

텍스트 기반 리뷰 분석을 통한 선호 영화 소재 탐색

F조

1811111 조민우(조장)
2171277 송지윤
2171550 조서진
1871375 어현우

목차 a table of contents

- 1 연구 배경
- 2 연구 목적
- 3 활용 데이터
- 4 연구 과정
- 5 연구 결과
- 6 결과 해석
- 7 결론 및 한계



연구 배경



관객이 떠난다: 한국 영화 매출·관람 급감

팬데믹 이후 회복을 기대했던 한국 영화는 오히려 침체기에 접어들었다. 2025년 3월 기준, 매출과 관객 수는 3개월 연속 하락 중이다. OTT 성장과 콘텐츠 다양성 부족 등이 맞물리며 관객의 이탈이 가속화되고 있다. 이는 산업의 구조적 위기와 지속 가능성에 대한 경고로 볼 수 있다.



세계도 외면: 국제 영화제 초청 0편

2025년 칸 영화제에서 한국 영화는 단 한 편도 초청되지 않았다. 이는 일시적 결과가 아닌, 국제적 평가에서의 신뢰 상실을 의미한다. '기생충' 이후의 기대감은 사라졌고, 작품성이 외면받는 현실이 드러났다. 한국 영화는 대중뿐 아니라 세계 영화계에서도 관심 밖으로 밀려나고 있다.

"영화산업붕괴수준..2026년,더심각할것"[종합]

영화진흥위원회, '한국영화 활력충전 토크콘서트' 개최
영화인들이 본 한국영화의 위기와 극복 방안

"3000억 규모의 공적 자금 필요"

"홀드백 제도 법제화 해야"



"불편한 진실을 말하자면, 영화 산업이 거의 붕괴됐습니다."

영화 '명량'의 김한민 감독은 팬데믹 이후 한국 영화의 현주소에 대해 이같이 진단했다.

- 2024년 기준, 극장 시장은 팬데믹 이전의 60% 수준만 회복
- 투자 위축 → 제작 편수 감소 → 산업 구조 붕괴 위기
- OTT 확산과 소비 패턴 변화로 기존 유통 구조 한계
- 도달산업 내부에서도 붕괴 수준이라는 평가 등장

따라서 본 연구는 텍스트 기반 리뷰 분석을 통해 관객과 평단의 긍정 평가 요소를 파악하고, 대중성과 작품성이 교차하는 표현·감성적 공통점을 도출하고자 함

1

대중과 평론가의 평가 기준 차이를 규명

2

양측 모두에게 긍정 평가를 받은 리뷰의 공통적 특성 도출

3

대중과 평론가 모두에게 긍정적으로 받아들여진 키워드·주제 분석

연구 목적

본 프로젝트와 기존 접근 방식과의 차이점

기존 연구의 한계

- 대부분의 기존 연구는 영화소비자전체를 하나의 집단으로 보고 분석함
- 설문조사나 통계 수치에 기반한 결과 해석이 중심
- 세부 평가 집단 간의 의미 차이까지는 충분히 반영되지 않음

>>

본 프로젝트의 차별점

- 대중 리뷰와 평론가 리뷰를 명확히 분리하여 분석
- 긍정 리뷰를 중심으로 자주 등장하는 어휘와 주제를 비교 분석
- 텍스트 기반 분석 기법(LDA 등)을 활용하여 차이점 도출

연구 목적

최종 목표

A

리뷰에 사용된 단어와 표현을 분석해, 긍정적인 평가를 받은 글에서 자주 등장하는 어휘의 특징을 찾는다. 관객과 평론가 리뷰에 공통적으로 나타나는 언어적 특성에 주목한다.

B

많은 사람들에게 긍정적인 평가를 받은 리뷰에서 공통적으로 언급된 주제나 관심 요소를 정리하고, 흥행성과 작품성을 모두 고려한 콘텐츠 방향성에 대한 단서를 도출하고자 한다.

C

이러한 분석 결과를 바탕으로, 한국 영화가 대중성과 작품성 사이의 균형을 이루며, 국내외 시장 모두에서 경쟁력 있게 생존할 수 있는 방향성을 제안하는 것이 본 연구의 최종 목표다.

대중 리뷰 데이터



출처: IMDB 50K Movie Reviews (Kaggle)

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

형태: 각 리뷰는 긍정/부정 라벨이 명확하게 붙어 있음

특징:

- 리뷰 길이 짧고, 감정 표현이 직설적
- “재밌다”, “지루했다”, “배우가 잘생김” 등 감각 중심 표현 다수
- 서사나 주제보다는 인상적 장면 감정 중심
- 키워드 중심 분석에 적합

대중이 무엇에 공감하고 반응하는지를 파악할 수 있는 데이터

평론가 리뷰 데이터



출처: Rotten Tomatoes Movies & Critic Reviews (Kaggle)

<https://www.kaggle.com/datasets/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset>

형태: 리뷰 + 평가 정보 (Fresh / Rotten, Top Critic 여부 등)

특징:

- 문장 구조가 복잡하고, 비유수사적 표현 다수
- 감상보다는 평가 중심 (예: 연출력, 메시지 전달력, 사회적 함의 등)
- 전반적인 완성도나 '예술성'에 집중

작품성 중심의 평가 언어를 분석하기 적합한 데이터

활용 데이터

항목	대중 리뷰 	평론가 리뷰 
출처	IMDB (Kaggle)	Rotten Tomatoes (Kaggle)
수량	50,000건	110,000건
라벨	긍정/부정	Fresh/Rotten(긍정/부정)
특징	감정 위주, 직설적	평가 위주, 문어체
분석 초점	공감, 감성 키워드	작품성, 비평 언어

연구 개요

대중/평론가 리뷰 모두 동일한 분석 방법을 적용

1단계	2단계	3단계	4단계	5단계
리뷰 텍스트 전처리	감성 분류 및 키워드 분석	LDA 토픽 모델링	LDA 시각화 및 키워드 정리	문서별 토픽 분포 및 대표 문서 추출

1. 리뷰 텍스트 전처리

```
# 전처리 함수
def preprocess_text(text):
    text_no_html = BeautifulSoup(text, "html.parser").get_text()
    expanded_text = contractions.fix(text_no_html)
    text_no_specials = re.sub(r'^a-zA-Z', ' ', expanded_text)
    text_clean = re.sub(r'\s+', ' ', text_no_specials).strip()
    text_lower = text_clean.lower()
    tokens = word_tokenize(text_lower)
    filtered_tokens = [word for word in tokens if word not in stopwords.words('english')]
    stemmer = PorterStemmer()
    stemmed_tokens = [stemmer.stem(word) for word in filtered_tokens]
    return ' '.join(stemmed_tokens)
```

처리 과정

- HTML 태그 제거 (BeautifulSoup)
- 축약어 복원 (contractions)
- 특수문자 제거 및 소문자화
- 토큰화 후 불용어 제거 (nltk)
- 어간 추출 (PorterStemmer)

- IMDB 리뷰 원문에는 HTML 태그, 특수문자, 축약어 등 다양한 불필요 요소가 존재함.

- 전처리 과정을 통해 텍스트를 분석 가능한 형태로 정제함.

- 정제된 텍스트는 cleaned_review 컬럼에 저장됨.

	review
3	Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents are fighting all the time. This movie is slower than a soap opera... and suddenl...



	cleaned_review
	basic famili littl boy jake think zombi closet parent fight time movi slower soap opera suddenli jake decid becom rambo kill zombi ok first go make film must decid thriller drama drama movi watcha...

2. 감성 분류 및 키워드 분석

```
# positive와 negative 나누기
df_pos = df_copy[df['sentiment'] == 'positive'][['sentiment', 'review', 'cleaned_review']]
df_neg = df_copy[df['sentiment'] == 'negative'][['sentiment', 'review', 'cleaned_review']]

# 명사, 형용사 필터링 함수
def filter_pos_adj(tokens):
    tagged = nltk.pos_tag(tokens)
    filtered = [word for word, pos in tagged if pos in ACCEPTED_POS]
    return filtered
```

처리 과정

- 긍정/부정 리뷰 분리
- 명사/형용사 필터링

- 긍정/부정 리뷰 분리

- sentiment 컬럼을 기준으로 positive와 negative 리뷰를 나눔.
- 이후 각 그룹을 별도로 처리하여 긍정/부정 리뷰의 특징을 파악.

- 명사/형용사 필터링

- 명사와 형용사 중심으로 필터링함.
- 감정 표현이나 핵심 주제를 담고 있는 중요 단어만 추출.

2. 감성 분류 및 키워드 분석

```
# 단어 빈도 계산 함수
def get_filtered_word_freq(text_series):
    # 텍스트 합치고 토큰화
    tokens = ' '.join(text_series).split()
    # 품사 필터링
    filtered_tokens = filter_pos_adj(tokens)
    # 단어 빈도 계산
    word_counts = Counter(filtered_tokens)
    # DataFrame 변환 후 빈도 내림차순 정렬
    df_freq = pd.DataFrame(word_counts.items(), columns=['Word', 'Frequency'])
    df_freq_sorted = df_freq.sort_values(by='Frequency', ascending=False)
```

처리 과정

- 단어 빈도 계산
 - 긍정/부정 그룹 각각에 대해 자주 등장하는 단어들을 카운트.
 - 상위 30개의 단어를 단어-빈도 테이블로 정리.

Positive reviews top keywords:

	Word	Frequency
107	film	50878
214	movi	40304
111	time	16624
258	good	15142
167	charact	13740
135	great	13124
277	stori	11146
2	watch	11127
9	scene	9839
13	show	8931
154	way	8914
423	realli	8400
778	end	8353
137	life	8147
217	peopl	7958
182	year	7819
143	play	7702
510	much	7292
252	best	7290
274	work	7258
6	thing	7038
352	man	6817
257	act	6599
...		

Negative reviews top keywords:

	Word	Frequency
10	movi	52484
19	film	44975
9	time	15329
57	bad	14778
284	good	14545
131	charact	14020
30	watch	12776
141	scene	11115
92	act	10132
244	realli	9771
45	thing	9470
105	end	8795
183	stori	8765
200	peopl	8587
170	plot	8509
88	way	8456
100	look	8051
166	much	7633
438	actor	7154
37	show	7111
136	tri	6598
122	actual	5938
190	work	5818
...		

- 공통 키워드 :

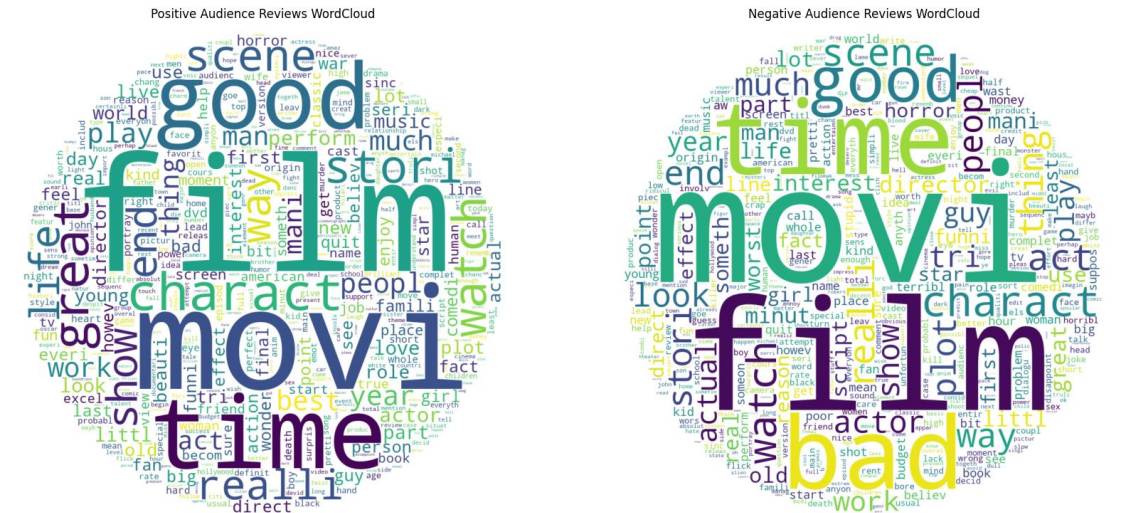
movi, film, charact, good, watch, scene, peopl, act, end, show

2. 감성 분류 및 키워드 분석

```
# 워드클라우드 생성 함수
def create_wordcloud(word_counts, mask):
    return WordCloud(
        width=1000,
        height=1000,
        max_words=500,
        mask=mask,
        background_color='white'
    ).generate_from_frequencies(word_counts)
```

처리 과정

- 워드클라우드 시각화
- 추출된 단어 빈도 데이터를 바탕으로 워드클라우드 생성



3. LDA 토픽 모델링

```

from gensim.models.ldamodel import LdaModel
from gensim.models.coherencemodel import CoherenceModel
import matplotlib.pyplot as plt

# 최적 토픽 수 찾기
def evaluate_lda_model(tokenized_doc, dictionary, corpus, label):
    coherence_values = []
    perplexity_values = []

    for num_topics in range(2, 13, 2):
        model = LdaModel(corpus=corpus,
                        num_topics=num_topics,
                        id2word=dictionary,
                        passes=4,
                        alpha='symmetric',
                        eta='auto',
                        random_state=100)

        coherence_model = CoherenceModel(model=model,
                                        texts=tokenized_doc,
                                        dictionary=dictionary,
                                        coherence='c_v')

        coherence = coherence_model.get_coherence()
        log_perplexity = model.log_perplexity(corpus)
        perplexity = 2 ** (-log_perplexity)

        coherence_values.append(coherence)
        perplexity_values.append(perplexity)

    print(f"[{label}] num_topics={num_topics} | Coherence={coherence:.4f} | Perplexity={perplexity:.2f}")

    return {
        "coherence": coherence_values,
        "perplexity": perplexity_values
    }

```

- 리뷰에서 반복적으로 나타나는 주제를 추출하기 위해 LDA 모델을 적용함

처리 과정

- 단어 사전 및 BoW 벡터 생성
- LdaModel로 2~12개 토픽 학습
- 각 토픽 수마다 Coherence, Perplexity 계산
- 최적 토픽 수 결정
 - 긍정 리뷰: 8개
 - 부정 리뷰: 8개

[POSITIVE]	num_topics=2	Coherence=0.3136	Perplexity=286.08
[POSITIVE]	num_topics=4	Coherence=0.3165	Perplexity=284.26
[POSITIVE]	num_topics=6	Coherence=0.3348	Perplexity=293.60
[POSITIVE]	num_topics=8	Coherence=0.3541	Perplexity=315.56
[POSITIVE]	num_topics=10	Coherence=0.3637	Perplexity=363.72
[POSITIVE]	num_topics=12	Coherence=0.3970	Perplexity=416.83

[NEGATIVE]	num_topics=2	Coherence=0.2946	Perplexity=264.10
[NEGATIVE]	num_topics=4	Coherence=0.2952	Perplexity=268.27
[NEGATIVE]	num_topics=6	Coherence=0.3121	Perplexity=279.33
[NEGATIVE]	num_topics=8	Coherence=0.3409	Perplexity=301.94
[NEGATIVE]	num_topics=10	Coherence=0.3923	Perplexity=352.13
[NEGATIVE]	num_topics=12	Coherence=0.3796	Perplexity=396.06

3. LDA 토픽 모델링

```
# 시각화 함수
def plot_lda_metrics(result_dict, label=""):
    coherence = result_dict["coherence"]
    perplexity = result_dict["perplexity"]
    topics = range(2, 13, 2)
    fig, axes = plt.subplots(1, 2, figsize=(14, 5)) # 1행 2열

    # Coherence 그래프
    axes[0].plot(topics, coherence, marker='o', color='blue')
    axes[0].set_title(f'Coherence Score - {label}')
    axes[0].set_xlabel('Number of Topics')
    axes[0].set_ylabel('Coherence')
    axes[0].grid(True)

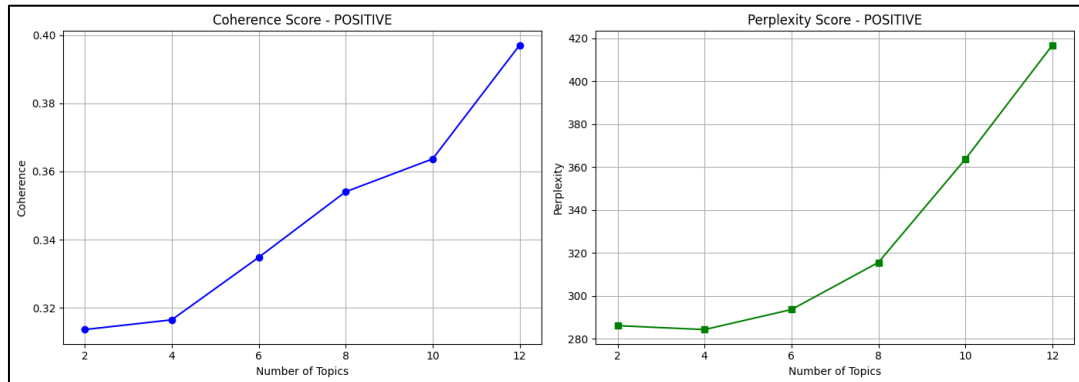
    # Perplexity 그래프
    axes[1].plot(topics, perplexity, marker='s', color='green')
    axes[1].set_title(f'Perplexity Score - {label}')
    axes[1].set_xlabel('Number of Topics')
    axes[1].set_ylabel('Perplexity')
    axes[1].grid(True)

    plt.tight_layout()
    plt.show()
```

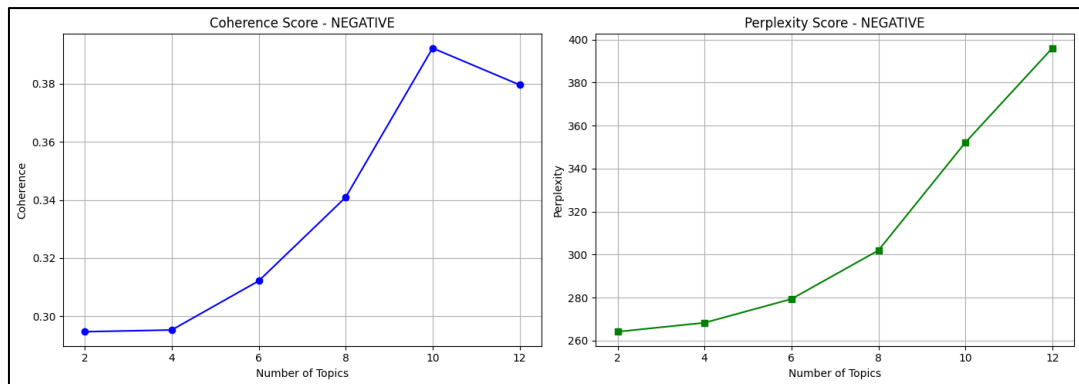
LDA 토픽 모델 시각화

- Coherence: 토픽의 일관성을 나타냄 (높을수록 좋음)
- Perplexity: 모델의 예측 혼란도 (낮을수록 좋음)
- 시각화를 통해 각 토픽 수의 성능을 직관적으로 비교

긍정



부정

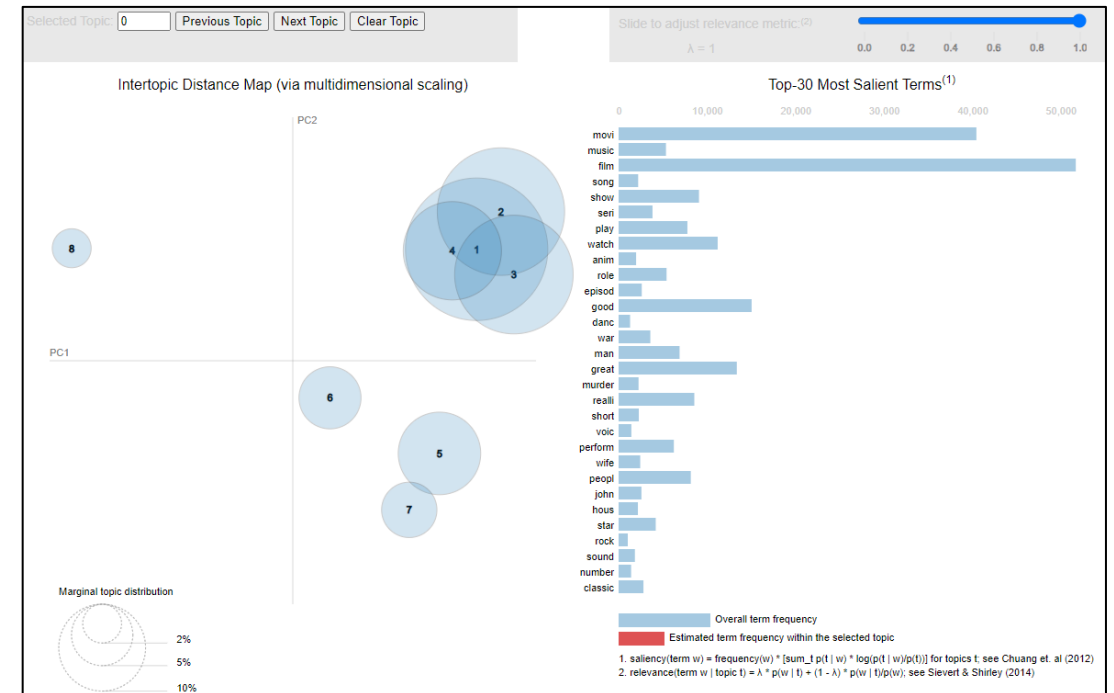


4. LDA 시각화 및 키워드 정리

```
# 최종 토픽 수 지정
pos_num_topics = 8
neg_num_topics = 8 # 10개->8개 (결과에 중복 키워드 이슈로 줄임)

# LDA 모델 생성
pos_ldamodel = LdaModel(pos_corpus, num_topics=pos_num_topics, id2word=pos_dictionary, passes=4, alpha='sym
neg_ldamodel = LdaModel(neg_corpus, num_topics=neg_num_topics, id2word=neg_dictionary, passes=4, alpha='sym
```

- 각 토픽은 원으로 나타나며, 원의 크기는 해당 토픽의 전체 문서에서의 비중(중요도)을 의미.
- 원 사이의 거리는 토픽 간의 주제적 유사도를 나타냄.
- 가까울수록 유사한 토픽, 멀수록 구분된 주제
- 오른쪽 바 그래프에서는 선택한 토픽의 주요 키워드 분포를 확인할 수 있음.



4. LDA 시각화 및 키워드 정리

```
# 토픽별 키워드와 가중치를 데이터 프레임으로
pos_topics_df = pd.DataFrame()

for i, topic in enumerate(pos_ldamodel.show_topics(formatted=False, num_words=40)):
    topic_keywords = " ".join([word for word, prop in topic[1]])
    pos_topics_df = pd.concat([pos_topics_df, pd.DataFrame({'Topic_Num': "Topic " + str(i+1),
                                                            'Keywords': topic_keywords}, index=[0])], ignore_index=True)

print(pos_topics_df)
```

처리 과정

- 긍정/부정 리뷰에 대해 LDA를 활용하여 8개의 주요 토픽 추출
- 각 토픽별로 빈도 기반 상위 키워드 40개를 나열하여 주제 성격 파악
- 키워드는 리뷰 내 등장 확률 기준으로 자동 추출됨

긍정

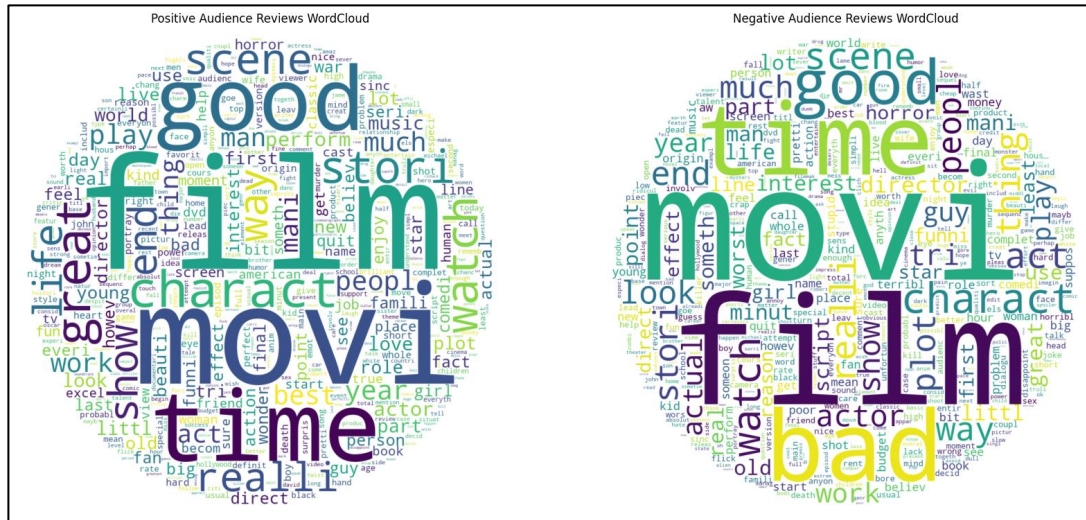
```
0 show, seri, episod, time, charact, tv, good, season, thing, anim, watch, year, new, way
1 film, time, charact, scene, movi, great, perform, work, stori, director, good, actor, b
2 film, life, war, world, stori, peopl, charact, way, man, human, live, time, end, power,
3 play, role, star, john, film, year, great, perform, georg, man, girl, best, time, bond,
4 man, play, murder, wife, role, perform, good, robert, son, father, polic, woman, young,
5 hous, drug, jane, scene, street, joe, doctor, horror, man, woman, hospit, new, murder,
6 movi, good, watch, great, time, realli, film, charact, stori, peopl, end, thing, way, l
7 music, song, danc, anim, disney, voic, cartoon, rock, band, cat, number, record, singer
```

부정

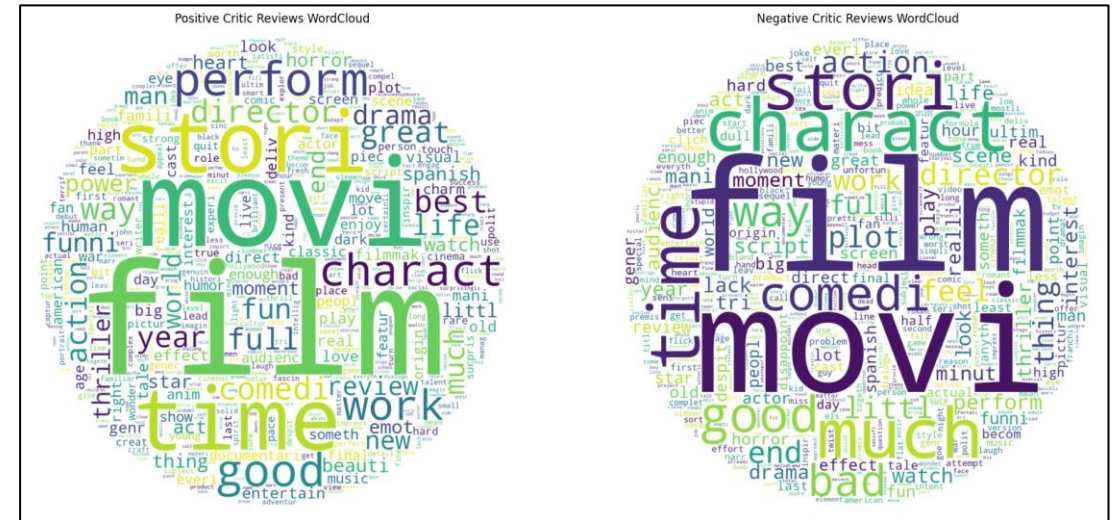
```
0 show, episod, seri, tv, watch, charact, kid, peopl, season, cartoon, old, televis, time, y
1 movi, film, bad, watch, good, time, act, realli, scene, thing, plot, charact, end, peopl,
2 film, charact, stori, time, interest, scene, work, much, director, end, point, littl, way,
3 girl, man, woman, hous, famili, old, end, young, wife, mother, daughter, murder, kill, yea
4 martial, ninja, art, fu, chan, keaton, kung, stan, hardi, cat, myer, wayn, dougla, laurel,
5 funni, comedi, movi, film, joke, origin, sequel, humor, good, time, star, laugh, game, fun
6 movi, book, film, scene, time, charact, good, play, stori, actor, look, action, version, r
7 film, movi, music, charact, actor, women, director, play, song, peopl, sex, gay, good, sto
```


키워드 분표 비교

대중



평론가



키워드 분표 비교

대중

Positive reviews top keywords:

	Word	Frequency
107	film	50878
214	movi	40304
111	time	16624
258	good	15142
167	charact	13740
135	great	13124
277	stori	11146
2	watch	11127
9	scene	9839
13	show	8931
154	way	8914
423	realli	8400

Negative reviews top keywords:

	Word	Frequency
10	movi	52484
19	film	44975
9	time	15329
57	bad	14778
284	good	14545
131	charact	14020
30	watch	12776
141	scene	11115
92	act	10132
244	realli	9771
45	thing	9470
105	end	8795

평론가

Positive reviews top keywords:

	Word	Frequency
164	film	139043
121	movi	82040
407	time	35318
254	stori	34127
268	perform	33637
57	good	32479
208	charact	30533
129	work	27607
228	director	25192
288	way	24909
543	comedi	24180
95	best	24129

Negative reviews top keywords:

	Word	Frequency
239	film	62647
12	movi	57200
98	charact	19304
165	time	18134
180	much	17405
382	stori	16532
275	good	16014
477	comedi	14675
27	bad	12008
208	way	11877
294	director	11498
54	end	11467

1. 공통점

양쪽 모두 film, movie, character, story, time, good, watch, scene, great 등의 단어가 핵심 키워드로 자주 등장함
 → 기본적인 영화 평가 요소(스토리, 연기, 시간, 연출 등)가 공통적으로 중요하게 다뤄짐

키워드 분표 비교

대중

긍정 키워드 : film, movi, time, good, charact, great, stori, watch, scene, show, way, realli, end, life, peopl, year, play, much, best, work, thing, man, act

부정 키워드 : movi, film, time, bad, good, charact, watch, scene, act, realli, thing, end, stori, peopl, boring, plot, way, look, much, actor, show, tri, actual, work

평론가

긍정 키워드 : film, movi, time, stori, perform, good, charact, work, director, way, comedi, best, great, full, year, life, action, fun, much, drama, funni, world, new

부정 키워드 : film, movi, charact, time, much, stori, good, comedi, bad, way, director, end, littl, plot, action, feel, thing, work, full, perform, script, interest, watch

2. 차이점

[대중 리뷰 키워드 특징]

- 긍정: watch, scene, show, play, life 등 감정 중심의 체험 표현이 많고, 재미·공감 위주의 반응이 두드러짐
- 부정: boring, actual, tried 등 직설적이고 주관적인 불만 표현이 주를 이룸

[평론가 리뷰 키워드 특징]

- 긍정: performance, director, drama 등 영화의 구성 요소나 연출에 대한 평가가 중심이며, world, life, new 등 주제적·문화적 관점도 일부 반영됨
- 부정: script, plot, feel, interest, littl 등 이야기 구조, 감정선 부족 등 비평적 관점에서의 문제 제기가 많음

키워드 분표 비교

구분	대중 리뷰 키워드 특징	평론가 리뷰 키워드 특징
긍정	watch, scene, show, play, life 등 → 감정 중심, 재미·공감 위주	performance, director, script, drama 등 → 구성 요소 중심, 작품성 평가 강조
부정	boring, try, actual, tri 등 → 직설적 불만, 주관적 반응	script, plot, feel, interest, littl 등 → 구조적 문제 비판, 감정선 부족 지적

- 대중 리뷰는 감정과 체험 중심으로, 재미나 지루함 같은 즉각적인 반응을 표현한다.
- 반면 평론가 리뷰는 연출, 구성 등 영화의 기술적 요소를 분석적으로 평가한다.
- 같은 작품도 대중과 평론가는 서로 다른 기준과 시선으로 바라본다.

LDA 분석 (대중 리뷰)

긍정

1. 프로레슬링 / 스포츠 경기

- 키워드: royal rumble, match, undertaker, batista, cena, jbl, championship
- 평가: 다양한 경기 구성, 극적 연출, 몰입감 높은 전개

2. 무협 판타지 영화

- 키워드: martial art, wirework, fox spirit, taoist, enchant, fantasy
- 평가: 몽환적인 분위기, 시각적 화려함, 감성적 연출

3. 전쟁 영화

- 키워드: world war, soviet, nazi, symbolic, black and white
- 평가: 역사적 상징성, 묵직한 주제의식, 내면 연기 강조

4. 감성 드라마 / 로맨스

- 키워드: romantic, cry, love, heart, life, robert downey jr
- 평가: 감정선 중심의 이야기, 따뜻한 메시지, 감동 유발

5. 클래식 뮤지컬 / 댄스 영화

- 키워드: gene kelly, dance, sinatra, hollywood, music, piano
- 평가: 경쾌한 분위기, 퍼포먼스 중심, 음악·댄스의 시너지

부정

1. 아동용 콘텐츠의 부적절성

- 키워드: power ranger, toy, kid show, violent
- 비판: 과도한 상업성, 선정성, 유치한 연출

2. 저예산 외계 공포 영화

- 키워드: alien, infection, fx, bad dialog
- 비판: 조악한 특수효과, 설정 혼란, 연기력 부족

3. 디스토피아 철학 드라마

- 키워드: dystopia, loneliness, society, betrayal
- 비판: 무거운 분위기, 전달력 부족, 몰입 실패

4. 실화 리메이크 / 정치 느와르

- 키워드: remake, death, cinematography / godfather, power, helpless
- 비판: 원작 감동 상실, 깊이 부족, 명성 의존 연출

LDA 분석 (평론가 리뷰)

긍정

1. 가족 애니메이션 / 크리스마스 영화

- 키워드: christmas, family, fun, kid, adult
- 평가: 가족 친화적인 따뜻한 분위기와 감성적 메시지로 호평

2. 사회 비판 드라마

- 키워드: seoul station, narrative, powerful, visual
- 평가: 사회 문제를 강렬한 연출로 효과적으로 전달

3. 고전 괴물 / 호러 영화

- 키워드: classic, horror, monster, night
- 평가: 고전적 공포 연출과 긴장감 유지로 완성도 인정

4. 인생 / 현실 드라마

- 키워드: story, truth, reality, life
- 평가: 삶의 진실성과 감정을 잔잔하게 표현

5. 감성 예술 영화

- 키워드: sensibilidade, interessante, cinema
- 평가: 예술적 구성과 감정선이 섬세하게 전개됨

부정

1. 허술한 플롯의 드라마

- 키워드: plot, message, drama, melodrama
- 비판: 전개의 인위성, 공감 실패, 감정 과잉

2. 무미건조한 리메이크

- 키워드: remake, humorless, elit, filthi, rich, women
- 비판: 해석 부재, 유머 없음, 불친절한 연출

3. 난해한 외국 영화

- 키워드: system, scenario, biography, learn
- 비판: 맥락 부족, 몰입 저해, 메시지 불분명

4. 저예산 공포 영화

- 키워드: terror, profan, scare, origin
- 비판: 잔혹성 과잉, 구조 미흡, 자극에 의존

긍정 평가 분석 요약

대중의 관점

흥미로운 전개

→ 지루하지 않고 몰입감 있는 흐름 선호

배우의 매력적인 연기

→ 캐릭터에 감정을 이입할 수 있는 연기력

감정을 자극하는 따뜻한 이야기

→ 웃음과 눈물이 공존하는 감성 중심

서사시각적으로 완성도 높은 연출

→ 미장센, 장면 구성 등에서의 만족감

평론가의 관점

철학적 서사구조

→ 이야기 전개에 담긴 사유와 의미 중시

사회적 메시지 전달력

→ 현실 문제에 대한 통찰력과 비판 의식

예술적 연출과 감정선의 섬세함

→ 표현 기법의 창의성과 감정 묘사의 깊이

>>

양측 공감 장르

- 가족 영화
- 감성 예술 영화
- 인디 성장 드라마

→ 감정적 몰입과 주제의식이
모두 살아있는 작품들

부정 평가 분석 요약

공통된 부정적 요소

감정 몰입 실패

→ 캐릭터나 이야기 흐름에 이입이 어려움

스토리/연출 완성도 부족

→ 개연성, 구성력, 연출력에서의 허술함

자극성 혹은 지루함

→ 과한 폭력/고어 또는 전개가 너무 느림

주요 비판 장르

리메이크

→ 원작의 감동을 살리지 못하고 새로움 부족

저예산 공포

→ 시각 효과 부실, 전개의 반복성과 단조로움

외국 영화(난해한 설정)

→ 현지 정서와 문맥 이해의 어려움, 불친절한 서사

텍스트 기반 감정 분석의 한계

주요 한계 요약

이진 분류(긍/부정)의 한계

→ 복합적이고 미묘한 감정(예: 슬픔+감동)을 포착하기 어려움

도메인 차이 문제

→ 대중 리뷰 기반 모델은 평론가의 언어적 스타일을 완전히 반영하지 못함

→ 평론가 감정 예측의 정확도에 한계 발생

모델 해석력 부족

→ 사용한 랜덤포레스트 모델은 '왜 그렇게 분류했는가'를 설명하기 어려움

시사점

- 감정 분류를 보다 정교하고 다층적으로 확장할 필요
- 대중 리뷰와 평론가 리뷰의 언어 차이를 반영한 별도 모델 고려
- 해석 가능한 딥러닝 기법 도입 필요성

결론 및 한계

한국 영화의 방향성 제안

1

감정과 서사 모두에 공감할 수 있는 시나리오 개발

2

배우 중심의 드라마 구조와 캐릭터 감정선 강화

3

연출의 예술적 완성도와 시각적 퀄리티 확보

4

대중과 평론가가 모두 공감 가능한 주제 선택

참고 자료 및 출처

<데이터>

출처: IMDB 50K Movie Reviews (Kaggle) – 대중 리뷰

<https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

출처: Rotten Tomatoes Movies & Critic Reviews (Kaggle) – 평론가 리뷰

<https://www.kaggle.com/datasets/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset>

<뉴스>

출처: [3월 영화산업 결산] 침체기 계속... 한국 영화 매출액·관객수 3개월 연속 감소

<https://www.sisaweek.com/news/articleView.html?idxno=224399>

출처: "0편" 올해 칸영화제 한국 영화 공식 초청 불발...위기감↑

<https://www.fnnews.com/news/202504110841419817>

"영화 산업 붕괴 수준...2026년, 더 심각할 것" [종합]

<https://www.hankyung.com/article/202411064057H>

<PPT 템플릿>

<https://blog.naver.com/seiru523/223800888213>