

# WHICH MOVIE ???

A movie recommender





400TM

46



400TM

47

400TM

48

400TM

49



# BUSINESS OVERVIEW

This project is aimed at providing personalized suggestions to movie-lovers based on their previously watched movies. The model is able to reduce a user's time in searching for movies by recommending to them top five movies

# PROBLEM STATEMENT

---

Leisure and entertainment have since time immemorial been part of society and with advancement in technology, movie production has been on the rise over the past two centuries.

Computer-Generated Imagery (CGI), compelling storylines, great actors and a massive audience have also led to streaming media platforms developed to accommodate users' week-to-week busy schedules that give them no time for live cinema experiences.

Yet the constant question probing such users is

**WHAT MOVIE SHOULD I WATCH?**

Instead of perusing through and through or getting bored after some minutes into a movie, the project attempts to suggest to such a person, movies they might be interested in. Streaming media companies are using machine learning systems to optimize and give every movie-lover an easy time to select preferential movies.

# MAIN OBJECTIVE

To utilize advanced machine learning techniques to understand a user-movie relationship and seek to predict the ratings they'd have made on unwatched movies and recommend the movies with the highest predicted ratings.



# SPECIFIC OBJECTIVES

- (a) To understand user-movie viewership and movie-viewership and ratings.
- (b) To predict ratings for unwatched movies.
- (c) To recommend unwatched movies from the predicting model.

# DATA UNDERSTANDING

movieId	title	genres	no. of views	mean_rating
356	Forrest Gump (1994)	Comedy Drama Romance War	329	4.164134
318	Shawshank Redemption, The (1994)	Crime Drama	317	4.429022
296	Pulp Fiction (1994)	Comedy Crime Drama Thriller	307	4.197068
593	Silence of the Lambs, The (1991)	Crime Horror Thriller	279	4.161290
2571	Matrix, The (1999)	Action Sci-Fi Thriller	278	4.192446

The data used to create the model consisted of information about movies and how users rated the movies they watched.

The picture above shows the top most watched movies and their average rating.

The picture besides shows the users who watched the most movies.

The users' ID keep the users anonymous.

userId	no. of watched movies
413	414
598	599
473	474
447	448
273	274

# MODELLING AND EVALUATION

---

The project was generally based on collaborative filtering methods where a user's preference is based on the other movies they have watched or what other users similar to them have watched.

The best model was Singular Value Decomposition(SVD) as it gave the least RMSE metric of 0.87. In comparison to the other models SVD's predictions were more accurate with the least error distance between the actual ratings and predicted ratings.

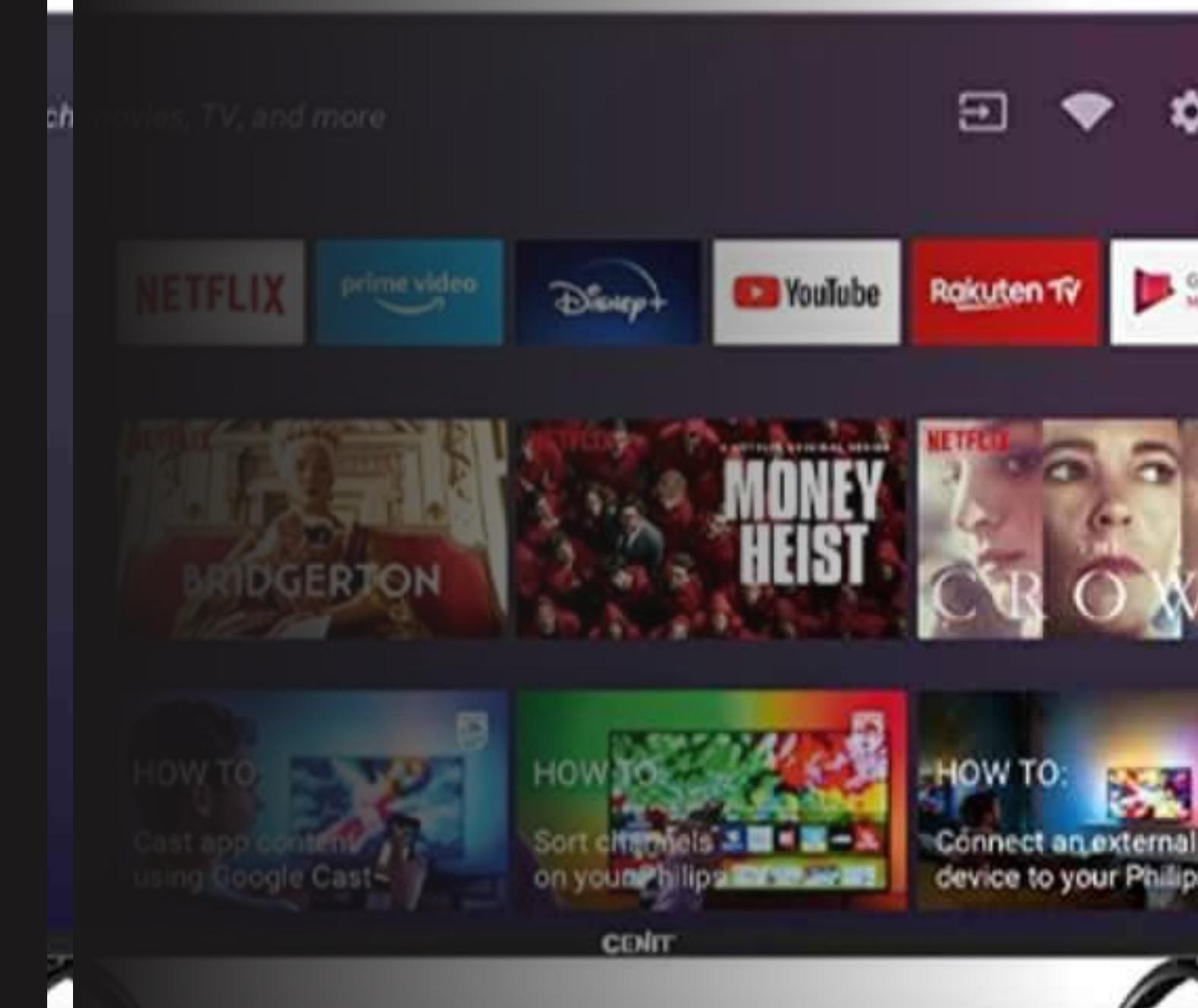
# RECOMMENDER

A **recommender** function was built that takes in a user's ID, the rating data and movie data.

It returns the top five movies suggested to the specific user based on the models prediction of the movies they have not watched.

```
[47] recommender(414,df,movies)
```

	movieId	predicted ratings	title	genres
0	2324	4.406175	Life Is Beautiful (La Vita è bella) (1997)	Comedy Drama Romance War
1	1203	4.294802	12 Angry Men (1957)	Drama
2	7387	4.261250	Dawn of the Dead (1978)	Action Drama Horror
3	1237	4.256279	Seventh Seal, The (Sjunde inseglet, Det) (1957)	Drama
4	2788	4.195378	Monty Python's And Now for Something Completel...	Comedy
5	3730	4.186059	Conversation, The (1974)	Drama Mystery



**THANK YOU**

