

The background of the slide is a stylized theater stage. On the left and right sides, there are vertical red curtains. A spotlight from the left illuminates a red carpet that leads towards the center of the stage. The main area of the stage is a dark blue color with a diagonal grey band running from the top right towards the bottom left. The title 'Recommendation System' is written in a large, white, serif font, centered on the stage.

Recommendation System

By: Tommy Phung

Overview

Source: MovieLens

- **Authors:** GroupLens

Main Feature: Ratings

Goal: Recommendation system that suggested five movies based off prior users' ratings

grouplens

UNIVERSITY OF MINNESOTA



Business Problem

- ❖ Over 8 million movies
- ❖ Limited time
- ❖ Prone to suggestion
 - Users watches recommended videos 70% of the time on Youtube

Improve streaming services by offering better recommendations



Targeted Streaming Services

The image depicts a stage with a dark blue background and red curtains at the top. A spotlight from the top right corner illuminates the center stage. On the stage, there are three logos: 'max' in white on a blue square on the left, 'NETFLIX' in red on a black square in the center, and 'prime video' with the Amazon smile logo in black on a light blue square on the right. The foreground shows the tops of several rows of red theater seats.

max

NETFLIX

prime video

Data Understanding

Number of users: 610

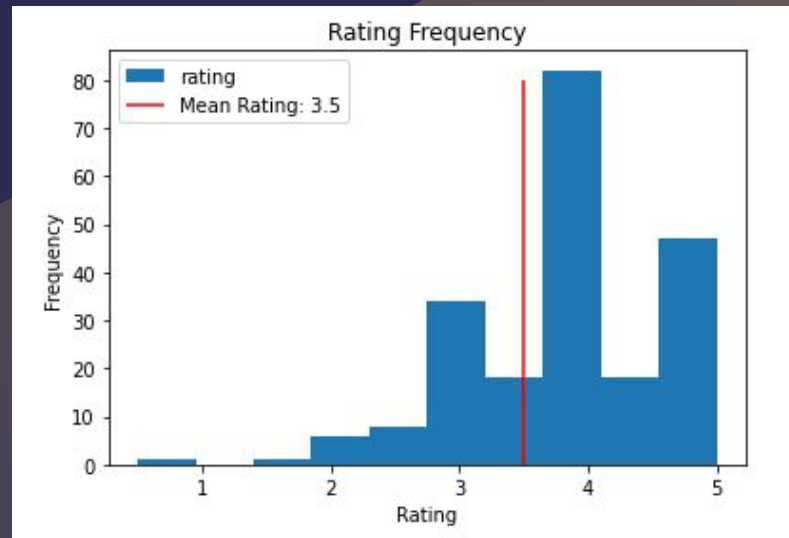
Number of Movies: 9,742

Number of reviews: 100,836

Dataset features: **Rating, Genre and Tags***

Rating Range: 0 (Bad) → 5 (Good)

- 0.5 steps



Data Preparation

1. Remove timestamp from dataset (Not Needed)
2. Create training and testing set (Test Accuracy)
3. Standardize Genres (Content Filtering)

Methods

Main Library: **Surprise**

Memory Based Modeling:

KnnBasic, KnnBaseline, KnnWithMean

- Clusters grouping similar items

Model Based Modeling:

SVD Singular Value Decomposition

- Matrix decomposition

Best Performing: KnnBaseline

Measurement

RSME: (Root Square Mean Error)

- Average error for the predicted rating

NDCG: (Normalized Discounted Cumulative Gain)

- Method to compare ranked list and relevance

Used RSME for interpretability

The background of the slide is a dark blue stage with red curtains on the left and right sides. A spotlight from the top right corner illuminates a diagonal path across the stage floor. At the bottom of the slide, there is a row of red theater seats.

Additional Conditions

1. Model won't recommend seen movies
2. Consider new users and movies (Cold Start)

User 1 Recommendations

Model: KnnBaseline

Added filtering: Top Favorite Genres

Includes Favorite Genres

Ghost in the Shell (Kôkaku kidôtai) (1995)

Singin' in the Rain (1952)

Notorious (1946)

Sicario (2015)

Big Short, The (2015)

Cold Start

A cold start is a problem when a **new user or item** is added without **prior history** in the current system.

User Case: A new user is added needs a recommendation.



Item Case: A new movie is added and need to be recommended.



Ex. New User Recommendations

Most Rated Movies	Most Popular Movies
Paper Birds (Pájaros de papel) (2010)	Pulp Fiction (1994)
Act of Killing, The (2012)	Shawshank Redemption, The (1994)
Jump In! (2007)	Forrest Gump (1994)
Human (2015)	Silence of the Lambs, The (1991)
L.A. Slasher (2015)	Matrix, The (1999)

Most Rated: All rated 5 stars

Most Popular: Movies with the most users rated.

New Movies

1. Average Action Movie Rating: ~3
 - Never be recommended...
2. User Favorite Genre weight?

User 1 Top Watched Genres → Recommend Action Movies

action adventure sci-fi	11
comedy	11
action adventure thriller	8
action drama war	8
comedy drama	6

Conclusion

Best Performing Model: KnnBaseline (.8708)

Cold Starts: Content Filtering and other method are needed for new users and items.

Two filtering are needed for a recommendation system

- Collaborative for current users and movies
- Content for new users and movies

Limitations

21 users → 26,663 of the total reviews (25%)

- Bias toward certain genres, series, age, etc.

5 - 10% write reviews unprompted

- More likely to leave negative reviews if any.

Review Bombing

- A group of people leaving a large amount of negative reviews to negatively impact the movies

Next Steps

Factor Genres

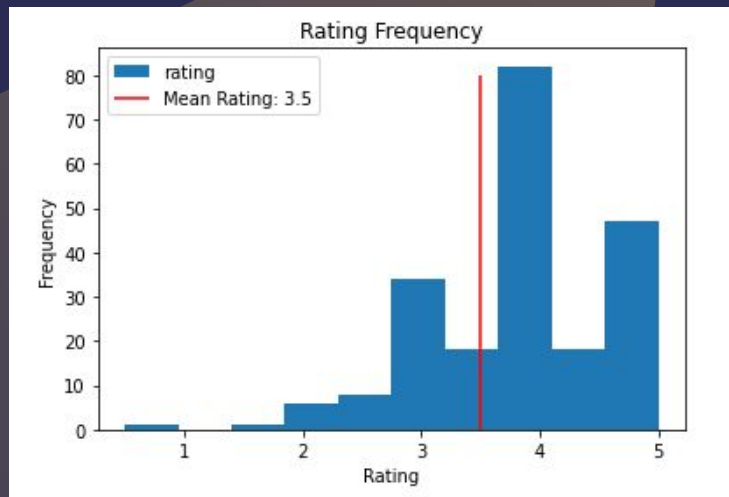
➤ Certain combination might be favorable

More Data

➤ More data = Better modeling

Diverse Users

➤ More variety in ratings



Questions?

Original Source:

<https://grouplens.org/datasets/movielens/latest/>

Email: phungtommy109@gmail.com

Github: <https://github.com/Tommyphung1>

Notebook: https://github.com/Tommyphung1/Project_4