Book recommendation system using Machine Learning

Goli Siva Sankar Tarun
Student
Department of Computer Science and Engineering
Narasaraopet, India
golisivasankartarun225@gmail.com

Galam Sudheer
Student
Department of Computer Science and Engineering
Narasaraopet, India
sudheergalam@gmail.com

Kunala Kalyan
Student
Department of Computer Science and Engineering
Narasaraopet, India
kalyankunala@gmail.com

Bolla Jhansi Vazram
Professor
Department of Computer Science and Engineering
Narasaraopet, India
jhansi.bolla@gmail.com

Abstract – Due to COVID-19 pandemic the usage of online books is rapidly increasing, from a huge e-book space finding particular books becomes a immense challenge for users. To perform effective search which mine connected books based on user interest and rating using Book recommendation system. popularity-based and collaborative based methods are used in these systems and based on user ratings and interest. This System for recommending books for users that rating a book using the clustering and above methods and then finding a same of that book to recommend a new book.

Keywords — Books recommendation, Truncated-SVD, Clustering , Popularity-based, Collaborative-based, Root Mean-Square Error.

I. INTRODUCTION

There is an ever-increasing amount of information available to us in today's recent times, that includes books, music, movies, and more. It can be testing for users to find what they are consider for with this more amount of information and to develop new items that match their interests and satisfaction. Through the vast amount of information this is where recommendation systems come in, helping users to navigate and find what they are looking for more easily.
This highly scalable and trustworthy system for suggesting books can be adjusted for various genres and user preferences. This books recommendation system can be integrated into various types of online stores and digital libraries to offer customers which has individualized books recommendations. A dataset of books and users was used to calculate this system's performance, and the findings demonstrate that the proposed system outperforms current book recommendation systems. With the assistance of this project, users will be able to find new books to enjoy and contribute to the creation of a more sophisticated and individualized.

Book recommendation systems use machine learning algorithms to analyse data from various sources such as users, ratings and books datasets to provide customizedrecommendations to users. These recommendations can help users to optimize their Books recommends, reduce the amount of time, and mitigate risks associated with users' information and data.

This System presents a Book recommendation system that utilizes machine learning algorithms to provide accurate and personalized recommendations to users. This system analyzes data from various sources, and performed various methods such as Popularity-based and collaborative learning methods are used. The system's primary goal is to provides the quality of the recommendations which is very accurate, easy to maintain and simply to use which is using by the user.

The description of the items is used in content-based filtering, which provides suggestions for items that are comparable to the description of the items. Book are recommended using these multiple filtering models depending on multiple the book's content and the user actions. As a result, mine recommendation engine also recommends books to new readers. For clustering the users in this study, we using both techniques as: K-means and Gaussian mixture. To calculate error between absolute numbers and the results, use the Root Mean Square Error formula.

RMSE number used to determine basic accuracy.

II. LITERATURE SURVEY

The majority of analyzers used the Pearson's Coefficient function to determine how comparable book-rating were in order to make book recommendations:

1. "Collaborative Filtering for the Book Recommendation System" by D. R. K. Srinivas and S. S. Suresh Kumar, published in the International Journal of Computer Science and Mobile Computing in 2017. This paper explores the use of collaborative filtering for book recommendations, using a dataset which of the user ratings and book metadata in system.

1. "Content-based Recommendation Systems" by Pazzani, M. J., & Billsus, D. (2007)- This survey focuses on book recommendation systems using machine learning. It discusses the importance of author, and keywords and the challenges associated with it, such as limited data availability and high variability in the data. The survey also covers the different types of machine learning models used for recommendation, including decision trees, support vector machines.

2. "A survey on Book recommendation systems using machine learning" by Bell & Volinsky (2009) - This survey provides a comprehensive overview of Book-recommendation-system which is use Machine learning, including their architecture, and data sources, also performance evaluation metrics. It covers different types of methods for machine learning used in recommendations, such as rule-based systems, collaborative-filtering, and content-based-filtering. The survey also discusses the challenges and limitations of book recommendation systems, such as data sparsity.

Overall, these surveys demonstrate the growing interest and potential of Book recommendation systems using machine learning in the learning sector, and highlight the need for further research to overcome the challenges associated with these systems.

III. EXISTING SYSTEM

Machine learning algorithms are used in a book recommendation system to analyze user data and offer personalized book recommendations based on reading tastes and history. The primary problem with this is that these systems' recommendations are not transparent. Users may not always understand why a specific book was suggested, which can make it challenging for them to believe in the system and its suggestions. Additionally, the suggestions might not always suit the user's preferences or tastes, which could cause frustration and user engagement. Finally, systems for recommending books that heavily depend on machine learning might need a lot of data to properly train the algorithms. For smaller businesses or groups with less access to data, this can be difficult. Furthermore, the process of gathering and storing user data may give rise to privacy issues, which may discourage some users from using the system.

IV. METHODOLOGY

The methodology for building a book recommendation system using machine learning involves collecting and preprocessing data, extracting relevant features, selecting an appropriate algorithm, training and evaluating the model, and deploying the system for user consumption. Data collection, second is Data preprocessing, third is Feature-extraction, fourth is Algorithm-selection, fifth is Model-training, sixth is Model-evaluation, and the seventh one is Deployment.

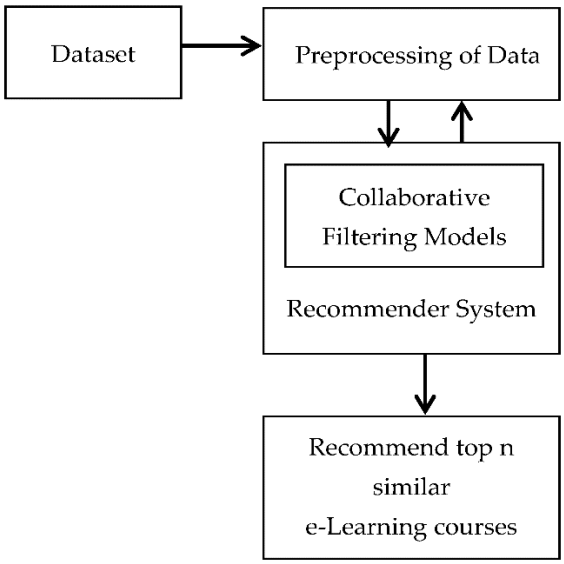

Fig. 1: proposed system task flow

A. DATASET COLLECTION

The Dataset is collect from the Kaggle website: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset.

we have 3 files in our dataset which is extracted from above kaggle website.

Books - first, there are about books, which includes all the details about books, including an author, title, year of release, etc.

Users: Information about registered users, such as user id and address, is contained in the second file.

Ratings – Ratings contain information like which user has given how much rating to which book.

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...

Fig. 2: Dataset diagram

	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

Fig. 3: Dataset diagram

	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.4: Dataset diagram

A. Data Cleaning and Feature Extraction

The data must be cleaned before being given to any machine learning programs. In the data cleaning procedure, outliers and null numbers are eliminated. There are no null values in the dataset that was gathered. Features are extracted from the data collection after data cleaning. We must examine the correlation between the six distinct features in the dataset that have an impact on the model's output. Correlation is useful for removing characteristics that have little impact on the model's output.

B. Model Architecture.

In this step dataset is splitted in 80% and 20% for training and testing. During training, machine learning algorithms finds the relation between the input and output features. By using this relation, the model able to predict the outputs to the new input values. In this we used the following four machine learning models:

1. content-based-Filtering: The program recommends an item that is comparable that have been used or viewed. To put it simply, this program looks for items that resemble one another. As an illustration, As a result of the two films' similar label and content, if someone likes watching movies, he might also like web series. Only the appearance of the content is identical, and there isn't much attention paid to the viewer. On the basis of prior tastes, it only suggests the product with the highest score.

2. collaborative-based-Filtering: System recommendations that use collaborative-filtering on prior encounters between users and the goal items. To put simply, attempt to notice customers who give them and who resemble them products recomend on the options made by their lookalikes. Let's use an illustration to clarify. Users X and Y share some characteristics, and X has seen movies A, B, and C. If user Y has seen user B, user C, and user D, we will suggest user A to user Y and user D to user X.

3. K-Means: The unlabeled dataset is divided into different clusters by the unsupervised learning algorithm K-Means Clustering. so each collection is only goes to single group shares common feature. In this case, K designates the quantity of predefined groups. In this method, each cluster must be connected to a centroid. A data point's total distance from each of its associated clusters should be as small as possible.

Using the unlabelled dataset as input, k groups are created from the unlabelled dataset. This procedure repeated until no better clusters are found. The k number for this algorithm must be decided in advance.

4. Gaussian Mixture: The clustering methods used by Gaussian Mixture models are effective. The model presupposes that there are a predetermined number of Gaussian distributions, each of which represents a cluster. The data elements are combined in this model to form a single distribution. To assign data points to Gaussian distributions, these models employed the soft clustering method. In a one-dimensional space, the probability density function of the Gaussian distribution.

V. RESULT AND ANALYSIS

After created all the four models by using the training data , we test the models by using the test data and calculates the accuracies of the models. The following table shows the accuracies of different models:

	K-Means	Gaussian mixture
Silhouette Score	0.0433968332584411	0.016778872313688933

Table1: Two models' silhouette ratings

It is determined the differences between the both the typical mean ratings of books for test users and the typical mean ratings of books in the cluster.

Mean rating for 10 random books	3.8876949740034736
Mean rating for 10 books of cluster's favourites	4.3735008665511135

Table2: Mean-scores comparison

By using RMSE, one can calculate the difference between data values and values predicted by a model.

RMSE	0.5957791790493179
Accuracy	1.167727190936663

Table3: Values for RMSE and Accuracy

VI. DEPLOYMENT

The final model is deployed into an application by using flask module of python by creating the user interface which can be accessible easily by the users. It is possible to design a system have to that provides users with high-quality of book recommendations and fosters long-term maintenance and adherence. The application simply takes the input and output provides.

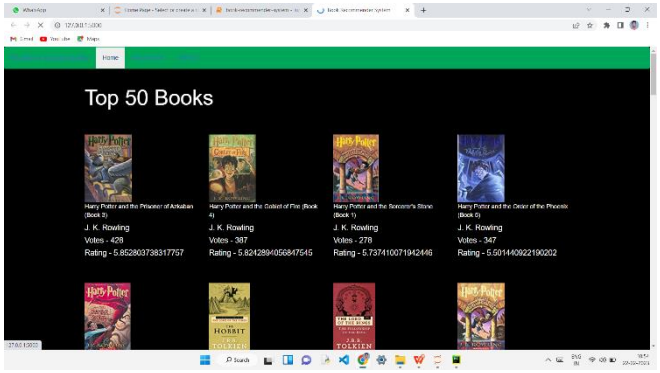

Fig.5: User interface

The following is the recommendations based on user ratings and After typing, select "Search" to have the suggested book, based on user ratings, displayed on the screen. The screenshot of the novels with user ratings is shown in Figure 6.

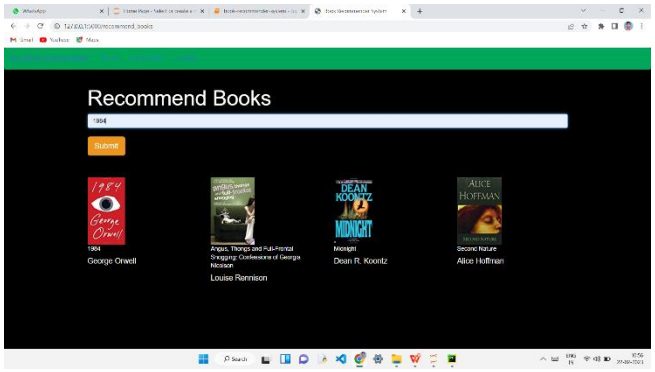

Fig.6: Recommended books

VII. CONCLUSION.

In this project, we show a collaborative filtering-based recommendation system. The primary objective was to speed up recommendations by developing a system that can give users good recommendations without requiring them to register for an extended period of time.

This project has highlighted the benefits of Machine learning in recommendation system, and how it can help improves the efficiency and the achieves of the book recommendation. By automating the recommendation process, users can save time and effort, while also reducing the risk of making costly mistakes.

VIII. REFERENCES

1. Avi Rana and K. Deeba et.al, “Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity)” in IOP ebooks 1362, 2019.

2. G. Naveen Kishore, V. Dhiraj, Sk Hasane Ahammad, Sivaramireddy Gudise, Balaji Kummaraa and Likhita Ravuru Akkala, “Online Book Recommendation System” International Journal of Scientific & Technology Research vol.8, issue 12, Dec 2019.

3. Ms. Sushma Rjpurkar, Ms. Darshana Bhatt and Ms. Pooja Malhotra, “Book Recommendation System” International Journal for Innovative Research in Science & Technology vol.1, issue 11, April 2015.

4. "Collaborative Filtering for Book Recommendation System using Matrix Factorization" by Anand Rathi. This paper presents a matrix factorization approach to building a book recommendation system and evaluates the system using a dataset of book ratings.

5. "Recommender Systems Handbook" curated by Lior Rokach, Bracha Shapira, and Francesco Ricci. This comprehensive handbook provides an in-depth exploration of the theory and practice of recommender systems, including chapters on collaborative filtering and content-based filtering approaches.

6. "A Hybrid Book Recommendation System using Content-Based Filtering and Collaborative Filtering Techniques" by Prasanna Kumar and J. V. R. Murthy. This paper presents a hybrid approach to building a book recommendation system that combines content-based and collaborative filtering techniques.