



**Module Code & Module Title**  
**CU6051NI Artificial Intelligence**

**Assessment Weightage & Type**  
**25% Coursework 1**

**Semester**  
**2020-21 Autumn**

**Student Name: Mimona Karki**

**London Met ID: 19031154**

**College ID: np01cp4a190214**

**Assignment Due Date: 22<sup>nd</sup> December 2021**

**Assignment Submission Date: 22<sup>nd</sup> December 2021**

*I confirm that I understand my coursework needs to be submitted online via Google Classroom under the relevant module page before the deadline in order for my assignment to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a mark of zero will be awarded.*

## Table of contents

1. Introduction .....	1
1.1. Recommendation systems .....	1
1.2. The flourishing market of e-commerce book sites .....	2
1.3. Book recommendation systems and problem domain .....	3
1.4. The role of Artificial intelligence (AI) in book recommendation systems .....	4
2. Background .....	5
2.1. A brief history on book recommendation systems .....	5
2.2. Existing works on book recommendation systems .....	6
2.3. Pros and Cons of book recommendation systems .....	11
3. Solution .....	13
3.1. Collaborative filtering (CF) .....	13
3.2. Nearest Neighbour algorithm .....	15
3.3. Distance metrics .....	17
3.3.1. Euclidean distance .....	17
3.3.2. Cosine similarity .....	17
3.3.3. Pearson coefficient .....	18
3.4. Flowchart of the proposed algorithm .....	19
3.5. System overview diagram .....	20
4. Conclusion .....	21
5. References .....	22

## Table of figures

Figure 1: Overview of a recommendation system, (Agrawal, 2021) .....	1
Figure 2: The US online book services market size, (Region, And Segment Forecasts, 2020-2027).....	2
Figure 3: Overview of content-based and collaborative filtering, (Das, 2020) .....	4
Figure 4: System Overview, (Nursultan Kurmashov, 2015).....	6
Figure 5: Experiment results, (Nursultan Kurmashov, 2015).....	7
Figure 6: System architecture, (Prof. Deepali Patil, 2015).....	8
Figure 7: E-reader Whichbook, (Whichbook, 2021) .....	9
Figure 8: Online bookstore BookBub, (BookBub, 2021).....	9
Figure 9: E-reader Goodreads, (Goodreads, 2021) .....	10
Figure 10: Example of collaborative filtering, (Grover, 2017) .....	13
Figure 11: Formula for calculating the average rating .....	15
Figure 12: Cosine similarity, (Grover, 2017).....	17
Figure 13: Pearson correlation coefficient, (Thakur, 2021).....	18
Figure 14: Flowchart of the proposed algorithm .....	19
Figure 15: System overview diagram .....	20

## 1. Introduction

### 1.1. Recommendation systems

A recommendation system is a subcategory of information filtering systems that is based on finding out the preference of a user by predicting their ratings for an item. In basic terms, recommendation systems make use of an algorithm that suggests relevant items to the users (Agrawal, 2021). For instance, e-commerce websites such as Amazon, eBay, Walmart, and streaming service providers such as Netflix, Hulu, and Amazon Prime Video are some of the popular examples of services using recommendation systems to recommend products/items to a user. Moreover, the recommendations could be based on various parameters, such as past buying records of top customers, most popular items on the company's website, user's geographical location and demographic information, and so on. As these systems help in generating a huge amount of income in business industries, recommendation system has become unavoidable in today's cyberspace. (Rocca, 2019).

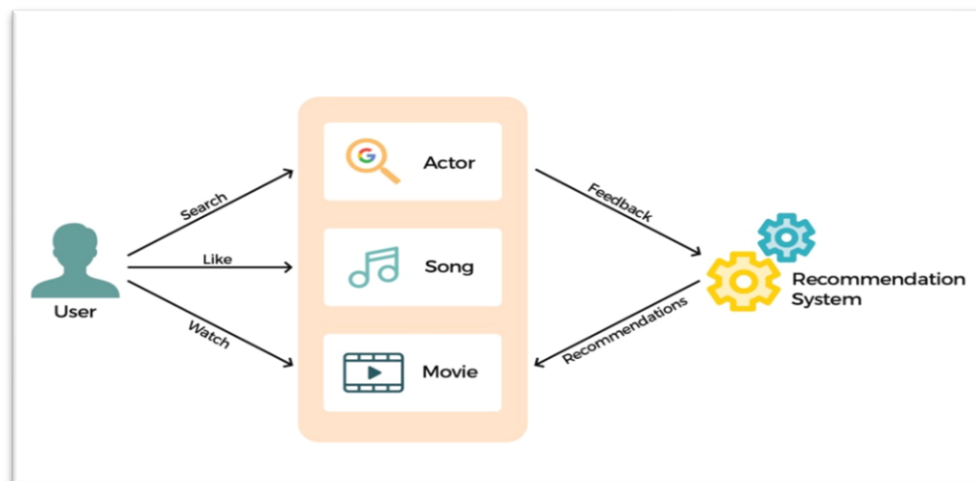


Figure 1: Overview of a recommendation system, (Agrawal, 2021)

The following research report provides a full conspectus of a Book Recommendation System which has emerged as a major research interest that is highly used in the field of e-commerce. The report focuses on the different paradigms of recommendation system, problem domain, advantages/disadvantages, review and analysis of the existing works, the solution to solving the stated problems using ML algorithms, and diagrammatic representations to make it more detailed.

## 1.2. The flourishing market of e-commerce book sites

In this modern world, technological advancement has given rise to a large number of book-selling online platforms due to the rising accessibility of books in digital format. E-book adoption is increasing enormously, and one of the contributing factors is the covid-19 pandemic. According to the good e-reader's blog, over 2 million new users have registered with the e-reader app Kobo during the pandemic (Kozlowski, 2020). Moreover, bibliophiles find the e-book option less expensive than the hardcover or paperback version. Additionally, online books can be downloaded easily and is also an eco-friendly option as it reduces paper consumption. As per the report by Online Book Services Market Size, Share & Trends Analysis, "the global online book services market size was valued at USD 17.7 billion in 2019 and is expected to grow at a compound annual growth rate (CAGR) of 5.8% from 2020 to 2027". (Region, And Segment Forecasts, 2020-2027)

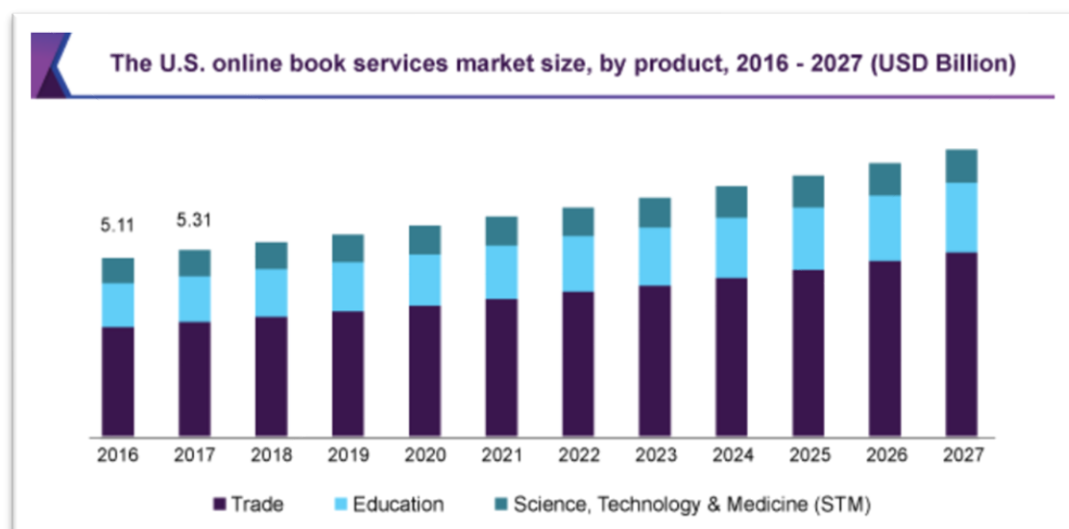


Figure 2: The US online book services market size, (Region, And Segment Forecasts, 2020-2027)

### **1.3. Book recommendation systems and problem domain**

The growing online book markets are found competing against each other based on different attributes, and a book recommendation system is one of them. A book recommendation system is designed to suggest books of interest to the customer. The major goal of developing such systems is to analyze the buyer's interest and recommend books to them suitably. Nowadays, it has become difficult to find relevant books in a vast e-book space due to the exponential growth of online books. Moreover, there are varieties of genres and different kinds of books are being published every day. As a result, the buyers might not know which book to pick from among thousands of books displayed on an e-commerce site. Also, the booksellers are unknown about the books that have the highest demand in the market. It would be great if such businesses utilize book recommendation engines to display books according to the reader's preferences and pattern of searching books. It would clear up the dilemma of the customers and at the same time help in increasing the business revenue rate. As such, both the seller and buyer would be benefitted. Considering the huge market of online book sites and the problem scenario a book recommendation system has been chosen as the project topic.

#### 1.4. The role of Artificial intelligence (AI) in book recommendation systems

The online book recommendation system applies numerous AI techniques in order to generate effective suggestions for buyers. The most popular techniques include Collaborative filtering, content filtering, and association mining. Collaborative filtering explores the user's opinions and provides recommendations based on the ratings provided to a book. The content filtering technique involves providing recommendations based on the past search records of a buyer. Finally, the association mining approach involves the mining of association and correlation relationships for determining the best outcome. (ProjectsGeek, 2018). Fig.3 below demonstrates the content-based filtering and collaborative filtering strategies.

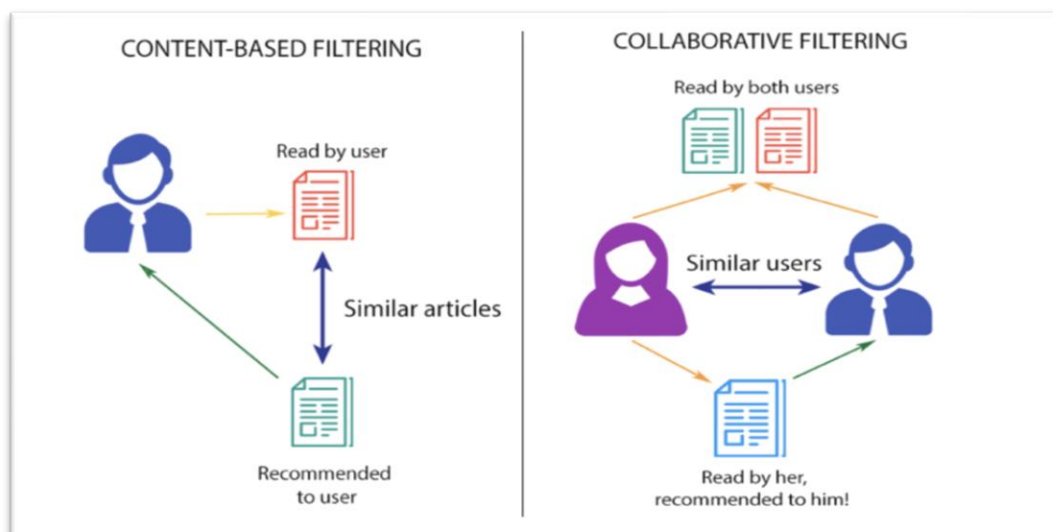


Figure 3: Overview of content-based and collaborative filtering, (Das, 2020)

## **2. Background**

### **2.1. A brief history on book recommendation systems**

An article by Sandor Apathy at Onespire unfolds the history of book recommendation systems where it is found that the concept of recommendation system was not out until the 20<sup>th</sup> century when Claude Shannon realized that the information can be transformed into numbers i.e. 0 or 1. Then, in the 1970s the first Usenet communication system was created at Duke University where the users were able to share textual contents which were classified as newsgroups and subgroups. But it was not based on user preferences precisely. Then, the first solution for recommending books to the reader evolved in 1979 by a computer librarian Grundy. The working mechanism of this system was simple where Grundy interviewed the readers about their preferences and recommended books to them according to the information. Specifically, the system classified the users into different groups based on the information and recommended books to them according to their age groups (Apáthy, 2019-2021). Thus, the history of the book recommendation system goes a long time back in the 20<sup>th</sup> century and continues to evolve.



## 2.2. Existing works on book recommendation systems

Referring to a conference paper published at Suleyman Demirel University, Faculty of Engineering and Natural Sciences, it has been found that an online book recommendation system has been developed which learns user preferences by asking the consumers to rate books and choose favorite categories. Then, the system generates a list of books that the user might like to read. More specifically, a user gets registered into the system and is asked to choose 10 favorite books from at least 3 genres. According to this information, the system recommends books to the users. As the system focuses on simplicity and speed, the collaborative filtering method is used to make recommendations (Nursultan Kurmashov, 2015). Fig.4 illustrates the developed system overview.

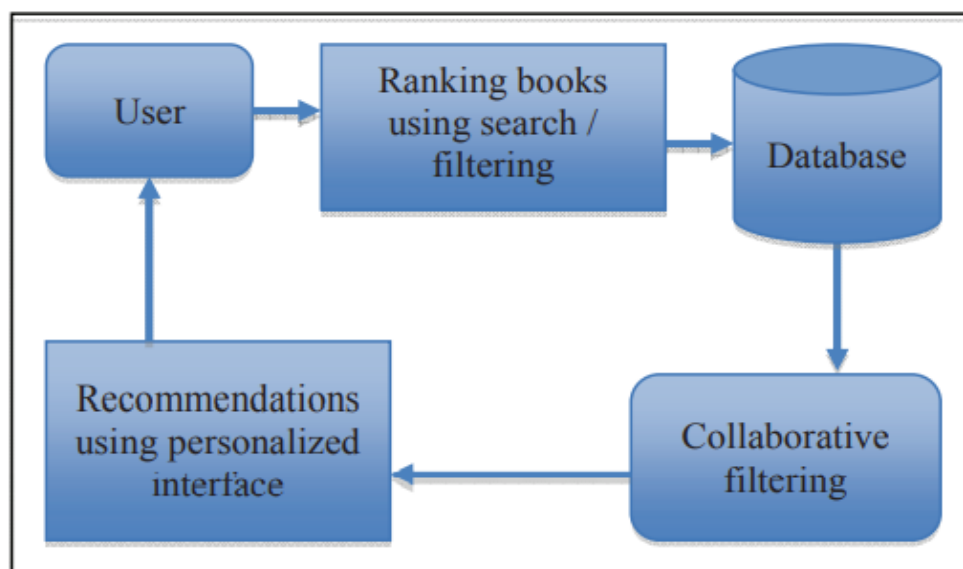


Figure 4: System Overview, (Nursultan Kurmashov, 2015)

As per the experiments conducted in the research paper, it can be seen that the users are satisfied with the convenience and qualitative book recommendations given by the system. The graph below presents the outcomes of the experiment. (Nursultan Kurmashov, 2015)

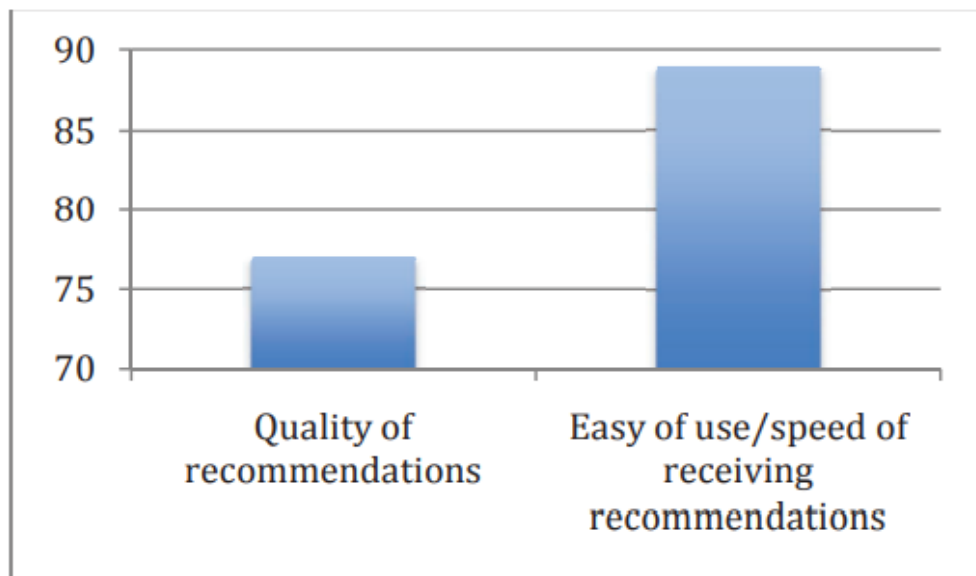


Figure 5:Experiment results, (Nursultan Kurmashov, 2015)

Another Journal article published by the International Journal of Advanced Research in Computer and Communication Engineering demonstrates a book recommendation engine in which three techniques are used for recommendation. The working mechanism of the proposed system is very simple where users log in to the system, search, and rate the books. The system stores the searching and ratings of the user and utilizes them for further operations. Based on the user's history, the collaborative and content-based strategy is used. If null results are generated from these methods, then the demographic recommendation is used. The results from all these strategies are combined and books are recommended to the user (Prof. Deepali Patil, 2015). Fig.6 presents the system architecture of the developed book recommendation system.

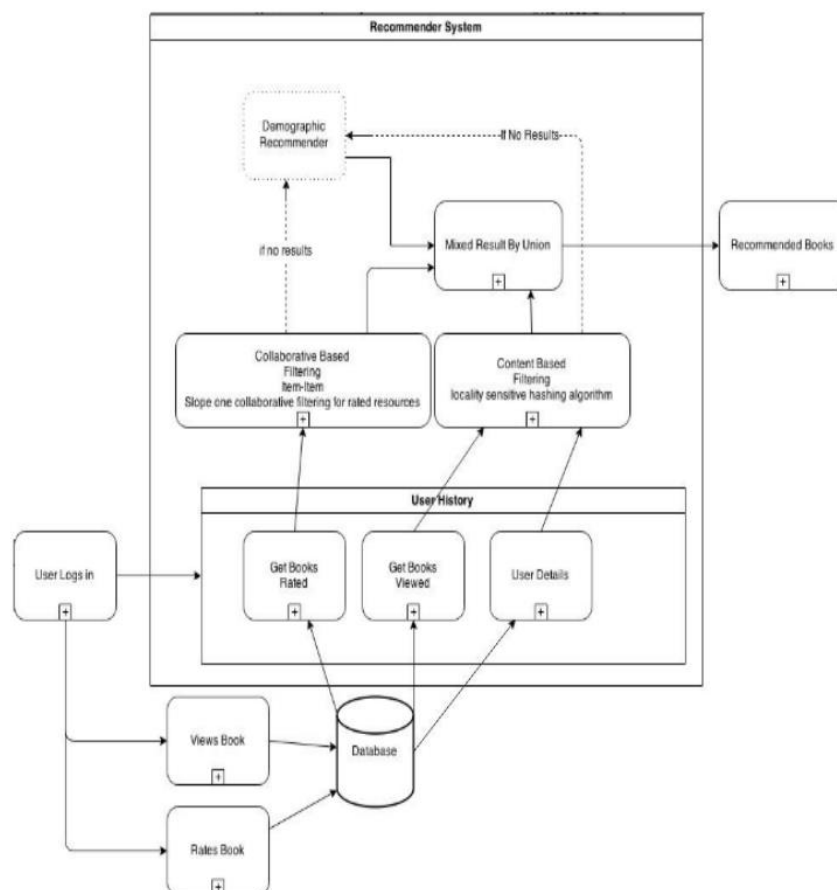


Figure 6: System architecture, (Prof. Deepali Patil, 2015)

Also, most online e-readers have existing recommendation systems to suggest relevant books to their readers and Whichbook is one of them. This site started with a clever search algorithm in 2000 and has been experimenting with different AI and Machine Learning algorithms to date (Whichbook, 2021).

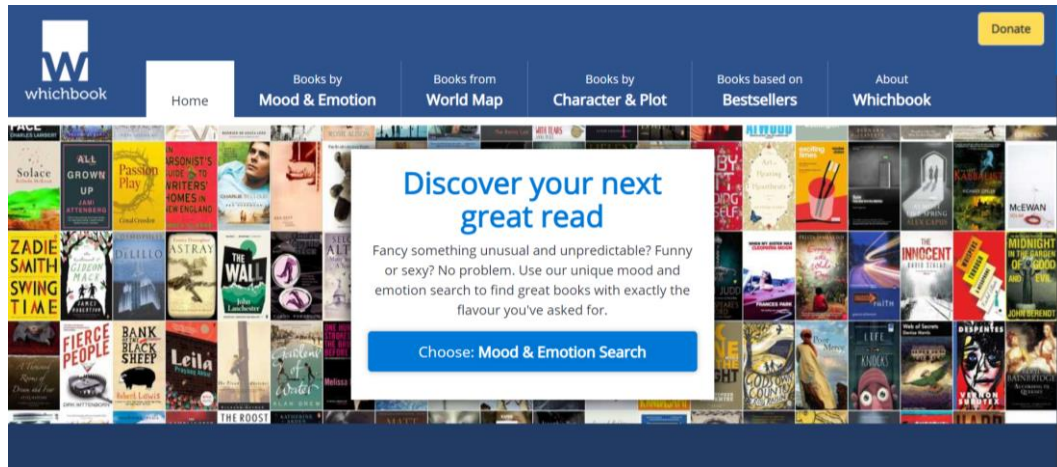


Figure 7: E-reader Whichbook, (Whichbook, 2021)

Another free online book service is BookBub, which has a huge market size and large numbers of happy readers because of its personalized book recommendation feature. It recommends books based on the reader's favorite author, friends, families, and BookBub editors (BookBub, 2021).

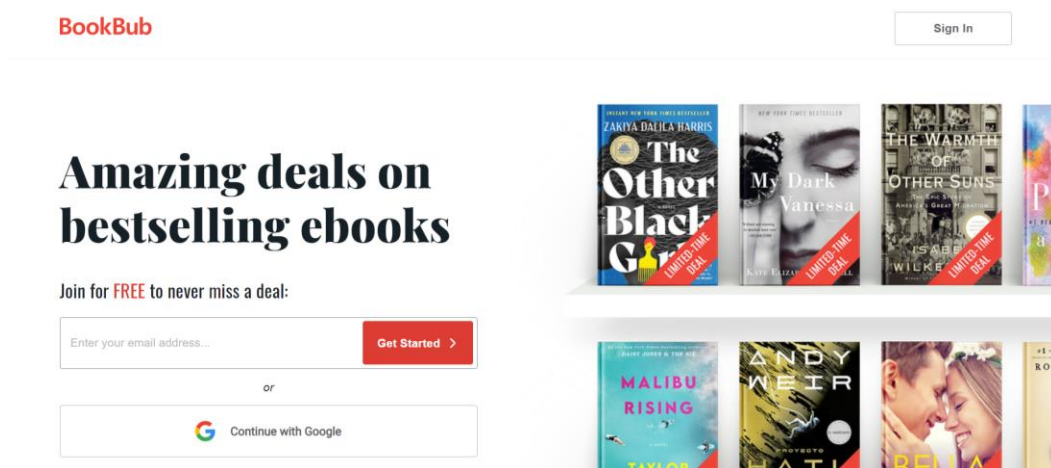


Figure 8: Online bookstore BookBub, (BookBub, 2021)

Launched in January 2007, Goodreads is the largest book reading site that provides personalized book recommendations for its users. As per their website, Goodread's recommendation engine analyses about 20 billion data points to suggest the perfect books to the readers (Goodreads, 2021).

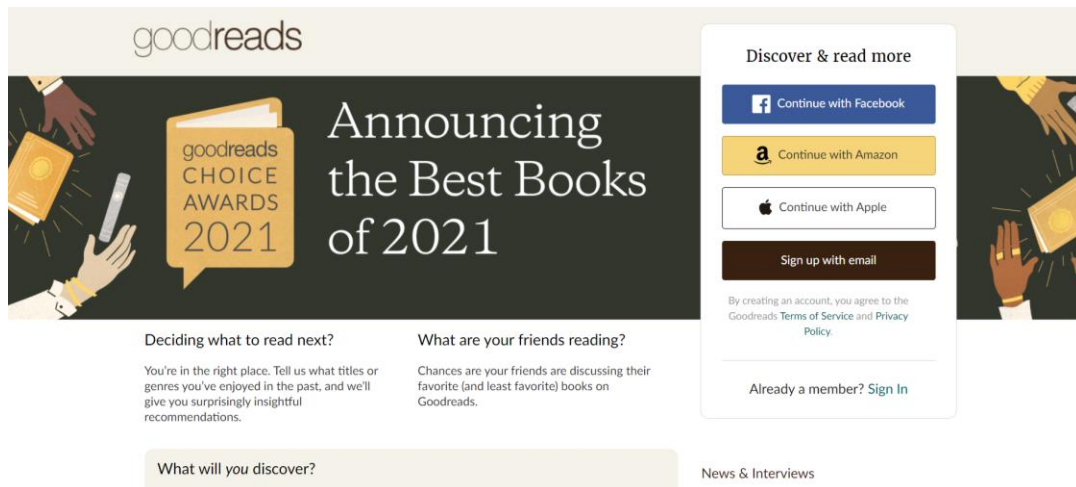


Figure 9: E-reader Goodreads, (Goodreads, 2021)

### **2.3. Pros and Cons of book recommendation systems**

The advantages of book recommendation system are summed up in the following points:

#### **Drive traffic**

Sending personalized emails and recommendations can drive more traffic into the bookselling/e-reader website which allows more data collection to enhance the customer's profile. (Alonso, 2020)

#### **Reduced time and effort**

As the recommendation system delivers relevant books to the users, it reduces the time and effort of the customer to search for books of their choice. (Alonso, 2020)

#### **Engage readers**

Individualized book recommendations make the readers more engaged as they are always excited to read new books which are recommended to them. (Alonso, 2020)

#### **Convert readers into customers**

When books are recommended to the readers daily, it makes the readers think that they are valued and as a result, it gives them a special touch that converts readers into customers. (Alonso, 2020)

#### **Increase revenue rate**

Recommendation systems can increase sales as it produces more happy customers which leads to an increased revenue rate. (Alonso, 2020)

#### **Aid in decision making**

Book recommendation systems can provide the business owner with detailed reports about book sales which aid in informed decision making. (Alonso, 2020)

Although book recommendation systems can seem beneficial, there are some potential disadvantages as well:

**Lack of data**

The lack of enormous users and their data is one of the major drawbacks of recommendation systems. For generating strong recommendations, the book sites need a large amount of data to work with (macmanus, 2009).

**Changing data**

The recommendation systems are based on the past behavior of the user and it might not be a great tool to depend on as the trends are always changing (macmanus, 2009).

**Changing readers preferences**

A reader might be willing to read books in different genres on different days. So, the reader might sometimes be incorrectly labeled by the recommendation system (macmanus, 2009).

### 3. Solution

This section explores the algorithms and techniques that will be used for implementing the proposed book recommendation system. The planned system will be developed using the collaborative filtering technique which is discussed in the section below.

#### 3.1. Collaborative filtering (CF)

Collaborative filtering refers to the process of filtering items based on the opinions of other people. The technique of collaborative filtering has established its root from the general human behavior of sharing opinions and is frequently used in recommender systems. For instance, if *user A* and *B* both like *item A* and *user B* also likes *item B*, then, *item B* is recommended to *user A* (Edpresso Team, 2021). Fig.10 demonstrates an example of collaborating filtering technique.

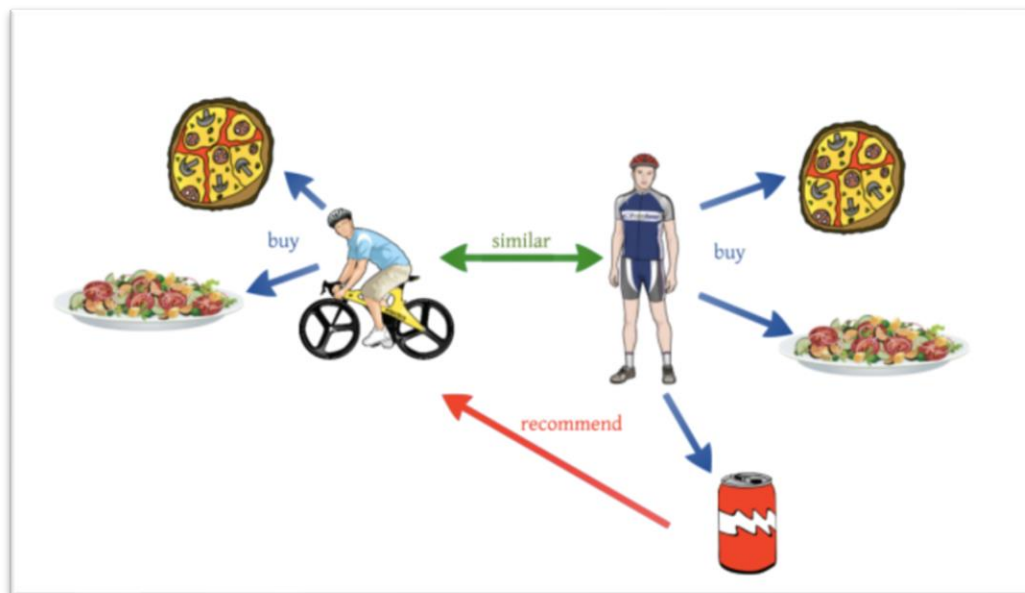


Figure 10: Example of collaborative filtering, (Grover, 2017)



There are two types of collaborative filtering approaches:

1. Nearest neighbour approach

This approach is based on finding similar users according to their similarity score and the final decision is made using the average of ratings. Similarity scores can be calculated using Pearson coefficient, Euclidean distance, cosine similarity, and so on. (Grover, 2017)

2. Model-based approach

This approach is based on rating the unrated items with the help of machine learning. PCA, SVD, Neural Nets are some of the examples of the model-based approach. (Grover, 2017)

### 3.2. Nearest Neighbour algorithm

The proposed book recommendation system will use the nearest neighbour algorithm for collaborative filtering which can be further divided into two categories user-item filtering i.e., User-based nearest neighbour algorithm, and item-item filtering i.e., Item-based nearest neighbour algorithm. User-item filtering takes a user and finds users that are similar to that user. According to their similarity of ratings, it recommends the items. Contrary to user-item filtering, the item-item filtering will first take an item and find users who like that item. Then, it finds other items that similar users liked and recommends them.

The nearest-neighbour algorithms use statistical techniques to calculate the predictions. Referring to a blog post at Real Python, to find the rating  $R$  that a user  $U$  would give to an item  $I$ , the users similar to  $U$  who have rated  $I$  should be searched first. Then, the  $R$  is calculated according to the ratings of the users found previously. The list of users similar to  $U$  can be calculated using different kinds of distance metrics such as Euclidean distance, cosine similarity, and a few more. Then, the rating  $R$  for an item  $I$  will be close to the average of the ratings given to  $I$  by the users most similar to  $U$ . Mathematically, the average rating given by  $n$  users can be calculated as shown in the figure below.

$$R_U = \left( \sum_{u=1}^n R_u \right) / n$$

Figure 11: Formula for calculating the average rating

Also, the similarity score can be calculated by using the formula:

$$\text{Sim}(a,b) = 1 / (1 + d(a,b))$$

Where  $a$  and  $b$  are two users and  $d$  gives the distance between them.

The major difference between the nearest-neighbour and model-based approach is that it does not require any kind of optimization algorithms. The similarity between the users can be calculated simply by using non-parametric Machine Learning approaches such as cosine similarity or Pearson correlation coefficient (Grover, 2017).

### 3.3. Distance metrics

For the proposed recommendation system, the considered distance metrics for calculating the similarity score are discussed below:

#### 3.3.1. Euclidean distance

Euclidean distance is one of the most popular distance metrics used by machine learning algorithms. The main purpose of using it in ML is to measure the similarity between observations (Mishra, 2020). The Euclidean distance can be calculated using the formula shown below:

Given two dimensions  $p$  and  $q$ ;  $p, q \in R^d$ :

$$p = (p_1, \dots, p_d)$$

$$q = (q_1, \dots, q_d)$$

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_d - q_d)^2}$$

#### 3.3.2. Cosine similarity

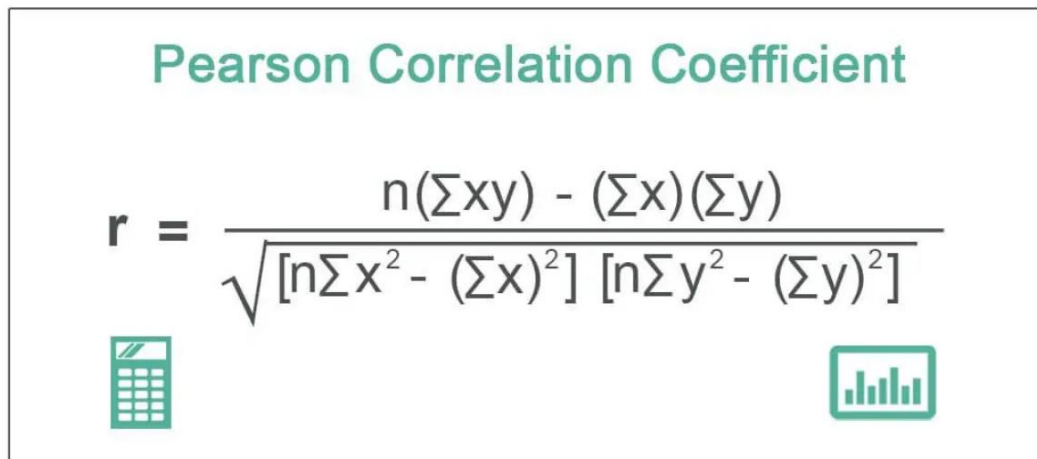
In this approach of distance metric, the similarity of two users is measured as the cosine of the angle between their vectors (Grover, 2017). For users  $u$  and  $u'$ , the cosine similarity is calculated as shown in the figure below.

$$\text{sim}(u, u') = \cos(\theta) = \frac{\mathbf{r}_u \cdot \mathbf{r}_{u'}}{\|\mathbf{r}_u\| \|\mathbf{r}_{u'}\|} = \sum_i \frac{r_{ui} r_{u'i}}{\sqrt{\sum_i r_{ui}^2} \sqrt{\sum_i r_{u'i}^2}}$$

Figure 12: Cosine similarity, (Grover, 2017)

### 3.3.3. Pearson coefficient

Pearson coefficient helps find out the relationship between two quantities. It allows us to measure the strength of association between two variables. The value of the Pearson coefficient ranges from -1 to +1. -1 indicates that there is a negative correlation, 0 indicates that there is no correlation, and +1 means that the variables are highly correlated (Chatterjee, 2021). Fig.11 presents the mathematical formula for calculating the Pearson coefficient.



**Pearson Correlation Coefficient**

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$






Figure 13: Pearson correlation coefficient, (Thakur, 2021)

Where,

- $r$  = Pearson Coefficient
- $n$  = number of the pairs of the stock
- $\sum xy$  = sum of products of the paired stocks
- $\sum x$  = sum of the  $x$  scores
- $\sum y$  = sum of the  $y$  scores
- $\sum x^2$  = sum of the squared  $x$  scores
- $\sum y^2$  = sum of the squared  $y$  scores

### 3.4. Flowchart of the proposed algorithm

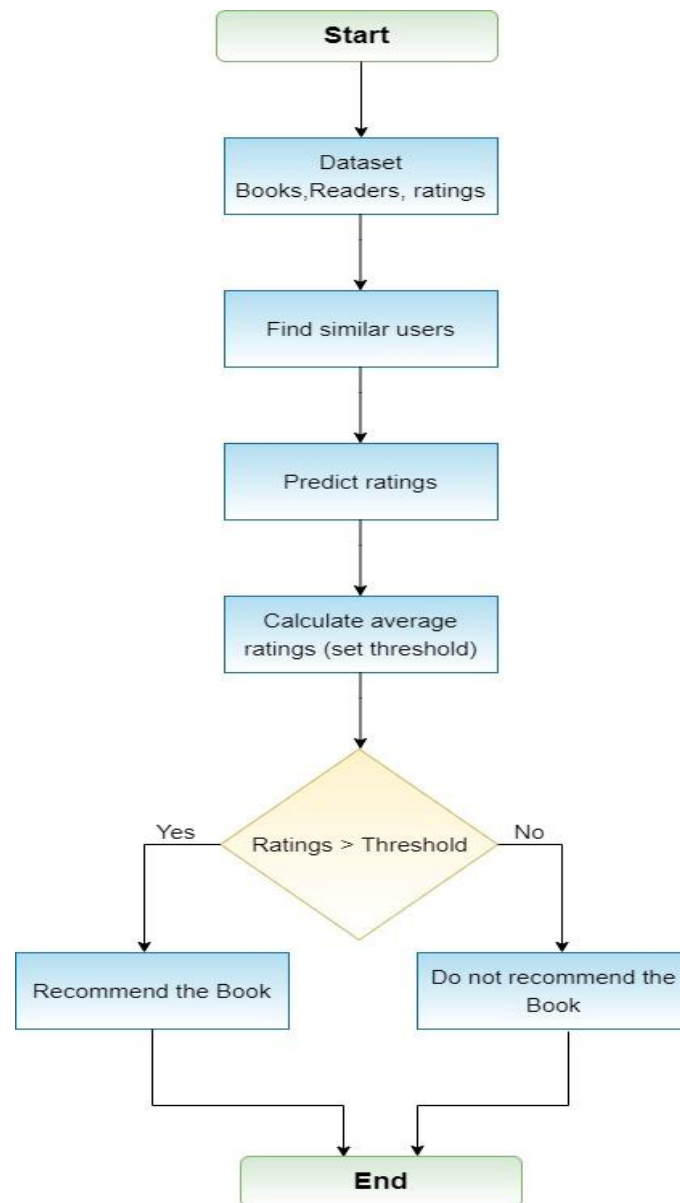


Figure 14: Flowchart of the proposed algorithm

Fig.14 above represents the flow of the chosen algorithm for developing a book recommendation system. The first step is to pass the dataset of books into the algorithm. Then, similar users are found using the distance metrics. The considered distance metrics are Euclidean distance, cosine similarity, and Pearson coefficient. The next step is to predict the user ratings for the books which they have not rated. A threshold is set by calculating the average ratings of the users which will be used to determine whether a book with a certain rating should be recommended to the reader or not.

### 3.5. System overview diagram

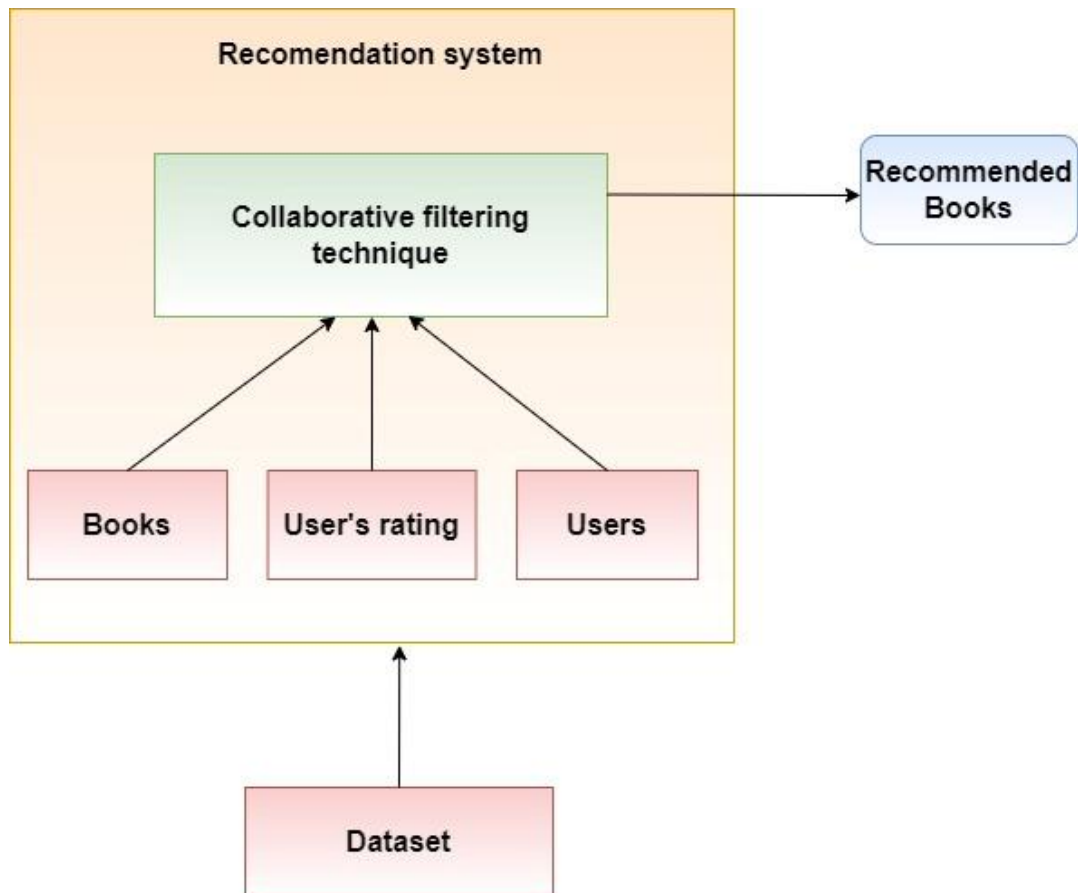


Figure 15: System overview diagram

Fig.15 shows the system architecture of the proposed book recommendation system. As it can be seen, the dataset is fed into the recommendation engine which uses a collaborative filtering technique to produce recommended books.

## 4. Conclusion

The presented book recommendation system is focused on simplicity and speed. The main goal of this system is to provide effective and relevant book recommendations without having to use an enormous amount of data such as a big user profile, browsing history, and readers attribute data which includes age, gender, geography, and so on. As the proposed system is based on a collaborative filtering method, it does not depend upon contents that are capable of being analyzed by machines. One of the key advantages of this system is that it does not require an understanding of the item itself ("books" in this case) to be recommended. So, the book recommender can solve all the problems and is advantageous in the field of e-commerce bookselling sites. In future work, the proposed book recommendation system will be developed using the books dataset and nearest neighbour algorithm. The model will be trained, tested, and validated for producing effective recommendations. After the successful development of the system, a final report analyzing the work done will also be prepared.

All in all, recommendation systems are an extremely beneficial tool that can be utilized to assist the users with their selection process. Whether it be an e-commerce bookstore or any other industry such as finance, medical, or sports, recommendation systems have proved to be the ultimate technique for consumer satisfaction and retention. Customer satisfaction and financial success are the two parameters in business that are closely linked with each other. If consumer satisfaction levels up, so does the financial success. On the other hand, if consumer satisfaction levels down, so does financial success. Hence, this book recommender is a competent system for e-sellers that addresses both the business parameters. As the market size continues to increase, such systems will be of great use for bibliophiles as it will reduce the work of finding the best books among the plenty. Also, the future development of recommendation systems is seen to implement deep learning and neural networks which shall increase the business value even more.



## 5. References

Agrawal, S. K., 2021. *Analytics Vidhya*. [Online]  
Available at: <https://www.analyticsvidhya.com/blog/2021/07/recommendation-system-understanding-the-basic-concepts/>

[Accessed 15 December 2021].

Alonso, B., 2020. *Kibo*. [Online]  
Available at: <https://kibocommerce.com/blog/recommendation-engine-benefits-aov/>

[Accessed 17 December 2021].

Apáthy, S., 2019-2021. *Onespire*. [Online]  
Available at: <https://www.onespire.net/news/history-of-recommender-systems/>

[Accessed 17 December 2021].

BookBub, 2021. *BookBub*. [Online]  
Available at: <https://www.bookbub.com/home/overview.php>

[Accessed 17 December 2021].

Chatterjee, S., 2021. *Medium*. [Online]  
Available at: <https://towardsdatascience.com/data-correlation-can-make-or-break-your-machine-learning-project-82ee11039cc9>

[Accessed 18 December 2020].

Das, S., 2020. *Analytics Vidhya*. [Online]  
Available at: <https://www.analyticsvidhya.com/blog/2020/11/create-your-own-movie-movie-recommendation-system/>

[Accessed 16 December 2021].

Edpresso Team, 2021. *educative*. [Online]  
Available at: <https://www.educative.io/edpresso/what-is-collaborative-filtering>

[Accessed 18 December 2021].

Goodreads, 2021. *Goodreads*. [Online]  
Available at: <https://www.goodreads.com/about/us>

[Accessed 17 December 2021].

[Accessed 17 December 2021].

Grover, P., 2017. *towards data science*. [Online]  
Available at: <https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0>

[Accessed 18 December 2021].

Kozlowski, M., 2020. *Good e-reader*. [Online]  
Available at: <https://goodereader.com/blog/kobo-ereader-news/over-2-million-new-users-have-registered-with-kobo-during-the-pandemic>

[Accessed 17 December 2021].

macmanus, r., 2009. *readwrite*. [Online]  
Available at: [https://readwrite.com/2009/01/28/5\\_problems\\_of\\_recommender\\_systems/](https://readwrite.com/2009/01/28/5_problems_of_recommender_systems/)

[Accessed 17 December 2021].

Mishra, A., 2020. *Data science foundation*. [Online]  
Available at: <https://datascience.foundation/datatalk/distance-metrics-in-machine-learning>

[Accessed 18 December 2021].

Nursultan Kurmashov, K. N. L. A. N., 2015. *ReseachGate*. [Online]  
Available at: [https://www.researchgate.net/publication/300412849\\_Online\\_book\\_recommendation\\_system](https://www.researchgate.net/publication/300412849_Online_book_recommendation_system)

[Accessed 16 December 2021].

Prof. Deepali Patil, A. S. ,. K. V. R., 2015. A Proposed Book Recommender System. *nternational Journal of Advanced Research in Computer and Communication Engineering*, 4(2), pp. 481-483.

ProjectsGeek, 2018. *ProjectsGeek*. [Online]  
Available at: <https://projectsgeek.com/2018/07/online-book-recommendation-system-project.html>

[Accessed 16 December 2021].

Region, And Segment Forecasts, 2020-2027. *Grand view research*. [Online]  
Available at: <https://www.grandviewresearch.com/industry-analysis/online-book->

services-market

[Accessed 17 December 2021].

Rocca, B., 2019. *towards data science*. [Online]  
Available at: <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>

[Accessed 15 December 2021].

Thakur, M., 2021. *WallStreetMojo*. [Online]  
Available at: <https://www.wallstreetmojo.com/pearson-correlation-coefficient/>

[Accessed 18 December 2021].

Whichbook, 2021. *Whichbook*. [Online]  
Available at: <https://www.whichbook.net/about/our-story/>

[Accessed 17 December 2021].