Hybrid Movie Recommendation System (Approach-1)

Group Members:

1. Abhishek Pratap Singh (23911008)
2. Chanda Gupta (23911009)
3. Sanjeev Saharan (23911013)

Problem Statement:

Build a movie recommendation system based on ‘MovieLens’ dataset.

Hybrid merges both Collaborative Filtering and Content Based Filtering by reducing the limitations of them individually.

DataSet:

MovieLens review dataset (ml-latest-small)

○ Ratings: 100k

○ Movies: 9k

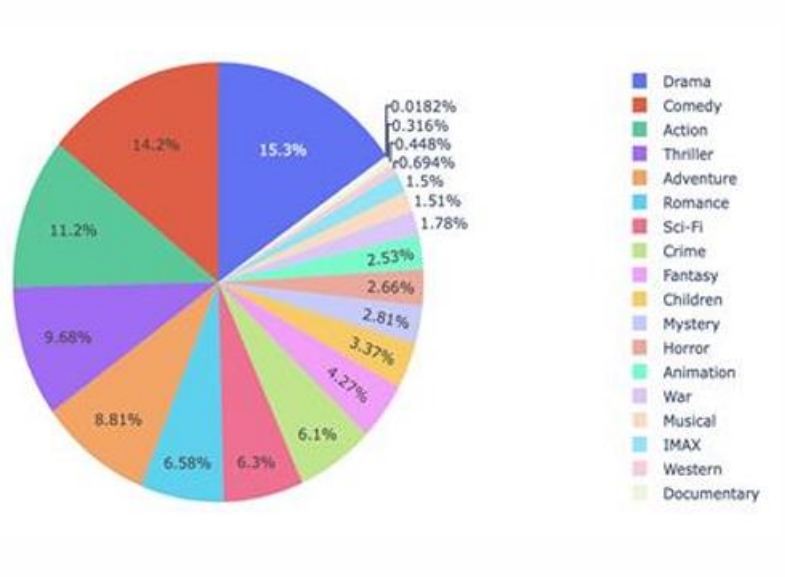
○ Users: 600

Models:

1. Popularity based model
2. Content based model
3. Collaborative Filtering
4. Matrix Factorization method
5. Combined model (SVD + CF)
6. Hybrid model

Split the dataset into 80% training and 20% testing based on the User ID.

Genre Distribution:



1. **Popularity based model:**

- Genre wise popular movies
- Computed on:
    - Popularity metric from TMDB data
    - Weighted Rating from IMDB

W= (v/(v+m) * R) + (m/(m+v) * C)
where, W=Weighted Rating
R = Average rating of a movie (scale: 1-10)
v = number of votes for the movie
m = minimum votes required to be listed in top
C = Mean vote average

Action Movies:

	title	popularity
788	Deadpool	514.569956
94	Guardians of the Galaxy	481.098624
127	Mad Max: Fury Road	434.278564
28	Jurassic World	418.708552
199	Pirates of the Caribbean: The Curse of the Bla...	271.972889

Animated Movies:

	title	popularity
88	Big Hero 6	203.734590
124	Frozen	165.125366
506	Despicable Me 2	136.886704
77	Inside Out	128.655964

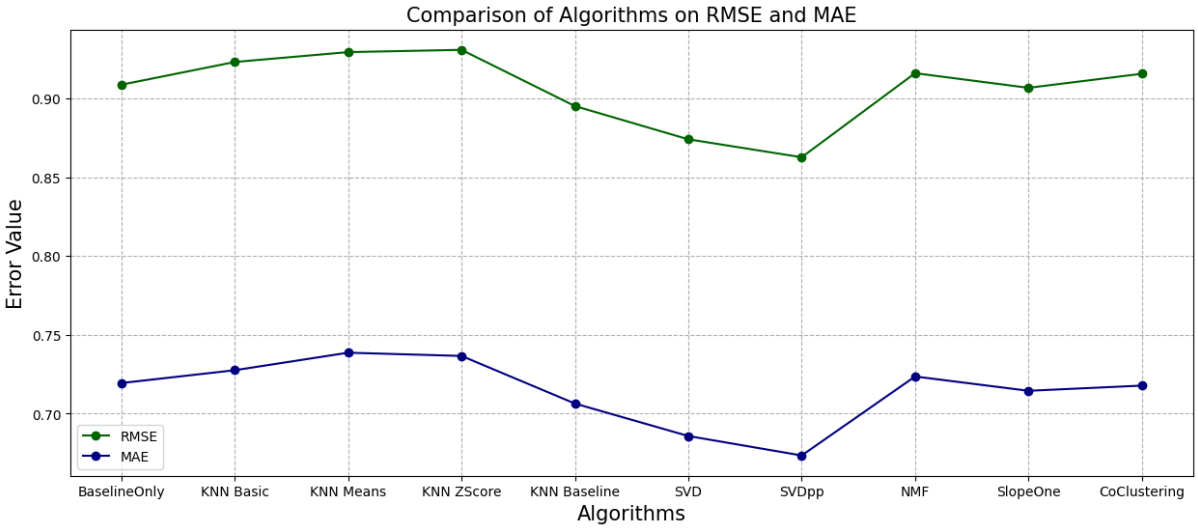
2- Content-Based Recommendation:

1. User profile based on item profiles
    a. Genre
    b. Year of release of movie
2. Movie - Movie similarity

Metric	Content-based (Genre)
RMSE	0.935837
MAE	0.717817

3- Collaborative Filtering:

- KNN (k-nearest neighbours) algorithm using Surprise library.
- Variations of KNN based approaches:
    - KNNBasic
    - KNNwithMeans
    - KNNWithZScore
    - KNNBaseline: integrates the baseline estimate rating.
- **Similarity metrics:**
    - Cosine similarity
    - Mean square difference-based similarity.
    - Pearson coefficient (mean-centred cosine similarity)
    - Pearson Baseline (uses global baselines for centring instead of means)



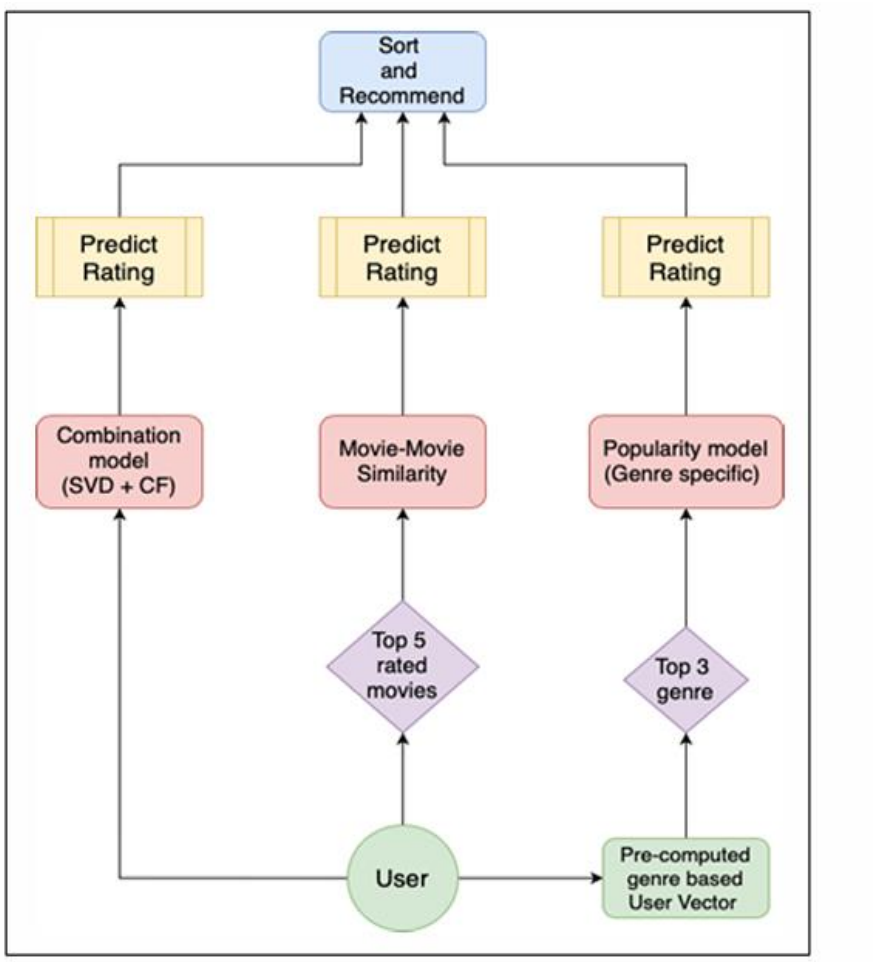
4 - CF and Latent Factor models (Combined Model):

- Matrix Factorization + CF
- Weighted linear combination of prediction ratings.
- Combined:
  - KNNBaseline (with pearson baseline similarity)
  - SVD
  - SVDpp : This model is an extension of SVD which considers implicit ratings.
  - BaselineOnly (takes the baseline predictor means of global ratings, user and the item/movie)

5- Hybrid Model:

Combination of recommendations using:

1. Combined model (SVD + CF)
2. Content Based Movie-Movie Similarity
3. Popularity model + User Profile (Genre Based)



Flow of the Hybrid recommendation model

REFERENCES

1-[Adomavicius G.Adomavicius G.[n.d.]] A. Adomavicius G., Tuzhilin. [n.d.]. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.. In Regional Conference Series in Mathematics.

2-[FermanFerman2002] Errico J. H. Beek P. V. Sezan M. I. Ferman, A. M. 2002. Content-based filtering and personalization using structured meta data. In Regional Conference Series in Mathematics.

3-[HugHug2017] Nicolas Hug. 2017. Surprise, a Python library for recommender systems. http://surpriselib.com.