المجلة العلمية لكلية التربية – المجلد الثالث– العدد الأول – يناير 2024
Faculty of Education journal – Volume 3 Issue: 1 – January 2024
الموقع الالكتروني للمجلة: http://journal.su.edu.ly/index.php/edujournalj/index
DOI: https://doi.org/10.37375/foej.v3i1.2589

تطبيق خوارزمية التصفية التعاونية لنظام التوصية

Muftah Ali Mohamed Elmorabet Eniba
Department of Information Technology
Faculty of Education – Misurata University
Misurata – Libya
eniba_muf@edu.misuratau.edu.ly

Ezeddin Ali Mohamed Eyad
Department of Computer Technologies
Higher Institute for Science & Technology
Misurata – Libya
ez1.ayad@gmail.com

الملخص:

تناولت هذه الدراسة جانباً مهماً وحيوياً في البحث العلمي، ألا وهو موضوع اقتناء مصادر المعلومات كالكتب من مواقع التجارة الإلكترونية، فعندما تكون أنواع الكتب المعروضة على هذه المواقع الإلكترونية كثيرة جدا، يكون من الصعب على المستخدم اختيار الكتب المناسبة له من بين الكم الهائل من الكتب الموجودة التي تزداد أعدادها سنوياً، وهذا ينعكس سلبا على المستخدمين في اهدار كثير من الوقت في اختيار ما يحتاجون منها.

لذلك اتجهت هذه الدراسة للمساهمة في حل هذه المشكلة وذلك بمساعدة المستخدمين في استكشاف واقتناء مصادر المعلومات المناسبة لاهتماماتهم، ذلك من خلال استخدام خوارزمية التصفية التعاونية المبنية على سلوك المستخدم وهي احدى طرق نظم التوصية، وللتحقق من دقة النظام فقد تم استخدام معادلة فرق المسافة لقياس مدى التشابه بين الكتب الموصي بها والكتاب المطلوب.

وتوصلت الدراسة الى مجموعة من النتائج أهمها الحصول على معلومات مفيدة تساعد المستخدمين في اتخاذ القرارات المناسبة لتلبية احتياجاتهم من خلال سلوكهم واقتراح عناصر ذات صلة باهتماماتهم، كما تجنب المستخدم اتخاذ القرارات التي لا تلبي اهتماماته او الغير مرغوب فيها، ومن ثم التنبيه بالاختيارات (القرارات) التي تساعده للوصول الى نتائج أفضل.

ويوصي هذه الدراسة باستخدام خوارزميات أخرى كخوارزمية التوصية باستخدام المحتوى للمقارنة بينها وبين الخوارزمية المعتمدة على سلوك المستخدم، وتوضيح أفضلهما من حيث النتائج.

الكلمات المفتاحية: سلوك المستخدم، خوارزمية التصفية التعاونية، اهتمامات المستخدم، الكتب المقترحة، أنظمة التوصية.

Applying Collaborative Filtering Algorithm for Recommendation Systems

Ezeddin Ali Mohamed Eyad
Department of Computer Technologies
Higher Institute for Science & Technology
Misurata – Libya
ez1.ayad@gmail.com

Muftah Ali Mohamed Elmorabet Eniba
Department of Information Technology
Faculty of Education – Misurata University
Misurata – Libya
eniba_muf@edu.misuratau.edu.ly

Abstract

This study addresses a vital aspect of scientific inquiry, namely the acquisition of information sources such as books from e-commerce websites. When the variety of books offered on these websites is extensive, it becomes challenging for user to select suitable books from the vast available number, which continues to grow annually. This negatively affects users, as they waste a significant amount of time in the selection process.

Therefore, this study aims to contribute to solving this problem by assisting users in exploring and acquiring information sources that align with their interests. The collaborative filtering algorithm, based on user behaviour, was employed as one of the recommendation system methods, to verify the accuracy of the system, the distance difference equation was used to measure the similarity between the recommended books and the required book.

The study obtained several results, including providing users with helpful information to make appropriate decisions that meet their needs based on their behaviour. It also suggests related items based on their interests, thereby avoiding decisions that do not align with their preferences. Additionally, the study alerts users to choices that assist them in achieving better outcomes.

Furthermore, this study recommends comparing the collaborative filtering algorithm, which relies on user behaviour, with other algorithms such as content-based recommendation algorithms. The aim is to highlight the superior algorithm in terms of results.

Keywords: User behavior, collaborative filtering algorithm, user interests, recommended books, and recommendation systems.

1. Introduction:

The significant advancements in web technologies in recent years have led to a rapid increase in the amount of information available online. As we live in the age of technology, the use of the internet has become essential in our daily lives, providing various advanced services tailored to users' needs, such as the "Amazon" e-commerce application and the "Google News" app for reading news articles, among others.

Simultaneously, user behaviour has evolved and become more complex. Users are no longer just consumers of information; they have become producers as well. Their activities extend beyond searching and browsing to include interaction, content creation, and sharing. The proliferation of vast amounts of big data on the internet does not necessarily improve the quality of our lives or provide us with useful information. Instead, it can be overwhelming and time-consuming, leading to distraction, increased stress, and pressure. Therefore, it has become necessary to use tools to filter the available information for users and select the relevant and useful parts for them. This is where Recommendation Systems (RS) come into play. (Jussi, 1990).

RS is designed to recommend items that users may be interested in based on several factors, including user preferences. They can predict user preferences and the products they are searching for on the internet. (Ricci, Rokach, Shapira, 2015).

These systems utilize their algorithms to find solutions to the problem of providing users with suitable products. They achieve this by predicting user preferences based on their previous data or by using data from similar users, depending on the methodologies employed by the system. Recommendation engines, which are essentially data filtering tools, aim to derive user satisfaction with specific products.

The main purpose of a recommendation system is to suggest items of interest to users and support them in decision-making. This can be achieved by employing various types of recommendation systems, such as content-based recommendation systems, collaborative filtering systems, utility-based recommendation systems, and demographic-based recommendation systems.

2. Study Problem and Questions:

Given the vast amount of data and the difficulty in identifying and acquiring appropriate books, and understanding user behavior to assist them in meeting their requirements, the study problem can be summarized by the following questions:

1. Can user interests be categorized and recommendations made accordingly?
2. How can users be assisted in exploring and finding information sources that align with their interests?
3. Can recommendation systems be a suitable solution to help users choose the right book?
4. Is it possible to engage and encourage researchers and users to use internet search engines and online web services to obtain information sources matching their interests?

3. Study Objectives:

The study aims to achieve the following objectives:

1. Categorize user interests by predicting their behaviour and provide recommendations accordingly.
2. Assist users in exploring and finding information sources that align with their interests using recommendation systems.
3. Aid users in making suitable choices by suggesting relevant items based on their interests using specific types of recommendation systems designed for this purpose.
4. Raise awareness among researchers and users and encourage them to use internet search engines and online web services to obtain information sources that match their interests.

4. concepts and terminology of study:

4.1. Recommendation Systems: Automatic systems that utilize machine learning algorithms to provide users with suggestions that meet their requirements as shown in (Figure 1). The main task of a recommendation system is to generate a list of recommendations in a specific domain, using one of the filtering methods, including content-based filtering, collaborative filtering, and hybrid recommendation as shown in (Figure 2). (Ricci, Rokach, Shapira, 2015).

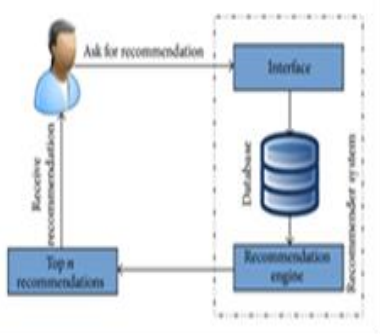

Figure 1: Systems work structure

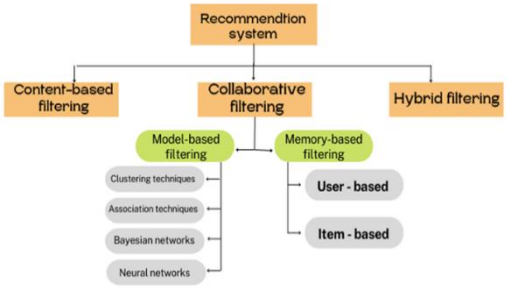

Figure 2: Types of recommendation systems

4.1.1. Collaborative Filtering: is one of the common approaches used in designing recommendation systems, as shown in (Figure 3). It involves gathering and analyzing a large amount of user information and interests using artificial intelligence algorithms, as illustrated in (Figures 4) and (figure 5), to make automatic predictions about user interests. It also involves aggregating the preferences and activities of multiple users and predicting user interests based on the similarity of their activities and preferences with other users. One distinctive feature of collaborative filtering is that it does not rely on content analysis. It is a method of obtaining information or patterns using collaborative techniques to make recommendations, such as book recommendations, without the need to understand the content of the book. (Ekstrand, Riedl, Konstan, 2011).

Figure 3: collaborative filtering structure

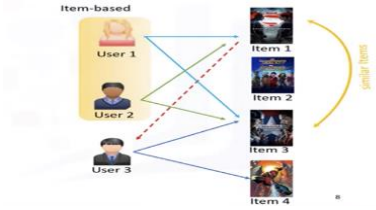

Figure 4: diagram of a collaborative filtering algorithm based on an element

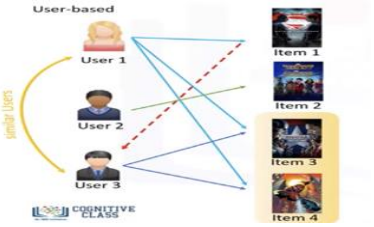

Figure 5: diagram of a collaborative filtering algorithm based on an user

4.2. E-commerce: It is the process of buying and selling goods and services over the Internet.

4.3. Web Services: It is a type of web service that provides electronic services between one application and another or one system and another. The application connects and requests the service from the web service of another application through its programming interface, and then the web service sends the result of the request to the user. (Papazoglou, Georgakopoulos, 2008).

4.4 Flask: Flask is a micro-framework or a so-called Micro Framework used in Python programming to build simple websites and web applications. (Grinberg, 2018).

4.5. User: the individual who engages in buying and selling goods and services online, using their own computers or other communication devices. (Oracle Commerce, 2022).

5. Previous Studies:

1. A study by (Yahya, 2022), titled " Recommendation Systems with Data Mining Integration", investigates the utilization of recommendation systems to comprehend individual interests and behavioral patterns. The study employs these systems to deduce user preferences, utilizing the acquired insights to predict products aligned with those patterns and subsequently present them as recommendations. This fosters a user's perception that the system comprehends their preferences, thereby enhancing trust in service providers. The

researcher addresses the issue of information overload, characterized by the pressure resulting from receiving more information than necessary for decision-making, by employing advanced time management techniques. This challenge hinders the ability to evaluate product details and make choices from a diverse array of online options. The researcher attempts to design a Recommendation System (RS) to address unique challenges in recommendation technology and solve issues such as reducing large dimensions and handling cold starts. The alignment in objectives, methodology, and outcomes with this study.

2. A study by (Diab, 2019), titled "Context-Based Recommendations in Social Web Applications", aimed to propose effective recommendation methods that assist in addressing the issue of information overload by leveraging users' social context. The study indicated that social context information can be utilized to enhance performance. It also highlighted the existence of related studies on cooperative recommendation techniques. Experimental results demonstrated that the proposed method significantly improves recommendation performance compared to traditional approaches.

3. A study by (Othman, 2018), titled "Using the recommendation system algorithm for Collaborative Filtering to apply to Sudanese universities", Recommendation systems are mechanisms or technologies aimed at anticipating the elements that may be of interest to the user. They involve software tools and techniques to provide optimal suggestions for user utilization. This paper presents the design of a model for a recommendation system and provides an explanation of the Collaborative Filtering algorithm applied in the field of education. The focus of this paper is on the behavior of students applying to Sudanese universities through the electronic application website, with the goal of selecting the most suitable colleges based on their interests, preferences, grades, specialization, and courses. Additionally, the paper addresses the user interaction with the new electronic application website, which is characterized by its smoothness and ease of use. The Collaborative Filtering algorithm facilitates decision-making for students, offering ease and flexibility. The system is built using the Oracle 12 C Developer ADF framework. The paper presents several results, notably highlighting the benefits of using the Collaborative Filtering algorithm in the management of electronic applications for Sudanese universities. It enables students to effortlessly obtain useful recommendations at any time, facilitating their access to valuable suggestions.

4. A study by (Mokhtar, 2015), titled "Cooperative Filtering Approach for Recommending Elective Courses", The study was carried out at the University of Sudan and aimed to recommend elective courses and integrate them into the current registration system of the College of Business Administration, whether online or offline. The recommendation system is based on several different collaborative filtering algorithms such as item-based and user-based. These algorithms offer item recommendations or predictions based on similar opinions of other users who think similarly to the current user. The approach requires minimal information and provides accurate recommendations. Two cooperative filtering approaches were compared to predict students' grades in different courses based on their performance in previous courses. The primary difference between item-based and user-

based approaches is that item-based relies on an item-to-item similarity model, while user-based relies on a user-to-user similarity model, primarily used to solve the prediction problem. The highly encouraging experimental results obtained from real-life datasets suggest that the student-based cooperative filtering approach for grade prediction is accurate. It was observed that there is not a significant difference in accuracy between item-based and user-based approaches.

6. System Description and Components:

Collaborative filtering was used, which studies user behaviour in choosing a specific product that aligns with their interests, by approaching user interests to facilitate the selection process. It was implemented in a book-specific application website, which required HTML pages to display the algorithms and how they work. Python language was also used and connected with HTML using the Flask tool.

6.1 Used Database:

Adatabase from the website Amazon.com was used, which contains a large collection of data divided into three files: Book data file, Interaction data file, and User data file, as shown in (Figure 6) and (Table 1.) These files contain important information such as book ratings, book title, author, publication year, publisher, book summary, and other information. The CSV database extension contains over 60,831 records, and this database was used to test the recommendation system.

Table 1: used Database

Name	Volume
book-data	65363 K.B
Ratings	4202 K.B
Users	10692 K.B

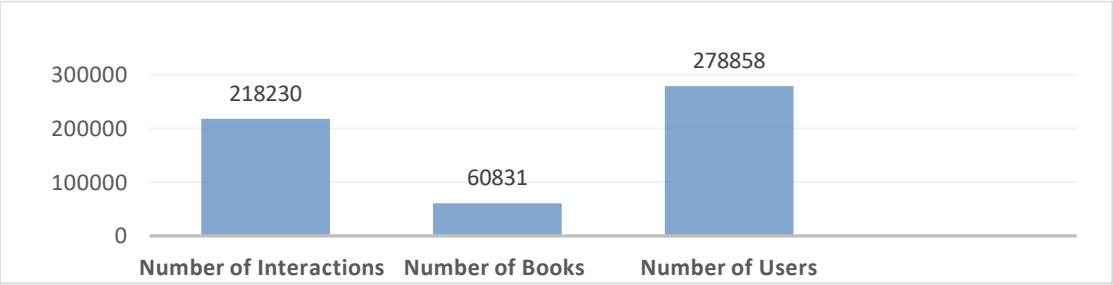

Figure 6: analysis of a collaborative filtering database

6.2. Building the Recommendation Website:

A simple website was designed using HTML pages and operated using the Python language to execute the Python code and illustrate the working concept of recommendation systems. A set of web pages was prepared, directly linked to a database extracted from the collaborative filtering recommendation algorithm, as shown in (Figure 10).

Figure 10: Homepage of the Proposed System.

(Figure 11) illustrates the book browsing page from the database used in this research.

Figure 11: Book Review Page.

(Figure 12) illustrates the book recommendation page, where the system searches for books similar to the requested book, and the result will be as shown in (Figure 13).

Figure 12: Recommended Books Search Page.

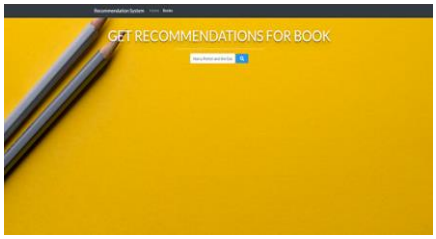

Figure 13: Result Page.

6.3. Operation Mechanism of Collaborative Filtering Algorithm

The operation of the collaborative filtering algorithm follows a set of steps that can be divided into three main stages: (Initial Processing Stage, Calculation and Processing Stage, and Result Presentation Stage), as shown in (Figure 7).

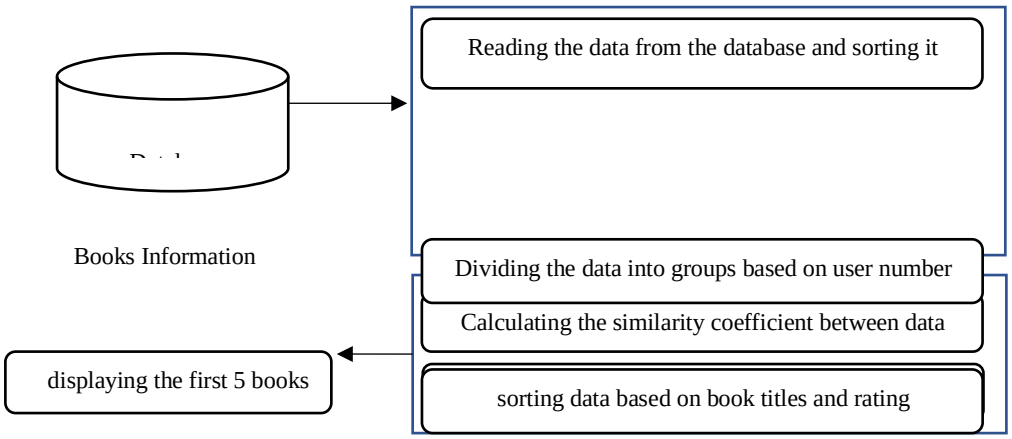

Figure 7: the working mechanism of a collaborative filtering algorithm.

6.3.1. Initial Processing Stage:

The initial processing stage involves obtaining the data (Get Data and Stored) and preparing it. This includes removing unnecessary data, retaining the required data, and eliminating duplicates, as shown in (Table2).

Table 2: Importing Data from the Books Database after Processing

BookId ... price			
0	27827010-a-gift-of-love-and-joy	...	9.85
1	11783587-a-mere-six-degrees-of-separation	...	NaN
2	4335624-a-structured-approach-to-basic-program...	...	34.69
3	40548137-adventures-of-love	...	NaN
4	52830931-after-the-storm	...	NaN
...	...	...	...
50553	4671.The_Great_Gatsby	...	NaN
50554	2657.To_Kill_a_Mockingbird	...	NaN
50555	41865.Twilight	...	2.1
50556	2767052-the-hunger-games	...	5.09
50557	3. Harry_Potter_and_the_Sorcerer_s_Stone	...	NaN

We observe that the number of books decreased from 60,831 to 50,557 books as a result of removing duplicate data.

6.3.2. The data is divided into groups based on user ratings (Books Rating Group by User ID), as shown in (Table 3).

Table 3: Data after being divided into groups based on User ID

User-ID	ISBN	Book-Rating	...	bbeScore	bbeVotes	price
0		276726		155061224	5	
...	101	2		3.29	.	
1	276726	155061224	5..	88	1	
3.13						
2		159181		155061224	0	
...	101	2			3.29	
3		159181		155061224	0	
...	88	1			3.13	
4		276727		446520802	0	
...		228630		2582	NaN	
....						
173329		250405		1559945044	0	
...	53	1			8.66	
173330		250506		151874301	0	
...	90	1			2.17	
173331		250510		312856830	0	
...	1091	13			3.73	
173332	250510	877738726	0	...	96	1
27.19						
173333	250645	1581823045	0	...	187	2

5.**2**

6.3.3. After that, the data is further divided into groups, but this time based on the (Book Title), as shown in (Table 4).

Table 4: Data after being divided into groups based on Book Title

User-ID	ISBN	Book-Rating	...	bbeScore	bbeVotes	price
5	278418	446520802	0 ...	228630	2582	NaN
7	3363	446520802	0 ...	228630	2582	NaN
8	7158	446520802	10 ...	228630	2582	NaN
9	8253	446520802	10 ...	228630	2582	NaN
11	11676	446520802	10 ...	228630	2582	NaN
...						
168060	150498	671508482	0 ...	91	1	7.34
168061	150498	671508482	0 ...	490	6	47.07
168062	241306	671508482	0 ...	97	1	3.12
168063	241306	671508482	0 ...	91	1	7.34
168064	241306	671508482	0 ...	490	6	47.07

6.4. Calculation and Final Processing Stage:

In this stage, the similarity between items is calculated using cosine similarity, a fundamental algorithm in collaborative filtering that relies on measuring similarity. It determines the degree of similarity between two items (or users). However, in user-based collaborative filtering, the focus is on attributes that align closely with the perspectives of multiple users to calculate similarity.

The similarity between users can be calculated using cosine similarity, which is a measure of the difference between two individuals by using the cosine of the angle between their vectors. The smaller the angle between the vectors, the more similar the vectors are.

If the vectors are similar in direction, meaning the angle between them is close to zero, then the vectors are considered similar. To determine whether the vectors' directions are similar, we use cosine similarity to calculate the angle between the vectors.

Cosine similarity can be calculated as in equation (1).

$$\cos(\theta) = \frac{A.B}{\|A\|\|B\|} \qquad (1)$$

Where A and B are vectors in the sample space, and in this study, we rely on cosine similarity.

Applying this equation to the items will yield a matrix, where the elements represent the degree of association between the items relative to a specific item. (Table 5) illustrates a portion of this matrix.

Table 5: Subset of the Similarity Matrix

```
Similarity                                                                    =
[[1.      0.10767638 0.1253876  ... 0.       0.09215122 0.15488062]

[0.10767638 1.    0.08625819 ... 0.1        0.11410887 0.11986583]

[0.1253876  0.08625819 1.          ... 0.03450328 0.04724556 0.09925833]
                  ...

[0.         0.1    0.03450328 ... 1.        0.04564355 0.11986583]

[0.09215122 0.11410887 0.04724556 ... 0.04564355 1.    0.03282661]

[0.15488062 0.11986583 0.09925833 ... 0.11986583 0.03282661  1. ]]
```

6.5. Results Presentation:

A recommended number of books are suggested when comparing the desired book with the rest of the items through the similarity matrix. Then, the elements that have the highest correlation and contain the highest value are selected, where a value of one indicates perfect similarity, a value of zero indicates no similarity, and values in between represent the degree of similarity. When selecting the book "A Heartbreaking Work of Staggering Genius," the result will be as shown in (Table 6).

Table 6: Recommended Books for the book "A Heartbreaking Work of Staggering Genius"

1.The Corrections
2.Snow Falling on Cedars
3.Stones from the River
4.Ella Minnow Pea: A Novel in Letters
5.House of Sand and Fog
6.Year of Wonders
7.Stolen Lives: Twenty Years in a Desert Jail
8.The Stone Diaries
9.Girl with a Pearl Earring
10.An Instance of the Fingerpost

When selecting the book "The Corrections," the result will be as shown in (Table 7).

Table 7: Recommended Books for the book "The Corrections"

1.A Heartbreaking Work of Staggering Genius
2.Snow Falling on Cedars
3.Running with Scissors
4.Stolen Lives: Twenty Years in a Desert Jail
5.A Fine Balance
6.House of Sand and Fog
7.Breath, Eyes, Memory
8.The Little Friend
9.A Map of the World
10.Oryx and Crake

6.5.1. Results and Evaluation:

The results and the website, which were prepared, are presented to illustrate the utilized recommendation system and the subsequent evaluation.

To evaluate the system, a set of examples was tested by labelling a book for the system and then interpreting and analyzing the results.

1. Book: "A Heartbreaking Work of Staggering Genius"
   When the system is given the book titled "A Heartbreaking Work of Staggering Genius," it will search for the top 5 most similar books to the requested book and display the results with the following book recommendations:

1. The Corrections
2. Snow Falling on Cedars
3. Stones from the River
4. Ella Minnow Pea: A Novel in Letters
5. House of Sand and Fog

These are the top five books that are most similar to the requested book, and (Figure 8) illustrates the similarity distance value between the requested book and the five most similar recommended books.

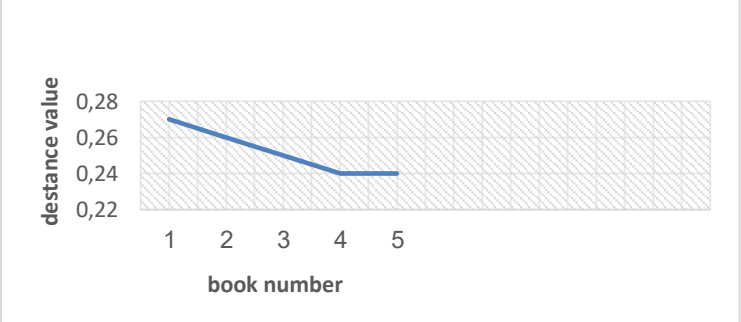

Figure 8: Similarity Distance between the Target

Book and the Recommended Books

2. Book: "Year of Wonders"
   When suggesting another book, let's say "Year of Wonders," the system searches for the top 5 most similar books and suggests the following:
   1. Empire Falls.
   2. A Heartbreaking Work of Staggering Genius.
   3. The Virgin Blue.
   4. Ella Minnow Pea: A Novel in Letters.
   5. A Fine Balance.

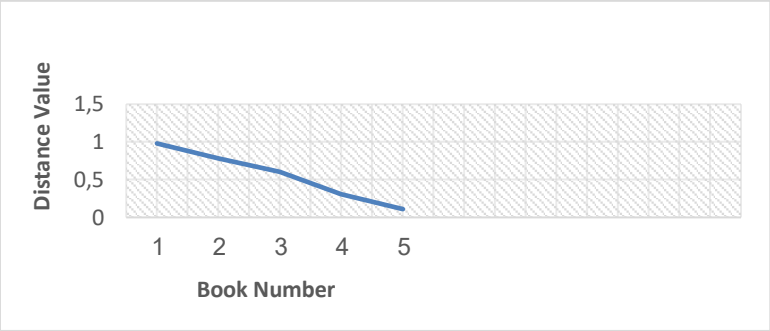

Figure 9: Similarity Distance between the Target Book and the Recommended Books

In this study, the measurement of results relies on the distance between the elements of the similarity matrix. It represents the distance between the requested book and the suggested books, providing a clear indicator of which approach is closer to the user and more effective. Therefore, when choosing the book "A Heartbreaking Work of Staggering Genius," the distance plot for collaborative filtering will be as shown in (Figure 8), and when choosing the book " Year of Wonders", the distance plot for collaborative filtering will be as shown in (Figure 9).

7. Conclusions:

Through the practical implementation of a collaborative filtering recommendation algorithm, designed and employed on a website that aligns with collaborative filtering technology, several key findings have been obtained. These include the utilization of Internet and website services to acquire valuable information, thereby raising researchers' awareness of the significance of recommendation systems. The system effectively guides users in exploring relevant information sources, assisting them in making appropriate decisions based on their behaviour, and suggesting items of interest that enhance user satisfaction. Moreover, it significantly saves researchers' time and effort in accessing information sources within their areas of interest.

8. Recommendations:

This study recommends the utilization of alternative algorithms, such as content-based recommendation algorithms, to compare them with user behaviour-based algorithms and determine their superior performance in terms of results. Furthermore, it suggests the development of a system through web applications that incorporate these algorithms in other domains, such as product recommendations and scientific movies. The hope is to attract the attention of relevant stakeholders to support and further develop these ideas. It is also recommended to provide training and preparation for specialized personnel in this field. Additionally, the study highlights the importance of conducting related studies to maximize the benefits derived from recommendation systems.

9. References

- jussi, Karlgren, (1990), for recommendations, retrieved from http://www.lingvi.st, Date of visit14/3/2023.

- Ricci, F., Rokach, L., & Shapira, B. (2015), Introduction to Recommender Systems Handbook. In Recommender Systems Handbook, Springer. Boston, MA.

- Ekstrand, M. D., Riedl, J., & Konstan, J. A. (2011), Collaborative filtering recommender systems. Foundations and Trends® in Human-Computer Interaction, 4(2).

- Papazoglou, M. P., & Georgakopoulos, D. (2008), Service-oriented computing. Communications of the ACM, 46(10).

- Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python. O'Reilly Media.

- Oracle Commerce, (2022), retrieved from the https://www.oracle.com /ae-ar/cx/ , Date of visit7/3/2023.

- Diab, E., (2019) Context-Based Recommendations in Social Web Applications. (Unpublished doctoral dissertation). Al-Baath University.

- Othman, N. M. Moussa, (2018), Using the recommendation system algorithm for Collaborative Filtering to apply to Sudanese universities, High study magazine, f.11, N, 43, pp. 76-93.

- Mokhtar, Wafa. (2015), Cooperative Filtering Approach for Recommending Elective Courses, Khartoum, Sudan: Sudan University.

- Yahya Omran Alaa (2022), Recommendation Systems with Data Mining Integration, (Master's Thesis), Iraq: university of Kufa, faculty of computer and Mathematics, retrieved from the .
https://mathcomp.uokufa.edu.iq, Date of visit 9/5/2023.