iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42927

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

MediaRec: A Hybrid Media Recommender System

Tanmay Bhuskute[1], Amit Jeve[2], Nihal Shah[3], Tejas Shah[4], B. A. Patil[5]
[1,2,3,4]U.G. student, [5]Assistant Professor, Department of Computer Engineering, All India Shri Shivaji Memorial Society's College of Engineering, Pune, India411001, Contact No: 9960382442, 9923036196

Abstract: ***This paper discusses about a hybrid recommendation platform for Movies, Books and Songs in one roof. A recommender system is a subgroup of information filtering systems that helps in predicting the "rating" or "Preference" that a user would give to any item. It also helps users to get media of their choice based on their experiences of self and other users in a productive and efficacious manner without wasting time in useless browsing. Previous approaches in recommender system (RS) include Content based filtering and Collaborative filtering. These approaches have a particular limitation as like the necessity of the user history as they visit. So as to overcome such dependencies, the Hybrid Recommendation System is introduced. It uses both Collaborative based filtering system and Content based filtering system for recommending media. In this way, the system performance will be greatly improved through the integration of these two.***

Keywords: Media Recommender System, Movies, Books, Songs, Recommender, TFIDF, Cosine Similarity, Pearson Correlation, KNN, K-Means Clustering.

I. INTRODUCTION

The exponential increase in the volume of information accessible to the internet has produced a potential information overburden challenge that impedes timely access to items of interest on the Internet. As a result, the demand for recommender systems is higher than ever. Recommender systems address the issue of information overburden by excluding important fragments from the large volume of dynamically produced information according to the user's preference, interests, or perceived behaviour associated with the article. Information filtering system. Recommender system can predict whether a specific user will like an item emanate from the user's profile.

What is Recommender System? Recommendation system is so common today that most of the people use it unknowingly. Since it's not possible to browse all products and content on the site, recommender systems are important in providing a better user encounter while uncovering more archive that may not else be found. Play a role. Practical examples of the recommendation system include product recommendations on Amazon, Netflix suggestion for movies and TV series in feeds, recommended video on platforms like YouTube, music on Apple Music, Google news feeds, and Google ads.

Existing recommender system can be divided into three categories: content-based recommenders, collaborative filtering-based recommenders, and hybrid recommendations. However, in some scenarios, profiles with many user words are not sufficient to accurately capture user preferences. Collaborative Filtering Recommendations: The system recommends movies based on the user's movie rating and usually does not include content. It is known as a cold boot problem because many users do not have enough history behaviours or there are not enough users in the system. Hybrid Recommendations: As mentioned above, content-based collaborative filter recommendation systems can provide meaningful results, but they have some drawbacks.

II. LITERATURE SURVEY

"For watching favourable movies online, we can utilize movie recommendation systems, which are more reliable, since searching of preferred movies will require more and more time which one cannot afford to waste. In this paper, to improve the quality of a movie recommendation system, a Hybrid approach by combining content based filtering and collaborative filtering, using Support Vector Machine as a classifier and genetic algorithm is presented in the proposed methodology and comparative results have been shown which depicts that the proposed approach shows an improvement in the accuracy, quality and scalability of the movie recommendation system than the pure approaches in three different datasets. Hybrid approach helps to get the advantages from both the approaches as well as tries to eliminate the drawbacks of both methods" (SCAD Institute of Technology et al., n.d.).

"In this paper, we design and implement a movie recommendation system prototype combined with the actual needs of movie recommendation through researching of KNN algorithm and collaborative filtering algorithm. Then we give a detailed principle and architecture of JAVAEE system relational database model. Finally, the test results showed that the system has a good recommendation effect" (Cui, n.d.).

©IJRASET: All Rights are Reserved

"This paper analyses and compares the performance of the Collaborative Filtering and Hybrid based approaches in generating movie recommendations. The Collaborative Filtering approach uses the rating data consisting of two main phases: (1) movie similarity and (2) movie rating prediction. Meanwhile, the Hybrid based approach adds the benefit of a Content-based to the Collaborative Filtering based approach. Thus, it uses both the rating and movie data and is consisting of four main phases: (1) text pre-processing, (2) term weighting, (3) movie clustering, and (4) Collaborative Filtering based approach. Empirical results show that the recommendation performances of both approaches are linear to the size of the movie neighbourhood. However, the Hybrid-based approach's required neighbourhood size is naturally a lot smaller than that of the Collaborative Filtering since the former employs a clustering technique. The performance comparisons show that the Collaborative Filtering based approach always outperforms the Hybrid based at any top-N position in Precision and NDCG metrics. These findings conjecture that the Hybrid approach does not always improve the Collaborative Filtering approach in movie recommendation" (Ifada et al., 2020).

This study compares the recommendations obtained with and without taking into account movies that have never received an above-average rating, where average rating is defined as the midpoint between 0 and the maximum rating, such as 2.5 on a 1 to 5 rating scale. Collaborative filtering technique is used for recommendations and Pearson correlation coefficient technique is used for similarity measure. Movie-Lens-100k dataset is used in this paper. "This experiment result shows that low rated movies are not significant in finding the movie predictions. So, it's suggestable to ignore them while calculating movie predictions" (Reddy et al., 2020).

III. METHODOLOGY

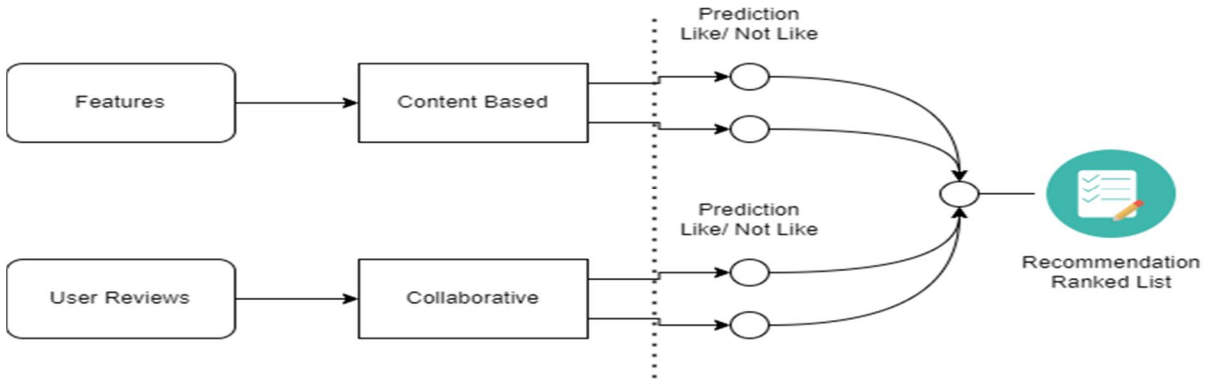


Fig1: Architecture Diagram of Media Recommender System

A. Hybrid Filtering Method

Content-based filtering method and Collaborative filtering methodologies can be used to recommend movies, songs, and books to consumers in a variety of ways. The content-based filtering strategy focuses on item similarity such as movie similarities whereas collaborative filtering focuses on building a link between different individuals who have similar movie viewing preferences. Movies with comparable plots can be recommended to the user based on the plot of a media that the user has previously viewed. This method is classified as content-based filtering because the recommendations are solely formed on the user's previous behaviour.

In the process of movies & books recommendations (Awawdeh et al., n.d.), user-based clustering models are improved by using Nearest neighbours for a given users database. The algorithms used to combine user-based Collaborative filtering models with K nearest neighbour. (KNN) is a supervised classification algorithm. It is also known as a lazy learning type algorithm. It does not construct any models while training the dataset. As it has no model creation, training usually takes lesser time as compared to the testing of the algorithm. KNN is applied on labelled data. Based on the neighbours it forecast the class label for any new unseen test data. In order to predict the label, KNN uses the nearest neighbour of the test sample. Euclidean distance, Cosine similarity, and other proximity measures are used to find the closest neighbours. K-Nearest neighbour's labels are used to forecast the class labels of test samples. Fig 1. shows system architecture of a Media Recommendation systems using both Content & Collaborative based filtering techniques (Hybrid Model).

In this procedure of recommendation, user-based cluster is created (Reddy et al., 2020). In order to improve the performance of recommender systems, other components are also reviewed. One such attribute considered in the research paper is correlation among the users. Pearson correlation coefficient (PCC) is used to find the correlation.

B. Cosine Similarity

Cosine similarity measures similarity between two objects. It uses mathematics to calculate the cosine of the angle created by two vectors projected in a multi-dimensional space. The output value ranges from 0 to 1. 0 indicates no similarity, while 1 indicates that both elements are identical. Cosine Similarity or cosine kernel computes the normalised dot product of input samples X and Y. Sklearn cosine similarity is used to calculate the cosine for the two vectors in the count matrix.

C. TF-IDF Vectorization

The TF-IDF subtask's principal functions are material retrieval and data extraction. In a document that is part of a corpus of papers, an attempt is made to highlight the significance of a term. The text is turned into a vector that can be used when you employ term frequency-inverse document frequency. The frequency of terms (TF) is paired with the frequency of documents (DF) (DF). The term frequency is used to determine the frequency of a term in a document. The frequency with which a word appears in an essay or other piece of literature might reflect how important it is. Each document is stored as a matrix in the data, with rows and columns indicating the number of distinct phrases in each text.

D. K-means Clustering

The goal of the k-means algorithm is to arrange data by increasing similarity within a cluster while decreasing similarity between clusters. A distance function is used to calculate similarity. The k-means approach uses the Euclidean distance to calculate the distance between data x in the Ci cluster and the ci centroid. The shortest gap between the data and the centroid position indicates maximum data similarity. The initial centroid, which is picked at random, has a significant impact on the k-means output. This study uses k-means clustering, a partition-based method that is basic, straightforward to use, and has a short calculation time. As a result, it is ideal for handling complex computational challenges in recommendation system similarity metrics. Clustering is accomplished by grouping user and rating data.

IV. IMPLEMENTATION

A. Dataset

We have picked a movies dataset which comprises of 45466 entries, books dataset which has 271360 entries, music dataset has 8010 entries.

B. Pre-Processing of Data

Pre-processing includes converting data into clean and quality data. Data integration, data cleaning, and data transformation are the main steps in data pre-processing. In data integration step, the data is collected from various sources & is integrated. During data cleaning, data containing null values and unnecessary rows with null values will be deleted. Data transformation includes object scaling, smoothing, aggregation, discretization of data. Good data pre-processing in machine learning is the most important factor that can make the difference between a good model and a bad machine learning model. So, we have to pre-process the data to get perfect accuracy.

Users with less than 50 ratings and books less than 100 ratings are excluded. (ONE TIME RUN)

```
[3]: # Users with less than 50 ratings are excluded:
     counts1 = ratings['User-ID'].value_counts()
     ratings1 = ratings[ratings['User-ID'].isin(counts1[counts1 >= 50].index)]
     # Books less than 100 ratings are excluded:
     counts2 = ratings['ISBN'].value_counts()
     ratings2 = ratings[ratings['ISBN'].isin(counts2[counts2 >= 100].index)]
```

Fig.11: Pre-Processing of Data

C. Fitting the Model in Algorithms

In order to predict, we split the dataset into train and test datasets for model fitting. The model is trained on the training set and predictions are made on the test set. Accuracy is checked using actual and predicted values after training. After training the model, a pickle file is made to reuse the model in the future.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. RESULTS AND DISCUSSIONS

To overcome the shortcomings of existing systems, we propose a mixed recommendation system. Combined recommendations make collaborative and content-based predictions, it then combines them single model by adding content-based capabilities to a collaboration-based method (and vice versa). For example, using a hybrid recommendation system, the site makes recommendations by comparing viewing and searching habits and search made by similar users.

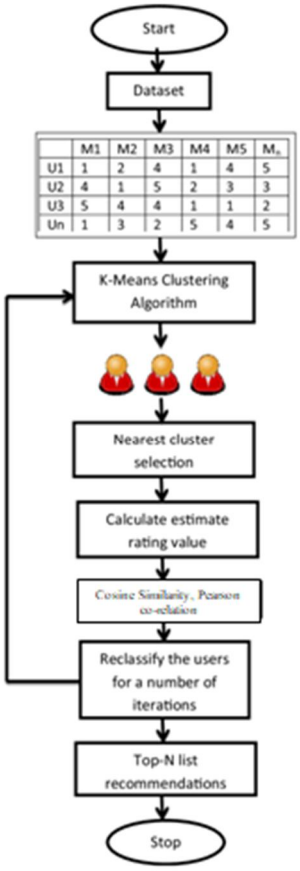


Fig2: Methodology of recommender system

Following is the output:

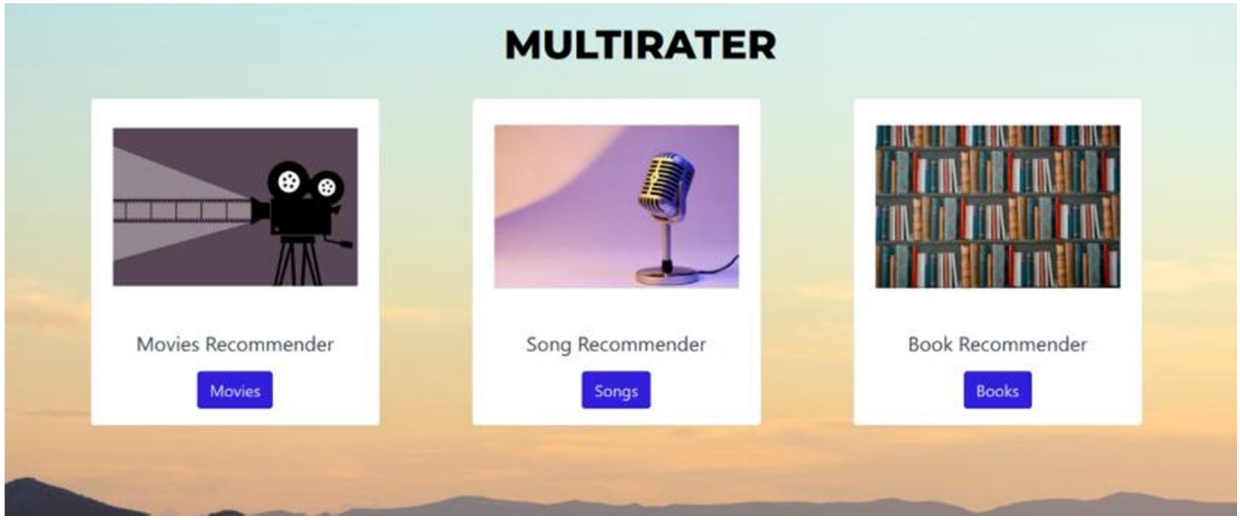


Fig3: Home Page UI

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

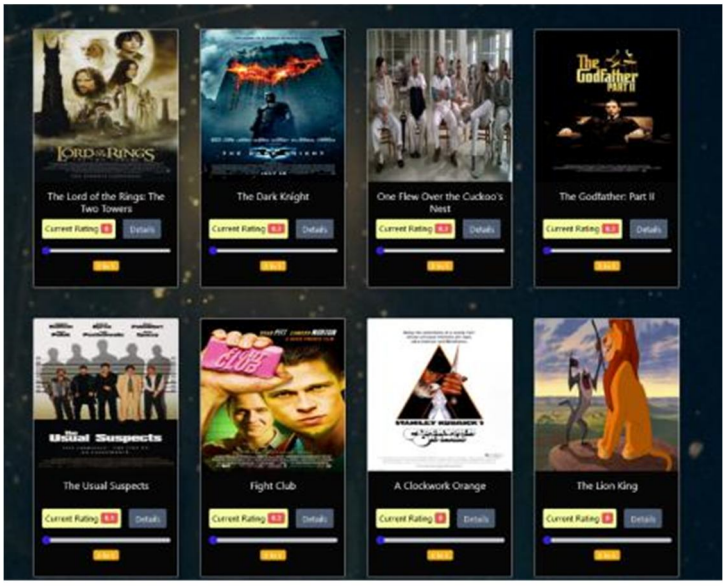


Fig.4: Rate Movies

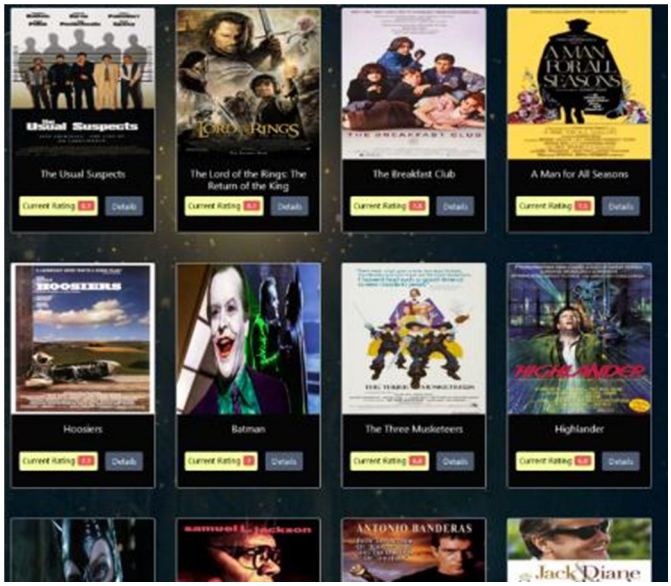


Fig.5: Recommended Movies



Fig.6: Movies Card



Fig.7: Rate Songs



Fig.8: Songs Recommendation

©IJRASET: All Rights are Reserved



Fig.9: Rate Books



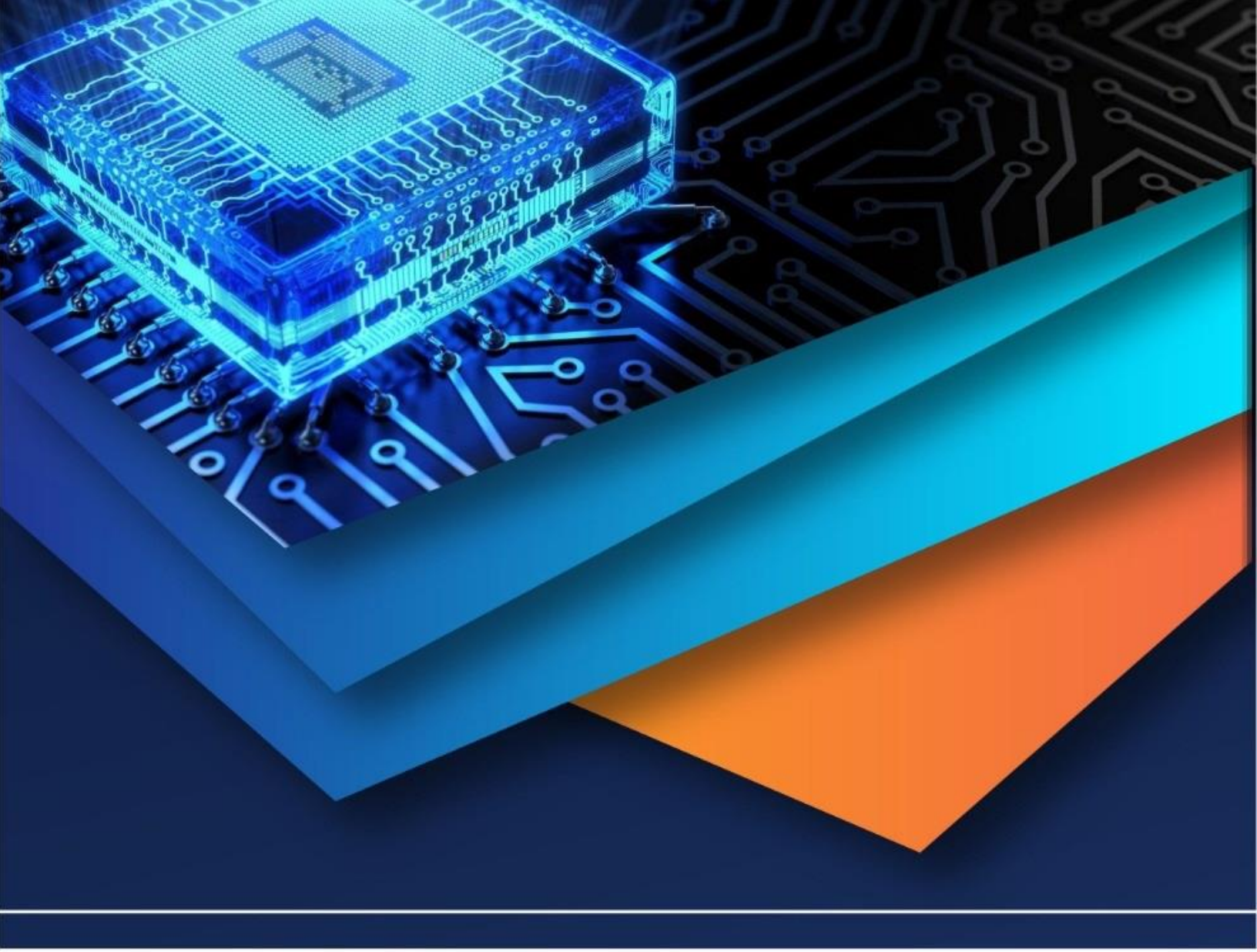
Fig.10: Books recommendation

VI. CONCLUSION

The media recommendation system allows a user to rate a number of movies, books and songs and then recommend respective media lists based on the user input. Due to the nature of the system, evaluating performance is a difficult task because there is no incorrect recommendation; it is simply a matter of opinion. Recommender systems open up new avenues for obtaining personalised information on the Internet. We devise a strategy that focuses on the user's personal interests, and media is recommended to users based on his previous reviews. This strategy aids in the improvement of recommendation accuracy. It also aids in the collection of authentic data with greater accuracy, making the system more responsive. We received a positive response from a small group of users after conducting informal evaluations. A large dataset is required by our system to produce efficient results. Furthermore, we would like to compare the results by combining various machine learning and clustering algorithms. We hope to eventually create a web-based application with a database and a learning model specific to each user.

REFERENCES

[1] Cui, Bei-Bei. (2017). Design and Implementation of Movie Recommendation System Based on Knn Collaborative Filtering Algorithm. ITM Web of Conferences. 12. 04008. 10.1051/itmconf/20171204008.
[2] 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.
[3] Kashyap, A., Sneha Srivastava and Anup Jung Shah. “A Movie Recommender System: MOVREC using Machine Learning Techniques.” (2020).
[4] Panwala, Kalp. (2020). "Various Types of Movies Recommendation System". 10.13140/RG.2.2.29566.41283.
[5] W. Wenzhen, "Personalized Music Recommendation Algorithm Based on Hybrid Collaborative Filtering Technology," 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA), 2019, pp. 280-283, doi: 10.1109/ICSGEA.2019.00071.L
[6] Ifada, Noor & Rahman, Triyani & Sophan, Mochammad. (2020).Comparing Collaborative Filtering and Hybrid based Approaches for Movie Recommendation. 219-223. 10.1109/ITIS50118.2020.9321014.
[7] 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.
[8] Muppana, Mahesh & Kanmani, R. & B, Surendiran. (2020). Analysis of Movie Recommendation Systems; with and without considering the low rated movies. 1-4. 10.1109/ic-ETITE47903.2020.453.
[9] M. Sunitha, T. Adilakshmi, Comparison Of User-Based Collaborative Filtering Model For Music Recommendation System With Various Proximity Measures, International Journal Of Innovative Technology And Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6S2, April 2019.
[10] Awawdeh, S.; Edinat, A.; Sleit, A. An Enhanced K-Means Clustering Algorithm for Multi-Attributes Data. Int. J. Comput. Sci. Inf. Secur. 2019, 17. Available online: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=An+Enhanced+K-means+ Clustering+Algorithm+for+Multi-+attributes+Data&btnG= (accessed on 14 April 2021).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)