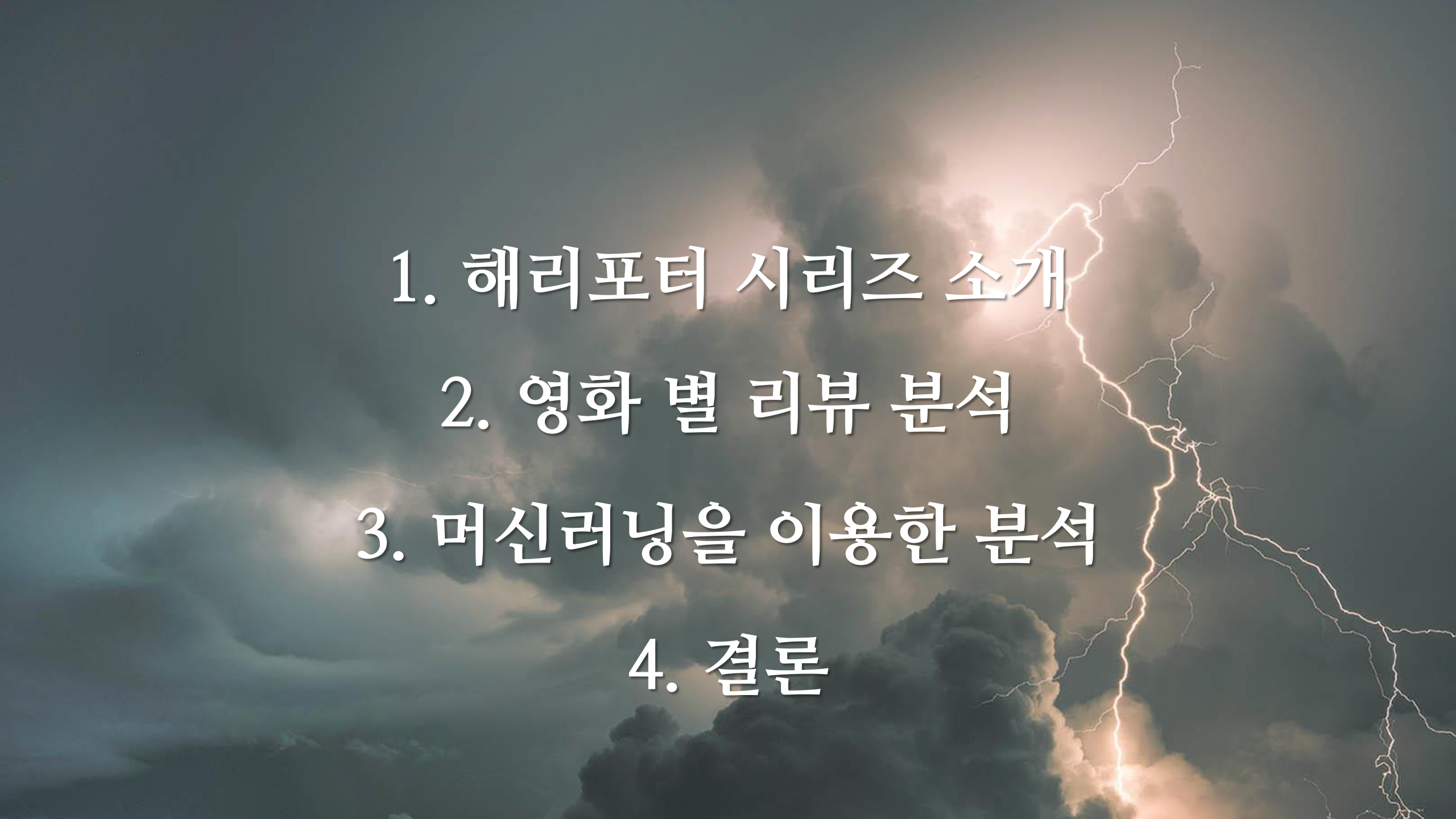




Naïve Bayes를 이용한 해리포터 시리즈 분석

- 
- A dramatic, dark sky filled with heavy, swirling clouds. A bright, jagged lightning bolt strikes down from the upper right, illuminating the surrounding clouds with a golden glow. The overall atmosphere is intense and powerful.
1. 해리포터 시리즈 소개
 2. 영화 별 리뷰 분석
 3. 머신러닝을 이용한 분석
 4. 결론

1. 해리포터 시리즈 소개



1. 해리포터 시리즈 소개



해리포터 시리즈 소개

1. 세계에서 가장 많이 팔린 소설책
2. 영국을 대표하는 문학 작품
3. 한국에서 1,000만부 이상 판매
4. 영화 수입 세계 2위 등재
5. 해리포터 후속 영화 꾸준히 개봉

2. 영화 별 리뷰 분석

평점 6점 이하 부정, 7점 이상 긍정으로 설정



2. 영화 별 리뷰 분석 (공통 분석 방법)

개발자 모드(F12) → network → pointwriteForm
추출 가능 주소가 별도로 존재

관람객 평점 3,580건 **내 평점 등록**

☒ 공감순 ☒ 최신순 ☒ 평점 높은 순 ☒ 평점 낮은 순 ☐ 스포일러 보기 ☐ 관람객 평점만 보기 ?

★★★★★ 10 설렘을 준 영화. 그 어린 나이에 즐거운 상상을 하게 해준. 판타지의 시작. 지금은 추억이 되어버렸
다는게 믿기힘들 정도로 내 삶의 큰 부분을 차지했었던 영화. 해리포터 개봉날은 연중행사였어...

달자씨(cvx1****) | 2013.06.29 03:20 | 신고

2340 38

★★★★★ 10 전설의 시작

약속님(jik****) | 2013.06.19 04:34 | 신고

2161 48

★★★★★ 10 전 이 영화를 잊지않을겁니다. 영원히요. 그리고, 제겐 역대 최고의 영화로 남을지도??? 모르겠네
요.

lovely gill(jeju****) | 2013.07.14 00:56 | 신고

1407 43

★★★★★ 10 이때로 돌아가고 싶다...

현정씨(sher****) | 2013.07.14 00:44 | 신고

1380 35

★★★★★ 10 애들 너무 귀여워 ㅠㅠ

클레오(song****) | 2013.06.21 01:18 | 신고

1046 28

영화 별 평점 게시물

관람객 평점 3,580건 **내 평점 등록**

☒ 공감순 ☒ 최신순 ☒ 평점 높은 순 ☒ 평점 낮은 순 ☐ 스포일러 보기 ☐ 관람객 평점만 보기 ?

★★★★★ 10 설렘을 준 영화. 그 어린 나이에 즐거운 상상을 하게 해준. 판타지의 시작. 지금은 추억이 되어버렸
다는게 믿기힘들 정도로 내 삶의 큰 부분을 차지했었던 영화. 해리포터 개봉날은 연중행사였어...

달자씨(cvx1****) | 2013.06.29 03:20 | 신고

2340 38

★★★★★ 10 전설의 시작

약속님(jik****) | 2013.06.19 04:34 | 신고

2161 48

★★★★★ 10 전 이 영화를 잊지않을겁니다. 영원히요. 그리고, 제겐 역대 최고의 영화로 남을지도??? 모르겠네
요.

lovely gill(jeju****) | 2013.07.14 00:56 | 신고

1407 43

★★★★★ 10 이때로 돌아가고 싶다...

현정씨(sher****) | 2013.07.14 00:44 | 신고

1380 35

★★★★★ 10 애들 너무 귀여워 ㅠㅠ

클레오(song****) | 2013.06.21 01:18 | 신고

1046 28

★★★★★ 7 호그와트... 그곳에 가고 싶다.

Elements Console Sources Network

Filter XHR JS CSS Img Media Font Doc WS Manifest Other Has blocked cookies

Blocked Requests

200 ms 400 ms 600 ms 800 ms 1000 ms 1200 ms 1400 ms 1600 ms

Name Headers Preview Response Initiator

search_bg.png

btn_search.png

sp_nav.png

btn_sp_bg4.png

sp_nav2.png

sp_btn.png

sp_jco3.png

h_tx_sp4.png

bg_tx_line2.gif

ico_arrow2.gif

sub_tab_bg.gif

spuginjs20201012

pointWriteFormList.nhn?code...

spri_grade_v2.png

blank.gif

ico_arrow4.gif

gnb_bg.gif

pointExtJs01.nhn?movieCod...

likelt.v2.desktopinit.js?y=Mon...

common.css?20200806170632

movie_tablet.css?20200806170632

movie_all.js?20200806170632

movie_end.css?20200806170632

likelt.v2.desktop.js

General

Request URL: https://movie.naver.com/movie/b1/mi/
tWriteFormList.nhn?code=30688&type=afterHisActua
ntWriteExecute=false&isMileageSubscriptionAlread
Is&isMileageSubscriptionReject=false

Request Method: GET

Status Code: 200

Remote Address: 210.89.168.33:443

Referrer Policy: unsafe-url

Response Headers

cache-control: no-cache

cache-control: no-store

content-encoding: gzip

content-language: ko-KR

content-length: 6930

content-type: text/html; charset=UTF-8

date: Sun, 11 Oct 2020 15:11:46 GMT

expires: Thu, 01 Jan 1970 00:00:00 GMT

p3p: CP="ALL CURa ADMa DEVa TAIa OUR BUS IND PHY

UNI PUR FIN COM NAV INT DEM CNT STA POL HEA PRE

OTC", CP="ALL CURa ADMa DEVa TAIa OUR BUS IND PH

L UNI PUR FIN COM NAV INT DEM CNT STA POL HEA PR

OTC"

진짜 주소

리뷰와 평점을 추출하기 위한 방법



2. 영화 별 리뷰 분석 (공통 분석 코드)

평점, 리뷰 추출

```
#
# 해리포터와 마법사의 돌

stone_df = pd.DataFrame(columns = ['review', 'score'])

stone_review_lst = []
stone_score_lst = []
for i in range(1, 160):
    url = 'https://movie.naver.com/movie/bi/mi/pointWriteFormList.nhn?code=30688&type=a'
    li = url.format(i)
    res = req.urlopen(li)
    soup = BeautifulSoup(res, 'html.parser')
    review = soup.select('div.score_reple')
    score = soup.select('div.star_score')
    for j in review:
        stone_review_lst.append(j.find('span').get_text().strip())
    for k in score:
        stone_score_lst.append(k.find('em').get_text().strip())

stone_review_lst
stone_score_lst

stone_df['score'] = stone_score_lst
stone_df['review'] = stone_review_lst
```

리뷰와 평점 추출 코드

※ 숫자, 영어 또한 의미가 있다고 판단

데이터 정제 ※ 특수문자와 인터넷 용어 제거

```
harryporter_room['review'] = harryporter_room['review'].str.replace('!!', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('π', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('τ', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('ㅋ', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('ㅎ', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('.', '')
harryporter_room['review'] = harryporter_room['review'].str.replace('[~!@#$%^&*~&
```

```
harryporter_room.columns = ['리뷰', '평점']
harryporter_room['긍정/부정'] = ''
harryporter_room = harryporter_room.reset_index(drop=True)
```

```
harryporter_room.loc[harryporter_room['평점'] >= 7, '긍정/부정'] = '긍정'
harryporter_room.loc[harryporter_room['평점'] <= 6, '긍정/부정'] = '부정'
```

```
harryporter_room.to_csv('c:/data/harryporter_room.csv', encoding='cp949')
harryporter_room
```

데이터 정제작업 및 긍정/부정 컬럼 추가



2. 영화 별 리뷰 분석 해리포터와 마법사의 돌

긍정 : 약 99.9 %

		리뷰	평점	긍정/부정
0	설렘을 준 영화 그 어린 나이에 즐거운 상상을 하게 해준 판타지의 시작 지금은 추억...	10	긍정	
1		전설의 시작	10	긍정
2	전 이 영화를 잊지않을겁니다 영원히요 그리고 제겐 역대 최고의 영화로 남을지도 모르겠네요	10	긍정	
3		이때로 돌아가고 싶다	10	긍정
4		애들 너무 귀여워	10	긍정
5		호그와트 그곳에 가고 싶다	7	긍정
6	나를 바꿔준 해리포터 정말 역대 최고의 영화이다 정말 이것을 만들게 해준 모든 사람...	10	긍정	
7	옛 추억이 갑자기 떠올라 오랜만에 봤는데 정말 재밌다 앞으로 이만한 판타지 영화가 ...	10	긍정	
8		제일 해리포터다운 해리포터 다들 뽕뽕송해	10	긍정
9		십대를 함께 걸어준 소중한 친구	10	긍정
10		그때 참 진짜 어딘가에 마법세계가 있었으면 좋겠다고 생각함지금도	10	긍정
11	오랜만에 봤는데 다시 봐도 재밌다 어릴 때 봤던 기억이 새록새록 떠오른다 그래서 더...	10	긍정	
12		시작부터 갑	10	긍정
13		윙가리움레비오우사아	10	긍정
14		그냥 말이 필요없음	10	긍정
15		해리포터를 만난건 우연이었다 그렇지만 인연이 되었다	10	긍정
16		추억의영화	10	긍정
17		다니엘 레드클리프	10	긍정
18	정말 대작이고 재밌습니다 제가 지금까지 본 영화 중에서 가장 재밌습니다	10	긍정	
19		마법사에돌 해리포터정말 귀엽고 멋져	10	긍정
20		다시는 없을 환상적인 작품	10	긍정
21		해리포터 시리즈중 가장 명작	10	긍정
22	아직 어린나이인데 나에게 너무 재미를 준 영화라고 생각하고 나는 이영화를10점으로 ...	10	긍정	
23		상상이 눈 앞에서 펼쳐졌을 때의 감동이란	10	긍정
24		The legend began here	10	긍정
25		진짜 초등학교 때 봤는데 그때는 새로운 뭔가 었다	10	긍정

부정 : 약 0.1 %

추출 결과(1420개)



2. 영화 별 리뷰 분석 해리포터와 비밀의 방

긍정 : 약 95 %

	리뷰	평점	긍정/부정
0	십점만점에 십점	10	긍정
1	1편만큼 대단한 2편	9	긍정
2	겁나 재밌음	10	긍정
3	크리스 콜럼버스가 12 시리즈를 맡은게 정말 잘한 일이라 생각함 해리포터의 기둥을 ...	10	긍정
4	어렸을때 맨날 이것만 봤었다	10	긍정
5	해리포터잖아요 해리포터니까요 해리포터라는 해리포터의 해리포터같은 이야기 다시는 나오...	10	긍정
6	짱짱	10	긍정
7	해리포터는 몇번이나 봐도 안질린다 경주행중	10	긍정
8	아이와 어른 모두를 꿈꾸게 하는 영화	8	긍정
9	이건 10점 만점에 1000점을 주어도 모자르지 않다	10	긍정
10	속편은 재미 없다는 이야기가 많은데 그걸 깬 영화	9	긍정
11	어떻게 이 세계 속에 완벽히 빠져들 수 있게 하는 이런 수작이 나올 수가 있을까 정...	10	긍정
12	다시 오지 못할 내 어린시절아 사랑해	10	긍정
13	추억의 해리포터	10	긍정
14	이 영화는 정말 좋다연기도 실감나서 정말 재미있다해리포터 영화는 30번도 넘게 봤는...	10	긍정
15	1편에 이어 2편마저도 훌륭하게 만들었다고 생각한다	10	긍정
16	1편의 분위기를 잘 계승 한듯	9	긍정
17	짱	10	긍정
18	어렸을 때 이거 보고 무서워서 잠 못잤었는데 재미있는 영화	10	긍정
19	무조건 해리 포터 시리즈는 10 점	10	긍정
20	2편 레알이지 쫄았었어	10	긍정
21	불멸의 명작 시리즈	10	긍정

부정 : 약 5 %

추출 결과(1460개)



2. 영화 별 리뷰 분석 해리포터와 아즈카반의 죄수

긍정 : 약 91 %

부정 : 약 9 %

		리뷰	평점	긍정/부정
0	해리포터 시리즈 중에 제일 좋아하는 편 이걸 대체 몇 번 봤는지 모르겠다진짜 레알임	10	긍정	
1	해리포터 시리즈 중에 아즈카반의 죄수가 제일 재밌음	10	긍정	
2	개인적으로 해리포터 모든시리즈중 가장 명작정말 재밌다	10	긍정	
3	틴무비 장르의 틀을 벗겨낸 알폰소 쿠아론의 연출력 점점 더 깊어만 가고 어두워지며 ...	9	긍정	
4	해리포터 시리즈 중에 내가 제일 좋아하는 편 어둡고 음침한 분위기가 정말 좋다 아동...	9	긍정	
5	왜 해리포터 시리즈 중 아즈카반의 죄수가 가장 흥행하지 못했는지 의문이다	10	긍정	
6	12편보다 더 흥미진진해서 재미있다	10	긍정	
7	이때부터 유치한해리포터가 사라짐 해리포터시리즈중 가장 몰입도가 높고 재밌음	8	긍정	
8	12편의 기술적 촌스러움이 사라진편 제일재밌던 편이다	10	긍정	
9	해리포터 복습하고 있는데 12편에 비해 훨씬 흥미진지하고 감동적임 최고	10	긍정	
10	해리포터 시리즈 중에서 제일 재미있다 해리포터 작가가 말했던 우스꽝스럽고 기괴한 분...	10	긍정	
11	알폰소 쿠아론감독은 진짜 최고다 라는 말밖에 안나옴 분위기며 연출이며그냥 편다	10	긍정	
12	알고보면 이 편이 가장 재밌다 그리고 해리포터 시리즈중 해리가 가장 잘생기게 나온...	10	긍정	
13	정말탄탄한스토리 영화다시보면 복선이정말많이깔려있음을알수있다 엔딩크레딧도멋지고 해리포...	10	긍정	
14	해리포터 시리즈 중에서 가장 완성도가 높은 편중 하나라고 꼽을 수 있는 아즈카반의 죄수	9	긍정	
15	해리포터는 책으로도 영화로도 보는 사람을 빠져들게 해 그것이 매력적이고 내가 지금까지...	10	긍정	
16	진짜 살면서 제일많이본영화가 아즈카반에죄수 8 번은봤다 대사도외울뻔	10	긍정	
17	와 나도 이 편이 가장 신나고 재밌던데 한 마음을 가진 사람이 이렇게도 많다니	10	긍정	
18	시리즈 중 가장 좋아	10	긍정	
19	알폰소 쿠아론이 이후로도 계속 만들었으면 어땠을까	10	긍정	
20	아즈카반 편은시간구성 때문에 흥미로운 명작임	10	긍정	
21	해리 포터 시리즈중 가장 원작을 잘살려내면서 감독의 매력을 충분히 보여준 작품 알폰...	10	긍정	
22	아 스네이프교수가 뒤돌아 늑대인간 보이자 애들 보호하려고 팔 뒤로 빼는 거에서부터 ...	10	긍정	
23	해리포터 시리즈 중 최고다 역시 알폰소쿠아론	10	긍정	
24	1 2 3 편이 재밌는데 그 중에서도 3편을 제일 재밌다고 생각한다	10	긍정	
25	해리포터 시리즈 중엔 제일 좋았지	9	긍정	

추출 결과(1442개)



2. 영화 별 리뷰 분석 해리포터와 불의 잔

긍정 : 약 79 %

부정 : 약 21 %

		리뷰	평점	긍정/부정
0		문화와 언어는 다르지만 뛰는 맥박은 똑같다	9	긍정
1	ooo 난 이편 재미있던데 쿼디치 경기에다 트리위저드 시합에다 드래곤 난 진짜 재미...		10	긍정
2		론이 엑소보다 잘생김	10	긍정
3		이 영화에게 나는 무엇을 바라는가 존재만으로도 나는 즐겁다	10	긍정
4	기대했던장면들이 나오지않아 원작에비해 아쉬운점이너무많지만 그래도 해리포터시리즈는 다치고		10	긍정
5		스네이프의 속을 도무지 알수가없음	8	긍정
6		전 작에 비해 상당히 화려해졌다	8	긍정
7		4편이 개인적으로 난 제일 재미있었음	10	긍정
8		어쩔수없다 책이 너무 길었어 하지만 해리포터니깐	8	긍정
9	12편에 비해 특수 효과와 영상미는 풍부해졌으나 모든 것이 3편보다 못 하다		7	긍정
10		내평생최고의영화언제나재미있다	10	긍정
11		이 영화 다 좋은데 누가 해리랑 론 머리 좀 잘라주지 그랬냐	10	긍정
12		해리포터는 강 10점 만점에 10	10	긍정
13	스킵 너무 심하잖아볼거리 살리려고 스토리 포기함 쿼디치스토리 디테일도 다빠지고 해리...		6	부정
14	하 해리포터라서 10점만점을 줘도 모자라다 생각하지만 원작 소설의 팬으로써는두편으로...		7	긍정
15	여전히 화려한 영상미를 자랑하면서 주인공들의 사춘기 시절을 잘 그려냈다 다만 트라이...		7	긍정
16	덤블도어 캐릭터가 상당히 터프해졌다 그 전작에서는 조심스러웠는데 상남자 교장선생...		7	긍정
17	개인적으로 진짜진짜좋아하는 로버트패턴슨이 나와서 더훈훈하게봤다는점해리포터와 로버트가...		10	긍정
18		헤르미온느잘 컸구낭	10	긍정
19		전작들보다 재밌는거 같은데 평점은제가 책을 안읽어서 그럴까요	8	긍정
20	영화와 책의 차이 불의 잔 부터 두편으로 찍었어야 했다 책에 비해 잘려나간 부분이...		9	긍정
21		엠마왓슨의 연기력 스토리의 다름 3편의 위엄	5	부정
22		하해리의비쥬얼이근데이편은정말재밌었음무엇보다도 엠마왓슨	10	긍정
23	감독이 바뀌니 역시 3편과 비교가 많이 된다 알폰소 쿠아론이 최고의 걸작을 만들었...		8	긍정
24		그래도 소설을 영상으로 만난다는 기쁨	6	부정
25		과도기의 배우들과 같은 시기에 놓인 영화	6	부정

추출 결과(1471개)



2. 영화 별 리뷰 분석 해리포터와 불사조 기사단

긍정 : 약 75 %

부정 : 약 25 %

	리뷰	평점	긍정/부정
0	엠브릿지 캐스팅 진짜 신의 한수다 ㄷㄷ 어쩔 저렇게 싫냐	10	긍정
1	여성부 장관 뺨치는 엠브릿지 연기	9	긍정
2	엠브릿지 캐스팅 최고이 캐릭터를 이토록 찰지게 연기할 수 있는 배우가 있을까	8	긍정
3	해리포터와 초창의 키스가 있었던 영화 엠브릿지가 특히 돋보이는 영화 시리우스가 죽어...	10	긍정
4	전체적으로어둡지만 스틸넘치는편네빌과루나의매력이두드러진	10	긍정
5	데이빗에이즈 감독은 책을 보긴 한것일까 헤르미온느와 론의 반장직과 DA회원들의 동전...	3	부정
6	조앤롤링은 영화에 만족하는지 궁금하다	5	부정
7	야 이새끼드라 해리포터불사조기사단 책이 총5편이야 어떻게 그걸 2시간 17분영화에 ...	8	긍정
8	불사조 기사단 이후로 잘 알지도 못하는 데이빗 에이즈라는 사람이 싫어졌다	3	부정
9	어렸을때 손에 땀을 쥐고 봤었는데 나는 해리포터 시리즈중 가장 재밌었다	10	긍정
10	데이빗 에이즈 감독이 영화를 다 망치네 1 2 3 4 편은 재미있었는데 5편 이후로...	1	부정
11	돌로레스 엠브리지 죽이고 싶었음	10	긍정
12	당시 알바 상당했나봄 기본적으로 해리포터 시리즈는 배우들도 캐릭터 위해 몸관리하느...	10	긍정
13	여시서부터 해리포터가 꼬이기 시작했다	1	부정
14	가장 긴 책인데 영화 중 가장 짧단 애길 듣고 더 빠졌달 진짜 책에서 나온 내용 중...	10	긍정
15	원작에 반도 못 미쳤다	1	부정
16	해리 포터가 진리 아니겠습니까물론 개인적인 생각입니다만 해리 포터 짱	10	긍정
17	시리즈중 최고 해리의 운명이지닌 고독이 이렇게 절절히 와닿았던 적이 없었고 그 옆에...	9	긍정
18	왜 평점이 낮은진 의문6점짜리 영화는 절대로 아님이때부터가 해리포터의 시작인데	10	긍정
19	너무 많이 압축되었다 하이라이트를 보는듯	3	부정
20	난 왜 원작 소설을 읽었는데도 영화는 날라댕기는 것밖에 기억에 없나	1	부정
21	소설을 영화가 감당하지 못한다	6	부정
22	친구들 단합 오졌고난 이편이 젤 재밌는거같다다 재밌지만 해리포터는 여태 안재밌는 편...	10	긍정
23	조금 지루했지만 엠브릿지가 돋보였다 시리우스가 허무하게 죽어서 아쉽다	8	긍정
24	시리우스 죽을때 너무 슬펐고 명장면도 너무많은데 평점이 왜 이따구인지개인적으로 해리...	10	긍정
25	러닝타임을 너무 짧게 잡았어— 크레딧 제외 상영시간만 2시간 반은 잡아야 했는데	7	긍정

추출 결과(1462개)



2. 영화 별 리뷰 분석 해리포터와 혼혈 왕자

긍정 : 약 73 %

		리뷰	평점	긍정/부정
0	왜 평점 일부러 낮추냐 이상한 놈들아	1990년2000년대 생들은 이 영화랑 성장을 ...	10	긍정
1		스네이프	10	긍정
2		영화를 다시 봤는데 덤블도어의 죽음이 이렇게 슬픈지 몰랐다	10	긍정
3	해리 포터와 죽음의 성물 시리즈로 나아가기 위한 발판을 잘 마련하다	칙칙하고 어두운...	7	긍정
4		알버스의 죽음이 영화를 빛나게 하였다	10	긍정
5		벨라트릭스 처죽이고싶다	10	긍정
6	책보다 재미없으면 책만보세요 왜영화보고 평점을 이리주시지난	개인적으로 인물들 캐스팅...	9	긍정
7		혼혈왕자 보세요 두번 보세요 말이 필요없음	10	긍정
8		다시 봤는데 왜이렇게 재밌지	9	긍정
9		이때는 스네이프를 진짜 최악의 악당이라고 생각했지	10	긍정
10		덤블도어 교수님이 돌아가셔서 너무 슬픔	10	긍정
11		평점이 왜이럼어두운 분위기가 정말 매력적이었는데	10	긍정
12		짱이다	10	긍정
13		혼혈 왕자가 스네이프교수 였던게 가장 인상깊었다	10	긍정
14	내용도 흥미진진했고 배우들이 성숙해져서 좋았다근데	평점이 왜이리 낮아	10	긍정
15		7편을 위한 분위기를 잘 깔았음 영상미랑 동굴에서의 씬은 압권	10	긍정
16	개쎈 진짜 다시봐도 명작이다 다른편에비해서는	지루한감이 있었지만 뭐어떤가 발판도 잡...	10	긍정
17		덤블도어 죽을때 진짜 눈물나더라	10	긍정
18	책은 안봐서 잘 모르겠지만 초중반의	멜로모드도 재미났고 그 후에도 역시 좋았습니다 ...	10	긍정
19	긴 말 필요없고 말도 안 되는	6점대의 평균별점을 높여야 하는건 사실시리즈물은 탄탄...	10	긍정
20		잠시 쉬어가며 훌륭한 마무리를 위한 준비	8	긍정
21	마법사의돌부터 복습중인데	아즈카반의죄수와 더불어 제일 재밌었다덤블도어 교장 존경스러웠음	10	긍정
22	책이랑 비교하기에는 책이 너무 많은 것을 담고있다	이정도면 잘 풀어나간 편 무엇보다...	10	긍정
23		어두운분위기가 정말 매력적이었던화	10	긍정

부정 : 약 27 %

추출 결과(1422개)



2. 영화 별 리뷰 분석 해리포터와 죽음의 성물 part 1

긍정 : 약 81 %

부정 : 약 19 %

		리뷰	평점	긍정/부정
0		도비왜죽어	10	긍정
1	한 편짜리 영화를 반으로 나눠 보는 느낌이라 초반이 지루한건 사실 대단원의 막을 위...		7	긍정
2		헤르미온느 진짜 넘 이쁨	9	긍정
3		2부를 향한 관문이지만 혼혈왕자보다 몰입이나 분위기가 더 낫다	8	긍정
4		도비때문에 폭풍눈물	9	긍정
5		혼혈왕자 2부를 있는 연결고리	10	긍정
6	릴리와 같은 패트로누스인 암사슴을 보여주면서 라고 말하던 장면이 잊혀지지가 않는다 ...		8	긍정
7	해리랑 헤르미온느 텐트안에서 서로를 의지하며 춤추는 장면에 나도 모르게 눈물이 흘렀다		10	긍정
8	여기서 잠깐 한숨돌리고 휴게소에 들리 겠습니다 그 휴식을 통해 힘차게 달리기를		7	긍정
9	죽음의성물 인터넷에 돌아다니는 자막들 오류가 너무 많아서 팬으로서 화난다 어떤 상황...		9	긍정
10		영화가 수학공식이야 한순간이라도 놓치면 다음 내용이 헛갈려	7	긍정
11		배우들과 함께 시리즈도 성장한다	7	긍정
12	첨엔 별로였지만 짐 다시보니까 지난시리즈와는 확실히 다른 분위기가 있어서 맘에드넹		8	긍정
13	해리포터 전 시리즈중에 제일 인물 심리묘사나 구체적인 이야기 그리고 어린애들 영화같...		9	긍정
14		긴말할것없이 최고 였는데	10	긍정
15			10	긍정
16		아 도비 1편 마지막 부분은 진짜 눈물 펄펄	10	긍정
17		책의 느낌을 잘 살려준 것 같다	10	긍정
18	너무 재밌게 봄 4편 5편은 너무 내용 많이 쳐내서 엄청 실망했었는데 일케 나눠 찍...		10	긍정
19	중3때 영화관에서 봤던 그 영화 참 추억 마지막에 허무했지만 2부를 향한 관문이었을 뿐		10	긍정
20		해리포터는 어느편이든 재미있다	10	긍정
21		난 이번 편 엄청 몰입되던데 특히 도비죽을때 너무슬퍼따	10	긍정
22	론 팔 분리될뻔했을때는 마법으로 잘만 고치더라 도비 죽어갈때는 왜 아무 힘도 안써보는거임		5	부정
23		그렇게 욕할 거면 차라리 자기가 영화 보지 말던가	10	긍정
24	원작에 충실한 해리포터 영화는 마지막까지 볼 수 없는 건지 착잡했는데 최근의 해리포...		7	긍정
25		12부로 나눈것치고 정말 재밌었는데어둡고 또 원작을 그대로 잘 그려내서 난 정말 좋았다	10	긍정

추출 결과(1456개)



2. 영화 별 리뷰 분석 해리포터와 죽음의 성물 part 2

긍정 : 약 97 %

부정 : 약 3 %

		리뷰	평점	긍정/부정
0	역대 최강의 반전 스네이프 7편을 보면 해리포터의 진정한 주인공은 스네이프인거 같...	10	긍정	
1	10년간의 마법기차여행을 마칩니다 이 긴 여행이 우리에게 추억이 되어버렸네	8	긍정	
2	해리 아들로 또 만들었음 좋겠어요 영원히 안끝날까 같았는데 허전하고 아쉽다 해리포터...	10	긍정	
3	12부는 제목답게 이어지는 거니까 스토리 부족은 전혀 생각나지 않고 완벽했다고 본다...	10	긍정	
4	닥쳐 년10점이야	10	긍정	
5	시리즈는 끝났지만 뭔가 그 뒷이야기가 무척궁금해지는 10년간 성장뒤에 숨겨진 아픔과...	10	긍정	
6	아 씨발 세베루스 스네이프 스네이프성님 아 이런 완벽한 반전의 나쁜놈같으니라고 하 ...	10	긍정	
7	어른이된 헤르미온드가 호그와트행 급행열차를 타고 떠나는 딸을 보면서 짓는 아련한 표...	10	긍정	
8	영화가 끝나고 엔딩컷이 나오는데 난쟁지모르게 눈물나더라 내 인생중 10년을 해리포터로 보냈...	10	긍정	
9		10	긍정	
10	나와같이성장한영화 평생기억에남을것같다	10	긍정	
11	원래 시리즈영화는 갈수록망하는데 해리포터는 아닌듯 해리포터와 같이 성장한1인 이런영...	10	긍정	
12	다 때려치고 스네이프 짱짱맨	10	긍정	
13	대망의 장식 끝은 완벽했다	10	긍정	
14	영화관에서 안본게 후회된다	10	긍정	
15	주인공은 해리포터가 아니라 스네이프였다	10	긍정	
16	스네이프가 패트로누스를 꺼내는 장면을 잊을 수가 없다	10	긍정	
17	내가 소장하고 싶은 영화 를 작년부터 모아왔었는데 단연 내가 제일 처음으로 모은 시...	10	긍정	
18	해리포터 시리즈 중 제일 잘만듬	10	긍정	
19	17년 동안 칼을 갈다가 두대 맞고 운명하신 볼드모트 경 을 추모하며	10	긍정	
20	볼드모트 진짜 해리 확인사살도 안하고 ㅋㅋ 멍청하구만 멍청해	9	긍정	
21	영원한 내 친구	10	긍정	

추출 결과(1438개)



2. 영화 별 리뷰 분석 분석 결과

영화 별 리뷰 분석 결과

리뷰/평점	마법사의 돌	비밀의 방	아즈카반의 죄수	불의 잔	불사조 기사단	혼혈 왕자	죽음의 성물 1	죽음의 성물 2	평균
긍정 비율	99.9%	95%	91%	79%	75%	73%	81%	97%	86%
부정 비율	0.1%	5%	9%	21%	25%	27%	19%	3%	13%
전체 리뷰 수	1420	1460	1442	1471	1462	1422	1456	1438	1446

※ 6점 이하 부정, 7점 이상 긍정

3. 머신러닝을 이용한 분석



3. 머신러닝을 이용한 분석 (공통 분석 코드)

정확도

```
# 정확도
cv = CountVectorizer()
x_train = cv.fit_transform(harryporter_stone['리뷰'])
x_train
y_train = harryporter_stone['긍정/부정']

encode_cv = x_train.toarray()
cv.inverse_transform(encode_cv[0:2])
cv.get_feature_names()

nb = MultinomialNB()
nb.fit(x_train, y_train)

x_test = cv.transform(harryporter_stone['리뷰'])
y_predict = nb.predict(x_test)
accuracy_score(harryporter_stone['긍정/부정'], y_predict)
```

NaiveBayes

```
# NaiveBayes
harry_word = Okt()
stone_allword = set(word for i in harryporter_stone['리뷰'] for word in word_tokenize(i))
stone_allword
stone_data = harryporter_stone[['리뷰', '긍정/부정']].values.tolist()
stone_t = [{word : (word in word_tokenize(x[0])) for word in stone_allword}, x[1]] for x in stone_data
stone_t

classifier = nltk.NaiveBayesClassifier.train(stone_t)
classifier.show_most_informative_features()
```

classification_report

```
stop_word = []
with open('c:/data/stop_words.txt', 'r', encoding='utf-8') as s:
    for word in s:
        stop_word.append(word.strip())
stop_word
from konlpy.tag import Okt
pos = Okt()
def text_tokenizing(doc):
    return [word for word in pos.morphs(doc) if word not in stop_word and len(word) > 1]
contents_token = [text_tokenizing(i) for i in harryporter_stone['리뷰']]
contents_token[0:2]
contents = [' '.join(i) for i in contents_token]
contents[0:2]
X = contents
Y = harryporter_stone['긍정/부정']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2)
Counter(Y_train)
Counter(Y_test)
cv = CountVectorizer()
X_train = cv.fit_transform(X_train)
nb = MultinomialNB()
nb.fit(X_train, Y_train)
X_test = cv.transform(X_test)
print(classification_report(Y_test, nb.predict(X_test)))
```

classification_report와 NaiveBayes 코드



3. 머신러닝을 이용한 분석 마법사의 돌

	precision	recall	f1-score	support
긍정	0.99	1.00	1.00	283
비정상 부정	0.00	0.00	0.00	2
accuracy			0.99	285
macro avg	0.50	0.50	0.50	285
weighted avg	0.99	0.99	0.99	285

Most Informative Features			
방에서 = True	부정 : 긍정	=	100.6 : 1.0
비밀의 = True	부정 : 긍정	=	100.6 : 1.0
실제 = True	부정 : 긍정	=	100.6 : 1.0
못하는 = True	부정 : 긍정	=	100.6 : 1.0
화려한 = True	부정 : 긍정	=	100.6 : 1.0
장면은 = True	부정 : 긍정	=	100.6 : 1.0
장면 = True	부정 : 긍정	=	100.6 : 1.0
시작은 = True	부정 : 긍정	=	100.6 : 1.0
많아서 = True	부정 : 긍정	=	100.6 : 1.0
약간 = True	부정 : 긍정	=	100.6 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 비밀의 방

	precision	recall	f1-score	support	
	긍정	0.94	1.00	0.97	274
비정상	부정	0.00	0.00	0.00	18
accuracy			0.94	292	
macro avg	0.47	0.50	0.48	292	
weighted avg	0.88	0.94	0.91	292	

Most Informative Features

어린이 = True	부정 : 긍정 =	47.9 : 1.0
않은 = True	부정 : 긍정 =	28.7 : 1.0
방이라는 = True	부정 : 긍정 =	20.5 : 1.0
보다는 = True	부정 : 긍정 =	20.5 : 1.0
성인이 = True	부정 : 긍정 =	20.5 : 1.0
원작을 = True	부정 : 긍정 =	20.5 : 1.0
바실리스크와의 = True	부정 : 긍정 =	20.5 : 1.0
마법사가 = True	부정 : 긍정 =	20.5 : 1.0
한번 = True	부정 : 긍정 =	20.5 : 1.0
이젠 = True	부정 : 긍정 =	20.5 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 아즈카반의 죄수

	precision	recall	f1-score	support
긍정	0.88	1.00	0.94	254
비정상 부정	0.00	0.00	0.00	35
accuracy			0.88	289
macro avg	0.44	0.50	0.47	289
weighted avg	0.77	0.88	0.82	289

Most Informative Features

안보고 = True	부정 : 긍정 =	17.3 : 1.0
드한 = True	부정 : 긍정 =	17.3 : 1.0
안 = True	부정 : 긍정 =	17.3 : 1.0
3편이 = True	부정 : 긍정 =	17.3 : 1.0
점수 = True	부정 : 긍정 =	10.4 : 1.0
담기에 = True	부정 : 긍정 =	10.4 : 1.0
전환 = True	부정 : 긍정 =	10.4 : 1.0
이하 = True	부정 : 긍정 =	10.4 : 1.0
전편들에 = True	부정 : 긍정 =	10.4 : 1.0
아니다 = True	부정 : 긍정 =	10.4 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 불의 잔

	precision	recall	f1-score	support
긍정	0.82	0.97	0.89	240
부정	0.27	0.05	0.09	55
accuracy			0.80	295
macro avg	0.54	0.51	0.49	295
weighted avg	0.72	0.80	0.74	295

Most Informative Features

최악 = True	부정 : 긍정 =	14.3 : 1.0
안 = True	부정 : 긍정 =	11.7 : 1.0
관찰했는데 = True	부정 : 긍정 =	9.1 : 1.0
없고 = True	부정 : 긍정 =	9.1 : 1.0
확실히 = True	부정 : 긍정 =	9.1 : 1.0
원작의 = True	부정 : 긍정 =	9.1 : 1.0
시리즈가 = True	부정 : 긍정 =	9.1 : 1.0
최악의 = True	부정 : 긍정 =	9.1 : 1.0
감독 = True	부정 : 긍정 =	7.0 : 1.0
엠마왓슨의 = True	부정 : 긍정 =	6.5 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 불사조 기사단

	precision	recall	f1-score	support
긍정	0.79	0.93	0.85	217
부정	0.57	0.28	0.37	76
accuracy			0.76	293
macro avg	0.68	0.60	0.61	293
weighted avg	0.73	0.76	0.73	293

Most Informative Features

최악 = True	부정 : 긍정	=	35.5 : 1.0
도대체 = True	부정 : 긍정	=	11.2 : 1.0
지루하다 = True	부정 : 긍정	=	9.1 : 1.0
무슨 = True	부정 : 긍정	=	9.1 : 1.0
별로 = True	부정 : 긍정	=	7.9 : 1.0
예이츠 = True	부정 : 긍정	=	7.9 : 1.0
그래도 = True	긍정 : 부정	=	7.1 : 1.0
못한 = True	부정 : 긍정	=	7.1 : 1.0
소설 = True	부정 : 긍정	=	7.1 : 1.0
원작의 = True	부정 : 긍정	=	7.1 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 혼혈 왕자

	precision	recall	f1-score	support
긍정	0.80	0.95	0.87	211
부정	0.71	0.32	0.44	74
accuracy			0.79	285
macro avg	0.75	0.64	0.66	285
weighted avg	0.78	0.79	0.76	285

Most Informative Features

없고 = True	부정 : 긍정	=	19.9 : 1.0
평점이 = True	긍정 : 부정	=	9.1 : 1.0
그저 = True	부정 : 긍정	=	8.5 : 1.0
지루하고 = True	부정 : 긍정	=	8.5 : 1.0
내용도 = True	부정 : 긍정	=	8.5 : 1.0
중요한 = True	부정 : 긍정	=	8.5 : 1.0
실망 = True	부정 : 긍정	=	6.9 : 1.0
별로 = True	부정 : 긍정	=	6.9 : 1.0
지루 = True	부정 : 긍정	=	6.6 : 1.0
건 = True	부정 : 긍정	=	6.6 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 죽음의 성물 part 1

	precision	recall	f1-score	support
긍정	0.84	0.99	0.91	240
부정	0.73	0.15	0.25	52
accuracy			0.84	292
macro avg	0.79	0.57	0.58	292
weighted avg	0.82	0.84	0.79	292

Most Informative Features

지루하다 = True	부정 : 긍정	=	26.5 : 1.0
지루하고 = True	부정 : 긍정	=	12.6 : 1.0
재미가 = True	부정 : 긍정	=	12.6 : 1.0
봤다 = True	부정 : 긍정	=	10.9 : 1.0
재미는 = True	부정 : 긍정	=	9.8 : 1.0
재미없음 = True	부정 : 긍정	=	7.0 : 1.0
매우 = True	부정 : 긍정	=	7.0 : 1.0
집중이 = True	부정 : 긍정	=	7.0 : 1.0
없고 = True	부정 : 긍정	=	7.0 : 1.0
내내 = True	부정 : 긍정	=	7.0 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 죽음의 성물 part 2

	precision	recall	f1-score	support
긍정	0.94	1.00	0.97	272
부정	0.00	0.00	0.00	16
accuracy			0.94	288
macro avg	0.47	0.50	0.49	288
weighted avg	0.89	0.94	0.92	288

Most Informative Features

평점 = True	부정 : 긍정 =	49.0 : 1.0
평점음 = True	부정 : 긍정 =	45.4 : 1.0
소셜 = True	부정 : 긍정 =	27.2 : 1.0
당연한 = True	부정 : 긍정 =	27.2 : 1.0
1점도 = True	부정 : 긍정 =	27.2 : 1.0
아나 = True	부정 : 긍정 =	27.2 : 1.0
이영화 = True	부정 : 긍정 =	27.2 : 1.0
1점 = True	부정 : 긍정 =	27.2 : 1.0
모르겠다 = True	부정 : 긍정 =	27.2 : 1.0
차라리 = True	부정 : 긍정 =	27.2 : 1.0

classification_report와 NaiveBayes 결과



3. 머신러닝을 이용한 분석 종합 결과

영화 별 NaiveBayes 결과

		마법사의 돌	비밀의 방	아즈카반의 죄수	불의 잔	불사조 기사단	혼혈 왕자	죽음의 성물 1	죽음의 성물 2	평균
Precision (정밀도)	긍정	99%	96%	90%	82%	79%	80%	83%	97%	88%
	부정	0%	0%	0%	27%	57%	80%	62%	0%	28%
Recall (재현율)	긍정	100%	100%	100%	97%	93%	97%	98%	100%	98%
	부정	0%	0%	0%	5%	28%	32%	14%	0%	9%
F1-score (조화 평균)	긍정	100%	98%	95%	89%	85%	88%	90%	98%	92%
	부정	0%	0%	0%	9%	37%	45%	23%	0%	14%
Accuracy (정확도)		99%	96%	90%	80%	76%	80%	82%	97%	87.5%

※ 원인 : 긍정, 부정의 비율의 큰 차이, 부족한 데이터 수



3. 머신러닝을 이용한 분석 문제점

```
text = pd.Series('이 영화 정말 재미없고 극혐이야')
z_test = cv.transform(text)
nb.predict(z_test)
```

```
text = pd.Series('이 영화 똥이야')
z_test = cv.transform(text)
nb.predict(z_test)
```



```
text = pd.Series('이 영화 정말 재미없고 극혐이야')...
```

```
array(['긍정'], dtype='<U2')
array(['긍정'], dtype='<U2')
```

```
text = pd.Series('이 영화 똥이야')...
```

```
array(['긍정'], dtype='<U2')
```

		리뷰	평점	긍정/부정
0	역대 최강의 반전 스네이프 7편을 보면 해리포터의 진정한 주인공은 스네이프인가 같...	10		1
1	10년간의 마법기차여행을 마칩니다 이 긴 여행이 우리에게 추억이 되어버렸네	8		1
2	해리 아들로 또 만들었음 좋겠어요 영원히 안끝날까 같았는데 허전하고 아쉽다 해리포터...	10		1
3	12부는 제목답게 이어지는 거니까 스토리 부족은 전혀 생각나지 않고 완벽했다고 본다...	10		1
4	닥쳐 년10점이야	10		1
...
1433	이제 딱히 영화관 가고 싶지가 않다	10		0
1434	언제나 봐도 재밌는 영화	10		0
1435	이건 뭐 개스토리인가 이런 쓰레기영화를 만드니까 판타지물이 발전을못하는거야 이건 그...	10		0
1436	백점만점에 백점	10		0
1437	진짜 아무리봐도 진짜 재미있어	10		0

일괄적으로 긍정/부정을 나누는 것은 신뢰를 훼손함
또한 정확도가 높아지지도 않음



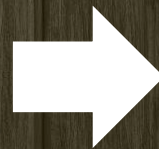
해리포터 시리즈의 리뷰를 이용한 효율적인 분석은 어렵다.

4. 결론



4. 결론

1. 긍정 / 부정 분류의 낮은 정확도 문제
2. 영화 별 다른 리뷰 개수로 신뢰도 하락
→ 추후 추가적인 데이터로 정확도 향상 필요
3. 제한적인 컴퓨터 성능으로 인하여 작업능력 하락
4. 긍정/부정의 압도적인 차이로 정확도 하락
5. 의미 있는 데이터 추출의 어려움



갈 길이 멀다
열심히 하자



Thank You