IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BOOK RECOMMENDATION SYSTEM WITH RELEVANT TEXT AUDIOBOOK GENERATION

[1]Bhagyashree Rao, [2]Nikhitha Jayaprakash, [3]Mahalakshmi Thevar, [4]Ujwala Ravale

[1]Student, [2]Student, [3]Student, [4]Assistant Professor

[1]Computer Engineering,

[1]SIES Graduate School Of Technology, Nerul, India

Abstract: In e-commerce today, contents available for users to explore are overwhelming as an average e-commerce website is about seventy percent (70%) more than a physical store in a total number of users and items. Thus, there arises the need to filter, prioritize and efficiently deliver relevant information. We use recommender systems for this purpose. Recommender systems are at the forefront of the ways in which content-serving websites like Facebook, Amazon, Spotify, etc. interact with their users. Book Recommendation system is such a recommender system built for book lovers. Using your profile, our system uses Machine Learning methods to provide you with highly personalized book recommendations. We are using a hybrid recommender system in order to provide recommendations from the dataset given by Goodbooks users (ratings and item features). We also offer more "conventional" suggestions that focus solely on the book's strengths. Audiobooks can improve readers' literacy by changing the way they listen, read, and learn. Our system allows users to download the audiobook and enjoy their reading time.

Index Terms **- Component, formatting, style, styling, insert.**

1. INTRODUCTION

A recommender system is a type of information filtering system which predicts the rating or preference that a user would give to an item. Recommender systems are sometimes referred to as recommendation systems. Recommendation system describes web applications that predict response to options.

We make use of a hybrid recommender system to power our recommendations. Here content-based filtering and collaborative filtering are two types of recommender systems that can be combined to form our hybrid system. A content-based recommender system aims to predict a user's characteristics or behavior based on the item's characteristics, which he or she responds positively to [3]. For example, if a user likes the first book of Harry Potter, and if the second book is similar to the first, then it recommends the second book to the user. Collaborative filtering on the other hand does not need the features of the items to be given. A feature vector characterizes every user and item. It collects user feedback on different items and uses them for recommendations [3]. Here user ratings are used to determine the user or item similarity in collaborative filtering. If there is a high correlation of users rating the first Harry Potter book and the second Harry Potter book, then they are deemed to be similar. Our hybrid system uses both of these approaches for more efficiency.

Now-a-days anything can be done on foot, which may include listening to the news, attending meetings, and even taking notes (with voice dictation). Reading is the only thing you can't do while walking. One of the best things about audiobooks is that they allow individuals to fit more reading into their already hectic schedules. Listening to audiobooks allows you to read in the “in-between” moments of your day when you would otherwise have to do something else.

2. LITERATURE SURVEY

This section provides an overview of the techniques available in the literature for recommendation systems, classification and clustering. Several techniques have been presented in the last decade, but there were considerable differences in each, with respect to the classification and accuracy. We could make a plan to go about the project with this initial survey. A summary of the different approaches and classification techniques for the book recommendation system is presented below.

In 2020, M. Kommineni, P. Alekhya, T. M. Vyshnavi, V. Aparna, K. Swetha and V. Mounika had presented their work in paper [4], which provides an overview of the Recommendation Systems that is currently employed in the operations of the online book shopping domain. This paper proposes an easy comprehensible machine for book hints that assist readers to signify the proper book, that's to be studied next.. They have used a User Based Collaborative Filtering (UBCF) approach and measured the performance of similarity measures in recommending books to a user.

In 2016, P. Mathew, B. Kuriakose and V. Hegde had presented their ideas in a research paper[5] having a Book Recommendation System (BRS) based on combined features of content based filtering (CBF), collaborative filtering (CF) and association rule mining to produce efficient and effective recommendations. They have proposed a hybrid algorithm in which we combine two or more algorithms, which use the recommendation system to recommend the book based on the reader's interest.

In 2014, K. Tsuji, F. Yoshikane, S. Sato and H. Itsumuran had proposed a method to recommend books through machine learning modules based on several features, including library loan records[3]. Several methods for automated book recommendation have been proposed in the literature, the effectiveness of these methods has not been researched very thoroughly. The study has examined the effectiveness of book recommendation methods that use Support Vector Machines (SVMs), Random Forest modules.

The paper[6] provides a thorough review of the research literature on audiobooks and ebooks. The review encompasses literature from various disciplines. Additionally, the destiny of those formats is explored in addition to the results for libraries.

After conducting the Literature survey, we found that most existing book recommendation systems lead to extraction of irrelevant information and lead to lack of user satisfaction. We could find that very limited research is conducted on audiobook generation projects. Thus, we decided to build our Book Recommendation System in such a way that the User gets the Recommendation of the books and can download an audiobook of the same if required.

3. PROPOSED WORK

Our main aim is to provide the best book recommendations to our users. We use a hybrid recommender system for this purpose. Hybrid systems are the combination of two other types of recommender systems: content-based filtering and collaborative filtering. Using collaborative filtering we recommend some books to the user on login. Using content based filtering we recommend books based on the book name searched by the user. We use the K-means clustering algorithm to cluster books based on rating. To extract the relevant text to include in the audio book, we use Jenks Optimization Algorithm and we generate the book with the help of Google Vision API. The block diagram of the system is shown below:

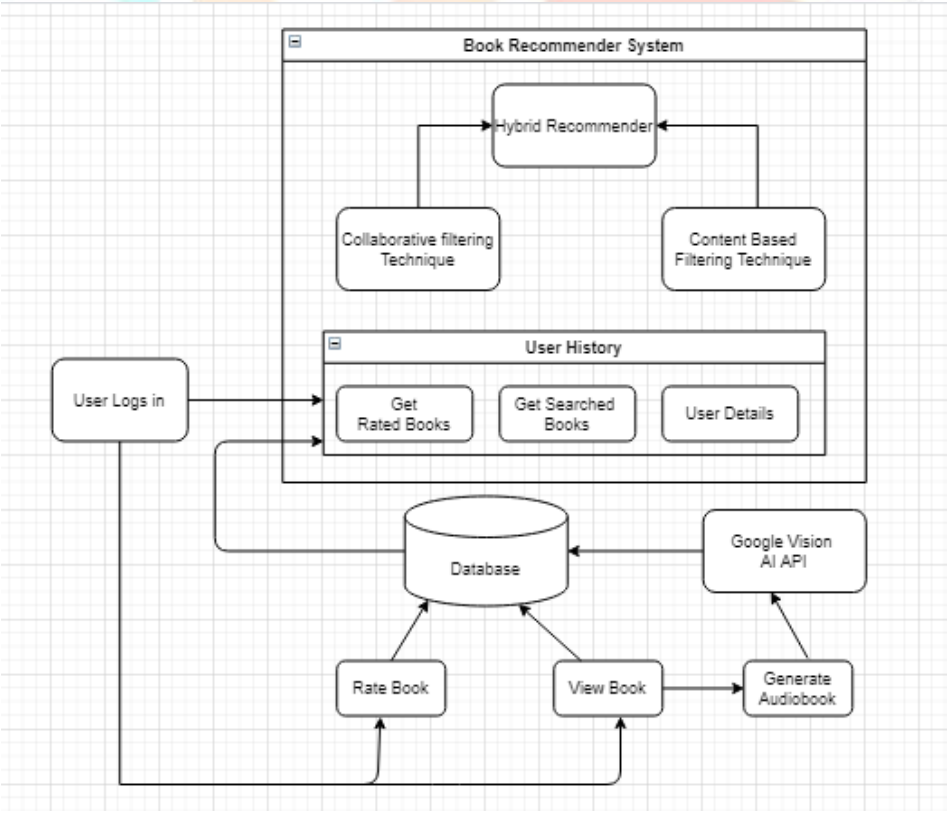

Fig-1: Block diagram of the recommender system.

For the audiobook generation, we require Google Vision AI API and Jenks Break Algorithm.Google Cloud Vision API returns not just text on the page, but also its layout. It groups text into chunks (pages, blocks, paragraphs, words, and characters) and returns its location on the page. In particular, for each word, it returns a bounding box[4].

3.1 RECOMMENDATION SYSTEM

In this section we have described how each of the algorithms/methods mentioned in the previous section have been utilized effectively with respect to our project.We use a hybrid recommender system which is the combination of two other types of recommender systems: content-based filtering and collaborative filtering.

3.1.1 Collaborative filtering

Collaborative filtering is a method by which user ratings are used in order to determine user or item similarities[2].For collaborative filtering, the primary features necessary are user_id, book_id, and ratings. Collaborative filtering does not need the features of the items to be given. Every user and item is described by a feature vector. It collects user feedback on different items and uses them for recommendations [2].We have applied K-means clustering and Gaussian mixture modelling to cluster the users and reach the best silhouette score. We have used the euclidean similarity distance measure. The formula for Euclidean similarity distance measure is given as follows:

$$\frac{1}{1 + d(p_1, p_2)}$$

Fig-2: Formula for euclidean distance similarity.[4]

3.1.2 Content-Based filtering

Content-based filtering is a method of recommending items by the similarity of the said items..A content-based recommender system aims to predict a user's characteristics or behaviour based on the item's characteristics, which he or she responds positively to[2]. For example, if a user likes the first book of Harry Potter, and if the second book is similar to the first, then it recommends the second book to the user.We use cosine similarity and our item similarities are a combination of user ratings and features derived from books themselves. The formula for cosine similarity is given as:

$$\cos(x, y) = \frac{(x \bullet y)}{\|x\| \|y\|}$$

Fig-3: Formula for cosine distance similarity.[7]

3.2 AUDIOBOOK GENERATION

When we try to generate the audiobook, from an existing pdf of the book, a conventional audiobook converts each and every text present in the pdf to audio. To avoid including irrelevant text in our audiobook, the three given steps are followed:

3.2.1. Extract text from PDFs (or images)

Using the Google Vision API, we are going to convert the pdf text or image into text. When you pass a document through the Vision API, you're returned both raw text as well as layout information. The API returns not just the raw text on the page, but also each character's (x, y) position.

3.2.2. Decide which parts of the text to include in the audiobook

Using the layout information, we can calculate the font size of the text and classify the font heights using machine learning. To figure out the cut offs Jenks Natural breaks algorithm can be used.

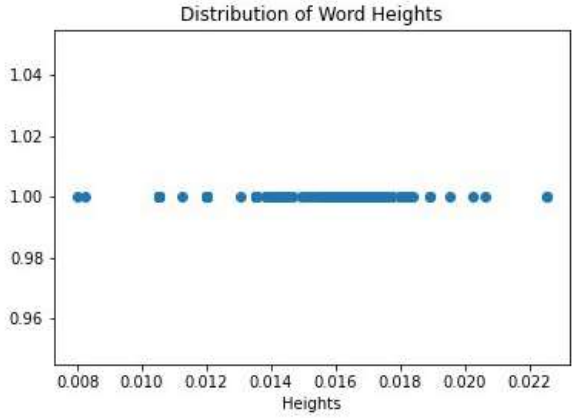

Fig-4: Sample demonstration for the distribution of
font sizes from the selected document to be converted into audio book

3.2.3. Convert the text into spoken words

Using the Google text-to-speech API, the relevant text extracted can be converted into audio.

4. DATASET

While there are many book datasets available to use, we decided to work with Goodreads Book data. There are several full Goodreads data sets available in kaggle, and we initially worked with this data to analyze metadata for books, authors, series, genres, reviews, and the interactions between users and items. Once we began building the models, we realized that the dataset was too large, so we processed the dataset and removed the books with the least rating.This dataset contains six million ratings for ten thousand most popular (with most ratings) books.

Table 4.1: Fields in the dataset

S.No	Attribute	Type of Data
1	book_authors	String
2	book_desc	String
3	book_format	Nominal
4	book_rating	Numeric
5	book_rating_count	Numeric
6	book_title	String
7	genres	String
8	image_url	String

5. EXPERIMENTAL RESULTS

5.1 Login Page

The user has to login to the website in order to view the various functionalities the website offers.

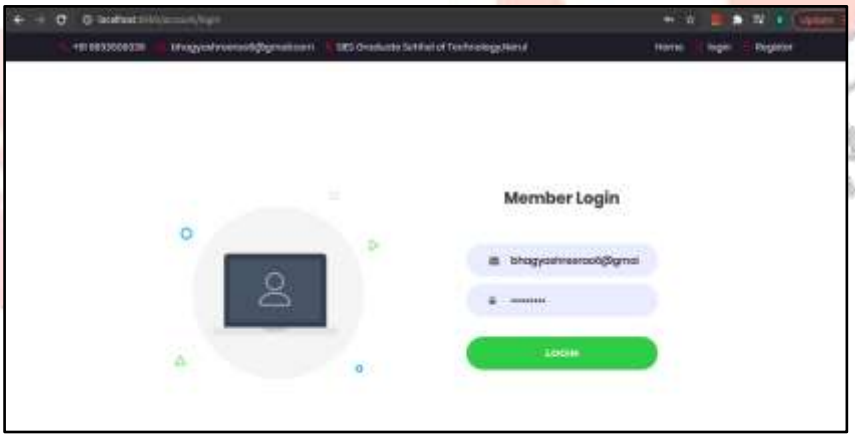

Fig-5: Login page of website

5.2 Home Page

This is the homepage of the website. The recommended books and popular books are displayed here.

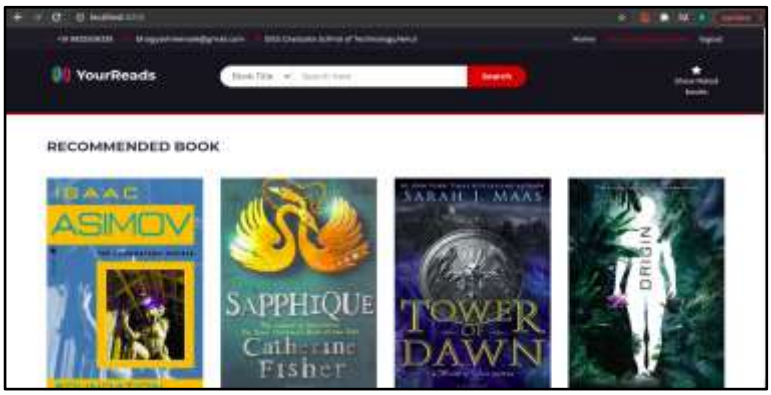

Fig-6: Homepage of website

Fig- 7: Popular books displayed on home page of website

5.3 Search a book

The search bar allows the user to search a book with the help of the book name.

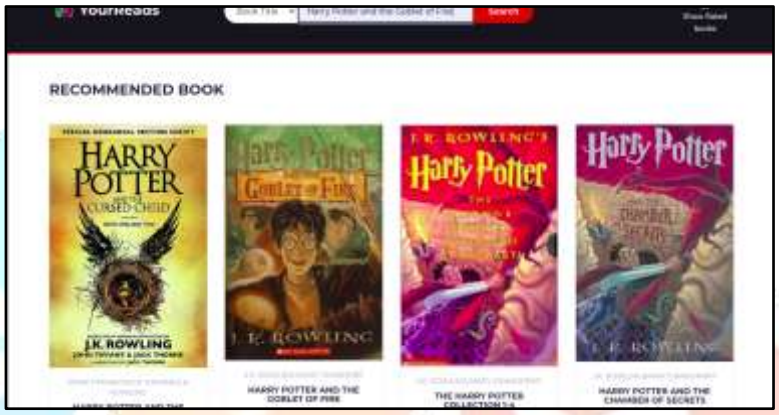

Fig-8: User search a book using the book name

5.4 Obtained Book Recommendations

Based on the book name searched by the user, the system will recommend some books and display the same to the user.

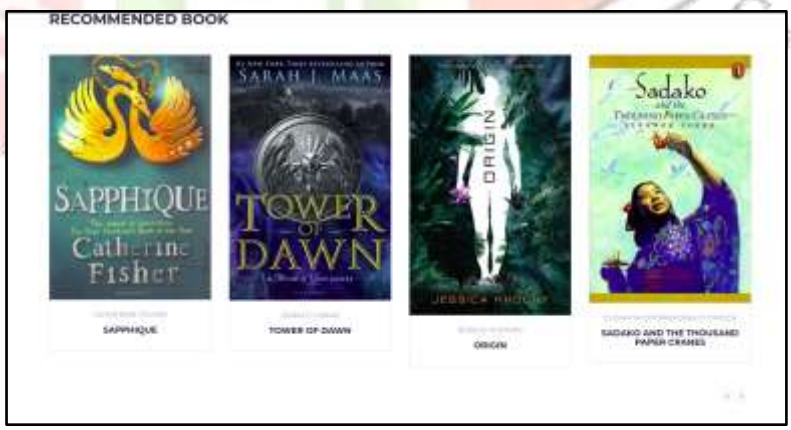

Fig-9: User gets book recommendations based on the search

5.5 View a book

The user can view a book by clicking on it. The book description, average rating, total rating, Author name, Genre, Format, Edition, Price is displayed. The user can submit the rating for the book. The user may also click on the Generate Audiobook button if he/she wants to download or listen to the audiobook.

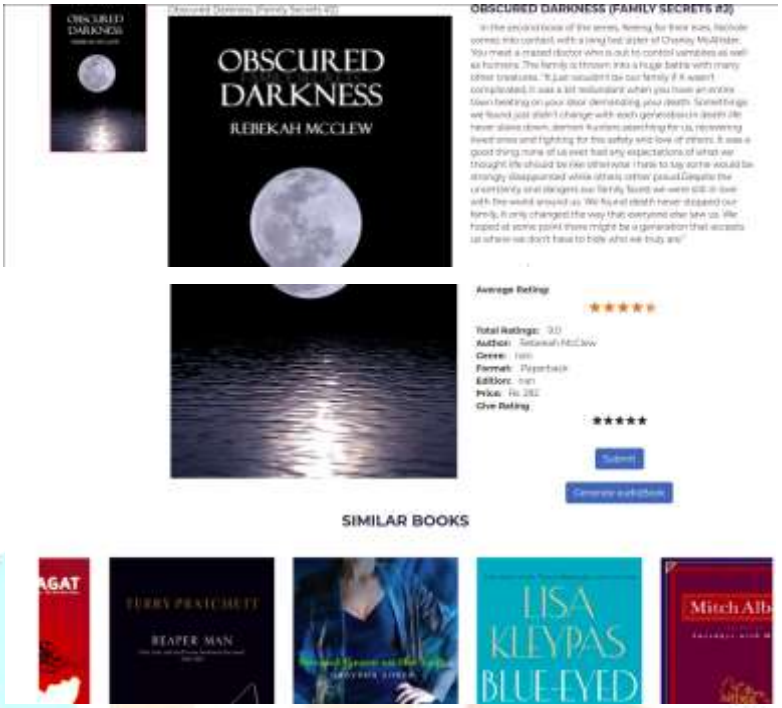

Fig-10: Page displayed on viewing a book

5.6 Audiobook

On clicking the 'Generate Audiobook' button, the user gets a link below the button. On clicking the open book button the user can listen to or download the book.

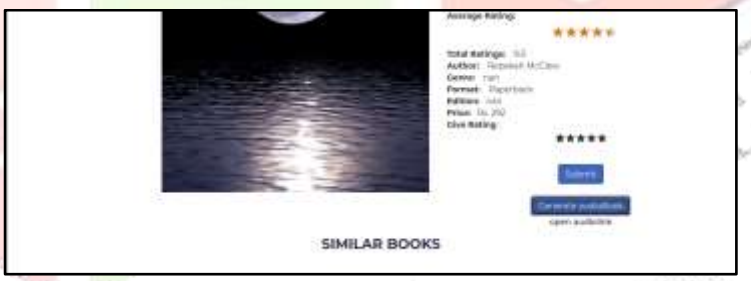

Fig-11: Generate Audiobook

Fig-12: Listen or Download the Audiobook

5.7 Rate a book

The user can rate a book if he/she wishes to and can submit the rating for the same. On the home page, there is a button 'Show Rated Books' where the user can view the books rated by him/her.

Fig-13: Submit Rating for a book

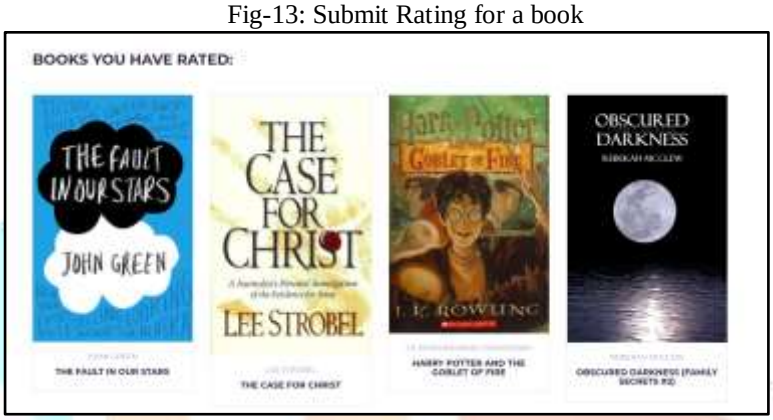

Fig-14: View rated books

6. CONCLUSIONS

This project was able to present a comprehensive review on research previously targeted on improving recommender systems. From the results and visualizations, we can deduce that the accuracy of rating followed a normal distribution which suggests consistency and efficiency.

We were able to train efficient models that had high accuracy. The model performed well on random inputs outside the dataset as well, we had tried to generate an audiobook for a research paper. We also tried to generate an audiobook, given a handwritten document. The former and later experiments gave us accuracy of above 70%.

The System has adequate scope for improvement in the future. We can develop and launch a Mobile app. Also System security, data security and reliability could be the major features which can be done in future.We can also add the API for the shopping and payment gateway. In the existing system there are only some selected categories like we search using book names, so as an extension to the site we can add more categories. Also we can add admin side with some functionalities like books management, User management etc.

REFERENCES

[1] Resnick, P., and Sami, R. (2007). The influence limiter: provably manipulation-resistant recommender systems. RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems, ACM, New York, USA. 25–32.

[2] https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421

[3] K. Tsuji, F. Yoshikane, S. Sato and H. Itsumura, "Book Recommendation Using Machine Learning Methods Based on Library Loan Records and Bibliographic Information," 2014 IIAI 3rd International Conference on Advanced Applied Informatics, pp. 76-79.

[4] M. Kommineni, P. Alekhya, T. M. Vyshnavi, V. Aparna, K. Swetha and V. Mounika, "Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm," 2020 Fourth International Conference on Inventive Systems and Control (ICISC), pp. 66-71.

[5] P. Mathew, B. Kuriakose and V. Hegde, "Book Recommendation System through content based and collaborative filtering method," 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), pp. 47-52.

[6] Jessica E. Moyer, "Audiobooks and E-books: A Literature Review",2012 Reference & User Services Quarterly, vol. 51, no. 4, pp. 340–54.

[7] Saikat Bagchi, "Performance and Quality Assessment of Similarity Measures in Collaborative Filtering Using Mahout", 2015 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15) .