.

0/140

등록

베스트 평

유승호의 눈이 파란건 도데체 무엇을 의도한걸까 ㅋㅋㅋ



네이버 평점

✓도검순 ✓최신순 ✓평점 높은 순 ✓평점 낮은 순

□ 관람객 평점만 보기 [?]

★★★★★ 1 [베스트] 유승호의 눈이 파란건 도데체 무엇을 의도한걸까 ㅋㅋㅋ

바둑왕(kunu\*\*\*\*) 2015.12.30 20:34 | 신고

공감 435 비공감 110

★★★★★ 2 [베스트] 최근에 본 영화 중 최악. 내용은 전집도 재밌는 전집하듯이 폭풍 끊어지는 느낌이고 심지어 음악도 우렘. 개연성도 부족하고, 작가가 어디서 끌어온 대사화 설정만 가득 모아놓은 느낌. 손발 오그라드는 대사 때문에 미칠것 같았음. 배우들이 아까워서 2점..

yoona\*\*\*\* | 2016.01.01 01:02 | 신고

공감 172 비공감 43

★★★★★ 1 [베스트] 18년 살면서 이렇게 쫓겨 나가고 싶었던 영화는 처음이어서 신선했다. 관객들도 웃기 성는지 다들 속속장려에서 웃고 계시더라

MOVIE RANKING 2016.06.29 기준

영화인기검색어 [?]

- 1 연디컨텐츠...
- 2 비밀은 없다
- 3 아가씨
- 4 사냥
- 5 특별수사: 사...
- 6 정글북

영화인기검색어 [?]

- 1 홍상수
- 2 쿠니무라 준
- 3 김민희
- 4 나홍진
- 5 천우희
- 6 김태리

티켓예매순 [?]

- 1 사냥 27.46%
- 2 굿바이 싱글 16.86%
- 3 레전드 오브... 11.35%
- 4 연디컨텐츠 데... 6.33%
- 5 정글북 5.04%
- 6 아가씨 4.83%









영화 **공작** 4주년 기념 **3D 리메이크** by the same director www.dadema.com **COMING 2008.08** **최승환** **감독**

[illegible]

# 한국영화의 미래 예측하기





모 델 링

---

# Modeling

## Predicting the Future With Social Media

Sitaram Asur

Social Computing Lab

HP Labs

Palo Alto, California

Email: sitaram.asur@hp.com

Bernardo A. Huberman

Social Computing Lab

HP Labs

Palo Alto, California

Email: bernardo.huberman@hp.com

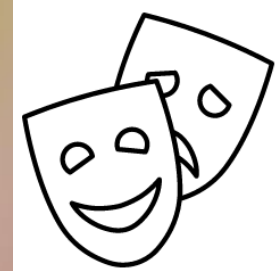
**Abstract**—In recent years, social media has become ubiquitous and important for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, we demonstrate how social media content can be used to predict real-world outcomes. In particular, we use the chatter from Twitter.com to forecast box-office revenues for movies. We show that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. We further demonstrate how sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media.

### I. INTRODUCTION

This paper reports on such a study. Specifically we consider the task of predicting box-office revenues for movies using the chatter from Twitter, one of the fastest growing social networks in the Internet. Twitter, a micro-blogging network, has experienced a burst of popularity in recent months leading to a huge user-base, consisting of several tens of millions of users who actively participate in the creation and propagation of content.

We have focused on movies in this study for two main reasons.

- The topic of movies is of considerable interest among the social media user community, characterized both by







자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



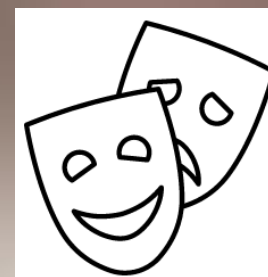
결론



2010 ~ 2015년 개봉영화



WEB PARSING



Actor



Director

 자료수집

 이론적 배경

 자료분석

 분석결과

 2016년 개봉영화 예측

 결론

# -Machine learning-

## -Time series estimation-



# 1. PARSING



자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



결론

	A	B	C	D	E	F	G	H	I
1	명량	김한민	최민식 류	오늘" 명량" 시사회로" 방금" 보고" 왔음." 단연코" 올해의" 수작임."					
2	명량	김한민	최민식 류	이순신	그는 신이다."				
3	명량	김한민	최민식 류	군도보다" 훨씬" 잦있음..극장가서" 꼭" 보세요"					
4	명량	김한민	최민식 류	29일" 시사회로" 봤어요냉정하게이건" 꼭" 봐야해..굿"					
5	명량	김한민	최민식 류	정말" 재미있네요" 잘보고" 왔습니다" 시사회"					
6	명량	김한민	최민식 류	오늘" 아침에" 조조로" 보고왔습니다." 8월4일에" 해군에" 입대하는					
7	명량	김한민	최민식 류	극찬할" 수 영상, 연출" 전부" 꽤나" 뛰어난" 작품이지" 않은가?" 평					
8	명량	김한민	최민식 류	국가에" 버림을" 받아도" 의리로" 지키려는" 충신." 피를" 흘리면서"					
9	명량	김한민	최민식 류	보는" 내내" 가슴찡하고" 울컥하는" 진짜" 최고의" 영화ππ한" 순간					
10	명량	김한민	최민식 류	작금의" 이시점에" 국가가" 혼란하고" 위태할수록국가를" 직간접으					
11	명량	김한민	최민식 류	평론가" 진짜" 싫억ㅋㅋ" πππ난" 진짜πππ" 감동적이었는데"					
12	명량	김한민	최민식 류	감동을" 넘 말 그대로" 명량" 그" 자체였다." 영화에" 정말" 공을" 많					
13	명량	김한민	최민식 류	""내가" 제 가장 미운 가장 흠모 가장 죽이 가장 차를" 함께" 하고"					
14	명량	김한민	최민식 류	전쟁씬에서" 입벌리고봄"					
15	명량	김한민	최민식 류	우리나라에서도이런해상전투씬이" 나오다니" 압권이다" 더불어" 0					
16	명량	김한민	최민식 류	이순신을" 5만원짜리에다가"		..넌라			
17	명량	김한민	최민식 류	평론가놈들" 정체가뭐냐" 명량이" 어떻게" 군도보다" 평점이낮냐"					
18	명량	김한민	최민식 류	안봤으면" 후회할" 영화!" 안볼수" 없었더" 영화!" 모두가" 봤으면"					
19	명량	김한민	최민식 류	올해" 최고의" 수작." 성웅" 이순신" 그는" 위대한" 조선의" 장수"					

```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))
```



```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))  
Out[78]: '0'
```

```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))
```

```
Out[78]: '0'
```

```
In [79]: classifier.classify(term_exists(tokenize('ㅋㅋㅋ스토리도 탄탄하고 연기자들 연기도 누구 하나 나쁘지 않았는데 왜 저렇게 별로라고 그러지 ㅋㅋ남들이 별로라 하니까 괜히 휩쓸려서 나쁘다 하는거 같은데 웃김ㅋㅋ이 정도면 평타는 쳤음')))
```



```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))
```

```
Out[78]: '0'
```

```
In [79]: classifier.classify(term_exists(tokenize('ㅋㅋㅋ스토리도 탄탄하고 연기자들 연기도 누구 하나 나쁘지 않았는데 왜 저렇게 별로라고 그러지 ㅋㅋ남들이 별로라 하니까 괜히 휩쓸려서 나쁘다 하는거 같은데 웃김ㅋㅋ이 정도면 평타는 쳤음')))
```

```
Out[79]: '0'
```

```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))
```

```
Out[78]: '0'
```

```
In [79]: classifier.classify(term_exists(tokenize('ㅋㅋㅋ스토리도 탄탄하고 연기자들 연기도 누구 하나 나쁘지 않았는데 왜 저렇게 별로라고 그러지 ㅋㅋ남들이 별로라 하니까 괜히 휩쓸려서 나쁘다 하는거 같은데 웃김 ㅋㅋ이 정도면 평타는 쳤음')))
```

```
Out[79]: '0'
```

```
In [80]: classifier.classify(term_exists(tokenize('아 진짜 대박재밌음 레알 도둑들 짱짱맨')))
```



```
In [78]: classifier.classify(term_exists(tokenize('와 이런것도 영화라고 만들다니 차라리 이거 만들 돈으로 아프리카 후원하는게 훨 낫겠다')))
```

```
Out[78]: '0'
```

```
In [79]: classifier.classify(term_exists(tokenize('ㅋㅋㅋ스토리로 탄탄하고 연기자들 연기도 누구 하나 나쁘지 않았는데 왜 저렇게 별로라고 그러지ㅋㅋ남들이 별로라 하니까 괜히 휩쓸려서 나쁘다 하는거 같은데 웃김ㅋㅋ이 정도면 평타는 쳤음')))
```

```
Out[79]: '0'
```

```
In [80]: classifier.classify(term_exists(tokenize('아 진짜 대박재밌음 레알 도둑들 짱짱맨')))
```

```
Out[80]: '1'
```

## 2. 감성분석



자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



결론

C:\Users\Baek\Miniconda3\Scripts\ipython.exe

```
In [81]: person_score("sim.txt")
('1', '커다란 사건속에서 선원들이 변화가는 과정에서 그들의 내면을 좀 더 자세하고 디테일하게 묘사했다면 좀 더 그들의 심정을 이해
그점이 조금 아쉽지만 그래도 참신한 스토리가 맘에드는 영화. 배우들 연기도 괜찮았음')
('1', '초반에는 전개는 좋았는데 갈수록 설득이 안되는 부분들이 있어서 아쉬웠습니다. 배우분들의 연기는 만족스러웠습니다.')
('1', '개봉할땐 몰랐는데 다시보니 박변기님의 메소드 연기가 일품입니다')
('1', '배우들 연기가 압도적이네요! 스토리도 긴장감 넘치고 올 해 본 영화 중 가장 몰입해서 본 영화예요!! 꼭 한번 보시길~')
('0', '연기력을장난아니저...')
('1', '연출과 시나리오에 있어서 해무는 아쉬운 부분들이 분명히 존재하는 영화입니다. 하지만 극중 배우들의 신들린 연기가 그 모든 것
들의 연기력이 작품의 몰입도와 완성도에 얼마나 지대한 영향을 끼치는지를 보여준 영화.')
('1', '이 영화 정말 괜찮습니다. 확실히 가볍게 즐길만한 영화는 아니라서 호불호가 갈리겠지만 전체적으로 잘 만든 수작이라고 생각됩니
리에서 쉽게 지워지지 않네요. 영화 잘봤습니다')
('1', '몰입감편다는말이 뭔지 제대로 보여준다. 긴장감에 두시간동안 잠시도 눈을떨수없었고한시간처럼 짧게 느껴질만큼 집중시키는영화
간본성이 밀바닥치는 광경을 보게될것이다')
('1', '오랜만에 진지하고 무거운 영화를 봤습니다. 영화보고 나뉘는데도 아직도 배위에 있는 느낌이 들어요. 가슴이 멍멍하고 슬픔니다. 한
sim의 선호도 점수는 100점 만점에 80점입니다.
Out[81]: 0.8016032064128257

In [82]: person_score("sim.txt")
('0', '그렇게 하정우는 지겨웠다...')
('1', '김태리라는 신인이 정말 잘함')
('1', '미술감독이 상 받을만 하네.')
('1', '당시 남성이 추악한 면을 사정없이 보여주는게 참 좋았고 아름다운 배경관 달리 지저분한 내부 상황이 좋았습니다. 여성들의 정사
진실이 느껴졌습니다. 아쉬운건 정사신의 시간을 좀 줄이고 스토리라인에 투자 했다면...')
('0', '뭔가이상하면서도이상했다')
('1', '외로우신 아가씨 역할에 김민희는 너무나 적역. 영화를 보고난뒤 놀랍도록 애잔하고, 통쾌하고, 울컥하고... 미묘하고도 복합적인
씨!!!')
('0', '어둡고 변태적으로 만들어놓고 일반인들은 이해하기 어렵습니다 라고 말하는 박찬욱 영화는 이제 지겹다.')
('0', '보는내내 지루하진 않았는데 이상하게 재미없는기분')
('0', '이제 박찬욱 영화는 안 보는 걸로.. 그냥 올드보이가 좋았던 것일 뿐. 박찬욱이 만드는 모든 영화가 내 스타일인 건 아닌가 봄.
영화관을 찾았지만.. 이젠 그 집착을 놓아줄 때가 된 듯.')
('0', '모호성과 노출이 많으면 예술성과 화제성이 강해진다는 3류감독들이 쓰던 기법을 박 감독이 써서 매우 씁쓸함. 이야기가 매우 험악
체관계로 가까워졌단 생각이 듦. 저급한 것과 고급화하고 변태성을 예술로 승화하려함')
park의 선호도 점수는 100점 만점에 56점입니다.
Out[82]: 0.5667833916958479

In [83]: _
```



# 1-1. PARSING AGAIN



자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



결론

	A	B	C	D	E
1	감독	영화	평점	정보	평점 평
2	김한민	사냥	6.66	2015 한국	6.6825
3	김한민	명량: 회오	4.94	2015 한국	6.931429
4	김한민	명량	8.49	2014 한국	6.671667
5	김한민	충신	5.38	2012 한국	6.93
6	김한민	최종병기	7.97	2011 한국	6.67
7	김한민	핸드폰	7.26	2009 한국	6.473333
8	김한민	극락도 살	6.92	2007 한국	6.25
9	김한민	갈치괴담	7	2003 한국	5.5
10	김한민	그렇게 김	5.5	1999 한국	0
11	최동훈	암살	8.97	2015 한국	7.37
12	최동훈	거장, 한국	5.97	2013 한국	7.51
13	최동훈	도둑들	7.6	2012 한국	7.5
14	최동훈	벽	7.15	2010 한국	7.54375
15	최동훈	전우치	8.18	2009 한국	7.452857
16	최동훈	중천	5.61	2006 한국	7.76
17	최동훈	타짜	9.1	2006 한국	7.492
18	최동훈	소년, 천국	7.62	2005 한국	7.46
19	최동훈	그때 그사	7.15	2004 한국	7.563333
20	최동훈	범죄의 재	8.76	2004 한국	6.965
21	최동훈	바람난 가	6.86	2003 한국	7.07
22	최동훈	눈물	7.07	2000 한국	0
23	추창민	7년의 밤	9.16	2015 한국	8.12875
24	추창민	광해, 왕0	9.22	2012 한국	7.972857
25	추창민	그대를 사	9.33	2010 한국	7.746667

	A	B	C	D	E
1	감독	영화	평점	정보	평점 평
2	김한민	사냥	6.66	2015 한국	6.6825
3	김한민	명량: 회오	4.94	2015 한국	6.931429
4	김한민	명량	8.49	2014 한국	6.671667
5	김한민	충신	5.38	2012 한국	6.93
6	김한민	최종병기	7.97	2011 한국	6.67
7	김한민	핸드폰	7.26	2009 한국	6.473333
8	김한민	극락도 살	6.92	2007 한국	6.25
9	김한민	갈치괴담	7	2003 한국	5.5
10	김한민	그렇게 김	5.5	1999 한국	0
11	최동훈	암살	8.97	2015 한국	7.37
12	최동훈	거장, 한국	5.97	2013 한국	7.51
13	최동훈	도둑들	7.6	2012 한국	7.5
14	최동훈	벽	7.15	2010 한국	7.54375
15	최동훈	전우치	8.18	2009 한국	7.452857
16	최동훈	중천	5.61	2006 한국	7.76
17	최동훈	타짜	9.1	2006 한국	7.492
18	최동훈	소년, 천국	7.62	2005 한국	7.46
19	최동훈	그때 그사	7.15	2004 한국	7.563333
20	최동훈	범죄의 재	8.76	2004 한국	6.965
21	최동훈	바람난 가	6.86	2003 한국	7.07
22	최동훈	눈물	7.07	2000 한국	0
23	추창민	7년의 밤	9.16	2015 한국	8.12875
24	추창민	광해, 왕0	9.22	2012 한국	7.972857
25	추창민	그대를 사	9.33	2010 한국	7.746667

### 3. 시계열 분석

자료수집

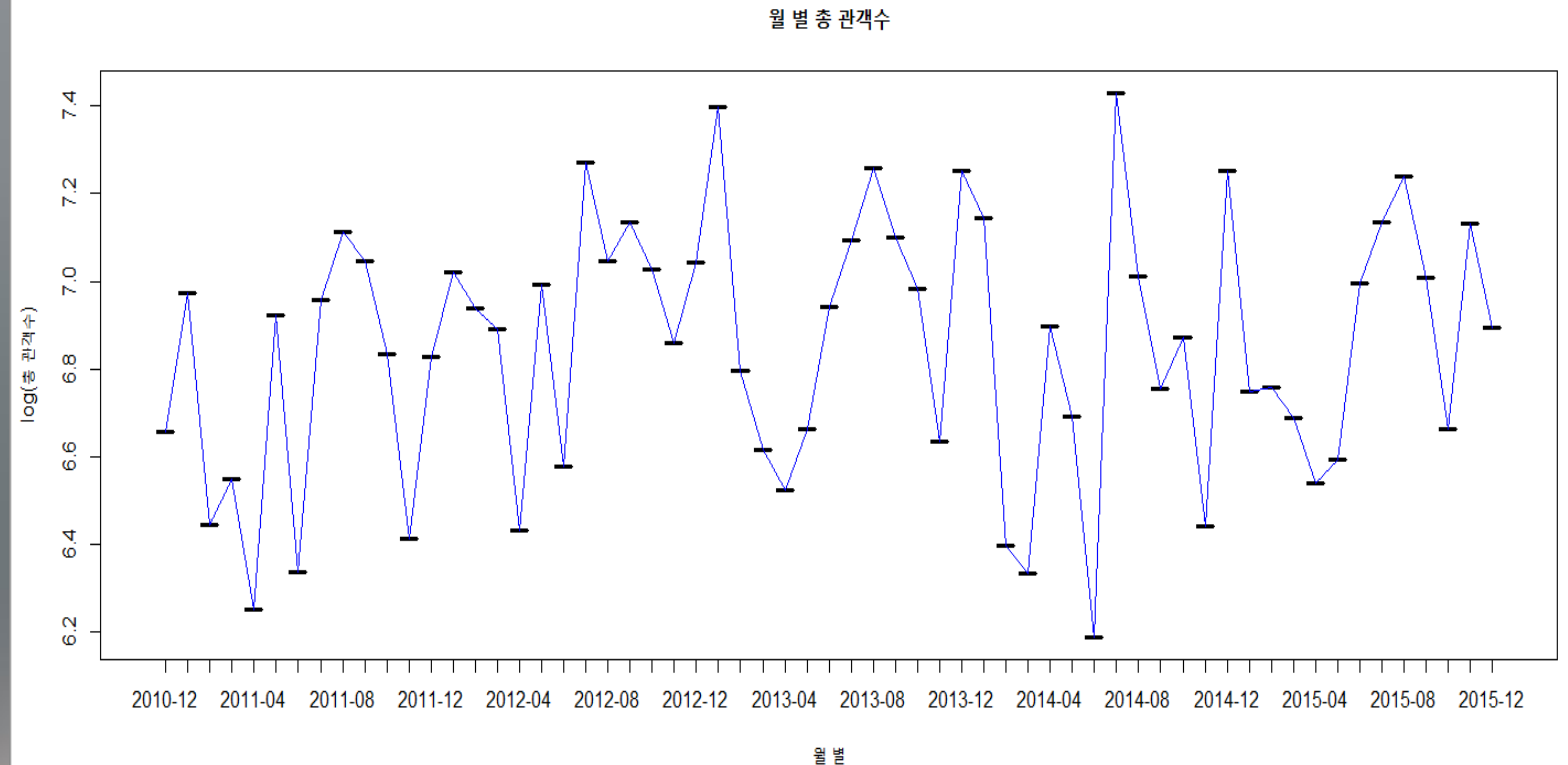
이론적 배경

자료분석

분석결과

2016년 개봉영화 예측

결론







...Time series estimation...

---



## ...Time series estimation...

---

Regression model :

$$f(t) = P(t) + T(t),$$

$$\text{where } P(t) = \sum_{n=0}^3 \alpha_n t^n$$

$$T(t) = \sum_{\substack{n|12 \\ n \neq 12}} \beta_n \cos(n\pi t) + \sum_{\substack{n|12 \\ n \neq 12}} \gamma_n \sin(n\pi t)$$

### 3. 시계열 분석

자료수집

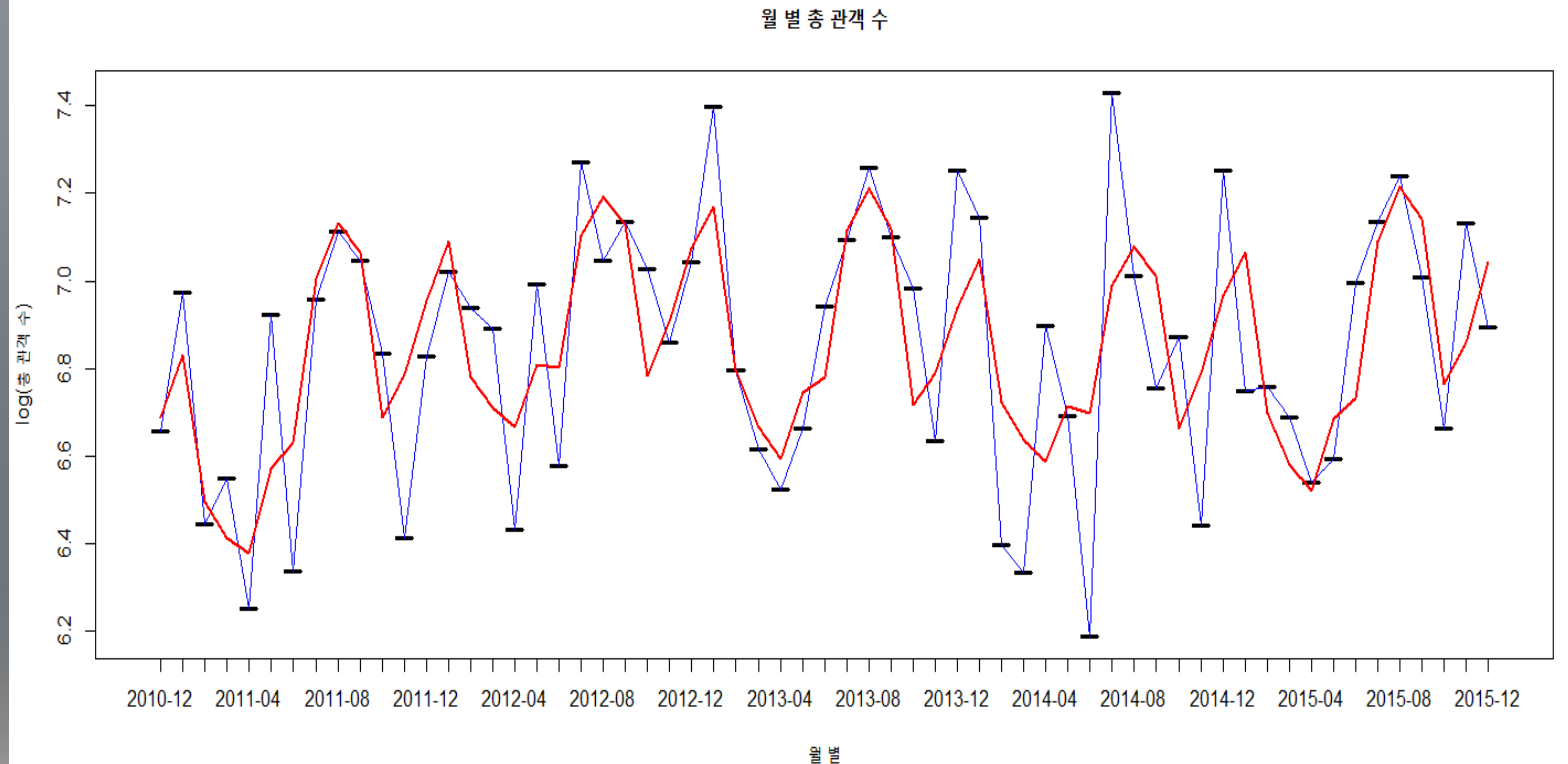
이론적 배경

자료분석

분석결과

2016년 개봉영화 예측

결론



# 분석결과



자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



결론

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}$$

$y_i$  : 관객 수 => 영화 평점

$x_1$  : 감독 평점

$x_2$  : 배우 평점

$x_3$  : 장르

$x_4$  : 등급

$x_5$  : 개봉일 데이터 + MODEL 2



		영화 제목	실제 평점	예측 평점 MODEL 1	예측 평점 MODEL 2
<div>📁 자료수집</div> <div>👤 이론적 배경</div> <div>👤 자료분석</div> <div>📊 분석결과</div>		아가씨	6.99	6.65	6.76
		검사외전	7.95	6.95	7.05
		곡성	7.61	7.29	7.38
		귀향	9.18	7.62	7.70
📄 2016년 개봉영화 예측		시간이탈자	7.20	7.23	7.31
📁 결론		그날의 분위기	7.89	7.14	7.22



자료수집



이론적 배경



자료분석



분석결과



2016년 개봉영화 예측



결론

# 결론

크롤링 스킴

감성 분석

시간

병렬화

시계열 분석

목적 사이트

# References

---

## 1. Data

- 영화진흥위원회(KOFIC) . <http://www.kobis.or.kr/>
- 네이버 영화. <http://movie.naver.com/>

2. Predicting the Future With Social Media . Bernardo A. Huberman (2010)

3. 한국어와 NLTK, Gensim의 만남. 박은정 (2015)



. 질문 이수과?.

---

Q & A

.고맙습니다.

---