International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 10, Issue 4, April 2021)

An Approach for Movie Recommendation Using Rapid Miner

Vishal Paranjape[1], Dr. Neelu Nihalani[2], Dr. Nishchol Mishra[3]
Research Scholar[1], Prof. Dept.of MCA[2], Assoc. Prof. SOIT[3], RGPV, Bhopal, India

***Abstract* --Recommendation system helps user in making decisions efficiently among the different available alternatives. The central idea of recommender system is to deal with similarity among the items. The present paper is based on the approach for recommendation of a movie based on the concept of collaborative recommendation where past ratings given by users on a particular product is used for making decision and proposing an item to a new user. The present paper implements the movie recommender system based on collaborative filtering using rapid miner. A comparison between prediction model for k=1 and k=100 for User k-KK is done and results are discussed moreover a comparison result for correlation mode with respect to parameter is done and results are discussed. The dataset for our proposed works is taken from MovieLens where data pertaining to several thousands of movies with their ratings are kept over various periods of time. At the end results using user k-NN for user prediction and performance evaluation is shown.**

***Keywords*-- Collaborative filtering, Recommender system, User k-NN (Item Rating Prediction), MovieLens.**

I. INTRODUCTION

In today's world with the boom of internet and growth of e-commerce and online platforms there is tremendous increase in the suggestions and options to the user for a particular search. This matter would be a serious concern if the suggestions are inappropriate so there needs to be a reliable system for recommending things which are of particular interest to a given user and thus recommender system comes into play with its reliability to suggest items of particular interest to the user[1]. These systems retrieve the most vital information's out of a vast pool of information. The main purpose of these systems is to sense the interest of the user and suggest the most appropriate item [2].

There are three basic types of recommender systems, they are content based, collaborating filtering and hybrid systems. The major concept for recommendation in a content based system is the similarity between content of items and user profile. Collaborative filtering based approach on the other hand is based on the ratings given to a particular item by different users which become a basis for recommendation. Whereas hybrid approach is a combination of both content based and collaborative approach [3,4].

Today there are numerous fields where recommender system is used like LinkedIn advises us to connect with people we may know, similarly Facebook uses friend suggestions using 'You may know', YouTube recommends us the trending videos of our interest, online shopping site like Amazon user recommender system to recommend a particular item and NetFlix is a widely used online movie recommender system.

II. REVIEW OF LITERATURE

A thorough study of the various research paper on collaborative filtering for movie recommendation was done and their findings are discussed below:

Leidy Esperanza Molina Fernández [5] worked on a movie recommendation mechanism for Netflix. The dataset that was used here consists of over 17K movies and 500K+ customers. The prominent recommender algorithm used in the work is Popularity, Collaborative Filtering, Content-based Filtering and Hybrid Approaches. The algorithm that best fit to the data are selected and implemented.

Steven Postmus [6] suggested approach for movies recommendation which has practical implications in Netflix ratings. There are ratings interms of Likert scale from 1-5. Further as a technique for a recommender system collaborative filtering (item-based, user-based and singular value de-composition) is discussed. It was seen that singular value decomposition model was found to be most suitable for this dataset.

Mehdi Elahi et.al. [7] told that rating of items forms the basis for knowing user preference in collaborative filtering recommender systems. Ratings are the basis for giving extra knowledge to the system for giving more accurate results.

Hu Jimning et.al. [8] discussed about the e-commerce recommendation system and existing collaborative filtering technique. An improved collaborative filtering algorithm is designed and implemented. The recommendation quality can be greatly improved as the proposed algorithm can solve the scalability problem.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 10, Issue 4, April 2021)

III. Using Extensions of RapidMiner for Collaborative Item Based Recommendations

Three types of operators as Item Recommendation, Item Rating Prediction & Recommender Performance are the basic recommender extensions. We use operators 'Item Rating Prediction' in Collaborative Filtering, We also applied user k-NN and apply model and performance operator. The Performance operator compute the assessment of rating prediction error methods: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE) [9]. We have used the MovieLens dataset where ratings for movies prediction is on the scale of 1-5 There are three files: 'ratings.csv', 'users.csv' and 'movie.csv'. For prediction of user ratings using collaborative filtering we requisite data with three fields: userid, movieid and ratings.

	A	B	C	D	E
1	userId	movieId	rating	timestamp	
2	1	1	4	964982703	
3	1	3	4	964981247	
4	1	6	4	964982224	
5	1	47	5	964983815	
6	1	50	5	964982931	
7	1	70	3	964982400	
8	1	101	5	964980868	
9	1	110	4	964982176	
10	1	151	5	964984041	
11	1	157	5	964984100	
12	1	163	5	964983650	
13	1	216	5	964981208	
14	1	223	3	964980985	
15	1	231	5	964981179	
16	1	235	4	964980908	
17	1	260	5	964981680	
18	1	296	3	964982967	
19	1	316	3	964982310	
20	1	333	5	964981179	

Fig.1: Rating Dataset from Movie Lens

IV. Methodology

In our present work we have focused on the analysis of collaborative item based recommender system using RapidMiner. Our present work is done on RapidMiner 8.2 version. In Rapid Miner, the recommendation extension has 26 recommendation operators.

These operators are categorized into the following groups: Item Recommendation, Item Rating Prediction and Recommendation performance.

In our present work we have used Item k-NN with Item Recommendation and Item Rating Prediction.

Firstly we have taken operator Read CSV in which we imported data from movie.csv and rating.csv file. Now we used the join operator to get a consolidated data of both the files into single, further we used select attribute and selected the attributes rating, title and userid. Now we specified the roles of each attribute using Set Role operator. On applying Split Data further using shuffled sampling and partition in the ratio of 0.9 and 0.1. Now we applied User k-NN with k=80 and setting the parameter with Min Rating 1 and correlation mode Cosine. Finally we applied model for Rating Prediction. The workflow of the above steps is depicted below:

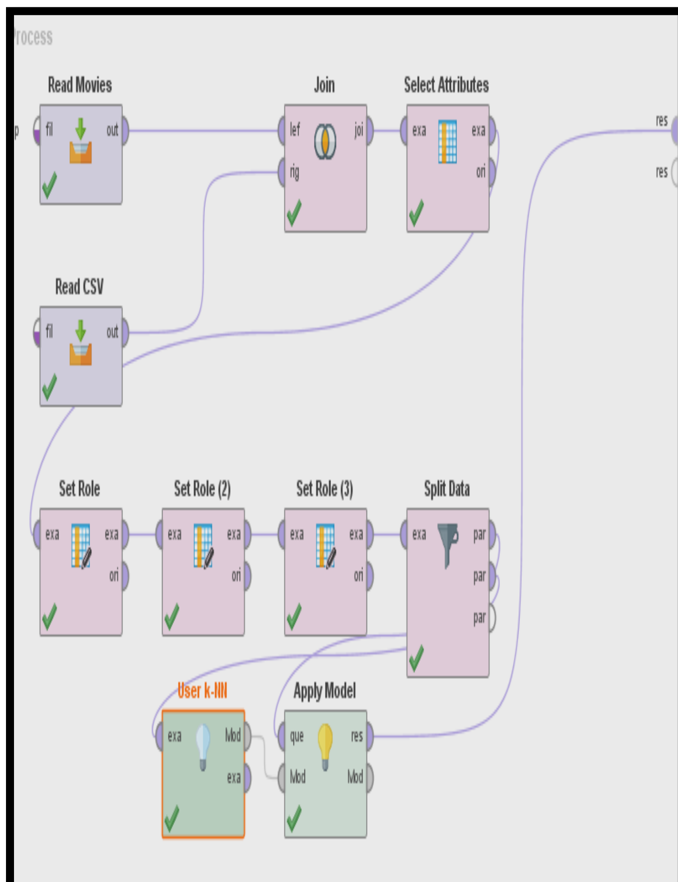

Fig.2: User k-NN & Apply Model Snapshot

The result of the above applied model is depicted as below where the prediction is generated on the basis of rating for a given title.

Row No.	rating	userId	title	prediction
1	4	5	Toy Story (1995)	3.427
2	5	7	Toy Story (1995)	3.632
3	3	15	Toy Story (1995)	3.365
4	4	18	Toy Story (1995)	3.868
5	4	19	Toy Story (1995)	2.402
6	4	21	Toy Story (1995)	3.386
7	5	31	Toy Story (1995)	3.629
8	3	32	Toy Story (1995)	3.073
9	5	40	Toy Story (1995)	3.688
10	5	43	Toy Story (1995)	3.789
11	3	44	Toy Story (1995)	3.601
12	4	45	Toy Story (1995)	3.663
13	5	46	Toy Story (1995)	3.119
14	3	50	Toy Story (1995)	3.178
15	3	54	Toy Story (1995)	4.346

Fig.3: Rating & Prediction Snapshot

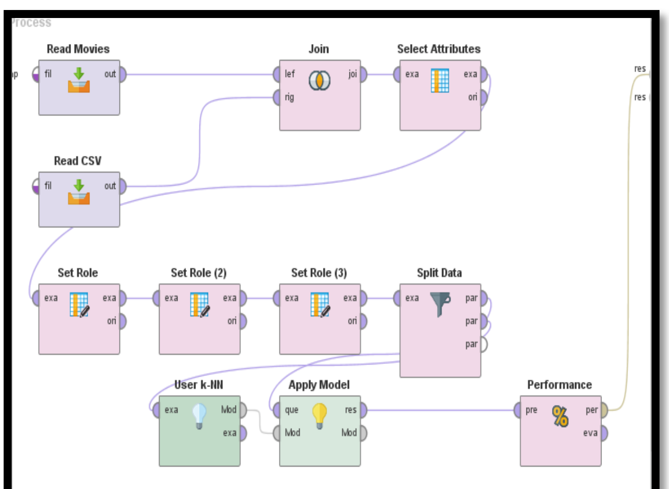

Fig.4: Performance Prediction Snapshot

Now we further apply the performance operator in the above model to depict the overall performance vector in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) & Normalized Mean Absolute Error (NMAE). A comparison between correlation mode and parameter on the three parameters are done and it is observed that Pearson mode gives better results with less error as compared with Cosine mode which is depicted below.

Table 1:
Comparison Table of Mode & Parameter

Performance Vector Comparison			
Correlation Mode Vs Parameter	RMSE	MAE	NMAE
Pearson	0.996	0.786	0.197
Cosine	1.037	0.815	0.204

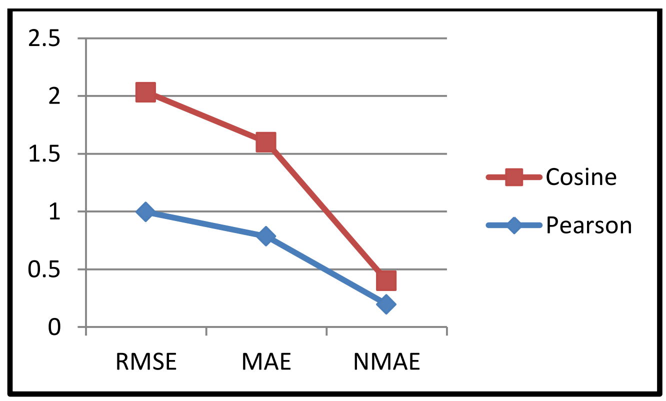

Fig.5: Graph showing Pearson to have more accuracy than correlation mode

Further we show the comparison of accuracy of rating for movie recommendation on varying value of parameter user k-NN where we take first 10 rating and its prediction for k=1 and k=100. Now we plot a graph to see the nature and find that as the value of 'k' in user k-NN increases bias also increases and the results are more inclined towards actual rating thus we see we are more close and predictions have greater accuracy.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 10, Issue 4, April 2021)

Table 2:
Comparison of Rating Prediction for k=1 & k=100

S.No.	Rating	k=1	k=100
1	4	4.688123781	4.466215071
2	5	4.688123781	4.5563862
3	5	5	4.971108185
4	5	3.858443366	4.244599287
5	5	3.30200458	3.548607904
6	3	3.858443366	4.040135138
7	5	5	4.912120635
8	4	3.858443366	4.117791054
9	5	4.440067136	4.871295417
10	5	5	4.879037171

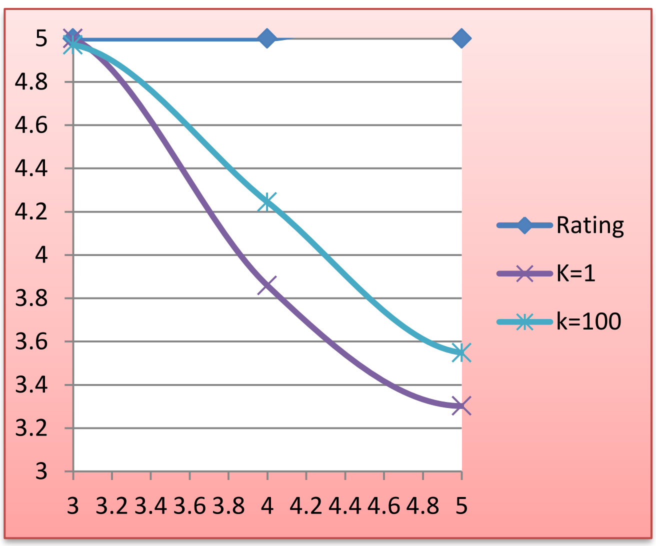

Fig.6: Graph showing Prediction for k=1& k=100

V. RESULT & DISCUSSION

The present paper works on the implementation of Collaborative Filtering based movie recommendation system using RapidMiner. Te paper discusses two results where one is the comparison between errors RMSE, MAE & NMAE for cosine and pearson parameters and it is found that pearson parameter gives better accuracy with less error with respect to cosine.

Another comparison on rating prediction is made where we have taken two values of ‘k’ in user k-NN , k=1 and k=100 and found that k=100 has more inclination towards actual rating which shows that in k-NN as the value of ‘k’ increases bias also increases. Our experiment is conducted on RapidMiner 8.2 and we have used the dataset from MovieLens.

VI. CONCLUSION

In this work we have worked on predicting the preference rating with respect to similar items with Item based collaborative filtering. There are two main problems with item based collaborative filtering cold start and data sparsity. Cold start problem are of two types cold item and cold user, in the first if a new item which has been added to the database and not yet been rated while in the other if the user is new the recommender system is unaware of his taste. The future scope of our work lies with recommending movies to users on the basis of their demographics and to deal with the problem of cold start and data sparsity.

REFERENCES

[1] D J.Bobadilla, F. Ortega, A.Hernando, A. Gutierrez, Recommender systems survey,Knowledge-Based Systems,46(2013) 109-132.
[2] P.Resnick,H.R. Varian, Recommender systems, Communications of the ACM, 40(1997) 56-58.
[3] G. Linden, B. Smith, and J. York, “Amazon.com recommendations:item-to-item collaborative filtering,” IEEE Internet Computing, 2003, vol. 7, no. 1, pp. 76–80.
[4] Mojdeh Talabeigi et.al, “A Hybrid Web Recommender System Based on Cellular Learning Automata”, International Conference on Granular Computing, IEEE, 2011.
[5] Leidy Esperanza Molina Fernández, “Recommendation System for Netflix”, Faculty of Science Business Analytics, January 2018.
[6] Steven Postmus,” Recommender system techniques applied to Netflix movie data”, Vrije Universiteit Amsterdam, February 2018.
[7] Mehdi Elahi , Francesco Ricci , Neil Rubens, “A survey of active learning in collaborative filtering recommender systems”, http://dx.doi.org/10.1016/j.cosrev.2016.05.002 1574-0137/© 2016 Elsevier Inc. All rights reserved.
[8] Hu Jimning,“Application and Research of Collaborative Filtering in E-commerce Recommendation System”, 978-1-4244-5540-9/10/$26.00 ©2010 IEEE.
[9] Qian Wang, Xianhu Yuan, Min Sun “Collaborative Filtering Recommendation Algorithm based on Hybrid User Model”, FSKD, 2010.