

넷플릭스 데이터를 통한 영상(영화, 드라마 등) 추천 서비스

- '콘텐츠 기반 필터링' 을 활용하여 -

1

서비스 목표

2

데이터 및 추천 알고리즘

3

전처리

4

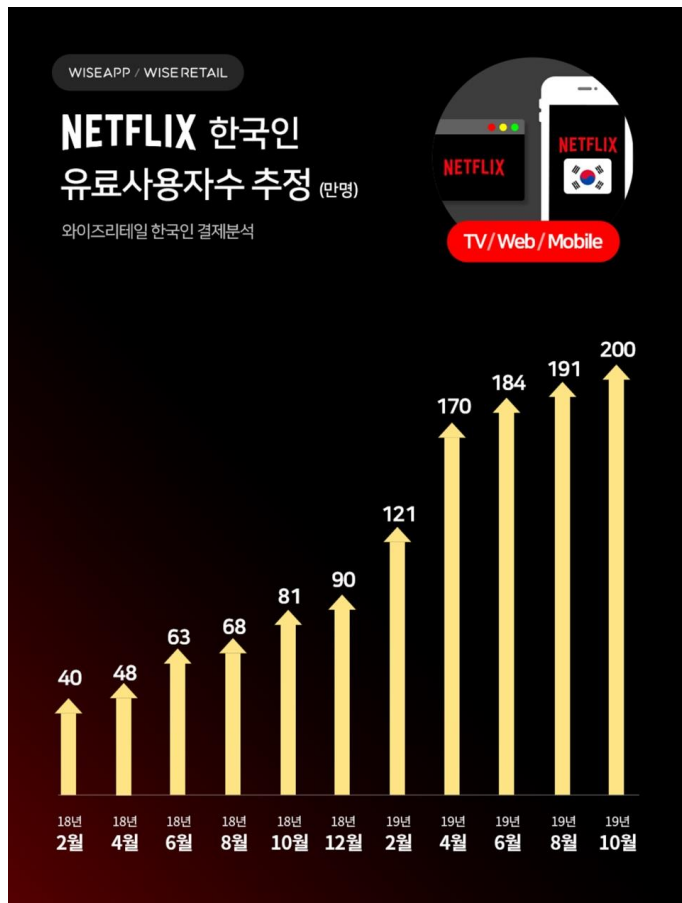
모델 구축

5

최종 결과

1

서비스 목표

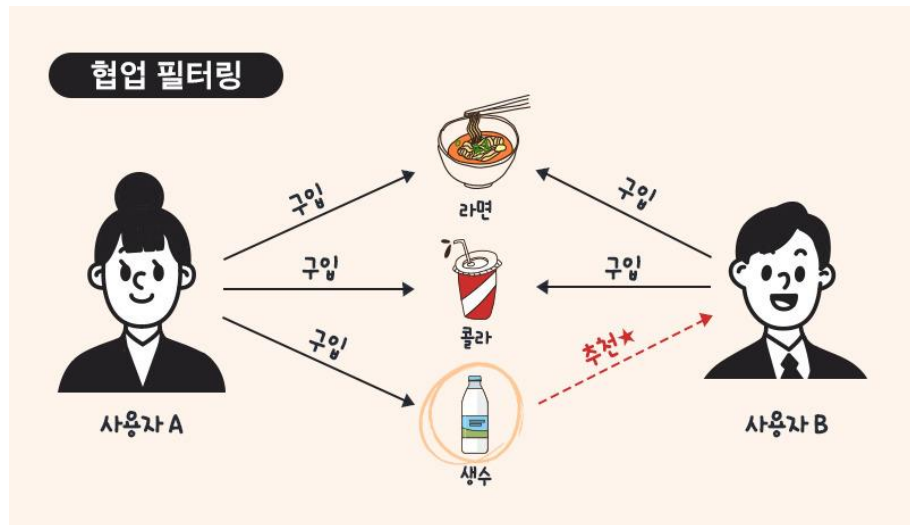


- 대다수의 사람들이 영상 콘텐츠를 접하는 방법은?
- 점차적으로 증가하고 있는 국내 넷플릭스 사용자
- 언제, 어디서든, 쉽고, 빠르게 볼 수 있음

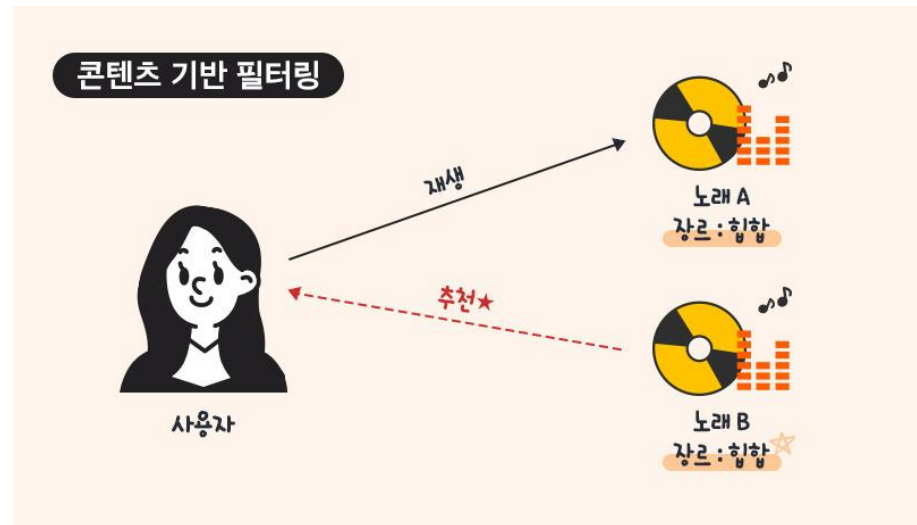


추천을 받고 바로 볼 수 있는 접근성이 편한
넷플릭스의 영상 콘텐츠를 추천

데이터 및 추천 알고리즘 _ 콘텐츠 기반 필터링



- 현재 넷플릭스에서 사용하고 있는 방법
- 비슷한 성향의 다른 사용자들의 결과를 참고
- 사용자들 간의 행동 기록이 중요




- 우리 팀에서 사용하고자 하는 방법
- 콘텐츠 항목 자체를 분석한 결과를 참고
- 각 콘텐츠의 특징들이 중요

데이터 및 추천 알고리즘 _ 콘텐츠 기반 필터링

소소한 하루 >

넷플릭스 추천 드라마 부탁 드립니다


 동천라안 성실수지맘 1:1 채팅
2020.08.25. 19:08 조회 864

요새같은 집콕에 혹시 추천해주실 정주행 드라마 영화 넷플릭스에 있을까요? ㅜㅜ
이번 주말 집콕 대비요 ㅜㅜㅎㅎ

참고로 전 키싱부스1,2, 비밀의 숲1, 정직한후보, 너무 재미있게 보았어요~
남편은 종이의 집, 기묘한 이야기 재미있다 하더라고요~

TV/영화/넷플/유튜브 >


[넷플릭스] 넷플릭스 추천좀 부탁드립니다

 햄스터 정식멤버 ★ 1:1 채팅
2020.07.25. 18:32 조회 3,687

넷플릭스 드라마 추천좀 부탁드립니다!
제가 본거는 종이의집, 기묘한 이야기 봤네요 ㅎㅎ
미스터리나 범죄, 추리물 좋아합니다!

자유게시판 >

넷플릭스 추천해주세요~

 jellybeans 열정회원 1:1 채팅
2020.08.09. 23:24 조회 343

재밌는 넷플릭스 미드 한드 추천 부탁드립니다~

음 제 취향은 법정(굿와이프) 범죄수사물(잔인한거 빼구요ㅠ 멘탈리스트, ncis 이정도 수준요) 약간 특이한 시트콤(커뮤
니티, 굿플레이스, 모던패밀리) 좋아해요 ㅋ

한드는 간단한 잡일 하면서 볼 수 있도록 가벼운거요 최근에 리갈하이 봤고 우리 사랑했을까 보구 있어요 ㅎ

넷플릭스 전문가님들 재밌는걸로 추천 부탁드립니다 :~)

“ **자신이 재미있게 봤던 영상 콘텐츠를 입력하면,**
그에 따라 비슷한 영상들을 추천 받을 수 있도록 하자! ”

데이터 및 추천 알고리즘 _ 넷플릭스 크롤링

- 미국 TV 프로그램, 드라마, 오리지널, 영화 등 중복된 것을 제거한 영상 콘텐츠 **약 2,600여개** 크롤링
- 콘텐츠 기반 필터링에 활용될 영화의 특징 - '**영화 제목, 줄거리, 장르, 특징**' 이라는 총 4개의 변수들을 활용
- 4개의 변수들 중에서 **줄거리 변수를 중점으로 추천 서비스 구축**

	Title	Summary	Genre	Feature
0	비밀의 숲	공감 능력을 잃은 검사가 살인 사건을 파헤친다 검사는 열혈 여형사의 도움을 받으며...	TV 드라마 범죄 한국 드라마 TV 프로그램 스릴러 TV 드라마	긴장감 넘치는
1	사이코지만 괜찮아	사랑 따위 못 느끼면 어때 태어날 때부터 온기가 없는 동화 작가 그녀가 사랑을...	TV 드라마 로맨스 TV 프로그램 코미디 로맨스 한국 드라마 TV 드라마 TV 프로...	영동 기발 가슴 뭉클 풍부한 감정 로맨스
2	킹덤	병든 왕을 둘러싸고 흥흥한 소문이 떠돈다 어둠에 뒤덮인 조선 기이한 역병에 신음...	웹툰 원작 한국 드라마 TV 프로그램 정치 한국 드라마 시대물 TV 프로그램 스릴러...	폭력 흥미진진
3	슬기로운 의사생활	탄생과 죽음이 공존하는 인생의 축소판이라고 불리는 병원에서 특별한 하루하루를 살아...	한국 드라마 TV 드라마 TV 프로그램 코미디	가슴 뭉클
4	인간수업	학교 안에서는 모범생 학교 밖에서는 영리한 범죄자 그의 비밀스러운 이중생활은 완...	TV 드라마 범죄 한국 드라마 TV 프로그램 10대 TV 드라마	다크
5	루시퍼	지긋지긋한 지옥 생활을 청산하고 은퇴를 선언한 악마 LA로 와 나이트클럽을 열고 ...	TV 드라마 범죄 TV 프로그램 판타지 TV 프로그램 만화 원작 TV 드라마 미국 ...	흥미진진
6	모범형사	내가 직접 잡아넣은 사형수가 무죄라면 그 억울한 누명 뒤에 사실을 은폐하려는 막강...	TV 드라마 범죄 한국 드라마 TV 프로그램 스릴러 TV 드라마	긴장감 넘치는
7	사랑의 불시착	뜻밖의 돌풍은 행운일까 불운일까 패러글라이딩 사고로 북한에 불시착한 재벌 딸 ...	TV 드라마 로맨스 TV 프로그램 코미디 로맨스 한국 드라마 TV 드라마 TV 프로...	가슴 뭉클 로맨스
8	이태원 클라쓰	활력과 욕망이 넘치는 이태원 한 전과자와 친구들이 꿈을 위해 뭉쳤다 일단 시작은...	웹툰 원작 한국 드라마 한국 드라마 사회 문제 TV 드라마 TV 드라마	감동 풍부한 감정
9	빨간 머리 앤	마음에 불꽃을 품은 여자아이 앤 나이 든 오누이의 초록지붕집에 발을 들이게 된다 ...	도서 원작의 TV 프로그램 가족이 함께 보기 좋은 TV TV 프로그램 10대 시대물...	가슴 뭉클 풍부한 감정
10	위쳐	세계적인 판타지 대작이 넷플릭스 시리즈로 다시 태어난다 어둠과 위험의 세계 진실...	TV 프로그램 판타지 도서 원작의 TV 프로그램 TV 드라마 미국 TV 프로그램 T...	흥미진진

전처리 _ Stopwords 제거 및 토큰나이저

- 줄거리 변수에서 유의미한 내용을 추출하기 위해 단어 및 어절로 나누는 **토큰화** 수행
- 문장의 의미를 파악하는데, 의미가 없는 '**구두점**' 및 '**Stopwords**' 등을 제거

['공감', '능력', '읽다', '검사', '살인', '사건', '파헤치다', '검사', '형사', '도움', '발다', '수사', '방해', '부패하다', '정치', '상황', '맞서다']
 ['사랑', '따위', '뜻', '느끼다', '어떨다', '태어나다', '때', '부터', '온기', '없다', '동화', '작가', '그녀', '사랑', '외면', '정신', '병중', '보호', '사오다', '마주치다', '이상하', '지만', '꽤', '괜찮다', '힐링', '로맨틱'
 ['병들다', '왕', '둘러싸다', '흥하다', '소문', '뜨다', '돈', '어둠', '뒀이다', '조선', '기이하다', '역병', '신음', '사다', '정체', '모르다', '악', '맞서다', '백성', '구원', '희망', '오직', '세자']
 ['탄생', '죽음', '공존', '인생', '축소', '판이', '라고', '불리다', '병원', '특별하다', '하루하루', '살아가다', '사람', '이야기', '그런', '드라마']
 ['학교', '모범생', '학교', '영리하다', '범죄자', '비밀스럽다', '이중생활', '완벽하다', '반', '친구', '위험하다', '사업', '관심', '갖다', '전', '까지는']
 ['지긋지긋하다', '지옥', '생활', '청산', '하고', '은퇴', '선언', '악마', 'la', '오다', '나이트클럽', '열다', '인생', '즐기다', '어느', '날', '강력', '계', '형사', '만나다', '새롭다', '삼', '살기', '시작']
 ['직접', '잡아넣다', '사형수', '무죄', '라면', '억울하다', '누명', '사실', '은폐', '막강', '권력', '있다', '그냥', '조용하다', '묻다', '상책', '일', '것', '하지만', '모범', '형사', '다르다', '어떨다', '결과', '기다리다',
 ['웃밖', '돌풍', '행운', '불운', '패러글라이딩', '사고', '북한', '불시', '착하다', '재벌', '알', '곳', '간간하다', '북한', '군', '장교', '만나다', '와중', '피어', '오르다', '낮선', '감정', '지']
 ['활', '력', '욕망', '넘치다', '이태원', '전과자', '와', '친구', '꿈', '위해', '뭉치다', '일단', '시작', '술집', '창업', '불합리', '세상', '막강', '적', '앞길', '막다', '기르다', '쓰다', '맞서다', '싸우다']
 ['마음', '불꽃', '품다', '여자아이', '앤', '나이', '들다', '오누이', '초록', '지붕', '집', '발', '들이다', '되다', '커스버트', '남매', '와', '낮선', '고아', '소녀', '가족', '되다', '성장하다', '과정', '그리다', '내다',
 ['세계', '적', '인', '판타지', '대작', '넷플릭스', '시리즈', '다시', '태어나다', '어둠', '위험', '세계', '진실', '어디', '있다', '운명', '묵인', '위', '치다', '마법사', '홀로', '남다', '공주', '싸움', '시작', '되다']
 ['나쁘다', '끌리다', '잘나다', '짤나다', '어리다', '설레다', '무섭다', '섹시하다', '뭔', '소리', '나', '고', '되다', '일', '하나', '없다', '14년', '차', '싱글맘', '4', '명의', '남자', '나타나다', '말씀', '로맨스', '기대하
 ['눈부시다', '발전', '정단', '기술', '하지만', '인간', '어', '두운', '본능', '기술', '이용', '기이하다', '악몽', '시작', '되다', '세계', '환상', '특급', '이르다', '불리다', 'sf', '시리즈']
 ['소심하다', '소년', '오', '티스', '대', '반전', '아웃사이더', '소녀', '와', '함께', '섹스', '클리', '닉', '연', '성', '상당사', '엄마', '귀', '동냥', '산지식', '모두', '전수', '해드리다', '오', '티스', '절찬', '상당',
 ['애초', '부터', '어울리다', '알다', '웃', '스페인', '최고', '명문', '사람', '학교', '전학', '오다', '되다', '세', '명의', '평범하다', '아이', '행운', '불잡다', '줄', '알다', '살인', '사건', '깊이', '연루', '되다']
 ['누군가', '나르다', '좋아하다', '사람', '나', '다가오다', '알람', '울', '린다', '마음', '알다', '앱', '좋다', '알람', '시대', '힘겹다', '살아가다', '소녀', '조조', '에게도', '푹푹하다', '사랑', '찾아오다']
 ['일과', '가정', '모두', '놓치다', '알다', '그녀', '다정하다', '이웃', '친구', '있다', '더없이', '행복하다', '그러나', '그녀', '세계', '비로소', '완성', '세월', '에도', '변하다', '없다', '남편', '사랑', '그럴다', '사랑',
 ['천진하다', '강단', '있다', '싱글', '맘', '동백', '씩씩하다', '고', '싫다', '세상', '편견', '여전하다', '힘들다', '이럴다', '그녀', '찾아오다', '설레다', '어쩌면', '사랑', '이제', '부터', '조금', '더', '융감하다', '되다'
 ['웃밖', '행운', '아니다', '뛰어나다', '재능', '뉴욕', '거대', '로펌', '들어가다', '대학', '중퇴', '청년', '엘리트', '변호사', '와', '의기', '투합', '시시하다', '따분하다', '세상', '상대로', '소송']
 ['가공', '테러', '삼시간', '아수라', '장이', '되다', '워싱턴', '이제', '미함중국', '운명', '하위', '각료', '사람', '손', '달리다', '긴장', '늦추다', '수', '없다', '정치', '스릴러']

모델 구축 _ 유사도 행렬

- TF-IDF 및 코사인 각도 기반의 벡터 유사도를 활용하여 유사도 행렬을 계산
- 유사도 값이 높은 값을 기준으로 인덱스 값을 추출

```
[[1.         0.         0.01190298 ... 0.06778533 0.         0.         ]
 [0.         1.         0.         ... 0.         0.         0.00310872]
 [0.01190298 0.         1.         ... 0.         0.         0.         ]
 ...
 [0.06778533 0.         0.         ... 1.         0.         0.         ]
 [0.         0.         0.         ... 0.         1.         0.         ]
 [0.         0.00310872 0.         ... 0.         0.         1.         ]]
```

< 유사도 행렬 >

```
[[ 0  356 1682 ... 1174 1172 2625]
 [ 1  248  527 ... 1066 1063 1312]
 [ 2  716 1879 ...  956  958 2625]
 ...
 [2623 2588 2496 ... 1318 1349 2625]
 [2624 1506 2151 ...  945  947 2625]
 [2625  850 1065 ... 1138 1154 1312]]
```

< 인덱스 추출 >

모델 구축 _ 1차 시도 및 문제점 도출

- 사용자가 관심 있어 하는 영상 콘텐츠 제목을 입력하면, **줄거리에 따라 유사도가 높은 콘텐츠 추천**

영화 제목	줄거리	특징
7번 방의 선물 (Input 값)	홍악범 들이 득실대는 7번 방에 이상한 놈이 들어온다. (중략) 교도소 역사상 전무후무한 외부인 반입 작전! 지금까지 이런 감동은 없었다.	가슴 뭉클, 감동
집행자	남아공의 한 교도소 에서 살인을 저지른 젊은 간수. (후략)	다크
살인자 만들기	홍악범 으로 기소되어 복역 중인 두 남자. (후략)	도발, 수사, 다크
아블루	(전략) 교도소 내 알력 다툼과 살인 사건에 휘말려도, 결코 꺾이지 않겠다. (후략)	긴장감 넘치는

그런데, '7번 방의 선물'을 Input으로 넣었을 때 나왔던 추천 콘텐츠 중 일부가 '홍악범, 교도소'를 인식하여 **특징이 전혀 다른 콘텐츠를 추천하는 문제점**을 발견!

모델 구축 _ 보완 및 2차 시도

- 1차 시도의 문제 원인은 결국 '줄거리' 변수 한 가지만 사용해서 생긴 문제점
- 더욱 풍부하게 영화의 특징을 이해하기 위해 '특징' 변수도 함께 고려하여 이중으로 필터링
- 즉, '줄거리'가 비슷한 것들을 기반으로 나열한 후, '특징' 변수에서 절반 이상 동일한 항목들을 최종적으로 추천

영화 제목	줄거리	특징
7번 방의 선물 (Input 값)	(전략) 6살 지능을 가진 딸을 교도소 안으로 데려오려 하는 전무후무한 반입 작전...	가슴 뭉클, 감동
글러브	(전략) 음주 폭행 물의를 일으킨 프로 야구 선수가 청각 장애 야구부 코치로 부임해 무려 전국 대회 출전이라는 감동 스토리, 지금 시작된다.	가슴 뭉클, 풍부
아이 엠 샘	7살 지능밖에 갖지 못했지만 누구보다 딸을 사랑하는 샘. (중략) 딸을 되찾으려면...	가슴 뭉클, 풍부
인스턴스 패밀리	(전략) 열결에 하나도, 둘도 아닌 세 아이를 데려온 부부. 아이들은 상처가 많아 새 가정의 사랑을 온전히 받아들이지 못하는데...	가슴 뭉클

최종 결과

- 깊이 있는 '콘텐츠 기반 필터링' 추천 알고리즘을 활용하기 위하여 '줄거리'와 '특징' 이라는 두 가지 변수를 활용해 최종적으로 사용자의 Input 값에 대하여 가장 유사한 영상 콘텐츠를 추천
- 사용자가 자신이 선호하는 영상 콘텐츠 취향에 대해 더욱 깊이 이해할 수 있도록 최종적으로 추천된 영화들의 '장르'를 이용하여 워드클라우드 형태로 결과를 도출
- 더 많은 종류의 영상 데이터와 관련된 보다 풍부한 변수들이 있다면 즉, 데이터의 양적 및 질적 퀄리티가 증가된다면, 사용자에게 더 잘 맞는 결과를 도출할 수 있을 것



< '7번 방의 선물' 워드클라우드 결과 >