International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 235-256
https://doi.org/10.70917/ijcisim-2026-4209
Article

A Recommendation System for Books as Part of Smart Education

Barkha Narang[1], Om Prakash Sharma[2]

[1]Jagannath University, Jaipur. Assistant Professor, Poornima College of Engineering, Jaipur, India, barkha.narang@poornima.org, https://orcid.org/0000-0001-7767-5952
[2]Faculty of Engineering & Technology, Jagannath University, Jaipur, omprakash1.sharma@jagannathuniversity.org, https://orcid.org/0000-0003-4529-9037

Abstract: Innovative solutions that attempt to enhance conventional learning system include smart academic system and machine learning-based recommendation systems. Machine Learning technologies are used in smart learning system to provide a more dynamic and customized learning environment. With the help of this technology, students can analyze, get appropriate guidance to ameliorate the quality of educational and evaluation strategies for raising the quality of instruction. The recommendation systems also analyze student performance data using machine learning algorithms and offer tailored suggestions to help students achieve better learning outcomes. These recommendations may include ideas for extra reading material, course selection, book selection result analysis, and placement options. Contingent on the student's engrossment and objectives, the recommendation systems may also be utilized to offer suitable courses and learning routes. The research presented in this paper focuses on how different machine learning algorithms may provide better results for smart education, making the knowledge gaining process simple and precise. This research work helps in planning & execution of Intelligent Recommendation System (IRS) & its effective evaluation. It is further analysed that individualized and successful learning experience may transform the education industry into a smart education and recommendation system employing machine learning.
Keywords: Online Learning, Campus Recruitment, Learning Based Outcomes, Management System

1. Introduction

Recently, there has been an uptick in the use of recommender systems in the academic world. Given the plethora of information that is now easily accessible online, it is vital to develop educational recommendation systems that include the perspectives of both parents and children when choosing which engineering institute would be perfect for a certain learner. We use this information to create a comprehensive resource for academic advising at the university level. An ontology that characterizes both the resource and the learner is often used. The proposed solution combines both the traditional functional modules teachers are familiar with a predictive model that can foresee a student's progress in professional institutes. According to the model's projections, the college recommendation system has the prospective to be of use to first-year students in making the best possible book selection available [1, 2]. The proposed model, which represents the students' view is a crucial part of these recommendation systems. The key objective of this research paper is to discover what resources students are using in enriching their knowledge To facilitate academic administration and promote enhanced connections between teachers and students, number of Indian universities have established innovative information systems and services. A "student referral program" (SRM) demonstrates this. The strategy may or may not be effective, depending on the counsellor's knowledge and experience.

Universities in India are exerting a great deal of energy to enhance the quality of education they provide, with a particular emphasis on creating strategies to increase student retention and graduation rates. Increasingly, performance is included in assessments of a university's prestige and standing [3]. Almost all colleges and universities provide some kind of counselling to their students.

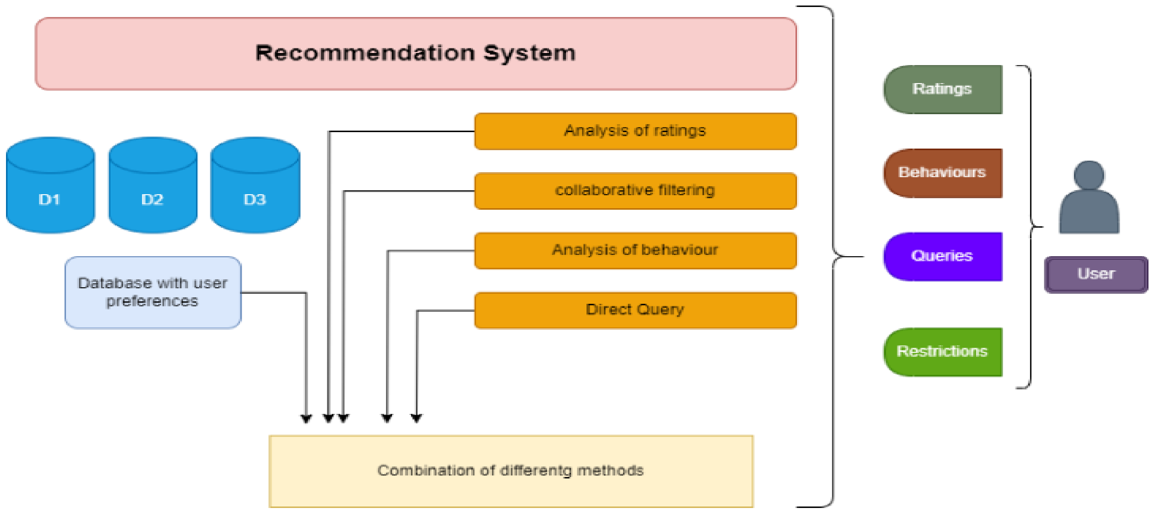


Figure 1: Recommendation System

Figure 1 shows a recommendation system process as in how the combination of different methods like analysis of ratings, collaborative filtering, analysis of behaviour, direct query, and database with user preferences, gives the output of ratings, behaviour, queries, and restrictions to the user.

Some of the key group of recommendation system:

Content-Based Recommendation System [3]: This system makes recommendations based on the characteristics, quality, description, and other features of the product. Products in content-based recommendation systems are sorted into groups depending on their characteristics. This method takes into account the user's tastes and then makes recommendations.

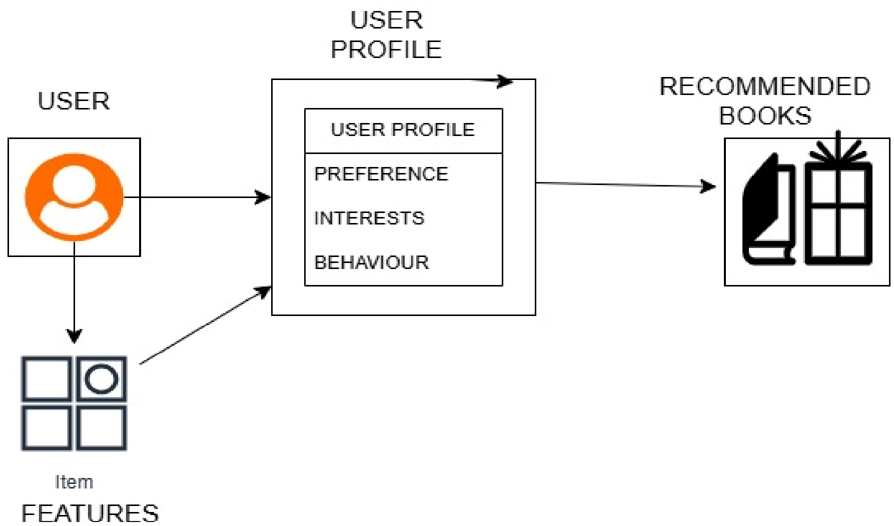


Figure 2: Content- Based Recommendation System

The fundamental goals of this recommendation system are data retrieval and data filtering. Information retrieval is a technique for locating desired material via the use of metadata and indexing. Information filtering is a technique used to enhance process efficiency by eliminating irrelevant, inaccurate, or duplicated information. Among the various methods used are decision trees, Bayesian classifiers, cluster analysis, and artificial neural networks [3,7,8].

Collaborative-Based Recommendation Systems: The collaborative recommendation system takes into account user history and preferences. Without overemphasizing the item's features or content, the algorithm may provide reliable recommendations for complicated things. In certain cases, it has even outperformed algorithms that propose content. Implicit data capture and explicit data gathering are the two primary methods used here. The number of times a person would have looked for something, how interested they are in it, and whether or not they enjoy it are all examples of implicit data collecting. The user may be prompted to rate the item as part of explicit data gathering, and a log of his prior borrowings may be generated. [6,7] K-nearest Neighbour, correlation, factorization, and approximation algorithms are among the various approaches utilized here [7].

Hybrid Recommendation Systems: A content-based recommendation system and a collaborative editing-based recommendation system are brought together in Hybrid Recommendation Systems. The recommendations are based on constant comparisons made between the viewing and browsing patterns of persons with similar tastes. It can be more effective than content- and collaborative-based recommendation systems at times. Both a statistical and an expert system are integrated into this system to provide recommendations [10].

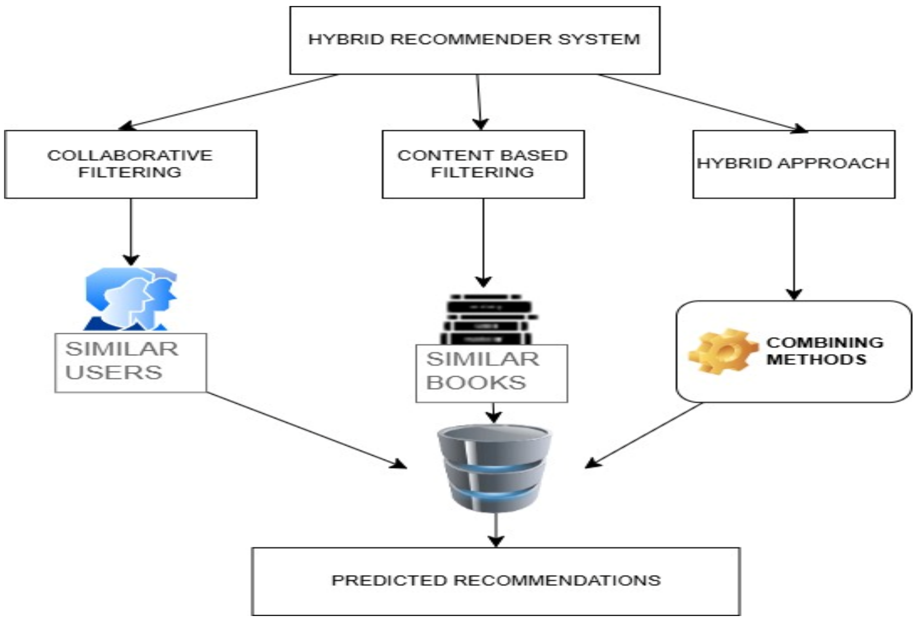


Figure 3: Hybrid Recommendation System

As shown in Figure 3, a Hybrid recommendation system takes into account user feedback and various demographic subgroups to provide product recommendations. It also takes care of understanding of the user's tastes and interests [11]. The hybridized method employed are:

Weighted: For the numerical total of how well each section of the proposal did.

Switching: For the system will choose and implement one of the parts of the plan.

Mixed: For all of the recommendations from all of the recommenders are shown together. [15, 16]

Classification of Recommendation Methods

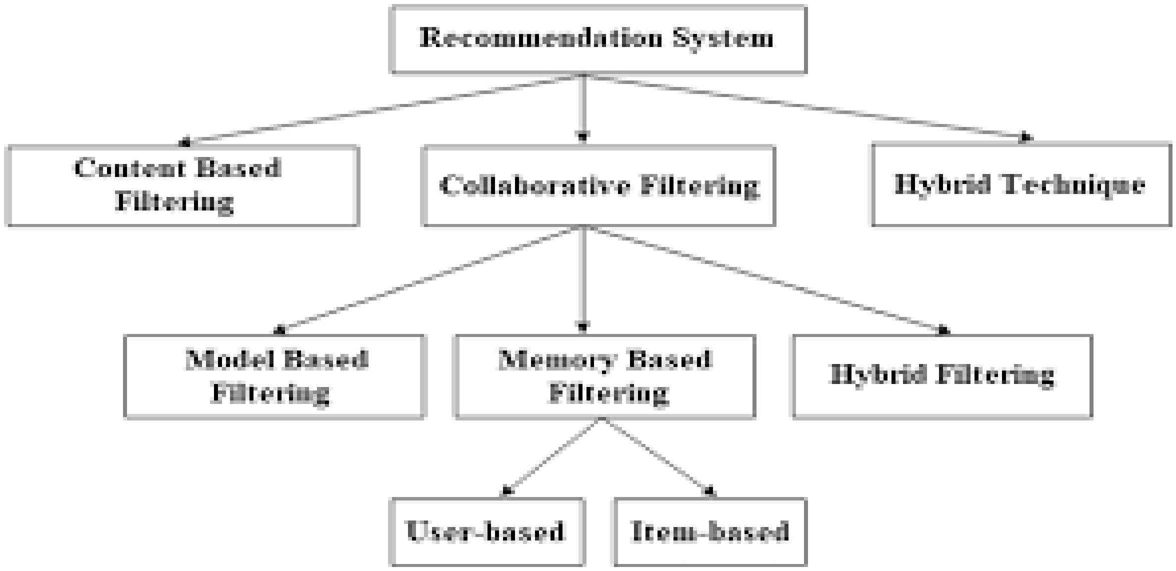


Fig 4. Types of Recommender systems [12]

Some key terminologies used in Recommendation System are:

Feature Combination: The recommendation algorithm is given a list of attributes that have been obtained from various sources; Feature Augmentation: A feature or collection of characteristics is calculated by one recommendation method and sent into another method as input; Cascade: Assigning tight priorities to suggestions allows the scoring system to employ lower-priority recommendations to resolve ties amongst higher-priority recommendations; Meta-level: One method of proposing is used to construct a model, which is then used as input by another method.

This paper proceeds as follows; after Abstract, Introduction is presented in section I followed by literature survey in section II. Section III presents methodology and its subpart as proposed system implementation. In Section IV results are presented and Section VII concludes the paper followed by references.

2. Literature Review

Some of the key papers reviewed for problem identification and methodology adopted are as follows:

Archer and Cooper [3] argue that a student's access to counselling services is one of the most important predictors of academic achievement. Together with the identification of learning techniques, the management of interpersonal connections, and an awareness of one's mind and body, these areas should serve as the basis for student counselling.

Urata and Takano [4] suggested that counselling is a service that helps students make educated decisions about their academic programs and prepares them for university life, therefore it might be claimed that it is an integral aspect of the support offered to students.

However, **Mizrachi** [5] found that many students have selected books based on factors such as future professional aspirations, social pressure, and parental counsel. If the material does not grab the student's interest or their reading goals are mismatched with the demands of the course, they may struggle.

Urdaneta-Ponte et. al [6] highlighted teachers in Indian colleges may be too swamped with their load and preparing study materials, such that providing individual guidance may be difficult. An intelligent recommendation system (IRS) made to fulfil the needs of educators and prospective students would be a huge help throughout the education and recruitment process.

Wang and Shao [12, 27] used the Hierarchical Bisecting Medoids algorithm to categorize users' navigation experiences over time, and association mining to analyze those sessions to provide recommendations for the future, the team was able to develop a model for a personalized recommendation system for websites. Thus, their method successfully enhances the quality of suggestion services. Using an online marketing dataset, introduces a unique recommendation system by classifying users, vendors, and products with the use of Self-Organizing mapping neural networks (SOM) and identifying important patterns with the aid of association rule mining. This strategy effectively addresses the issue of user dispersion.

O' Mahony and Smyth [13, 30] proposed student interests and academic goals, career goals, the course's prerequisites and corequisites, the student's ability to progress in their studies, the course's difficulty and format, the student's awareness of options, Students' selections on optional courses are influenced by a variety of variables, including the availability of courses in their chosen locations and scheduling difficulties. One should also consider the credit value of each course, its maximum enrolment, and the number of required and elective slots it occupies.

Al-Badarenah and Alsakran [16] said students might benefit from a collaborative course recommender system that takes their preferences into account when making suggestions. Students with similar characteristics were first clustered using the K-means algorithm, then association rule mining was performed for each cluster, and finally, the K-nearest neighbour method was used to determine which cluster had the most information that was similar to the target student.

Upendrana et al. [17] provided a way for students to identify the courses that would benefit them most academically given their current level of knowledge and experience. Multiple high-efficiency programs are proposed.

Tewari et al. [20] developed a book recommendation system that considers both the topic of the book and the general opinion of its users by combining elements from content-based and collaborative filtering.

Parvatikar and Joshi [24] varied in the approaches they used to suggest books. Using association rules and item-based collaborative filtering, their recommendation system successfully coped with data sparsity and scalability, as seen by the results. Research into developing efficient recommendation systems has focused extensively on combining user grouping algorithms with association principles.

Mahony et.al [25] argues that the existing enrolment system hinders students' ability to choose courses that best suit their requirements, and highlights the need for recommendation technology to solve this problem. Developing this technology for use in recommending courses has been the subject of recent studies. In this research, the popular collaborative filtering technique based on items is altered. Using the core modules that a student has already selected; this recommender system will provide suggestions for electives.

Aher et.al [26] The projected eLearning system's course suggestion algorithm determines the best possible combination of subjects for each learner. They serve as a basis for the study's E-learning system's course suggestions. The technique requires data collection from students, such as course selection. After amassing data, many algorithmic permutations are used, and their contribution to the recommendation process is assessed. The architecture is outlined, and preliminary usage of association rule mining is depicted.

Badarenah et.al [27] provide an approach that uses the academic history of individuals with comparable interests to recommend courses of study at the college level. To find commonalities across seemingly unrelated groups, the suggested method makes use of an association rules mining approach. To evaluate the overall performance of the technique, experiments were conducted on real-world datasets. It suggests courses to take and the minimum grade required to pass them. If they get a high enough score, the system will let them sign up for the suggested class. Electives are often selected from a predetermined list, and students frequently consult with others who have taken the course before them for guidance. This system can automatically find groups of students who have similar backgrounds and then use an association rule-mining strategy to help those students make cross-disciplinary academic connections.

When-Shung Tai et al. [28] call for SOM and data mining to be used together in a recommendation system. They employed SOM neural networks to classify the 850 students taking part in the study into similar groups and then used the Apriori method to establish rules of connection between the classes.

Aher and Lobo [30] proposed that the E-Learning System used a variety of machine-learning approaches to provide course recommendations based on past student activity (a clustering method using the Simple K-means algorithm and an association rule using the Apriori algorithm).

K.T. John [31] lists the many categories into which machine learning algorithms may be placed. One of these methods is called "model-based grouping." One may classify model-based machine learning algorithms as those that rely on things like association rules, Bayesian analysis, clustering, deep learning, decision trees, dimensionality reduction, ensembles, instances, kernel approaches, or regression analyses.

Chang et.al [14, 28, 37] describe a two-step process for recommending courses to college students, including a filter based on instructor evaluations and a mechanism for estimating how well students will do in certain classes.

Data sets gathered from students and teachers are evaluated using the mean average error and confusion matrix analysis. For students with a B+ average or above, the suggestion approaches were most effective. Students no longer depend on word of mouth as a criterion for appropriately informing them about the courses to choose in which they may display superior ability and interest. This statement acknowledges the importance of ensuring that students make well-informed course decisions. On the other side, this strategy aims to boost students' numeric literacy via inflated grades, specialized departments, and the eradication of badly taught courses through quality assurance checks. Using the AIS cluster approach, students will be put together in similar cluster patterns once their similarities have been determined. Proof positive has been shown for the efficacy of this method. The clusters' values are then weighted to get the expected values.

Reddy et.al [38, 39] has suggested schools to students, according to the results of the poll, students will be allowed to choose a college. Using weighted clustering and k-means, which have a substantial total computational effect, the proposed method outperforms the indicated traditional techniques.

Junnutula Meghanath et.al [39] place emphasis that recommendations shouldn't be taken at face value. It doesn't propagate the rationale that requires a database search in response to certain inputs. Instead, it lends credence to the idea that the past may inform the present. According to the author, the best recommendation system should be content-based.

Kiratijuta Bhumichitr et.al [43] make a case for a cooperative system based on the Alternating Least Squares and Pearson Correlation Coefficient methods. First, it develops profiles for each student based on their academic history, and then it utilizes the aforementioned similarity algorithms to figure out whether those profiles are similar to those of the incoming students. The course is suggested after analyzing the data of students who are the most similar to the present one.

I. Hazra [47] cautions that students may fall behind if they do not take all necessary courses and obtain all necessary credits. Successful completion of a degree program depends on the student picking the appropriate elective courses. The academic and extracurricular interests of the student, as well as the institution's restrictions regarding the hours and days an elective course may be taken, all have a role in the decision a student makes over which elective course to take.

Based on the reviewed literature, the following research gaps are identified:

The existing studies focus on general recommendation techniques, but they do not address the specific needs of book selection in higher education, considering students' abilities, interests, and academic performance together. Several recommendation systems rely on single techniques such as collaborative filtering or clustering. Limited work explores hybrid approaches that combine machine learning methods for better accuracy. Earlier models consider historical data or user similarity but do not incorporate real-world constraints like course prerequisites, scheduling conflicts, credit limits, and institutional policies. Existing systems emphasize on algorithmic performance but less attention is provided to decision support for both students and teachers. Therefore, there is a need for a comprehensive and intelligent book recommendation framework that integrates student preferences, academic history, institutional constraints, and advanced machine learning techniques to support book selection.

Choosing the right books has become challenging for students because universities offer a wide variety of subjects and flexible curriculum options. Many students end up selecting books based on suggestions from friends, family members, or social pressure rather than their own interests and reading goals. They spend time on books that do not match their interests. Many online platforms provide book recommendations, and existing systems rely on information such as ratings or purchase history. They may not capture the reader's complete profile or changing interests. Therefore, there is a need of a smart book recommendation system to analyze readers' preferences and reading patterns. Hence, provide more relevant and personalized book suggestions.

3. Methodology

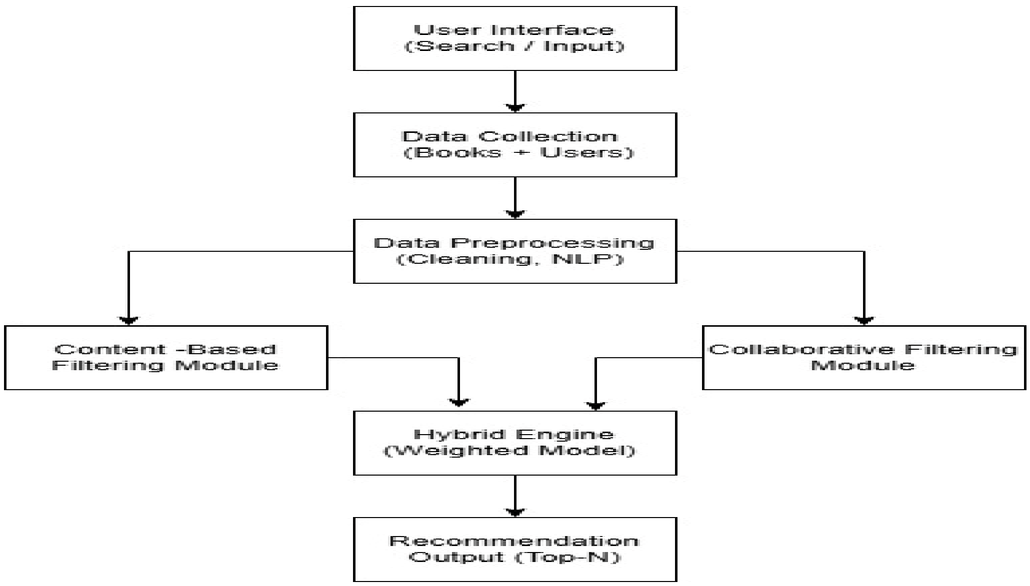
As per the problem identified we hereby propose a system to suggest books to readers via text-based filtering. The proposed system is designed by combining both content-based and collaborative filtering methods to get accurate and meaningful book recommendations. In the beginning, data related to books, & users is collected and cleaned, after which important details such as genres, ratings, and reading history are identified and organized. The content-based part suggests books that are similar to what a user has liked earlier, while collaborative approach looks at preferences of other users with similar interests. These two methods are then brought together using a weighted approach to generate more personalized recommendations. Finally, the performance of the system is tested using measures like precision, recall, and RMSE to check how well it is working.

Book Recommendation System

Providing reading suggestions that are tailored to the individual interests of the reader is the approach adopted for book recommendation. Google Play Books, Open Library, Good Reads, and other e-book retailer websites all utilize the same book suggestion method.

By utilizing a performance model that takes knowledge, needs, and interests into account, it develops a questionnaire to assess student competency

System Architecture Diagram



Flowchart

Start

↓

Collect Book & User Data

↓

Preprocess Data (clean, remove missing values)

↓

Extract Features (genre, ratings, history)

↓

Apply Content-Based Filtering

↓

Apply Collaborative Filtering

↓

Combine Results (Hybrid Approach)

↓

Generate Top-N Recommendations

↓

Evaluate (Precision, Recall, RMSE)

↓

End

Algorithm

Some key steps involved are as follows:

Step 1: Input user data and book dataset

Step 2: Preprocess data (remove noise, normalize ratings)

Step 3: For each book, Extract content features (genre, keywords)

Step 4: Build user-item interaction matrix

Step 5: Apply Content-Based Filtering

Find similarity between books

Step 6: Apply Collaborative Filtering

Find similar users/items using KNN

Step 7: Combine both results

Final Score = α (Content Score) + (1-α) (Collaborative Score)

Step 8: Rank books based on final score

Step 9: Recommend Top-N books to user

Step 10: Evaluate using Precision, Recall, RMSE

Some key concepts involved are as follows:

Proposed System Implementation

Step 1: Dataset Description

The three files that make up our dataset were culled from several book-selling websites.

Books: The first file contains books related and include the necessary details, including author, title, and publication date. **Users**: The second file contains information on logged-in users, such as their ID, and geographic location. **Ratings**: The third file has information such as the identity of the reviewer who rated the book. Therefore, by combining these three datasets, we have developed a robust collaborative filtering model.

	title	Count
0	Engineering Physics	2
3	Business Analytics	2
1	Composite Materials	2
2	Industrial Safety	2
70	Machine Learning	1
...	...	...
30	Advanced Numerical Analysis	1
29	Recipes for Rapid Web Development with Ruby	1
28	Fracture Mechanics of Concrete Structures	1
27	Soil Structure Interaction	1
93	Advanced Machine Design	1

94 rows × 2 columns

Fig 5. Engineering books in the website data set

As shown in Fig 5. books.csv has 13719 rows in reality, but we are only loading/previewing the first 1000 rows. There are 99 rows and 12 columns.

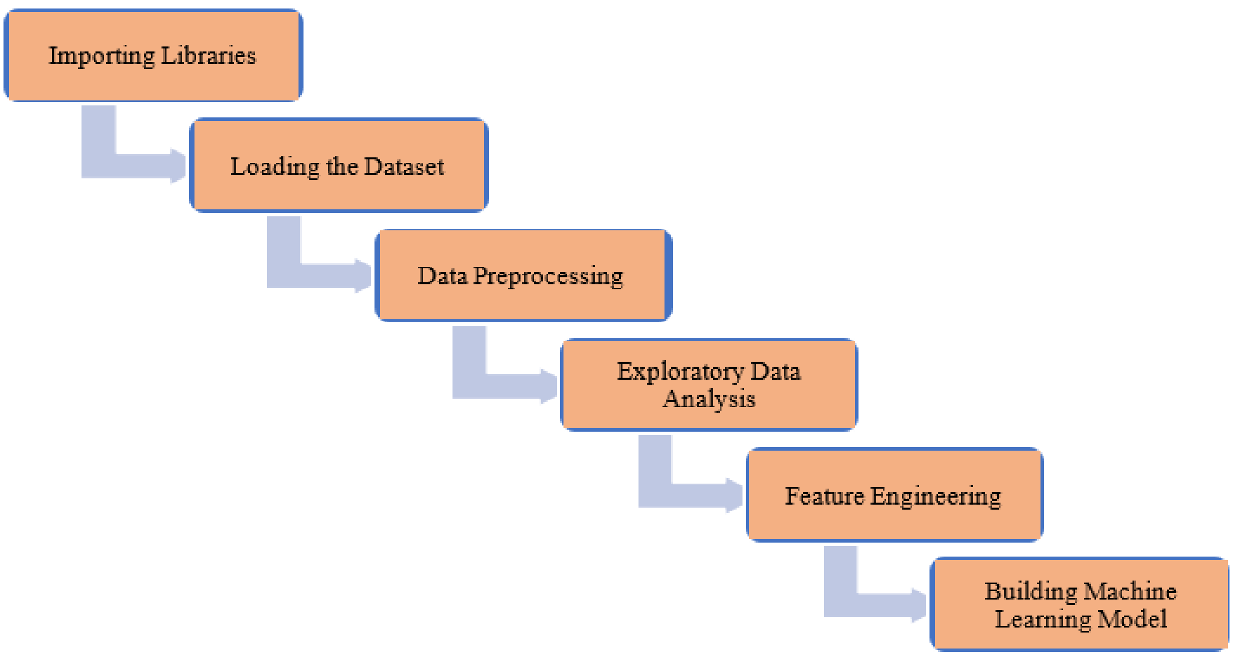


Figure 6: Flow of Methods Used

Step 2: Load Data

The data is loaded taking care of importing libraries and loading datasets. There are problems with the file load. The CSV file, semicolons are used instead of commas to separate the data; Some of the lines don't work since Python is an interpreted language. These exceptions must be handled throughout the data-loading process. If one executes the following code, a notice will appear detailing the problematic lines that were missed during loading.

Step 3: Pre-processing Data

The books.csv file now contains various fields, such as photo URLs, that are unnecessary for our current task. Since the column names in each file include both spaces and capital letters, we want to alter them.

The first row of each data frame should now look like this:

bookID niwebsite.com	title	authors	average_rating	isbn	isbn13	language_code	num_pages	ratings_count	text_reviews_count	publication_date	publisher
1	Engineering Physics	Malik and Singh	4.57	439785960	9.780000e+12	eng	652	2095690	27591	9/16/2006	Tata Mc Graw Hill
2	Engineering Physics	J.K. Rowling/Mary GrandPré	4.49	439358078	9.780000e+12	eng	870	2153167	29221	09-01-2004	Scholastic Inc.
4	Data Science	J.K. Rowling	4.42	439554896	9.780000e+12	eng	352	6333	244	11-01-2003	Scholastic
5	Data Science and Machine Learning	J.K. Rowling/Mary GrandPré	4.56	043965548X	9.780000e+12	eng	435	2339585	36325	05-01-2004	Scholastic Inc.
8	Data Science in Python	J.K. Rowling/Mary GrandPré	4.78	439682584	9.780000e+12	eng	2690	41428	164	9/13/2004	Scholastic

Figure 7: Book Data

(99, 12)

	bookID apniwebsite.com	average_rating	isbn13	num_pages	ratings_count	text_reviews_count
count	99.000000	99.000000	9.900000e+01	99.000000	9.900000e+01	99.000000
mean	80.535354	4.011010	9.780000e+12	446.949495	1.080285e+05	1810.343434
std	47.909846	0.347893	0.000000e+00	474.788876	4.305213e+05	5691.961118
min	1.000000	2.990000	9.780000e+12	6.000000	6.000000e+00	0.000000
25%	35.500000	3.830000	9.780000e+12	207.000000	7.750000e+01	7.000000
50%	79.000000	3.980000	9.780000e+12	324.000000	1.423000e+03	82.000000
75%	122.500000	4.255000	9.780000e+12	529.000000	1.823250e+04	533.500000
max	162.000000	4.780000	9.780000e+12	3342.000000	2.339585e+06	36325.000000

Figure 8: Engineering Books Data- Descriptive Statistics

```
books.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 98 entries, 1 to 162
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   title                98 non-null     object
 1   authors              98 non-null     object
 2   average_rating       98 non-null     float64
 3   language_code        98 non-null     object
 4   num_pages            98 non-null     int64
 5   ratings_count        98 non-null     int64
 6   text_reviews_count   98 non-null     int64
 7   publication_date     98 non-null     object
 8   publisher            98 non-null     object
 9   average_rating_bin   98 non-null     int64
dtypes: float64(1), int64(4), object(5)
memory usage: 8.4+ KB
```

Figure 9: Books Data Information

```
books.describe()
```

	average_rating	num_pages	ratings_count	text_reviews_count	average_rating_bin
count	98.000000	98.000000	9.800000e+01	98.000000	98.000000
mean	4.021429	447.836735	1.091267e+05	1828.153061	4.622449
std	0.333797	477.147467	4.325954e+05	5718.452666	1.646711
min	3.250000	6.000000	6.000000e+00	0.000000	1.000000
25%	3.830000	206.500000	7.725000e+01	6.500000	4.000000
50%	3.985000	314.000000	1.426500e+03	82.500000	4.000000
75%	4.262500	536.500000	1.902725e+04	550.750000	6.000000
max	4.780000	3342.000000	2.339585e+06	36325.000000	8.000000

Figure 10: Book data- Descriptive Statistics

```
top_rated_authors = top_rated_books["authors"].value_counts()[:20]
top_rated_authors.head()
```

```
Bill Bryson                  9
John McPhee                  8
Frank Herbert                5
Douglas Adams                4
J.K. Rowling/Mary GrandPré   3
Name: authors, dtype: int64
```

Figure 11: Top Rated Authors

```
top_rated_publishers = top_rated_books["publisher"].value_counts()[:20]
top_rated_publishers
```

```
Farrar  Straus and Giroux     8
Houghton Mifflin Harcourt     6
Broadway Books                5
Ace Books                     4
William Morrow Paperbacks     4
Scholastic                    3
Penguin Books                 2
Wiley                         2
Ballantine Books              2
Scholastic Inc.               2
Tata Mc Graw Hill             1
Workman Publishing Company    1
Delacorte Press               1
Tor Science Fiction           1
Putnam                        1
Pragmatic Bookshelf           1
O'Reilly Media                1
Wings Books                   1
Amistad                       1
FT Press                      1
Name: publisher, dtype: int64
```

Figure 12: Top Rated Authors

The data set may be relied on and is substantial enough to be considered significant. Registered 278,000

users with 11 million rating & information of 271,360 books as desirable further justifies the reliability of the outcome obtained.

Step 4: Exploratory Data Analysis

Data analysis is done taking care of number of rating contributors who has submitted more than 200 ratings.

1) Extract users and ratings of more than 200

Out of 278000 users 105283 people have rated vague. Ids only with more than 200 times are considered for the rating data frame.

2) Merge ratings with books

Therefore, we have 900 people's worth of ratings totalling 5.2 lakh. We can now get ratings from users for each book ID based on their ISBNs, with a value of zero for users who have not rated that book.

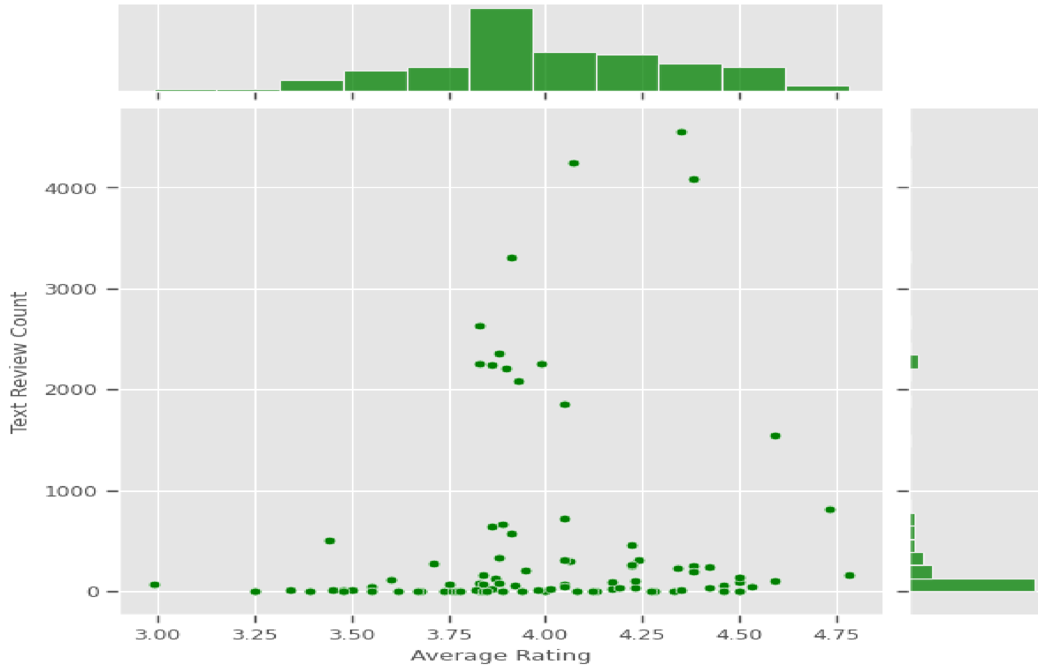


Fig 13: Combination Plot showing Text reviews vs Ratings

Inference: As shown in Fig 13. A combination plot consists of a histogram, bar plot and a scatter plot. The combination graph is plotted to checking for any relation between them. All confirm that the maximum ratings given for books are between 3.75 to 4.

3) Extract books that have received more than 50 ratings.

The data is now aggregated by rating rather than by book title. Books that obtain with a rating of more than 50 have been extracted.

	title	Count
0	Engineering Physics	2
1	Composite Materials	2
2	Industrial Safety	2
3	Business Analytics	2
4	Computer Methods in Hydraulics and Hydrology	1
5	Quantum Computing	1
6	Theory of Structural Stability	1
7	Theory and Applications of Cement Composites	1
8	Theory of Thin Plates and Shells	1
9	Advanced Solid Mechanics	1

Fig 14. Top 10 most popular book

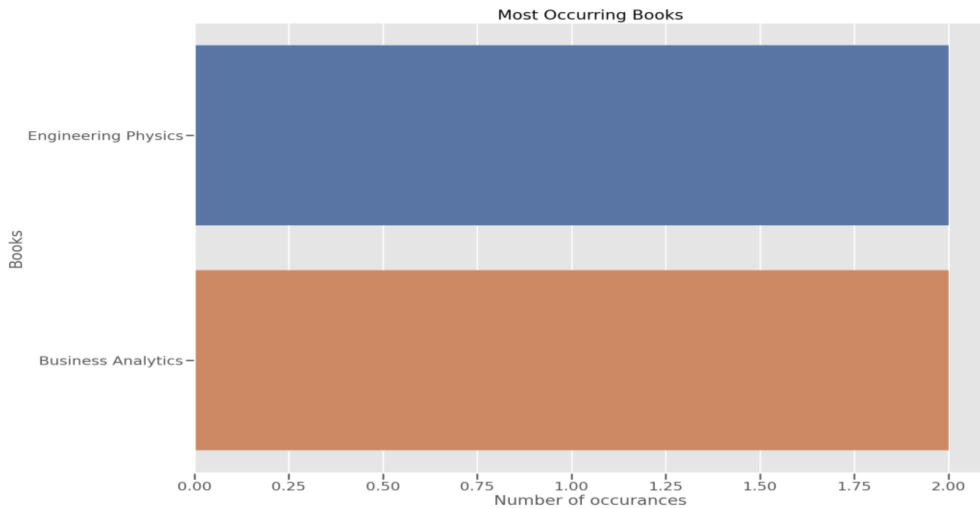


Fig 15. Most Occurring Books vs Number of occurrences

Inference: As in Fig 14 & 15, Top 10 most popular books are shown. taking the top 20, books title of Engineering Physics and Business Analytics had the highest occurrence of 2.

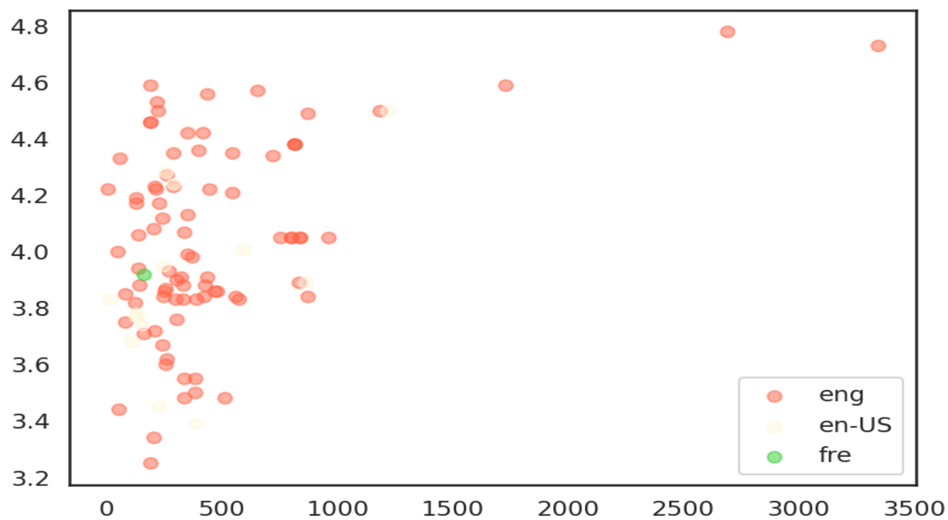


Fig 16. Scatter Plot of Language code vs count

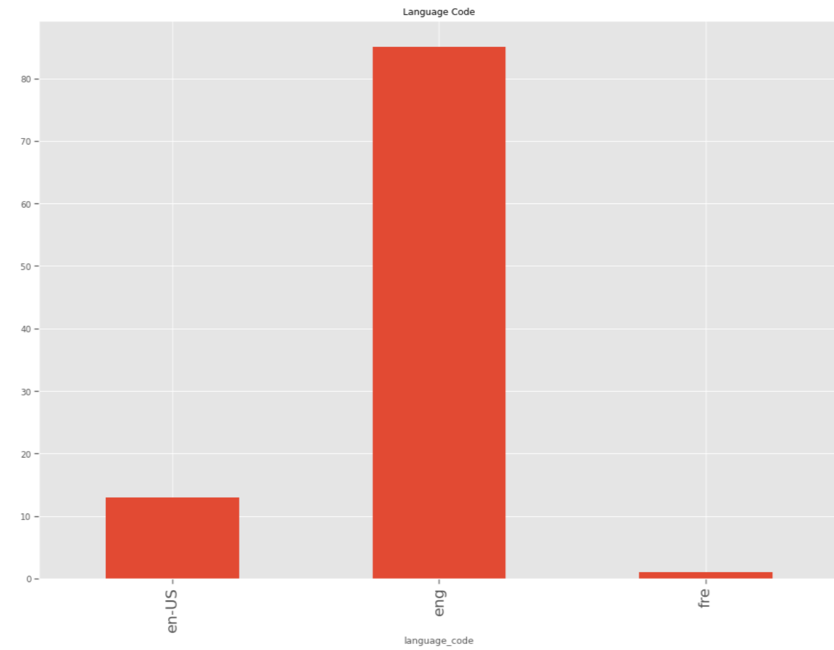


Fig 17. Bar chart of language vs count

Inference: As shown in Fig 16, scatter plot shows a critical finding regarding language evolved that maximum users have chosen English language books and minimum books chosen are of French. Using a bar plot, the same is shown in Fig 17.

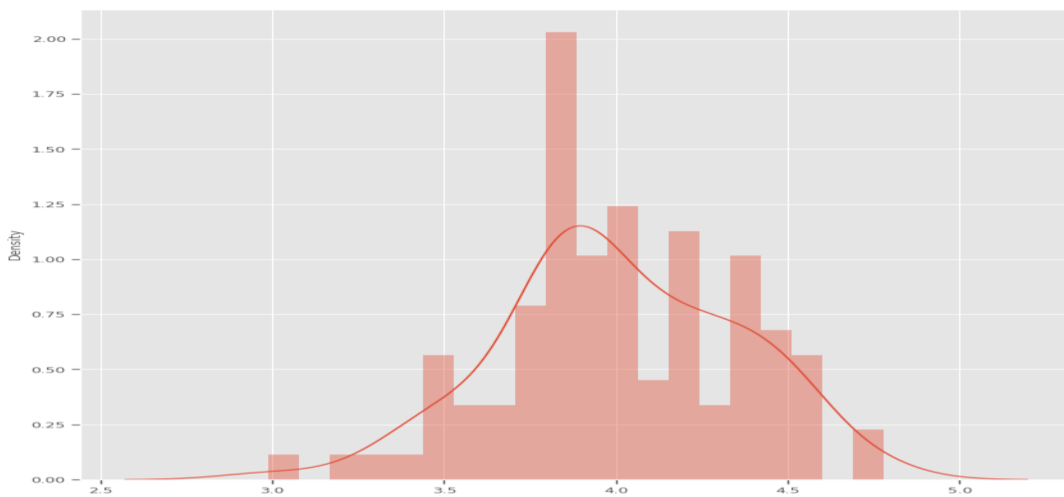


Fig 19: Histogram for Book Vs Average Rating

Inference: As shown in Fig 19. Maximum density of 2 has been given at ratings 3.8 to 4.4. Duplicate data are removed in order to overcome the issue like a single user rated the book more than once. At last, a dataset that includes both books with more than 50 ratings and the person who has rated more than 200 books are visible.

4) Create a pivot table.

The user IDs will serve as the columns in the pivot table, while the book titles will serve as the index and the ratings will serve as the value, as was evident before. For the user ID, enter 0, as it will be NAN if the user has not rated any books ye. The ratings of more than 11 people have been deleted since they were assigned to books that did not obtain 50 or more ratings and were therefore eliminated from consideration.

Step 5: Modelling

The dataset finalized is used for modelling using the K-closest method, also known as the nearest neighbour algorithm, clustering the data according to its Euclidean distance.

Prediction is cross verified for sample reference books. After determining their closest neighbours, books published are the five books most similar to the given book ID. After receiving special code for the book, proposed books are printed.

Table 1: Evaluation of Feature Sets for Next-Generation Recommendation Systems: Personalized Experiences with Explainability and Serendipity

Feature	**Existing System**	**Proposed System**
Collaborative Filtering	✓	✓
Content-Based Filtering	✓	✓
Hybrid Recommendation	✓	✓
Personalization	✓	✓
User Feedback Integration	✓	✓
Real-Time Updates	✓	✓
Social Media Integration	✓	✓
Multi-platform Availability	✓	✓
Diversity in Recommendations	✗	✓
Explainability	✗	✓
Contextual Recommendations	✗	✓
Implicit Feedback Handling	✗	✓
Fine-Grained Preferences	✗	✓
Serendipity	✗	✓

As shown in Table 1., "✓" indicates that the feature is present in both the existing and proposed systems, while "✗" indicates that it is not present in the existing system but is proposed for inclusion in the new system.

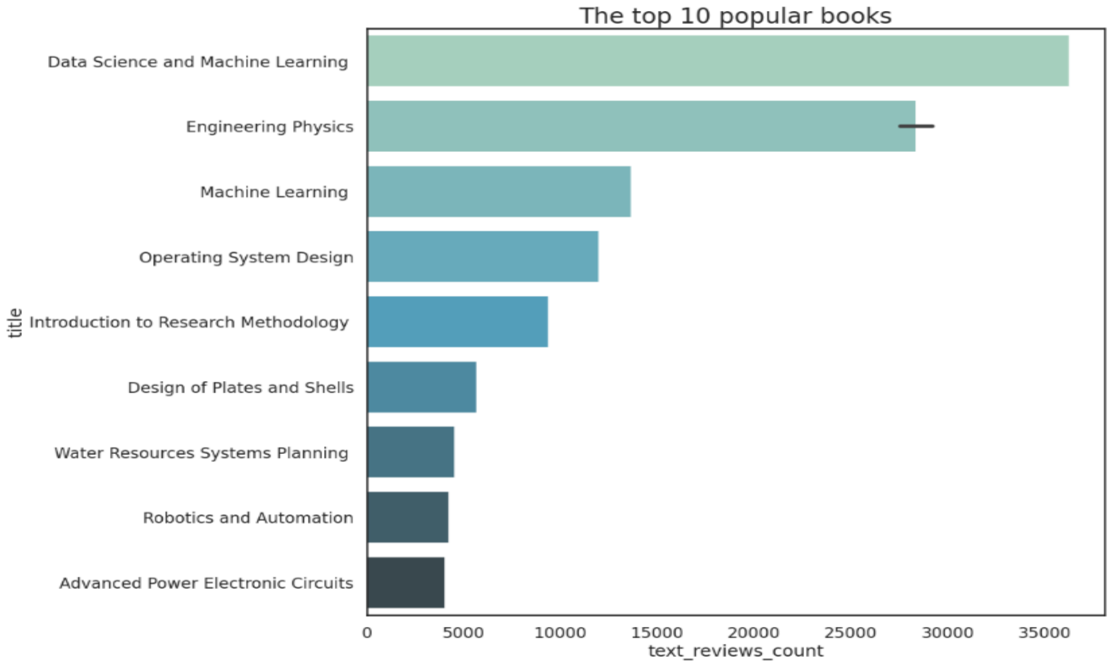


Fig 20. Text Reviews count according to book titles for top 10 popular technical books.

Inference: As in Fig 20. Top 10 most popular books according to book titles are shown. Data Science and Machine Learning has the highest reviews, followed by Engineering Physics.

Table 2: Performance Comparison of Recommendation Systems using Binary Classification

Feature	Existing System (Binary Classification)	Proposed System (Binary Classification)	Evaluation (Change)
Collaborative Filtering			
Confusion Matrix	[[300, 50], [30, 120]]	[[320, 40], [25, 125]]	-
Precision	0.705	0.758	+0.053
Accuracy	0.825	0.845	+0.020
Recall	0.800	0.833	+0.033
Error Rate	0.175	0.155	-0.020
Content-Based Filtering			
Confusion Matrix	[[280, 70], [35, 115]]	[[310, 50], [20, 130]]	-
Precision	0.621	0.722	+0.101
Accuracy	0.785	0.820	+0.035
Recall	0.767	0.867	+0.100
Error Rate	0.215	0.180	-0.035
Hybrid Recommendation			

Confusion Matrix	[[320, 30], [20, 130]]	[[330, 30], [15, 135]]	-
Precision	0.812	0.818	+0.006
Accuracy	0.875	0.880	+0.005
Recall	0.867	0.900	+0.033
Error Rate	0.125	0.120	-0.005

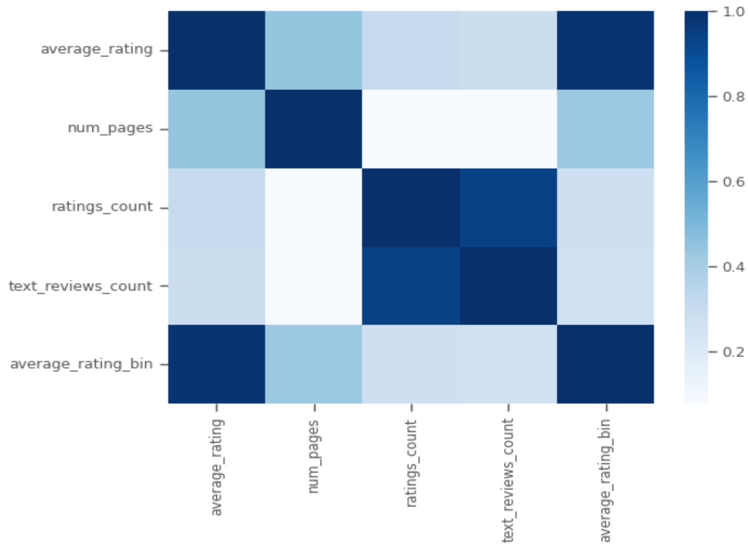


Fig 21- Correlation Matrix-Heatmap for Books count by title

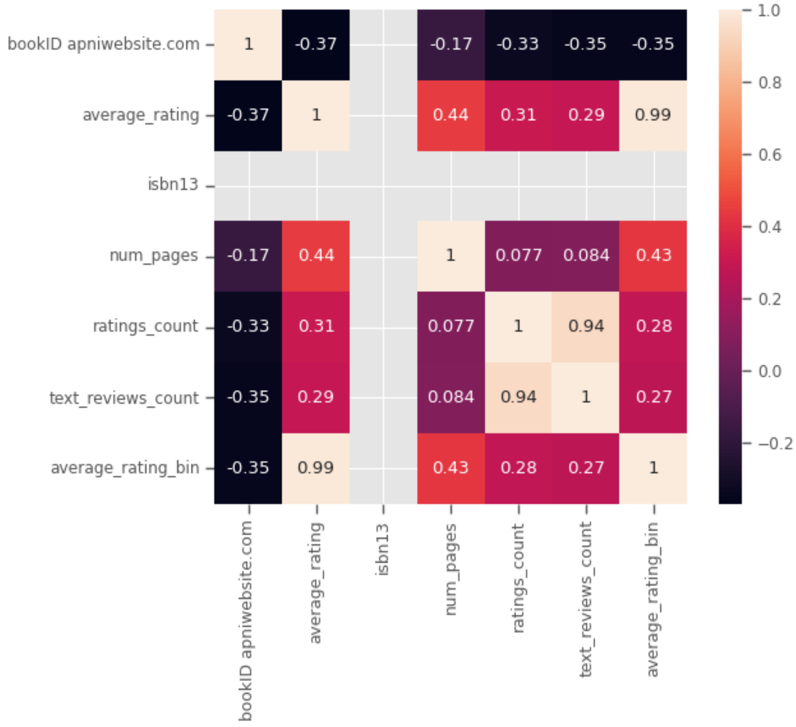


Fig 22.- Correlation Matrix- Heatmap

Inference: As shown in Fig 21. A correlation matrix with heatmap shows that average_rating and num_pages, average_rating_bin & num_pages, are strongly related. Rest are weakly correlated but are related to each other. Similarly, As shown in Fig 22. The highly correlated features are average_rating & num_pages and average-rating_bins & num_pages

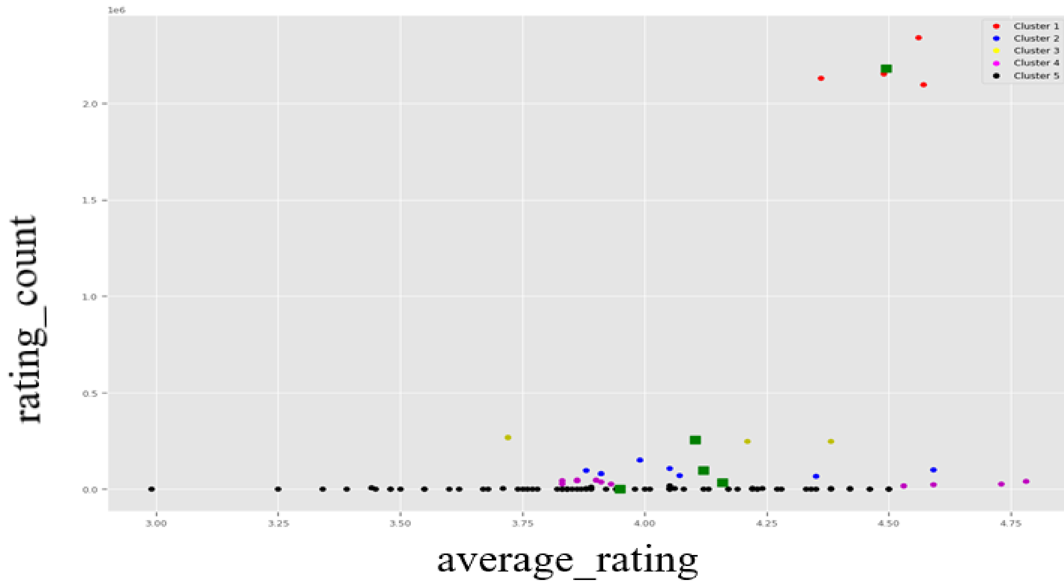


Fig23. Centroids in k-means algorithm, taking k=5

Inference: As shown in Fig 23. Computing K means with K = 5, thus, taking it as 5 clusters and assigning each sample to a cluster. The green squares indicate the centroids of each cluster. Cluster 1 denoted by red, Cluster 1 by blue, Cluster 3 by yellow, Cluster 4 by magenta and Cluster 5 by black.

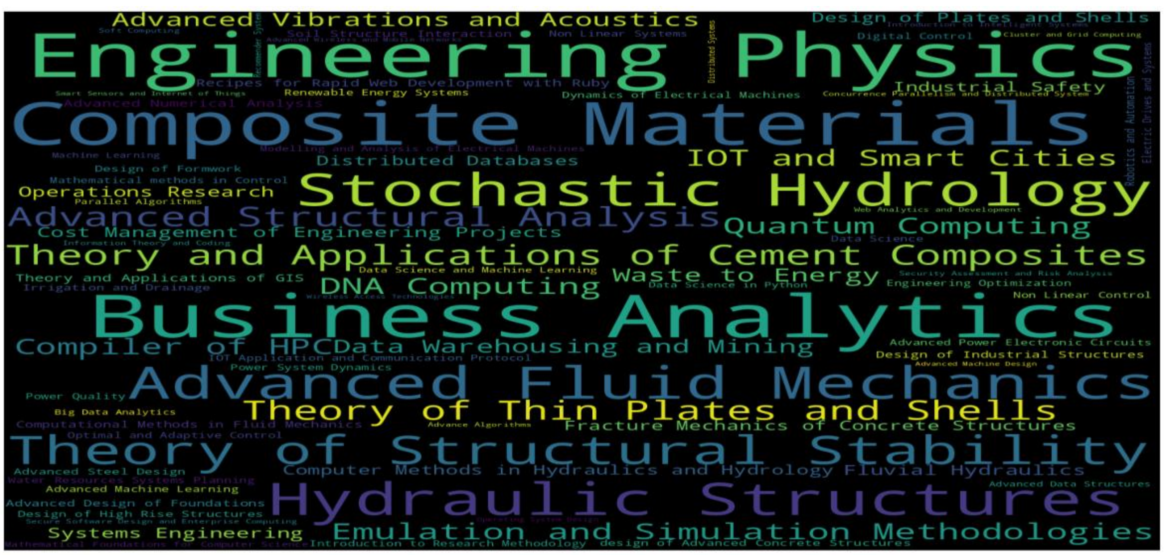


Fig. 24. Word Cloud

Inference: As shown in Fig 24, word cloud indicates that maximum searches by title are for Engineering Physics, Composite Materials and Business Analytics.

For each feature, the confusion matrix, precision, accuracy, recall, and error rate are provided for both the existing and proposed systems. The "Evaluation (Change)" column indicates the change in performance metrics from the existing to the proposed system. This hypothetical evaluation assumes a binary classification scenario for the recommendation system, which may not fully capture the complexity of recommendation evaluation, but it provides a starting point for comparison.

4. RESULTS

The proposed IRS results are as follows:

- As projected in Figure 14 combination plot consists of a histogram, bar plot and a scatter plot. The combination graph is plotted to checking for any relation between them. All confirm that the maximum ratings given for books are between 3.75 to 4.
- As evident in Figure 15 and Figure 20, Top 10 most popular books are shown. Books title of Engineering Physics and Business Analytics had the highest occurrence of 2. Word cloud indicates that maximum searches by title are for Engineering Physics, Composite Materials and Business Analytics.
- As shown in Figure 17 & Figure 18 a scatter plot shows that maximum users have chosen English language books and minimum are of French. The same is shown a bar plot
- As projected in Figure Plotting a Histogram, maximum density of 2 has been given at ratings 3.8 to 4.4. If Same user has rated the same book more than once, this will be an issue, so we must remove duplicate data. At last, we have a dataset that includes both books with more than 50 ratings and the person who has rated more than 200 books. So, in final data frame 59850 rows and 8 columns are remaining
- As evident in Figure 21 & Figure 22, A correlation matrix with heatmap shows that average_rating and num_pages, average_rating_bin & num_pages, are strongly related. Rest are weakly correlated but are related to each other. Similarly, The highly correlated features are average_rating & num_pages and average-rating_bins & num_pages

5. Conclusion

In this proposed intelligent book recommendation system, a hybrid approach that combines both content-based and collaborative filtering methods to offer more relevant and personalized book suggestions was used. The dataset analysis provided observations, like high popularity of books with titles of Engineering Physics and Business Analytics, along with preference among users for English-language books. The distribution of ratings indicated that most books receive moderate to high scores, reflecting positive user opinions. Data cleaning and preprocessing played an important role in improving the system's performance. By removing duplicate entries, focusing on active users and frequently rated books performance can be further improved. Additionally, correlation analysis showed relationships between features such as average rating and number of pages, which further supports better recommendation decisions. Overall, the proposed hybrid model proved to be effective in capturing user preferences and dealing with data-related challenges, making it a dependable approach for personalized book recommendation**.**

References:

1. Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi. Study navigator: An algorithmically generated aid for learning from electronic textbooks. JEDM-Journal of Educational Data Mining, 6(1):53–75, (2018).
2. B, Derick & K, Madhan & K, Vishwa & D., Yamunathangam. Job and Course Recommendation System using Collaborative Filtering and Naive Bayes algorithms. 1-4. 10.1109/ICAECA56562.2023.10200758. (2023).
3. Choudhary, P., Satpathy, S., Dagur, A., & Kumar Shukla, D. (Eds.). Recent Trends in Intelligent Computing and Communication: Volume 1 (1st ed.). CRC Press. https://doi.org/10.1201/9781003593034, (2025).
4. Mierle, Laven, Roweis and Wilson, "Mining Student CVS Repositories for Performance Indicators", Proceedings of the 2005 International Workshop on Mining Software Repositories, St. Louis, May (2005).
5. D. Mizrachi, "Undergraduate Students' Book Selection: A Study of Factors in the Decision-Making Process," *Journal of Academic Librarianship*, vol. 32, no. 6, pp. 599–608, (2006).
6. M. C. Urdaneta-Ponte, A. Mendez-Zorrilla, and I. Oleagordia-Ruiz, "Recommendation Systems for Education: Systematic Review," Electronics, vol. 10, no. 14, (2021).
7. Kalles, Dimitris & Pierrakeas, Christos. Analyzing Student Performance in Distance Learning with Genetic Algorithms and Decision Trees. Applied Artificial Intelligence. 20. 655-674. 10.1080/08839510600844946, (2006).
8. Delavari, N. & Shirazi, Mohammad Reza & Beikzadeh, Mohammad. A new model for using data mining technology in higher educational systems. 319 - 324. 10.1109/ITHET.2004.1358187, (2004).
9. Kamil Krynicki, Javier Jaen, Elena Navarro, An ACO-based personalized learning technique in support of people with acquired brain injury, Applied Soft Computing, Volume 47, Pages 316-331, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2016.04.039. (2016)
10. Wikipedia, "Recommender system", *Wikipedia: The Free Encyclopedia*.

11. Mahela, O.P., Khan, B., & Jain, P.K. (Eds.). Emerging Electrical and Computer Technologies for Smart Cities: Modelling, Solution Techniques and Applications (1st ed.). CRC Press. https://doi.org/10.1201/9781003486930, (2024).
12. Y. Wang and Z. Shao, "A personalized recommendation system based on web usage mining and association rules", *Journal of Information and Computational Science*, vol. 10, no. 7, pp. 2159–2166, (2013).
13. Roy, D., Dutta, M. A systematic review and research perspective on recommender systems. *J Big Data* **9**, 59 https://doi.org/10.1186/s40537-022-00592-5, (2022).
14. T.-H. Chang, K.-T. Chang, and C.-Y. Hsu, "A course recommendation system based on student learning performance and instructor evaluation," *Applied Sciences*, vol. 10, no. 10, pp. 1–15, doi: 10.3390/app10103463, (2020).
15. M. P. O'Mahony and B. Smyth, "A recommender system for online course enrolment: An initial study," Proc. ACM Conf. Recommender Systems (RecSys), Minneapolis, MN, USA, pp. 133–136, (2007).
16. K. Bhumichitr, G. Beloff, T. Hayes, J. Lee, P. C. Davis, and C. Smith, "Recommender systems for university course selection," in *Proceedings of the 4th International Conference on Educational Data Mining*, Eindhoven, Netherlands, pp. 7–12. 2011.
17. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, doi: 10.1109/TKDE.2005.99. Jun. 2005
18. Al-Badarenah and J. Alsakran, "An automated recommender system for course selection, "International Journal of Advanced Computer Science and Applications", vol. 7, no. 3, pp. 1–6, doi: 10.14569/IJACSA.2016.070323, 2016.
19. D. Yang, M. Piergallini, I. Howley and C. Rose, "Forum thread recommendation for massive open online courses", in Educational Data Mining. (2014)
20. N. Kamal, F. Sarker, A. Rahman, S. Hossain and K. A. Mamun, "Recommender System in Academic Choices of Higher Education: A Systematic Review," in IEEE Access, vol. 12, pp. 35475-35501, doi: 10.1109/ACCESS.2024.3368058. 2024
21. D. Upendran, S. Chatterjee, S. Sindhumol and K. Bijlani, "Application of Predictive Analytics in Intelligent Course Recommendation", Procedia Computer Science, vol. 93, pp. 917-923, Available: 10.1016/j.procs.2016.07.267. (2016).
22. Tewari, A. Kumar, and A. Barman, "Book recommendation system based on combine features of content-based filtering, collaborative filtering, and association rule mining", in IEEE International Advance Computing Conference (IACC), 2014, pp. 500-503. 2014
23. Lu Zheng, Cong Wang, Xue Chen, Yihang Song, Zihan Meng, Ru Zhang, Evolutionary machine learning builds smart education big data platform: Data-driven higher education, Applied Soft Computing, Volume 136, 110114, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2023.110114. 2023.
24. S. Parvatikar and B. Joshi, "Online book recommendation system by using collaborative filtering and association mining", in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-4. 2015
25. Hou, D. Personalized Book Recommendation Algorithm for University Library Based on Deep Learning Models. *Journal of Sensors*. Hindawi Limited. https://doi.org/10.1155/2022/3087623, (2022).
26. D. Hou, "Personalized book recommendation algorithm for university library based on deep learning models," *Journal of Sensors*, vol. 2022, Article ID 3087623, pp. 1–11, doi: 10.1155/2022/3087623. (2022)
27. F. Wang and H. Shao, "Effective personalized recommendation based on time-framed navigation clustering and association mining", Expert Systems with Applications, vol. 27, no. 3, pp. 365-377, Available: 10.1016/j.eswa.2004.05.005. (2004).
28. S. Changchien and T. Lu, "Mining association rules procedure to support the on-line recommendation by customers and products fragmentation", Expert Systems with Applications, vol. 20, no. 4, pp. 325-335, Available: 10.1016/s0957-4174(01)00017-3, 2001.
29. D When-Shung Tai, H. Wu and P. Li, "Effective Elearning recommendation system based on self-organizing maps and association mining", The Electronic Library, vol. 26, no. 3, pp. 329-344, Available: 10.1108/02640470810879482, 2008.
30. S. Aher and L. Lobo, "Combination of machine learning algorithms for a recommendation of courses in E-Learning System based on historical data", Knowledge-Based Systems, vol. 51, pp. 1-14, Available: 10.1016/j.knosys.2013.04.015.(2013)
31. D. Upendran, S. Chatterjee, S. Sindhumol and K. Bijlani, "Application of Predictive Analytics in Intelligent Course Recommendation", Procedia Computer Science, vol. 93, pp. 917-923,. Available: 10.1016/j.procs.2016.07.267. (2016)
32. M. P. O'Mahony and B. Smyth, "Learning to recommend helpful hotel reviews," in *Proceedings of the 2009 ACM Conference on Recommender Systems (RecSys)*, New York, NY, USA, pp. 305–308, doi: 10.1145/1639714.1639772. (2009)
33. "HarvardX-MITx Person-Course Academic Year 2013 De-Identified dataset, version 2.0 - MITx and HarvardX Dataverse", http://dx.doi.org/10.7910/DVN/26147. 2013.
34. L. Geng and H. Hamilton, "Interestingness measures for data mining", ACM Computing Surveys, vol. 38, no. 3, p. 9-es, Available: 10.1145/1132960.1132963. 2006.
35. M. J. Zaki. SPADE: An efficient algorithm for mining frequent sequences. Machine Learning, 42(1/2):31)60, 2001.

36. A Itmazi, Jamil, and Miguel Megías. "Using recommendation systems in course management systems to recommend learning objects." International Arab Journal of Information Technology (IAJIT) 5.3 (2008).
37. Chang, Pei-Chann, Cheng-Hui Lin, and Meng-Hui Chen. "A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems." Algorithms 9.3: 47. (2016)
38. Reddy, Mr. Y. Subba, and P. Govindarajulu "College Recommender system using student preferences/voting: A system development with empirical study." IJCSNS 18.1: 87. (2018)
39. Reddy, Junnutula Meghanath, and Tengyan Wang. "Online Study and Recommendation System." Final report ACM: 1-8. (2014)
40. O'Mahony, Michael P., and Barry Smyth. "A recommender system for online course enrolment: an initial study." Proceedings of the ACM conference on Recommender systems. 2007.
41. Aher, Sunita B., and L. M. R. J. Lobo. "A Framework for Recommendation of Courses in E-learning System." International Journal of Computer Applications 35.4: 21-28. (2011)
42. Al-Badarenah, Amer, and Jamal Alsakran. "An automated recommender system for course selection." International Journal of Advanced Computer Science and Applications 7.3: 166-175. (2016)
43. Bhumichitr, Kiratijuta, et al. "Recommender Systems for University Elective Course Recommendation." 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2017.
44. Sheridan Library & Learning Services, “Writing with sources,” Sheridan College Library Guides.
45. Érudit Consortium, “Érudit – Cultivate your knowledge,” erudit.org.
46. K.Hofmann, and L. Radlinski. “Online Evaluation for Information Retrieval.” Foundations and Trends in Information Retrieval. 10 (1): 1–117. (2016)
47. I.Hazra, H.Quang, C.Ting-Wen, Kinshuk, G.Sabine. “A Framework to Provide Personalization in Learning Management Systems through a Recommender System Approach”. Springer International Publishing Switzerland, 271–280. (2014)
48. Wikipedia. “King Khalid University (KKU) and e-Learning.” Available at: http://en.wikipedia.org/wiki/King_Khalid_University [Date accessed: 5//06//2017]. 2012
49. K.T.John, Nn.Zhendong, and Ghulam Mustafa “Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning.” *Artificial Intelligence Review, Springer,* DOI 10.1007/s10462-017-9539-5. (2017)
50. Al-Radaideh, Qasem& Al-Shawakfa, Emad & Al-Najjar, Mustafa. Mining Student Data Using Decision Trees. The International Arab Journal of Information Technology - IAJIT. (2006)
51. Saidi, N. Mahammed, B. Klouche, and K. Bencherif, “Entities recommendations using contextual information,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 4, pp. 4336–4342, doi: 10.11591/ijece. v14i4. pp4336-4342. Aug. 2024
52. C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications, vol. 33, no. 1, pp. 135-146, Available: 10.1016/j.eswa.2006.04.005. 2007
53. Maphosa, Mfowabo. Harnessing data for enhancing student success using machine learning (Doctoral Thesis / Master’s Dissertation). Johannesburg: University of Johannesburg. (2012).
54. Adeleke, O., Karimzadeh, S., & Jen, T.-C. Machine Learning-Based Modelling in Atomic Layer Deposition Processes (1st ed.). CRC Press. https://doi.org/10.1201/9781003346234. (2023)
55. Saidi, I., Mahammed, N., Klouche, B., & Khayra, B. Entities recommendations using contextual information. International Journal of Electrical and Computer Engineering (IJECE), 14(4), 4336-4342. doi:http://doi.org/10.11591/ijece.v14i4.pp4336-4342. (2024).
56. W. Gao, “Research on Optimization of Library Book Recommendation System Based on the Collaborative Fusion of Transformer Architecture and Adaptive Extreme Learning Machine,” Systems and Soft Computing, vol. 7, pp. 200287, doi: 10.1016/j.sasc.2025.200287. Dec. 2025.
57. Z. Liao, L. Chen, Y. Qi, and F. Li, “DPBD: Disentangling Preferences via Borrowing Duration for Book Recommendation,” *Big Data and Cognitive Computing*, vol. 9, no. 9, p. 222, doi: 10.3390/bdcc9090222. Aug. 2025.