International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Movie Recommendation System

Abhiraj Agrawal, Assistant Professor Mr. Rudrendra Bahadur Singh, Ankit Dwivedi, Dhruv Rishi
Dept of Computer Science and Engineering
B.Tech , Babu Banarasi Das National Institute of Technology and Management

Abstract

Recommendation System is a major area which is very popular and useful for people to take proper automated decisions. It is a method that helps user to find out the information which is beneficial to him/her from variety of data available. When it comes to Movie Recommendation System, recommendation is done based on similarity between users (Collaborative Filtering) or by considering particular user's activity (Content Based Filtering) which he wants to engage with. To overcome the limitations of collaborative and content based filtering generally, combination of collaborative and content based filtering is used so that a better recommendation system can be developed. Also various similarity measures are used to find out similarity between users for recommendation. In this paper, we have surveyed state-of-the-art methods of Content Based Filtering, Collaborative Filtering, Hybrid Approach and Deep Learning Based Methods for movie recommendation. We have also reviewed different similarity measures. Various companies like facebook which recommends friends, LinkedIn which recommends job, Pandora recommends music, Netflix recommends movies, Amazon recommends products etc. use recommendation system to increase their profit and also benefit their customers. This paper mainly concentrates on the brief review of the different techniques and its methods for movie recommendation, so that research in recommendation system can be explored. .

Keywords—Recommender Systems; Content-Based Filtering; Movie recommender; TF-IDF; Cosine Similarity; Vector Space Mode

1. Introduction

Recommendation System is an information tool which helps users to find out the items which they want from the large no of items available , . Main goal of recommendation system is to forecast the rating which a specific user gives to an item. It helps the user to find the best solution from the available list of items. Many companies use recommendation system so that they can serve their user and raise their profit like Netflix, YouTube, Amazon and others Still now it is a good topic of research because to find what the user wants from available resource is a big challenge, as our choice keeps on changing with time. Nowadays what we purchase online is recommendation. For example, if we want to buy books, listen music, watch movies etc there is one recommendation system that is working in background which suggest the user based on his previous actions . Many platforms like Netflix which suggest movies, Amazon which suggest products, Spotify that suggest music, LinkedIn that is used for recommending jobs or any social networking sites which suggest users, all these work on recommendation system ,. By using these recommendation engine users can easily find out what he wants according to his/her choice. So to build an effective recommender system is also a challenge because user's preference keeps on changing with time.

1.1 Applications

Recommendation System is a vast area which is used everywhere in every field. People use recommendations as it saves time, so it plays a vital role in various areas. It is used in many real life applications like Entertainment, E-Commerce, Services, Social Media etc.. In Entertainment area recommendation system is widely used in watching movies or listening music or any TV program. When we talk about E-Commerce field, Amazon is the world's largest shopping site. Some use it for purchasing books, for buying any household products or any products, some use it for clothing. So this way whole world is dependent on these E-commerce sites for one or the other work. Some other E-Commerce sites like Flip cart, E bay, Myntra, Shop clues etc. also provides recommendations. Some other applications of recommendation system are listed below:

• Movie Recommendation: Netflix uses algorithm for recommending movies according to their interest. Other such platforms that provide recommendations include hotstar, sonyLIV, voot, ALTBalaji etc

• Music Recommendation: Pandora generates a radio station. It uses the properties of songs to recommend other songs. Other medium in this field that suggest music recommendation are Spotify, JioSavan, Gaana etc .

• News: Various applications that provide news recommendation can be Google News, Apple News(integrated into IOS and macOS), Flip board, Feedly, Tweet Deck, Pocket, Mix, Zig, News360. All these suggest news, articles, blog post, content from top publishers etc .

• Fashion: People can buy various clothing items of their choice. This section include various shopping sites like Myntra, Amazon, Club Factory, SHEIN, Lime Road, Flip cart and others .

• Travel service: Recommendation helps here to suggest various travelling sites to safeguard journey. This includes Road trippers which leads you plan any road trip with ease. Using Hooper, users can input their travel plans, and the app will tell them when is the best time to book their flight .

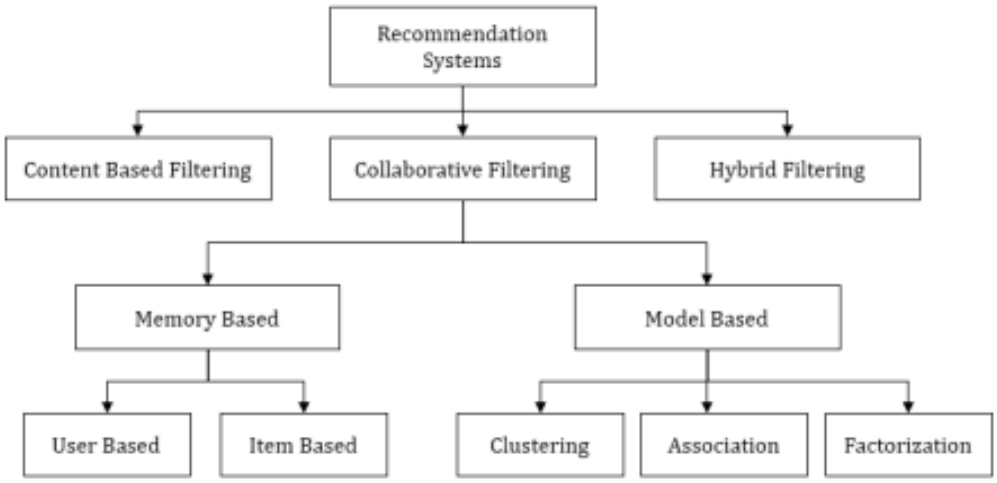
1.2 Challenges

There are various challenges faced by Recommendation System. These challenges are Cold Start problem, Data Sparsity, Scalability.

Cold Start Problem: It needs enough users in the system to find a match. For instance, if we want to find similar user or similar item, we match them with the set of available users or items. At initial stage for a new user, his profile is empty as he has not rated any item and the system do not know about his taste, so it becomes difficult for a system to provide him recommendation about any item. Same case can be with new item, as it is not rated by any user because it's new for the user. Both these problem can be resolved by implementing hybrid techniques .

Data Sparsity: The user or rating matrix is very sparse. It is very hard to find users that have rated the same items because most of the user does not rate the items. So it becomes hard to find set of users who rate the items. To give recommendation is really tough when there is less information about any user .

Scalability: Collaborative Filtering use massive amount of data to make reliable better which require more number of resources. As information grows exponentially processing becomes expensive and inaccurate result from this Big data challenge



1.3 Related Work

- Reference [11] worked on a restaurant recommender system that was based on case-based recommendation technique. The adopted technique was used to select and rank restaurants. It was implemented to serve as a guide to attendees of the 1996 democratic national convention in Chicago and operated as a web utility.

- Reference [12] applied content-based technique in paper recommendation system. The author used Jaccard similarity coefficient or jaccard index to compute similarity between users' query (users' attributes) and the attributes of the papers.The recommendations suggested by the system were sent via emails to the intended users.

- Reference [13] designed a group recommender system for Facebook. He used hierarchical clustering and decision techniques to suggest or recommend the most suitable Facebook group (s) to Facebook users. He extracted profile information of the Facebook members at University of North Texas and used it as a test data.

- Facebook recommendation system provides friends recommendations or suggests friends as "people you may know". These suggestions or recommendations are based on mutual friends, work and educational information, groups you are part of, contacts you have imported using friends finder and many other factors. This recommendation system uses facebook users' profile [14].

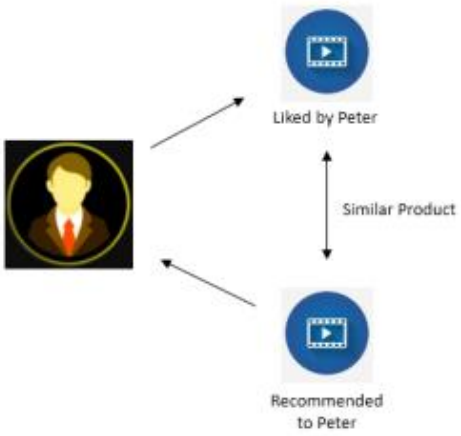
- Amazon's customers who bought, CDNOW.com's Album Advisor, MovieFinder.com's Match Maker, and Reel.com's Match Maker use item to item correlation as recommendation technology to provide recommendations to their customers. Amazon's customers who bought feature recommend products to its customers. CDNOW.com's Album Advisor suggests music to its customers. MovieFinder.com's Match Maker, and Reel.com's Match Maker recommend Videos to their customers

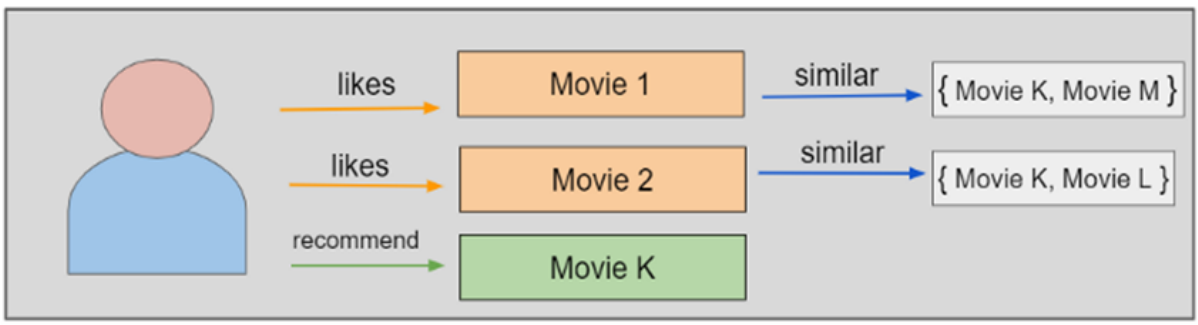
2.Proposed Work

Approach:--Algorithm we have used for recommendation system is

Content-Based Filtering

Content-Based Filtering are also known as cognitive filtering . This filtering recommends item to the user based on his past experience. For example, if a user likes only action movies then the system predicts him only action movies similar to it which he has highly rated. The broader explanation could be suppose the user likes only politics related content so the system suggests the websites, blogs or the news similar to that content. Unlike collaborative filtering, content based filtering do not face new user problem. It does not have other user interaction in it. It only deals with particular user's interest. Content based filtering first checks the user preference and then suggest him with the movies or any other product to him. It only focus on single user's ideas, thoughts and give prediction based on his interest. So if we talk about movies, then the content based filtering technique checks the rating given by the user. The approach checks which movies are given





high ratings by the user by checking the genre categories in the user profile. After analysing user profile, the technique recommends movies to user according to his taste. Content based filtering whole process is shown by giving an example of Geometric Shapes . Here in the figure, first an Item Profile is developed based on the liking of the user. Here the user likes circle and triangle of blue colour. Now based on item profile, user profile is build. This user profile is generated by getting the data from item profile. As we can see in item profile, user likes circle and triangle of blue colour so user profile is also having circle, triangle and blue colour. Now we will match this user profile with the collection of different shapes available. In the shapes collection, we have pentagon of blue colour then circle of yellow colour and two square of yellow colour. So here system finds which of these shapes matches with the user profile. So here blue colour pentagon matches with the user's interest. Jieun Son and Seoung Bum Kim proposed Content-based filtering for recommendation systems using multiattribute networks that contain attribute

information about item . By using all attribute based on network analysis various items are recommended and overspecialization problem is resolved. Results show that problems like sparsity and scalability are also addressed when compared with pure Content Based Filtering, Linked Open Data and Feature Weighting. For conducting experiments Movie lens dataset is used, where on a random basis 100 users are taken for experiment purpose and accuracy is also improved when compared with the above mentioned methods. In the paper, Providing Entertainment by Content-based Filtering and Semantic Reasoning in Intelligent Recommender Systems by Yolanda Blanco-Fernandez et al. solved overspecialization problem . For this, hidden semantic association between user and the product are known then applying Spreading Activation technique to detect a node that is strongly connected. There is no requirement about other user's data to make recommendations. It is easy for content based approach to recommend new items. It provides recommendation to the user with unique taste. There is no first rater problem. It also provide content feature which helps us to explain reason for an item recommended. To find any particular feature like images or movies of any specific genre, it sometimes becomes a problem. Generally it is referred as overspecialization problem. User is never recommended anything outside user profile. It is easy to miss recommending item to user as there is not enough information about that item.

Steps:

We have provided the following steps for recommending the movies:
1. Get the data of movies containing attributes title, movieid and genre.
2. Get the data from of rating containing attributes userid, movieid and rating stramp.
3. Merge these 2 datasets together
4. Create a rating dataframe with average rating and number of rating.
5. Set the number of rating column with rating

Methodology

The use of collaborative-filtering technique in recommending research papers has been criticized by some authors. Authors like suggest that collaborative-filtering technique is ineffective in domains where items (e.g. research papers) are more than users. Said; "Users are not willing to spend time to rate items explicitly". Hence, content-based approach is adopted for the design and implementation of research paper recommender system. This approach does not depend on the ratings of other users but uses the contents describing the items and the users' taste or needs. The researchers used the following data collection procedure and methods in representing the research papers, users' profile of interest, and also in providing recommendations to the users

- Dataset for the system: Sources of the movies are the dataset provided by the movieslens , and also the ones from open sources obtained on the internet. Information about users' profile of interest is collected from the users during their transactional behaviors or the usage
- Keyword-Based Vector-Space Model: The researchers used this model with basic TF-IDF weighing technique to represent a movie as a vector of weights, where each weight indicates the degree of association between a movie and a term or keyword.
- Item Representation: The items (research papers) are represented by a set of features (also called attributes or properties). This attributes are: title of the movie, abstract, keywords, research area, movie ID, and the productions. The abstract represents the movie when the frequency of a term in the movie dataset is being determined.
- TF-IDF and Cosine Similarity: The researchers used TF-IDF and cosine similarity to determine how relevant or important a research paper is to a user's query. The importance increases proportionally to the number of times a term (in the user's query) appears in the research paper. TF-IDF is given by:

  □□ − □□□(□, □,□) = □□(□, □) × □□□(□,□) (1)

  Where t=term in the user's query
  d= a movie in the collection,
  D= a collection of movies
  TF=Term Frequency given by:

  $$□□(□, □) = \frac{□□,□}{□□}$$

  IDF= Inverse Document Frequency which is given by:

  $$□□□(□,□) = □□□\ \frac{□}{|□ ∈ □:□ ∈ □|}.$$

  Where
  N = number of movies in the collection
  □□= Number of terms in the movie d
  □□,□ = Number of times term t appears in movie d

$$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \log\left(\frac{N}{\text{df}_i}\right)$$

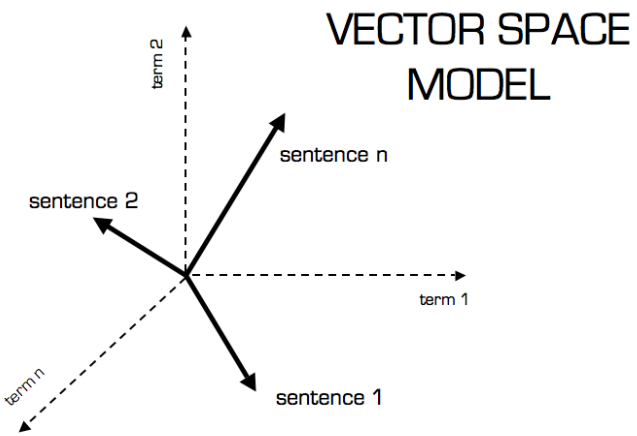
$\text{tf}_{i,j}$ = total number of occurences of i in j
df_i = total number of documents (speeches) containing i
N = total number of documents (speeches)

The Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The researchers used this method to determine how similar a research paper is to a user's query or paper that a user has liked in the past. The research papers are represented as vectors of weights, where each weight indicates the degree of association between the research papers and the Tern. Given two research papers or documents dj , dk represented as vectors of weights, their similarity is measured by:

$$\text{Sim}(d_j, d_k) = \frac{\overrightarrow{d_j}.\overrightarrow{d_k}}{|\overrightarrow{d_j}|.|\overrightarrow{d_k}|} = \frac{\sum_{i=1}^{n} w_{i,j}.w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2}\sqrt{\sum_{i=1}^{n} w_{i,k}^2}}$$

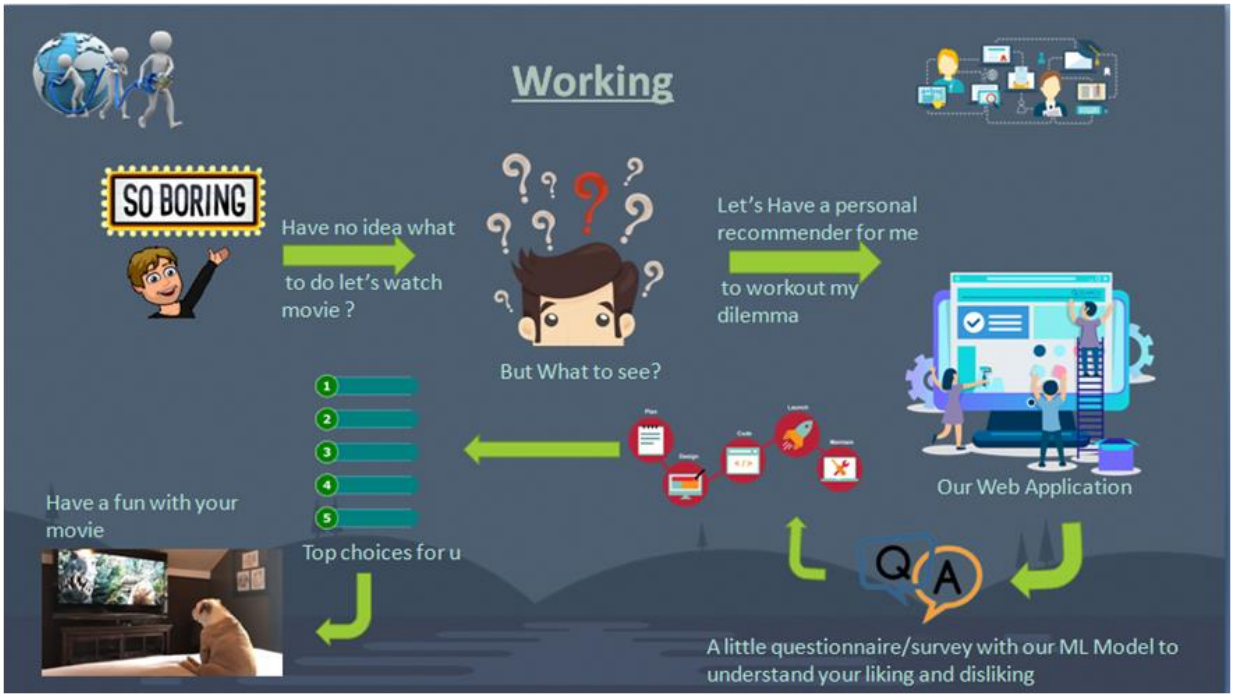
Where Wi, j = Weight of term i in movie j
Wi, k = Weight of term i in movie k

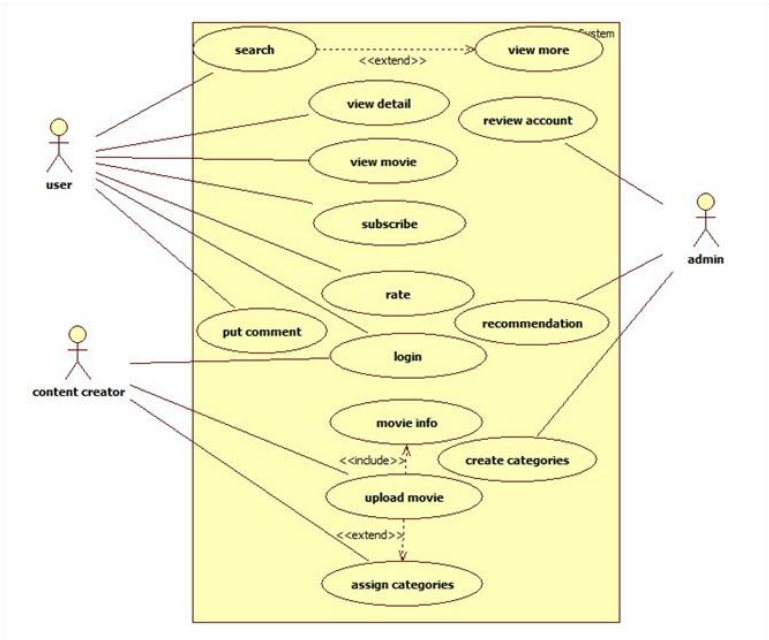
After calculating TF-IDF scores, how do we determine which items are closer to each other, rather closer to the user profile? This is accomplished using the Vector Space Model which computes the proximity based on the angle between the vectors. In this model, each item is stored as a vector of its attributes (which are also vectors) in an n-dimensional space and the angles between the vectors are calculated to determine the similarity between the vectors. Next, the user profile vectors are also created based on his actions on previous attributes of items and the similarity between an item and a user is also determined in a similar way.



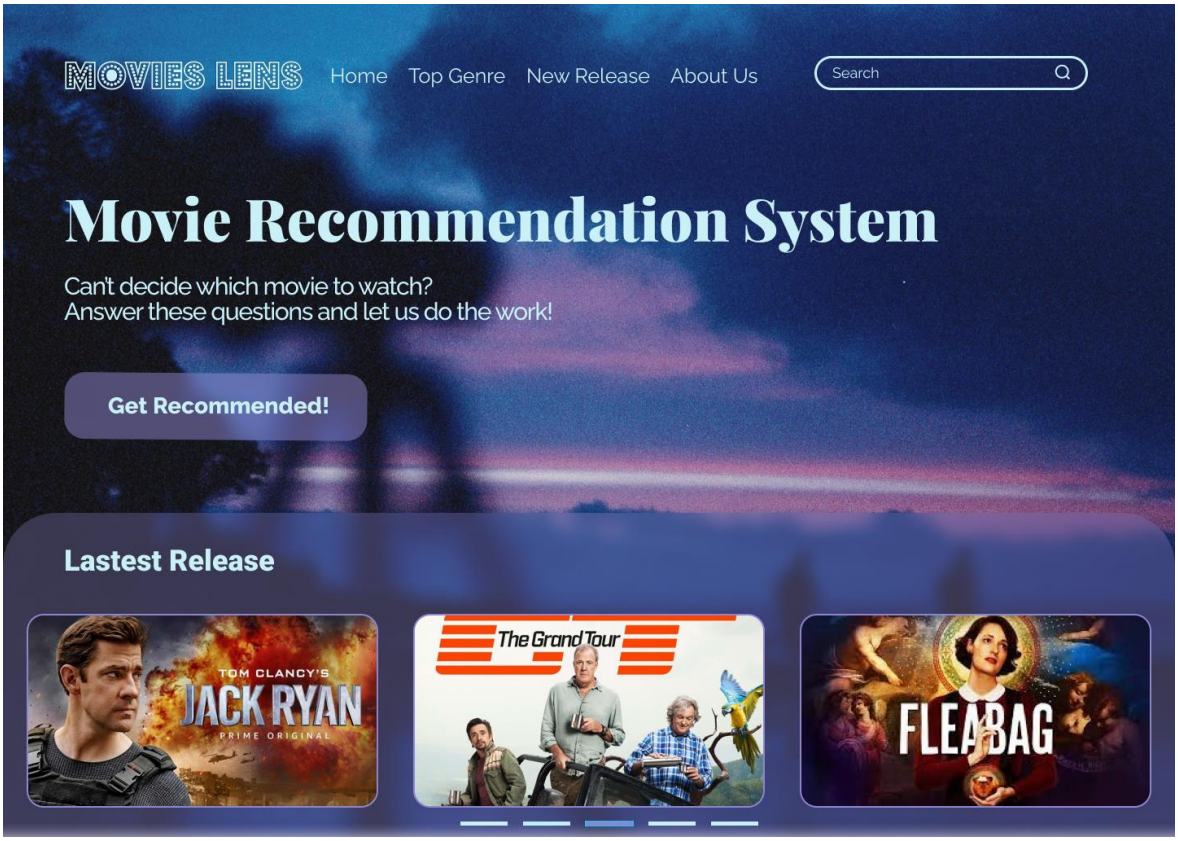
Result and Analysis

- With the completion of registration and login module, we have completed the user side and admin side entry pages of the application which results into the successful logins.
- The user are shown different movies with respective movie rating which completes basic viewing module resulting into view part from user side. The review module resulted to show the user name and their corresponding review on a particular movie. The movie list shows all the appropriate details about every movie enclosed in the list.
- Hence according to the ratings, the recommendation is provided to the logged in users. The recommendation is fully based on the good rating of other members in the clusters.

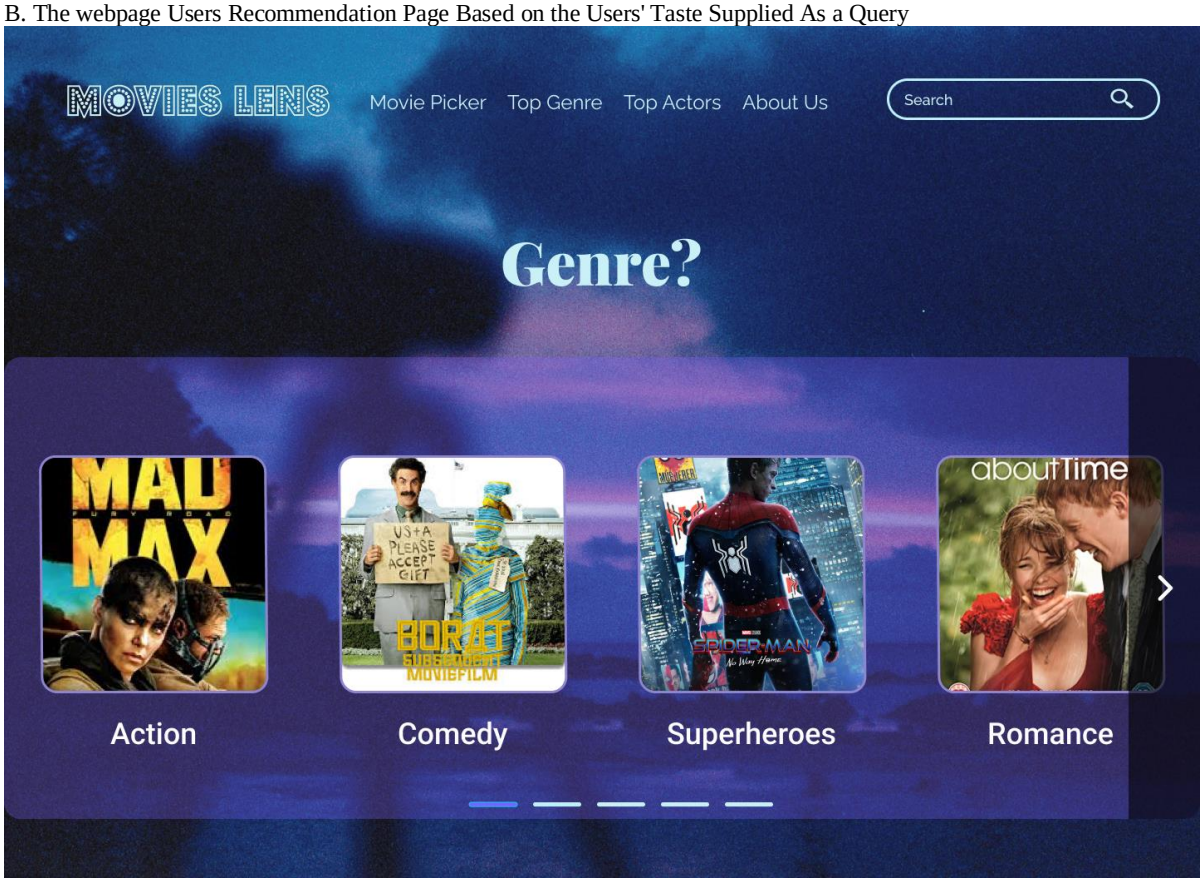


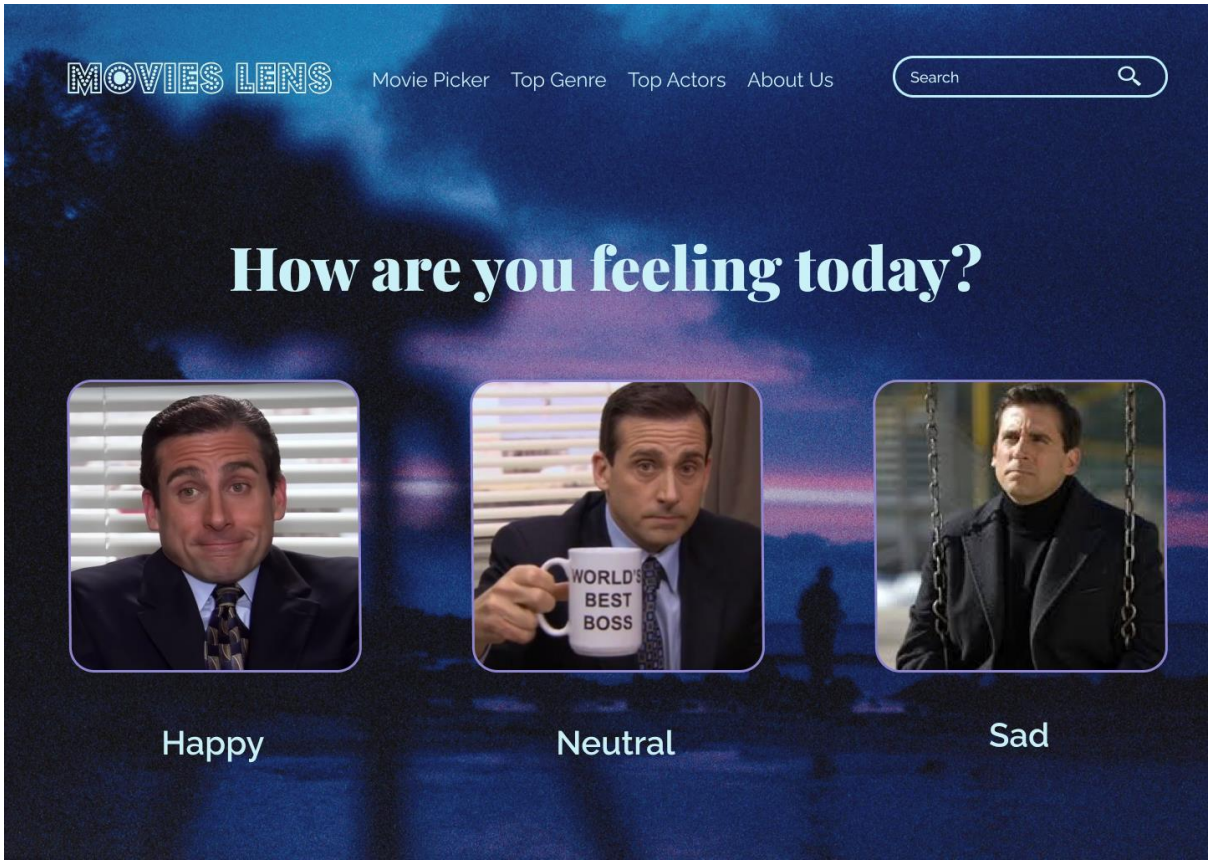


A.The movie recommendation ' Search Page

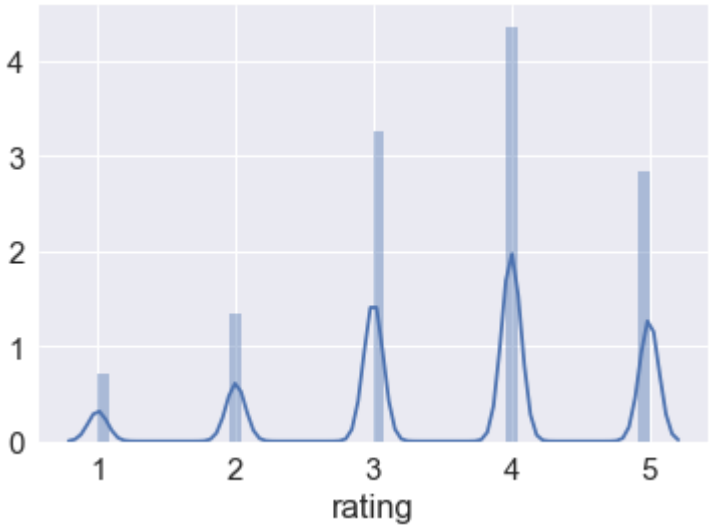


B. The webpage Users Recommendation Page Based on the Users' Taste Supplied As a Query





C. The backend dataset graphs



Facilities required for proposed work

- **Hardware Requirements:**
  Recommendation scheme provides availability to a large extent through cloud storage of your sample information. This highlights the need for a computer network interface. User should have a valid web link machine. To keep the collection synchronized with the cloud, Wi-Fi or 3G.
- **Software Requirements:**
  For the frontend web development, we use HTML, CSS, and JavaScript languages.
- **HTML:** Eventually all code in an internet implementation is converted into HTML. It is the language internet browsers comprehend and use to show user data. The knowledge of HTML by a computer developer is similar to the comprehension of a screwdriver by a carpenter. It is so essential and essential that for jobs it is often assumed.

- **CSS:** HTML provides some fundamental style alternatives, but designers need to have CSS knowledge to construct a nice frontend. CSS offers painting, templates, glitter, and buttons, tassel, lamps, and many other items that can be used to enhance internet page presentation. CSS is so widely used that textbooks were created to simplify CSS writing. These scripts are also recognized as CSS precompiled, such as Sass and LESS, but they are merely used to compose CSS CODE more effective and manageable.
- **JavaScript:** It's not just a frontend language; it's the most widely used language. JavaScript is software running on a customer machine, i.e. the computer of a user. This implies JavaScript can be used to program customer experiences quickly, intuitively and funnily without requiring a user to refresh their internet site. JavaScript can program all the drag-and-drop, infinite-scroll and photos that come to life on an internet site. JavaScript is so common that full frameworks were constructed to facilitate the construction of application frontends. For JavaScript-heavy frontends, frameworks such as Angular, Mongo DB and Node.js are all commonly used. Mongo DB database is used to arrange information in a way that enables us to search information, type information, and manipulate information in different respects. We need some manner of collecting, storing, retrieving and sorting and manipulate the data.

3 Dataset Descriptions

A dataset is just a unit to measure information released in a public open data repository. The researchers use the dataset to perform experiments. For this, they divide the dataset into training and testing set to obtain the desired result. A dataset corresponds to one or more database tables where every column of a table represents a particular variable and each row corresponds to a given record of the data. The model is run using the training set which is then compared with the target result. Then testing data is applied to provide estimation to the final model. There are various datasets available for movies which are widely available. The datasets like Movielens100k, Movielens-10M, Movielens-10k etc are used. Other dataset are IMDB, Netflix dataset, Film Trust dataset etc. The movie dataset contains the fields like user id, item id, ratings, tags, timestamps etc.

Name	Movies	Ratings	Users	tags	URL
Movielens-100k [52]	9115	1,00,005	671	1197	http://grouplens.org/datasets/movielens/100k/
Movie lens [54]	27,278	9,999,999	69139	-	http://grouplens.org/datasets/movielens/latest/
Movielens-1M [76]	40,000	14 M	1,60,000	670,000	https://grouplens.org/datasets/movielens/
Yahoo Web scope (R4) [22]	11,915	1,11,131	7,641	-	http://webscope.sandbox.yahoo.com
Movielens-100k [53]	1682	1,00,000	943	-	http://movielens.org
Movielens-15M	17,000	10M	1,38,000	4,65,000	https://grouplens.org/datasets/movielens/15m/
Movie lens Full	58,000	1,80,000	600	1,100,000	https://grouplens.org/datasets/movielens/latest/
Movielens-100k	1700	1,00,000	1000	-	http://grouplens.org/datasets/movielens/100k/
Movielens-10M	10,000	10M	71,000	1,00,000	http://grouplens.org/datasets/movielens/10m/

Conclusions and Future Scopes

This paper describes different types of filtering techniques. Various uses, advantages, disadvantages are also discussed. To build an efficient recommender system a hybrid combination of different methods of recommendation is must. It is concluded that by using combination of similarity measure a better user similarity can be generated rather than using single similarity measure and efficiency of the system is also increased. One of the fact that similarity measure like RJMSD is evolved by the author and up till now it is only used in movie recommendation. The author also showed that this similarity measure is better than the other in terms of efficiency parameters. Accuracy of any recommender system can be improved if we add extra movie features. Generally, most of the papers have shown the combination of collaborative filtering and content-based filtering. By combining methods the problems related to the two methods are tried to resolve. So hybrid filtering is the most well-known technique in any recommendation system. Because using this helps to build an effective recommendation system.

There are various areas of recommendation system as discussed earlier. Various techniques are also discussed which works in giving recommendation. So, scope of any recommender system is to build a model in such a way that their user gets proper recommendation and efficiency of the system is maintained.

Future work:

The next step of our future work is to adopt hybrid algorithm to see how the combination of collaborative and content-based filtering techniques can gives us a better recommendation compared to the adopted technique in this paper

References

[1] J. A. Konstan and J. Riedl, "Recommender systems: From algorithms to user experience", User Modeling and User-Adapted Interaction 22(1-2):101-123, 2012.

[2] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search", Egyptian Informatics Journal 18(2):105-112, 2017.

[3] S. H. Min and I. Han, "Detection of the customer time-variant pattern for improving recommender systems", Expert Systems with Applications 28(2):189-199, 2005
[4] S. Wattal, Y. Hong, M. Mandviwalla, and A. Jain, "Technology diffusion in the society: Analyzing digital divide in the context of social class", IEEE Proc. of 44th Hawaii International Conference on System Sciences, 1-10, 2011.
[5] M. Goldmann and G. Kreitz, "Measurements on the spotify peer-assisted music-on-demand streaming system", IEEE International Conference on Peer-to-Peer Computing, 206-211, 2011.
[6] P. N. V. Kumar and V. R. Reddy, "A Survey on Recommender Systems (RSS) and Its Applications", International Journal of Innovative Research in Computer and Communication Engineering 2(8):5254- 5260, 2014.
[7] C. A. Gomez-Uribe and N. Hunt, "The Netflix Recommender System: Algorithms, Business Value, and Innovation", ACM Transactions on Management Information Systems 6(4):1-19, 2015.
[8] O. Celma, ` Music Recommendation, Springer, Berlin, Heidelberg, 43-85, 2010.
[9] F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie, and Z. Li "DRN: A Deep Reinforcement Learning Framework for News Recommendation", Proceedings of the World Wide Web Conference, 167-176, 2018.
[10] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, "Visually-aware fashion recommendation and design with generative image models", IEEE International Conference on Data Mining, 207-216, 2017.
[11] J. L. Kolodner: A Restaurant Recommender System, 1993
[12] B.S. Oladapo: A Research Paper Recommender System, 2013 unpublished.
[13] E.Baatarjav, J.Chartree, and T. Meesumrarni: Group Recommendation System for Facebook, 2010
[14] http://www.facebook.com

Author's Profile

Mr. Rudrendra Bahadur Singh. , received a Bachelor of Engineering (Computer Science and Engineering) Presently she is working as Assistant professor at Babu Banarasi Das National Institute of Technology and Management. In this project she worked as the Team Guide.
E-mail:- Rudra.rathor20@gmail.com

Abhiraj Agrawal, is currently pursuing his B.Tech from Babu Banarasi Das National Institute of Technology and Management in Computer Science and Engineering.He is considerably fascinated by the Computer engineering subjects and have grip on basics. In this project he worked as the role of Backend Developer.
E-mail:- abhi9559137@gmail.com

Ankit Dwivedi, is currently pursuing his B.Tech from Babu Banarasi Das National Institute of Technology and Management in Computer Science and Engineering. An ambitious problem solver with a passion for online business, has experience of creating both logical & innovative solutions to complex web problems. Working for 2.5 years from home on different web-projects.
E-mail:- dwivedia009@gmail.com

Dhruv Rishi, is currently pursuing his B.Tech from Babu Banarasi Das National Institute of Technology and Management in Computer Science and Engineering. Worked on UI and UX from home for about 3 years mainky focusing on user experience.
E-mail:-dhruvrishi945@gmail.com