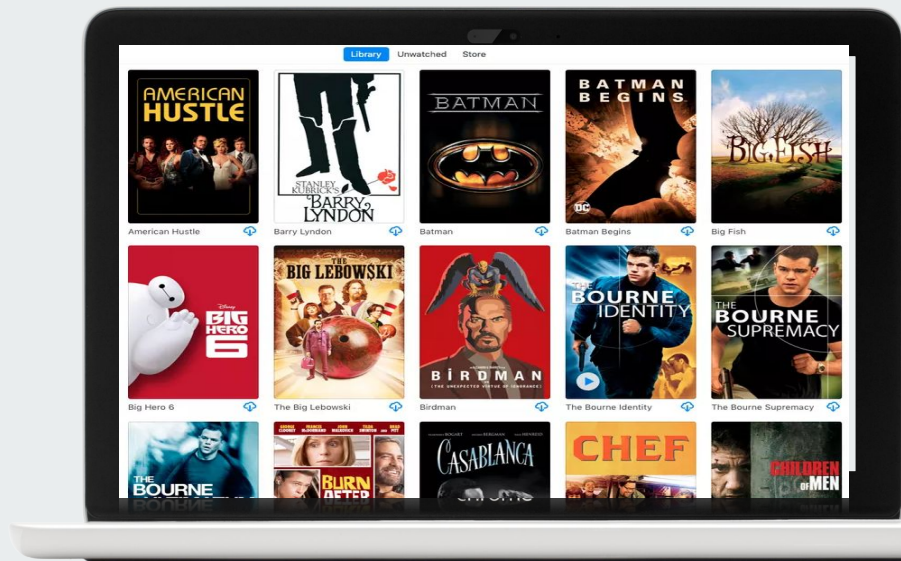# MOVIE RECOMMENDATION SYSTEM

Problem statement and solution proposal

The Problem

For the viewer, the problem is of choice as there are thousands of movies to choose from. The viewer experience is better when movies can be recommended according to his tastes and preferences in the best possible manner. The purpose of this project is to build a recommendation engine for movies.Recommender System seeks to predict or filter movies according to the user's choices.
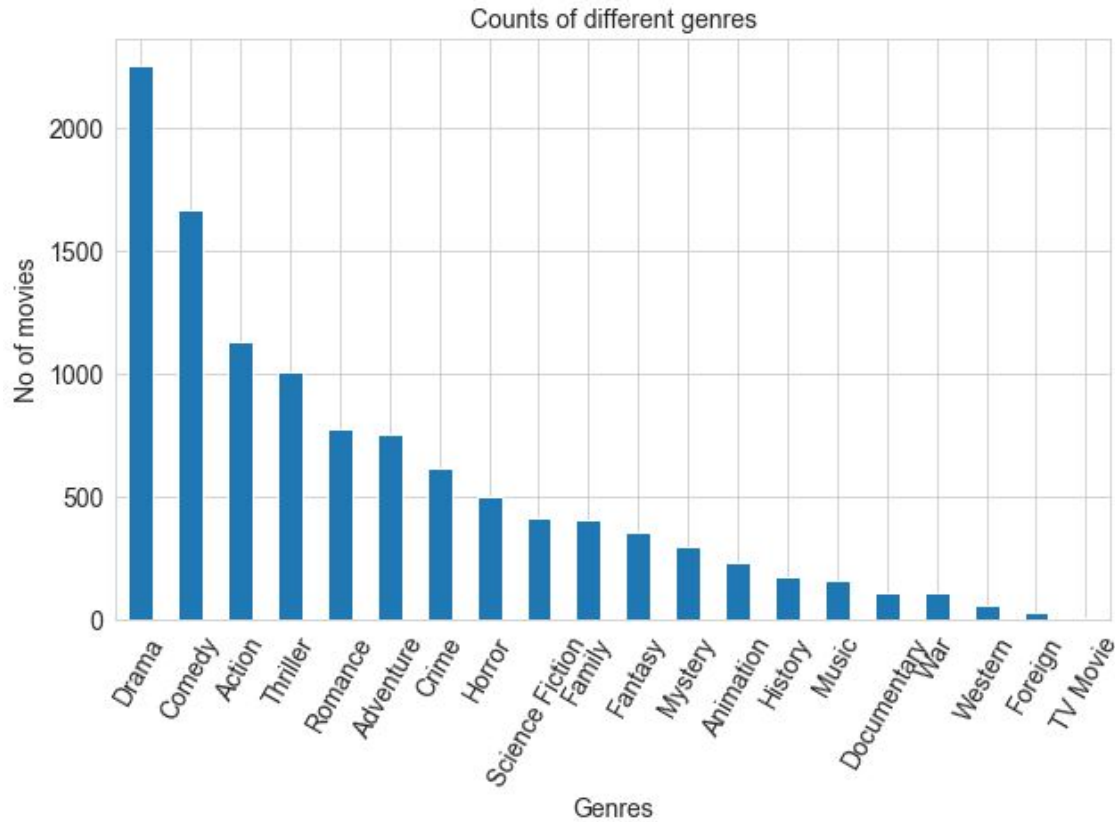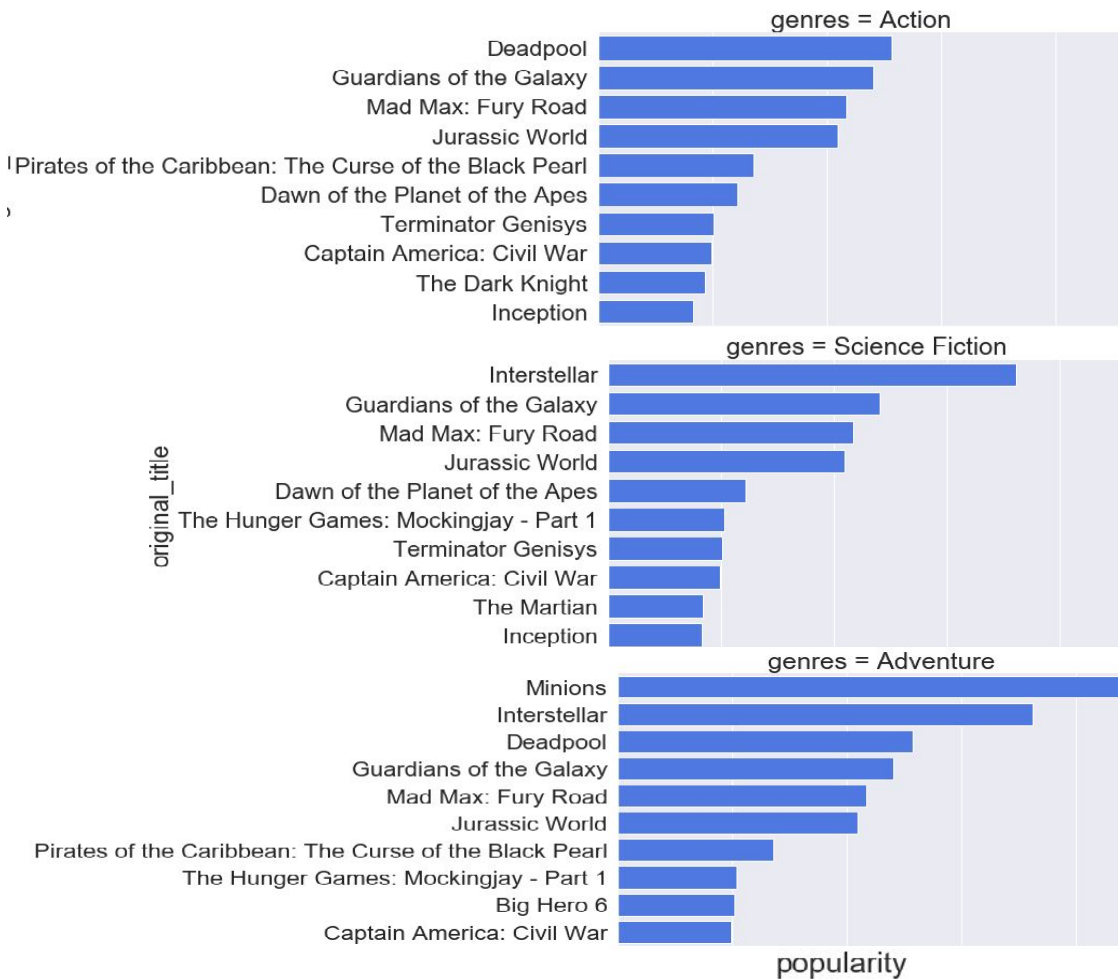
STEPS

1. Get global rating data:

   I have worked on 2 datasets. For content based recommendations I have used the TMDB 5000 Movies dataset. For applying collaborative filters I have used data from movielens dataset of 10000 movies.

2. Creating generic recommendations. These recommendations are in general for all viewers.

3. Content based recommendations: It recommends movies similar to what the user likes.

4. Implement collaborative filtering: It recommends movies specific to a viewer based on his tastes

Counts of different genres

Most of the movies are in the genres:- Drama, Comedy and Action

genres = Action

- Deadpool
- Guardians of the Galaxy
- Mad Max: Fury Road
- Jurassic World
- Pirates of the Caribbean: The Curse of the Black Pearl
- Dawn of the Planet of the Apes
- Terminator Genisys
- Captain America: Civil War
- The Dark Knight
- Inception

genres = Science Fiction

- Interstellar
- Guardians of the Galaxy
- Mad Max: Fury Road
- Jurassic World
- Dawn of the Planet of the Apes
- The Hunger Games: Mockingjay - Part 1
- Terminator Genisys
- Captain America: Civil War
- The Martian
- Inception

genres = Adventure

- Minions
- Interstellar
- Deadpool
- Guardians of the Galaxy
- Mad Max: Fury Road
- Jurassic World
- Pirates of the Caribbean: The Curse of the Black Pearl
- The Hunger Games: Mockingjay - Part 1
- Big Hero 6
- Captain America: Civil War

original_title

popularity

Most popular movies listed on TMDB are of genres:- Action, Sci-Fi and Adventure

FINDINGS FROM TMDB DATASET

Vote Average as an Indicator

Title	Vote Count	Vote Average
Dancer, Texas Pop. 81	1	10
Little Big Top	1	10
Stiff Upper Lips	1	10
Me You and Five Bucks	2	10
Sardaarji	2	9.5
One Man's Hero	2	9.3
The Shawshank Redemption	8205	8.5
There Goes My Baby	2	8.5
The Godfather	5893	8.4
The Prisoner of Zenda	11	8.4

Vote Average, used alone as an indicator for good movies shows movies watched by single viewers. As these scores are topping the charts, it could be misleading.

A more reliable Indicator

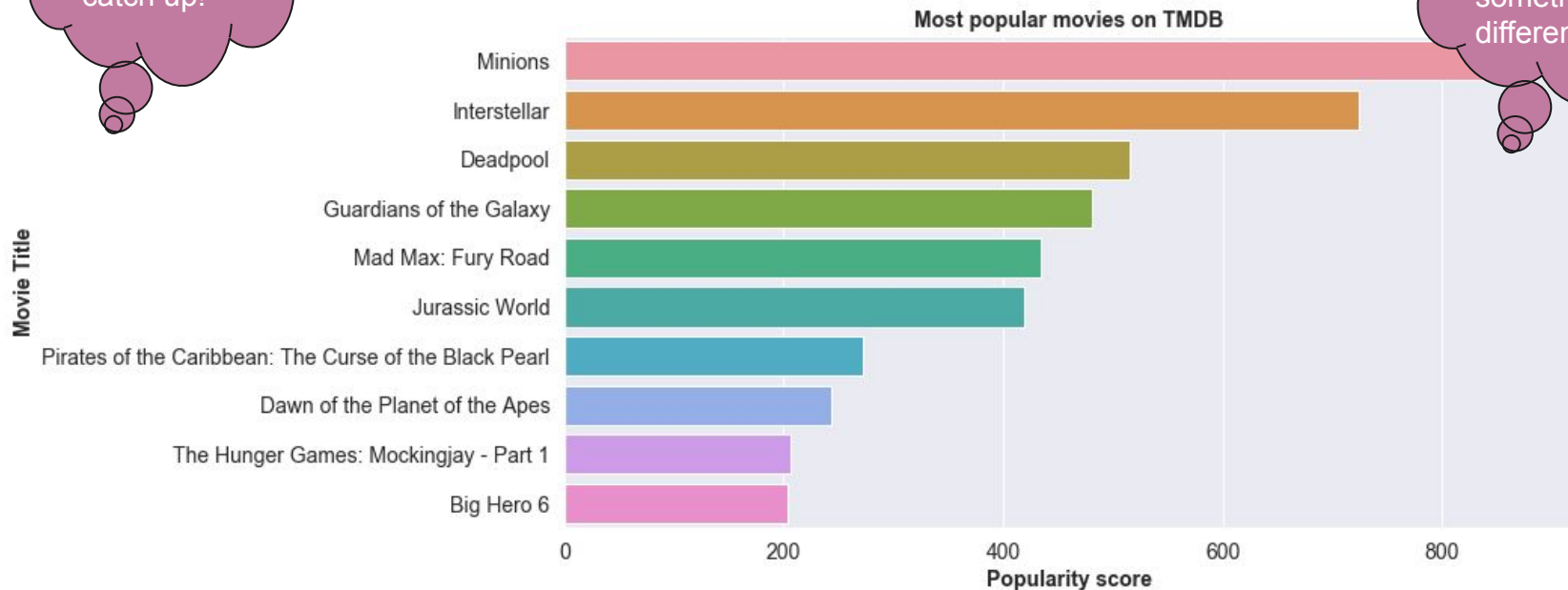| Movie ID	Title	Vote Count	Vote Average
96	Inception	13752	8.1
65	The Dark Knight	12002	8.2
0	Avatar	11800	7.2
16	The Avengers	11776	7.4
788	Deadpool	10995	7.4
95	Interstellar	10867	8.1
287	Django Unchained	10099	7.8
94	Guardians of the Galaxy	9742	7.9
426	The Hunger Games	9455	6.9
127	Mad Max: Fury Road	9427	7.2

If Vote Average along with vote count is chosen as an indicator for good movies, it makes recommendations of movies more reliable.
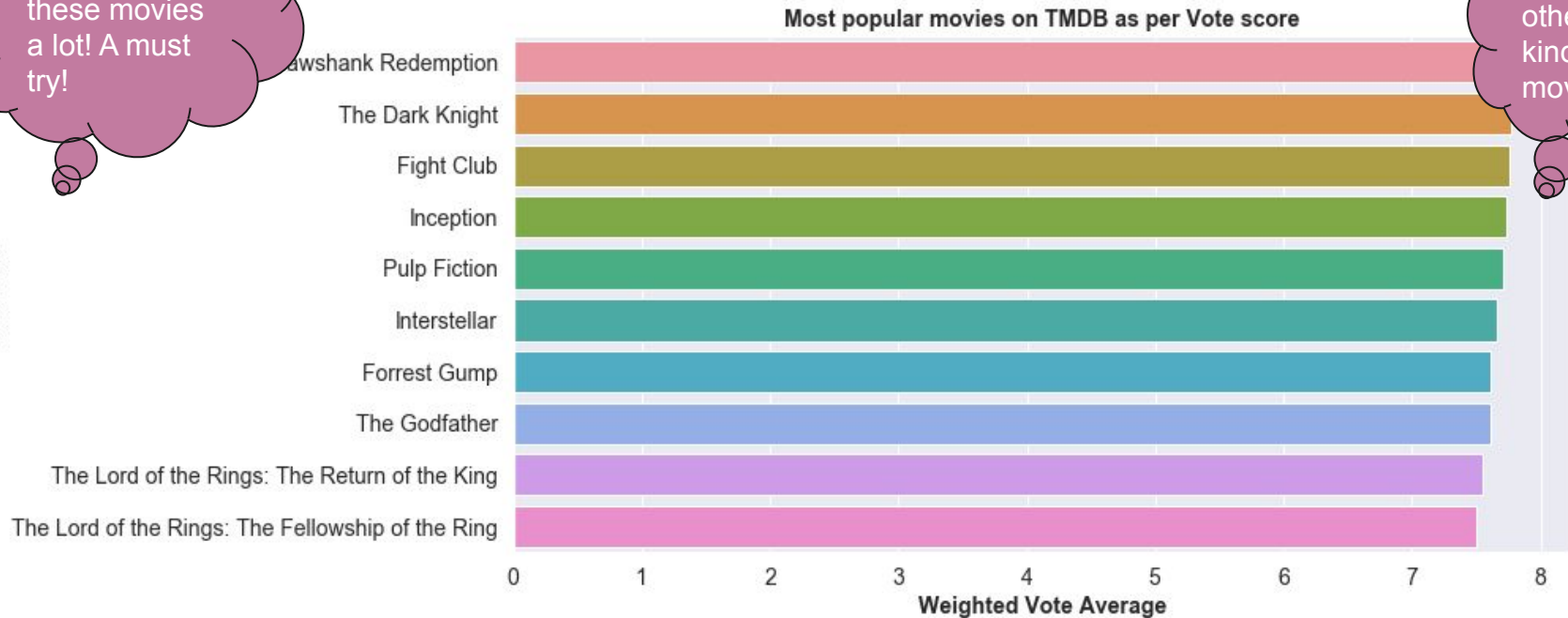
HOW TO IDENTIFY GOOD MOVIES?

GENERIC RECOMMENDATIONS

RECOMMENDATIONS BASED ON WEIGHTED VOTE AVERAGE:
We have been able to improve the movie chart by being selective. The indicator used is a weighted vote average that takes into account, vote average, vote count and benchmark in viewership. These movies indicate the best impressions of larger number of viewers. However, generic recommendations are impersonalised.

Heard about these movies a lot! A must try!

But I like some other kinds of movies!

GENERIC RECOMMENDATIONS

RECOMMENDATIONS BASED ON CONTENT FILTERING:
This recommendation is designed to match movies of the same description.The steps are:
- Clean text data of movie descriptions.
- Convert text into Tfidf vectors.
- Find similarities between movies by computing cosine similarities.
- Group movies that are most similar to a selected movie.

Batman movies are my favourites. Looking for movies similar to 'Dark Knight Rises'

'Dark night Rises' is a well made movie. But I am looking for movies of same make or same director.

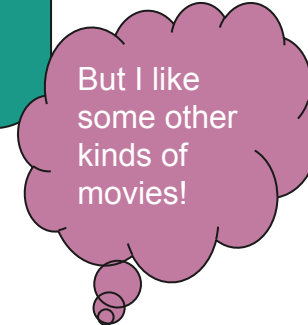| Title |
| --- |
| Batman Forever |
| The Dark Knight |
| Batman |
| Batman Returns |
| Slow Burn |
| Batman Begins |
| Batman: The Dark Knight Returns, Part 2 |
| JFK |
| Batman & Robin |
| Batman v Superman: Dawn of Justice |

CONTENT BASED RECOMMENDATIONS

RECOMMENDATIONS BASED ON CONTENT FILTERING:

This recommendation is designed to match movies of the same make. A text metadata is created including director, genres, keywords and lead actors. The steps are:

- Clean text data of movie descriptions.
- Convert text into word vectors.
- Find similarities between movies by computing cosine similarities.
- Group movies that are most similar to a selected movie.

Heard about these movies a lot! A must try!

But I like some other kinds of movies!

Title	Weighted Vote Average
The Dark Knight	7.77
Inception	7.74
Interstellar	7.66
The Dark Knight Rises	7.22
The Prestige	7.22
Batman Begins	7.09
Memento	7.24
Kick-Ass	6.70
Hitman	6.05
Insomnia	6.29

CONTENT BASED RECOMMENDATIONS

RECOMMENDATIONS BASED ON USER BASED COLLABORATIVE FILTERING:
These recommendations initially grab attention as it shows all the possible movies that the viewer is likely to rate high. These movies are grouped together based on collaborative ratings of similar kinds of viewers. Knn Basic Model from Surprise Library has been used to learn and predict from the data by Movielens.

Movie ID	Title	Estimated Ratings	Watched Status	Weighted Vote Average
491	Shawshank Redemption, The (1994)	4.7	unseen	7.85
6710	Dark Knight, The (2008)	4.6	unseen	7.77
7372	Inception (2010)	4.4	unseen	7.74
8376	Interstellar (2014)	4.3	unseen	7.66
847	Godfather, The (1972)	4.5	unseen	7.61
4800	Lord of the Rings: The Return of the King, The (2003)	4.3	unseen	7.55
3639	Lord of the Rings: The Fellowship of the Ring, The (2001)	4.2	unseen	7.51
8475	Guardians of the Galaxy (2014)	4.2	unseen	7.47
4137	Lord of the Rings: The Two Towers, The (2002)	4.3	unseen	7.45
8063	Django Unchained (2012)	4.2	unseen	7.40

These movies look interesting! It's a must watch?

USER SPECIFIC RECOMMENDATIONS

These are my favourites. I talk about it all the time

These are exactly my kind of movies i'd like to watch!

Title	Estimated Rating	Watched Status	Weighted Vote Average	Genres
Star Wars: Episode V - The Empire Strikes Back (1980)	4.4	seen	7.48	['Action', 'Adventure', 'Sci-Fi']
Star Wars: Episode IV - A New Hope (1977)	4.5	seen	7.47	['Action', 'Adventure', 'Sci-Fi']
Back to the Future (1985)	4.3	seen	7.36	['Adventure', 'Comedy', 'Sci-Fi']
Gladiator (2000)	4.3	seen	7.25	['Action', 'Adventure', 'Drama']
Star Wars: Episode VI - Return of the Jedi (1983)	4.5	seen	7.19	['Action', 'Adventure', 'Sci-Fi']
Toy Story (1995)	4.1	seen	7.11	['Adventure', 'Animation', 'Children', 'Comedy', 'Fantasy']
Jurassic Park (1993)	3.9	seen	7.02	['Action', 'Adventure', 'Sci-Fi', 'Thriller']
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)	4.5	seen	6.99	['Action', 'Adventure']
Indiana Jones and the Last Crusade (1989)	4.3	seen	6.86	['Action', 'Adventure']
Monty Python and the Holy Grail (1975)	4.2	seen	6.71	['Adventure', 'Comedy', 'Fantasy']

Workings:
- Identified the genres that user likes based on his ratings.
- Applied model's predictions on all unseen movies.
- Selected movies of genres liked by user
- Filtered movies based on weighted vote Average.

Title	Estimated Rating	Watched Status	Weighted Vote Average	Genres
Lord of the Rings: The Return of the King, The (2003)	4.35	unseen	7.55	['Action', 'Adventure', 'Drama', 'Fantasy']
Lord of the Rings: The Fellowship of the Ring, The (2001)	4.22	unseen	7.51	['Adventure', 'Fantasy']
Guardians of the Galaxy (2014)	4.19	unseen	7.47	['Action', 'Adventure', 'Sci-Fi']
Lord of the Rings: The Two Towers, The (2002)	4.25	unseen	7.45	['Adventure', 'Fantasy']
Inside Out (2015)	3.91	unseen	7.40	['Adventure', 'Animation', 'Children', 'Comedy', 'Drama', 'Fantasy']
Lion King, The (1994)	4.30	unseen	7.31	['Adventure', 'Animation', 'Children', 'Drama', 'Musical', 'IMAX']
WALL-E (2008)	4.21	unseen	7.24	['Adventure', 'Animation', 'Children', 'Romance', 'Sci-Fi']
Dark Knight Rises, The (2012)	4.23	unseen	7.22	['Action', 'Adventure', 'Crime', 'IMAX']
Up (2009)	4.18	unseen	7.21	['Adventure', 'Animation', 'Children', 'Drama']
The Martian (2015)	4.06	unseen	7.16	['Adventure', 'Drama', 'Sci-Fi']

USER SPECIFIC RECOMMENDATIONS

HYBRID RECOMMENDATION BASED ON COLLABORATIVE FILTERS AND CONTENT BASED FILTERS.

I liked 'Gladiator 'a lot! Let me explore movies like this one !

Ah! These are are my kind of movies I am looking forward to watch.

Title	Watched Status	Weighted Vote Average	Estimated Rating
Seven Samurai (Shichinin no samurai) (1954)	unseen	6.56	4.27
Blade Runner (1982)	unseen	7.06	4.17
Nebraska (2013)	unseen	6.32	4.12
The Martian (2015)	unseen	7.16	4.06
Shin Godzilla (2016)	unseen	6.11	4.00
Raise the Titanic (1980)	unseen	6.09	4.00
Ip Man 3 (2015)	unseen	6.14	4.00
Gunman, The (2015)	unseen	6.02	4.00
Nicholas Nickleby (2002)	unseen	6.10	4.00
Bourne Supremacy, The (2004)	unseen	6.63	3.98

Workings:
- Selected a movie from viewer's liked genres.
- Applied models predictions on unseen movies
- Similar movies based on content were filtered.
- Applied filter of weighted vote average

Why collaborative filtering is better ?

- It's efficient, provided all relevant user data is available.
- This recommender is able to match users and generate reasonable predictions on ratings.
- User data helps to customise the recommendations.

References

https://www.kaggle.com/rounakbanik/movie-recommender-systems

https://www.datacamp.com/community/tutorials/recommender-systems-python

https://blog.dominodatalab.com/recommender-systems-collaborative-filtering/

Contact:

aroonima12@gmail.com

Contact

aroonima12@gmail.com