WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Streamlining Movie Recommendations with Machine Learning

Soleti Youvasri

B. Tech Bobbili,535558, India

A B S T R A C T

Recommender systems generate for the user an ordered list of items most likely to be purchased. These are nothing but recommender systems that are utilized in numerous e-commerce websites to draw in the customers to buy the things which they are selling. In movie recommendation, it recommends the relevant movie to the users which depend on their past preferences. In the recommender systems, the frequently used approach is Collaborative Filtering Approach. It is the process of predicting the interests of a user according to the previous selections of the other users. The underlying psychological assumption of this is that 'people who think similarly tend towards making similar choice'. Various websites like Netflix, Amazon Prime Video also uses Recommender system in recommending movies to the customers based on the browsing history of the customer in the past. There are different algorithms which are used in the movie recommending, which include K-Means Clustering, K-Nearest Neighbour, and Singular Value Decomposition. CF Recommendation systems have disadvantages like "cold start problem," "data sparsity," and "grey sheep problem."

Keywords: *Recommendation systems, personalized suggestions, collaborative filtering , machine learning, user satisfaction.*

Introduction

Recommender systems are life-blooming applications in this information technology era. They actually help users navigate the number of overwhelming choices and perhaps guide people into enjoying the content, services, or products when faced with paralyzing options. In the age of choice, this power makes decision much easier and dramatically increases user experience. Based on a detailed report of purchase or search history, RS helps users locate goods to which they are likely to be interested.

Recommender Systems are broadly classified into Personalized and Non-Personalized. The Personalized RS generates a personalized recommendation based on the profile and preferences of various users and predicts user needs by drawing inferences from past behavior and choices. These systems use the knowledge gathered from past user activities to offer relevant options that share their tastes. Non-personalized RS: Draws patterns at a particular time for everybody, and is not pointed to individual preferences. In such systems, suggestions go to individual profiles rather than trending options liked by most. Here, Non-Personalized RS ignores specific user interests and targets general preferences.

The three major approaches of Content-Based Filtering, Collaborative Filtering and Hybrid Methods are used in a Personalized Recommender System. Content-based filtering suggests items like those a user liked or on which she/he interacted before. This approach bases its process on what the item features are; thus, suggested items should always be relevant to the user based on past preferences. This could be very well personalized to specific interest qualities.

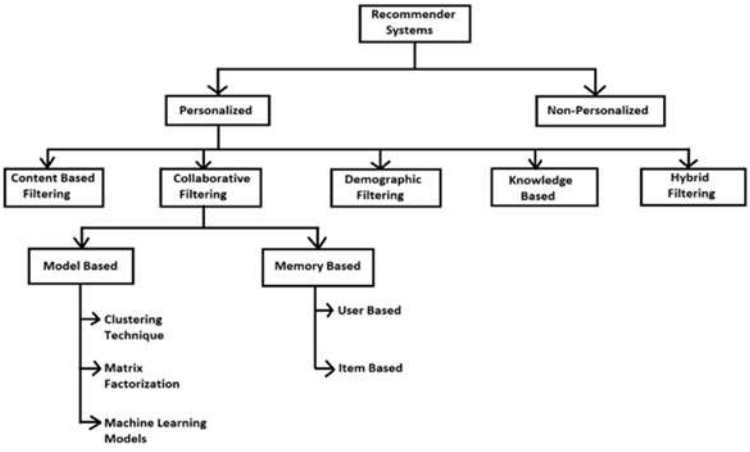


Fig 1: Classification of Recommender Systems

It predicts what a user will like by analyzing the behavior of similar users. There are two categories: Memory-Based techniques, including User-Based and Item-Based filtering, and Model-Based techniques. The Memory-Based techniques depend on historical information about similar users or items, whereas Model-Based techniques use algorithms and statistical models in order to analyze user data for accurate preference predictions. These methods are extensively used in modern digital platforms, especially in the recommendation system of Netflix, which uses high-order RS algorithms based on user habits and preferences to select movies and TV shows.

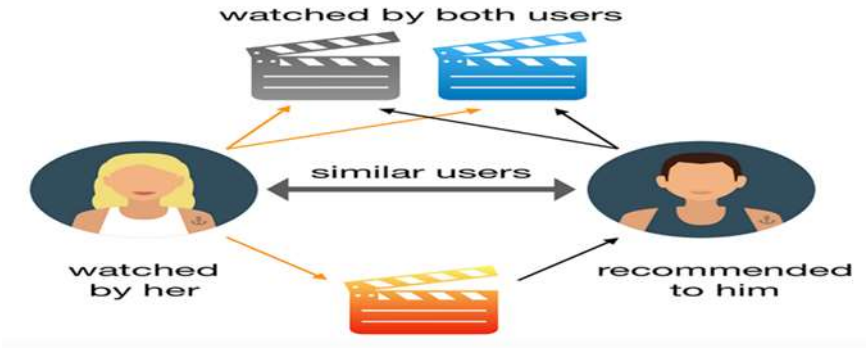


Fig 2: Collaborative VS Content-Based Filtering

By analyzing past user behavior, Netflix will be able to make the appropriate viewing suggestions as this increases enjoyment and satisfaction on the site. Including Recommender Systems in our daily digital life improves the quality of user satisfaction and interactions and makes these systems an integral part of the digital world. With growing efficiency of RS, an even higher human requirement follows to obtain personalized advice in our complex digital

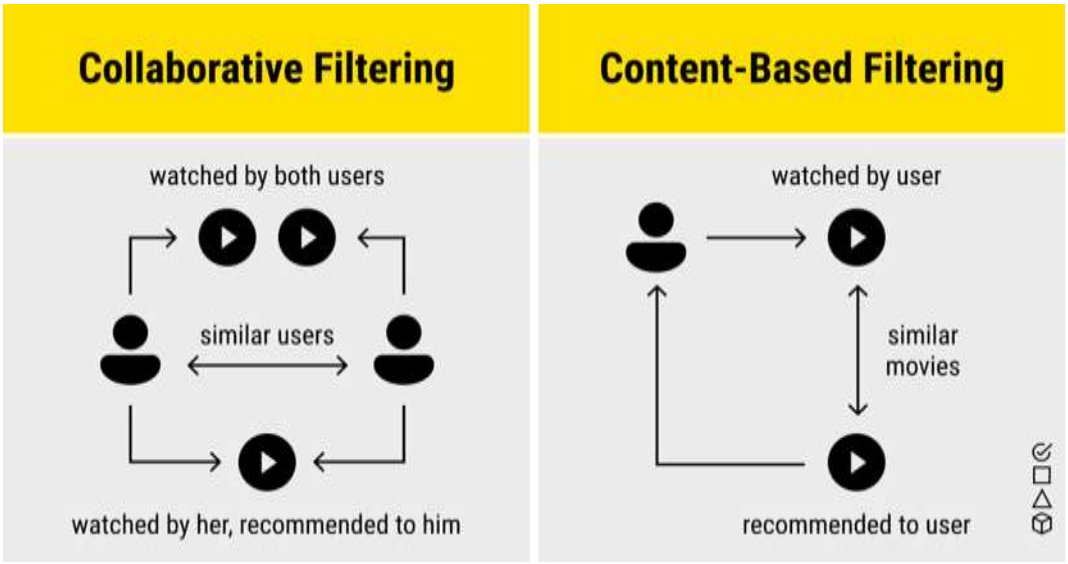


Fig 3:Illustrating Collaborative Filtering

environment. The future changes will guarantee to make user experiences through relevance and richness of recommendations shape digital content and services. . .

LITERATURE SURVEY:

It will use machine learning as a collaboration and content-based filtering service with MovieLens, Film Trust, and Kaggle. It employs text vectorization, which minimizes user bias and also minimizes the cold-start problem with the use of new user demographics for first suggestions[1]. Using Genre, Cast, and Description Based on Content-Based Recommender System, this paper predicts movie popularity. A CNN model predicts the popularity of groups based on junior, teenage, mid-age, and the senior categories. The model was able to achieve 96.8% accuracy as it used IMDb and TMDb datasets, doing greater than benchmark models[2]. This paper investigates a collaborative filtering-based movie recommendation system with the ALS model on the MovieLens 100K dataset, which achieves an RMSE of 0.9761 for the top 25 predictions in 517 seconds, and increases this to 97% accuracy with 1000 predictions. RMSE refers to Root Mean Square Error[3]. Wibowo and Baizal (2022) propose a movie recommendation system using knowledge-based filtering and K-Means clustering. It personalizes suggestions based on genre, director, and cast, allowing users to find similar films. This hybrid method enhances accuracy and user satisfaction over traditional techniques[4]. This paper presents the UPC Sim algorithm for movie recommendations using User Profile Correlation-based Similarity alongside collaborative filtering. It incorporates k-nearest neighbors and three similarity modules rating value, behavior value, and weighting[5]. This paper proposes hybrid filter movie recommendations with regard to genre labels obtained from the ML-100K and ML-25ml datasets. This algorithm predicts unrated movies with regard to expert genres, focusing on nonlinear genre similarities in ML-100K for user ratings. This algorithm merges content-based and collaborative filtering for better feature extraction[6]. This paper would provide a method to recommend unreleased movies via sentiment analysis and a hybrid engine. It analyses the comments on YouTube regarding the trailer, measures the sentiment, predicts their performance, aggregates the top movies from users, and provides recommendations based on preferences[7]. Movie tweets analyze trends and sentiment for a recommendation system. Sentiment analysis classifies opinions as positive, negative, or neutral. The TextBlob library measures review polarity and subjectivity using datasets such as Movielens 100K, Movielens 20M, IMDb, and Netflix[8]. This paper presents a movie recommendation system using personal data and ratings, applying the k-clique method and normalized discounted cumulative gain to address cold-start problems and enhance accuracy by clustering users in social networks[9]. This paper introduces a movie recommendation system that analyzes user reviews to gauge sentiment for suggestions. Reviews are categorized as positive, negative, or neutral. The JUMRv1 dataset involves three steps: 1. Word embedding 2. Feature selection 3. Classification. IMDB reviews are organized accordingly. Future work can look into different feature extraction methods, such as transformers and n-grams[10]. This paper applies LDA to identify movie comment topics and emotional trends. BERT was applied for the sentiment classification, while LDA extracts topics from a Douban movie review database. Sentiment analysis assesses the emotional polarity of texts through mining techniques[11]. Hybrid Recommender Systems with cooperative as well as content-based methods for movie recommendations. Here I am presenting a method that uses recent purchases and demographics to improve the recommendations, which eventually could help remove the cold start problem. Future work: Location aware as well as contextual features[12]. The paper discusses a movie recommendation system utilizing machine learning and filtering methods on the MovieLens datasets to predict user preferences, using techniques like Alternating Least Squares, with performance evaluated by RMSE and MAE[13]. Sajwan (2023) presents a hybrid movie recommendation system that merges collaborative and content-based filtering via machine learning. It preprocesses data, uses SVD and deep learning for clustering, and evaluates with MAE and RMSE, addressing cold start and sparsity while allowing real-time updates and user feedback[14]. This paper introduces a neural network model to movie recommendation using explicit and implicit data: NCF with weighted perceptrons. MovieLens contains 20 million entries and has a Hit-Ratio of 87. Neural networks increase the efficiency of recommendations but, as it stands, cannot bypass the cold start problem[15]

METHODOLOGY:

Singular Value Decomposition:

The Singular Value Decomposition (SVD), a method from linear algebra that has been generally used as a dimensionality reduction technique in machine learning. SVD is used as a collaborative filtering technique. It uses a matrix structure where each row represents a user, and each column represents an item.SVD is a technique, which decomposes any matrix into 3 generic and familiar matrices.The matrix is decomposed into a product of a square matrix, a diagonal (possible rectangular) matrix, and another square matrix.The concept of eigenvalues and eigenvectors is used in Singular Value Decomposition.Eigenvalues are the unique set of scalar values associated to with set of linear equations most commonly found in the matrix equations. The term "characteristic roots" also refers to the eigenvectors.When a certain matrix is multiplied, the eigenvector of a square matrix is said to be a non-vector that is equal to a scalar multiple of that vector.

'$Av = \lambda v$' is known as eigen vector equation

1. Where v = Eigenvector and λ be the scalar quantity that is termed as eigenvalue associated with given matrix A.

Steps involved in finding SVD of a matrix:

Let us take a sample matrix to which SVD is to be determined and denote it with some letter as '**A**'.Firstly we compute the singular values σ_i by finding the eigenvalues of $\mathbf{AA^T}$.

The characteristic polynomial equation to find singular values is $\det(AA^T - \lambda I) = 0$.

Now we find the right singular vectors (the columns of V) by finding an orthonormal set of eigenvectors of A^TA. It is also possible to proceed by finding the left singular vectors (columns of U) instead.

After performing some row reductions and unit length vector calculation we determine $\mathbf{A = U\Sigma V^T}$.

The final matrix is divided into 3 matrices in the form as $A = U\sum V^T$

Here, U = Left Singular Vectors

$\sum$ = Singular Valu

V^T = Right Singular Vectors

Where, *C* is a *m x n* utility matrix

U is a *m x r* orthogonal left singular matrix, which represents the relationship between users and latent factors

$\sum$ is a *r x r* diagonal matrix, which describes the strength of each latent factor

V is a *r x n* diagonal right singular matrix, which indicates the similarity between items and latent factors.

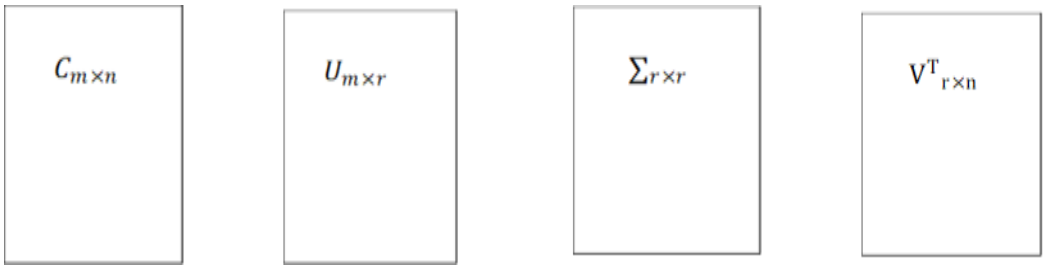


Fig 2: Singular Value Decomposition Division

K-Means Clustering Algorithm:

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. This groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process. It is a centroid-based algorithm, where each cluster is associated with a centroid.This algorithm's primary goal is to reduce the total distances between each data point and its corresponding clusters.The k-means clustering algorithm mainly performs two tasks:1.Iteratively chooses the optimal value for K centre points or centroids.2.Identifies the nearest k-center for each data point. A cluster is formed by the data points that are close to a given k-center.

Steps involved in K-Means Clustering Algorithm:

Initially select **'K'** number of random datapoints which may or may not be of input dataset as centroids.

Now calculate the distance between datapoints and the centroids.

Assign the datapoints to closest centroid to form **'K'** predefined clusters.

Since we need to find the closest cluster, so we will repeat the process by choosing a new centroid. Again calculate the distance between datapoints and new centroids.This process of calculating the distance and repositioning the centroid continues until we obtain our final cluster.

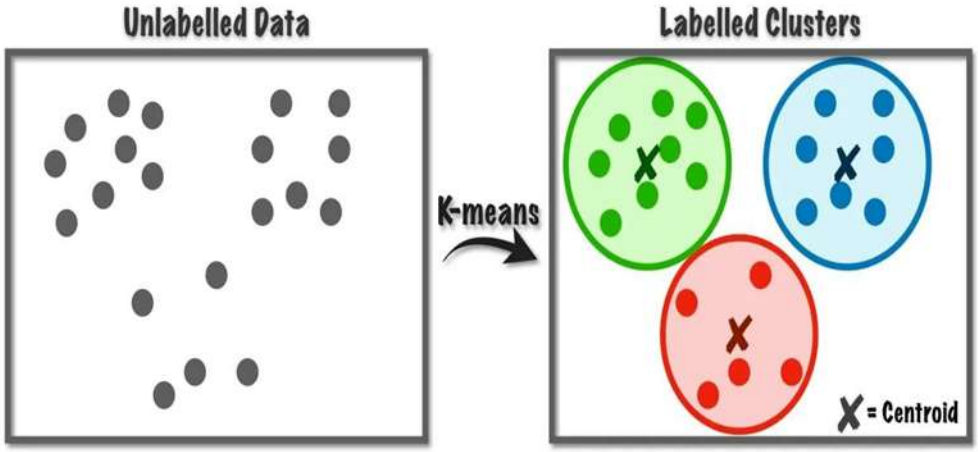
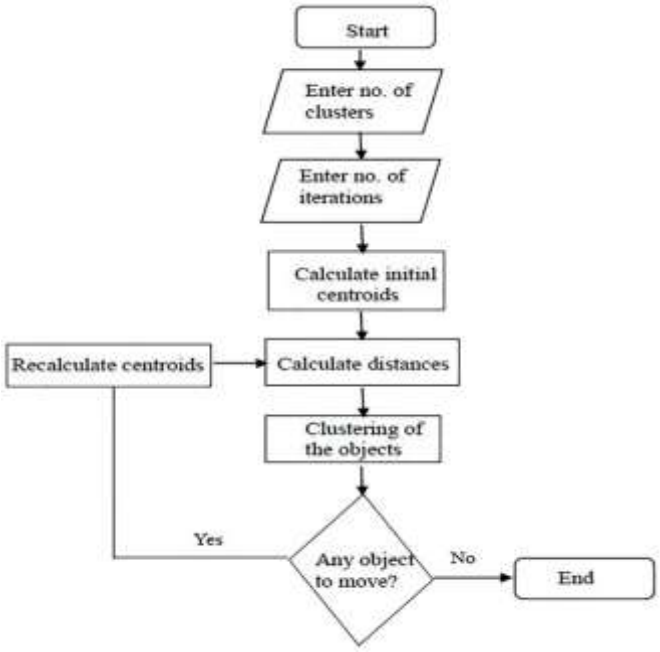


Fig 4: Labelled vs Un-labelled Data

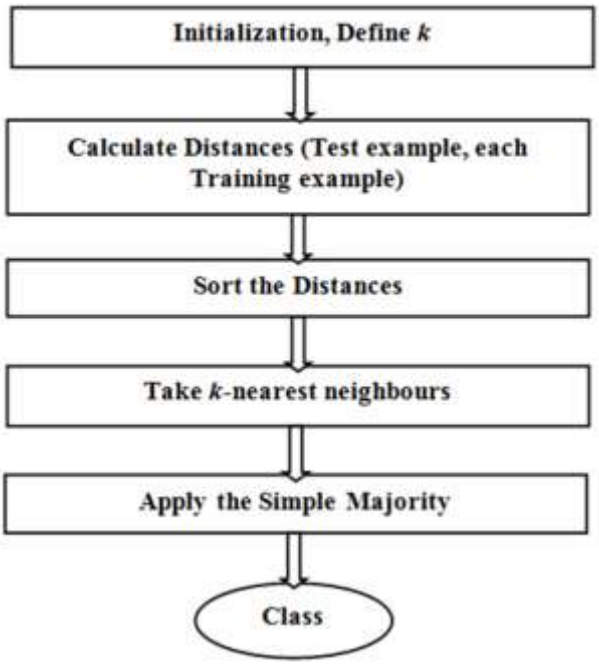


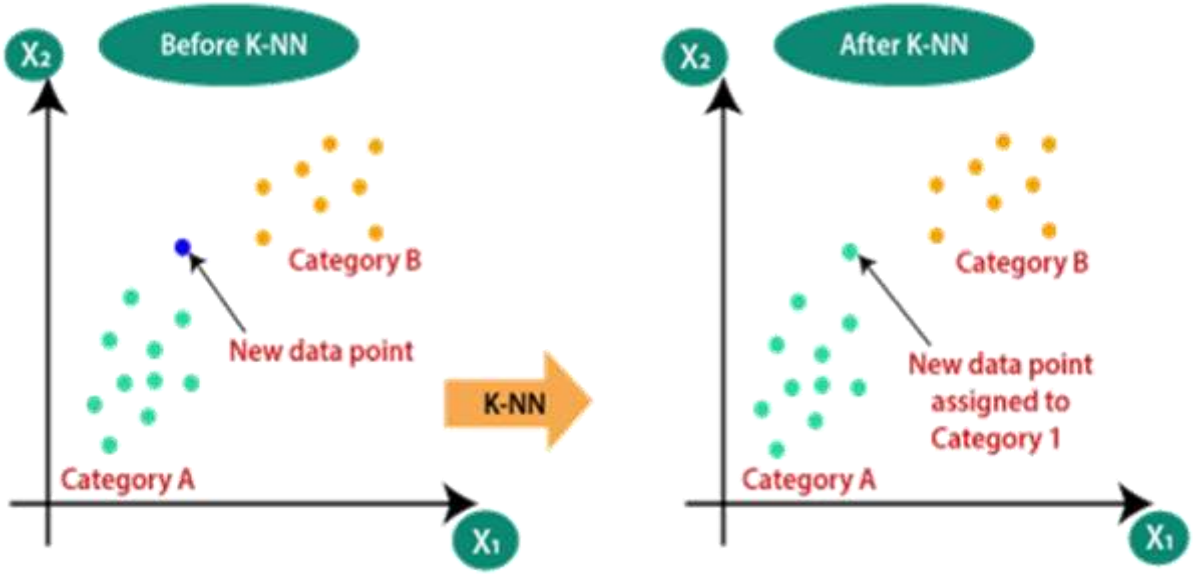
K-Nearest Neighbour (KNN):

It is a supervised machine learning algorithm which can be used for both classification as well as regression predictive problems.K-NN method assumes similarity between new case/data and existing cases and assigns new case into category most similar to existing categories.K-NN algorithm is used to store datasets at training phase and based on new data available later it further classifies the new data which is more similar to existing data.The K-NN algorithm is **non-parametric**, which means it doesn't make any assumptions about the underlying data.It is also known as **a lazy learner algorithm** since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data.The value of **K** is predefined in this algorithm. The most preferred value for K is 5. A very low K number, such as K=1 or K=2, might be noisy and cause outlier effects in the model. Although K may have large values, but there might be some challenges.

Steps Involved in working of KNN:

Load the dataset.Initialize K to your chosen number of neighbors.Calculate the distance between the new datapoint and the existing datapoints.Now based on distance value obtained sort them in ascending order.Next choose top K rows from sorted order.Now, assign a class to the new datapoint based on most frequent class of these rows.





RESULTS:

The aggregated MovieLens dataset of the University of Minnesota had already been applied to test the performance of the adopted model in this study. K-means clustering is a technique for the categorization of users based on their individual interests, thus allowing the grouping of similar preferences across the users. This will allow the uncovering of patterns of similarity among them and eventually support the generation of recommendations customized for the active user. The K-means clustering method showed the promise of getting a good accuracy of around 78.33%. On the other hand, the KNN algorithm works in an entirely different way; it computes the "distance" of the movie the user is interested in and all other movies stored within its database.The algorithm then ranks the movies based on proximity to the target movie and returns the best K movies closest, considered the most similar recommendations for the user. However, this algorithm tends to do relatively less in terms of accuracy in results and because of this limitation, it is usually regarded as a lazy learner algorithm that relies on distances rather than having intrinsic knowledge of the data. Finally, we also make use of Singular Value Decomposition (SVD) to devise another movie recommendation algorithm based on user ratings.SVD is one of the best known among matrix factorization techniques that assist the extraction of latent features from the given data. However, the speed of the model cannot be computed exactly and hence creates a problem for it to be efficient. Thus, K-means and SVD share two unique strengths and weaknesses in the realm of recommendation systems: accuracy in recommendations versus speed in processing algorithm.

Algorithm	Accuracy
K-Means Clustering	78.33%
K- Nearest Neighbor	96.67%

CONCLUSION:

This paper evaluates various CF approaches for movie recommendation systems, which are content-based and collaborative filtering methods. Content-based filtering provides personalized recommendations without any user data; however, this leads to a 'filter bubble' around past interests. Collaborative filtering generates better recommendations based on similar tastes, but the other problems it experiences are 'sparsity' and the 'new user' problem. The algorithms considered include K-Means Clustering, K-Nearest Neighbor, Singular Value Decomposition, and Alternating Least Squares. As much, clustering proves to demonstrate good accuracy. There are improvements introduced, yet challenges arise in the form of dealing with cold starts and making the provided recommendations more personalized.

REFERENCES

[1] Giridharan, N. & Nathan, K. & Swetha, M.. (2022). Movie recommendation system using machine learning. International journal of health sciences. 498-505. 10.53730/ijhs.v6nS8.9728.

[2] S. Sahu, R. Kumar, M. S. Pathan, J. Shafi, Y. Kumar and M. F. Ijaz, "Movie Popularity and Target Audience Prediction Using the Content-Based Recommender System," in IEEE Access, vol. 10, pp. 42044-42060, 2022, doi: 10.1109/ACCESS.2022.3168161.

[3] Wibowo, K. D., & Baizal, Z. K. A. (2022). Movie Recommendation System Using Knowledge-Based Filtering and K-Means Clustering. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 460-465.

[4] Widiyaningtyas, T., Hidayah, I., & Adji, T. B. (2021). User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system. *Journal of Big Data*, *8*(1), 1-21.

[5] Roy, A., & Ludwig, S. A. (2021). Genre based hybrid filtering for movie recommendation engine. *Journal of Intelligent Information Systems*, *56*(3), 485-507.

[6] Awan, M. J., Khan, R. A., Nobanee, H., Yasin, A., Anwar, S. M., Naseem, U., & Singh, V. P. (2021). A recommendation engine for predicting movie ratings using a big data approach. *Electronics*, *10*(10), 1215.

[7] Sahu, S., Kumar, R., MohdShafi, P., Shafi, J., Kim, S., & Ijaz, M. F. (2022). A Hybrid Recommendation System of Upcoming Movies Using Sentiment Analysis of YouTube Trailer Reviews. *Mathematics*, *10*(9), 1568.

[8] Kumar, S., De, K., & Roy, P. P. (2020). Movie recommendation system using sentiment analysis from microblogging data. *IEEE Transactions on Computational Social Systems*, *7*(4), 915-923.

[9] Vilakone, P., Xinchang, K., & Park, D. S. (2020). Movie recommendation system based on users' personal information and movies rated using the method of k-clique and normalized discounted cumulative gain. *Journal of Information Processing Systems*, *16*(2), 494-507.

[10] Chatterjee, S., Chakrabarti, K., Garain, A., Schwenker, F., & Sarkar, R. (2021). JUMRv1: A Sentiment Analysis Dataset for Movie Recommendation. *Applied Sciences*, *11*(20), 9381.

[11] Zhang, Y., & Zhang, L. (2022).Movie Recommendation Algorithm Based on Sentiment Analysis and LDA. *Procedia Computer Science*, *199*, 871-878.

[12] Sujithra Alias Kanmani, R., Surendiran, B., & Ibrahim, S. P.(2021). Recency augmented hybrid collaborative movie recommendation system. *International Journal of Information Technology*, *13*(5), 1829-1836.

[13] Tahir and K. B. Ali, "Movies Recommendation System Using Machine Learning Algorithms," 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, 2023, pp. 1324-1328, doi: 10.1109/ICAICCIT60255.2023.10465879.

[14] A. Sajwan, L. Gopal, R. Chauhan and A. Balodi, "Movie Recommendation System Using Machine Learning," 2023 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC), Bengaluru, India, 2023, pp. 1-7, doi: 10.1109/TEMSCON-ASPAC59527.2023.10531374.

[15] Jena, K. K., Bhoi, S. K., Mallick, C., Jena, S. R., Kumar, R., Long, H. V., & Son, N. T. K. (2022). Neural model based collaborative filtering for movie recommendation system. International Journal of Information Technology, 1-11.

[16] Lavanya, R., & Bharathi, B. (2021, March). Systematic analysis of movie recommendation system through sentiment analysis. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (pp. 614-620). IEEE.

[17] rokiaraj, P., Sandeep, D. K., Vishnu, J., & Muthurasu, N. (2023). Movie Recommendation System Using Machine Learning. Advances in Science and Technology, 124, 398-406

[18] Marappan, R., & Bhaskaran, S. (2022). Movie recommendation system modeling using machine learning. International Journal of Mathematical, Engineering, Biological and Applied Computing, 12-16

[19] Liu, D., & Li, H. B. (2022). A Matrix Decomposition Model Based on Feature Factors in Movie Recommendation System. *arXiv preprint arXiv:2206.05654*.

[20] M. Chenna Keshava , P. Narendra Reddy , S. Srinivasulu , B. Dinesh Naik, 2020, Machine Learning Model for Movie Recommendation System, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue04(April2020)