Sistem Rekomendasi Buku Untuk Perpustakaan SDN 2 Ganungkidul Menggunakan Metode *Content-Based Filtering*

1st Alfan Didan Septiandri Argandi
S1 Teknik Informatika
Telkom University
Purwokerto, Indonesia
alfandidan@student.telkomuniversity.ac.id

2nd Paradise
S1 Teknik Informatika
Telkom University
Purwokerto, Indonesia
paradise@telkomuniversity.ac.id

3rd Nicolaus Euclides Wahyu Nugroho
S1 Teknik Informatika
Telkom University
Purwokerto, Indonesia
nicolausn@telkomuniversity.ac.id

***Abstrak* — Penelitian ini bertujuan untuk merancang dan mengembangkan sistem rekomendasi buku berbasis Content-based Filtering guna mendukung layanan peminjaman buku di Perpustakaan SDN 2 Ganungkidul Nganjuk, serta mengevaluasi kinerja sistem rekomendasi yang dikembangkan berdasarkan tingkat akurasi prediksi. Sistem rekomendasi dikembangkan menggunakan algoritma TF-IDF untuk pembobotan teks dan cosine similarity untuk mengukur kemiripan antar buku berdasarkan metadata, seperti judul, kategori, dan deskripsi singkat. Implementasi sistem dilakukan dengan arsitektur client–server, di mana backend dikembangkan menggunakan TypeScript dengan Hono.js dan dijalankan melalui Bun Runtime, sedangkan antarmuka pengguna dibangun menggunakan Vue.js dan Tailwind CSS. Dataset yang digunakan berasal dari koleksi buku Perpustakaan SDN 2 Ganungkidul yang telah diolah dalam format terstruktur. Evaluasi kinerja sistem dilakukan melalui pengujian langsung terhadap 19 pengguna dengan menggunakan metrik Normalized Discounted Cumulative Gain (NDCG@10) dan Precision@10. Hasil evaluasi menunjukkan nilai rata-rata NDCG@10 sebesar 0,61 dan Precision@10 sebesar 0,65. Hasil tersebut mengindikasikan bahwa sistem mampu menempatkan buku yang relevan pada posisi peringkat atas rekomendasi serta memberikan tingkat ketepatan rekomendasi yang cukup baik.*

***Kata kunci* — sistem rekomendasi, content-based filtering, TF-IDF, cosine similarity**

I. PENDAHULUAN

Kemampuan literasi membaca merupakan indikator fundamental dalam menilai kualitas pendidikan. Laporan Programme for International Student Assessment (PISA) menunjukkan bahwa tingkat literasi membaca siswa di Indonesia masih berada di bawah rata-rata negara OECD, yang menandakan perlunya upaya sistematis untuk meningkatkan minat dan kemampuan membaca sejak jenjang pendidikan dasar. Salah satu sarana strategis dalam mendukung literasi adalah perpustakaan sekolah, yang menyediakan akses terhadap sumber bacaan yang beragam dan relevan dengan kebutuhan siswa.

Namun, pemanfaatan perpustakaan sekolah pada umumnya masih dilakukan secara konvensional. Proses pencarian dan pemilihan buku bergantung pada inisiatif pengguna tanpa adanya dukungan sistem yang mampu memberikan rekomendasi secara personal. Kondisi ini berpotensi menyebabkan ketidaksesuaian antara minat siswa dan buku yang dipilih, sehingga menurunkan minat baca dan efektivitas layanan perpustakaan.

Perkembangan teknologi informasi memungkinkan penerapan sistem rekomendasi sebagai solusi untuk mempersonalisasi layanan informasi. Sistem rekomendasi telah banyak digunakan pada berbagai domain untuk mengatasi permasalahan information overload dengan menyarankan item yang relevan berdasarkan preferensi pengguna. Dalam konteks perpustakaan sekolah, sistem rekomendasi buku diharapkan dapat membantu siswa menemukan bacaan yang sesuai dengan minat dan tingkat pemahamannya.

Penelitian ini mengusulkan pengembangan sistem rekomendasi buku berbasis Content-Based Filtering untuk Perpustakaan SDN 2 Ganungkidul. Pendekatan ini dipilih karena mampu memberikan rekomendasi berdasarkan kemiripan konten buku tanpa bergantung pada data pengguna lain, sehingga sesuai diterapkan pada lingkungan dengan jumlah pengguna dan data interaksi yang terbatas. Kontribusi utama penelitian ini terletak pada penerapan Content-Based Filtering berbasis TF-IDF dan Cosine Similarity pada konteks perpustakaan sekolah dasar serta evaluasi kinerjanya menggunakan metrik peringkat rekomendasi.

II. KAJIAN TEORI

A. Sistem Rekomendasi

Sistem rekomendasi merupakan suatu perangkat lunak yang dirancang untuk memberikan prediksi item kepada pengguna berdasarkan informasi yang tersedia. Tujuan utama dari sistem ini adalah untuk mengatasi permasalahan kelebihan informasi (*information overload*) dengan mempersonalisasi pengalaman pengguna secara akurat serta menyarankan item atau produk yang sesuai dengan preferensi pengguna[6][7].

Sistem rekomendasi telah banyak diimplementasikan dalam berbagai domain, terutama pada platform *e-commerce*, guna menyarankan produk yang relevan kepada pengguna. Selain itu, sistem ini juga diterapkan secara luas pada layanan hiburan digital, seperti Netflix, YouTube, Spotify, dan platform sejenis lainnya. Berikut ini adalah beberapa pendekatan umum yang digunakan dalam pengembangan sistem rekomendasi.

Content-based Filtering merupakan pendekatan dalam sistem rekomendasi yang memanfaatkan minat atau preferensi personal pengguna dengan mengidentifikasi pola dari riwayat interaksi pengguna terhadap item, seperti item

yang pernah diakses atau disukai[8]. Berdasarkan pola tersebut, sistem kemudian merekomendasikan item lain yang memiliki kemiripan dengan item yang sebelumnya diminati oleh pengguna. Pendekatan ini bersifat *user-independent*, karena sistem hanya memerlukan data dari satu pengguna tanpa bergantung pada informasi dari pengguna lain[9].

B. Arsitektur Sistem

Sistem rekomendasi dikembangkan dengan arsitektur client–server. Backend bertanggung jawab terhadap pemrosesan data, perhitungan TF-IDF, dan penghitungan kemiripan antar buku, sedangkan frontend berfungsi sebagai antarmuka interaksi pengguna. Backend diimplementasikan menggunakan TypeScript dengan framework Hono.js dan dijalankan pada Bun Runtime, sementara frontend dibangun menggunakan Vue.js dan Tailwind CSS.

C. Dataset

Dataset penelitian berasal dari koleksi buku Perpustakaan SDN 2 Ganungkidul yang telah disusun dalam format terstruktur. Data buku mencakup atribut judul, kategori, dan deskripsi singkat. Data ini digunakan sebagai dasar pembentukan representasi konten buku dalam sistem rekomendasi.

D. Pemrosesan Teks dan Representasi Fitur

Tahapan pemrosesan teks meliputi *lowercasing*, penghapusan tanda baca, tokenisasi, penghapusan *stop words*, dan lemmatisasi. Hasil pemrosesan teks selanjutnya direpresentasikan dalam bentuk vektor menggunakan metode TF-IDF. Pembobotan TF-IDF digunakan untuk menilai tingkat kepentingan setiap term dalam dokumen relatif terhadap keseluruhan korpus buku.

E. Pengukuran Kemiripan Fitur

Tingkat kemiripan antar buku dihitung menggunakan *Cosine Similarity* terhadap vektor TF-IDF. Nilai kemiripan digunakan untuk menentukan daftar buku yang paling relevan terhadap buku acuan atau riwayat bacaan pengguna.

F. Evaluasi Sistem

Evaluasi kinerja sistem dilakukan secara offline menggunakan metrik *Precision@K* dan *Normalized Discounted Cumulative Gain* (NDCG@K). Pengujian dilakukan terhadap 19 pengguna dengan memanfaatkan data histori bacaan sebagai *ground truth* untuk menilai relevansi rekomendasi yang dihasilkan.

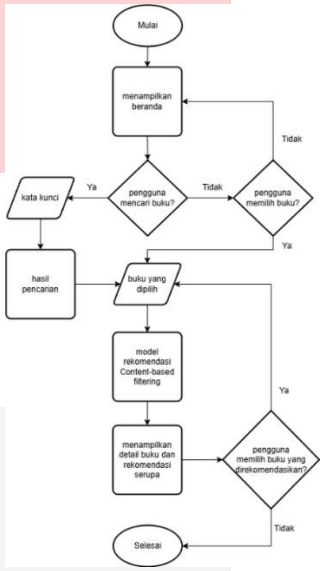
III. METODE

Penelitian ini menggunakan pendekatan eksperimental dengan tujuan merancang dan mengevaluasi sistem rekomendasi buku berbasis *Content-Based Filtering* pada lingkungan perpustakaan SD Negeri 2 Ganungkidul. Metodologi penelitian difokuskan pada proses pengolahan data buku, perancangan sistem rekomendasi, serta evaluasi kinerja sistem berdasarkan metrik akurasi rekomendasi.

A. Alur Kerja Sistem Rekomendasi

Diagram alur sistem rekomendasi buku berbasis *Content-Based Filtering* pada gambar di atas menggambarkan proses kerja sistem yang dimulai dari interaksi awal pengguna hingga dihasilkannya rekomendasi buku yang relevan. Proses diawali pada tahap *Mulai*, ketika sistem menampilkan halaman beranda sebagai titik awal interaksi pengguna. Pada halaman ini, pengguna dapat melakukan pencarian buku atau langsung memilih buku yang tersedia dalam koleksi perpustakaan sebagai dasar interaksi selanjutnya.

Apabila pengguna melakukan pencarian, sistem akan memproses kata kunci yang dimasukkan, seperti judul, genre, atau topik buku, kemudian menampilkan daftar buku yang sesuai dengan input tersebut. Dari hasil pencarian atau dari daftar buku pada beranda, pengguna dapat memilih salah satu buku untuk dilihat lebih lanjut. Buku yang dipilih ini kemudian dijadikan sebagai buku acuan dalam proses rekomendasi. Sistem selanjutnya menganalisis karakteristik konten buku berdasarkan atribut judul, genre, dan sinopsis, lalu menghitung tingkat kemiripan dengan buku lain dalam koleksi menggunakan metode *Cosine Similarity*.



GAMBAR 1
(Diagram Alur Sistem Rekomendasi)

Berdasarkan hasil perhitungan kemiripan tersebut, sistem menampilkan detail buku yang dipilih beserta daftar rekomendasi buku serupa dengan tingkat kemiripan tertinggi. Sistem kemudian mengevaluasi interaksi lanjutan pengguna terhadap rekomendasi yang diberikan. Jika pengguna memilih salah satu buku rekomendasi, maka buku tersebut kembali diproses sebagai buku acuan untuk menghasilkan rekomendasi lanjutan, sehingga membentuk siklus rekomendasi yang berkelanjutan. Apabila pengguna tidak memilih buku rekomendasi, proses interaksi dengan sistem diakhiri dan alur sistem berakhir pada tahap *Selesai*.

B. Term Frequency-Inverse Document Frequency

TF-IDF adalah metode pembobotan teks yang digunakan untuk merepresentasikan dokumen dalam *vector space model*. *Term Frequency* (TF) mengukur frekuensi kemunculan suatu term dalam satu dokumen, yang dihitung dengan rumus :

$$tf=T/L \quad (2.1)$$

dimana :

T = frekuensi kemunculan term

L = jumlah kata unik dalam dokumen j

Inverse Document Frequency (IDF) digunakan untuk menilai sejauh mana suatu term memberikan informasi yang unik dalam koleksi dokumen (korpus) secara keseluruhan [15], [16]. IDF dihitung menggunakan rumus:

$$idf=log\frac{N}{n(i)} \quad (2.2)$$

dimana :

N = total dokumen

$n(i)$ = jumlah dokumen yang mengandung term i

Kombinasi antara TF dan IDF menghasilkan skor TF-IDF, yang mencerminkan tingkat kepentingan suatu term dalam dokumen tertentu relatif terhadap seluruh dokumen dalam korpus[15]. Rumusanya adalah :

$$tf\text{-}idf = tf * idf \quad (2.3)$$

Semakin tinggi nilai TF-IDF suatu term, semakin penting term tersebut dalam konteks dokumen tersebut.

C. Item Similarity

Dalam sistem rekomendasi, pengukuran tingkat kemiripan antar item merupakan komponen esensial dalam menghasilkan rekomendasi yang sesuai dengan preferensi pengguna. Algoritma *Cosine Similarity* digunakan sebagai pendekatan untuk mengukur sejauh mana kesamaan antara item yang telah disukai oleh pengguna dengan item lainnya.

Cosine similarity merupakan metode pengukuran kemiripan antara dua buah vektor dalam ruang berdimensi banyak, berdasarkan nilai dari sudut kosinus antara vektor-vektor tersebut[17]. Semakin mendekati nilai 1, maka sudut antara kedua vektor semakin kecil, yang berarti kedua item tersebut memiliki tingkat kemiripan yang tinggi[18]. Artinya, kedua vektor tersebut semakin mirip. Rumus matematis dari cosine similarity dituliskan sebagai berikut[18]:

$$cos(\vec{x},\vec{y}) = \frac{\vec{x},\vec{y}}{\|\vec{x}\| \, \|\vec{y}\|} \quad (2.4)$$

dimana :

$\vec{x},\vec{y}$ = hasil perkalian *dot product* dari x dan y

$\|\vec{x}\|$ = panjang vektor x

$\|\vec{y}\|$ = panjang vektor y

D. Evaluasi Sistem

Evaluasi sistem diperlukan untuk menilai kinerja dan efektivitas suatu sistem dalam mencapai tujuan yang telah ditentukan. Pada sistem rekomendasi, yang menghasilkan keluaran berupa prediksi item yang disarankan kepada pengguna, diperlukan proses evaluasi untuk mengukur sejauh mana rekomendasi yang dihasilkan relevan terhadap preferensi pengguna.

Untuk menilai performa sistem, digunakan metrik evaluasi yang dapat mengukur kebenaran dan relevansi dari hasil prediksi sistem. Dalam penelitian ini, metrik yang digunakan meliputi *Precision at K* dan *Normalized Discounted Cumulative Gain at K*. Dimana *Precision@K* hanya digunakan sebagai perbandingan nilai evaluasi.

Dalam konteks sistem rekomendasi, sistem umumnya menghasilkan daftar rekomendasi terbatas yang dikenal dengan top-K list. *Normalized Discounted Cumulative Gain at K* (NDCG@K) memberikan bobot pada setiap item dalam top-K. Ketika item dalam daftar dinilai relevan, maka akan ada bobot yang diperoleh oleh item tersebut berdasarkan posisinya dalam daftar. NDCG@K dinyatakan dalam rumus sebagai berikut:

Discounted Cumulative Gain (DCG):

$$DCG@K = \sum_{i=1}^{K} \frac{rel_i}{log_2(i+1)} \quad (2.5)$$

Ideal Discounted Cumulative Gain (IDCG):

$$DCG@K = \sum_{i=1}^{|REL_K|} \frac{1}{log_2(i+1)} \quad (2.6)$$

Normalized Discounted Cumulative Gain (NDCG) :

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (2.7)$$

dimana:

i = posisi item dalam daftar rekomendasi

rel_i = relevansi item pada posisi ke-i

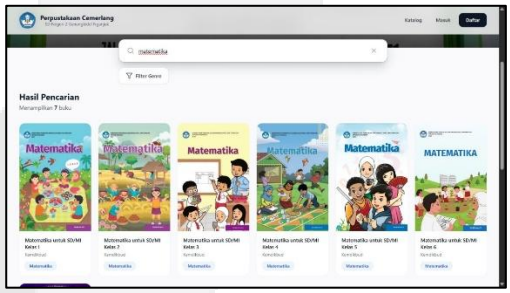
$|REL_K|$ = jumlah item relevan yang tersedia

Sementara *Precision at K* (Precision@K) digunakan untuk mengukur ketepatan prediksi dalam daftar rekomendasi tersebut, dengan K menyatakan jumlah item teratas yang direkomendasikan. Precision at K didefinisikan secara matematis sebagai:

$$Precision@K = \frac{jumlah\ item\ relevan\ di\ top\text{-}K}{K} \quad (2.8)$$

IV.HASIL DAN PEMBAHASAN

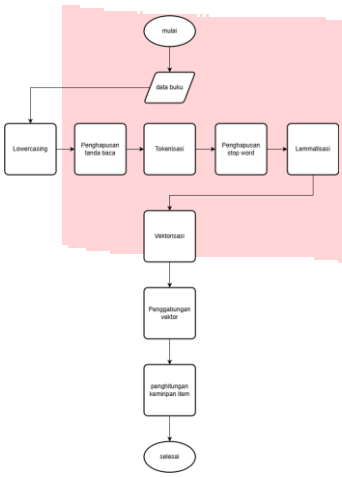
Evaluasi sistem rekomendasi buku dilakukan untuk menilai kemampuan sistem dalam menghasilkan rekomendasi yang relevan berdasarkan konten buku dan histori bacaan pengguna. Pengujian dilakukan secara *offline* dengan memanfaatkan data histori bacaan pengguna sebagai *ground truth*, tanpa melibatkan umpan balik eksplisit atau studi persepsi pengguna. Pendekatan ini sejalan dengan tujuan penelitian yang berfokus pada evaluasi akurasi prediksi sistem rekomendasi berbasis *Content-Based Filtering*.



GAMBAR 2
(Tangkapan Layar Website E-Perpus)

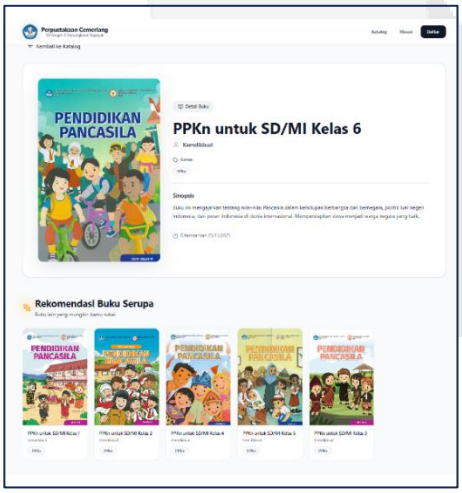
Pengujian dilakukan terhadap 19 pengguna dengan skema rekomendasi *top-10*, sehingga setiap pengguna memperoleh sepuluh buku rekomendasi berdasarkan tingkat kemiripan tertinggi terhadap buku acuan. Kinerja sistem dievaluasi menggunakan dua metrik utama, yaitu *Precision@10* dan *Normalized Discounted Cumulative Gain* (*NDCG@10*). Precision@10 digunakan untuk mengukur proporsi buku relevan dalam sepuluh rekomendasi teratas, sedangkan NDCG@10 digunakan untuk mengevaluasi kualitas peringkat rekomendasi dengan mempertimbangkan posisi buku relevan dalam daftar rekomendasi.

Hasil evaluasi menunjukkan bahwa sistem rekomendasi memperoleh nilai rata-rata Precision@10 sebesar 0,65. Nilai ini mengindikasikan bahwa, secara rata-rata, sekitar 65% buku yang direkomendasikan pada sepuluh peringkat teratas termasuk dalam kategori relevan berdasarkan histori bacaan pengguna. Hasil ini menunjukkan bahwa pendekatan *Content-Based Filtering* yang diterapkan mampu menghasilkan rekomendasi yang cukup tepat meskipun hanya mengandalkan atribut konten buku tanpa memanfaatkan data interaksi antar pengguna.



GAMBAR 3
(Diagram, Alur Pembentukan Rekomendasi)

Selain itu, nilai rata-rata NDCG@10 yang diperoleh sebesar 0,61. Nilai NDCG@10 yang berada di atas 0,5 menunjukkan bahwa sistem tidak hanya mampu merekomendasikan buku yang relevan, tetapi juga menempatkan sebagian besar buku relevan tersebut pada posisi peringkat yang relatif tinggi. Hal ini penting dalam konteks sistem rekomendasi, karena pengguna cenderung lebih memperhatikan item yang muncul pada peringkat awal daftar rekomendasi.



GAMBAR 4
(Hasil Rekomendasi di Website E-Perpus)

Kombinasi hasil Precision@10 dan NDCG@10 tersebut menunjukkan bahwa sistem rekomendasi yang dikembangkan memiliki keseimbangan antara ketepatan rekomendasi dan kualitas pengurutan peringkat. Penerapan metode TF-IDF dalam merepresentasikan konten buku terbukti efektif dalam menangkap karakteristik utama buku berdasarkan judul, kategori, dan deskripsi singkat. Selanjutnya, penggunaan *Cosine Similarity* memungkinkan sistem mengukur kedekatan semantik antar buku secara sederhana namun efisien.

TABEL 1
(Akurasi Sistem Rekomendasi)

User	NDCG @K per user	Precision @K per user
1	1,00	1,00
2	0,69	0,80
3	0,71	0,70
4	0,82	0,80
5	0,42	0,40
6	0,42	0,40
7	0,64	0,80
8	0,65	0,50
9	0,80	0,80
10	0,47	0,60
11	0,43	0,50
12	0,42	0,50
13	0,24	0,30
14	0,37	0,50
15	0,83	0,80
16	0,78	0,70
17	0,67	0,70
18	0,53	0,70
19	0,78	0,90
Rata-rata	**0,61**	**0,65**

Namun demikian, hasil evaluasi juga menunjukkan adanya keterbatasan sistem. Karena sistem hanya mengandalkan kesamaan konten, rekomendasi yang dihasilkan cenderung bersifat homogen dan terbatas pada buku dengan karakteristik konten yang serupa. Sistem belum mampu menangkap variasi preferensi pengguna yang bersifat implisit atau kontekstual, seperti perubahan minat baca dari waktu ke waktu. Selain itu, evaluasi dilakukan pada skala data dan jumlah pengguna yang terbatas, sehingga hasil yang diperoleh belum dapat digeneralisasi untuk koleksi buku yang lebih besar atau lingkungan pengguna yang lebih beragam.

Secara keseluruhan, hasil pengujian menunjukkan bahwa sistem rekomendasi buku berbasis *Content-Based Filtering* yang dikembangkan telah mampu memberikan rekomendasi dengan tingkat relevansi dan kualitas peringkat yang cukup baik untuk konteks perpustakaan sekolah dasar. Temuan ini menguatkan bahwa pendekatan *Content-Based Filtering* dapat menjadi solusi awal yang layak untuk mendukung layanan peminjaman buku pada lingkungan dengan keterbatasan data interaksi pengguna.

V. KESIMPULAN

Penelitian ini berhasil merancang dan mengembangkan sistem rekomendasi buku berbasis *Content-based Filtering* untuk mendukung layanan peminjaman buku di Perpustakaan SDN 2 Ganungkidul Nganjuk. Sistem memanfaatkan algoritma TF-IDF dan cosine similarity untuk mengukur kemiripan konten buku berdasarkan metadata, serta diimplementasikan dengan arsitektur *client–server* yang memisahkan *backend* dan *frontend*. Hasil pengembangan menunjukkan bahwa sistem mampu memberikan rekomendasi buku yang relevan dan mudah digunakan oleh siswa sekolah dasar.

Hasil evaluasi kinerja sistem rekomendasi menggunakan metrik NDCG@10 dan Precision@10 menunjukkan bahwa sistem memiliki performa yang cukup baik. Nilai rata-rata NDCG@10 sebesar 0,61 mengindikasikan bahwa item relevan cenderung muncul pada posisi peringkat atas rekomendasi, sedangkan nilai rata-rata Precision@10 sebesar 0,65 menunjukkan tingkat ketepatan rekomendasi yang memadai. Dengan demikian, sistem rekomendasi yang dikembangkan telah memenuhi tujuan evaluasi akurasi prediksi.

REFERENSI

[1] T. Anwar, V. Uma, dan Shahjad, "Book Recommendation for eLearning Using Collaborative Filtering and Sequential Pattern Mining," dalam *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*, Institute of Electrical and Electronics Engineers Inc., Okt 2020. doi: 10.1109/ICDABI51230.2020.9325599.

[2] V. Nuipian dan J. Chuaykhun, "Book Recommendation System based on Course Descriptions using Cosine Similarity," dalam *ACM International Conference Proceeding Series*, Association for Computing Machinery, Des 2023, hlm. 273–277. doi: 10.1145/3639233.3639335.

[3] P. Arunruviwat dan V. Muangsin, "A Hybrid Book Recommendation System for University Library," dalam *ICSEC 2022 - International Computer Science and Engineering Conference 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, hlm. 291–295. doi: 10.1109/ICSEC56337.2022.10049318.

[4] S. K. Dirjen *dkk.*, "Sistem Rekomendasi Buku untuk Perpustakaan Perguruan Tinggi Berbasis Association Rule," *masa berlaku mulai*, vol. 1, no. 3, hlm. 304–312, 2019.

[5] M. Alkaff, H. Khatimi, dan 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, hlm. 193–202, Sep 2020, doi: 10.30812/matrik.v20i1.617.

[6] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, dan R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Applied Sciences (Switzerland)*, vol. 10, no. 21, hlm. 1–20, Nov 2020, doi: 10.3390/app10217748.

[7] F. Ricci, L. Rokach, dan B. Shapira, *Recommender Systems Handbook*. Springer US, 2011. doi: 10.1007/978-0-387-85820-3.

[8] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, dan B. Venkatesh, "Content-based movie recommendation system using genre correlation," dalam *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2019, hlm. 391–397. doi: 10.1007/978-981-13-1927-3_42.

[9] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, dan R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Applied Sciences (Switzerland)*, vol. 10, no. 21, hlm. 1–20, Nov 2020, doi: 10.3390/app10217748.

[10] R. Logesh, V. Subramaniyaswamy, D. Malathi, N. Sivaramakrishnan, dan V. Vijayakumar, "Enhancing recommendation stability of collaborative filtering recommender system through bio-inspired clustering ensemble method," *Neural Comput Appl*, vol. 32, no. 7, hlm. 2141–2164, Apr 2020, doi: 10.1007/s00521-018-3891-5.

[11] *Recommender Systems Handbook*. Springer US, 2011. doi: 10.1007/978-0-387-85820-3.

[12] K. Kmgs dan R. Rahm, "Learning to Use Normalization Techniques for Preprocessing and Classification of Text Documents."

[13] D. J. Ladani dan N. P. Desai, "2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)," 2020, doi: 10.1109/ICACCS48705.2020.9074166.

[14] *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020.

[15] D. Wang, Y. Liang, D. Xu, X. Feng, dan R. Guan, "A content-based recommender system for computer science publications," *Knowl Based Syst*, vol. 157, hlm. 1–9, Okt 2018, doi: 10.1016/j.knosys.2018.05.001.

[16] K. Samosir dan F. Ginting, "A Comparative Analysis of Content-Based Filtering and TF-IDF Approaches for Enhancing Sports Recommendation Systems," 2024. [Daring]. Tersedia pada: http://innovatics.unsil.ac.id

[17] C. Fiarni, H. Maharani, dan N. Calista, "Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods," 2019.

[18] M. Chiny, M. Chihab, O. Bencharef, dan Y. Chihab, "Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms," Scitepress, Mei 2022, hlm. 15–20. doi: 10.5220/0010727500003101.

[19] S. Sarica dan J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8 August, Agu 2021, doi: 10.1371/journal.pone.0254937.

[20] Y. HaCohen-Kerner, D. Miller, dan Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLoS One*, vol. 15, no. 5, Mei 2020, doi: 10.1371/journal.pone.0232525.

[21] V. Balakrishnan dan L.-Y. Ethel, "Stemming and Lemmatization: A Comparison of Retrieval Performances," *Lecture Notes on Software Engineering*, vol. 2, no. 3, hlm. 262–267, 2014, doi: 10.7763/lnse.2014.v2.134.