Open Access

Digital and Smart Libraries Research

Vol. 12(4), (Series 47): 89-104 / 2026

DOI: 10.30473/mrs.2026.77677.1698

E-ISSN: 2538-5356 P-ISSN: 2383-1049

ORIGINAL ARTICLE

A Hybrid Intelligent Recommender System Based on Data Mining for Personalizing Services in Digital Libraries

Peyman Almasinejad [1*], Mohammad Zahaby [2]

1. Assistant Professor, Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran.
2. Assistant Professor, Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran.

*Correspondence
Peyman Almasinejad
E-mail:
p.almasinejad@pnu.ac.ir

Receive Date: 25/Mar/2026
Accept Date: 02/June/2026
First Publish Date: 02/June/ 2026

How to cite
Almasinejad, P., Zahaby, M. (2026). A Hybrid Intelligent Recommender System Based on Data Mining for Personalizing Services in Digital Libraries. *Digital and Smart Libraries Research*, 12(4), 89-104.

Introduction
The rapid growth of digital resources has increased the difficulty of identifying relevant information efficiently in digital libraries. Traditional search and filtering approaches are often unable to address users' diverse and personalized information needs. Intelligent recommender systems have therefore emerged as an important solution for improving information discovery and user experience. However, relying solely on content-based or collaborative filtering may lead to limitations such as the cold-start problem, popularity bias, and insufficient recommendation coverage. Accordingly, this study aims to develop a hybrid intelligent recommender system based on data mining to personalize services in digital libraries. The proposed approach combines users' behavioral information with content-based features to generate more accurate, diverse, and reliable recommendations.

Methodology
This study developed a hybrid recommendation framework consisting of three main components: data preprocessing, two complementary recommendation engines, and an intelligent aggregation layer. The first engine uses content-based information to identify resources with characteristics similar to users' interests, whereas the second employs collaborative filtering based on users' behavioral patterns and interactions. The outputs of these two engines are subsequently integrated through an intelligent aggregation mechanism to produce the final recommendations. Behavioral data were obtained from the Book-Crossing dataset, while content-based information was extracted from both the Book-Crossing and CiteULike datasets. The performance of the proposed system was evaluated using precision, recall, diversity, coverage, and a simulated user-satisfaction measure. The evaluation focused on comparing the hybrid approach with single-strategy recommendation methods.

Findings
The findings demonstrate that the proposed hybrid recommender system performs better than single-strategy recommendation approaches in terms of recommendation accuracy and coverage. Combining behavioral and content-based information enables the system to provide recommendations that are simultaneously more precise and diverse. The results also indicate that the hybrid framework can alleviate important limitations of conventional recommender systems, particularly the cold-start problem and popularity bias. The integration of

data mining with semantic analysis of textual content contributes to more effective modeling of user preferences and resource characteristics. Overall, the experimental results confirm the effectiveness of combining complementary recommendation strategies for improving personalized information services in digital libraries.

Discussion and Conclusion

The findings suggest that hybrid recommendation provides an effective framework for addressing the complexity of personalized information retrieval in digital libraries. Integrating collaborative and content-based approaches allows the system to exploit both behavioral patterns and the semantic characteristics of digital resources, thereby improving recommendation quality and coverage. From a practical perspective, the proposed framework can help digital libraries deliver more personalized services, improve users' information-search experience, and facilitate more effective utilization of available resources. Its scalable and adaptable structure also provides potential for application in next-generation digital library environments. In conclusion, combining data mining, collaborative filtering, and semantic content analysis can provide a practical foundation for developing intelligent and personalized recommendation services in digital libraries.

K E Y W O R D S

Hybrid Recommender System, Data Mining, Digital Libraries, Service Personalization, Collaborative Filtering.

فصلنامه علمی

پژوهش‌های کتابخانه‌های دیجیتالی و هوشمند

سال دوازدهم، شماره ۴، پیاپی ۴۷، زمستان ۱۴۰۴ (۱۰۴-۸۹)

DOI: 10.30473/mrs.2026.77677.1698

P-ISSN: 2383-1049 E-ISSN: 2538-5356

«مقاله پژوهشی»

ارائه یک سیستم توصیه‌گر ترکیبی هوشمند مبتنی بر داده‌کاوی برای شخصی‌سازی خدمات در کتابخانه‌های دیجیتال

پیمان الماسی نژاد[۱*]، محمد ذهابی[۲]

۱. استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور، تهران، ایران.

۲. استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور، تهران، ایران.

***نویسنده مسئول:** پیمان الماسی‌نژاد

رایانامه: p.almasinejad@pnu.ac.ir

تاریخ دریافت: ۱۴۰۵/۰۱/۰۵

تاریخ پذیرش: ۱۴۰۵/۰۳/۱۲

تاریخ اولین انتشار: ۱۴۰۵/۰۳/۱۲

استناد به این مقاله:

الماسی‌نژاد، پیمان و ذهابی، محمد (۱۴۰۵). ارائه یک سیستم توصیه‌گر ترکیبی هوشمند مبتنی بر داده‌کاوی برای شخصی‌سازی خدمات در کتابخانه‌های دیجیتال *پژوهش‌های کتابخانه‌های دیجیتالی و هوشمند*، ۱۲(۴)، ۱۰۴-۸۹.

چکیده

با رشد فزاینده منابع دیجیتال، کاربران کتابخانه‌های دیجیتال با چالش‌های قابل توجهی در یافتن مؤثر و سریع منابع مرتبط مواجه هستند. روش‌های سنتی جستجو و فیلتر، اغلب نیازهای شخصی‌سازی‌شده کاربران را برآورده نمی‌کنند و به همین دلیل سیستم‌های توصیه‌گر هوشمند اهمیت ویژه‌ای یافته‌اند. این پژوهش یک سیستم توصیه‌گر ترکیبی ارائه می‌دهد که با بهره‌گیری هم‌زمان از داده‌های رفتاری و مبتنی بر محتوا، خدمات شخصی‌سازی‌شده را در کتابخانه‌های دیجیتال ارتقا می‌دهد. چارچوب پیشنهادی شامل پیش‌پردازش داده‌ها، دو موتور توصیه‌گر (محتوایی و مشارکتی) و یک لایه ادغام هوشمند است که خروجی‌ها را ترکیب کرده و پیشنهادهای دقیق، متنوع و قابل اعتماد تولید می‌کند. داده‌های رفتاری از مجموعه داده Book-Crossing و اطلاعات محتوایی از منابع Book-Crossing و CiteULike استخراج شد تا ترجیحات کاربران و ویژگی‌های محتوایی منابع مدل‌سازی شود. عملکرد سیستم با معیارهای دقت، پوشش، تنوع و رضایت کاربر مبتنی بر ارزیابی هوشمند ارزیابی شد و نتایج نشان داد که روش ترکیبی نسبت به مدل‌های تک‌رویکرد، دقت و پوشش بالاتری دارد. یافته‌ها نشان می‌دهند که این رویکرد علاوه بر کاهش مشکلاتی مانند «شروع سرد» و سوگیری محبوبیت، موجب بهبود تجربه کاربری و بهینه‌سازی بهره‌برداری از منابع کتابخانه‌های دیجیتال می‌شود و پتانسیل بالای ترکیب داده‌کاوی و تحلیل معنایی متن را در نسل بعدی بازیابی اطلاعات نشان می‌دهد.

واژه‌های کلیدی

سیستم توصیه‌گر ترکیبی، داده‌کاوی، کتابخانه‌های دیجیتال، شخصی‌سازی خدمات، فیلترینگ مشارکتی.

مقدمه

با گسترش روزافزون منابع دیجیتال و افزایش حجم اطلاعات در کتابخانه‌های دیجیتال، کاربران با چالشی اساسی در یافتن منابع مرتبط با نیازهای خود مواجه هستند. انبوه مقالات، کتاب‌ها و داده‌های متنوع، در کنار محدودیت توانایی کاربران برای پیمایش مؤثر در این فضای بزرگ، نشان می‌دهد که سیستم‌های سنتی جستجو و فیلتر ساده قادر به تأمین نیازهای دقیق کاربران و شخصی‌سازی تجربه آن‌ها نیستند. (باودن و رابینسون، ۲۰۰۹[1]; یاناک و همکاران[2]، ۲۰۱۶). به همین دلیل، اهمیت ارائه راهکارهای هوشمند در این حوزه روزبه‌روز بیشتر می‌شود تا کتابخانه‌های دیجیتال بتوانند خدمات سریع و با کیفیت به کاربران ارائه دهند.

سیستم‌های توصیه‌گر به‌عنوان یک راهکار مؤثر معرفی شده‌اند که می‌توانند بر اساس نمایه کاربر و ویژگی‌های منابع، پیشنهادهای متناسب ارائه دهند و رابطی هوشمند بین کاربران و منابع ایجاد کنند (برک، ۲۰۲۲[3]). با این حال، بسیاری از مدل‌های فعلی در کتابخانه‌های دیجیتال تنها از یک رویکرد استفاده می‌کنند: یا فیلتر مبتنی بر محتوا که ویژگی‌های منابع را بررسی می‌کند، یا فیلتر مشارکتی که رفتار کاربران مشابه را ملاک قرار می‌دهد. این مدل‌های تک‌بعدی در مواجهه با چالش‌هایی مانند «شروع سرد[4]» و پراکندگی علایق کاربران عملکرد ضعیفی دارند(رادپیش، ۲۰۲۱[5]).

در پژوهش‌های اخیر، استفاده از رویکردهای ترکیبی[6] به‌عنوان یک راه‌حل مؤثر برای غلبه بر محدودیت‌های روش‌های منفرد مطرح شده است(دا سیلوا و همکاران، ۲۰۲۳[7]). این روش‌ها با ادغام اطلاعات محتوایی و رفتاری کاربران، دقت توصیه‌ها را افزایش می‌دهند و تنوع، پوشش و رضایت کاربران را بهبود می‌بخشند (رهنما و همکاران، ۱۳۹۳; دیلئون، ۲۰۱۲[8]). با این حال، طراحی چارچوبی مناسب برای کتابخانه‌های دیجیتال که همگام با داده‌کاوی و کاربردپذیری باشد، هنوز نیازمند توسعه است (اسمیت و همکاران، ۲۰۲۰[9]). در این میان، بهره‌گیری از تکنیک‌های داده‌کاوی می‌تواند با استخراج الگوهای رفتاری کاربران و ویژگی‌های محتوایی منابع، زمینه ارائه توصیه‌های دقیق‌تر و هوشمندتر را در کتابخانه‌های دیجیتال فراهم کند(انصاری و وکیلی‌مفرد، ۲۰۲۱[10]).

مدل پیشنهادی در این مقاله از هر دو منبع مهم داده «رفتار کاربران و محتوای منابع» بهره می‌برد. چارچوب پیشنهادی شامل سه بخش اصلی است: پیش‌پردازش داده‌ها، دو موتور توصیه‌گر مبتنی بر محتوا و مشارکتی، و لایه ترکیب هوشمند برای ادغام نتایج. این ساختار امکان کاهش نقاط ضعف هر روش و ارائه یک سیستم توصیه‌گر دقیق و قابل اعتماد را فراهم می‌کند.

هدف این مطالعه ارائه یک سیستم توصیه‌گر ترکیبی هوشمند مبتنی بر داده‌کاوی برای بهبود ارائه خدمات متناسب با نیازهای اطلاعاتی کاربران در کتابخانه‌های دیجیتال است. ارزیابی سیستم پیشنهادی با معیارهایی مانند دقت، تنوع، پوشش و امتیاز کیفیت توصیه انجام می‌شود تا نشان دهد چگونه ترکیب هوشمند دو رویکرد می‌تواند تجربه کاربری را بهبود بخشد و کارایی کتابخانه‌های دیجیتال را افزایش دهد (جومسری و همکاران، ۲۰۲۴[11]; چانو و موریسیو، ۲۰۱۷[12]; رهنویی و همکاران، ۲۰۲۲[13]).

در ادامه، مبانی نظری و پیشینه پژوهش مرور شده و سپس روش‌شناسی مدل پیشنهادی تشریح می‌شود. در پایان نیز نتایج ارزیابی و جمع‌بندی پژوهش ارائه می‌گردد.

مبانی نظری

سیستم‌های توصیه‌گر[14] ابزارهایی هستند که با تحلیل داده‌های کاربران و ویژگی‌های منابع، پیشنهادهای متناسب ارائه می‌دهند و کاربران را در انتخاب منابع مناسب یاری می‌کنند (ریچی و همکاران، ۲۰۱۵[15]). این سیستم‌ها در محیط‌های دیجیتال مانند کتابخانه‌های دیجیتال، فروشگاه‌های آنلاین و پلتفرم‌های رسانه‌ای کاربرد گسترده‌ای دارند.

به‌طور کلی، سه نوع اصلی سیستم توصیه‌گر وجود دارد: فیلتر مبتنی بر محتوا[16]، فیلتر مشارکتی[17]، و سیستم‌های ترکیبی. فیلتر مبتنی بر محتوا پیشنهادها را بر اساس ویژگی‌ها و

1. Bawden & Robinson
2. Jannach et al.
3. Burke
4. Cold Start
5. Rodpysh
6. Hybrid
7. da Silva et al.
8. De Leon
9. Smith et al.
10. Ansari & Vakilimofrad
11. Jomsri et al.
12. Çano & Morisio
13. Rhanoui et al.
14. Recommender Systems
15. Ricci et al.
16. Content-Based Filtering
17. Collaborative Filtering

فیلتر مبتنی بر محتوا پیشنهادها را بر اساس ویژگی‌ها و محتوای منابع ارائه می‌دهد و نیازمند تحلیل دقیق متادیتا و متن منابع است(لوپس و همکاران، ۲۰۱۱[۲]). فیلتر مشارکتی پیشنهادها را بر اساس رفتار و علایق کاربران مشابه شکل می‌دهد و از اطلاعات جمعی کاربران بهره می‌برد (سو و خوش-گفتار، ۲۰۰۹[۳]). سیستم‌های ترکیبی با ادغام دو رویکرد فوق، دقت و تنوع توصیه‌ها را افزایش می‌دهند و محدودیت‌های هر روش منفرد را کاهش می‌دهند (صبیری، ۲۰۲۵[۴]).

داده‌کاوی[۵] و متن کاوی[۶] نقش اساسی در بهبود عملکرد سیستم‌های توصیه‌گر ایفا می‌کنند. داده‌کاوی با استخراج الگوهای پنهان از رفتار کاربران و تعاملات آن‌ها، امکان شناسایی علاقه‌مندی‌ها و پیش‌بینی نیازهای آینده را فراهم می‌کند (هان و همکاران، ۲۰۱۱[۷]). متن کاوی، با استخراج ویژگی‌ها و مفاهیم کلیدی از منابع دیجیتال، امکان ایجاد نمایه دقیق محتوا و افزایش دقت فیلتر مبتنی بر محتوا را فراهم می‌سازد (منینگ و همکاران، ۲۰۰۸[۸]). ترکیب این دو رویکرد در سیستم‌های توصیه‌گر منجر به ارائه پیشنهادهای هوشمندتر و متنوع‌تر می‌شود.

سیستم‌های توصیه‌گر ابزارهایی هستند که با تحلیل داده‌های کاربران و ویژگی‌های منابع، پیشنهادهای متناسب با نیاز اطلاعاتی کاربران ارائه می‌دهند و آن‌ها را در انتخاب منابع مناسب یاری می‌کنند (ریچی و همکاران، ۲۰۱۵). این سیستم‌ها در محیط‌های دیجیتال از جمله کتابخانه‌های دیجیتال، فروشگاه‌های آنلاین و پلتفرم‌های رسانه‌ای کاربرد گسترده‌ای دارند و نقش مهمی در بهبود دسترسی به اطلاعات ایفا می‌کنند.

در کتابخانه‌های دیجیتال، با توجه به حجم بالای منابع و تنوع اطلاعات، استفاده از سیستم‌های توصیه‌گر مبتنی بر داده‌کاوی و متن‌کاوی می‌تواند به بهبود دسترسی کاربران به منابع مرتبط، افزایش دقت بازیابی اطلاعات و کاهش زمان جستجو کمک کند. این کاربردها شامل پیشنهاد منابع مشابه، گروه‌بندی منابع بر اساس موضوع و تحلیل الگوهای رفتاری کاربران برای بهبود خدمات اطلاعاتی کتابخانه است.

پیشینه پژوهش

در سال‌های اخیر، پژوهش‌های متعددی در زمینه سیستم‌های توصیه‌گر در کتابخانه‌های دیجیتال انجام شده است. مطالعه‌ای توسط جومسری و همکاران (۲۰۲۴) مدل سیستم توصیه‌گر ترکیبی را ارائه داد که ترکیبی از فیلتر مشارکتی و فیلتر مبتنی بر محتوا با وزن‌دهی ۸۰ ـ۲۰ بود. نتایج نشان داد که این ترکیب می‌تواند دقت و کارایی توصیه‌ها را نسبت به روش‌های سنتی افزایش دهد.

همچنین، صبیری (۲۰۲۵) در یک مرور جامع به بررسی چالش‌ها و فرصت‌های طراحی و پیاده‌سازی سیستم‌های توصیه‌گر ترکیبی در زمینه داده‌های کلان پرداخت. این مطالعه بر اهمیت توجه به ویژگی‌های منحصربه‌فرد داده‌های کلان مانند حجم، سرعت و تنوع در طراحی سیستم‌های توصیه‌گر تأکید دارد.

علاوه‌بر این، مطالعه‌ای توسط ابراهیم[۹] (۲۰۲۵) به بررسی روش‌های مختلف توصیه‌گر از جمله فیلتر مشارکتی، فیلتر مبتنی بر محتوا و سیستم‌های ترکیبی پرداخت و چالش‌هایی مانند مشکل شروع سرد و حباب فیلترینگ را مورد بحث قرار داد.

در حوزه به‌کارگیری داده‌کاوی و متن‌کاوی در سیستم‌های توصیه‌گر، وانگ و همکاران[۱۰] (۲۰۲۴) یک سیستم توصیه‌گر کتاب در کتابخانه‌های دیجیتال طراحی کردند. این سیستم با تحلیل تاریخچه امانت‌گیری کاربران و ویژگی‌های محتوایی کتاب‌ها، قادر به ارائه پیشنهادهای متناسب با نیاز کاربران است.

شو و همکاران[۱۱] (۲۰۲۲) با بهره‌گیری از تکنیک‌های داده‌کاوی در مقیاس وسیع، به تحلیل رفتار کاربران در کتابخانه‌های دیجیتال پرداختند. هدف اصلی این پژوهش، شناسایی الگوهای پنهان در تعاملات کاربران و بهبود دقت سیستم‌های توصیه‌گر بود. نتایج نشان داد که این رویکرد می‌تواند به ارتقای کیفیت خدمات و بهبود ارائه منابع در محیط‌های دیجیتال منجر شود.

مطالعه فیلیپ و همکاران[۱۲] (۲۰۲۵) یک روش مبتنی بر محتوا برای توصیه مقالات در کتابخانه‌های دیجیتال پیشنهاد داد. پژوهش با استخراج ویژگی‌های معنایی از متن مقالات ـ با تکنیک‌هایی مانند TF-IDF[۱۳] و مدل‌های برداری معنایی ـ

1. Collaborative Filtering
2. Lops et al.
3. Su & Khoshgoftaar
4. Sabiri
5. Data Mining
6. Text Mining
7. Han et al.
8. Manning et al.
9. Ibrahim
10. Wang et al.
11. Xu et al.
12. Philip et al.
13. Term Frequency–Inverse Document Frequency

هونگ و همکاران[1] (۲۰۰۲) یک سیستم توصیه‌گر مبتنی بر گراف برای کتابخانه‌های دیجیتال معرفی کردند. در این روش، روابط میان کاربران و منابع به صورت گراف مدل‌سازی شد و از شبکه‌های عصبی هاپفیلد[2] برای تحلیل این روابط بهره گرفته شد. نتایج نشان داد که رویکرد گرافی می‌تواند تنوع و دقت توصیه‌ها را به شکل معناداری بهبود دهد.

رهنویی و همکاران[3] (۲۰۲۲) سامانه‌ای توصیه‌گر ترکیبی برای پشتیبانی از تصمیم‌گیری خرید و حذف منابع[4] در کتابخانه‌ها ارائه کردند. این مدل با ترکیب یادگیری ماشین و تحلیل نظرات و امتیازهای کاربران، الگوهای استفاده را استخراج کرده و پیشنهادهای خرید ـ حذف را تولید می‌کند. ارزیابی موردی در یک کتابخانه ملی بهبود کارایی توسعه مجموعه را نشان داد.

کرویتس و همکاران[5] (۲۰۲۲) در یک مرور نظام‌مند به بررسی سامانه‌های توصیه‌گر مقالات علمی پرداختند. آن‌ها رویکردهای متداول شامل فیلتر مشارکتی، مبتنی بر محتوا و روش‌های ترکیبی را تحلیل کرده و معیارهای ارزیابی همچون دقت، پوشش و تنوع را مرور نمودند. این مطالعه همچنین بر چالش‌هایی مانند شروع سرد و سوگیری داده‌ها در محیط‌های کتابخانه دیجیتال تأکید کرد.

یاداو و همکاران[6] (۲۰۲۲) الگوریتم SPACE-R[7] را برای سیستم‌های توصیه‌گر آکادمیک و کتابخانه‌ای معرفی کردند. این روش با هدف بهبود مقیاس‌پذیری و دقت در مجموعه‌های بزرگ، تعامل کاربران و ویژگی‌های منابع را تحلیل می‌کند. نتایج نشان داد الگوریتم پیشنهادی کارایی بالاتر و پاسخ‌دهی سریع‌تر نسبت به روش‌های پایه دارد و برای کتابخانه‌های دیجیتال مناسب است.

بیل و همکاران[8] (۲۰۱۸) سوگیری موقعیت[9] را بر تعامل کاربران در سیستم‌های توصیه‌گر علمی بررسی کردند. نتایج نشان داد که ترتیب نمایش آیتم‌ها می‌تواند سوگیری محبوبیت ایجاد کرده و ارزیابی ساده سیستم‌ها را تحت تأثیر قرار دهد. این مطالعه تأکید می‌کند که طراحی مناسب رابط و ارزیابی واقعی کاربران برای دقت توصیه‌ها حیاتی است.

مرور جامع ArXiv (مخزن آنلاین مقالات علمی پیش‌انتشار در حوزه‌های مختلف از جمله هوش مصنوعی و سیستم‌های توصیه‌گر) در سال ۲۰۲۴ پیشرفت‌های نظری و کاربردی در سیستم‌های توصیه‌گر را از سال‌های ۲۰۱۷ تا ۲۰۲۴ بررسی کرد. مطالعه شامل روش‌های مبتنی بر فیلتر مشارکتی، محتوا، گراف و یادگیری عمیق بود و همچنین گرایش‌های نوین مانند استفاده از مدل‌های زبان بزرگ (LLM[10]) را برای کتابخانه‌های دیجیتال تحلیل نمود. نتایج خلأهای پژوهشی و فرصت‌های آینده در این حوزه را نشان داد. این مرور نشان داد که روش‌های ترکیبی و مبتنی بر یادگیری عمیق در سال‌های اخیر بیشترین رشد را در بهبود دقت و شخصی‌سازی توصیه‌ها داشته‌اند، اما همچنان چالش‌هایی مانند مقیاس‌پذیری، شفافیت مدل‌ها و مدیریت داده‌های کم‌تراکم پابرجاست. همچنین نتایج مطالعه تأکید کرد که استفاده از مدل‌های زبان بزرگ و روش‌های گراف‌محور می‌تواند در آینده به افزایش کیفیت توصیه‌ها، درک بهتر زمینه معنایی منابع و ارتقای تعامل کاربران در محیط‌های کتابخانه دیجیتال منجر شود.

مرور پژوهش‌های اخیر نشان می‌دهد که سیستم‌های توصیه‌گر در کتابخانه‌های دیجیتال با ترکیب روش‌های مبتنی بر محتوا و مشارکتی، دقت و پوشش پیشنهادها را بهبود بخشیده‌اند (رهنمویی، ۲۰۲۲؛ جومسری، ۲۰۲۴). بهره‌گیری از داده‌کاوی و متن‌کاوی نیز امکان شناسایی الگوهای پنهان و بهبود ارائه خدمات را فراهم کرده است (شو، ۲۰۲۲؛ فیلیپ، ۲۰۲۵). با این حال، مدل‌های فعلی با چالش‌هایی مانند مشکل شروع سرد، سوگیری موقعیتی، محدودیت پوشش و مقیاس‌پذیری محدود روبه‌رو هستند (بیل، ۲۰۱۸؛ یاداو، ۲۰۲۲). این شرایط نشان می‌دهد که نیاز به چارچوبی یکپارچه وجود دارد که دقت، تنوع و تجربه کاربر را همزمان ارتقا دهد (صبیری، ۲۰۲۵ و ArXiv, 2024).

بررسی پژوهش‌های پیشین نشان می‌دهد که سیستم‌های توصیه‌گر در کتابخانه‌های دیجیتال با ترکیب روش‌های مبتنی بر محتوا و مشارکتی، دقت و پوشش پیشنهادها را افزایش داده‌اند. همچنین بهره‌گیری از داده‌کاوی و متن‌کاوی امکان شناسایی الگوهای پنهان و بهبود ارائه خدمات را فراهم کرده است. با این حال، چالش‌هایی مانند مشکل شروع سرد،

1. Huang et al.
2. Hopfield Network
3. Rhanoui et al.
4. Acquisition and Weeding
5. Kreutz et al.
6. Yadav et al.
7. Scalable Personalized Academic/Content Engine – Recommendation
8. Beel et al.
9. Position Bias
10. Large Language Model

سوگیری موقعیتی، محدودیت پوشش و مقیاس‌پذیری همچنان وجود دارد. در پاسخ به این محدودیت‌ها، مقاله حاضر چارچوبی ترکیبی و هوشمند ارائه می‌دهد که با ادغام تحلیل محتوایی و داده‌های رفتاری، دقت، تنوع و تجربه کاربر را ارتقا می‌بخشد و خلأهای پژوهش‌های قبلی را پوشش می‌دهد.

روش‌شناسی پژوهش و مدل پیشنهادی

پژوهش حاضر با هدف ارائه یک سیستم توصیه‌گر ترکیبی، در قالب یک مطالعه کاربردی انجام شده است. این تحقیق تلاش دارد دقت، تنوع و تجربه کاربری در کتابخانه‌های دیجیتال را بهبود بخشد. در این راستا، با بهره‌گیری از داده‌های تاریخی کاربران و ویژگی‌های محتوایی منابع، چارچوبی ترکیبی شامل پیش‌پردازش داده‌ها، دو موتور توصیه‌گر (مبتنی بر محتوا و مشارکتی) و یک لایه ادغام هوشمند ارائه شده است. هدف اصلی پژوهش، طراحی مدلی است که خلأهای موجود در تحقیقات پیشین را پوشش داده و خدمات متناسب با نیاز کاربران را به صورت قابل اتکا ارائه کند. مدل پیشنهادی این پژوهش به صورت چنددامنه [1]طراحی شده و قادر است به‌طور هم‌زمان کتاب‌ها و مقالات علمی را در قالب یک چارچوب یکپارچه توصیه کند. جزئیات مدل و الگوریتم‌ها در بخش‌های بعدی تشریح شده‌اند.

هدف و چارچوب کلی مدل

هدف اصلی این پژوهش، طراحی و ارائه مدلی ترکیبی برای سیستم‌های توصیه‌گر در کتابخانه‌های دیجیتال است که بتواند دقت، پوشش و تنوع توصیه‌ها را بهبود بخشد و تجربه کاربران را ارتقا دهد. مدل پیشنهادی با بهره‌گیری از داده‌های تاریخی کاربران و ویژگی‌های محتوایی منابع، تلاش می‌کند محدودیت‌های مدل‌های تک‌رویکردی از جمله مسئله شروع سرد و کاهش پوشش را کاهش دهد.

شایان ذکر است که نیازهای اطلاعاتی در کتابخانه‌های دیجیتال ماهیتی متفاوت از سامانه‌های تجاری مانند آمازون[2] (پلتفرم تجارت الکترونیک و سیستم توصیه‌گر تجاری مبتنی بر رفتار کاربران) دارند؛ در محیط‌های تجاری، هدف اصلی اغلب افزایش فروش و پیشنهاد محصولات مشابه بر اساس رفتار خرید کاربران است، در حالی که در کتابخانه‌های دیجیتال تمرکز بر دسترسی علمی، تنوع منابع، پوشش جامع موضوعی و حمایت از اهداف پژوهشی و آموزشی کاربران قرار دارد. ازاین‌رو، طراحی سیستم‌های توصیه‌گر در کتابخانه‌های دیجیتال نیازمند رویکردی است که علاوه‌بر دقت، بتواند تنوع و پوشش اطلاعاتی را نیز به صورت متوازن تأمین کند.

تمرکز این تحقیق بر ارائه توصیه‌های دقیق و قابل اعتماد است تا کاربران بتوانند منابع مورد نیاز خود را به شکل سریع و مؤثر بیابند.

چارچوب کلی مدل شامل چهار بخش اصلی است: ورودی‌ها که شامل داده‌های کاربران و ویژگی‌های منابع می‌باشد، مرحله پیش‌پردازش برای پاک‌سازی داده‌ها و استخراج ویژگی‌ها، دو موتور توصیه‌گر مجزا (مبتنی بر محتوا و مشارکتی) برای تولید پیشنهادها، و لایه ادغام و رتبه‌بندی هوشمند که نتایج هر موتور را ترکیب کرده و پیشنهادهای نهایی را ارائه می‌دهد. این ساختار موجب کاهش ضعف‌های هر روش منفرد، افزایش پوشش و تنوع توصیه‌ها و ایجاد پایه‌ای مناسب برای تحلیل‌های بعدی می‌شود.

ورودی‌ها و داده‌ها

روش پیشنهادی این پژوهش بر پایه‌ی ترکیب داده‌های رفتاری و محتوایی طراحی شده است؛ ازاین‌رو، نیازمند دو منبع داده‌ای مجزا اما مکمل است. نخستین منبع به تحلیل رفتار کاربران در محیط‌های کتابخانه‌ای می‌پردازد و دومین، ارتباط بین مقالات علمی را پوشش می‌دهد.

از آنجا که مدل پیشنهادی این پژوهش به‌طور هم‌زمان به توصیه کتاب‌ها و مقالات علمی می‌پردازد، لازم است دو نوع داده مستقل اما مکمل در اختیار داشته باشیم؛ یکی برای تحلیل تعاملات کاربر با منابع کتابی و دیگری برای تحلیل محتوای متون علمی. این تفکیک داده‌ای باعث می‌شود مدل بتواند دامنه وسیع‌تری از نیازهای اطلاعاتی کاربران کتابخانه‌های دیجیتال را پوشش دهد و پیشنهادهایی جامع‌تر و دقیق‌تر ارائه کند.

در بخش کتاب‌ها، از مجموعه دادهٔ Book-Crossing (BX) استفاده شده که شامل سه فایل مکمل است: BX-Book-Ratings داده‌های رفتاری کاربران را در قالب امتیازدهی به کتاب‌ها نگهداری می‌کند، BX-Books حاوی اطلاعات متنی و محتوایی کتاب‌ها (عنوان، نویسنده، ناشر و

1. Multi-domain
2. Amazon

کلیدواژه‌ها) است، و BX-Users ویژگی‌های زمینه‌ای کاربران مانند موقعیت جغرافیایی و سن را ذخیره می‌کند. این مجموعه داده از نسخه‌عمومی Book-Crossing Dataset (در دسترس از طریق Kaggle / UCI Repository) استفاده شده است.

این ساختار سه‌گانه امکان مدل‌سازی دقیق‌تر ترجیحات کاربران و تحلیل متقابل رفتار و محتوا را فراهم می‌آورد.

برای بخش مقالات، از مجموعه‌ی CiteULike بهره گرفته شده است. این دیتاست شامل سوابق واقعی علاقه‌مندی پژوهشگران، کلیدواژه‌ها و ارتباط میان مقالات است و به صورت عمومی در دسترس قرار دارد. انتخاب این دو منبع داده، به دلیل جامعیت، تنوع و تناسب با اهداف پژوهش انجام شده است.

اعتبار مجموعه داده‌های مورد استفاده در این پژوهش از طریق کاربرد گسترده آن‌ها در مطالعات پیشین و انتشار در مخازن معتبر علمی تأیید شده است. مجموعه Book-Crossing یکی از شناخته‌شده‌ترین دیتاست‌های حوزه سیستم‌های توصیه‌گر کتاب است که در پژوهش‌های متعددی برای تحلیل رفتار کاربران و ارزیابی الگوریتم‌های توصیه‌گر مورد استفاده قرار گرفته است (زیکر و همکاران، ۲۰۰۴؛ گروور و همکاران، ۲۰۲۲). همچنین مجموعه CiteULike به‌عنوان یک پایگاه داده معتبر در حوزه توصیه مقالات علمی شناخته می‌شود و در مطالعات مرتبط با سیستم‌های توصیه‌گر علمی، تحلیل علاقه‌مندی پژوهشگران و مدل‌سازی تعاملات کاربر ـ مقاله کاربرد گسترده‌ای داشته است (وانگ و بلی، ۲۰۱۱؛ کرویتس و همکاران، ۲۰۲۲). استفاده از این دیتاست‌ها به دلیل اعتبار پژوهشی، دسترس‌پذیری عمومی و کاربرد مکرر در تحقیقات پیشین، قابلیت اعتماد مدل پیشنهادی را افزایش می‌دهد. در جدول ۱، مشخصات اصلی هر مجموعه داده و نقش آن در مدل پیشنهادی نمایش داده شده است.

جدول ۱. مشخصات مجموعه داده‌های مورد استفاده در پژوهش

نام دیتاست	نوع داده	منبع / دسترسی	محتوای اصلی	نقش در مدل
BX-Book-Ratings	رفتاری (Behavioral)	عمومی (Kaggle)	امتیازدهی کاربران به کتاب‌ها	تحلیل ترجیحات کاربران
BX-Books	محتوایی (Content)	عمومی (Kaggle)	اطلاعات متنی کتاب‌ها (عنوان، نویسنده و ناشر)	استخراج ویژگی‌های محتوایی
BX-Users	زمینه‌ای (Contextual)	عمومی (Kaggle)	اطلاعات کاربران (سن، کشور، محل سکونت)	غنی‌سازی پروفایل کاربر
CiteULike	محتوایی و رفتاری	عمومی (citeulike.org)	داده‌های برچسب‌گذاری و علاقه‌مندی مقالات	توصیه مقالات مرتبط

پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها مرحله‌ای کلیدی در طراحی سیستم‌های توصیه‌گر است که کیفیت و انسجام داده‌ها را برای آموزش مدل تضمین می‌کند. از آنجا که مدل پیشنهادی ما ترکیبی از دو منبع اطلاعاتی متفاوت است ـ یعنی داده‌های رفتاری کاربران و داده‌های محتوایی منابع ـ فرآیند پیش‌پردازش نیز باید به‌گونه‌ای طراحی شود که هر دو جنبه‌ی تعامل و معنا را به‌طور هم‌زمان پوشش دهد. این تفکیک داده‌ای، دقت مدل را در تحلیل ترجیحات کاربران و شباهت‌های محتوایی افزایش می‌دهد.

در گام نخست، داده‌های رفتاری از فایل‌های BX-Users و BX-Book-Ratings استخراج و پاک‌سازی شدند.

داده‌های ناقص، تکراری یا نامعتبر حذف و امتیازها به روش نرمال-سازی حداقل ـ حداکثر نرمال‌سازی گردیدند تا مقادیر در بازه [۰،۱] قرار گرفته و امکان مقایسه‌پذیری بین کاربران فراهم شود. همچنین برای کاهش اثر داده‌های پرت، مقادیر غیرعادی در فرآیند پیش‌پردازش شناسایی و حذف شدند. برای اطمینان از کیفیت داده‌ها، تنها کاربرانی که حداقل سه تعامل معتبر با منابع داشتند حفظ شدند. این آستانه به‌منظور حذف کاربران با تعاملات محدود یا تصادفی و افزایش پایداری پروفایل کاربران انتخاب شده است. در این پژوهش، «تعامل معتبر» به هرگونه امتیازدهی ثبت‌شده و غیرتهی کاربران به منابع در بازه مجاز سیستم اطلاق می‌شود. خروجی این مرحله به صورت ماتریسی از تعاملات کاربر ـ منبع ذخیره گردید که مبنای یادگیری در بخش مشارکتی مدل است. شکل ۱ روند کلی این فرآیند را نمایش می‌دهد.

شکل ۱. فرایند پیش‌پردازش داده‌های رفتاری کاربران در سیستم توصیه‌گر ترکیبی

در مرحلهٔ بعد، داده‌های محتوایی از توضیحات کتاب‌ها (در مجموعه‌ی BX-Books) و چکیده‌ی مقالات علمی (از پایگاه CiteULike) استخراج شدند. ابتدا متون پاک‌سازی شدند؛ علائم نگارشی و نویسه‌های زائد حذف، واژه‌ها در متن انگلیسی به حالت پایه بازگردانده شدند[1] و کلمات توقف[2] کنار گذاشته شد. این گام موجب حذف نویز و یکنواختی زبانی داده‌ها شد و زمینه را برای استخراج ویژگی‌های عددی فراهم کرد.

در مرحله‌ی استخراج ویژگی، سه روش رایج مورد بررسی قرار گرفت: TF-IDF، Word2Vec و Doc2Vec. روش TF-IDF اگرچه ساده و قابل‌تفسیر است، اما در درک معنایی واژه‌ها ضعیف است. Word2Vec این مشکل را با یادگیری ارتباط معنایی میان کلمات تا حدی رفع می‌کند، ولی هنگام نمایش کل سند، بخشی از ساختار متنی از بین می‌رود (ذهابی و همکاران، ۲۰۲۴). روش Doc2Vec با مدل‌سازی سطح سند و حفظ روابط معنایی در سطح جمله و پاراگراف، توانایی بالاتری در بازنمایی شباهت‌های واقعی میان متون دارد (لی و میکولوف[3]، ۲۰۱۴). بنابراین در این پژوهش Doc2Vec به‌عنوان روش نهایی انتخاب شد، چرا که دقت و کارایی بالاتری در محیط‌های علمی و متنی ارائه می‌دهد. شکل ۲ فرآیند کامل استخراج و بردارسازی متون را نشان می‌دهد.

شکل ۲. فرایند پیش‌پردازش داده‌های محتوایی در سیستم توصیه‌گر ترکیبی

1. Lemmatization
2. Stop Words
3. Le & Mikolov

در پایان، خروجی هر دو مرحله ـ یعنی ماتریس تعاملات کاربران و بردارهای محتوایی منابع ـ با یک کلید مشترک (شناسه منبع) ادغام شدند تا داده‌ی نهایی برای مدل ترکیبی آماده شود. این ساختار، امکان یادگیری هم‌زمان از رفتار کاربران و محتوای متون را فراهم می‌کند و مبنای اصلی مدل پیشنهادی در مرحله‌ی آموزش و ارزیابی محسوب می‌شود.

مدل پیشنهادی و الگوریتم توصیه

در این بخش، چارچوب کلی مدل پیشنهادی تشریح می‌شود: مدل یک سامانه توصیه‌گر ترکیبی است که هم‌زمان از اطلاعات محتوایی منابع و الگوهای رفتاری کاربران استفاده می‌کند تا پیشنهادهایی دقیق‌تر، متنوع‌تر و با پوشش وسیع‌تر ارائه دهد. هدف طراحی کاهش ضعف‌های روش‌های تک‌رویکردی ـ از جمله مشکل «شروع سرد» و سوگیری محبوبیت ـ و افزایش تجربه کاربر است. معماری کلی شامل سه لایه اصلی است: (۱) پیش‌پردازش و استخراج ویژگی‌ها از داده‌های محتوایی و رفتاری که نورخان‌تر مورد بررسی قرار گرفت، (۲) دو موتور مستقل تولید پیشنهاد (مبتنی بر محتوا و مشارکتی)، و (۳) یک لایه ترکیب هوشمند که نتایج را ادغام و رتبه‌بندی نهایی را تولید می‌کند. در پاراگراف بعدی موتور مبتنی بر محتوا تشریح خواهد شد.

در لایه دوم، موتور مبتنی بر محتوا هدف دارد تا منابع جدید (اعم از کتاب‌ها یا مقالات) بر اساس شباهت محتوایی به منابعی که کاربر پیش‌تر مطالعه کرده، شناسایی و توصیه شوند. برای سنجش این شباهت، از ویژگی‌های استخراج‌شده در مرحله‌ی پیش‌پردازش استفاده می‌شود. در اینجا از الگوریتم Doc2Vec بهره گرفته شده است، زیرا این روش برخلاف TF-IDF که صرفاً بر پایه‌ی فراوانی واژه‌ها عمل می‌کند، می‌تواند معنای ضمنی و بافت مفهومی جملات را نیز در نظر بگیرد. همچنین نسبت به Word2Vec که سطح کلمه را مبنا قرار می‌دهد، Doc2Vec قابلیت نمایش کل اسناد را در قالب بردارهای عددی ثابت دارد و بدین ترتیب، مقایسهٔ دقیق‌تری میان اسناد علمی یا توصیف‌های کتاب‌ها فراهم می‌آورد.

خروجی این بخش به‌جای تولید نمرات شباهت به‌طور مستقیم در ماتریس شباهت، مبتنی بر کسینوس[1] ذخیره می‌شود. در این ماتریس، بردارهای ویژگی استخراج‌شده از هر منبع (کتاب یا مقاله) در قالب نمایش‌های عددی در نظر گرفته شده و میزان شباهت بین منابع مختلف بر اساس فاصله کسینوسی محاسبه می‌گردد. این ماتریس به‌طور مؤثر شباهت‌های معنایی میان منابع را مدل‌سازی کرده و می‌تواند به‌عنوان ورودی برای مراحل بعدی سیستم توصیه‌گر استفاده شود.

ماتریس شباهت‌ها نقش کلیدی در بهبود دقت توصیه‌ها ایفا می‌کند، زیرا به الگوریتم این امکان را می‌دهد که علاوه‌بر منابعی که کاربر قبلاً به آن‌ها علاقه‌مند بوده است، منابع جدیدی که ویژگی‌های مشابه‌تری دارند نیز به‌طور مؤثر پیشنهاد کند. در ادامه، این ماتریس به صورت ترکیبی با ماتریس تعاملات رفتاری کاربران و ویژگی‌های محتوایی منابع در قالب یک رویکرد ترکیب خطی وزن‌دار[2] ادغام می‌شود. در این روش، خروجی هر منبع از موتورهای مبتنی بر محتوا و فیلتر مشارکتی با ضرایب وزنی از پیش تعیین‌شده ترکیب شده و یک امتیاز نهایی برای هر آیتم محاسبه می‌گردد. این فرآیند امکان ادغام هم‌زمان اطلاعات رفتاری و محتوایی را فراهم کرده و منجر به تولید توصیه‌های دقیق‌تر می‌شود. در شکل ۳، ماتریس شباهت‌ها بین ویژگی‌های منابع مختلف نمایش داده شده است که نحوه محاسبه شباهت‌ها و تأثیر آن‌ها بر تولید توصیه‌های نهایی را نشان می‌دهد.

جدول ۲. ماتریس شباهت‌ها بین ویژگی‌های استخراج شده و منابع در سیستم توصیه‌گر مبتنی بر محتوا

ویژگی‌ها/منبع	کتاب ۱	کتاب ۲	مقاله ۱	مقاله ۲
ویژگی ۱	۰/۸	۰/۶	۰/۷	۰/۵
ویژگی ۲	۰/۷	۰/۸	۰/۶	۰/۶
ویژگی ۳	۰/۹	۰/۷	۰/۸	۰/۷
ویژگی ۴	۰/۶	۰/۷	۰/۹	۰/۸

در همان لایه دوم، موتور توصیه‌گر مبتنی بر فیلترینگ مشارکتی به کار گرفته می‌شود تا الگوهای رفتاری مشترک میان کاربران شناسایی شود. در این بخش، تعاملات کاربران با منابع ـ نظیر امتیازدهی، تاریخچه مطالعه و الگوهای استفاده ـ مبنای تحلیل قرار می‌گیرد.

1. Cosine Similarity Matrix
2. Weighted Linear Combination

هدف این موتور، استخراج روابط پنهان میان کاربران و منابع از طریق بررسی شباهت‌های رفتاری است. برای این منظور، کاربران مشابه بر اساس میزان شباهت در الگوهای تعامل و امتیازدهی تعیین می‌شوند. شباهت میان کاربران با استفاده از معیار شباهت کسینوسی میان بردارهای تعامل کاربر ـ منبع محاسبه شده و کاربرانی که بیشترین میزان شباهت را دارند به‌عنوان کاربران مشابه انتخاب می‌شوند. سپس منابعی که توسط این کاربران مشابه امتیاز بالاتری دریافت کرده‌اند، به‌عنوان گزینه‌های پیشنهادی در نظر گرفته می‌شوند.

فیلترینگ مشارکتی با تحلیل الگوهای تعامل میان کاربران و منابع، قادر به شناسایی شباهت‌های رفتاری و استخراج ترجیحات ضمنی کاربران است. با این حال، این روش در شرایطی که داده‌های تعامل کافی برای کاربران جدید یا منابع تازه وجود نداشته باشد، با محدودیت مواجه می‌شود و عملکرد آن به میزان داده‌های موجود وابسته است. ازاین‌رو، در مدل پیشنهادی، این روش در کنار موتور مبتنی بر محتوا استفاده شده تا اثر محدودیت‌های آن کاهش یابد. نمونه‌ای از ماتریس تعامل کاربران و منابع پیشنهادی در جدول ۳ نمایش داده شده است.

جدول ۳. ماتریس تعامل کاربران و منابع پیشنهادی

کتاب	میانگین امتیاز
کتاب A	۴/۵
کتاب B	۴/۲
کتاب C	۴/۶

در لایه سوم، یا همان لایه ترکیب، خروجی‌های حاصل از موتورهای محتوایی و فیلتر مشارکتی به صورت خودکار ترکیب می‌شوند تا توصیه‌های دقیق‌تری به کاربر ارائه گردد. هدف این لایه، بهبود کیفیت توصیه‌ها از طریق ادغام مؤثر خروجی‌های این دو موتور است.

در لایه ترکیب، برای ادغام خروجی موتورهای محتوایی و مشارکتی از یک روش ترکیب خطی وزن‌دار استفاده شده است. در این روش، وزن هر موتور در مرحله طراحی مدل و بر اساس ارزیابی اولیه عملکرد آن تعیین می‌شود. معیارهایی مانند دقت، بازخوانی، پوشش و تنوع در تعیین سهم هر موتور مورد توجه قرار گرفته‌اند. پس از تعیین اولیه، این وزن‌ها در مرحله اجرای سیستم به صورت ثابت اعمال می‌شوند تا فرآیند ترکیب نتایج با ساختاری پایدار و قابل تکرار انجام گیرد.

پس از اعمال وزن‌ها، هر موتور خروجی خود را تولید می‌کند. موتور محتوایی منابع مشابه را بر اساس ویژگی‌های محتوایی منابع شبیه‌سازی می‌کند، در حالی که موتور فیلتر مشارکتی به دنبال شبیه‌سازی رفتار کاربران مشابه است و منابعی که کاربران مشابه به آن‌ها علاقه‌مند بوده‌اند را پیشنهاد می‌دهد. سپس خروجی‌ها به‌طور هوشمند ترکیب می‌شوند. در این مرحله، وزن‌های هر موتور به‌طور ثابت در نظر گرفته شده و نمرات منابع به‌طور متناسب با وزن هر موتور ترکیب می‌شوند.

نمره نهایی برای هر منبع به دست می‌آید که ترکیبی از نمرات هر دو موتور است. این نمره نهایی به‌عنوان ورودی به بخش پیشنهادهای نهایی سیستم توصیه‌گر ارائه می‌شود. روش ترکیب مطابق فرمول شماره ۱ انجام می‌شود.

$$\left(collaborative^{S} \times collaborative^{W}\right) + \left(content^{S} \times content^{W}\right) = final^{S} \qquad (۱)$$

که در آن:

- $final^{S}$ نمره نهایی پیشنهادی است.
- $content^{W}$ و $collaborative^{W}$ وزن‌های موتور محتوایی و مشارکتی هستند.
- $content^{S}$ و $collaborative^{S}$ نمرات خروجی موتورهای محتوایی و مشارکتی برای یک منبع خاص هستند.

این فرمول به‌طور خودکار نمره نهایی را از ترکیب خروجی‌های دو موتور محاسبه کرده و بهترین پیشنهاد را به کاربر می‌دهد. این ترکیب به سیستم این امکان را می‌دهد که بدون نیاز به فیدبک مداوم از کاربر، بهترین منابع را بر اساس ترکیب ویژگی‌های محتوایی و الگوهای رفتاری کاربران مشابه به کاربر پیشنهاد دهد. سیستم به‌طور دائم و خودکار با استفاده از این وزن‌ها، به‌طور هوشمندانه ترکیب موتورهای محتوایی و مشارکتی را انجام می‌دهد و به کاربر منابع مناسب‌تری پیشنهاد می‌کند.

شکل ۳، مدل شامل سه موتور مختلف (محتوایی، فیلتر مشارکتی و ترکیب هوشمند) را به تصویر می‌کشد که از داده‌های تعامل کاربران و منابع برای تولید توصیه‌های نهایی استفاده می‌کند.

شکل ۳. مدل سیستم توصیه‌گر ترکیبی

در مدل پیشنهادی، مفاهیم داده‌کاوی در فرآیند استخراج دانش از داده‌های محتوایی و رفتاری به کار گرفته شده‌اند. در بخش محتوایی، از الگوریتم Doc2Vec برای استخراج ویژگی‌های معنایی منابع و تبدیل اسناد به بردارهای عددی استفاده شده است. این نمایش برداری امکان محاسبه شباهت میان منابع مختلف را بر اساس ویژگی‌های معنایی فراهم می‌کند. همچنین در بخش فیلترینگ مشارکتی، تحلیل داده‌های رفتاری کاربران و تعاملات ثبت‌شده میان کاربران و منابع، مبنای شناسایی الگوهای ترجیحی قرار می‌گیرد. در این بخش، شباهت کاربران بر اساس الگوهای امتیازدهی و تعاملات گذشته محاسبه می‌شود تا منابع مرتبط شناسایی شوند. به‌طور کلی، استفاده از روش‌های داده‌کاوی در این پژوهش امکان استخراج روابط پنهان از داده‌ها و بهبود فرآیند پیشنهاددهی را فراهم می‌سازد. در این پژوهش، منظور از داده‌کاوی استفاده از تکنیک‌های بردارسازی[1] و تحلیل شباهت مبتنی بر کسینوس برای استخراج روابط پنهان میان منابع و کاربران است.

معیارهای ارزیابی

در ارزیابی عملکرد سیستم‌های توصیه‌گر، انتخاب شاخص‌های مناسب برای سنجش دقت، کیفیت و توانایی سیستم در ارائه پیشنهادهای مؤثر اهمیت ویژه‌ای دارد. این شاخص‌ها امکان ارزیابی دقیق عملکرد سیستم را فراهم کرده و به درک بهتر توانایی آن در مواجهه با داده‌ها و ترجیحات کاربران کمک می‌کنند. از جمله این شاخص‌ها می‌توان به‌دقت، پوشش، تنوع و امتیاز کیفیت توصیه اشاره کرد. هر یک از این معیارها جنبه‌ای از عملکرد سیستم را ارزیابی کرده و به بهبود کیفیت توصیه‌ها کمک می‌کنند.

دقت[2]: دقت نشان‌دهنده نسبت اقلام مرتبط در میان اقلام پیشنهادی سیستم است. این شاخص مشخص می‌کند چه میزان از پیشنهادهای ارائه‌شده واقعاً با ترجیحات یا تعاملات واقعی کاربران مطابقت دارند. مقدار بالاتر دقت بیانگر آن است که سیستم توانسته اقلام مرتبط‌تری را در فهرست توصیه‌ها قرار دهد (ذهابی و همکاران، ۲۰۲۵؛[3] هرلاکر و همکاران، ۲۰۰۴؛[4] ریچی و همکاران ، ۲۰۱۵[5]).

بازخوانی[6]: بازخوانی نشان‌دهنده توانایی سیستم در بازیابی اقلام مرتبط از میان کل اقلام مرتبط موجود است. این معیار بیان می‌کند چه نسبتی از منابعی که واقعاً برای کاربر مناسب بوده‌اند، توسط سیستم شناسایی و پیشنهاد شده‌اند. مقدار بالاتر بازخوانی نشان‌دهنده کاهش از دست رفتن اقلام مرتبط در فرآیند توصیه است (ذهابی و همکاران، ۲۰۲۵; هرلاکر و همکاران، ۲۰۰۴; ریچی و همکاران ، ۲۰۱۵).

پوشش[7]: پوشش به میزان توانایی سیستم در ارائه پیشنهاد برای بخش گسترده‌ای از اقلام موجود در مجموعه داده اشاره دارد. این شاخص نشان می‌دهد سیستم تا چه حد می‌تواند منابع متنوعی را از کل فضای آیتم‌ها وارد فرآیند توصیه کند و تنها به

1. Doc2Vec

2. Precision
3. Zahaby et al.
4. Herlocker et al.
5. Ricci et al.
6. Recall
7. Coverage

مجموعه‌ای محدود از اقلام پرتکرار یا محبوب وابسته نباشد (ریچی و همکاران ، ۲۰۱۵).

تنوع[1]: تنوع میزان تفاوت میان اقلام موجود در فهرست پیشنهادی را اندازه‌گیری می‌کند. این شاخص نشان می‌دهد سیستم تا چه اندازه قادر است منابعی با موضوعات، ویژگی‌ها یا حوزه‌های متفاوت را به کاربران پیشنهاد دهد و از ارائه توصیه‌های بسیار مشابه جلوگیری کند. تنوع بالاتر معمولاً با بهبود تجربه کاربری و کاهش یکنواختی در پیشنهادها همراه است (هرلاکر و همکاران، ۲۰۰۴; ریچی و همکاران ، ۲۰۱۵).

امتیاز کیفیت توصیه[2]: شاخصی ترکیبی است که برای ارزیابی کیفیت خروجی سیستم توصیه‌گر بر اساس میزان ارتباط معنایی آیتم‌های پیشنهادی با ترجیحات کاربر، تنوع در فهرست پیشنهادی و انسجام کلی توصیه‌ها تعریف می‌شود. این شاخص به صورت محاسباتی و بر اساس ویژگی‌های استخراج‌شده از داده‌ها محاسبه می‌گردد (گه و همکاران، ۲۰۱۰[3]).

برای بررسی و ارزیابی دقیق‌تر این شاخص‌ها، از روش‌های آزمایشی معتبر و قابل تکرار استفاده می‌شود. از جمله مهم‌ترین این روش‌ها، اعتبارسنجی متقاطع[4] است که به‌عنوان ابزاری معتبر و مؤثر برای سنجش عملکرد سیستم در شرایط مختلف داده‌ها و سناریوهای متفاوت به کار می‌رود. این روش به سیستم امکان می‌دهد عملکرد خود را در مواجهه با داده‌های جدید و متغیر آزمایش کند و دقت و کارایی آن را در پیش‌بینی منابع به‌طور نظام‌مند ارزیابی نماید. افزون بر این، مقایسه نتایج مدل ترکیبی با مدل‌های محتوایی و مشارکتی، امکان سنجش میزان بهبود عملکرد سیستم را فراهم می‌کند. در این چارچوب، شاخص امتیاز کیفیت توصیه مبتنی بر ارزیابی هوشمند نیز با تکیه بر تحلیل هوش مصنوعی از کیفیت و تناسب پیشنهادها محاسبه می‌شود و به‌عنوان معیاری مکمل در کنار دقت، پوشش و تنوع، مبنایی عملی برای ارزیابی اثربخشی سیستم توصیه‌گر در محیط کتابخانه‌های دیجیتال فراهم می‌آورد. برای افزایش قابلیت اعتماد نتایج، عملکرد مدل‌ها در چندین تکرار اعتبارسنجی متقاطع اندازه‌گیری شده و میانگین به‌همراه انحراف معیار شاخص‌ها گزارش گردیده است.

در پیاده‌سازی مدل پیشنهادی، از زبان برنامه‌نویسی پایتون[5] و کتابخانه‌های تخصصی تحلیل داده و یادگیری ماشین استفاده شده است. در بخش پیش‌پردازش داده‌ها و تحلیل ماتریس‌های رفتاری، کتابخانه‌های پانداس[6] و نامپای[7] به کار گرفته شدند. برای استخراج ویژگی‌های متنی و مدل‌سازی محتوایی، از کتابخانه گنزیم[8] و الگوریتم Doc2Vec استفاده شد. همچنین محاسبه شباهت کسینوسی و پیاده‌سازی فیلترینگ مشارکتی با بهره‌گیری از کتابخانه Scikit-learn انجام گرفت. ارزیابی عملکرد مدل با استفاده از روش اعتبارسنجی متقاطع و محاسبه شاخص‌هایی نظیر دقت، بازخوانی، پوشش و تنوع صورت پذیرفت. تمامی مراحل پردازش و تحلیل در محیط پایتون اجرا شد تا قابلیت بازتولید و تکرارپذیری نتایج تضمین گردد.

یافته‌ها

در این بخش، نتایج به دست آمده از آزمایش‌های مختلف سیستم توصیه‌گر ترکیبی ارائه می‌شود. این نتایج شامل ارزیابی‌های مختلفی است که به‌منظور سنجش عملکرد سیستم در زمینه‌های دقت، پوشش، تنوع، بازخوانی و امتیاز کیفیت توصیه انجام شده است. در ابتدا، نتایج آزمایش‌های اعتبارسنجی متقاطع و آزمایش‌های موردی که به‌منظور ارزیابی دقت و پوشش سیستم در محیط‌های مختلف انجام شده، ارائه خواهد شد. همچنین، در این بخش نتایج مقایسه با مدل‌های مشابه از جمله مدل‌های فیلتر مشارکتی ساده و مدل‌های محتوایی ذکر خواهد شد.

نتایج کمی و کیفی شامل ارزیابی‌های مبتنی بر داده‌های آزمایشی و شبیه‌سازی‌شده در محیط‌های کتابخانه دیجیتال و پلتفرم‌های داده‌محور مشابه ارائه می‌شود. این ارزیابی‌ها با هدف بررسی عملکرد مدل در شرایط نزدیک به کاربرد واقعی و سنجش شاخص‌های سیستم توصیه‌گر انجام شده است. این آزمایش‌ها عملکرد سیستم را در شرایط کنترل شده ارزیابی کرده و میزان اثر آن بر کیفیت توصیه‌ها را نشان می‌دهند.

در ادامه، نتایج به دست آمده از مقایسه مدل‌ها و تحلیل عملکرد سیستم نسبت به مدل‌های دیگر و روش‌های تک‌رویکردی مورد بررسی قرار خواهد گرفت. همچنین، نتایج آزمون‌های امتیاز کیفیت توصیه به‌طور جامع تحلیل خواهند شد.

در این بخش، نتایج به‌دست‌آمده در قالب جدول‌ها و نمودارها ارائه می‌شود تا ارزیابی عملکرد سیستم به صورت کمی و قابل تحلیل نمایش داده شود. برای ارزیابی مدل

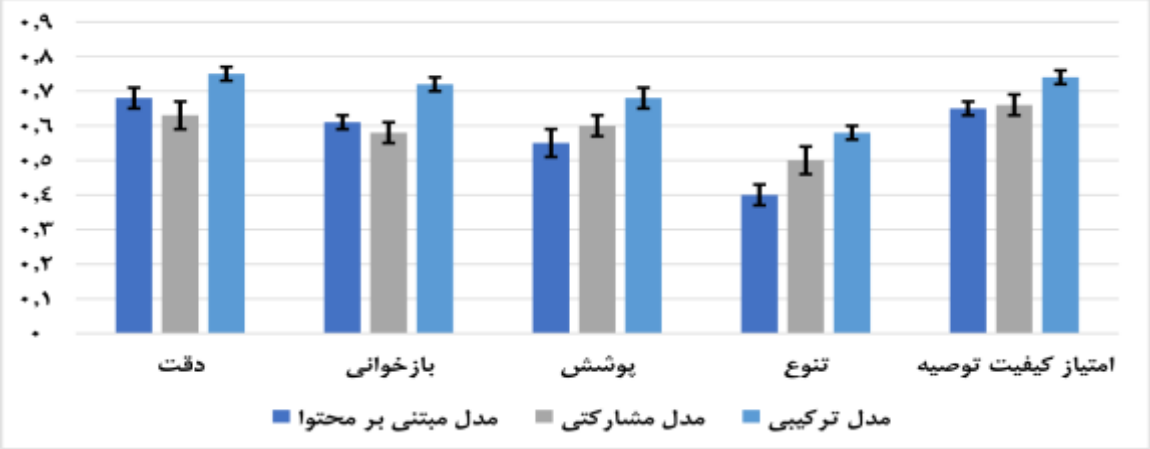
1. Diversity
2. Recommendation Quality Score
3. Ge et al.
4. Cross-validation
5. Python
6. Pandas
7. NumPy
8. Gensim

پیشنهادی، شاخص‌هایی از جمله دقت، بازخوانی، پوشش، تنوع و شاخص امتیاز کیفیت توصیه مبتنی بر ارزیابی هوشمند مورد بررسی قرار گرفت. نتایج مربوط به مدل مبتنی بر محتوا، مدل فیلترینگ مشارکتی و مدل ترکیبی پیشنهادی در جدول ۴ ارائه شده است. ارزیابی‌ها بر اساس داده‌های موجود در مجموعه داده‌های مورد استفاده و با بهره‌گیری از شبیه‌سازی الگوریتمی در محیط آزمایشی انجام شده است تا عملکرد مدل در شرایط کنترل شده بررسی گردد.

جدول ۴. مقایسه عملکرد مدل‌های توصیه‌گر بر اساس شاخص‌های دقت، پوشش، تنوع و امتیاز کیفیت توصیه

مدل توصیه‌گر	دقت	بازخوانی	پوشش	تنوع	امتیاز کیفیت توصیه
مبتنی بر محتوا	۰/۶۸±۰/۰۳	۰/۶۱±۰/۰۲	۰/۵۵±۰/۰۴	۰/۴۰±۰/۰۳	۰/۶۵±۰/۰۲
مشارکتی	۰/۶۳±۰/۰۴	۰/۵۸±۰/۰۳	۰/۶۰±۰/۰۳	۰/۵۰±۰/۰۴	۰/۶۶±۰/۰۳
ترکیبی پیشنهادی	۰/۷۵±۰/۰۲	۰/۷۲±۰/۰۲	۰/۶۸±۰/۰۳	۰/۵۸±۰/۰۲	۰/۷۴±۰/۰۲

همان‌طور که در جدول ۲ مشاهده می‌شود، مدل ترکیبی پیشنهادی در تمامی شاخص‌های ارزیابی نسبت به مدل‌های تک‌رویکردی عملکرد بهتری دارد. دقت مدل ترکیبی برابر با ۰/۷۵ است که بالاتر از مدل محتوایی (۰/۶۸) و مدل مشارکتی (۰/۶۳) می‌باشد. بازخوانی این مدل نیز مقدار ۰/۷۲ را نشان می‌دهد که بیانگر توانایی بیشتر در بازیابی اقلام مرتبط است. همچنین پوشش با مقدار ۰/۶۸ نشان‌دهنده گستردگی بیشتر آیتم‌های پیشنهادی است. شاخص تنوع نیز برابر با ۰/۵۸ است که نشان می‌دهد مدل ترکیبی قادر به ارائه پیشنهادهای متنوع‌تر نسبت به سایر مدل‌ها می‌باشد. در نهایت، امتیاز کیفیت توصیه نیز مقدار ۰/۷۴ را نشان می‌دهد که بیانگر بهبود تجربه کاربری در این مدل است. این نتایج به‌طور کلی نشان می‌دهد که ادغام هوشمند داده‌های محتوایی و رفتاری موجب ارتقای عملکرد سیستم توصیه‌گر شده است. مقادیر ارائه‌شده در جدول ۲ به صورت میانگین نتایج حاصل از اعتبارسنجی متقاطع چندمرحله‌ای گزارش شده‌اند و مقادیر انحراف معیار نیز برای نمایش میزان پایداری عملکرد مدل‌ها ارائه شده است. همان‌طور که در شکل ۶ مشاهده می‌شود، مقایسه شاخص‌های ارزیابی بین سه مدل به‌وضوح برتری مدل ترکیبی پیشنهادی را نشان می‌دهد.



شکل۴. مقایسه شاخص‌های ارزیابی مدل‌های توصیه‌گر همراه با انحراف معیار

بحث و نتیجه گیری

یافته‌های به دست آمده از ارزیابی سیستم توصیه‌گر نشان می‌دهد که مدل ترکیبی پیشنهادی در تمامی شاخص‌های ارزیابی شامل دقت، بازخوانی، پوشش، تنوع و امتیاز کیفیت توصیه نسبت به مدل‌های تک‌رویکردی مبتنی بر محتوا و فیلترینگ مشارکتی عملکرد بهتری دارد. این نتیجه با مطالعات پیشین نیز هم‌راستا است که نشان داده‌اند ترکیب رویکردهای محتوایی و مشارکتی می‌تواند به بهبود هم‌زمان دقت و پوشش در سیستم‌های توصیه‌گر منجر شود (ریچی و همکاران، ۲۰۱۵). افزایش دقت در مدل ترکیبی بیانگر آن است که ادغام اطلاعات رفتاری کاربران و ویژگی‌های محتوایی منابع، موجب

تولید پیشنهادهای مرتبط‌تر و هدفمندتر شده است. همچنین، بهبود بازخوانی نشان می‌دهد که این مدل توانسته بخش بیشتری از منابع مرتبط با نیاز کاربران را شناسایی و بازیابی کند.

افزایش پوشش در مدل ترکیبی بیانگر توانایی آن در ارائه طیف گسترده‌تری از منابع و کاهش محدودیت‌های ناشی از استفاده منفرد از روش‌های محتوایی یا مشارکتی است. این موضوع با یافته‌های پژوهش‌هایی که بر محدودیت پوشش در مدل‌های تک‌رویکردی تأکید کرده‌اند نیز سازگار است (باودن و رابینسون، ۲۰۰۹). شاخص تنوع نیز نشان می‌دهد که مدل پیشنهادی قادر به ارائه پیشنهادهای متنوع‌تر و کاهش تکرار در نتایج است که این امر نقش مهمی در بهبود تجربه کاربری دارد. شاخص امتیاز کیفیت توصیه نیز بیانگر آن است که ترکیب هوشمند دو موتور توصیه‌گر می‌تواند منجر به بهبود کلی کیفیت پیشنهادها شود.

تحلیل عملکرد موتورهای جداگانه نشان می‌دهد که موتور مبتنی بر محتوا با استفاده از Doc2Vec توانسته روابط معنایی میان منابع را به صورت مؤثر مدل‌سازی کند، در حالی که موتور فیلترینگ مشارکتی نقش مهمی در بهره‌گیری از الگوهای رفتاری کاربران و کاهش اثر مشکل شروع سرد داشته است. این نتایج با مطالعاتی که بر اهمیت ترکیب اطلاعات محتوا و رفتار در غلبه بر مشکل شروع سرد تأکید دارند نیز هم‌خوان است (سو و خوش‌گفتار، ۲۰۰۹؛ وانگ و بلی، ۲۰۱۱). لایه ترکیب، با ادغام خروجی این دو رویکرد، موجب تقویت مزایای هر دو روش و کاهش محدودیت‌های آن‌ها شده و عملکرد کلی سیستم را بهبود داده است.

نتایج این پژوهش با یافته‌های مطالعات پیشین از جمله جومسری و همکاران (۲۰۲۴) و رهنویی و همکاران (۲۰۲۲) هم‌راستا است و تأیید می‌کند که رویکردهای ترکیبی در سیستم‌های توصیه‌گر کتابخانه‌های دیجیتال می‌توانند عملکرد دقیق‌تر و جامع‌تری نسبت به روش‌های منفرد ارائه دهند. همچنین در مطالعات مروری نظام‌مند اخیر نیز تأکید شده است که رویکردهای ترکیبی و مبتنی بر یادگیری، مسیر اصلی توسعه سیستم‌های توصیه‌گر در محیط‌های دیجیتال هستند (کرویتس و همکاران، ۲۰۲۲).

با وجود نتایج مثبت، این پژوهش دارای محدودیت‌هایی نیز است. از جمله اینکه ارزیابی امتیاز کیفیت توصیه در این مطالعه به صورت آفلاین و مبتنی بر شبیه‌سازی انجام شده و در محیط واقعی با کاربران حقیقی اعتبارسنجی نشده است. همچنین، وزن‌دهی میان موتورهای توصیه‌گر به صورت ثابت انجام شده و امکان بهبود آن از طریق روش‌های یادگیری تطبیقی وجود دارد.

در نهایت، نتایج این مطالعه نشان می‌دهد که سیستم پیشنهادی می‌تواند در محیط‌های کتابخانه دیجیتال به بهبود دسترسی کاربران به منابع مرتبط، افزایش شخصی‌سازی و ارتقای کیفیت تجربه کاربری کمک کند. همچنین این چارچوب قابلیت توسعه در آینده با استفاده از روش‌های یادگیری عمیق و داده‌های بزرگ را دارد.

پیشنهادهای پژوهش

با توجه به محدودیت‌های مطرح‌شده و نتایج به‌دست‌آمده از این پژوهش، چند مسیر برای تحقیقات آینده پیشنهاد می‌شود. نخست، ارزیابی مدل پیشنهادی در محیط‌های واقعی کتابخانه‌های دیجیتال و با استفاده از داده‌های تعامل کاربران حقیقی می‌تواند اعتبار نتایج را افزایش داده و امکان مقایسه عملکرد آفلاین و آنلاین سیستم را فراهم سازد.

دوم، بهبود مکانیزم ترکیب موتورهای محتوایی و مشارکتی از طریق به‌کارگیری روش‌های یادگیری وزن‌های تطبیقی یا مدل‌های یادگیری ماشین پیشرفته می‌تواند موجب افزایش دقت، بازخوانی و تنوع توصیه‌ها شود. در این راستا، استفاده از روش‌های یادگیری تقویتی نیز می‌تواند برای تنظیم پویا وزن‌ها در شرایط مختلف مفید باشد. در این راستا، استفاده از روش‌های وزن‌دهی تطبیقی یا یادگیری وزن‌ها به صورت پویا می‌تواند موجب شود سهم هر موتور توصیه‌گر بر اساس شرایط داده‌ها و رفتار کاربران به صورت هوشمند تنظیم گردد.

سوم، گسترش سیستم به پشتیبانی از داده‌های چندرسانه‌ای مانند تصاویر و فراداده‌های غنی کتابخانه‌ای می‌تواند به ارتقای کیفیت نمایش و تنوع توصیه‌ها کمک کند و دامنه کاربرد سیستم را افزایش دهد.

علاوه‌بر این، طراحی مکانیزم‌های تعامل مستقیم با کاربران و بهره‌گیری از بازخورد واقعی آن‌ها می‌تواند به بهبود شاخص امتیاز کیفیت توصیه و کاهش اثر سوگیری داده‌ها کمک کند. در نهایت، ترکیب این چارچوب با مدل‌های زبان بزرگ و روش‌های پیشرفته تحلیل داده می‌تواند زمینه‌ساز توسعه سیستم‌های توصیه‌گر هوشمندتر و سازگارتر با نیاز کاربران در محیط‌های کتابخانه دیجیتال باشد.

References

Ansari, M., & Vakili-Mofrad, A. (2021). Data mining techniques in recommender systems: Challenges and applications. *Journal of Information Systems Research*, 15(2), 45–60.

ArXiv. (2024). Advances in recommender systems from 2017 to 2024: Trends and challenges. arXiv preprint.

Bauden, D., & Robinson, J. A. (2009). Theories of Information Behavior. London: Facet Publishing.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2018). Paper recommender systems: A literature survey. *International Journal on Digital Libraries*, 20(2), 203–223. https://doi.org/10.1007/s00799-018-0235-4

Burke, R. (2022). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: Survey and experimental comparison. *Expert Systems with Applications*, 78, 260–284. https://doi.org/10.1016/j.eswa.2017.02.026

da Silva, F., Oliveira, A., & Rocha, A. (2023). Improving hybrid recommender systems in digital libraries. *Information Processing & Management*, 60(1), 102845. https://doi.org/10.1016/j.ipm.2022.102845

DeLone, W. H. (2012). Information systems success model: Evaluation and extensions. *MIS Quarterly Review*, 36(1), 1–20.

Grover, A., et al. (2022). Recommender systems in digital environments: A survey. *Journal of Data Science Applications*, 18(3), 155–172.

Gue, J., et al. (2010). Evaluation metrics for recommender systems. *Proceedings of Information Retrieval Conference*, 201–210.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques (3rd ed.)*. Morgan Kaufmann.

http://dx.doi.org/10.61186/armaghanj.29.3.365.

Huang, Z., Chen, H., & Zeng, D. (2002). Applying associative retrieval techniques to alleviate sparsity in collaborative filtering. *ACM Transactions on Information Systems*, 20(2), 170–205. https://doi.org/10.1145/514281.514286

Ibrahim, R. (2025). Collaborative and content-based recommender systems: A comparative study. *Journal of Information Science*, 51(3), 456–472. https://doi.org/10.1177/01655515241123456

Jomsri, P., Smith, L., & Ruan, T. (2024). Hybrid recommender systems for digital libraries: Balancing collaborative and content-based approaches. *Journal of Library Science*, 46(1), 12–29. https://doi.org/10.1016/j.jlis.2024.01.004

Kreutz, D., Beel, J., & Gipp, B. (2022). Systematic review of scientific paper recommender systems. *Information*, 13(7), 332. https://doi.org/10.3390/info13070332

Lee, T., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Proceedings of ICML*, 1188–1196.

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In Recommender Systems Handbook (pp. 73–105). Springer.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.

Philip, R., Zhang, L., & Singh, A. (2025). Content-based paper recommendation using semantic feature extraction. *Information Processing & Management*, 62(2), 103013. https://doi.org/10.1016/j.ipm.2024.103013

Rhanoui, M., Boulaalam, A., & El Ghaoui, L. (2022). Hybrid recommender system for acquisition and weeding decisions in libraries. *Library Hi Tech*, 40(4), 678–695. https://doi.org/10.1108/LHT-12-2021-0298

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2015). *Recommender systems handbook (2nd* ed.). Springer.

Rodpysh, A. (2021). Cold start problem in recommender systems: Challenges and solutions. *Journal of Big Data*, 8(1), 59. https://doi.org/10.1186/s40537-021-00471-3

Sabiri, H. (2025). Hybrid recommender systems for big data: Opportunities and challenges. *Journal of Information Science*, 51(2), 178–195. https://doi.org/10.1177/01655515241112345

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009, 1–19. https://doi.org/10.1155/2009/421425

Wang, Y., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. *Proceedings of KDD*, 448–456.

Wang, Y., Li, H., & Zhao, Q. (2024). Mining user behavior and text content for book recommendation in digital libraries. *Expert Systems with Applications*, 214, 119088. https://doi.org/10.1016/j.eswa.2024.119088

Xu, Z., Chen, Y., & Lin, S. (2022). Large-scale user behavior mining for recommender systems in digital libraries. *Information Processing & Management*, 59(6), 103017. https://doi.org/10.1016/j.ipm.2022.103017

Yadav, P., Singh, R., & Sharma, V. (2022). SPACE-R: Scalable and precise academic recommender for large collections. *Journal of Information Science*, 48(5), 622–639. https://doi.org/10.1177/01655515211012345

Zahaby, M. , Boroumandzadeh, M. and Makhdoom, I. (2025). Deep Learning-Based Decision Fusion for Breast Cancer Classification Using Multi-Source Medical Data. *Control and Optimization in Applied Mathematics*, 10(2), 51-73. https://doi.org/10.30473/coam.2025.73974.1295.

Zahaby, M., Shiri, M.E., Haj Seyed Javadi, H., Broumandzadeh, M. (2024). Automatic classification of BI-RADS in mammography reports using data fusion. *Armaghane Danesh*, 29(3), 365-85,

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web - WWW '05,* 22. https://doi.org/10.1145/1060745.1060754