Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/gk6mg444

COLLABORATIVE FILTERING BASED TOP-N RANKED RECOMMENDER SYSTEM: AN IMPLEMENTATION AND COMPARISON METHODOLOGY

Sajid Ansari[1*], Md. Imran[2], Piyush Kumar Gupta[3]

[1*,2,3]Department of Computer Science and Engineering, Jamia Hamdard, New Delhi

Corresponding Author: Sajid Ansari
[*]Department of Computer Science and Engineering, Jamia Hamdard, New Delhi

Received-January 19, 2022 Acceptance- March 21, 2022

Abstract.
The trend of digital explorations especially on e-Commerce websites is being witnessed with the fast-growing use of the internet. It led commercials to understand their consumers' behaviour more accordingly. In this paper, we implemented an Artificial Intelligence based technique called Collaborative Filtering on the dataset we created to obtain top-n ranked recommendations for the books on Artificial Intelligence (AI). We also compared our work with other similar implementation of AI based recommender systems to show the differences in the methodologies and approaches that are published by other users.

Keywords: Top-n Ranked Recommendation, Collaborative Filtering, Recommender System Techniques, Comparative analysis.

1 INTRODUCTION

People's daily life has changed at a very fast rate due to changing in technology. As we are growing and moving towards novel and advanced technologies, these novel technologies fulfill our daily life's needs. These novel technologies are given birth to the internet. According to Wikipedia worldwide internet user are increased day by day, there are 16% worldwide internet user in the year 2005, 30% in 2010, 48% in 2017 and 53.6% in 2019 of their world population, in 2019 there are 47% of internet user in developing country and 86.6% of internet user in a developed country, the most internet user's countries like UK, Germany, USA, Japan has approximately 95% of internet use of their population according to 2020 data [1]. This rapid growth of internet users has changed people's living style. Forbes [2] says it pushed 70 % of internet users and streaming more than 12%. Due to the novel Coronavirus, almost all countries in the world have applied lockdown, in this situation dependencies on the internet have increased, all the Universities and school in the developed and developing countries have decided to provide online classes, except online classes they have no choices to provide classes to the students. In the lockdown situation, people preferred to buy almost all things through internets application like if they have to buy foods, vegetables, clothes, groceries items, books, etc. If they do not use these applications, then they have to walk to buy the products for living and if they walk to market then they have fear of Covid infection. people use internet's application like amazon prime, Netflix, etc. to watch movies, peoples play a game on the application of internet like PUBG, Minecraft, Fortnite, Dota2, Apex Legends, etc. Now we have a good number of applications for transfer money from one bank account to another bank account within

minutes, for example, Google pay, Phone Pay, Paytm, PayPal, etc. There are different applications for different purposes, people can play online games, people can watch movies, people can purchase online products, people can buy railway and Air tickets online. Almost every commercial internet-based application uses recommender systems to suggest the best things to their users for more sell. First of all, Amazon started using the recommender system and amazon saw an increment in business then almost every company like eBay, Walmart, Facebook Marketplace, Flipkart, etc., starts using recommender systems.

Recommender system (RS) is the software that suggests identical products to the user based on their earlier purchased history. RS examines huge data of user's items and compiles a list of those items which would fulfill the user's requirement. Most of the commercial companies are using recommender systems to draw attention to the user to purchase more products by offering likely products. There can be a Book recommender system, movie recommender system, songs recommender, Car recommender system, mobile phone recommender system, etc. The book recommender system is being used by several companies like Amazon, Flipkart, Myntra, Paytm Mall, etc. to recommend the most suitable books, the user would be examined to buy as they are matched with their choice. To make online shopping easy and attractive many researchers have implemented a good number of recommender techniques like collaborative filtering, opinion mining by Sohail et al. [3], content-based, Hybrid recommender system, etc.

Collaborative Filtering (CF): CF generates lists of products similar to the user's preferences. Collaborative filtering is based on the assumption that if user A has rated two books and another user B has rated two or more books, one book is similar to user A then CF(Collaboration) recommend a book to user A that book would out of them which book is previously rated by user B. recommender systems contains mainly three things: user data, products data, and recommender algorithms. In the products model, product features are analyzed similarly in the user model, user features are analyzed after that the characteristics of the users are matched with the features of products to judge which products to recommend using the recommender algorithm.

Sources of user and item feature: Implicit Feedback and Explicit Feedback

Implicit Feedback: In this method, user's choices are recorded based on their Internet search, purchased, and liked items. Explicit feedback: It is calculated through a review given by the users on items.

Collaborative filtering is dividing into two categories: User-based approach and item-based approach.

1. User-based approach: In this method, the user has a similar choice as the neighbor. If user A and B are neighbors, then if A likes some items for example x and y then items x and y will be recommended to user B.

2. Item-based approach: In this method groups of similar items are formed, for example, if there are four items in a group x, y, z, and p, if user A likes x then y, z, and p can be recommended to the user A.

Content-based recommender system: It tries to categories the user based on features or behavior of a user, which user has given the reaction to the different type of items. Once we know the choice of user, we can calculate similarities between the feature's vector of user based on previous record and feature's vector of products and recommend the products to the user based on their choices.

We also discuss a comparison methodology to make a comparative view of methods and approaches that have been used by authors and researchers in the context of recommender systems. For example, we introduce some factors:

- Dataset Used
- Collaborative filtering or other techniques applied (on dataset)
- Comparison of work done

- Extensive discussion of literature
- Real implementation

Comparison has been performed through these factors in Table 2 with the research articles mentioned there. Evaluation for each article has been marked in the respective row accordingly.

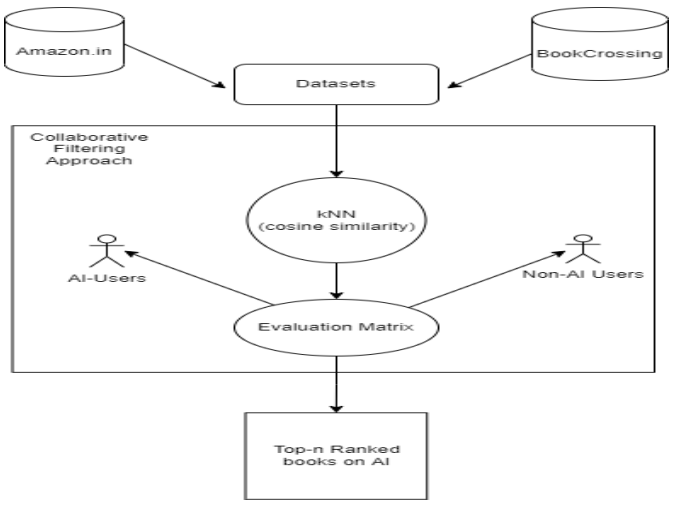
2 METHODOLOGY

The task is to find the top-N ranked list of books on Artificial intelligence which the user might like. We make user profiles by collecting review data of users. There are different types of books in our dataset, one is for artificial intelligence and another all we say is for Non-AI. We do many steps in this whole approach of the recommendation.

- First, with the help of K-nearest neighbors we do collaborative filtering. Collaborative filtering is defined as an algorithm of making predictions automatically about the interest of a user by collecting information of interest of different users. The assumption of collaborative filtering methodology is underlined is that if person A like to have mango and person B also like to have mango then if person B like to have an Apple then person A also like To have an Apple, it means that if person A has the same opinion as person B on an issue then there is a probability that if person B gives a view on another issue, person A will like this view also. From kNN algorithm Cosine similarity, we got five books for active books which are neighbors of active books.
- We choose artificial intelligence book in those five books which come from kNN process (step 1), we make the table of all AI book which comes from kNN process and of an active user for that above books come from kNN. For an active book, there is an active user in the dataset who has given a review for that active book, we search for the active user in the dataset. we make a table of books for different active users by repeating the above process.
- In the third step, we count how many users are there in the table for every book (i.e., totalCount) and we make a separate column for totalCount. We give top recommenders to those who have more totalCount. We arrange the book in decreasing order according to their totalCount, from this list we take the top N Book, hence we find the top-N book list of AI.
- Now, we make a comparative analysis based on the 5 factors we discussed in the introduction section. Table 2 is the result after analysis and provides as if the previous works had worked on Dataset, did it applied Collaborative filtering or any other technique on the dataset used. It also checks if the comparison of works had done or the in-depth discussion on literature background performed and lastly if it had a real implementation or not.

3 IMPLEMENTATION

In this section, we describe two datasets used in the experiment. The first is collected from Amazon.in and the other is from BookCrossing. The implementation details will also cover in this section.



3.1 Datasets

1. Amazon.in: We have collected 72 books, 230 reviewers and 230 ratings of these Artificial Intelligence books from these users [4]. This dataset is up to date and contains many recently published books on Artificial Intelligence. It also includes the book's authors, publishers and Amazon image-URLs and, User's demographic details and age. Each user or reviewer is distinguished by a unique user-id.

2. BookCrossing: Contains 271,379 books, 278,858 users with demographic information providing 1,149,780 ratings [5]. We filtered out all Artificial Intelligence books from it. All users who have rated any Artificial Intelligence books are only considered, although all other books (Non-AI) books rated by those users are considered.

Finally, our dataset consists of 462 books (95 books on Artificial Intelligence and the rest others Non-AI), 253 reviewers and 765 ratings provided to both Artificial Intelligence books and Non-AI books.

3.2 kNN-based Collaborative Filtering

This paper implements the kNN based Collaborative filtering algorithm upon the dataset of Artificial Intelligence books created. We recommend these artificial intelligence books to the set of users who already have read and reviewed the set of artificial intelligence books and the other Non-AI books. The recommendation for each user in the User dataset is noted and then the totalCount is performed on the evaluation table of the final recommender. This provides the final result of the top-N ranked recommender of Artificial Intelligence books.

3.3 Evaluation Metrics

The pivot table consists of indices as books by name, the rows by User-Ids, totalBookRating, totalRatingCount, etc. The final pivot matrix contains a set of books (both of artificial intelligence and Non-AI) and a set of users or reviewers. The values inside this matrix are the ratings provided by the users set and for null or unknown ratings is defaulted as 0. Our evaluation metric is the Cosine similarity to find the nearest neighbors of books which are picked from the above matrix to find k similar items (or books) to be recommended to the set of users who already read and reviewed these books.

4 RESULT

Table 1. Topmost recommended Books on Artificial Intelligence after implementation.

ISBN	Book	TotalCount
1976756340	Surfing the Tsunami	125
168330859X	TIME Artificial Intelligence	118
1541773756	The Big Nine	97
188886916X	The Artificial Kid (Context San Francisco)	65
9781633693296	Smart Business	39
692945407	Robot is the Boss	10
1472963881	Rage inside the Machine	9
312622376	Our Final Invention	1

We presented the Top N recommender approach and the above table shows the results of our approach, In the above table top Book for Artificial intelligence is **Surfing the Tsunami** with their

ISBN number, this book has is highest Total Count in the table so that **surfing the Tsunami** has top ranked in the list.

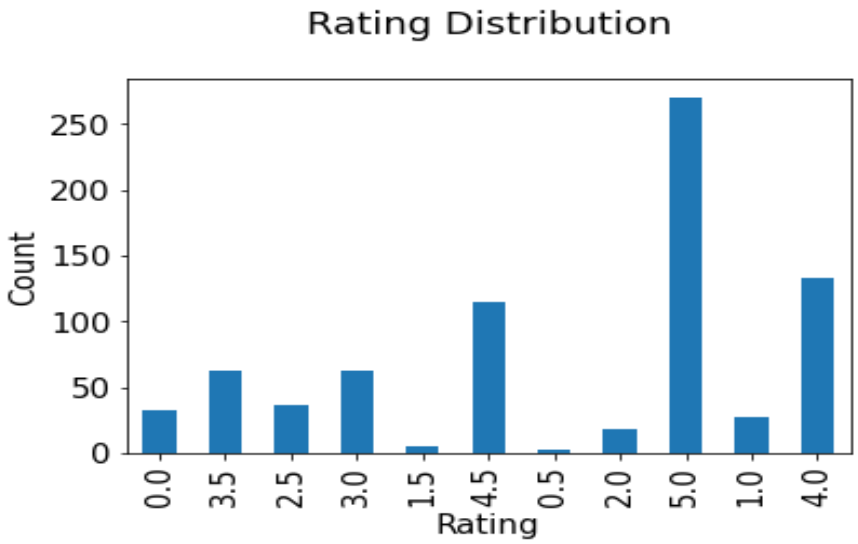


Fig. 1. Plot describes Rating Distribution of the dataset.

The ratings' axis varies from 1-5 with a 0.5 break value between the range. Similarly count axis has the range from 0 to 250 i.e., the maximum total count.

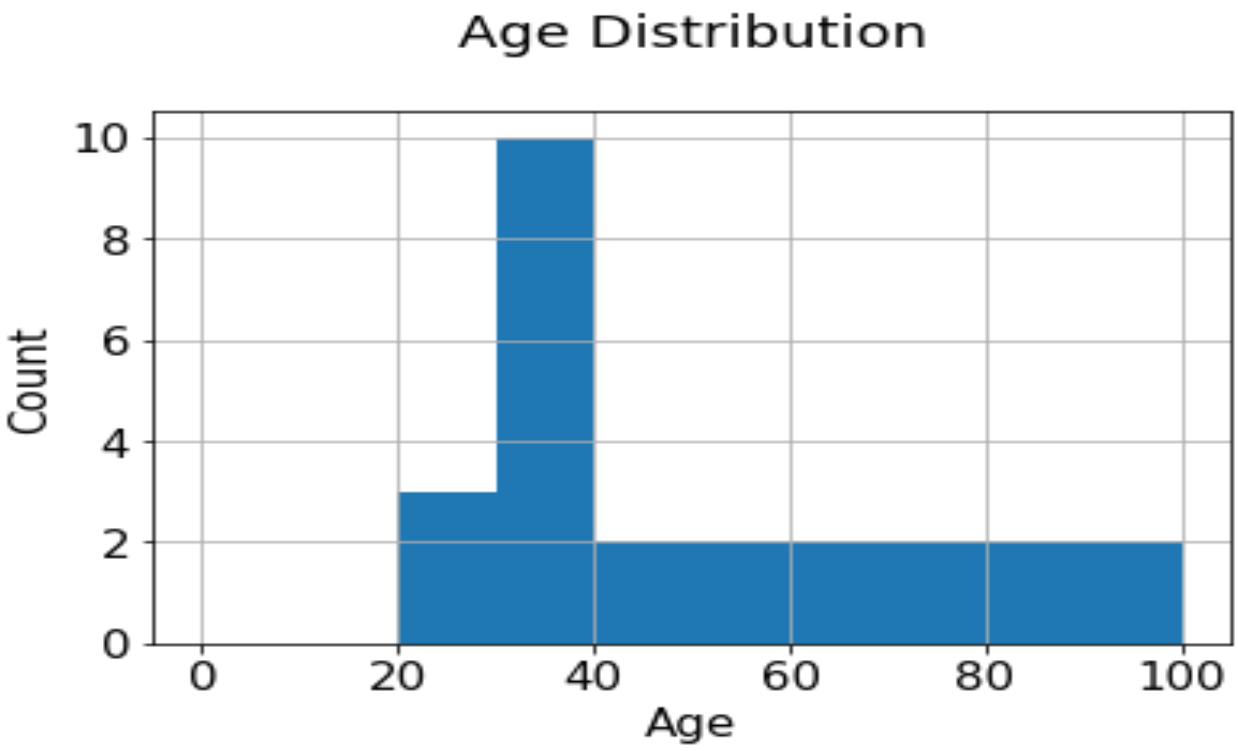


Fig. 2. Plot describes the Age Distribution of the users in the dataset.

This figure plots the age distribution of the users/readers present in our dataset. The range of age is 0-100 and that of the number of users/readers to this age category is on the count axis. It is clear that most of the users in our dataset are youths.

Table 2. Result of Comparative Analysis based on comparison factors.

Related Work	Dataset Used	Collaborative Filtering or other techniques applied	Comparison of work done	Extensive discussion on Literature	Real Implementation
Book recommender System Based on Collaborative Filtering and Association Rule Mining for College Students [6]	NO	NO	NO	NO	NO
Book recommender System Based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining [7]	NO	NO	NO	NO	NO
e-Learning Recommender System for Teachers using Opinion Mining [8]	NO	YES	NO	NO	NO
Personalised Book recommender System based on Opinion Mining Technique [9]	NO	YES	NO	YES	NO
Generating Top-N Items recommender Set Using Collaborative, Content Based Filtering and Rating Variance [10]	YES	YES	NO	YES	YES
Generating Collaborative Filtering based Top-n Ranked Recommender System: An Implementation and Comparison Methodology (Proposed by us)	YES	YES	YES	YES	YES

5 CONCLUSION

In this paper, we have implemented kNN based collaborative filtering on the dataset created and additionally on the dataset(s) mentioned in the implementation section. The result in Table 1 after implementation of the discussed methodology provides the topmost books in the field of Artificial Intelligence. These are the books recommended most of the time to the readers in the dataset. It is obtained by firstly collecting all the books in the dataset that have been recommended and then ranking all recommended books by counting the number of times the book has been recommended. This ultimately provides the final result of obtaining the Top-n ranked recommendation (Table 1). We also suggest the comparison method based on the factors discussed in the introduction section to make a comparative analysis of the previous similar works. The result of the analysis is collected in Table 2 with all the corresponding factors. It suggests the improvements and changes in approach required to be adopted by the authors of articles on recommender systems for the future development in the field of recommender systems. Our paper highly emphasis these comparison factors for the comparative analysis but we also invite authors to contribute to broadening this matrix of factors for comparison of works.

References

1. Available at. [https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users]. Accessed on (20/07/2021).

2. Available at. [https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal/?sh=3d6191573104]. Accessed on (11/06/2021).
3. Shahab Saquib Sohail; Jamshed Siddiqui; Rashid Ali, "Book recommendation system using opinion mining technique", 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), DOI: 10.1109/ICACCI.2013.6637421.
4. Available at. [https://www.amazon.in/s?k=ai+books&ref=nb_sb_noss_1]. Accessed on (30/06/2021).
5. Available at. [http://www2.informatik.uni-freiburg.de/~cziegler/BX/]. Accessed on (21/06/2021)
6. A.S Tewari, Kumari Priyanka, "Book recommender System Based on Collaborative Filtering and Association Rule Mining for College Students", 2014 International Conference on Contemporary Computing and Informatics (IC3I), DOI: 978-1-4799-6629-5/14.
7. A.S Tewari, Abhay Kumar, Asim Gopal Barman, "Book recommender System Based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining", 2014 IEEE International Advance Computing Conference (IACC), DOI: 978-1-4799-2572-8/14.
8. A.S Tewari, Anita Saroj, Asim Gopal Barman, "e-Learning Recommender System for Teachers using Opinion Mining", Springer-Verlag Berlin Heidelberg 2015 K.J. Kim (ed.), Information Science and Applications, Lecture Notes in Electrical Engineering 339, DOI 10.1007/978-3-662-46578-3_122.
9. Kumari Priyanka, A.S Tewari, Asim Gopal Barman, "Personalised Book recommender System based on Opinion Mining Technique", Proceedings of 2015 Global Conference on Communication Technologies(GCCT 2015), 978-1-4799-8553-1/15.
10. A.S Tewari, J.P Singh, A.G Barman, "Generating Top-N Items recommender Set Using Collaborative, Content Based Filtering and Rating Variance", Procedia Computer Science 132 (2018) 1678–1684.