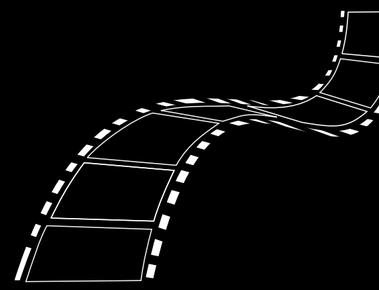
Movie Recommendation System

AU1940123 Yashvi Navadia
AU1940161 Kashish Jivani
AU1940165 Neel Popat



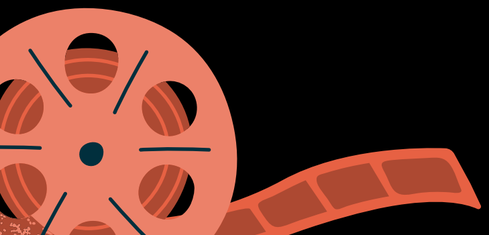
Introduction

Movie Recommendation system :

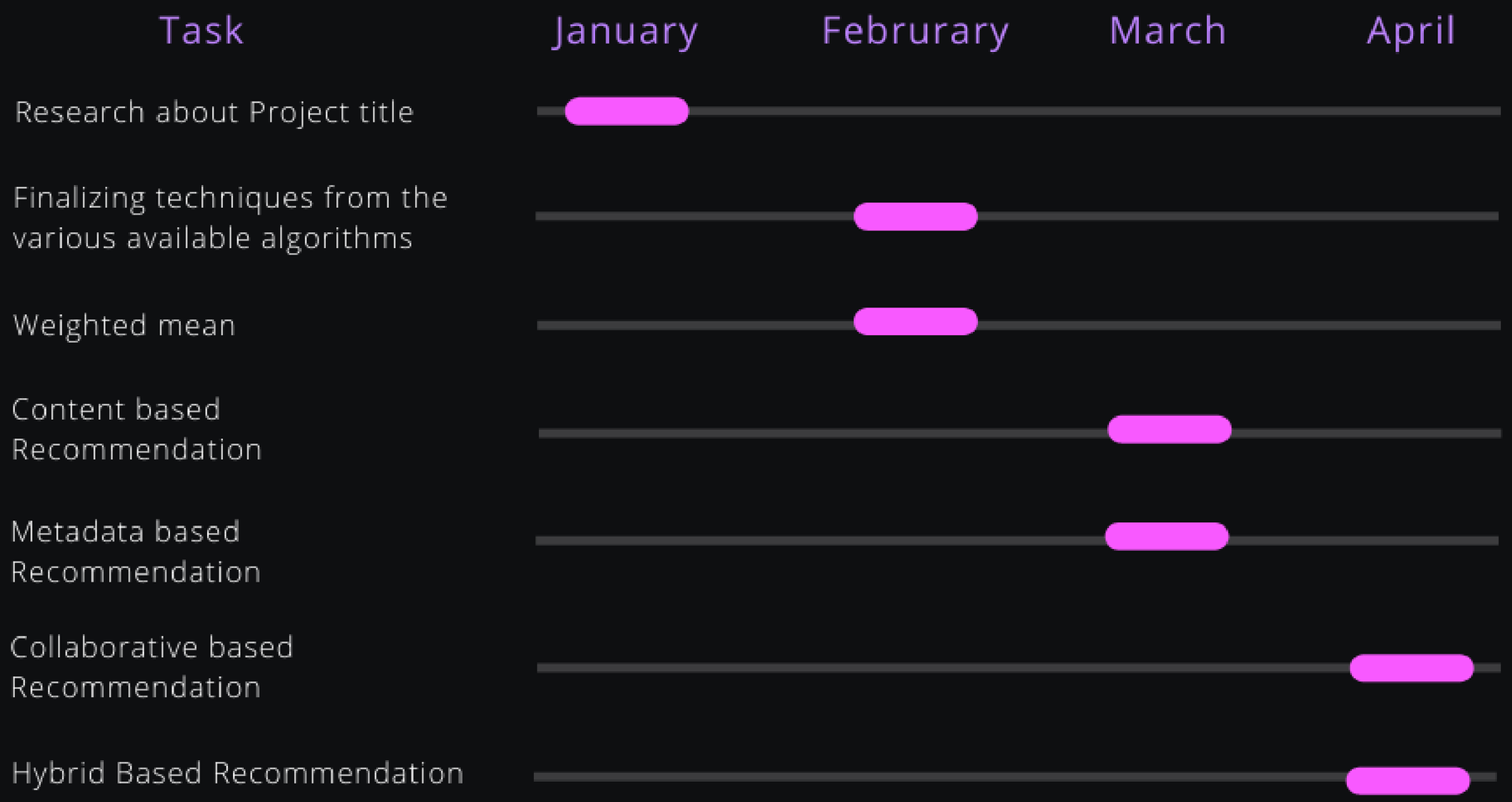
- Various OTT (over the top) platforms uses a recommendation system for its users making the interface user-oriented.
- Users spends a lot of time in finding what they might like because they have 1000s of options.
- Hence it is important to have such an algorithm which helps the users to find their interests.

Problem Statement

The project aims to assist the users in finding content which he/she likes hence reducing the time spent by the users to find appropriate content.



Progress Report



Existing body of work

- Currently, all the OTT platforms, google, youtube and all kinds of search engines use recommendation strategy to retain their users for maximum amount of time. Below are some articles and papers which have information regarding various recommendation systems.

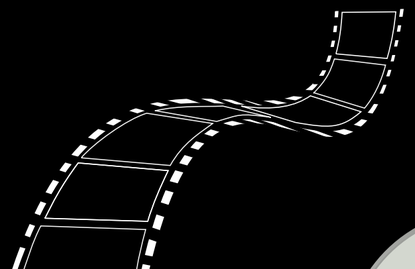
- S. India, “How netflix's recommendation Engine Works?,” Medium, 05-Nov-2019. [Online]. Available: https://medium.com/@springboard_ind/how-netflixs-recommendation-engine-works-bd1ee381bf81. [Accessed: 24-Apr-2022].

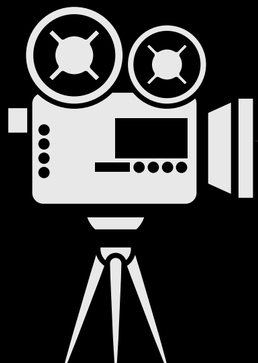
- Netflix, “Netflix prize data,” Kaggle, 13-Nov-2019. [Online]. Available: https://www.kaggle.com/netflix-inc/netflix-prize-data. [Accessed: 30-Mar-2022].

Data set

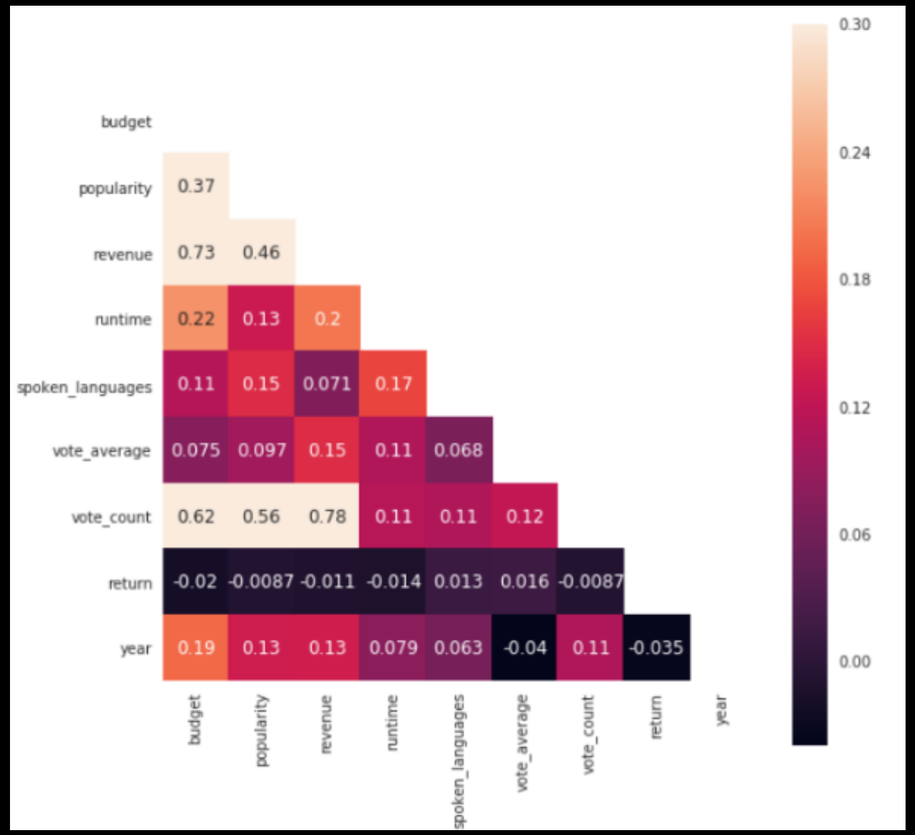
We have used two MovieLens dataset.

- **The Full Dataset:** 26,000,000 ratings, 45,000 movies, 270,000 users.
- **The Small Dataset:** 100,000 ratings, 9,000 movies, 700 users.

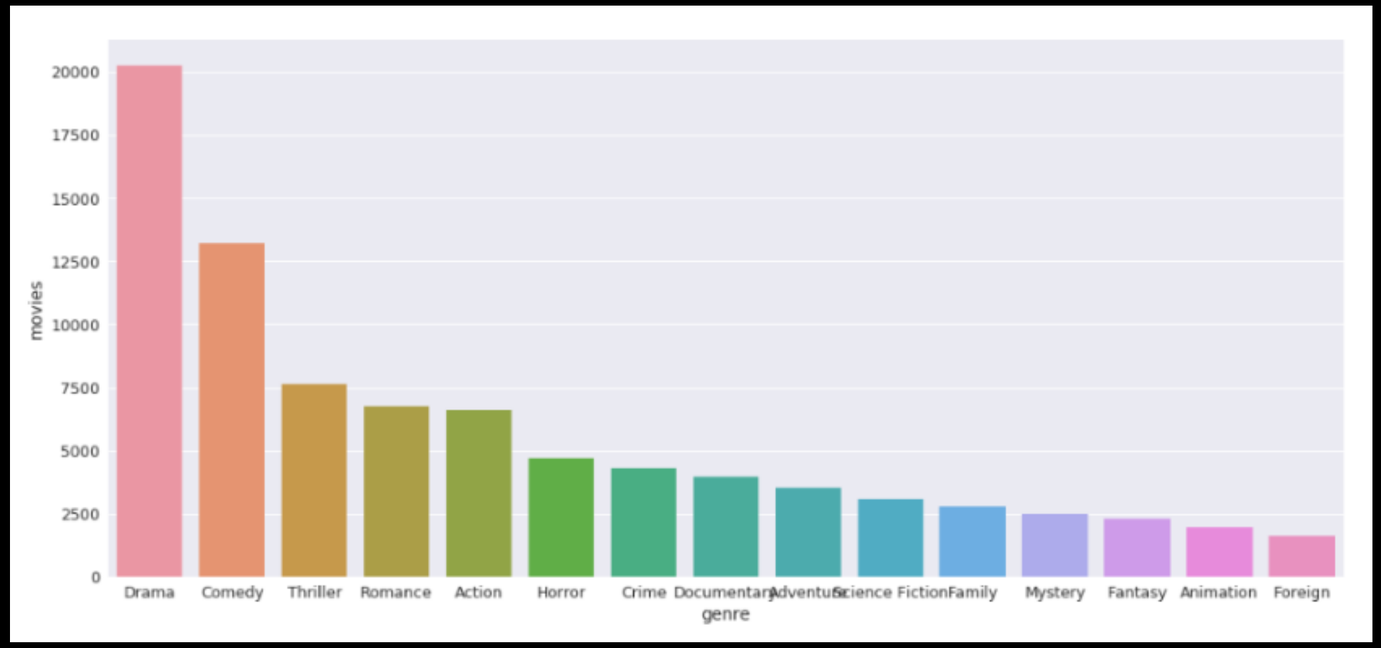




EDA



Analysis on budget, popularity, revenue, runtime, etc.



Analysis on different kinds of Genres

APPROACH

1

WEIGHTED MEAN

- Set priorities to movies to give recommendations.

2

CONTENT BASED FILTERING

- Consideration of overview,tagline, cast and director,etc

3

COLLABORATIVE FILTERING

- Estimate the movie rating using SVD.

4

HYBRID RECOMMMENDER

- Hybrid Implementation of user specific and content based recommendation engine

Weighted mean

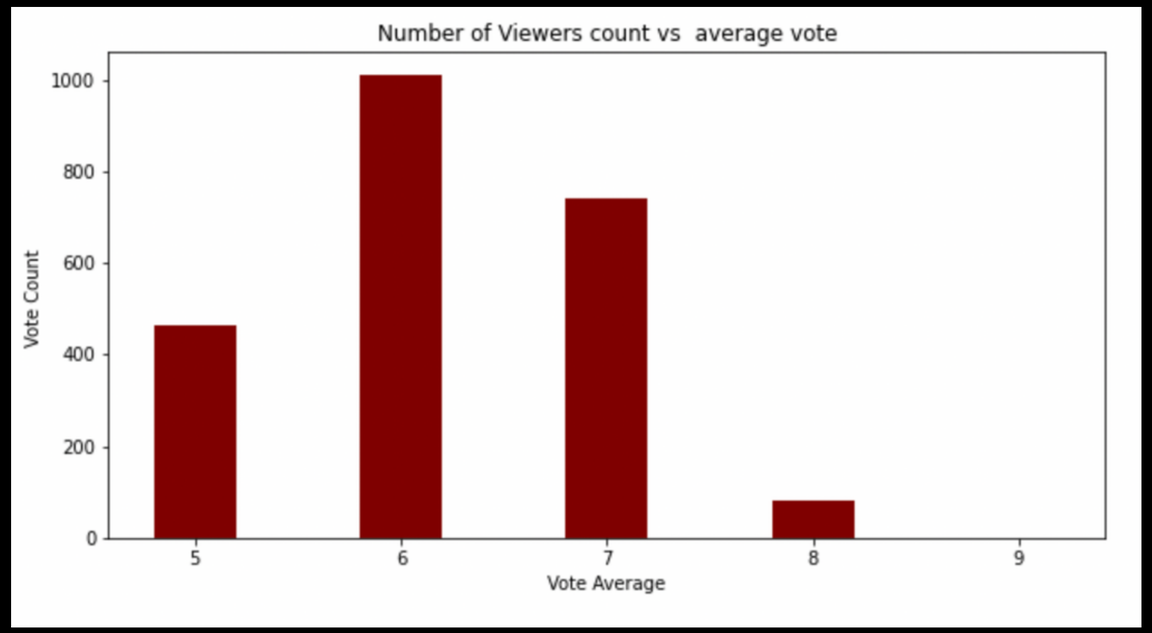
Weighted Rating (WR) = $(\frac{v}{v+m}.R) + (\frac{m}{v+m}.C)$

where,

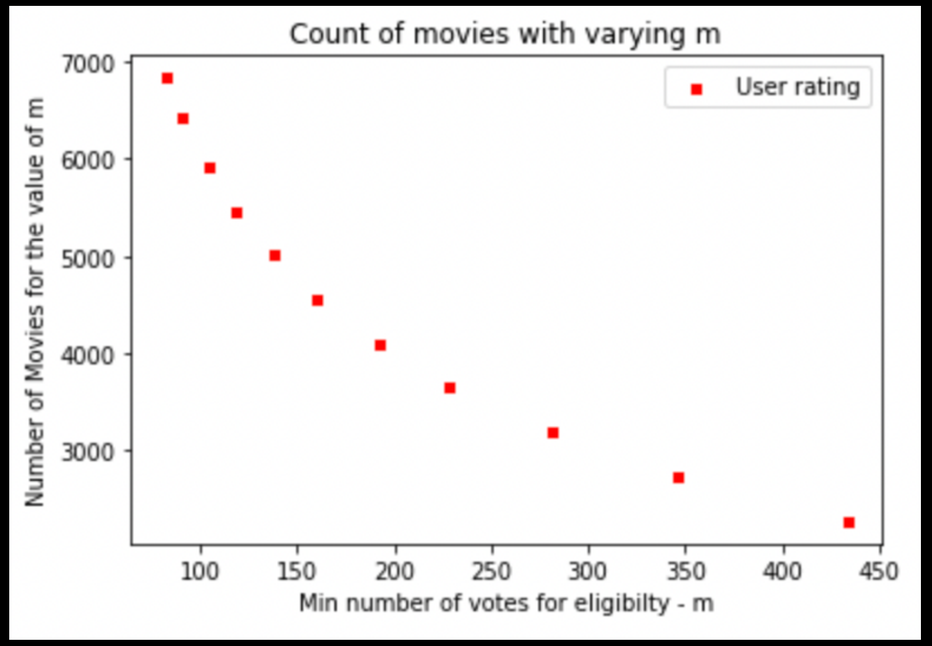
- v is the number of votes for the movie
- m is the minimum votes required to be listed in the chart
- R is the average rating of the movie
- C is the mean vote across the whole report

- It creates one paramter "weighted rating" which considers all 4 parameters above for a movie.
- Does not give personalized recommendations.

Analysis on parameters of Weighted mean



- The graph represents the 0.85 percentile of the data with the respective vote average and corresponding vote count.



- The graph represents the number of movies varying as the hyper-parameter m varies.

CONTENT BASED FILTERING

- As we do not have a quantitative metric to judge our machine's performance, we have done it qualitatively.
- Therefore we have used TF-IDF Vectorizer to transform the text data into a vector.
- Cosine similarity calculates a numeric quantity that denotes the similarity between two movies.

$$cosine(x, y) = \frac{x.yT}{||x||.||y||}$$

- We will now have a pairwise cosine similarity matrix for all the movies in our dataset and can list down similar movies by entering any movie of our choice.

Results of Content Filtering

Movie : 'The Dark Knight'

Sr.No	Movie
7931	The Dark Knight Rises
132	Batman Forever
1113	Batman Returns
8227	Batman: The Dark Knight Returns, Part 2
7565	Batman: Under the Red Hood
524	Batman
7901	Batman: Year One

INFERENCE :

- Recommendations based on Overview, Tagline

Sr.No	Movie
8031	The Dark Knight Rises
6218	Batman Begins
6623	The Prestige
2085	Following
7648	Inception
4145	Insomnia
3381	Memento

INFERENCE :

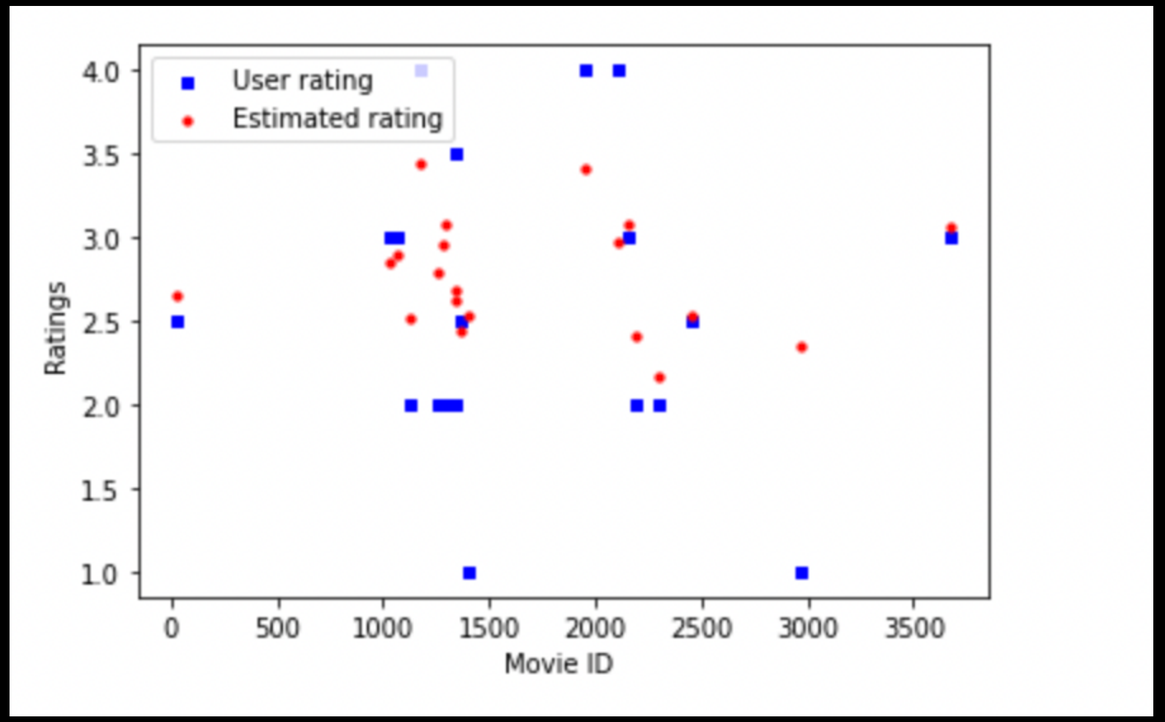
- Recommendations based on Cast, Crew, Keyword, Genre

Collaborative filtering

- Solution for the issue of bias towards specific users.
- Collaborative filtering makes use of SVD algorithm.

- Input : User id , Movie id
- Output : Estimated rating

- The RMSE obtained was less than 1 and the engine gave estimated ratings for a given user and movie.



- comparing observed user rating vs. estimated user rating for a particular movie ID for "**User 1**".

Hybrid Recommender system

- Merging of the content based recommender by adding the estimated score of the specified user.
- Provision of the user id of user and a movie name makes recommendation of similar movies sorted on the basis of expected ratings by that particular user.

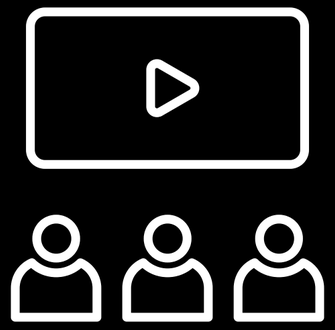
movie: "Avatar"

User id : 1

Movie id	Title	estimated rating
1011	The Terminator	3.083
522	Terminator 2: Judgment Day	2.947
8658	X-Men: Days of Future Past	2.935

User id : 500

Movie id	TItle	Estimated rating
8401	Star Trek Into Darkness	3.238
974	Aliens	3.203
7265	Dragonball Evolution	3.195



Contribution of each member

Kashish Jivani : Implemented cosine similarity concept in Content filtering. Compared two ouputs of content based filtering and derived appropriate conclusion from that. Helped in initial research work required for implementation of methods used.

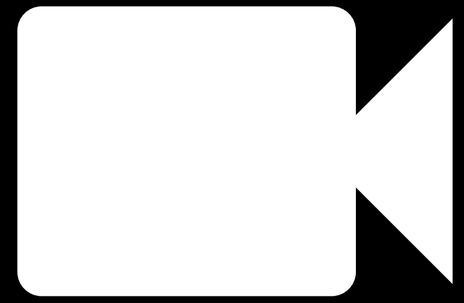
Neel Popat : implemented SVD in collaborative filtering and also implemented hybrid recommmender. Played a major role in coding and analyzing the methods using mathematical approach.

Yashvi Navadia : Focused more on analysis part including graphs and results of various concepts being used in the system. Helped in the implementation of hybrid recommender and in deriving appropriate results.

References

- C. Z. Omega and Hendry, "Movie Recommendation System using Weighted Average Approach," 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech), 2021, pp. 105-109, doi: 10.1109/ICITech50181.2021.9590147.

- S. Agrawal and P. Jain, "An improved approach for movie recommendation system," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 336-342, doi: 10.1109/I-SMAC.2017.8058367.

- C. M. Wu, D. Garg and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 11-15, doi: 10.1109/ICSESS.2018.8663822.

Thank you