



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 8, August 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Movie Recommendation System Employing Machine Learning Techniques

Mrs.S.Vanjimalar^{1*}, Dr.R.Saravana Moorthy², Dr.S.Velmurugan³ Mrs. N.Paviyasre⁴

Mrs. A. Immaculate⁵

^{*1}Assistant Professor, Department of Computer Science (SF), Kongunadu Arts and Science College,
Coimbatore, India

²Associate Professor and Head, Department of Computer Science (SF), Kongunadu Arts and Science College,
Coimbatore, India

³Assistant Professor, Department of Computer Science (SF), Kongunadu Arts and Science College,
Coimbatore, India

⁴Assistant Professor, Department of Computer Science (SF), Kongunadu Arts and Science College,
Coimbatore, India

⁵Associate Professor, Department of Computer Science (SF), Kongunadu Arts and Science College,
Coimbatore, India

ABSTRACT: Recommendation systems forecast consumer preferences for yet-to-be-viewed goods like movies, songs, or books. The purpose of movie recommendation systems is to foresee user interests and make recommendations for things that are likely to be of interest to them. Movie recommendation systems use a range of filtration methods and algorithms to help consumers find the most relevant films. From a business perspective, the more pertinent content or movies a user finds on a certain platform, the better their engagement and, consequently, the higher revenue. The most extensively used subcategories of machine learning algorithms for movie suggestions include content-based filtering, collaborative filtering systems, and hybrid filtering. Using machine learning, we can develop a model that can suggest movies based on prior data. We will look at various machine learning methods that can be used to recommend movies in this essay.

KEYWORDS: Movie recommendation system, collaborative filtering, content-based filtering, hybrid system, k-means, k-nearest neighbour (KNN), clustering algorithm

I. INTRODUCTION

As early as the late 19th century, the idea of machine learning had begun to take hold [1]. We need a recommender system mostly because there are simply too many options available to us in the modern world thanks to the internet. Today's globe has a rapidly expanding variety of entertainment options because to the entertainment industry's revolution. A recommendation system, often known as a recommendation engine, is a methodology for information filtering that aims to anticipate user preferences and offer suggestions in accordance with these choices [2]. The recommendation algorithm is mostly employed by e-commerce sites like Amazon, Flipkart, and eBay as well as digital entertainment services like Netflix, Prime Video, and IMDB. Web-based portals are what users rely on for movie recommendations. The genres of movies, such as comedy, suspense, animation, and action, can be used to distinguish them. The metadata for the movies, such as the release year, language, director, or actors, may also be used to categorize them most providers that stream videos online [3]. The fundamental concepts of "share" and "learn" have been developed into the major slogan of the present recommendation system. The heart of the recommendation system is thought to be the process of discussing and observing one another's ideas [4]. Because of its ability to provide enhanced entertainment, a movie recommendation is becoming increasingly popular as a part of our social lives. A Recommendation system is a type of information filtering system which is used to predict the "rating" or "preference" a user would give to an item [5]. The ability to predict user preferences and needs when

analyzing user behavior or other user behavior to produce a personalized recommender is a very helpful [6] feature in a recommendation system.

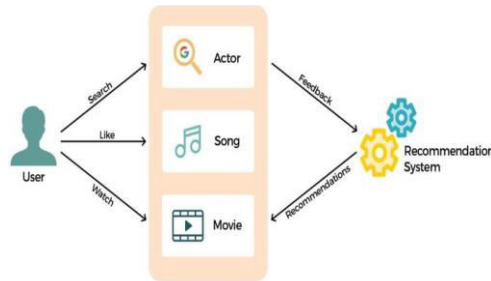


Fig 1: Recommendation system

Recommendation systems are typically classified into the following filtering techniques:

- Collaborative filtering
- Content-based filtering
- Hybrid filtering.

Collaborative Filtering

The majority of recommendation systems use collaborative filtering to identify similar user patterns or information [7]; this method can exclude items that users like based on the ratings or reactions of similar users

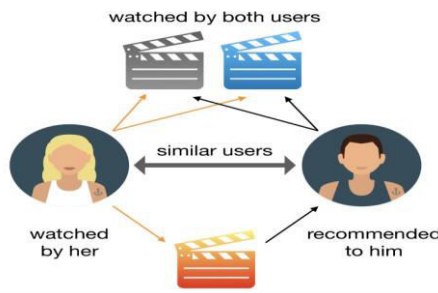


Fig – 2: Collaborative Filtering technique

Cooperative filtering comes in two flavors:

- A. **User-based:** measures the degree to which target users and other users are comparable.
- B. **Item-Based:** This method assesses how similar two items are that target user's rate or interact with.

Content-Based Filtering

According to the user's past behavior or explicit feedback, content-based filtering uses item features to suggest additional [8] items that are similar to what they already like.

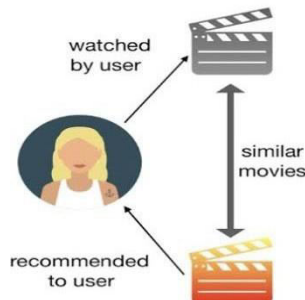


Fig – 3 : Content-based Filtering technique

The similarity between item vectors can be computed by three methods [9]:

- A. Cosine similarity
- B. Euclidian distance
- C. Pearson’s correlation

Hybrid Filtering

Which can be thought of as a hybrid [10] of the content- based and collaborative filtering methods. It avoids the flaws of all recommender techniques.

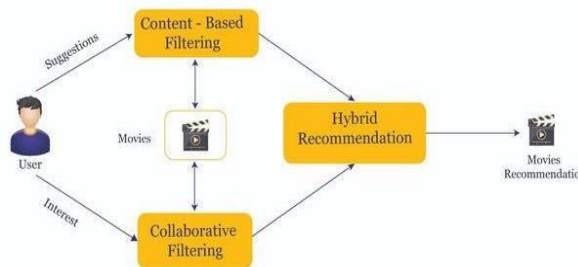


Fig – 4: Hybrid Filtering technique

Problems Related to the Recommendation System

There are some problems related to the recommendation system are According to [9, 11,12].

- A. **Cold-start problem:** when a user registers for the first time, he/she not watched any movie. So, the recommendation system does not have any movie based on which it can give result [9]. This called cold-start problem.
- B. **Data scarcity problem:** This problem occurs when the user has rated very few items based on which it is difficult for the recommendation system to accurate result [9].
- C. **Scalability:** In this, the encoding goes linearly on items. The system works efficiently when the data set is of limited size [9].

Machine Learning Algorithms

Which focuses on using data and algorithms to mimic how humans learn, gradually improving its accuracy. There are three types of machine learning techniques: supervised, unsupervised, and Semi-supervised and reinforcement learning. The machine learns undersupervision in Supervised Learning. It includes a model that can predict with the help of a labelled dataset. A labelled dataset is one in which the target answer is already known. The following supervised learning algorithms are widely used: decision tree, logistic regression, support vector machine, and k-nearest neighbors. Unsupervised Learning involves the machine learning on its own using unlabeled data. The machine attempts to find a pattern in the unlabeled data and responds. K-means clustering, Hierarchical clustering, and the apriority algorithm are the most commonly used unsupervised learning algorithms.

A. K-Means Algorithm

Clustering[8] is a process to group a set of objects in such a way that objects in the same clusters are more similar to each other than to those in other clusters.

Unsupervised learning algorithm K-Means Clustering [13] divides the unlabeled dataset into various clusters. Here, K specifies how many pre-defined clusters must be created as part of the process; for example, if K=2, there will be two clusters, if K=3, there will be three clusters, and so on.

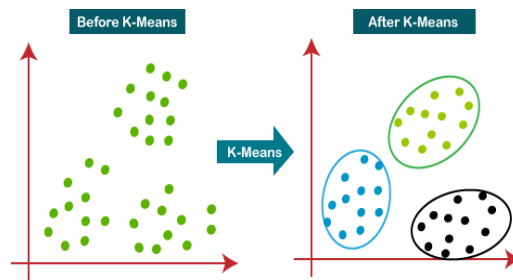


Fig – 5 : K-means clustering

B. K-Nearest Neighbors

The k-nearest neighbor algorithm, also known as KNN or k-NN, is a non- parametric, supervised learning classifier that uses proximity to classify or predict the grouping of a single data point. While it can be used for either regression or classification problems, it is most commonly used as a classification algorithm, based on the assumption that [14] similar points can be found nearby.

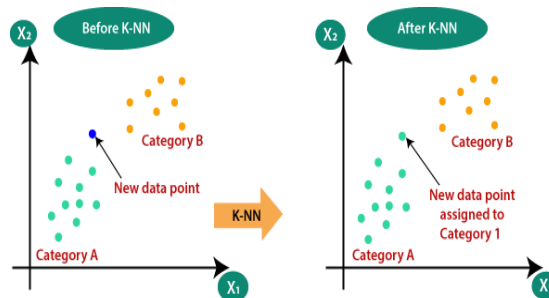


Fig – 6 : K- Nearest Neighbors

C. Support Vector Machine

A Support Vector Machine" (SVM) can be applied to classification or regression problems. However, classification issues are where it is most frequently used. When using the SVM algorithm, each data point is represented as a point in n- dimensional space, with each feature's value being the value of a specific coordinate. Next, we perform [15] classification by identifying the hyper-plane that effectively distinguishes.

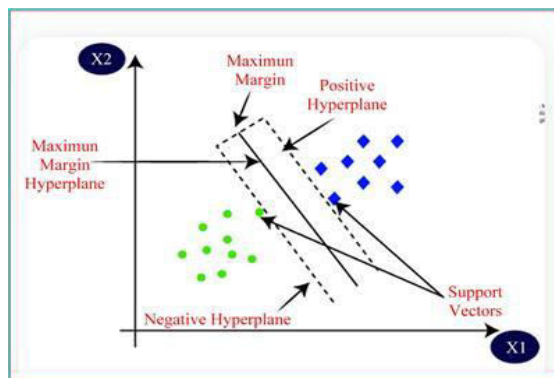


Fig – 7 : Support Vector Machine

II. LITERATURE REVIEW

Raghavendra C K and Srikantaiah K.C [16] are discussed about the various similarity metrics applied in recommendation tasks are compared by considering MovieLens dataset for developing a recommendation system for movies using user- based and item-based. that item-based CF gives better accuracy and faster than user based.



Accuracy respectively is 0.84 and 0.76. M Ashok, S Rajanna, PV Joshi [17] are discussed about Machine Learning and Sentiment Analysis based techniques are used for further optimizing search query results. This provides the user with quicker and more relevant data, thus avoiding irrelevant data and providing much needed personalization. Based on classification technique are naïve bayes, svm, maxEnt, Random forest more accurate result provide is maximum Entropy respectively 77.17% N Banik, MHH Rahman[18] discussed about polarity detection system on textual movie reviews in Bangla by using two popular machine learning algorithms

named Naïve Bayes (NB) and Support Vector Machines (SVM) and provided a comparative results where SVM performed slightly better than NB by considering stemmed unigram as feature with an excellent precision of 0.86. P Vilakone at. [19]All discussed about , to evaluate the performance; collaborative filtering methods are monitored using the k nearest neighbors, the maximal clique methods, the k-clique methods, and the proposed methods are used to evaluate. experiment and the output showed was more effective results is maximal clique method for accuracy is 59.63%. Choudhury, S. S., Mohanty, S. N., & Jagadev, A. K. [20] discussed about recommendation models such as Backpropagation (BPNN) model, SVD (Singular Value Decomposition) model, DNN (Deep Neural Network model) and DNN with Trust were compared to recommend the suitable movie to the user. Results imply that DNN with trust model proved to be the best model with high accuracy of 83% with 0.74 MSE value and can be used for best movie recommendation.

Problem Statement

According to literature study there are certain problem with the prediction movie for new user that called cold-start problem. It is at the bottom of the list of suggested products because it is a new item without any customer reviews. There for I used KNN and matrix factorization technique (singular value decomposition) to predict movie. Also, scalability problem in survey there for I compare my dataset on word most popular rating web IMDB.

III. PROPOSED SYSTEM

Our proposed model is based movie recommendation using two algorithms that fall into this category, K Nearest Neighbor (KNN) and Singular Value Decomposition (SVD). Here also compare two algorithm result compare the top 3 predictions given by the KNN vs. SVD. And also model evaluation using cross validation and test-train-split for measuresimilarity define between users and most accurate result providing.

- Step 1:** The first step is to use the Kaggle movie lens dataset to predict recommendedmovies for each individual.
- Step 2:** The Second step would be to introduce collaborative-based filtering algorithmsKNN and Matrices factorization SVM.
- Step 3:** The third step is to evaluate the model and apply the regression evaluation metricsto our recommendation system to train-test-split and cross validation
- Step 4:** final recommended best movie

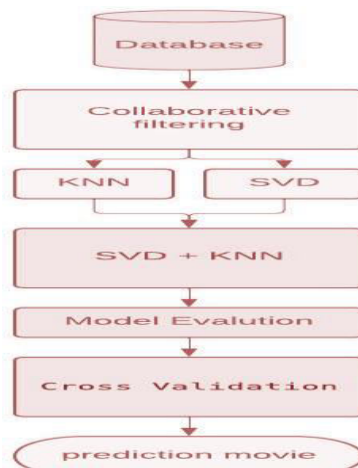


Fig – 8 : Proposed model

IV. EXPERIMENTAL RESULTS

As a result, we first perform SVD on the training data to generate a model. Then we repeat the process with KNN. We developed a user-based collaborative filtering model using KNN. We used cosine similarity to calculate the similarity between the K nearest neighbors in the KNN algorithm. We get forecasts for each user's movie ratings from both SVD and KNN. By averaging the estimated ratings of KNN and SVD, the findings are combined. In this splitting our data into train and testing groups, we used 5 cross-fold validation. We used Root Mean Square Error, Mean Absolute Error, precision, and recall as evaluation measures. For each algorithm, the precision and recall outcomes of the five cross fold validations were averaged.

Error Rate Measure

MAE: Using the MAE metric, we can determine the mean absolute difference between a dataset's predicted values and real values. The better a model matches a dataset, the lower the MAE.

$$MAE = 1/n * \sum |y_i - \hat{y}_i|$$

where:

Σ is a symbol that means "sum"

y_i is the observed value for the i th observation \hat{y}_i is the predicted value for the i th observation n is the sample size

RMSE: A measurement that reveals the square root of the typical squared discrepancy between a dataset's predicted values and real values. The better a model matches a dataset, the lower the RMSE. It is determined by:

$$RMSE = \sqrt{\sum (y_i - \hat{y}_i)^2 / n}$$

where:

Σ is a symbol that means "sum"

\hat{y}_i is the predicted value for the i th observation y_i is the observed value for the i th observation n is the sample size

Confusion Matrix

Precision: The proportion of properly classified positive samples (True Positive) to the total number of positively classified samples is known as precision. (either correctly or incorrectly).

$$\text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive}) \quad \text{Precision} = TP / (TP + FP)$$

Recall : The recall is calculated as the proportion of Positive samples that were properly identified as Positive to all Positive samples. The recall gauges how well the model can identify positive data. The more positive samples that are found, the higher the recall.

$$\text{Recall} = \text{True Positive} / (\text{True Positive} + \text{False Negative}) \quad \text{Recall} = TP / (TP + FN)$$

Dataset Description

The Movie lens 100k Dataset, which comprises of 100.000 ratings from 1000 users on 1700 movies, will be analyzed for this project. Each individual in this dataset has given at least 20 films a rating. Simple demographic data for the users, such as age, gender, and employment, is also provided in addition to this data. The collection is accessible via the permalink <http://grouplens.org/datasets/movielens/100k/>.

Results

In all metrics, the SVD model beats the KNN and the random predictor. It has the lowest RMSE, MAE, and recall, as well as the greatest precision. The KNN model is nearly as effective as the SVD model. Show In **Figure 9** and **Table 1** terms of RMSE, SVD is only 3.95% better, while MAE is 3.99% better. SVD also has a 3.94% higher accuracy and a 5.69% higher recall rate. Of course, both KNN and SVD outperform the random prediction algorithm. KNN, for example, has a 37.28% lower MAE and a 36.14% lower RMSE than the random prediction. It should also be observed that the difference in MAE and RMSE between the models is nearly identical.



ERROR RATES	SVD	KNN	SVD+ KNN
RMSE	0.9338	0.9728	0.9323
MAE	0.7354	0.7674	0.7379

Table 1 : Perform Measure for RMSE and MAE

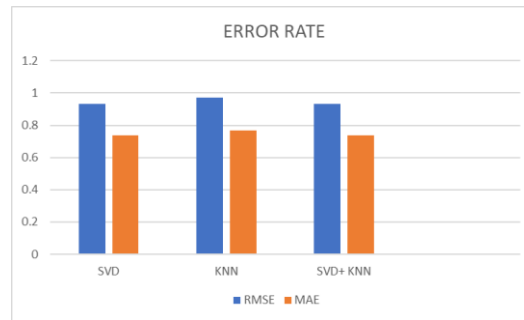


Fig – 9 : Compare Results RMSE and MAE

In all metrics, the SVD model beats the KNN and the random predictor. It has the lowest RMSE, MAE, and recall, as well as the greatest precision. Show in **Table 2** and **Figure 10** The KNN model is nearly as effective as the SVD model. In terms of RMSE, SVD is only 3.95% better, while MAE is 3.99% better. SVD also has a 3.94% higher accuracy and a 5.69% higher recall rate. Of course, both KNN and SVD outperform the random prediction algorithm. KNN, for example, has a 37.28% lower MAE and a 36.14% lower RMSE than the random prediction. It should also be observed that the difference in MAE and RMSE between the models is nearly identical.

Table 2 : Perform Measure for Precision and Recall

EVALUATION MATRIC	SVD	KNN	SVD+ KNN
PRECISION	0.8745	0.8436	0.8734
RECALL	0.2583	0.2739	0.2565

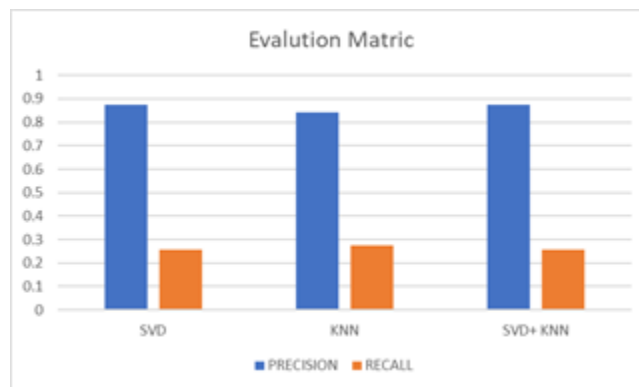


Fig – 10 : Compare Results Precision and Recall

V. CONCLUSION

Our combined model is quite precise. This signifies that the majority of the suggested things are relevant. The results demonstrate that the combination model, in which we summed the predicted ratings of the KNN and SVD models, does not outperform the SVD model alone. In reality, the results show that the SVD model outperforms the KNN



model on the 100k movie lens dataset, such that when the models are combined, the resulting model performs somewhat worse in most metrics (MAE, precision, and recall) than the SVD technique. As a result, combining the SVD and KNN models is not worth the effort, and we would fare better if we just employed one of them.

REFERENCES

- [1] Lin, C. H., & Chi, H. (2019, March). A novel movie recommendation system based on collaborative filtering and neural networks. In International Conference on Advanced Information Networking and Applications (pp. 895-903). Springer, Cham.
- [2] Wu, C. S. M., Garg, D., & Bhandary, U. (2018, November). Movie recommendation system using collaborative filtering. In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) (pp. 11-15). IEEE.
- [3] 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.
- [4] Subramaniam, R., Lee, R., & Matsuo, T. (2017, December). Movie Master: Hybrid Movie Recommendation. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 334-339). IEEE.
- [5] Ahuja, R., Solanki, A., & Nayyar, A. (2019, January). Movie recommender system using K-Means clustering and K-Nearest Neighbor. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 263-268). IEEE.
- [6] Singh, R. H., Maurya, S., Tripathi, T., Narula, T., & Srivastav, G. (2020). Movie recommendation system using cosine similarity and KNN. *International Journal of Engineering and Advanced Technology*, 9(5), 556-559.
- [7] Mahata, A., Saini, N., Saharawat, S., & Tiwari, R. (2016, December). Intelligent movie recommender system using machine learning. In International Conference on Intelligent Human Computer Interaction (pp. 94-110). Springer, Cham.
- [8] Ghazanfar, M., & Prugel-Bennett, A. (2010). Building switching hybrid recommender system using machine learning classifiers and collaborative filtering. *IAENG International Journal of Computer Science*, 37(3).
- [9] Raghavendra, C. K., & Srikantiah, K. C. (2021, May). Similarity Based Collaborative Filtering Model for Movie Recommendation Systems. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1143-1147). IEEE.
- [10] Banik, N., & Rahman, M. H. H. (2018, September). Evaluation of naïve bayes and support vector machines on bangla textual movie reviews. In 2018 international conference on Bangla speech and language processing (ICBSLP) (pp. 1-6). IEEE.
- [11] Ashok, M., Rajanna, S., Joshi, P. V., & Kamath, S. (2016, March). A personalized recommender system using machine learning based sentiment analysis over social data. In 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-6). IEEE.
- [12] Choudhury, S. S., Mohanty, S. N., & Jagadev, A. K. (2021). Multimodal trust based recommender system with machine learning approaches for movie recommendation. *International Journal of Information Technology*, 13(2), 475-482.
- [13] Vilakone, P., Park, D. S., Xinchang, K., & Hao, F. (2018). An efficient movie recommendation algorithm based on improved k-clique. *Human-centric Computing and Information Sciences*, 8(1), 1-15.
- [14] S. (2022, December 10). System Design Interview: Recommendation System Design (As Used By YouTube, Netflix etc.). Medium. Retrieved December 13, 2022, from <https://medium.com/double-pointer/system-design-interview-recommendation-system-design-as-used-by-youtube-netflix-etc-c457aacc3ab>
- [15] Kumar, A. (2022, August 19). Recommender Systems in Machine Learning: Examples - Data Analytics. Data Analytics. Retrieved December 13, 2022, from <https://vitalflux.com/recommender-systems-in-machine-learning-examples>
- [16] Primaldi, E. (2019, June 25). How to build a content-based movie recommender system with Natural Language Processing. Medium. Retrieved December 13, 2022, from <https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243>.
- [17] www.javatpoint.com.
- [18] <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>



- [19] <https://www.javatpoint.com/k-means-clustering- algorithm-in-machine-learning>
- [20] www.javatpoint.com/machine-learning-support-vector-machine-algorithm
- [21] Hinno, R. (2021, January 31). Simple SVD algorithms. Medium. Retrieved December 13, 2022, from <https://towardsdatascience.com/simple-svd-algorithms- 13291ad2eef2>



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details