Hybrid Book Recommendation System: Combining Popularity with Collaborative Filtering

Shashank Dwivedi, *Harshit Chaudhary, Ashwani Singh, Naveen Gupta, Ankit Mishra
Department of Computer Science and Engineering, UIT Naini, Prayagraj
**Corresponding Author: - Harshit Chaudhary*

Abstract- **Recommendation systems are software tools designed to deliver personalized suggestions based on user preferences, behaviour, and interactions. With the rapid growth of digital books, there is a pressing need for efficient systems to assist users in navigating large collections. This paper examines various book recommendation methodologies, highlighting their advantages and limitations, and explores ways to improve accuracy and user satisfaction. By synthesizing insights from recent research and applying advanced machine learning techniques, this study proposes a hybrid system that integrates popularity-based filtering, collaborative filtering, and cosine similarity measures. The developed model aims to provide reliable, scalable, and user-centred recommendations for enhanced user experience.**

Keywords- Recommender System, Popularity-Based Filtering, Collaborative Filtering, Cosine Similarity, Hybrid Model

I. INTRODUCTION

A book recommendation program suggests books to users on the basis of various algorithms and user preferences. Popular models such as collaborative filtration, content-based filtering, and hybrid treatment have been built to increase the accuracy of suggestions. The deep learning algorithm-based recommendation system is the latest to step up efficiency and personalization with the advances of AI and ML.

Recommender Systems keep users interested in the site. Earlier when there were no recommendation systems available, people used to take suggestions from other people. But in nowadays platforms like Google knows what you will likely to search for, YouTube knows what videos you are likely going to watch based on your search and watch history, likes and dislikes, Spotify recommends the songs based on user's listening , remixes and weekly song recommendations [1].

We are surrounded with Recommender Systems [RS]:

Examples of OTTs like Amazon, Netflix, Hotstar, Jio Cinema uses RS to keep their users engaged in their applications.

Examples of Music Apps include Spotify, Hungama, Jio Saavan, Apple Music, etc. Examples of Social Media Apps include Instagram, Facebook, YouTube, Twitter, etc.

All the apps which are commonly used in our daily life have recommendation systems which try their level best to keep customers engaged in their service as much as possible to generate revenue.

II. METHODOLOGY

To build an effective book recommendation system, we combine two well established techniques: Popularity-based Filtering and Collaborative-based Filtering. This hybrid approach ensures that users receive personalised recommendations while keeping overall book popularity.

2.1 Data Collection and Preprocessing:

1) Dataset: Data included Book-Crossing with ratings of users, book characteristics, and demographics. Users.csv has demographic information including users' age, their ids, and location. Books.csv has a combination of metadata on the books concerning ISBNs, title, author, publisher, and year of publication. Ratings.csv: The data in this set contains explicit user ratings with respect to various books, which are recommended for algorithms.
2) Data Cleaning: This involves the removal of duplicates and the entry of null with zeros for accuracy and better readability.
3) Feature Extraction: Find the key patterns from this data set and create the user- book matrix for how users rated the books.

2.2 Model Development:

1) Popularity Based Recommendation System: A popularity-induced recommendation is a type of

recommendation system that recommends items based on the items ' global popularity [3][5]. Examples of Popularity Based Recommendation Systems are Youtube Trending section, Twitter trending topics, IMDB Top 100 movies, and The Top 100 songs of the year.

2) Collaborative Filtering (CF): The term refers to a recommendation technique -process in which predictions about users' interest are made autonomously (filtering) by collaborating the preference /taste information of a large number of users. The fundamental philosophy of CF can be explained as-"If user A agreed with user B on some matter, then user A will also agree with user B on some other matter.a greater probability of sharing B's opinion on another matter than that of a randomly selected person [2].
3) Hybrid Model: By blending popularity-based recommendations with collaborative filtering, we create a system that suggests both popular and personalised books.

2.3 Training Process:

Cosine Similarity: Equation and Explanation

It is a metrics used to find the similarity between the two items without depending upon the size of dataset [3].

For each book different user ratings are written as vector and similarity between each books user vector is calculated using cosine similarity which is the cos value between each vector.

Mathematical Formula:

Where:

$$\cos(\theta) = \frac{\mathbf{A}.\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

- A·B is the dot product of vector A and vector B.
- |A| is the size (norm) of vector A.
- |B| is the size (norm) of vector B.
- θ is the angle between the two vectors.

Example Table Representation

Below is a sample table where each book's ratings by different users are represented as vectors:

Books	User1	User2	User3	User4	User5
Book 1	4	0	3	5	2
Book 2	5	2	0	3	0
Book 3	3	4	2	1	4
Book 4	0	3	4	2	1
Book 5	2	5	0	4	3

(Table 2.3.1)

Each row represents a book as a vector and cosine similarity is applied to compare book based on how user have rated them.

The technique enables the system to generate recommendations based on books that have similar rating patterns making this book recommendation most meaningful to the user.

III. LITERATURE REVIEW

In this respect , book recommendation systems help people find books they will probably enjoy in the future on the basis of their past choice or preferences . Such systems may be built in numerous ways and researchers work towards improving them to be able to give recommendations with a greater behaviour on the real-world-accurate scale. The different key approaches, their advantages and disadvantages , and the current boundaries under developments in this area will be discussed.

I) Common Approaches to Book Recommendation

- Collaborative Filtering (CF) – Learning from Others

Collaborative filtering and Word of Mouth are nearly identical functionalities.

The recommendations are based on the perception of other users having like-minded perspectives[3][7].

User-based CF: It recommends items for a user on the basis of similarity of preferences with other users [6][8].

Item-based CF: It recommends items for the user with respect to the similarity among the items. This is particularly useful when there are significantly more users than items available [6][9].

General recommendations can be made on characteristics such as age, location, and occupation [1].

An AI-based Knowledge Graph is how very subtle links are woven between users and books. User-based CF: Suggests items to a user based on similarities of preferences among that user and other users. Item-based CF: Suggests items to a user based on similarities among the items. This technique is especially good when the number of users is far larger than the number of items.

Problems with CF:

Cold Start Problem: It doesn't work well for new users or new books since there's no past data [1].

Sparsity: Many users don't rate books, making it hard to find similar users.

- Content-Based Filtering (CBF) – Matching Books to Interests

This method recommends books based on their content. If you like a book, the system finds others with similar genres, authors, or keywords [10].

Example: If you enjoy mystery novels, it will suggest more mystery books.

Techniques Used: It compares books using word matching and similarity scores.

Limitations of CBF:

Repetitive Suggestions: It keeps recommending books that are too similar to what you already read, making it hard to discover new genres [3].

- Improving Recommendations with Hybrid Models

Since both CF and CBF have weaknesses, researchers combine them into hybrid systems to make recommendations smarter.

Example: Amazon and Goodreads use a mix of CF and CBF to suggest books based on both user behaviour and book details.

- Challenges in Book Recommendation Systems

Cold Start Problem – Helping New Users and New Books

When a new user joins, the system doesn't know their preferences yet. Similarly, when a new book is added, there's no data to recommend it [2].

Proposed Solutions:

Firstly, it is possible to make general recommendations based on demographic data like age, location, and occupation [1].

With users and books, AI-based Knowledge Graphs figure out very subtle links between them

Data Sparsity – Lack of Ratings and Reviews

Since many users do not rate books, accurate recommendations become difficult to generate [1] [2].

Proposed Solutions:

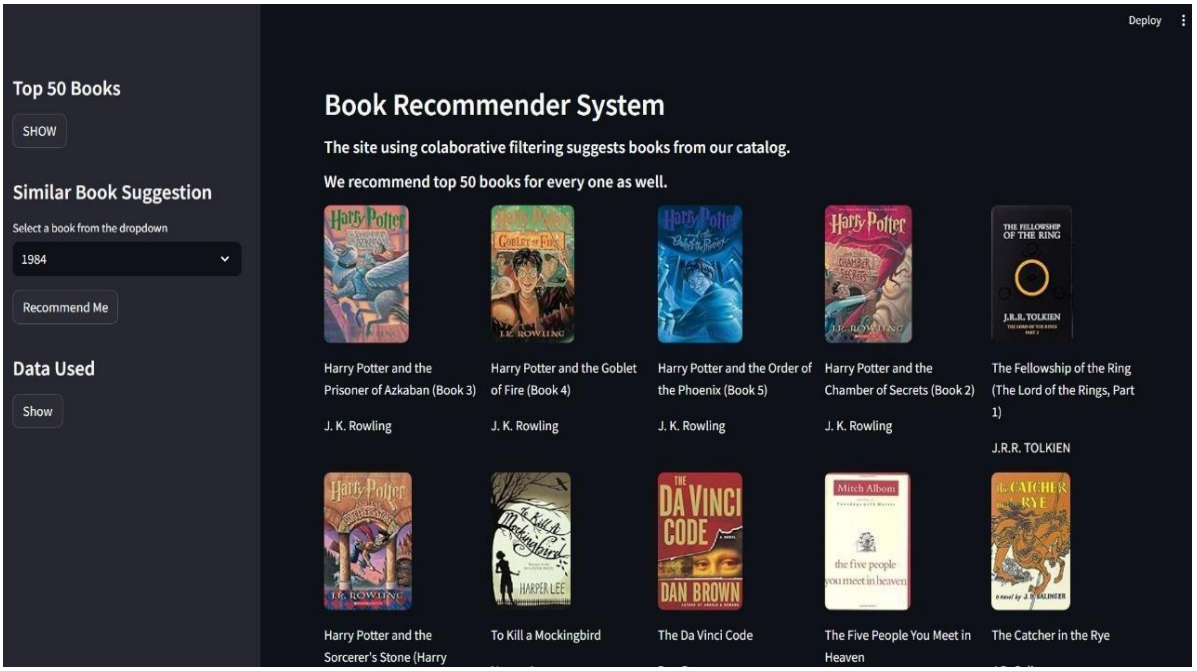
Use AI algorithms and matrix factorization methods to estimate the missing ratings.

IV. RESULT

It has a very huge range of user ratings related to books and was used for testing the developed system. The results revealed that better working strategy was employed by the hybrid model as compared to only collaborative filtering or popularity -based methods . Recommendations from a hybrid system were good and valuable because they were formulated on balanced use of both techniques . The recommendations were further made by creating clustering of books with similar rating distributions through cosine similarity just to make it more personalized according to each user's taste. This aspect of the hybrid model also resolves cold start problem when there is a new user or book which fails to gather sufficient data for optimal recommendations . This was important to ensure that users not only received good book suggestions but also new users.

The proposed hybrid book recommendation system is capable of improving the recommendation accuracy from popularity and collaborative filtering approaches. Resultant exhibit of hybrid approach composite conduct with both methods to be much better than employing single method. This recommendation system guides the users in word exploring interesting books through recommending books on general popularity, and...

Specific preferences of particular individuals were considered, so that different preferences clicked at two levels, which helped, in a way, to ensure that pertinent and rewarding ideas were spun out of the model.



(Fig. 4.1)

Here Fig 4.1 shows the result of top 50 books

(Fig. 4.2)

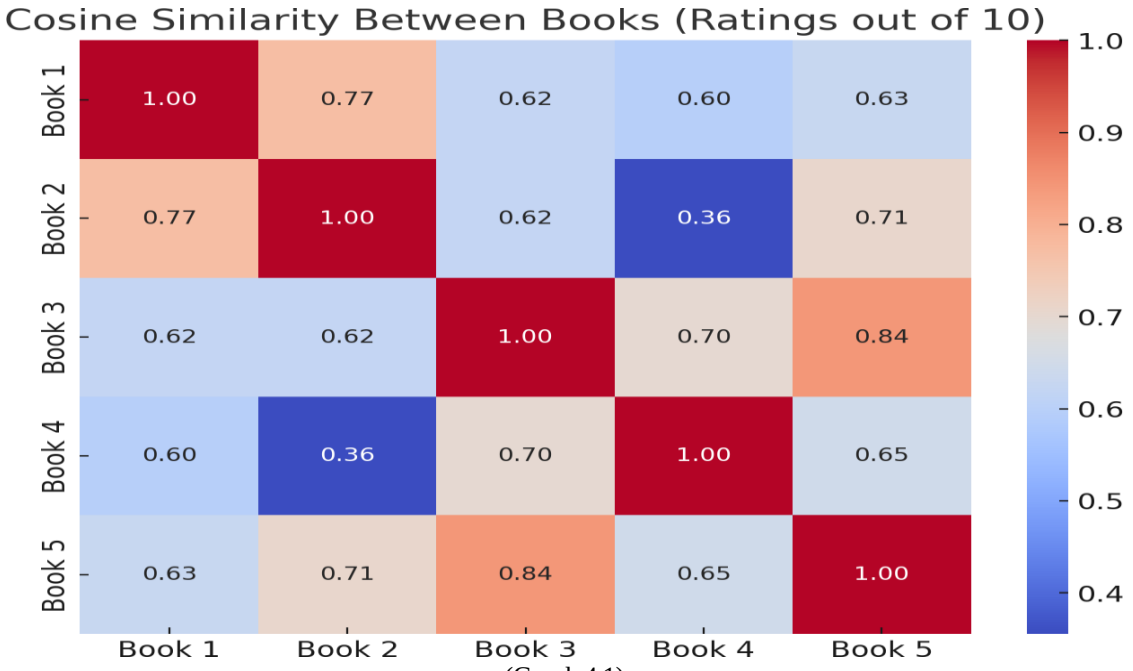
Here Fig4.2 shows the result of similar book of a particular category.

GRAPH:

Here is the cosine similarity heatmap for books, considering ratings out of 10. The color intensity represents the similarity between book rating vectors, where: Red (1.00) indicates perfect similarity (same ratings).

Blue (~0.3 - 0.5) indicates lower similarity. Neutral colors (~0.6 - 0.8) indicate moderate similarity.

This graph 4.1 helps visualize how similar different books are based on user ratings.



(Graph 4.1)

V. FUTURE SCOPE

Future improvements could include:

Deep learning: Using artificial intelligence to improve recommendations.

Natural Language Processing (NLP): Understanding book descriptions and user reviews to make better suggestions.

User interaction data: Tracking browsing behaviour to improve recommendations.

Real-time recommendations: Updating book suggestions instantly based on user activity.

REFERENCE

[1] Pooja M. Kerudi, Pratheeksha M., Raghavendra C. K., and T Sai Srujan, "Book Recommendation System: A Systematic Review and Research Issues," Journal of Huazhong University of Science and Technology, June 2021.

[2] Himanshu Mishra and Ashish Asthana, "Book Recommendation System," International Research Journal of Engineering and Technology (IRJET), May 2022.

[3] Abhishek Samar Singh, Progyajyoti Mukherjee, Syed Aamir Bokhari, and Prof. Sonia Maria D'Souza, "Collaborative and Popularity-Based Book Recommender System," International Journal of Scientific Research in Science and Technology, Nov-Dec 2022.

[4] Dominik Kowald and Emanuel Lacic, "Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems," arXiv, March 2022.

[5] Yuanhao Liu et al., "Mitigating Popularity Bias in Collaborative Filtering via Fair Sampling," arXiv, February 2025.

[6] Kushal Sharma and Sarita Patil, 2023, Comparative study of Content based and Collaborative Filtering Recommendation System for Social Media, International Journal of Novel Research and Development, 8(5), 609-612.

[7] Zhang, A., Zheng, J., Wang, X., Yuan, Y., & Chua, T. S. (2023). Invariant Collaborative Filtering to Popularity Distribution Shift. Proceedings of the ACM Web Conference 2023 (WWW '23), April 30-May 4, 2023, Austin, TX, USA, 12 pages. ACM, New York, NY, USA.

[8] Zheng, X., Niu, Z., & Chen, W. (2013). Opinion-Based Collaborative Filtering to Solve Popularity Bias in Recommender Systems. In H. Decker et al. (Eds.), DEXA 2013, Part II (LNCS 8056, pp. 426–433). Springer-Verlag Berlin Heidelberg.

[9] Zhang, X., Su, K., Qi, F., Zhang, Y., & Zhang, K. (2022). Collaborative Filtering Alg