Jurnal Ilmiah Informatika dan Komputer (Informatech)
Volume 3, Number 2026, June 2026, page 49 -57
DOI : 10.69533
Website : *https://ejournal.rizaniamedia.com/index.php/informatech*

E-ISSN: 3047-4752

Developing a Content-Based Book Recommendation System for School E-Libraries Using TF-IDF and Cosine Similarity

Muhammad Fatkul Roziq[1*], Riska Nurtantyo Sarbini[2], Halimahtus Mukminna[3] , Moh. Syaiful Anam[4]

[*,1, 2, 3, 4)] Progam Studi Teknik Komputer, Universitas Islam Kadiri

[*1]*email : fatkulroziq@gmail.com*

[2]*email : tyo@uniska-kediri.ac.id*

[3]*email : halimahtusm@uniska-kediri.ac.id*

[4]*email : anam@uniska-kediri.ac.id*

(*Article Received: 19 April 2026; Article Revised: 15 May 2026; Article Published: 1 June 2026;*)

ABSTRACT *– The rapid growth of e-libraries has increased the availability of digital book collections, making it difficult for users to identify relevant reading materials. This study develops an automated book recommendation system for the SMKN 1 Semen e-library using Content-Based Filtering with TF-IDF for term weighting and Cosine Similarity for similarity measurement. The dataset consisted of 35 book records containing metadata and book descriptions collected from the SMKN 1 Semen e-library. Text data were processed through case folding, tokenization, stopword removal, and stemming before recommendation generation. The system was developed using the Laravel framework and a MySQL database following the Research and Development (R&D) method with the Waterfall model. Recommendation performance was evaluated using a Top-5 recommendation scenario, where the recommended books were compared with manually identified relevant books based on content similarity using Precision and Recall metrics, while system functionality was verified through Black Box Testing. The experimental results achieved a Precision of 80% and a Recall of 57.14%, indicating that the proposed approach effectively generates relevant book recommendations. This study contributes by demonstrating that the integration of Content-Based Filtering, TF-IDF, and Cosine Similarity provides an effective recommendation approach for school e-libraries with limited user interaction data, enabling personalized book recommendations and improving the efficiency of book discovery in small-scale digital library environments.*

Keywords *- content-based filtering, TF-IDF, cosine similarity, recommendation systems, e-libraries*

Pengembangan Sistem Rekomendasi Buku Berbasis Konten untuk Perpustakaan Elektronik Sekolah Menggunakan TF-IDF dan Cosine Similarity

ABSTRAK *– Pertumbuhan pesat perpustakaan elektronik telah meningkatkan ketersediaan koleksi buku digital, sehingga menyulitkan pengguna untuk mengidentifikasi bahan bacaan yang relevan. Studi ini mengembangkan sistem rekomendasi buku otomatis untuk perpustakaan elektronik SMKN 1 Semen menggunakan Content-Based Filtering dengan TF-IDF untuk pembobotan istilah dan Cosine Similarity untuk pengukuran kesamaan. Dataset terdiri dari 35 catatan buku yang berisi metadata dan deskripsi buku yang dikumpulkan dari perpustakaan elektronik SMKN 1 Semen. Data teks diproses melalui case folding, tokenisasi, penghapusan stopword, dan stemming sebelum pembuatan rekomendasi. Sistem ini dikembangkan menggunakan framework Laravel dan database MySQL mengikuti metode Penelitian dan Pengembangan (R&D) dengan model Waterfall. Kinerja rekomendasi dievaluasi menggunakan skenario rekomendasi Top-5, di mana buku-buku yang direkomendasikan dibandingkan dengan buku-buku relevan yang diidentifikasi secara manual berdasarkan kesamaan konten menggunakan metrik Precision dan Recall, sementara fungsionalitas sistem diverifikasi melalui Black Box Testing. Hasil eksperimen mencapai Precision sebesar 80% dan Recall sebesar 57,14%, menunjukkan bahwa pendekatan yang diusulkan secara efektif menghasilkan rekomendasi buku yang relevan. Studi ini berkontribusi dengan menunjukkan bahwa integrasi Content-Based Filtering, TF-IDF, dan Cosine Similarity memberikan pendekatan rekomendasi yang efektif untuk perpustakaan elektronik sekolah dengan data interaksi pengguna yang terbatas, memungkinkan rekomendasi buku yang dipersonalisasi dan meningkatkan efisiensi penemuan buku di lingkungan perpustakaan digital skala kecil..*

Kata Kunci *– content-based filtering, TF-IDF, cosine similarity, sistem rekomendasi, e-perpustakaan*

Jurnal Ilmiah Informatika dan Komputer (Informatech)
Volume 3, Number 2026, June 2026, page 49 -57
DOI : 10.69533
Website : *https://ejournal.rizaniamedia.com/index.php/informatech*
E-ISSN: 3047-4752

1. INTRODUCTION

Several previous studies have applied Content-Based Filtering (CBF) to book recommendation systems using different weighting and similarity techniques. Alkaff et al. demonstrated that combining TF-IDF and Cosine Similarity can generate relevant book recommendations based on content similarity in regional digital library services. Rahmadhani et al. improved recommendation performance by replacing TF-IDF with TF-RF weighting, while Bhajantri et al. proposed a hybrid recommendation approach integrating Collaborative Filtering, Association Rule Mining, and Content-Based Filtering to enhance recommendation accuracy. Although these studies have shown promising results, they mainly focus on public libraries or large-scale recommendation platforms that generally have abundant user interaction data and more diverse collections.

However, limited attention has been given to the implementation and evaluation of Content-Based Filtering in school e-library environments, where the number of book collections is relatively small and user interaction data are limited. Existing studies also place greater emphasis on improving recommendation accuracy through alternative weighting methods or hybrid approaches, while providing limited discussion on the effectiveness of the conventional TF-IDF and Cosine Similarity combination for small-scale educational digital libraries. This research addresses this gap by implementing and evaluating a Content-Based Filtering recommendation system using TF-IDF and Cosine Similarity in the SMKN 1 Semen e-library. The study aims to determine whether this lightweight and computationally efficient approach can provide relevant book recommendations without relying on extensive user behavior data, making it more suitable for school library environments.

TF-IDF was selected as the text-weighting technique because of its simplicity, computational efficiency, and proven effectiveness in representing document content in recommendation and information retrieval systems. Compared with alternative weighting methods such as TF-RF, BM25, or word embedding approaches, TF-IDF does not require class labels, parameter tuning, or large training datasets. These characteristics make TF-IDF particularly suitable for school e-library applications with only 35 book records, where lightweight computation and straightforward implementation are more practical while still providing meaningful document representations. Combined with Cosine Similarity, TF-IDF enables effective measurement of content similarity and supports accurate recommendation generation in environments with limited data.

2. LITERATURE REVIEW

The development of recommendation systems has become an important solution for helping users discover relevant information from large collections of digital content. In the context of digital libraries, recommendation systems can assist users in finding books that match their interests and reading needs more efficiently. Bhajantri et al. [5] developed a hybrid book recommendation system by combining Collaborative Filtering, Association Rule Mining, and Content-Based Filtering, resulting in improved recommendation accuracy. Meanwhile, Unnisa and Akshaya [6] integrated Machine Learning with Content-Based Filtering to provide mood-centric book recommendations, demonstrating that personalized recommendations can increase user satisfaction and engagement. These studies indicate that recommendation systems play a significant role in improving user experience in digital information services.

Among various recommendation approaches, Content-Based Filtering (CBF) is widely used because it generates recommendations based on item characteristics rather than relying on preferences from other users. According to Ricci [6], Content-Based Filtering recommends items that possess attributes similar to those previously preferred by users. Lops et al. [9] further explain that the method analyzes item features to construct user profiles and identify similar items. In digital library environments, Content-Based Filtering is particularly suitable because book information such as titles, categories, authors, and descriptions can be utilized as content features for recommendation generation. Lops et al. [15] explain that the method analyzes item features to construct user profiles and identify similar items. In digital library environments, Content-Based Filtering is particularly suitable because book information such as titles, categories, authors, and descriptions can be utilized as content features for recommendation generation.

Recent studies demonstrate a growing interest in applying recommender systems to digital libraries using Content-Based Filtering, Collaborative Filtering, and Hybrid approaches. Hybrid recommendation methods generally achieve higher recommendation accuracy by combining multiple recommendation techniques. However, these approaches usually require extensive user interaction

histories and involve greater computational complexity. In contrast, Content-Based Filtering generates recommendations based on item characteristics rather than user behavior, making it more suitable for educational digital libraries where user interaction data are often limited.

Previous studies have shown that TF-IDF combined with Cosine Similarity is effective for representing textual information and generating relevant book recommendations. Other researchers have explored alternative weighting techniques such as TF-RF or integrated machine learning methods to improve recommendation performance. Nevertheless, most of these studies focus on public libraries or large-scale recommendation systems with abundant user interaction data. Limited research has investigated the effectiveness of conventional TF-IDF and Cosine Similarity in small-scale school e-library environments with limited collections. Therefore, this study evaluates whether this lightweight approach can provide effective personalized book recommendations in the SMKN 1 Semen e-library.

The effectiveness of Content-Based Filtering strongly depends on document representation and similarity measurement techniques. TF-IDF remains one of the most widely adopted term-weighting methods because it effectively represents the importance of terms within documents while maintaining computational efficiency. Compared with alternative techniques such as TF-RF, BM25, or word embedding methods, TF-IDF is easier to implement, does not require parameter tuning or training data, and performs well on relatively small datasets. After document vectors are generated, Cosine Similarity provides an efficient mechanism for measuring document similarity by calculating the angle between vectors. Considering that this research uses only 35 book records in a school e-library, the combination of TF-IDF and Cosine Similarity provides a practical, lightweight, and computationally efficient recommendation approach while maintaining satisfactory recommendation performance.

3. RESEARCH METHOD

This research employed a Research and Development (R&D) method with the Waterfall software development model. The R&D method was used because the research not only conducted analysis but also produced a product, a book recommendation system for an e-library. The Waterfall model was chosen because it has systematic and structured stages, encompassing needs analysis, system design, implementation, testing, and evaluation. The research phase began with identifying system requirements through observations and interviews with the e-library administrators of SMKN 1 Semen. Next, a recommendation system based on Content-Based Filtering (CBF) was designed, implemented using the Laravel and MySQL frameworks, and functional testing and recommendation quality evaluation were conducted.

3.1 DATASET DESCRIPTION

The dataset used in this study consisted of 35 book records obtained from the SMKN 1 Semen e-library database. Each record contains bibliographic and textual information used for recommendation generation. The dataset includes several book categories representing different subject areas available in the school library. The attributes utilized in the recommendation process are the book title, author, category, publisher, publication year, and book description. Among these attributes, the book description serves as the primary textual feature for preprocessing, TF-IDF weighting, and Cosine Similarity calculations, while the remaining metadata are used to support recommendation presentation within the system.

3.2 CONTENT-BASED FILTERING

Content-Based Filtering (CBF) is a method in recommendation systems that generates recommendations based on the characteristics or attributes of an item. This method works by analyzing the similarity of content between items so that users can get recommendations for items that have similar characteristics to items previously selected or interested in. In a book recommendation system, the attributes used can be the title, category, author, or description of the book. The advantage of the Content-Based Filtering method is that it does not require a large amount of user interaction data, making it suitable for systems with limited user data. Furthermore, this method is able to provide more personalized recommendations based on the characteristics of the item the user is accessing.

Although Content-Based Filtering offers several advantages, including personalized recommendations and independence from user interaction data, the method also has several limitations. One important limitation is the overspecialization problem, where users tend to receive recommendations that are highly similar to previously accessed books, thereby reducing recommendation diversity. Another limitation is the cold-start problem for newly added books that contain insufficient descriptive information, making similarity calculations less effective. Despite these limitations, Content-Based Filtering remains an appropriate approach for school e-library environments because user interaction data are limited while descriptive metadata and book descriptions are available. Future research may overcome these limitations by integrating

Collaborative Filtering or Hybrid recommendation approaches.

3.3 Text Preprocessing

Text preprocessing is the initial stage in text processing, which aims to clean and prepare data before analysis. This stage is necessary to reduce noise and improve the quality of document representation. The preprocessing process used in this study includes: (1) Case Folding, which is the process of changing all letters to lowercase so that the word format is uniform. (2) Tokenizing, which is the process of breaking down sentences into groups of words or tokens that can be processed further. (3) Stopword Removal, which is the process of removing common words that do not have important meaning in the information search process, such as "and", "which", "or", and "di". (4) Stemming, which is the process of changing affixed words into basic words so that variations in word forms can be represented in the same form. The results of the preprocessing stage are used as the basis for forming text features in the document weighting process.

The preprocessing stage was implemented in PHP within the Laravel framework. Case folding and tokenization were performed using native PHP string processing functions. Indonesian stop-word removal utilized the Sastrawi StopWordRemover library, while stemming was carried out using the Sastrawi Stemmer library, which is widely adopted for Indonesian Natural Language Processing (NLP). The processed text was subsequently used to construct TF-IDF vectors for similarity computation.

3.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is a word-weighting method used to determine the importance of a word in a document. This method combines the frequency of a word's appearance in a document (Term Frequency) with the word's rarity across the entire document (Inverse Document Frequency). The TF-IDF value is calculated using Equation.

$$TFIDF(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right)$$

Description:
(1) TF(t,d) = frequency of occurrence of term t in document d.
(2) DF(t) = number of documents containing term t.
(3) N = total number of documents.
(4) TF-IDF(t,d) = weight of term t in document d.
The greater the TF-IDF value of a word, the greater the contribution of the word in representing the contents of the document.

3.5 Cosine Similarity

Cosine Similarity is a method used to measure the level of similarity between two documents represented as vectors. This method calculates the cosine value of the angle between two vectors, resulting in a similarity value between 0 and 1. Values closer to 1 indicate a high level of similarity between the two documents. The Cosine Similarity equation is shown in Equation.

$$Similarity(A,B) = \frac{A \cdot B}{|A| \times |B|}$$

Description:
(1) A·B = dot product between vectors A and B
(2) |A| = magnitude of vector A
(3) |B| = magnitude of vector B.
The similarity value ranges from 0 to 1. A value closer to 1 indicates that two documents have highly similar content.

3.6 Precision

Precision was calculated as the proportion of relevant books among the five recommended books, whereas Recall was calculated as the proportion of relevant books successfully retrieved from all relevant books available in the dataset.

Precision measures the proportion of recommended books that are relevant to users. Precision was calculated using Equation.

$$Precision = \frac{Relevant\ Recommended\ Books}{Total\ Recommended\ Books} \times 100\%$$

3.7 Recall

The relevance assessment was conducted manually by comparing the recommended books with books that share similar categories and content descriptions. A recommendation was considered relevant when the recommended book belonged to the same subject area and contained substantially similar learning content as the reference book. Recommendation performance was evaluated using the Top-5 recommendation scenario

Recall measures the proportion of relevant books that were successfully recommended by the system. Recall was calculated using Equation.

$$Recall = \frac{Relevant\ Recommended\ Books}{Total\ Relevant\ Books} \times 100\%$$

The evaluation was conducted using the Top-5 Recommendation scenario. The obtained Precision and Recall values were used to assess the effectiveness of the TF-IDF and Cosine Similarity methods in generating relevant book recommendations in the e-library system.

3.8 Baseline

This study focuses on evaluating the implementation of Content-Based Filtering using TF-IDF and Cosine Similarity within a school e-library

environment. Therefore, no comparative recommendation algorithm was implemented as a baseline. The effectiveness of the proposed approach was assessed through functional testing and recommendation quality evaluation using Precision and Recall metrics. Comparative evaluation with alternative recommendation algorithms is proposed as future work.

3.9 Experimental Setup

The recommendation experiment was conducted using a Top-5 recommendation configuration, where the five books with the highest Cosine Similarity scores were presented to the user. The recommendation process was performed for every selected book in the dataset. During evaluation, the generated recommendations were compared with manually identified relevant books based on category similarity and content descriptions. Precision and Recall were then computed to evaluate recommendation quality. The Top-5 configuration was selected because it provides a concise recommendation list while remaining practical for users in a school e-library environment with a relatively small collection of 35 books.

Respone :

Thank you for your valuable comments. The Methodology section has been substantially revised. A detailed description of the dataset, including the number of book records and attributes, has been added. The preprocessing stage now specifies the libraries used for Indonesian stop-word removal and stemming. Complete mathematical formulations of TF-IDF and Cosine Similarity have been included. The relevance assessment procedure for Precision and Recall evaluation has been clarified, together with the experimental setup and Top-5 recommendation configuration. Finally, the manuscript explicitly states that no baseline method was implemented because the study focuses on the implementation and evaluation of a Content-Based Filtering approach in a school e-library, while comparative evaluation is identified as future work.

4 Results And Discussion

4.1 System Implementation

The login page is the initial display used by users to access the library system by entering their registered email/username and password.

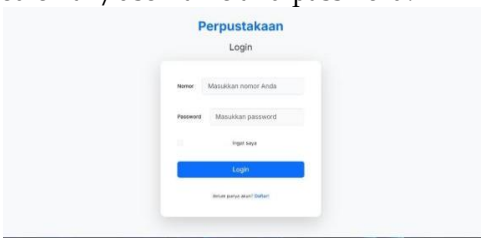

Figure 1. Login Page

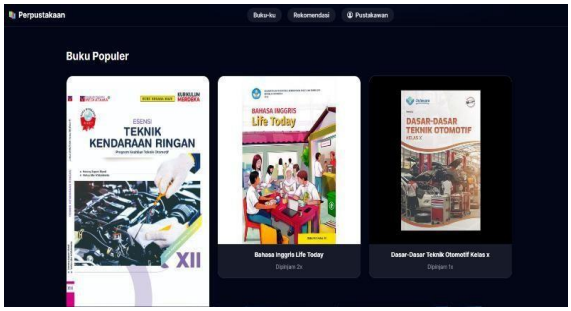

Figure 2. Book List Page

The book list page in the e-library system functions to display the entire book collection book detail page on the e-library system which functions to present complete information regarding the book selected by the user.

The book borrowing history page display a list of books that have been borrowed by the user as a form of recording activity in the e-library system.

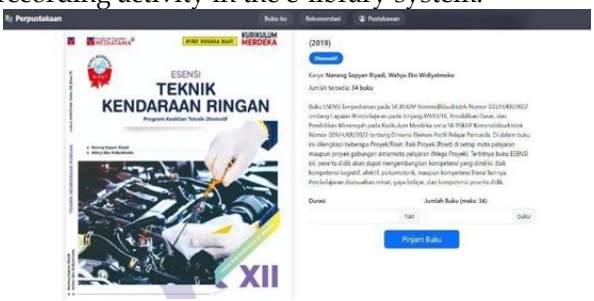

Figure 3. Book Detail Page

Book recommendation page in the e-library system generated automatically using the Content-Based Filtering method.

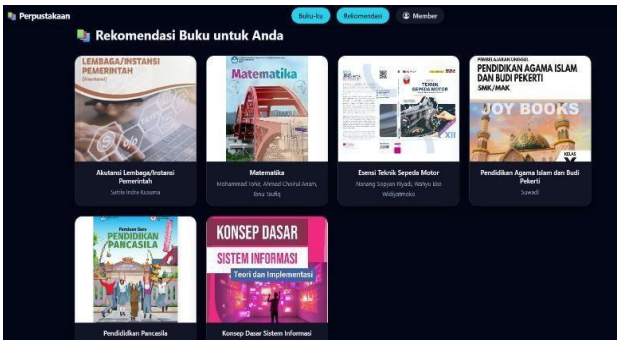

Figure 4. Book Recommendation Page

The developed recommendation system was implemented as a web-based application using the Laravel framework and a MySQL database. The system integrates book management, search functionality, borrowing history, and an automated recommendation module based on Content-Based Filtering. Rather than emphasizing the user interface, this study focuses on the recommendation mechanism, where book descriptions are preprocessed, weighted using TF-IDF, and ranked using Cosine Similarity. The recommendation results are displayed automatically on the book detail page by presenting the Top-5 books with the highest

similarity scores. This implementation demonstrates the feasibility of integrating a lightweight recommendation engine into a school e-library environment.

A book recommendation system was developed in a web-based e-library application using the Laravel framework and a MySQL database. The recommendation feature is displayed on the book detail page, displaying a list of books with the highest similarity to the book currently selected by the user. The system automatically performs TF-IDF weighting and Cosine Similarity calculations to generate relevant recommendations. The system's implementation allows users to obtain book recommendations without having to manually search. Recommendations are displayed in order of highest similarity, with the most relevant books at the top.

4.3 Text Preprocessing Results

Preprocessing was performed on the book description data before TF-IDF weighting. An example of book description data is shown below: Initial text
"Basic automotive technology studies the basic concepts of vehicle systems and automotive components." The preprocessing stage resulted in the following changes: (1). Case Folding, basic automotive technology studies the basic concepts of vehicle systems and automotive components. (2). Tokenizing, [technology, basic, automotive, study, concepts, basics, systems, vehicles, and, components, automotive]. (3). Stopword Removal, [technology, basic, automotive, study, concepts, basics, systems, vehicles, and, components, automotive]. (4). Stemming, [technology, basic, automotive, study, concepts, basics, systems, vehicles, and, components, automotive].

The preprocessing results show that words that do not contribute to the recommendation process are successfully removed so that document representation becomes more optimal.

4.4 TF-IDF Calculation

After preprocessing, each term is assigned a weight using the TF-IDF method. For example, if the term "automotive" appears twice in a document and has an IDF value of 0.69, the TF-IDF weight is calculated using Equation (1).
TF-IDF = 2 × 0.69 = 1.38

4.5 Calculation Cosine Similarity

The similarity value between books is calculated using the Cosine Similarity method. For example, there are two document vectors, A and B, formed from TF-IDF weighting.
A = (1.38; 0.69; 0.45)
B = (1.20; 0.72; 0.40).

Based on Equation (2), the similarity score is 0.89. This value indicates that the two books have a high level of similarity and are therefore worthy of being recommended over each other. The calculation process is performed on the entire book collection in the database. The system then performs a ranking process based on the highest similarity score and displays the five best recommendations (Top-5 Recommendations).

4.6 Black Box Test Results

Black Box testing is conducted to ensure all system functions operate according to user requirements. Test scenarios include the login process, book data management, category management, book search, and the book recommendation feature.

Table 1. Black Box testing results.

No.	System Features	Results
1	Login	Successful
2	Manage Book Data	Successful
3	Manage Categories	Successful
4	Book Search	Successful
5	Book Recommendations	Successful
6	Logout	Successful

Based on the test results, all system functions performed according to specified requirements. No functional errors were found that would impact the book recommendation process.

4.7 Evaluation of Recommendation Performance

The system was evaluated using Precision and Recall metrics in the Top-5 Recommendation scenario. Testing was conducted by comparing the system's recommendation results against predetermined relevance data.

Table 2. The evaluation results

No	Metrik	Nilai
1	Precision	80%
2	Recall	57,14%
3	F1-Score	66.67%

A Precision score of 80% indicates that most of the books recommended by the system match the relevant categories or content of the referenced books. In other words, the majority of recommended books have a good match.

Meanwhile, a Recall score of 57.14% indicates that the system was unable to find all relevant books available in the library collection. This score is influenced by limited information in the metadata and book descriptions used as the basis for calculating similarity.

In addition to Precision and Recall, the F1-score was calculated to provide a balanced evaluation of recommendation performance. The obtained F1-score of 66.67% indicates that the proposed

recommendation model maintains a reasonable balance between recommendation accuracy and coverage. Although the Precision is relatively high, the moderate Recall value reduces the overall F1-score, suggesting that further improvements are needed to retrieve a larger proportion of relevant books.

4.8 Evaluation

The relatively high Precision (80%) indicates that most recommended books are relevant to the selected reference book. This result demonstrates that TF-IDF combined with Cosine Similarity effectively identifies books with similar textual characteristics. However, the Recall value of 57.14% indicates that not all relevant books available in the collection were successfully retrieved. This occurs because Content-Based Filtering recommends books based solely on textual similarity. Relevant books that contain different vocabulary, shorter descriptions, or limited metadata may receive lower similarity scores and therefore are not included in the Top-5 recommendations. Furthermore, the relatively small dataset consisting of only 35 books limits the diversity of recommendation candidates, which also affects Recall performance.

Although this study does not directly compare the proposed approach with alternative recommendation algorithms such as BM25, Hybrid Recommender Systems, or deep learning-based methods, previous studies have reported that hybrid approaches generally achieve higher recommendation accuracy by combining content information with user interaction data. However, these methods require substantially more user data and computational resources. In contrast, the proposed TF-IDF and Cosine Similarity approach is computationally lightweight, easy to implement, and well suited to school e-library environments where user interaction data are limited.

4.9 Discussion

The experimental results demonstrate that the proposed recommendation model successfully generates relevant book recommendations for a school e-library environment. The high Precision value indicates that TF-IDF effectively represents important textual features, while Cosine Similarity accurately measures document similarity. These findings suggest that Content-Based Filtering is a practical recommendation approach for digital libraries with limited user interaction data.

Nevertheless, the moderate Recall value indicates that relying solely on textual similarity may overlook relevant books that contain different terminology or incomplete descriptions. This limitation reflects one of the inherent characteristics of Content-Based Filtering, where recommendation quality strongly depends on document representation. Future improvements may include enriching book metadata, incorporating semantic text representation methods, or integrating Collaborative Filtering within a Hybrid recommendation framework to improve recommendation diversity and retrieval coverage.

From a practical perspective, the proposed system can support librarians and students by reducing the time required to discover relevant books while improving user experience in accessing digital library collections. Therefore, the developed recommendation model provides both practical benefits for educational institutions and a foundation for further research on recommendation systems in small-scale digital libraries.

Response:

Thank you for your constructive comments. The Results and Discussion section has been substantially revised. The description of system interfaces has been reduced, while scientific analysis has been strengthened. An additional evaluation metric (F1-score) has been included to complement Precision and Recall. The discussion now explains the relationship between the high Precision and moderate Recall values and provides a deeper interpretation of the experimental findings. Although no direct comparison with alternative recommendation algorithms was conducted, the discussion now explains the advantages and limitations of the proposed approach relative to other recommendation methods reported in previous studies. The implications of the findings for school e-library environments have also been elaborated.

5 CONCLUSION

This study has successfully developed and evaluated a Content-Based Filtering-based book recommendation system for a web-based school e-library using TF-IDF for term weighting and Cosine Similarity for document similarity measurement. The experimental evaluation showed that the proposed system achieved a Precision of 80%, a Recall of 57.14%, and an F1-score of 66.67%, indicating that the recommendation model is capable of providing relevant book recommendations within a small-scale digital library environment.

The main scientific contribution of this research lies in demonstrating that the integration of Content-Based Filtering, TF-IDF, and Cosine Similarity provides a simple, computationally efficient, and practical recommendation approach for school e-libraries with limited user interaction data and relatively small book collections. Unlike recommendation methods that require extensive user behavior information, the proposed approach relies solely on textual content and book metadata while still producing satisfactory recommendation performance. These findings contribute to the application of content-based recommendation

techniques in educational digital libraries and provide evidence that lightweight recommendation models can effectively support personalized book discovery in resource-constrained environments.

Future research may improve recommendation performance by incorporating richer textual representations, semantic embedding techniques, or hybrid recommendation methods combining Content-Based Filtering with Collaborative Filtering. In addition, evaluation using larger datasets and more comprehensive recommendation metrics, such as MAP and NDCG, would provide a more complete assessment of recommendation quality.

Response:

Thank you for your valuable comment. The Conclusion section has been completely revised. Instead of merely restating the experimental results, the revised conclusion now explicitly highlights the scientific contribution of the study by emphasizing the applicability of Content-Based Filtering combined with TF-IDF and Cosine Similarity for school e-library environments with limited user interaction data and small-scale collections. The conclusion also discusses the broader implications of the findings and outlines several directions for future research, including the use of hybrid recommendation methods, semantic text representations, larger datasets, and additional evaluation metrics.

6 ACKNOWLEDGEMENTS

The author would like to thank all parties who provided support and assistance in conducting this research. Special thanks go to the supervisor who provided direction, input, and guidance throughout the research process and the writing of this article. Furthermore, the author would like to thank the school and all parties who provided the necessary data and information to ensure the successful completion of this research.

BIBLIOGRAPHY

[1] A. Bhajantri et al., "Personalized Book Recommendations: A Hybrid Approach Leveraging Collaborative Filtering, Association Rule Mining, and Content-Based Filtering," EAI Endorsed Transactions on Internet of Things, vol. 10, 2024, doi: 10.4108/eetiot.6996.

[2] S. Unnisa and A. N. S., "Personalized Mood-Centric Book Recommendation Integrating Machine Learning with Content Based Filtering," International Journal of Information Technology, Research and Applications, vol. 3, no. 3, pp. 15–22, Sep. 2024, doi: 10.59461/ijitra.v3i3.100.

[3] M. Alkaff, H. Khatimi, and A. Eriadi, "Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering," MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 1, pp. 193–202, Sep. 2020, doi: 10.30812/matrik.v20i1.617.

[4] S. Rahmadhani et al., "Sistem Rekomendasi Penelusuran Buku Berbasis Content-Based Filtering dengan Pembobotan TF-RF," JIP (Jurnal Informatika Polinema), 2024.

[5] R. Ardiansyah, M. A. Bianto, and B. D. Saputra, "Sistem Rekomendasi Buku Perpustakaan Sekolah Menggunakan Metode Content-Based Filtering," Jurnal CoSciTech (Computer Science and Information Technology), vol. 4, no. 2, pp. 510–518, Oct. 2023, doi: 10.37859/coscitech.v4i2.5131.

[6] F. Ricci, Recommender Systems Handbook. Boston, MA, USA: Springer, 2015, doi: 10.1007/978-1-4899-7637-6.

[7] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, Recommender Systems: An Introduction. Cambridge, United Kingdom: Cambridge University Press, 2016.

[8] 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, Jun. 2005, doi: 10.1109/TKDE.2005.99.

[9] P. Lops, M. de Gemmis, and G. Semeraro, "Content-Based Recommender Systems: State of the Art and Trends," in Recommender Systems Handbook, Boston, MA, USA: Springer, 2011, pp. 73–105, doi: 10.1007/978-0-387-85820-3_3.

[10] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," in The Adaptive Web, Berlin, Germany: Springer, 2007, pp. 291–324, doi: 10.1007/978-3-540-72079-9_9.

[11] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[12] H. Kurniadi, Y. B. Utomo, and M. S. Anam, "Penerapan Algoritma Naive Bayes Sebagai Sistem Pendukung Keputusan Dalam Penentuan Pemilihan Mobil Untuk Driver Online," Jurnal TeIKa, vol. 15, Apr. 2025.

[13] Y. B. Utomo, D. E. Yuliana, and H. Kurniadi, "Sistem Pendukung Keputusan dalam Menentukan Ketua Himaprodi Menggunakan Metode Weighted Product," Jurnal Teknik Informasi dan Komputer

(Tekinkom), vol. 5, no. 2, p. 501, Dec. 2022, doi: 10.37600/tekinkom.v5i2.703.

[14] S. Zhang et al., Proceedings of the 13th ACM Conference on Recommender Systems (RecSys 2019). New York, NY, USA: Association for Computing Machinery, 2019.

[15] N. Qazi and B. L. W. Wong, "An Interactive Human-Centered Data Science Approach Towards Crime Pattern Analysis," Information Processing and Management, vol. 56, no. 6, p. 102066, Nov. 2019, doi: 10.1016/j.ipm.2019.102066.

[16] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep Learning Based Recommender System: A Survey and New Perspectives," ACM Computing Surveys, vol. 56, no. 1, pp. 1–38, 2023.

[17] J. Wu, X. Wang, and Y. Zhang, "Recent Advances in Recommender Systems: A Survey," IEEE Access, vol. 11, pp. 85672–85695, 2023.

[18] H. Wang, F. Zhang, X. Xie, and M. Guo, "Large Language Models for Recommender Systems: A Survey," arXiv preprint arXiv:2308.10837, 2023.

[19] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, 3rd ed. Cham: Springer, 2022.

[20] P. Lops, M. de Gemmis, and G. Semeraro, "Content-Based Recommender Systems: State of the Art and Trends," in Recommender Systems Handbook, 3rd ed., Springer, 2022.