MOVIE RECOMMENDATION SYSTEM

Laura Miller

SUMMARY

The global video streaming market size was estimated at USD 50.11 billion in 2020 and is expected to reach 224 billion by 2028. These figures seem promising for new players so long as they can find their niche.

Project output is a simple recommender that:

- incorporates content-based filtering & collaborative filtering
- can work well as part of a larger hybrid recommendation system

OUTLINE

BUSINESS PROBLEM

Telebi is planning to enter the video streaming market and is developing recommendation system approaches prior to its launch. They are testing algorithms using MovieLens data made available by GroupLens.
Recommendation system should:

- **Address “cold start problem” of recommending films to new users**
- **Predict user ratings with collaborative filtering in order to provide recommendations**

METHODS

01 SCRUB

Address duplicates. Prepare text data for NLP.

02 EDA

Explore data. Descriptive statistics & visualization.

03 NLP

Content-based filtering using text similarity.

04 MODEL

Collaborative filtering. Select and tune best model.

05 EVALUATE

Examine quality of recommendations and performance metrics.

06 IMPLEMENT

Test how content-based filtering & collaborative filtering can work for new user.

DATA

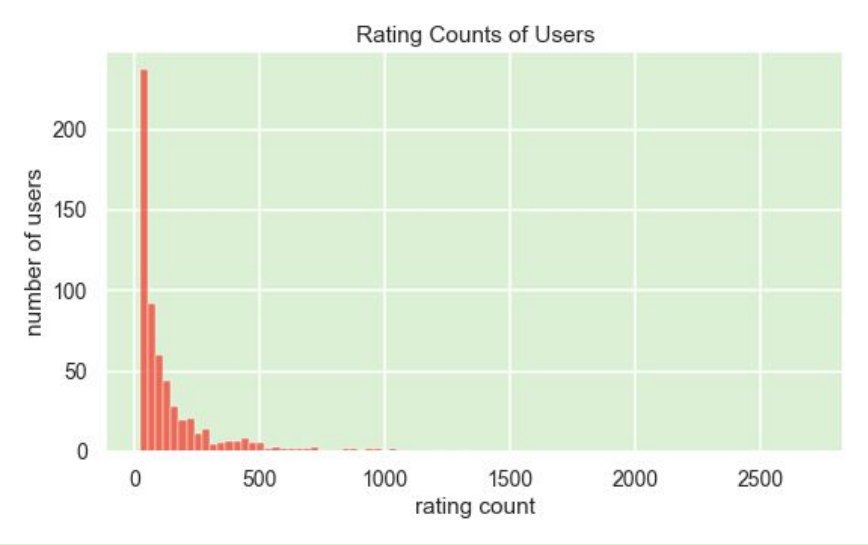
MovieLens “small” dataset from GroupLens research lab at University of Minnesota

Ratings
- 100K ratings by 610 users for 9,000 movies

Tags
- 1,590 tags by 58 users for 1,570 movies

Movies
- Movie titles and genre info

DATA



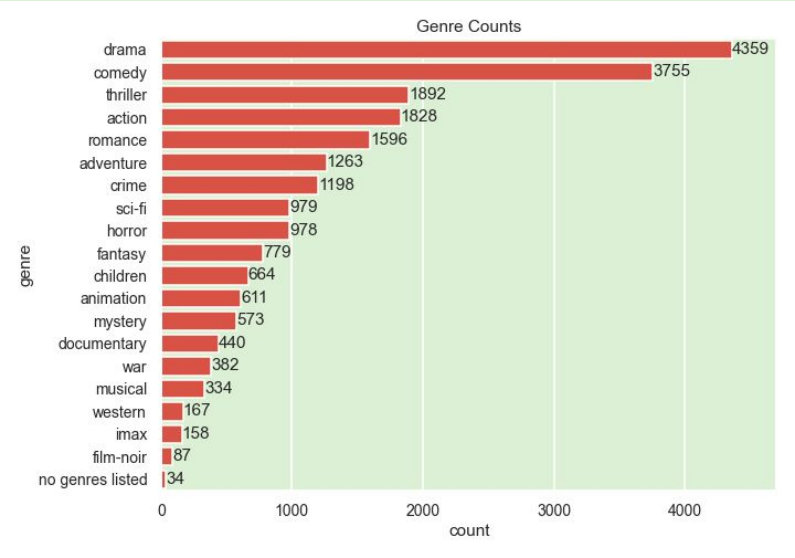
Users who have rated more films could be more engaged or have older accounts

20
Minimum # of ratings per user

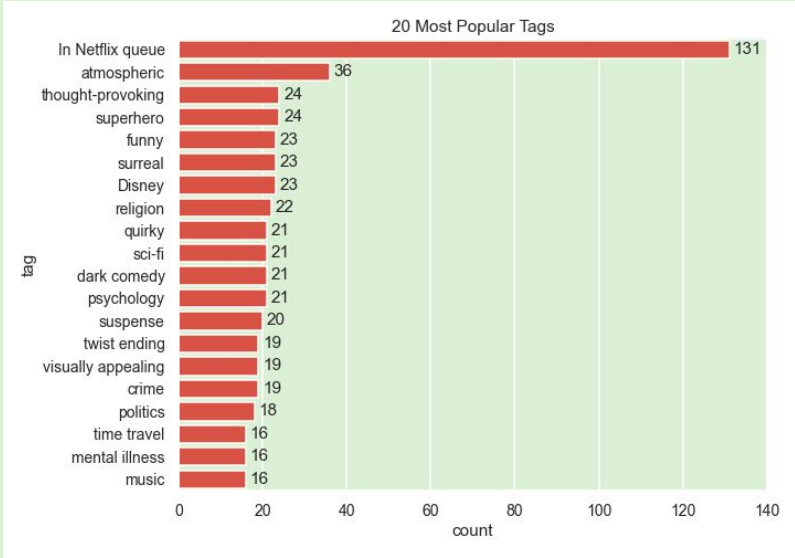
2698
Max # of ratings per user

70
Median # of ratings per user

DATA



Drama and comedy are the most common genre labels



“In Netflix queue” is most popular tag. Other tags are more descriptive

RESULTS: COMPONENTS

CONTENT-BASED FILTERING

Cold start solution using NLP. For given film, returns top n recommendations based on text similarity (cosine similarities between TF-IDF word vectors).

COLLABORATIVE FILTERING

BaselineOnly prediction algorithm. Returns top n recommendations based on predicted ratings.

RMSE = 0.8647

RESULTS: CONTENT-BASED FILTER

Movies like “Toy Story (1995)”

- “Toy Story (1995)”
- "Bug's Life, A (1998)"
- “Toy Story 2 (1999)”
- “Guardians of the Galaxy 2 (2017)”
- “Monsters, Inc. (2001)”
- “Turbo (2013)”

“Toy Story” BOW:
adventure animation children comedy fantasy pixar pixar fun

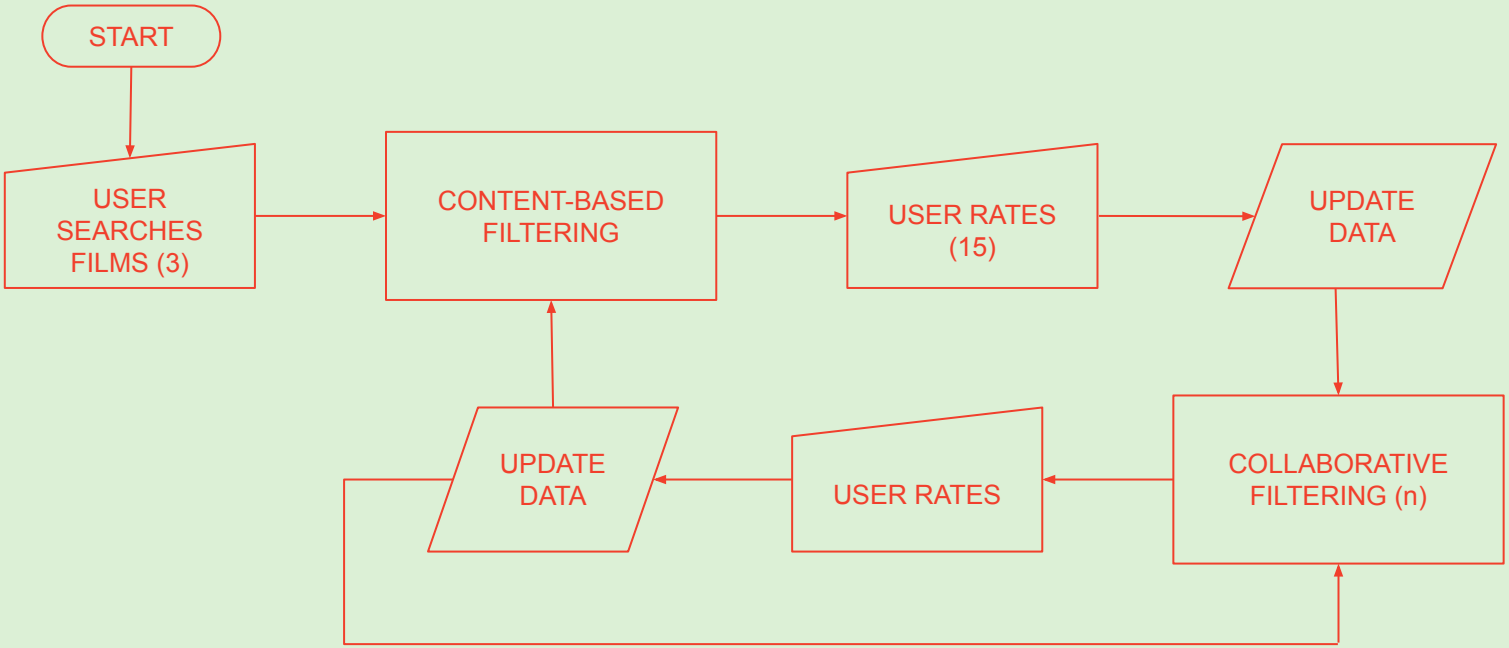
Movies like: “American Psycho (2000)”

- “Saw VI (2009)”
- "Bird with the Crystal Plumage, The (Uccello dalle piume di cristallo, L') (1970)"
- “American Psycho (2000)”
- “Book of Shadows: Blair Witch 2 (2000)”
- “Testament of Dr. Mabuse, The (Das Testament des Dr. Mabuse) (1933)”
- “From Hell (2001)”

“American Psycho” BOW:
crime horror mystery thriller

Results limited by descriptiveness of BOW. Consider adding synopses.

RESULTS: SYSTEM FLOW



CONCLUSIONS

Achieved recommendation system that uses content-based filtering and collaborative filtering (RMSE ~ 0.86) to recommend films to both new and current users. Telebi should strongly encourage users to rate every film they watch and provide descriptive tags.

Future work:

- Improve content-based recommendations with more text data (tags + plot synopses) and more visual/creative UI

THANKS!

Does anyone have any questions?

Email: laura@warrahdesign.com
GitHub: @NariMo91