

# An Efficient Algorithm for Movie Recommendation System

Anitha R <sup>a,b,1</sup>, Surya Koti Kiran A <sup>b</sup>, Anurag K <sup>b</sup> and Nikhil Y <sup>b</sup>

<sup>a</sup>Associate Professor Department of Computer Science And Engineering

<sup>b</sup>Koneru Lakshmaiah Education Foundation, Vaddeswaram

**Abstract.** Now a day's recommendation system has changed the fashion of looking the items of our interest. OTT Movie Application Recommendation for mobile users is crucial. It performs a complete aggregation of user preferences, reviews and emotions to help you make suitable movies. It needs every precision and time-liness, however, this can be info filtering approach that's accustomed predict the preference of that user. Recommender System may be a system that seeks to predict or filter preferences in keeping with the user's selections. The very common purpose where recommender system is applied are OTT platforms, search engines, articles, music, videos etc During this work we tend to propose a Collaborative approach-based Movie Recommendation system. it is supported collaborative filtering approach that creates use of the knowledge provided by users, analyzes them so recommends the flicks that's best suited to the user at that point. The suggested motion picture list is sorted in keeping with the ratings given to those movies by previous users. It conjointly helps users to search out of their selections supported the movie expertise of alternative users in economical and effective manner while not wasting a lot of time in useless browsing [1]. Therefore, we tend to offer the item-oriented methodology of the analysis of social network as the steering force of this method to further improve accuracy within the recommendation system. We tend to propose economic healthcare associates during this paper The algorithmic rule of the Film Recommendation supported improved KNN strategy that measures simpler advisory system accuracy. However, to evaluate performance, the k closest victimized neighbors, the maximum inner circles, as well as the basic inner strategies are used [2]. The exception to this is the projected results, which use algorithms to check for (supposedly) involvement. The performance results show that the projected strategies improve additional accuracy of the Movie recommendation system than the other strategies employed in this experiment.

**Keywords.** Movie Recommendation, KNN, Correlation, IMDB, Search by keywords.

## 1. Introduction

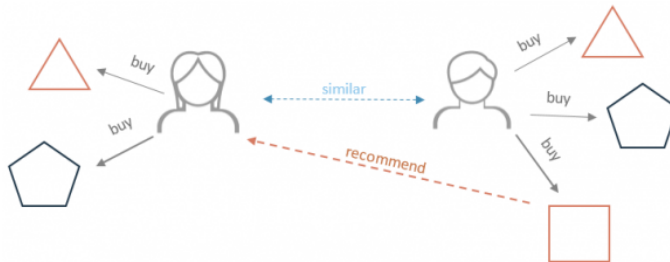
Individuals' everyday lives are getting progressively reliant on portable administrations as cell phones become more common. Shrewd gadgets furnish individuals with organization subtleties, item data, advertising data, and proposal data. Quite possibly the main employments of versatile administrations is film suggestion. A film recommender frame-

<sup>1</sup>Anitha R, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.  
E-mail: anitharaju15@gmail.com.

work has end up being a successful instrument for giving clients important film suggestions. The proposals are given to help clients in managing data over-burden by helping them in rapidly and effectively finding appropriate films. In contrast to the market for PCs (computers), portable administrations place a more prominent spotlight on idealness, requiring specialist organizations to measure and quantify information quickly. Subsequently, film proposals in portable administrations should be stressed regarding both guidance precision and, therefore, practicality. Recommender Frameworks produce proposals; the purchaser can favor them dependent on their inclinations, and ought to likewise give certain or unequivocal info, either quickly or later [3]. Clients' practices and ideas are routinely saved in the recommender data set and can be utilized to make new suggestions during ensuing client framework connections. Film proposal can be a complex and tedious cycle that requires an assortment of client inclinations, film types, etc. Accordingly, an assortment of idea strategies has been recommended to determine the issues [4]. customized ideas of the greatest quality give the client experience another measurement. Customized online suggestion frameworks have as of late been utilized to give different types of custom-made data to their clients. These frameworks are generally utilized in an assortment of utilizations and are getting more mainstream continuously. The recommender frameworks will be partitioned into two classifications:

1. Collaborative sifting approach
2. Content-based sifting approach

Things are suggested by a common sifting structure zeroed in on client as well as article closeness measures. The framework suggests things that are regular among comparable clients. There are a few advantages of Collaborative sifting. (i) It is content-free, depending exclusively on associations. (ii) Because individuals in CF give clear appraisals, genuine quality assessments of things should be possible. (iii) Since proposals depend on client likeness instead of thing closeness, it submits fortunate suggestions.



**Figure 1. Collaborative filtering block diagram**

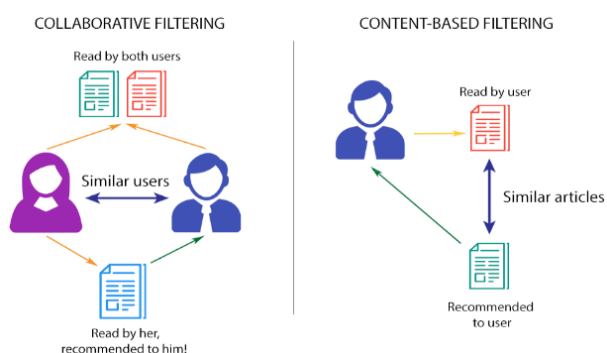
Besides, clients' absence of cognizance of slant is a significant issue in film suggestion. Individuals are progressively ready to post their own audits online nowadays. Clients will communicate their assumptions and sentiments about motion pictures in their audits. As a result, the assessments passed on in these audits affect the decisions made by different clients [5]. Clients can peruse the audits, assess their own encounters, pick the most important surveys, erase those that are negative or possibly inconvenient, and settle on their own decisions and choices. As an outcome, the sound of audits can be a huge thought to consider while assessing a film. Likewise, watchers are destined

to pick the movies that most of individuals like and overlook the movies that most of individuals' disdain. System extracts keywords from the search statement for various film descriptions, mines the keyword in the database and recommends a film based on a similar user view. We use a feelings database based on the keywords, together with the weight of the positivity or negativity in the database. We then classify it according to those feelings.

This is dependent on a suggestion process that gives clients different recommendations [6]. Clients would be prescribed movies using this system. When compared to existing systems, this system can yield more accurate results. The current system eats away at individual clients' evaluations. This system also has newly added features such as searching for recommendations [7] using keywords that matches with description. As every shopper could have completely different preferences, it should be pointless for shoppers World Health Organization have distinct tastes from the concepts given by the system. This method calculates the likenesses of shoppers so recommend films to them supported the ratings provided by different shoppers of comparable tastes. This will offer the client with a particular recommendation.

## 2. Methodology

The existing system works on individual users' rating and on the genre of the show. this could be someday useless, the users World Health Organization have completely different style from the recommendations shown by the system as each user might have different tastes. This technique calculates the similarities between completely different users then suggest show to them as per the ratings given by the various users of comparable tastes. The planned system may be a higher system than the other



**Figure 2. Difference between collaborative and content-based filtering**

existing systems. this technique has accessorial the positive options of existing systems and has overcome the drawbacks of existing systems. The system uses item-based cooperative based mostly algorithms. within the existing system picture shows solely will be searched mistreatment picture show title however within the planned system they will be searched through key words that matched with description of the movie [8,9]. There are 3 main types in movie recommendation system as following:

- Content Based: the advice system recommends different moving-picture shows that area unit just like that elite movie.  

$$f(\text{movie}) \rightarrow \text{movie}$$
- Collaborative: the advice system recommends movies that area unit rated extremely by the similar users.  

$$f(\text{movies}, \text{user}) \rightarrow \text{movie}$$
- This method combines previous approaches (that is, Approach 1 and Approach 2). most likely, most organizations use a hybrid approach to production recommender systems

In this paper we have executed the two methods namely Correlation Coefficient effect and K nearest neighbors. We have taken the dataset from the IMDB website where all the movies gets registered. We have implemented the program in a way that the movies will be recommended based on the similar user whose watching of movies will be similar.

### 2.1. Correlation coefficient

Correlation coefficients is a accustomed measure however sturdy a relationship is between 2 variables. There are many kinds of coefficient of correlation, however the foremost widespread is Pearson's. Pearson's correlation (also known as Pearson's R) could be a coefficient of correlation unremarkably utilized in statistical regression. If you're beginning get into statistics, you'll in all probability find out about Pearson's R initial. In fact, once anyone refers to the coefficient of correlation, they're typically talking concerning Pearson's.

The main element of our moving picture recommendation system depends on a learning conception known as cooperative filtering. cooperative filtering bases its suggestions solely on users' past knowledge and preferences, largely within the style of reviews (albeit their area unit alternative strategies of gathering user preferences). To understand this, I'll illustrate associate example of user-user cooperative filtering that utilizes the closest neighborhood rule. Say I enjoyed looking at moving picture A, movie B, and moving picture C, and my friend enjoyed looking at moving picture A, movie B, and moving picture D. The cooperative filtering rule can possibly counsel that I'll get pleasure from looking at moving picture D, and my friend can get pleasure from looking at moving picture C supported our previous positive preferences. It is smart on behalf of me and my friend to get pleasure from our moving picture recommendations as a result of we have a tendency to share similar preferences. Of course, this isn't continually the case, however the percentages become more and more in our favor once we begin analyzing larger datasets of user reviews.

### 2.2. KNN Nearest Neighbors

In this section, as a method to recommend movies based on collaborative approach, we present the KNN classification algorithm [10]. The KNN contribution is two-fold. Next, it computes a special K value ideal for that patient for each unknown patient  $u_0$ . Second, precise measurements of distance are employed. A normalization function is used for the set  $N \subset A$  to scale the numerical characteristics to the interval[0,1] to prevent wide range

characteristics from outweighing those with narrower ranges, given a set U of patients defined by a set of characteristics.

$$\widehat{V_{u_i a_j}} = \frac{V_{u_i a_j} - \min_{u_n}(v_n \cdot a_j)}{\max_{u_n}(v_n \cdot a_j) - \min_{u_n}(v_n \cdot a_j)} \quad (1)$$

As we have conveyed above, for both supervised and unsupervised issues[11], the above algorithm can be used. To forecast any latest data values, the above algorithm uses 'function similarity' [12]. This proposes that a rating relying on coordinates of the training set values is applied to the new phase. Computing the interval between the initial point and each point taken from the training dataset is the initial step. There are different ways to measure this distance, including different kind of distance such as Euclidean distance, Manhattan Distance and Hamming Distance. The three mentioned distance calculated using below equations.

$$EuclideanDistance = \sum_{i=1}^k (x_i - y_i)^2 \quad (2)$$

$$ManhattanDistance = \sum_{i=1}^k |x_i - y_i| \quad (3)$$

$$HammingDistance(D_H) = \sum_{i=1}^k |x_i - y_i| \quad (4)$$

$$x=y \Rightarrow D=0 \mid x \neq y \Rightarrow D=1$$

At an incredibly low approximation of k, the model overfits the training results, resulting in a high error rate on the range accepted. If closely observed, the validation error curve reaches a minimal value of k=9. This value of k is the optimal calculation of the model. This curve is known as the 'elbow curve' that can be used to determine the value of k.

### 3. Results and Discussion

As in the above discussed methods and operations. we are able to build the recommender system for recommending movies by user-based collaborative approach and an additional feature which is helpful to get recommendation of movies, based on keywords in search phrase. The movies are recommended when the keywords have similar words in the description of the movies. Both the correlation and KNN methods executed successfully, and they give accurate and timeliness result. This recommendation system recommends movies based on what similar users watch and genre the user usually watch. The execution of additional feature that is getting recommendations through searching keywords is successful. A test case was executed for finding recommendation of movies based on searching keywords. For the input "Football player trains a woman team" the

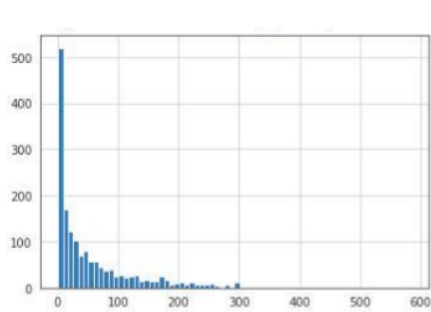


Figure 3. Histogram for movie Ratings

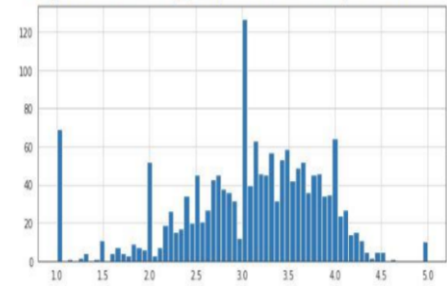


Figure 4. Number of movie ratings in the data set

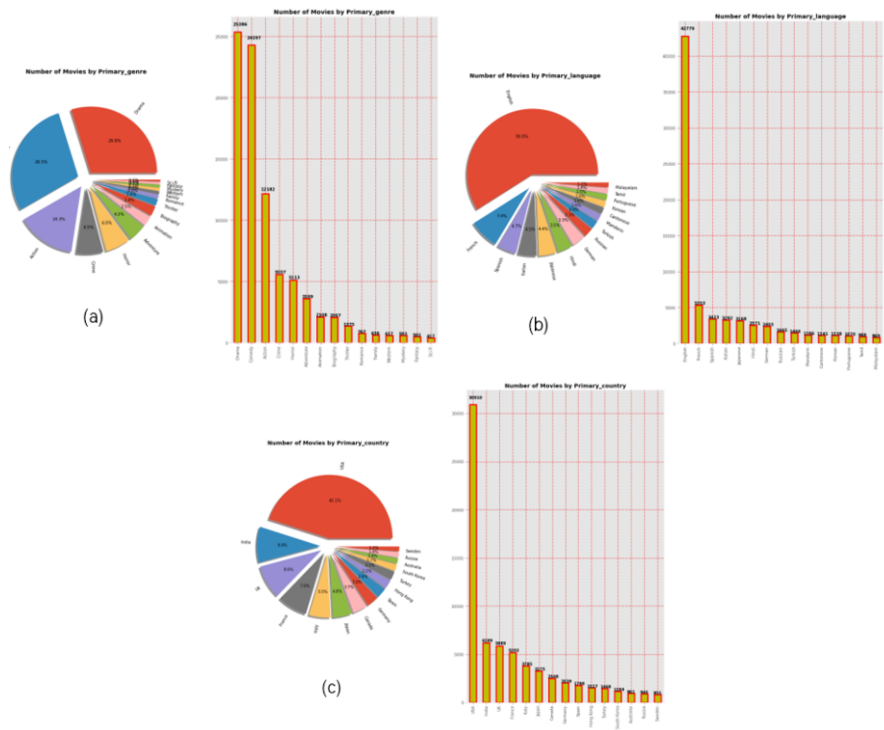


Figure 5. Data visualization.

output returned is “Bigil, What Men Want, The Sensei, Quan li Kou Sha Etc..”. The other test case for the input “Plane crashes immortal Amazonian war” the output returned is “Kosmos Anna 2000, Wonder Woman, Airport ‘80, Last Flight to Abuja Etc..”. This shows that this feature’s result is successful. The code implementation was done with the help of google colabs. Google colab is a software tool by Google Research which allow anyone to create and work the python code on a web-browser. It is very well suited to Data Analysis, Machine Learning(ML) and education. It is a free Jupyter notebook service and does not require a setup for use, also provides free access to computing resources along with GPUs. For the input “Bigil” output was successful. The Recommender sys-

tem recommended “Sarkar, Their, Pailwaan, Viswasam...etc.” for the input “Bigil”. For the input “Wonder Woman” output was successful. The Recommender system recommended “Black Panther, IT, A Quiet Place, Hereditary...etc.” for the input “Wonder Woman”. Importing the packages and reading the dataset was done. The required dataset

```
get_recommendations_for('Wonder Woman')
| Movie Name: Black Panther
| Director: Ryan Coogler
| Cast: Chadwick Boseman, Michael B. Jordan, Lupita Nyong'o, Danai Gurira, Martin Freeman, Daniel Kaluuya, Letitia
e, Sterling K. Brown, Angela Bassett, Forest Whitaker, Andy Serkis, Florence Kasumba, John Kani, David S. Lee
| Language: English, Swahili, Nama, Xhosa, Korean
| Genre: Action, Adventure, Sci-Fi
| Year: 2018
| Country: USA
| Description: T'Challa, heir to the hidden but advanced kingdom of Wakanda, must step forward to lead his peo
d must confront a challenger from his country's past.
| Rating: 7.3
| Total Votes: 597858
```

Figure 6. The Recommender system output for “Wonder Woman”.

was loaded. Operations like view head() of data, info of data were done. Importing the required datasets and analyzing the dataset done was used to find recommendations for movies based on keywords search. The visualization of data based on primary genre of the movie is done. The pie graph shows amount of movie each genre has. From the chart drama and comedy has highest with 29.8% and 28.5%. The Figure 4 is represents the visualization techniques to show ratings and number of ratings for the movies in the data set Figure 5 help us visualize the dataset effectively. The first plot tells us how the movies are segregated according to genre i.e drama, action, romance etc. The second plot tells us how the movies are segregated according to language i.e English, hindi, Spanish etc. The third figure tells us how the movies are segregated according to where the movie is released i.e US, UK, India etc. The correlation score tells us how close the movies above are related to the movie we provided as an input. Closer its value to 1, the more chances the user may like the movie recommended by our recommender. The Recommender system recommended “Black Panther, IT, A Quiet Place, Hereditary...etc.” for the input “Wonder Woman” can be verified in Figure 6. Movie recommendation for given key-

```
[29]: query='football player trains a woman team'
search_result=searchFAISSIndex(imdb_movies,'id',query,index,nprobe=10,model=model,topk=20)
search_result=search_result[['id','description','title','cosine_sim','L2_score']]
search_result
#expected bigil

Batches: 100% 1/1 [00:00<00:00, 22.04it/s]
```

	id	description	title	cosine_sim	L2_score
19	83217	A former football player struggles to train a ...	Bigil	0.567901	0.929622
17	80897	A woman is boxed out by the male sports agents...	What Men Want	0.547876	0.950919
9	46296	In 1985 Colorado, a long-absent woman, profici...	The Sensei	0.539643	0.959538
13	70509	A young woman trains to become a badminton cha...	Quan li kou sha	0.517158	0.982692
16	77080	A female recording artist encounters the misog...	Stadium Anthems	0.511985	0.987942
14	70644	A self-help seminar inspires a sixty-something...	Hello, My Name Is Doris	0.502626	0.967371
12	63611	A neuron-transfer scientist experiments with t...	Aurora	0.494476	1.005509
5	30229	A female Senator succeeds in enrolling a woman...	Soldato Jane	0.490444	1.009511
8	44292	A determined woman works with a hardened boxin...	Million Dollar Baby	0.486515	1.013396
6	43651	A single mom plays in the local ice hockey tea...	Chicks with Sticks	0.483045	1.016813

Figure 7. Movie Recommendation keyword “Football player trains a woman team” Output

words. For the input “Football player trains a woman team” the output returned is “Bigil, What Men Want, The Sensei, Quan li Kou Sha Etc..”. This shows that this feature’s result is successful.

#### 4. Conclusion and Future Scope

In this project, a better movie recommendation system was built successfully and new features such as searching movies through description were added. When user searches using keywords, the system will recommend movies which have similar keywords in its description. We studied what a recommender system is and how we can create it in Python using only the Pandas library. The algorithm we used is good but real time system use a very complex algorithm. This algorithm can be embedded into a movie streaming app. That way users can get the best out of the application. This algorithm can also be used just for viewing a movie details just by letting it know a part the story of story that will happen. Latest movies are given preference over older ones for better user experience. This project can be forked and can be given better features according to requirements.

#### References

- [1] Sadhana SA, SaiRamesh L, Sabena S, Ganapathy S, Kannan A. Mining target opinions from online reviews using semi-supervised word alignment model. In 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM) 2017 Feb 3 (pp. 196-200). IEEE.
- [2] Sulthana AR, Jaithunbi AK, Ramesh LS. Sentiment analysis in twitter data using data analytic techniques for predictive modelling. In Journal of Physics: Conference Series 2018 Apr 1 (Vol. 1000, No. 1, p. 012130). IOP Publishing.
- [3] Zhao ZD, Shang MS. User-based collaborative-filtering recommendation algorithms on hadoop. In 2010 third international conference on knowledge discovery and data mining 2010 Jan 9 (pp. 478-481). IEEE.
- [4] Jain A, Vishwakarma SK. Collaborative filtering for movie recommendation using RapidMiner. International Journal of Computer Applications. 2017 Jul;169(6):29-33.
- [5] Kumar M, Yadav DK, Singh A, Gupta VK. A movie recommender system: Movrec. International Journal of Computer Applications. 2015 Jan 1;124(3):7-11.
- [6] Bagci H, Karagoz P. Context-aware location recommendation by using a random walk-based approach. Knowledge and Information Systems. 2016 May 1;47(2):241-60.
- [7] Desrosiers C, Karypis G. A comprehensive survey of neighborhood-based recommendation methods. Recommender systems handbook. 2011:107-44.
- [8] Jeong WH, Kim SJ, Park DS, Kwak J. Performance improvement of a movie recommendation system based on personal propensity and secure collaborative filtering. Journal of Information Processing Systems. 2013;9(1):157-72.
- [9] Inan E, Tekbacak F, Ozturk C. Moreopt: A goal programming based movie recommender system. Journal of computational science. 2018 Sep 1;28:43-50.
- [10] Hamed A, Sobhy A, Nassar H. Accurate Classification of COVID-19 Based on Incomplete Heterogeneous Data using a K NN Variant Algorithm. Arabian Journal for Science and Engineering. 2021 Mar 4:1-12.
- [11] Selvakumar K, Ramesh LS, Kannan A. Enhanced K-means clustering algorithm for evolving user groups. Indian Journal of Science and Technology. 2015 Sep 1;8(24):1-8.
- [12] Dev AV, Mohan A. Recommendation system for big data applications based on set similarity of user preferences. In 2016 International Conference on Next Generation Intelligent Systems (ICNGIS) 2016 Sep 1 (pp. 1-6). IEEE.