

Movie for Mood

리뷰 키워드 기반 영화 추천 시스템

1. 기획 배경

2. 데이터 소개

3. 추천 시스템

4. 개선방안

단어 하나만으로 영화를 추천해 줄 수 있을까?



데이터 수집

영화 목록 : 영화진흥위원회 DB 2000년 ~ 2021년 영화 데이터 수집. (총 7633개)

크롤링 방법: Selenium / BeautifulSoup

출처:

- 네이버 블로그
- 네이버 영화 리뷰

데이터

	타이틀	리뷰
0	미나리	<p>(스포일러)</p> <p>선댄스 영화제서 공개된 이후로 오스카 레이스가 한창인 지금까지 북미 현지에서 만장일치에 가까운 극찬을 받고 있는 영화, 정이삭 감독의 신작 <미나리>를 지난 부산 국제 영화제서 미리 볼 기회가 있었다. 개인적으로 이 영화를 기다렸던 이유는 단 하나다. 시놉시스만 대충 훑어봐도 기승전결의 구조가 훤히 들여다보일 정도로 보편적인 것 같아 보이는 이 드라마 서사가 과연 무엇을 성취하였는지 궁금했기 때문이다. 80년대, 미국, 한국 이민자. 그 안에서 아버지와 어머니의 갈등, 할머니와 손주와의 갈등. 영화를 어느 정도 봐온 이라면 이 키워드만 들어도 대충 각자의 머릿속에서 4, 5개 즈음의 시나리오가 절로 상기될 법 하다.</p> <p>우선, <미나리>는 다행히도 내 머릿속의 시뮬레이션 중 최악의 각본을 피해 간 영화다. 나는 영화를 보기 전 배우 윤여정이 연기한 할머니 순자 캐릭터가 영 거슬렸다. 나는 부디 영화가 할머니의 죽음을 통한 가족의 성장이란 진부한 귀결을 택하지 않았기를 바랬다. 굳이 다른 국적의 영화에서까지 눈물로 축축해진 가족 드라마를 보기 싫었던 탓이다. 허나 나의 이러한 선부른 걱정과는 달리, 영화는 순자를 이용해 극중 인물과 관객의 눈물샘을 자극하는 안일함을 보이지 않으며 각본이 험거워지는 부분을 그녀의 유머만으로 벌충하려는 나태한 태도를 보이지도 않는다. 물론 순자의 유머러스한 태도는 퍽퍽해질 무렵의 영화를 거둬하여 환기하는 익숙한 장치로 사용되며 아울러 후반부에 그녀가 뇌졸중을 앓는다는 설정은 이 영화가 자칫 뻘한 노선을 택할지도 모를 전조처럼 보이기도 한다.</p> <p>허나 순자는 비단 그러한 기계적인 방식으로 소비되고 마는 캐릭터가 아니다. 두 부부간의 다툼을 중재하는 것, 데이빗과 앤의 성장, 그 사이에서 겪는 판이한 문화 차이, 모니카의 희망과 불안, 깨졌던 가족의 믿음을 재확인시켜주는 것. 이 모든 중심에 순자가 있으며 영화는 캐릭터의 이러한 포지셔닝을 위해 '전형적인 할머니 상'이라는 해묵은 성질을 빌려오는데 그치지 않고 스스로 만들어낸 순자라는 캐릭터 고유의 성향을 효과적으로 이용한다. 해외 평단의 반응 중 순자 캐릭터에게 유독 호평이 자자했던 까닭은 비단 그들에게 아시아계 할머니 캐릭터 상이 독특하게 느껴졌기 때문만이 아니었을 것이다. 전형적인 할머니상과 윤여정 배우 특유의 비정형화적 상이 부딪혔을 때 느껴지는 묘한 이질감, 그리고 그 이질감이 빚어내는 강력한 존재감. 그 느낌 하나로 윤여정은 극을 자신의 것으로 사로잡는다. 배우의 훌륭한 역량과 캐릭터를 조형하는 각본의 사려 깊은 공력이 더해진 순자라는 캐릭터는 <미나리>의 최대 장점 중의 하나가 틀림없다...</p>
1	해리포터와 비밀의 방	해리포터 시리즈의 두번째 영화인 비밀의 방에는 SNS에서 종종 쓰이는 "도비는 이제..."
2	클로젯	<p>영화 클로젯 스포 가득입니다. 안보신 분들은 줄거리 내용은 피해 주시기 바랍니다. 또한 영화 클로젯의 쿠키영상은 없어요. 쿠키영상 기다리지 마시고 엉덩이 때셔도 됩니다....11941531</p> <p>영화 클로젯 후기빼구이 열려 있는 옷장은 무한한 상상력을 우리에게 선사한다. 그런 상상력을 자극하는 영화가 나왔다. 영화클로젯은 실상 우리 영화에서보다 외국 영화에서 많이 등장하는 소품으로 대표적으로 나니아 연대기에서 다른 차원으로 들어 가는 문이 되어 주었다. 기억이 나지는 않지만 외국 영화에서 아이들 벌주는 용도로 옷장이나 좁은 공간에 가두는 영화도 심심치 않게 볼 수 있었다. 한국 영화 장화홍련에서도 옷장이 등장 했었던 걸로 기억한다. 공포 영화에서 옷장문이 천천히 빼그덕 열리면 공포감은 배가 된다. 아마도 감독님은 그런 공포 요소를 잘 알고 계신듯하다. 그래서 옷장속 이야기를 만들지 않았나 생각한다. 그렇지만 클리셰가 강해 예측한 부분을 뛰어 넘는 그런 부분이 1도 없이 연출을 하다보니 전형적인 공포 영화가 되어 버려 조금 실망스러운 건 사실이다. 영화 감상에서 중의적 표현을 좋아하고 떡밥 얻어 먹는 걸 좋아하는 깨비네의 취향에 문제가 있을 수도 있지만 그래도 그렇지 너무 대놓고 말하니 그 슬프고 아팠던 아이들에게 감정이입할 부분이 부족해 아쉬웠다. 아이들이 아픈 영화 볼때면 눈물 콧물 범벅이 기본인 깨비네가 멀쩡하게 나온 걸 보면 아무래도 클로젯은 이런쪽 연출은 영 아닌 듯하다. 강 공포영화로만 만족해야하는갑다...</p>

데이터 정제

1. 기존에 있는 한국어 불용어 사전 200개를 기본으로 추가적으로 불용어를 추출해서 불용어 사전 제작. (칙오, 격인, 내뿜는, 절로, 드드리 etc)
2. 한글자 단어는 다 삭제하고, 2번 미만으로 등장한 단어들을 불용어 사전에 추가.
3. Okt(Open Korean Text), 트위터에서 만든 오픈소스 한국어 처리기로 형태소 분석(Pos Tagger). 명사, 동사, 형용사 평선을 설정해 호출. 잘게 쪼개지지만 시간이 오래 걸림.

코사인 유사도 기반 추천 시스템

사용자가 선택한 영화를 리뷰 키워드 기반으로 비슷한 리뷰 키워드가 있는 영화 목록 10개 추천.

```
m_recommendation('찰리와 초콜릿 공장')
```

	title	review
575	알라딘	하나 만점 디즈니 공주 스미다 예쁘다 알라딘 주중 외모 일등 누가 영원하다 왕자 민...
2255	미스 페레그린과 이상한 아이들의 집	예바 그린 비중 없다 어리다 보기 잔인하다 성인 유치하다 체적 난국 이다 독특하다 ...
763	드래곤 길들이기 3	아이 성장하다 드래곤 전설 되다 순간 하다 복하다 세상 받아들이다 준비 돼다 보다 ...
3422	메이즈 러너	생각 놀란 민 비중 보급 상자 쓰다 왁스 들다 미로 설계 필요 이상 들이다 나머지 ...
3895	헝거게임: 캐칭 파이어	누가 편이 없다 하다 대박 난방 고장 나서다 추다 상영 덜덜 떨다 보다 기다 러닝 ...
6817	나비효과	나비 감당 하다 거대하다 태풍 인생 선택 끝없다 연쇄 부딪다 사건 고리 그제 바로 ...
6059	어거스트 러쉬	감동 하다 최소한 개연 감 생략 감독 기백 돋보이다 나다 기타 배우다 알다 프레디 ...
6004	원스어폰어타임	참하다 가십 세계 박제 돼다 하다 인물 영 대한 사랑 넘치다 영인 자리 옮기다 관람...
5550	업	추억 물건 아니다 마음 들다 오프닝 시퀀스 보다 눈물 펄펄 쏟아 만들다이순재 아저씨...
5203	토이 스토리 3	이별 순간 오다 하다 누군가 마음 변질 되어다 때문 아니다 어떻게 사이 시간 러가다...

개선방안

- 단어 기반 추천 시스템 (Word2Vec) 완성 예정.
- 영화 설명이 제공되지 않아서 **정보**를 알기 어려움.
→ 탑3 키워드로 줄여서 출력할 수 있게 설정.
- 유사도 정도를 보여줄 **시각화 자료**가 없어서 구조를 알기 어려움.
→ 사이트에 구현한다면 주요 단어를 시각화해 같이 보여줄 수 있도록 시도.
- 리뷰 키워드 안에서도 우선순위를 보기 어려움.
→ 유사도 순서대로 **랭킹**을 보여주는 칼럼 추가.

```
[159] model_result1 = model.wv.most_similar("안타깝다")
      model_result1

[('슬프다', 0.6412979960441589),
 ('아프다', 0.5750150680541992),
 ('실', 0.5609993934631348),
 ('편하다', 0.5415959358215332),
 ('서글프다', 0.5340200662612915),
 ('아쉬움', 0.5307308435440063),
 ('달다', 0.5171127319335938),
 ('고생', 0.51429682970047),
 ('야하다', 0.5023354291915894),
 ('마음', 0.49909502267837524)]
```


감사합니다.