IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Movie Recommendation System Using Content-Based Filtering

Dr. Ganesh D
Associate Professor
School of CS & IT
JAIN (Deemed-to-be-University)

Yash Bhansali
Department of MCA
School of CS & IT
JAIN (Deemed-to-be-University)

Abstract— Several advanced level platforms, such as Information Gathering, Learning Techniques, the Internet - Of - things (IoT), and Deep Learning, have emerged as a result of technological breakthroughs. We use technology almost everywhere we operate to meet social demands. In addition, new systems have been developed as a result of this. In recent times, recommendation engines have risen in importance, whether it be in entertainment, education, or other businesses. Previously, users had to decide which publications to buy, which films to watch, and which songs to listen to, among other things. A content-based algorithm's cornerstones are material collection and quantitative analysis. As the study of text acquiring and filtering has progressed, many modern content-based recommendation engines now offer recommendations based on text information analysis. This paper discusses the content-based recommender. The film has several characteristics that set it apart from other recommender systems, including diversity and uniqueness. These features are used to build a movie prototype and determine similarity. We present a novel method for calculating feature weights that improves movie representation. Finally, we examine the strategy to determine how it has progressed.

Keywords— Filtering, Recommender System, Recommender, vector similarity, and cosine matrix

I. INTRODUCTION

Recommendation Systems are an information platform that assists users in finding items that meet their needs from a large number of options. The main goal of developing a recommender system is to provide the relevant content out of irrelevant. This system also assists the user in selecting the best option from a variety of options. Recommendation systems are used by many platforms, including Amazon prime, YouTube, and Amazon, to better serve their customers and increase profits. It's still an interesting research area because determining what a consumer wishes from resources available is difficult, especially since our choices change over time. Nowadays, we buy based on recommendations. When looking for a video on YouTube about a specific topic, there are always a wide range of options. There wouldn't be much of a challenge if the results are correctly ranked, but if they aren't? In such a situation, we'd undoubtedly devote a significant amount of time to finding the best possible movies that fits us and meets our needs.

This is what pops up as a suggestion when you browse for something on a website. Even if you don't search the very next time you visit a particular website, the platform may be able to make recommendations that you will enjoy. Isn't this a fantastic feature? Essentially, a content - based recommendation platform's job was to provide user with more specific features. Video recommendations are made using recommendation systems on YouTube, product recommendations are made using recommendation systems on Amazon and Flipkart, movie recommendations are made using recommendation systems on Netflix and Amazon Prime, and so on. Whatever you do on these websites is tracked by a machine that analyses your behaviour and then suggests things/items that you might be interested in.

This research paper looks at movie recommendations and the reasoning behind them, as well as common movie recommendation systems, problems with traditional film recommendation engines, and other relevant topics. Among the well-known datasets are the Movielens dataset, the TMDB Movie Dataset, as well as the Netflix dataset. Websites like Netflix, Amazon Prime, and others use movie recommendations to raise revenue or profit margin by improving the customer experience. Netflix, in fact, held a competition in 2009 with a prize money of around one million us dollars ($1M) for someone who could enhance the current technique by at least 10%.

A recommendation engine in the background, for example, gives recommendations to a user who wants to listen to the music, pick up a book, or watch movies. Netflix suggests movies, Spotify recommendations music, Amazon recommends things, LinkedIn offers jobs, and another social media platform recommends users all employ a recommendation system based on a user's previous behaviour. These recommendation engines make it simple for people to find what they want depending on their preferences. As a result, it's tough to create an effective recommendation systems because interest of the majority change over time.

II. II. RELATED WORK

Over the last few decades, several methodologies for attempting to make new film recommendations have really been extensively researched. Filtering can be divided into two types: content-based (CB) and collaborative (CF). Based on a historical database of user ratings, recommender system generates recommendations for a particular user. Content-based systems, and from the other hand, make a recommendation by comparing representations of content in a given product (such as a book, movie, or song) to depictions of content that match a user profile. On the other hand, CB systems can better characterise website visitors.

A. Collaborative-Filtering Recommendation

Collaborative filtering suggestion is a popular algorithm in recommendation systems. A user's taste is determined by an algorithm model based on previous activity.

In 1991, Goldberg et al. [20] was the first one to propose the idea of collaborative filtering. Collaborative filtering is based on the assumption that people who previously agreed on a specific element will agree on it again in the future and will enjoy a similar sort of product or item.

2.1 User-Based Collaborative-Filtering

In this strategy, it is assumed that the consumer will love the goods that are also appreciated by other people who have similar tastes in a product. As a result, the initial step in this strategy is to find a person with similar tastes or inclinations. When users like comparable items, collaborative filtering considers them to be identical. The following formula is used to measure similarity between u and v:

$$suv = \frac{|N(u) \setminus N(v)|}{|N(u) \; [\; N(v)|}$$

User/Item	Item A	Item B	Item C	Item D
User A	✓		✓	recommend
User B		✓		
User C	✓		✓	✓

Table 2.1. User-based CF

User-Based Collaborative-Filtering is shown in table 2.1. Only User C is deemed a neighbour of User A based on User A's search history, hence the system's recommendation is for Item D..

2.2 Item-Based Collaborative-Filtering

The item-based strategy is different since it presume that the users will enjoy the products that are similar to what they previously selected.

As a result, the initial stage inside this filtration is to create a list of products that are comparable to goods that a user has previously enjoyed. The fundamental purpose of item-based collaborative filtering is to determine how similar two objects are. Favored items are considered by Item CF. The greater the number of users with almost the same name, the more identical they should be. Let's pretend that the two user sets, N(i) and N(j), both like the words I and j. The degree of resemblance between them would be estimated using the formula:

$$sij = \frac{|N(i) \setminus N(j)|}{|\mathrm{N(i)} \; [\; \mathrm{N(j)}|}$$

User/Item	Item A	Item B	Item C
User A	✓		✓
User B	✓	✓	✓
User C	✓		recommend

Table 2.2. Item-based CF

The Item-Based CF system is depicted in Table 2.2. People who like or are interested in Item A would like Item C, according to the interesting history of each and every user for Item A; so, it is believed how both Items A and C are one and the same. If user C likes item A, it's safe to assume that he or she also wants Item C.

2.3 Limitations

I) Scalability

To improve reliability, a huge amount of data is employed in collaborative filtering, necessitating a high number of resources. Processing becomes imprecise and expensive as data rises exponentially, providing a difficulty in Big Data.

II) Data Sparsity

There is a lot of empty space in a data matrix or a user. This is due to the fact that the majority of users are uninterested in evaluating an item, making it harder to discover others who have similar scores on the same items. As a result, locating users who've already rated similar things is tough. As a consequence of the shortage of user information, the recommendation becomes tough.

III) Cold Start Problem

To locate a match, the system requires a sufficient number of users. For instance, if we want to locate an item or even a user whom is the same as other users, we compare it to other goods or individuals. A fresh profile is initially empty because the user has not reviewed something, and the system has no idea what their tastes are, so the system has no idea what a user likes. This becomes challenging for any algorithm to make additional recommendations. This is especially true for new items that, due to their uniqueness, have yet to be rated by any user. Both of these challenges are addressed using hybrid techniques.

IV) Cannot be applied across content domains -

Users must choose and rank items in each domain separately.

B.Content-Based Filtering Recommendation

Content-based filtering algorithms are applied depending on user attributes. When knowledge about an item, including its identity, location, or descriptions, is available but not about the user, this method is employed. It predicts elements based upon that user information and fully ignores other users' efforts, just like collaborative approaches. It tends to make use of the information provided by the user, either expressly or impliedly. As the user gives more content-based filtering procedures actions on the suggestions, such as content-based recommender, the engine becomes more accurate.

In a content-based recommendation engine, each user is assumed to operate independently. When analysing the properties or attributes of the item, there is no requirement for information on other users; instead, it looks for commonalities among items and suggests the most similar option to other users. If we examine the movie's content, such as the director, writer, and cast, for example, each one of these elements might be regarded a feature. Users are recommended items that are substantially comparable to the item they voted for. The relationship S among items Oi and Oj is defined as follows:

S(Oi,Oj)=f(A1i,A1j)+f(A2i,A2j)+....f(Ani,Anj)

The characteristics for item I are A1i and A2i, and the function f reflects the distance similarities between the first attribute for item I and j.

III . DESCRPTION OF RELATED WORK

The fact that collaborative filtering bases item recommendations on ratings and reviews and references is one of the most fundamental issues. If no one has given a rating to a product, the recommendations will be incorrect, and also the product would not be suggested to a new user. If enough knowledge about the user is available, or if the user has given rating on any items or products, the system can identify the user and provide recommendations.A content-based recommendation engine overcomes the problems of collaborative filtering. Because the user's choices are ignored, the reliability of the suggestions is unchanged, and the user preserves their security by not having to give any personal information. This algorithm can swiftly alter its recommendations when multiple different customers have different interests for the same item. The user specifies the name of a film from which the method's outputs are generated. Movies that are comparable to the needs of the user will be shown as an output.

III . LITERATURE REVIEW

The drawbacks of collaborative filtering approaches, such as the sparsity problem and the cold-start problem, were mentioned by Sang-Min Choi, et al. [1]. The authors have presented a technique that uses category information to prevent this problem. The authors proposed a genre correlation-based movie recommendation system. The authors claim that the category description for newly developed content is present. As a result, even if original features does not yet have sufficient ratings or views, it can nevertheless appear in the recommendation list thanks to classification or genre information. The proposed system is unbiased when it comes to highly rated, most-watched content and new, less-watched content. As a result, even a brand-new film can be recommended.

S. Rajarajeswari and colleagues [2] explored Simple Recommender Systems, Content-based Recommender Systems, and Collaborative Filtering-based Recommender Systems before proposing a Hybrid Recommendation System as a solution. Cosine similarity and SVD were taken into account by the writers. Using cosine similarity, their method generates 30 movie choices. The movies are then filtered based on SVD and user ratings. Because the authors offered a solution that only takes one movie as input, the system only considers the most recent movie that the user has seen.

Ms. Neeharika Immaneni, et al. [3] suggested a hybrid recommendation technique that takes into account both content-based and collaborative filtering approaches in a hierarchical manner to provide users with individualised movie recommendations. The most distinctive aspect of this research is that the study had conducted movie suggestions based on a suitable sequence of images that accurately represent the movie tale storyline, which aids in better visuals.

Deldjoo et al [4.] The authors suggested a model that rewards a strategy for extracting a collection of stylistic elements from a video's content, such as brightness, colour, and motion.

Basilico and Hofmann[5] To learn a prediction function, the authors suggested a framework in which a unified method integrates all available training information, such as prior user-item ratings as well as properties of objects or users.

Deldjoo et[6], Efficient algorithms for Video on youtube reco6mmendation systems were proposed by the authors.

IV . PROPOSED METHODOLGY

The dataset must be preprocessed, and the relevant features must be combined into a single feature. We'll really do have to transform from such a feature into vectors later. We'll need to figure out how similar the vectors are later. Finally, obtain recommendations in accordance with the system architecture outlined below.

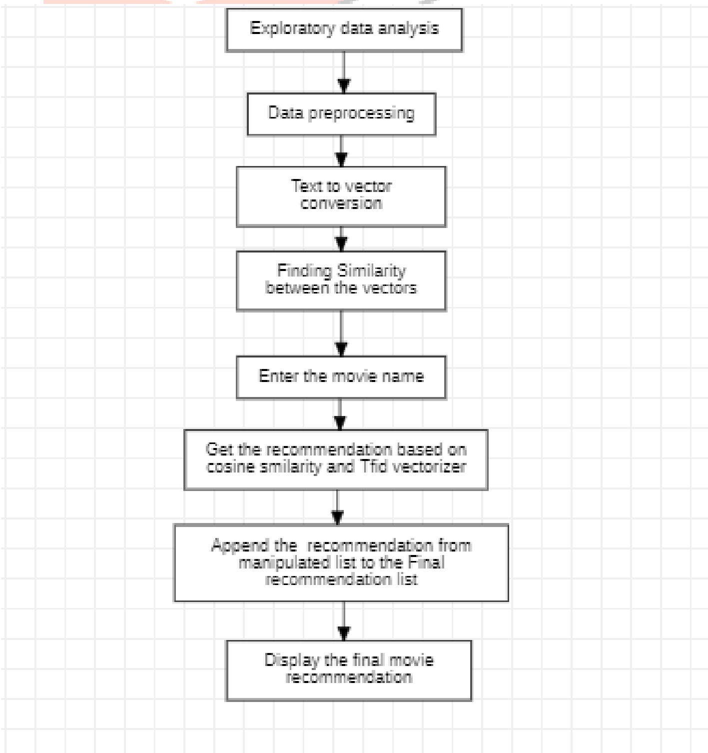

Fig1.System Architecture

We used a content-based filtering algorithm that uses a feature extraction method to determine how similar each item is. The distance metric is determined by the cosine similarity algorithm. The cosine of the angle between the two feature vector projected in a multi - dimensional vector space is used to calculate the similarity of two items. If the user types in a valid movie title, the python code will produce a list of movies to watch. When a movie's name matches one in the dataset, the soup column (all details concatenated into one string) of the each movie is used to generate recommendation.

Step 1	Import the dataset and perform the data pre-processing steps.
Step 2	Import Pandas and create count matrix using the count vectorizer method.
Step 3	Using Cosine similarity matrix determines the similarity of documents irrespective of their size.
Step 4	Create a directory setup for the website where the main input field is placed inside the form.
Step 5	Connect the page to flask and render it.
Step 6	In the terminal open the python file and provide the link in the browser.
Step 7	The user enters the movie name and if it is available in the dataset the cosine matrix is calculated and top 10 similar movies are sorted and displayed to the user.
Step 8	If the movie does not exist in the dataset then message regarding the reason for the same is displayed.

Fig2. Pseudo code for Proposed system

V. RESULT AND DISCUSSIONS

Because it analyses the qualities of items to generate predictions, the content-based recommendations is appropriate in cases where there is known data about the item rather than the user.

Figure 3.1 showcases our website's home page. The user can type a movie title into the text box provided and then press the "Recommend" button.

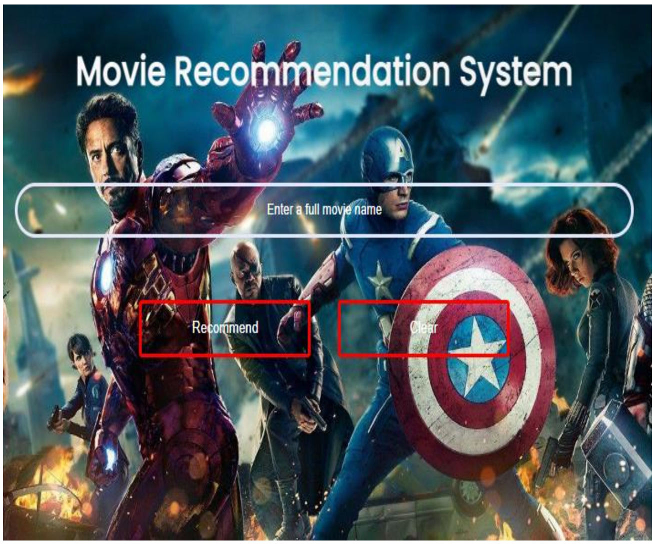

Figure 3.1

In ***Figure 3.2*** the recommendations that are given to the users are shown based on the input that they have provided to our proposed system. In the ***Figure 3.2*** we can see that the user want the recommendations for the spider man movie.

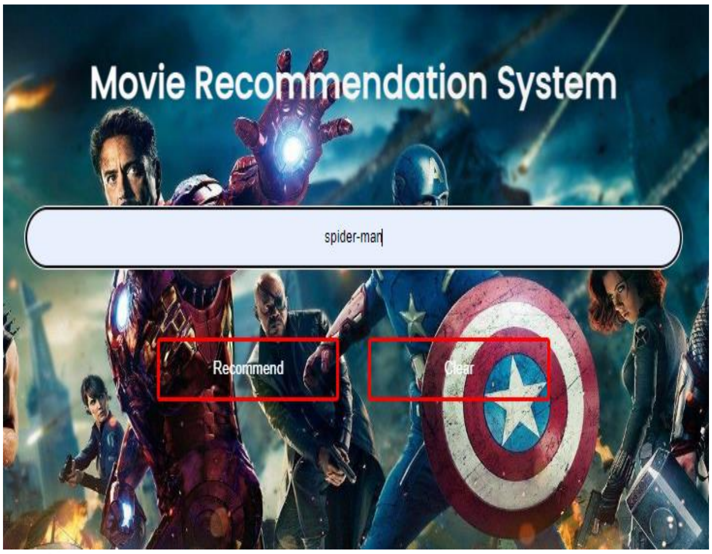

Figure 3.2

Figure 3.3 shows search result suggestions based on the search input "Spider-man".

Movie Title	Release Date	Director Name	Cast	Trailers	OTT Platforms
Spider-Man 3	2007-05-01	Samraimi	['Tobeymaguire', 'Kirstendunst', 'Jamesfranco']	Watch Trailer	NETFLIX, prime video, SONY PICTURES TELEVISION
Spider-Man 2	2004-06-25	Samraimi	['Tobeymaguire', 'Kirstendunst', 'Jamesfranco']	Watch Trailer	prime video, SONY PICTURES TELEVISION, NETFLIX
Charlotte's Web	2006-12-15	Garywinick	['Juliaroberts', 'Stevebuscemi', 'Johncleese']	Watch Trailer	BIG FLIX, HBO max
Cirque Du Freak: The Vampire's Assistant	2009-10-23	Paulweitz	['Johnc.Reilly', 'Joshhutcherson', 'Chrismassoglia']	Watch Trailer	hotstar, EROS NOW, HBO max
Oz: The Great And Powerful	2013-03-07	Samraimi	['Jamesfranco', 'Milakunis', 'Rachelweisz']	Watch Trailer	EROS NOW, YouTube Premium, NETFLIX
Daybreakers	2009-01-06	Michaelspierig	['Ethanhawke', 'Samneill', 'Willemdafoe']	Watch Trailer	YouTube Premium, SONY PICTURES TELEVISION
Army Of Darkness	1992-10-09	Samraimi	['Brucecampbell', 'Embethdavidtz', 'Marcusgilbert']	Watch Trailer	YouTube Premium, EROS NOW, STARZ
Harry Potter And The Order Of The Phoenix	2007-06-28	Davidyates	['Danielradcliffe', 'Rupertgrint', 'Emmawatson']	Watch Trailer	BIG FLIX, prime video, HBO max
Tower Heist	2011-11-02	Brettratner	['Benstiller', 'Eddiemurphy', 'Caseyaffleck']	Watch Trailer	NETFLIX, HBO max
The Ice Storm	1997-09-27	Anglee	['Kevinkline', 'Joanallen', 'Sigourneyweaver']	Watch Trailer	STARZ, BIG FLIX, SONY PICTURES TELEVISION

Figure3.3

In Figure 3.4 This set of standards would then return movie titles that seem to be equivalent to that same input that the user has decided to enter, if relevant. When the user enters an invalid movie name, it will also display that movie could not have been found. To search the most equivalent movie names, the system will compare the data to all existing movie titles.

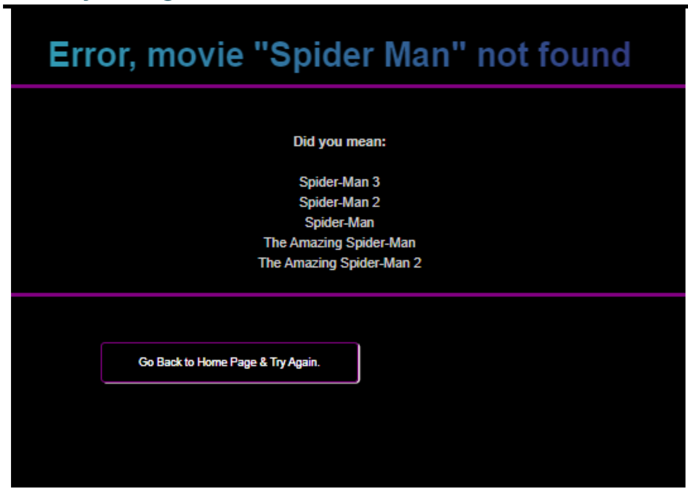

Figure 3.4

VI. CONCLUSION

Due to the overabundance of data, the recommendation systems framework has become extremely important. We are specifically looking for a better way to work on the exactness of the film agent for the content-based recommender framework. Proposal frameworks that use continuous data from wearable devices and snap stream to generate more advanced results are more successful in general.

For example, the suggested results from a health care field suggestion framework, such as analysis and therapy procedures, have a lower predilection than clinical information-based outcomes. It does, however, have significant value as solid data, which can provide quick support to convention administrations and quick advancements, because consistent data provides a more relevant output by mirroring patients' present state. In terms of innovation, the proposed plan framework is divided into two sections: an information mining portion that examines articles and clients based on data obtained, and a recommendation sifting model region. While utilising the proposal framework, each invention and model has been thoroughly investigated and developed to be more custom tailored to the assistance business.

VII. REFERENCES

[1] R. Kumar, S. A. Edalatpanah, S. Jha, and R. Singh (2019). A new method for solving gaussian-valued neutrosophic shortest path problems has been developed. 347-353 in Published In journal of Technology and Advanced Technology.

[2] Ahmad, Muyeed, Mir Tahsin Imtiaz, and Raiyan Khan are the members of the team. "A cluster and pattern recognition network-based movie recommendation system." IEEE's 8th International Computing and Communication Conference and Conference is taking place this year (CCWC). 2018 IEEE.

[3] "Recommender system utilising fuzzy c-means clustering and genetic algorithm which is based weighted similarity measure," Hirdesh Shivhare, Anshul Gupta, and Shalki Sharma, IEEE International Conference on Computer, Communications, and Control (2015).

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

[5] Manoj Kumar, D.K. Yadav, Ankur Singh, and Vijay Kr. Gupta, "A Movie Recommender System:MOVREC," Published In journal of Computer Applications (0975 – 8887) Vol 124 – No.3 (2015).

[6] "A new mobile highly personalized stock exchange recommender systems method based on a money flow model. (Survey)"Annual 2014 Issue of Mathematical Problems in Engineering, Qingzhen Wu, Jiayong Chen, Qiang Xu, Qingzhen Wu, Jiayong Chen, Qingzhen Wu, Jiayong Chen, Qingzhen Wu, Jiayong Chen, Qingzhen Wu, Jiayong Chen, Qingzhen Wu,

[7]"An Improved Collaborative Movie Recommender System Using Computational Intelligence," Zan Wang, Xue Yu*, Nan Feng, and Zhenhua Wang (2014), Journal of Visual Language and Computing, Volume 25, Issue 6.

[8]"Movie Recommendation System Using Collaborative Filtering," IEEE 9th International Conference on Software Engineering and Service Science, 2018, pg. C. M. Wu, D. Garg, and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," IEEE 9th International Conference on Software Engineering and Service Science, 2018, pg (11-15).

[9] "Movie recommender system using item-based collaborative filtering technique," 1-5, IEEE Conference on recent Trends in Engineering, Technology, and Science, 2016. L. T. Ponnam, S. D. Punyasamudram, S. N. Nallagulla, and S. Yellamati.

[10] "Reliable Movie Recommender System with Correct Evaluation and Emotion Assessment," Proceedings of the International MultiConference of Engineers and Computer Scientists Vol II, RyuRi Kim, Ye Jeong Kwak, HyeonJeong Mo, Mucheol Kim, Seungmin Rho, Woon Kian Chong,Ka Lok Man (2015).

[11] "Design of Movie Recommendation System by Collaborative Filtering," International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 4, 2013

[12] Dain Kaplan, Neil Rubens, Mehdi Elahi, Masashi Sugiyama (2016). "Active Learning in Recommender Systems." Ricci, Francesco; Rokach, Lior; Shapira, Bracha, in Ricci, Francesco; Rokach, Lior; Shapira, Bracha (Eds.). Recommender Systems Handbook (2 ed.). Springer US is a publishing company based in the United States.

[13] "A digitized bookshelf: original work into recommendation system," by Jussi Karlgren, was released in October 2017. The 27th of October, 2017.

[14] R. Singh ,S. Jha and R. Kumar(2017) . Find the shortest in a networks having type-2 triangle fuzzy arc length. Journal of Applied Research in Industrial Engineering, 4(1), 1-7.

[15] Chiru, Preda, Dinu, and Macri (C. G. Chiru, C. Preda, V. N. Dinu, and M. Macri) (2015, September). A movie recommendation system is built using the user's psychological profile. The IEEE hosted the World Conference on Innovative Computers Communication and Processing (ICCP) in 2015. (pp. 93-99). IEEE.

[16] R. Hande, A. Gutti, K. Shah, J. Gandhi, and V. Kamtikar (R. Hande, A. Gutti, K. Shah, J. Gandhi, and V. Kamtikar) (R. Hande, A. Gutti (2016). MOVIEMENDER is a website that provides movie recommendations. IJESRT, vol. 5, no. 11, p. 686. 686 in International Journal of Applied Sciences and Research Technology.

[17] J. Son and S. B. Kim (2017). Multiattribute networks with content-based filtering for recommender systems. 404-412 in Expert Systems with Applications.

[18] Rho, Ka Lok Man, and Woon Kian Chong (2015), "Trustworthy Recommendation System with Correct Evaluation and Emotion Evaluation," Proceedings of the National Academy of Sciences.