International Journal of Recent Development in Engineering and Technology
www.ijrdet.com (ISSN 2347-6435 (Online) Volume 14, Issue 06, June 2025 Crossref DOI 10.54380/IJRDET0625_13)

Enhancing Movie Recommendations through a Hybrid Machine Learning Framework

Urvi Upadhyay

Department of Computer Engineering, Shri Govindram Seksaria Institute of Technology & Science, Indore

Email: upadhyayurvi718@gmail.com

***Abstract*- In the era of digitalization where our most of day-to-day activities are accomplished with the help of mobile phones via internet recommender system plays a vital role in boosting the entire industry by making us feel the pleasure on physical shopping sitting at home. In this regard recommender system helps with providing correct choices to users on the basis of their interest and likings.**

The present paper proposes a hybrid recommender model by incorporating feature engineering. In the proposed approach we add the content-based filtering model (where user's ratings based on average rating for movies are calculated) with collaborative filtering model (where this model uses K Nearest neighbor algorithm). Prediction with an RMSE of 0.925 is obtained by our proposed hybrid recommender model.Our proposed technique gave an improvement over recommendations with an RMSE of 0.982.

***Keywords*-Collaborative filtering, movie recommendation, SVD, content-based filtering**

I. INTRODUCTION

Programs that predict items to user's on the basis of their likings and behavior are called Recommender systems.Today majority of recommender system are based on providing the most appropriate items which can influence and boost up sales of e-commerce. [1-4]. When creating any recommender system, the user should always come first in order to ensure that the things or products being recommended are providing the highest level of pleasure [5–6]. A large number of recommender systems are found in the literature. However, the bulk of recommender systems do tend to recommend popular, easily recognizable, or routine items in their list of suggested items [7-8].

Due to the absence of novelty and serendipity in these offerings, most recommender systems encounter problems like "Popularity Bias," "Long Tail" item ignorance, "Matthew Effect," and other related problems [5–12]. The popular products in the catalog have a tendency to become even more popular as a result of these shortcomings in the traditional recommender systems, adding to the ever-longening list of "Long Tail" or "Non-Popular" items that are waiting to be recommended for an infinite amount of time and eventually starve [5–10].

Numerous recommender systems are used by large corporations. Today a number of e-commerce companies are also making use of recommender systems to increase their sales. Since the middle of the 1990s, numerous studies have been conducted. It was during that time that the term

"Recommender System" first appeared. These systems can be viewed as an effective method of information filtering and data refinement (Adomavicius et al., 2020).Recommender systems are divided into many types based on the attributes they employ. For instance, the use of the location and the surrounding environment results in the development of a class of recommender systems known as domain-based systems, or system knowledge of the user, giving the context in which knowledge-based systems have emerged. The following categories apply to these systems according to Jannach et al. (2010):

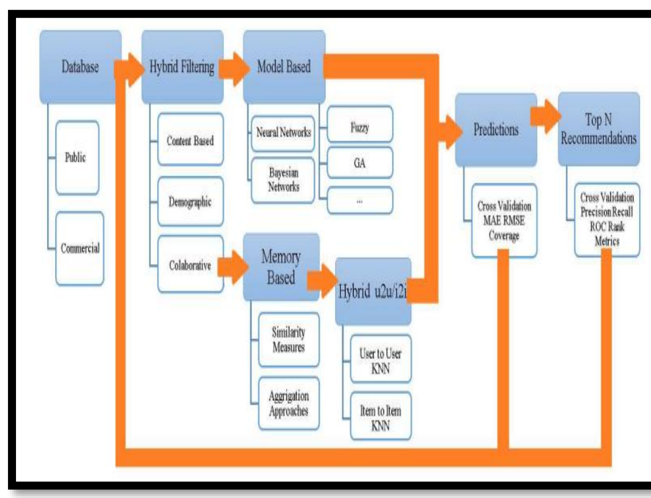

Fig. 1. Types of recommender systems and their relations

Content-based model: a user receives recommendations based on the content of his or her profile.

- Collaborative-based model: recommendations that other users have found useful with comparable circumstances (that is, comparable inclinations or preferences) are offered to the user,

International Journal of Recent Development in Engineering and Technology
www.ijrdet.com (ISSN 2347-6435 (Online) Volume 14, Issue 06, June 2025 Crossref DOI 10.54380/IJRDET0625_13)

- Knowledge-based approach: this model debates which things meet the demands of the user and offers recommendations.
- Combined model: this model combines the previously stated models.

II. Related work

Many studies have been conducted in the subject of recommender systems, with the majority of these studies focused on enhancing the prediction accuracy ofemploying methods such as content-based, collaborative filtering, and others to create a recommender system.

Numerous websites offer personalized recommender systems to their users, and e-commerce and academia in particular have focused a lot of attention on these systems and developed a number of effective algorithms to improve their performance. Some of the well-known recommender systems across several fields are shown in Table 1 below.

TABLE I.
RECOMMENDERSYSTEMS

Domain	Recommender Systems
Movie	Netflix, Movie Lens
Music	Ringo, CoCoA
Message	GroupLens, p-Tango
Electronic Mall	eBay, Amazon, Alibaba
Web page	R2P, Siteseer, QuIC

The section below enlists the works done by several researchers based on different techniques:

To address the scalability issue, Hu Jimning et al. [13] suggested an enhanced collaborative filtering technique. Regarding the tightly connected CF approach structure.

To address the CCS and ICS issues, Jian Wei et al. [14] presented two deep learning neural network-based recommendation algorithms.

As Choudhury et al. [15] noted, researchers have made numerous attempts to address these problems, such as using RS to buy a book or watch a movie. Nevertheless, the majority of these studies have not been able to deal with the problems of cold start concerns, malicious attacks, and data sparsity. There are 4 major recommendation models used to suggest best movie those are the Deep Neural Network (DNN) model, Singular Value Decomposition (SVD) model, the DNN with Trust modeland the Back propagation (BPNN) model. With an accuracy of 83% and a 0.74 MSE value, the results indicated that the DNN with trust model was the best model; nevertheless, it shouldn't be used to suggest the best films.

Recommendation system approaches and techniques were covered by Nitasha Soni et al. [16]. This article employs a content-based filtering strategy, however alternative approaches may also be used depending on user experience or business requirements.

Gupta Meenu and others [17] K-NN algorithms and collaborative filtering were utilized to increase the accuracy of the findings in contrast to the content-based approach.

Hong-Quan Do et al. [18] revealed the concept of weighted hybridization which is based on hybridizing CF & CBF combination for the purpose of providing recommendations.

Hamid Jazayeriy et al. [19] suggested an innovative approach for improving recommendations.

III. Proposed Methodology

Our work was done using google colab.The diagram below depicts the concept of hybridization for improvising item recommendations.

The following steps are taken into consideration in our proposed approach:

1. Information about users movie preferences are used for the content-based filtering (CBF) models to provide recommendations. Here we also calculate user's rating based on average rating.
2. Next we apply user based collaborative filtering model using K Nearest Neighbor algorithm.
3. Finally we make an improvised hybrid recommender system by retraining our model after feature engineering by adding release year.

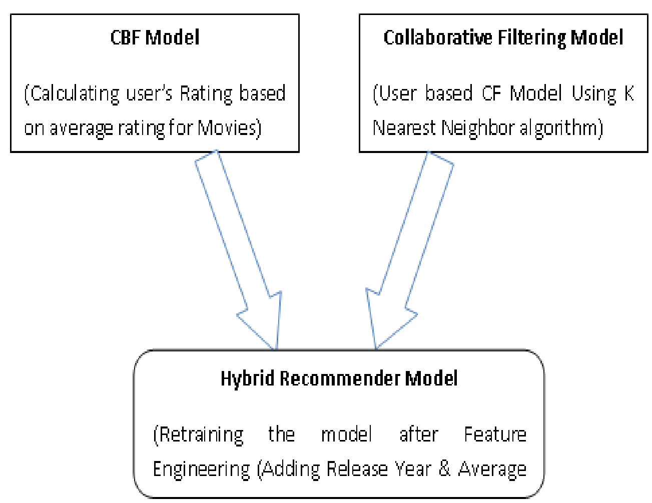

Fig. 2. : Flow of proposed recommender system

IJRDET
International Journal of Recent Development in Engineering and Technology
www.ijrdet.com (ISSN 2347-6435 (Online) Volume 14, Issue 06, June 2025 Crossref DOI 10.54380/IJRDET0625_13)

A. Dataset

In our proposed method we have used Movie Lens 1M Dataset, which contains 1 million movie ratings given by 6000 users on 4000 movies.

The figure below depicts distribution of users based on gender for the given dataset where we find that maximum number of males fall under the age group of 25-34 whereas maximum number of females fall under the age group of 18-24 years.

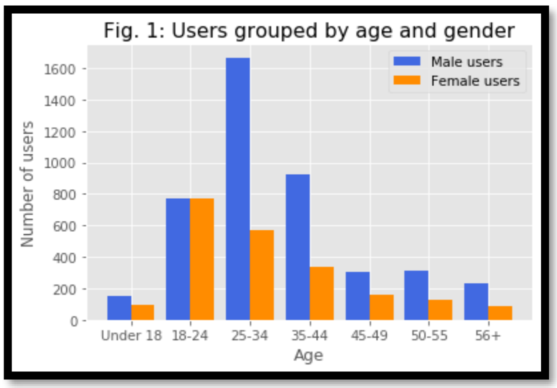

Fig 3: Distribution of User based on Gender

Table 2 :
Fields of Bruker Dataset

	Brukkr ID	Kjonn	Alder	Jobb	Postkode
0	0	F	45	6	92103
1	1	M	50	16	55405-2546
2	2	M	18	20	44089
3	3	M	35	1	33304
4	4	M	35	6	48105

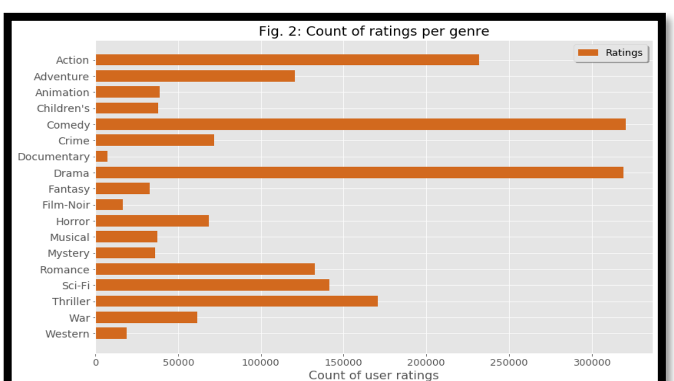

Fig.4: Count of ratings per genre

We can see from Figure 2 above, that there are significantly more action, comedy and drama movies in the dataset than anything else. We therefore expect the average rating for these genres to converge around the average rating for all movies.

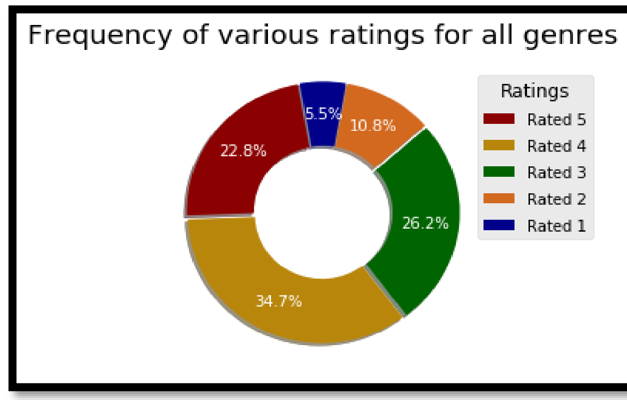

Fig.5: Count of ratings per genre

The Figure 4 above clearly shows that the MovieLens userbase rate movies highly on average. A score of 4 is very common, while 1's and 2's combined only take up 15% of the total ratings.

In order to build our proposed model, we must first divide our dataset into training and testing data. Additionally, we must first build a very basic baseline model in order to offer context for the evaluation of more complex recommendation models. Using a popularity-based method, we get the average rating for every movie and then assume that all users would evaluate all movies based on the average rating. The majority of users really rank movies in the neighborhood of the average movie rating, despite the fact that this methodology does not offer any customization.

As a result, we anticipate that it will be difficult to beat the RMSE for this model. Model baseline for RMSE: 0.981626106076701.

I use the movie tastes of the consumers to anticipate suggestions for the content-based filtering (CBF) algorithms. Specifically, the CBF models create a unique model for each user based on data on what each user has rated, what they have rated, and the genres of the reviewed movies. After then, the algorithm projects ratings for every other film in the dataset that the viewer hasn't viewed.

International Journal of Recent Development in Engineering and Technology
www.ijrdet.com (ISSN 2347-6435 (Online) Volume 14, Issue 06, June 2025 Crossref DOI 10.54380/IJRDET0625_13)

Table 3:
Comparison of RMSE of Models

	Model	RMSE
0	Linear Regression	1.071366
1	Lasso	1.037082
2	KKN_7	1.061543
3	Random Forest Regression	1.080374
4	Support Vector Regression	1.054499

With respect to KNN linear models performed much better, although baseline model performed the best of all. Lasso had a significantly lower RMSE than the rest, so I will therefore use Lasso for the hybrid recommender in section 2.5. It will however be interesting to see how feature engineering will affect both the classifier and the best linear models. I wonder whether the difference in RMSE between the different models will change or remain the same.

Collaborative Filtering

Using the K-Nearest Neighbors technique, we are now building collaborative filtering (CF) models that are expected to perform better than content-based filtering models. Our collaborative filtering technique will be user-based, meaning that user data (gender, age, location, etc.) won't affect the recommendations in any way.

On the basis of user preferences of most similar user according to the taste of users predictions are calculated by user-based CF models.

Table 4:
Models with a different number of neighbors

	Model	RMSE
0	kNN-5	1.013958
1	kNN-10	0.978301
2	kNN-20	0.962699
3	kNN-30	0.959578
4	kNN-40	0.958769
5	kNN-60	0.959030

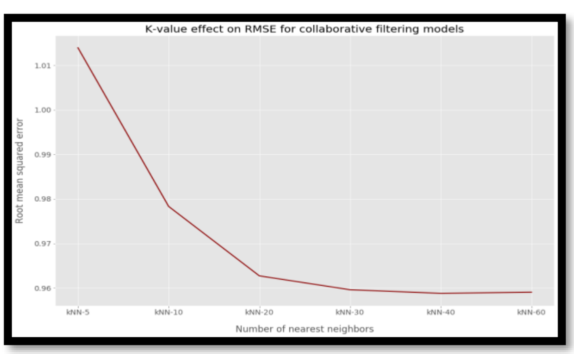

Fig.6: Graph of kNN Model wrt RMSE

The results obtained by collaborative filtering models were pretty better with respect to both the content-based and baseline models. RMSE of 0.9588 was achieved with a K-value of 40.With the value of K=60 that is higher K-values did not any further affect RMSE value.

A much more improvised recommender system is created by combiningCBF and the CF models.There are both pros and cons of CF and CBF therefore hybrid recommender model is proposed in our work.

Now as a part of feature engineering we will add release year to the movies dataset also we add average age of movie fan to the movie dataset.

Table 5:
RMSE values for Models

	Model	RMSE
0	Lasso	1.020758
1	KNN_7	1.040054
2	SVR	1.070242

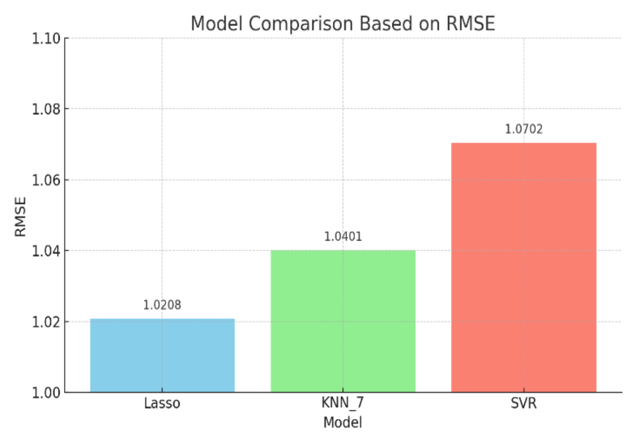

Fig.7: Bar Graph for Models comparison based on RMSE

Previously, Lasso performed with an RMSE of 1.037. It still outperforms both KNN and SVR after training with new features, with an RMSE reduced by 0.016242. The performance increase is not exactly astounding, but still significant enough to confirm that the feature engineering was successful.

Finally we test our improved hybrid recommender system and find the RMSE for KNN_40, Lasso & Hybrid model to get the results as :

Table 6:
Models RMSE values for our Proposed Models

	Model	RMSE
0	KNN_7	0.958015760014762
1	Lasso	1.019278388038111
2	Hybrid	0.9254351615504374

International Journal of Recent Development in Engineering and Technology
www.ijrdet.com (ISSN 2347-6435 (Online) Volume 14, Issue 06, June 2025 Crossref DOI 10.54380/IJRDET0625_13)

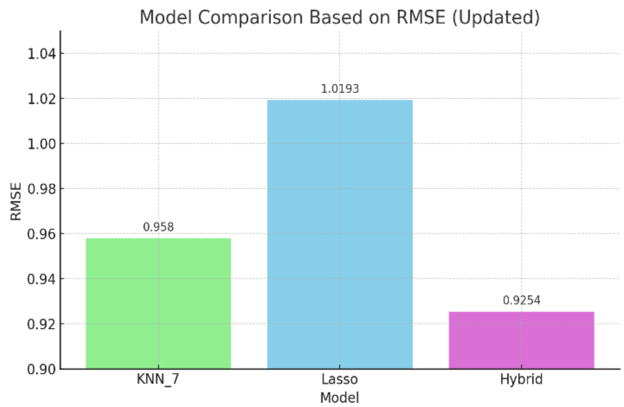

Fig. 8: Bar Graph for Models comparison based on RMSE

On test data an RMSE of 0.925 is obtained with hybrid recommender model. Earlier we saw that the RMSE performance of our baseline model came to be about 0.982.Therefore, we can see that there is an improvement in recommendations after implementing measures.

IV. CONCLUSION

We have proposed our work on the improvised hybrid recommender system. Our work uses machine learning technique for recommending items based on popularity and also make use of hybrid technique for recommendation to enhance the efficiency of recommendation. This work incorporates feature engineering to offer a hybrid recommender model. The suggested method combines a collaborative filtering model (which makes use of the K Nearest Neighbor algorithm) with a content-based filtering model (which bases user ratings on the average rating for movies).Our suggested method's hybrid recommender model generates predictions with an RMSE of 0.925 on test data. After taking adjustments for particular user preferences, such as combining individuals with similar tastes, the recommendations significantly outperform the baseline model, which had an RMSE of 0.982.

REFERENCES

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, Jul. 2013, doi: https://doi.org/10.1016/j.knosys.2013.03.012.

[2] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," Decision Support Systems, vol. 74, pp. 12–32, Jun. 2015, doi: https://doi.org/10.1016/j.dss.2015.03.008.

[3] D. Kotkov,, J. Veijalainen, and S. Wang, "Challenges of serendipity in recommender systems," presented at the International conference on web information systems and technologies(WEBIST 2016), vol. 2, pp. 251-256, Rome, Italy, April 23–25, 2016, doi: https://doi.org/10.5220/0005879802510256.

[4] B. B. Sinha and R. Dhanalakshmi, "Evolution of recommender system over the time," Soft Computing, vol. 23, no. 23, pp. 12169–12188,Jun. 2019, doi: https://doi.org/10.1007/s00500-019-04143-8.

[5] H. Abdollahpouri, R. Burke, and B. Mobasher, "Controlling Popularity Bias in Learning-to-Rank Recommendation," Proceedings of the Eleventh ACM Conference on Recommender Systems, Aug. 2017, doi: https://doi.org/10.1145/3109859.3109912.

[6] H. Abdollahpouri, R. Burke, and B. Mobasher, "Managing popularity bias in recommender systems with personalized re-ranking."presented at The Thirty-Second International Florida Artificial Intelligence Research Society Conference (FLAIRS-32),May 18–22, 2019, Sarasota, Florida, USA, doi: https://doi.org/10.48550/arXiv.1901.07555.

[7] H. Abdollahpouri, "Popularity Bias in Ranking and Recommendation," Proceedings of the2019 AAAI/ACM Conference on AI, Ethics, and Society, Jan. 2019, doi: https://doi.org/10.1145/3306618.3314309.

[8] C. Sun and Y. Xu, "Topic Model-Based Recommender System for Longtailed Products Against Popularity Bias," Jun. 2019, doi: https://doi.org/10.1109/dsc.2019.00045.

[9] M. Levy, and K. Bosteels, "Music recommendation and the long tail," presented at the workshop on music recommendation and discovery (WOMRAD),pp. 55-58, Barcelona, SPAIN, 2010.

[10] Sang-Min Choi, Sang-Ki Ko, and Yo-Sub Han.: A movie recommendation algorithm based on genre correlations, Expert Systems with Applications, 39(9) : 8079–8085, 2012.

[11] Antonio Hernando, Jes´us Bobadilla, and Fernando Ortega.: A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model, Knowledge-Based Systems, 97:188–202, 2016.

[12] Shouxian Wei, Xiaolin Zheng, Deren Chen, and Chaocha Chen.: A hybrid approach for movie recommendation via tags and ratings, Electronic Commerce Research and Applications, 18:83–94, 2016.

[13] Hu Jimning, ―Application and Research of Collaborative Filtering in E-commerce Recommendation System‖, 2010 IEEE.

[14] Jian Wei1, Jianhua He1, Kai Chen2, Yi Zhou2, Zuoyin Tang, ―Collaborative Filtering and Deep Learning Based Recommendation System For Cold Start Items‖, Expert Systems With Applications, 2016.

[15] 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), 475-4

[16] Nitasha Soni; Krishan Kumar; Ashish Sharma; Satyam Kukreja; Aman Yadav, ―Machine Learning Based Movie Recommendation System‖, 2021 IEEE 8th Uttar PradeshSection

[17] Meenu Gupta, Aditya Thakkar, Aashish, Vishal Gupta, Dhruv Pratap Singh Rathore, ―Movie Recommender System Using Collaborative Filtering‖, IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4, (ICESC 2020).

[18] Do, H.-Q., Le, T.-H., Yoon, B.,Dynamic weighted hybrid recommender systems. In: 2020 22nd International Conference on Advanced Communication Technology (ICACT). pp. 644–650.

[19] Hamid Jazayeriy, Saghi Mohammadi and Shahaboddin Shamshirband, ―A Fast Recommender System for Cold User Using Categorized Items‖, Math. Comput. Appl. 2018, 23, 1; doi:10.3390/mca23010001.