



Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526

Volume 14 Issue 01, 2025

Book Recommendation System using Machine Learning

Onkar A. More¹, Shubham S. Kore², Kabir G.Kharade³

^{1,2}Research Student, Department of Computer Science, Shivaji University, Kolhapur

³Assistant Professor, Department of Computer Science, Shivaji University, Kolhapur

onkarmore3091@gmail.com¹, koreshubham7007@gmail.com², kgk_csd@unishivaji.ac.in³

Peer Review Information	Abstract
<i>Submission: 17 Jan 2025</i> <i>Revision: 14 Feb 2025</i> <i>Acceptance: 15 March 2025</i> Keywords <i>Book Recommendation</i> <i>Collaborative Filtering</i> <i>Cosine Similarity</i> <i>Machine Learning</i>	The increasing volume of online texts made it imperative to develop intelligent systems that can help readers discover books of their choice. This article presents a Book Recommender System employing popularity-based and collaborative filtering algorithms to enhance personalized recommendations. The system employs datasets comprised of user ratings, book metadata, and user profiles to develop two distinct models. The popularity-based approach chooses top-rated books according to the number of ratings and mean rating, giving popular items to new users. The collaborative filtering approach employs user behavior to recommend books by discovering similar readers and their interests, giving a more personalized experience. Data preprocessing involved missing value handling, duplicate elimination, and removing users with large rating histories for improving recommendation accuracy. The models' performance was measured in terms of diversity and relevance of the suggested books. The popularity-based model effectively indicates best-selling books, and the collaborative filtering model indicated improved personalization using user interactions. This study emphasizes the significance of combining different recommendation methods to cater to diverse user needs and paves the way for further research in hybrid recommendation systems.

Introduction

In the current digital age, the wealth of literary work on the Web has revolutionized how consumers locate and select books. With so many publications dispersed over various platforms, locating books that suit individual tastes has become ever more problematic. Recommender systems have become a central solution, providing personally tailored recommendations that improve user experience and interactivity. A recommender system evaluates user behavior, interests, and

interaction history to provide personalized recommendations. They are extensively utilized in sectors like e-commerce, entertainment, and education, where personalization helps enhance user satisfaction. In the book world, book recommender systems fill the gap between readers and the massive collection of available books, making it easy to discover and getting users familiar with content that resonates with them. This research aims to design and test a Book Recommender System that combines popularity-based and collaborative filtering methods. The popularity-based approach

identifies best-selling books based on the number of ratings and average rating, which makes it suitable for suggesting popular books. On the other hand, collaborative filtering uses user interaction information to provide personalized recommendations by detecting similarities between users and their reading habits. The main goal of this study is to improve book recommendation accuracy and personalization with the incorporation of these two approaches. The methodology in the research is systematic and involves data preprocessing, model formulation, and system performance evaluation in determining the strength of the system.

Review Of Literature

In this paper, an online application is developed for the library book recommendations. They give ratings to all the books in the library. It uses an automated approach to help users choose, borrow, and receive books based on ratings. It uses a collaborative filtering algorithm that suggests books by analyzing user ratings and feedback [1]. The goal of the system is to recommend the top five most similar books to the user that suit their interests and preferences. The collaborative filtering here is used using the KNN algorithm. It also uses a Random Forest classifier. They describe the system architecture of the book recommendation system [2]. The paper presents a study focused on developing a book recommendation system, leveraging advanced machine learning methods to overcome the shortcomings of current recommendation strategies. It employs hybrid systems, which integrate both techniques, enhancing recommendation precision through the utilization of user behavior and book characteristics. The incorporation of supplementary data such as user demographics and reading history further enriches the system's capacity to provide personalized and pertinent recommendations [3]. The paper presents a study focused on developing a book recommendation system utilizing an Arabic dataset of book reviews. Collaborative filtering algorithms are employed for the system's implementation. The Surprise Library contains all the necessary algorithms once the dataset is imported [4]. The goal of the system is to recommend the top five most similar books to the user that suit their interests and preferences. The collaborative filtering here is used using the KNN algorithm. It also uses a Random Forest classifier. They describe the system architecture of the book recommendation system [5]. A survey says that there are various methods used for book recommendation, like content-based, cooperative, demographic, and hybrid. The content-based filtering methods are grounded on

an item description and the user's interest. Cooperative filtering ways produce user-specific recommendations of things supported by patterns of ratings. The demographic recommendation technique of the advice system uses user profiles like age, gender, demographic space, interest, etc. In a hybrid system, we merge two techniques: content-based and cooperative filtering [6]. The proposed ponder in this paper is utilized for a book suggestion framework. It says that each proposal framework comprises of two substances, one is the client and the other is the thing. In this, it employs affiliation run the show mining, which finds curiously affiliations and relationship connections among the huge dataset of things [7]. The goal of this paper is to improve retention, helping to form customer habits and trends and increased the cart value [8]. The main idea of this paper is that the book recommendation system initiative seeks to assist individuals in finding a book that interests them and encourages them to read more books. It uses a dataset with three tables: book ratings, users, and books. They use various types of machine learning algorithms like KNN, Decision Tree, Naïve Bayes, Logistics Regression, Random Forest, and XG Boost [9].

Proposed Methodology

1. Data Collection and Preprocessing:

The dataset consists of three CSV files: *Books.csv*, *Ratings.csv*, and *Users.csv*, which are loaded into a pandas DataFrame for analysis. Basic data exploration begins by checking for missing values using the `isnull().sum()` function to identify any columns with incomplete data. Additionally, the `duplicated().sum()` function is used to detect and quantify any duplicate entries in the dataset. Following this initial exploration, the data cleaning process is carried out. One key step in the cleaning process is converting the "Book-Rating" column to a numeric format, ensuring that all values are properly formatted for analysis. Finally, the *Ratings.csv* file is merged with the *Books.csv* file based on the ISBN to combine book details with corresponding ratings, providing a more comprehensive view of the dataset for further analysis.

2. Model Development:

A popularity-based recommendation system works by counting the number of ratings each book receives and calculating the average rating per book. It then filters books based on a threshold number of ratings to ensure that only well-reviewed or frequently rated books are considered for recommendation. On the other hand, collaborative filtering leverages user-item interaction data to generate personalized recommendations. This method often employs similarity measures, such as cosine similarity or matrix factorization, to identify patterns in user

preferences and recommend books based on the preferences of similar users. Together, these techniques offer both general and personalized approaches to book recommendations.

3. Model Compilation and Training:

When a machine learning model, such as collaborative filtering with matrix factorization, is used, it is trained using a ratings dataset. The model learns to predict book ratings for users by analyzing their past behavior, identifying patterns in their preferences and interactions. By doing so, the model is able to make personalized recommendations, predicting which books a user is likely to rate highly based on the preferences of similar users or their own historical ratings. This approach enhances the recommendation system by tailoring suggestions to individual users.

4. Model Performance and Evaluation:

To evaluate the performance of a recommendation system, several metrics are commonly used. RMSE (Root Mean Squared Error) is employed to assess the accuracy of rating predictions by measuring the difference between predicted and actual ratings. Precision

and Recall are used to evaluate the ranking performance of the system, with precision measuring the proportion of relevant items recommended, and recall assessing the ability of the system to identify all relevant items. Validation is typically performed through methods like train-test splitting or cross-validation, which help ensure the model generalizes well to unseen data and avoids overfitting.

5. Book Recommendation:

The recommendation system suggests books using two main approaches. The popularity-based approach highlights top-rated and frequently rated books, providing suggestions based on the overall preferences of a large number of users. Additionally, user-based recommendations offer personalized suggestions by analyzing the interests of similar users. This method identifies patterns in user behavior and preferences, recommending books that users with similar tastes have enjoyed. Together, these approaches help ensure that the system provides both general and tailored book suggestions to meet diverse user needs.

Dataset

ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
195153448	Classical I	Mark P. O. Morford	2002	Oxford University Press	http://images.ama	http://images.an	http://images.amazon.com/images/P/0195153448.01.LZZZZZZZ.jpg
2005018	Clara Calli	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.ama	http://images.an	http://images.amazon.com/images/P/0002005018.01.LZZZZZZZ.jpg
60973129	Decision ii	Carlo D'Este	1991	HarperPerennial	http://images.ama	http://images.an	http://images.amazon.com/images/P/0060973129.01.LZZZZZZZ.jpg
374157065	Flu: The St	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.ama	http://images.an	http://images.amazon.com/images/P/0374157065.01.LZZZZZZZ.jpg
393045218	The Mumn	E. J. W. Barber	1999	W. W. Norton & Company	http://images.ama	http://images.an	http://images.amazon.com/images/P/0393045218.01.LZZZZZZZ.jpg
399135782	The Kitch	Amy Tan	1991	Putnam Pub Group	http://images.ama	http://images.an	http://images.amazon.com/images/P/0399135782.01.LZZZZZZZ.jpg
425176428	What If?:	T Robert Cowley	2000	Berkley Publishing Group	http://images.ama	http://images.an	http://images.amazon.com/images/P/0425176428.01.LZZZZZZZ.jpg
671870432	PLEADING	Scott Turow	1993	Audioworks	http://images.ama	http://images.an	http://images.amazon.com/images/P/0671870432.01.LZZZZZZZ.jpg
679425608	Under the	David Cordingly	1996	Random House	http://images.ama	http://images.an	http://images.amazon.com/images/P/0679425608.01.LZZZZZZZ.jpg
074322678X	Where You	Ann Beattie	2002	Scribner	http://images.ama	http://images.an	http://images.amazon.com/images/P/074322678X.01.LZZZZZZZ.jpg
771074670	Nights Bel	David Adams Richards	1988	Emblem Editions	http://images.ama	http://images.an	http://images.amazon.com/images/P/0771074670.01.LZZZZZZZ.jpg
080652121X	Hitler's Se	Adam Lebor	2000	Citadel Press	http://images.ama	http://images.an	http://images.amazon.com/images/P/080652121X.01.LZZZZZZZ.jpg
887841740	The Middl	Sheila Heti	2004	House of Anansi Press	http://images.ama	http://images.an	http://images.amazon.com/images/P/0887841740.01.LZZZZZZZ.jpg
1552041778	Jane Doe	R. J. Kaiser	1999	Mira Books	http://images.ama	http://images.an	http://images.amazon.com/images/P/1552041778.01.LZZZZZZZ.jpg
1558746218	A Second	Jack Canfield	1998	Health Communications	http://images.ama	http://images.an	http://images.amazon.com/images/P/1558746218.01.LZZZZZZZ.jpg
1567407781	The Witch	Loren D. Estleman	1998	Brilliance Audio - Trade	http://images.ama	http://images.an	http://images.amazon.com/images/P/1567407781.01.LZZZZZZZ.jpg
1575663937	More Cunt	Robert Hendrickson	1999	Kensington Publishing Corp.	http://images.ama	http://images.an	http://images.amazon.com/images/P/1575663937.01.LZZZZZZZ.jpg
1881320189	Goodbye t	Julia Oliver	1994	River City Pub	http://images.ama	http://images.an	http://images.amazon.com/images/P/1881320189.01.LZZZZZZZ.jpg
440234743	The Testar	John Grisham	1999	Dell	http://images.ama	http://images.an	http://images.amazon.com/images/P/0440234743.01.LZZZZZZZ.jpg
452264464	Beloved	F Toni Morrison	1994	Plume	http://images.ama	http://images.an	http://images.amazon.com/images/P/0452264464.01.LZZZZZZZ.jpg
609804618	Our Dumb	The Onion	1999	Three Rivers Press	http://images.ama	http://images.an	http://images.amazon.com/images/P/0609804618.01.LZZZZZZZ.jpg
1841721522	New Vege	Celia Brooks Brown	2001	Ryland Peters & Small Ltd	http://images.ama	http://images.an	http://images.amazon.com/images/P/1841721522.01.LZZZZZZZ.jpg

`users.head()`

	User-ID	Location	Age
0	1	nyc, new york, usa	NaN
1	2	stockton, california, usa	18.0
2	3	moscow, yukon territory, russia	NaN
3	4	porto, v.n.gaia, portugal	17.0
4	5	farnborough, hants, united kingdom	NaN

Fig 1: Dataset of Books.csv

`ratings.head()`

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

```
ratings_with_name = ratings.merge(books,on='ISBN')

num_rating_df = ratings_with_name.groupby('Book-Title').count()['Book-Rating'].reset_index()
num_rating_df.rename(columns={'Book-Rating':'num_ratings'},inplace=True)
num_rating_df
```

	Book-Title	num_ratings
0	A Light in the Storm: The Civil War Diary of ...	4
1	Always Have Popsicles	1
2	Apple Magic (The Collector's series)	1
3	Ask Lily (Young Women of Faith: Lily Series, ...	1
4	Beyond IBM: Leadership Marketing and Finance ...	1
...	...	...
241066	Ä?Ä?piraten.	2
241067	Ä?Ä?rger mit Produkt X. Roman.	4
241068	Ä?Ä?sterlich leben.	1
241069	Ä?Ä?stlich der Berge.	3
241070	Ä?Ä?thique en toc	2

241071 rows × 2 columns

Fig 2: Popularity Based Recommender

The Figure displays a Python code snippet that processes book rating data. It merges two datasets (ratings and books) based on the ISBN, then groups the data by Book-Title to count the number of ratings each book received. The resulting DataFrame (num_rating_df) contains

two columns: Book-Title and num_ratings. The output table shows a list of books along with their respective rating counts, but some book titles appear to have encoding issues.

```
x = ratings_with_name.groupby('User-ID').count()['Book-Rating'] > 200
padhe_likhe_users = x[x].index

filtered_rating = ratings_with_name[ratings_with_name['User-ID'].isin(padhe_likhe_users)]

y = filtered_rating.groupby('Book-Title').count()['Book-Rating'] >= 50
famous_books = y[y].index

final_ratings = filtered_rating[filtered_rating['Book-Title'].isin(famous_books)]

pt = final_ratings.pivot_table(index='Book-Title',columns='User-ID',values='Book-Rating')

pt.fillna(0,inplace=True)

pt
```

Fig 3: Collaborative Filtering Recommender

The Figure shows a Python script filtering and pivoting book rating data. It first selects users (padhe_likhe_users) who have rated more than 200 books and then filters ratings from these users. Next, it selects books (famous_books)

with at least 50 ratings and further filters the dataset. Finally, it creates a pivot table (pt) where book titles are rows, user IDs are columns, and ratings are the values, replacing NaN values with 0.

Result

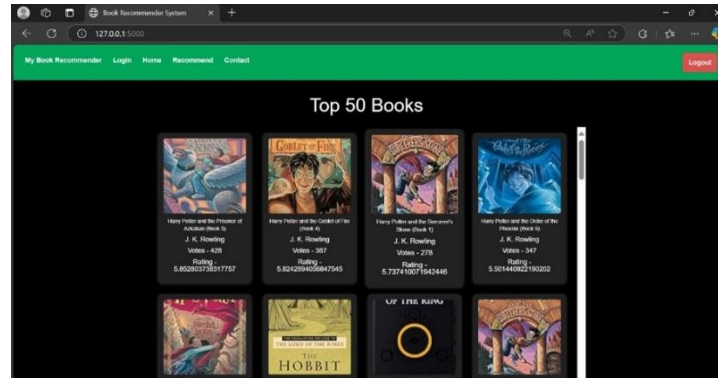


Fig 4: Top 50 Books

The Figure shows the top 50 books by using the Popularity based filtering. In this it give a top

50 books according to the users ratings.

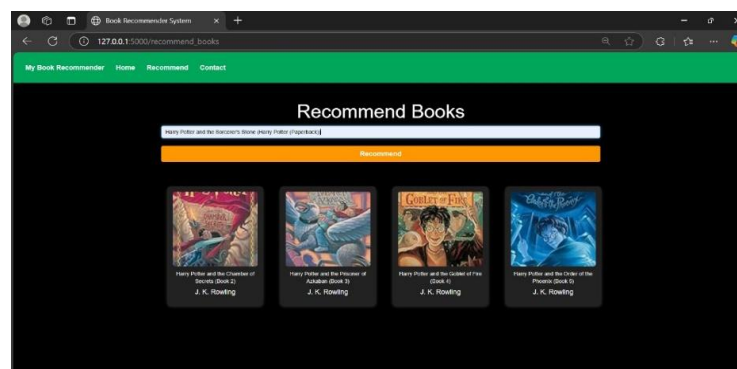


Fig 5: Recommended Books

The Figure shows that the system gives a recommended books using the collaborative filtering. In this the collaborative Filtering uses the user preferences for the recommendation.

Conclusion

This paper suggest a Book Recommendation System based on Machine Learning that aims to improve user experience through personalized book recommendations. The research discusses various recommendation strategies, such as popularity-based filtering and collaborative filtering, to enhance recommendation accuracy and user satisfaction. The approach consisted of data preprocessing and collection, such as missing value handling, duplicate removal, and data merging. The popularity-based model suggests books based on frequency of ratings and average rating, whereas collaborative filteringsuggests personalized recommendations based on user interaction. Model performance was measured using evaluation metrics like RMSE, precision, and recall to ensure accurate book suggestions.Future extensions can involve deep learning-based recommendation models, hybrid models blending content-based and collaborative filtering, and online recommendation updates in real time. This system can also be incorporated in online book

platforms in order to enhance user satisfaction and engagement.

References

- K. M. S. B. B. N. P. M. A. Anjali Sanjivanrao More, "Book Recommendation System Using Machine Learning," International Journal Of Creative Research Thoughts, p. 5, 2022.
- A. U. P. M. S. N. S. Ashlesha Bachhav, "Book Recommendation System using Machine learning and Collaborative Filtering," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), p. 5, 2022.
- D. K. A. J. Magesh.K, "Book Recommendation System Using Machine Learning," International Research Journal of Modernization in Engineering Technology and Science, p. 5, 2024.
- A. A. Sewar Khalifeh, "A Book Recommender System Using Collaborative Filtering methods," 2021.
- T. S. V. Helmina Mercy, "Book Recommendation System Using Machine Learning," IJCRT, p. 6, 2023.

P. Shahane, "A Survey on Book Recommendation Systems," p. 7, 2021.

M. D. B. P. M. Ms. Sushama Rajpurkar, "Book Recommendation System," IJIRST –International Journal for Innovative Research in Science & Technology, p. 3, 2015.

B. B. Himani Jain, "Book Recommendation Using Machine Learning," International Research Journal of Modernization in Engineering Technology and Science , p. 2, 2023.

N. A. Anant Duhan, "Book Recommendation System Using Machine," p. 7, 2023.