JSITIK, Vol.5 No.2 Juni 2027
ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

Sistem Rekomendasi Buku Berbasis Content-Based Filtering dengan Menggunakan TF-IDF dan Cosine Similarity

Diterima:
21 Juli 2026
Revisi:
05 Oktober 2026
Terbit:
28 Oktober 2026

Hensel Revy Vandela, *Ilham Saifudin, Reni Umilasari

Universitas Muhammadiyah Jember

Abstrak— **Latar Belakang:** Meningkatnya jumlah koleksi buku digital menyulitkan pengguna dalam menemukan buku yang sesuai dengan kebutuhannya. Oleh karena itu, diperlukan sistem rekomendasi yang dapat menyajikan buku yang relevan. **Tujuan:** Penelitian ini bertujuan merancang, mengimplementasikan, dan mengevaluasi sistem rekomendasi buku berbasis *Content-Based Filtering* (CBF) menggunakan TF-IDF dan *Cosine Similarity*. **Metode:** Sistem dibangun menggunakan metode *Content-Based Filtering* (CBF), dengan TF-IDF dan *Cosine Similarity*, serta dievaluasi menggunakan metrik *Precision*, NDCG, dan *Running Time* pada skenario Top-3, Top-5, dan Top-10. **Hasil:** Sistem menghasilkan nilai *Precision* sebesar 0,523 pada Top-3 dan 0,518 pada Top-5, serta 0,508 pada Top-10, serta nilai NDCG sebesar 0,666 pada Top-3 dan 0,688 pada Top-5, serta 0,706 pada Top-10. Nilai *Cosine Similarity* tertinggi yang diperoleh mencapai 0,550. **Kesimpulan:** Kombinasi TF-IDF dan *Cosine Similarity* terbukti efektif dalam menghasilkan rekomendasi buku yang relevan dan terurut dengan baik sesuai tingkat kemiripan konten.
Kata Kunci— Sistem Rekomendasi; Content-Based Filtering; TF-IDF; Cosine Similarity

Abstract*— *Background: *The increasing number of digital book collections makes it difficult for users to find books that meet their needs. Therefore, a recommendation system is needed to provide relevant book suggestions.* ***Objective:*** *This study aims to design, implement, and evaluate a book recommendation system based on Content-Based Filtering (CBF) using TF-IDF and Cosine Similarity.* ***Methods:*** *The system was developed using the Content-Based Filtering (CBF) method, incorporating TF-IDF and Cosine Similarity. Its performance was evaluated using Precision, NDCG, and Running Time metrics under Top-3, Top-5, and Top-10 recommendation scenarios.* ***Results:*** *The system achieved Precision scores of 0.523 for Top-3, 0.518 for Top-5, and 0.508 for Top-10. The NDCG scores were 0.666 for Top-3, 0.688 for Top-5, and 0.706 for Top-10. The highest Cosine Similarity score obtained was 0.550.* ***Conclusion:*** *The combination of TF-IDF and Cosine Similarity proved effective in generating relevant book recommendations ranked according to content similarity.*
Keywords*— Recommendation System; Content-Based Filtering; TF-IDF; Cosine Similarity*

This is an open access article under the CC BY-SA License.

Penulis Korespondensi:

Ilham Saifudin,
Teknik Informatika,
Universitas Muhammadiyah Jember,
Email: ilham.saifudin@unmuhjember.ac.id
ORCID ID: https://orcid.org/0000-0002-2063-4524

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

I. PENDAHULUAN

Perkembangan teknologi informasi menyebabkan jumlah koleksi buku digital terus meningkat. Kondisi ini memudahkan pengguna dalam mengakses berbagai pilihan buku, tetapi juga membuat kesulitan dalam menemukan buku yang sesuai dengan kemauan. Untuk mengatasi permasalahan tersebut, diperlukan sistem rekomendasi yang mampu menyajikan buku-buku yang relevan.

Sistem rekomendasi merupakan metode yang digunakan untuk memberikan saran item sesuai preferensi pengguna (Saifudin & Widiyaningtyas, 2024). Salah satu pendekatan yang banyak digunakan adalah *Content-Based Filtering*, yang menghasilkan rekomendasi berdasarkan kemiripan antar item. Pada data teks, kemiripan tersebut dihitung menggunakan TF-IDF untuk pembobotan kata dan *Cosine Similarity* untuk mengukur tingkat kemiripan dokumen.

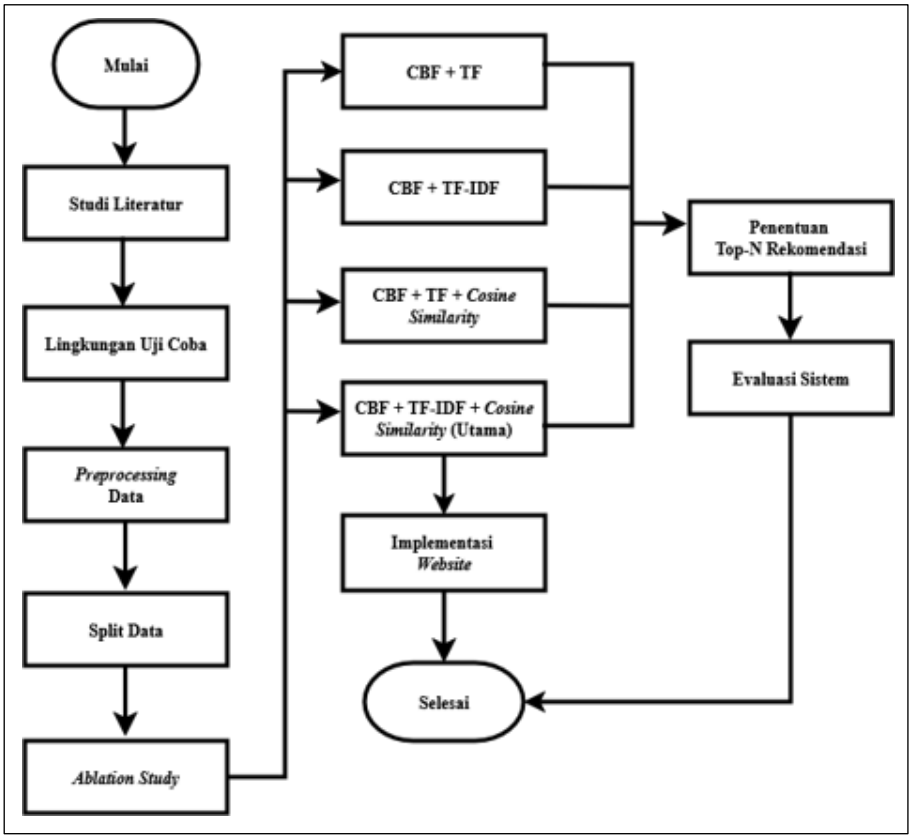
Penelitian yang dilakukan oleh (Alkaff et al., 2020) membahas pengembangan sistem rekomendasi buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan menggunakan metode *Content-Based Filtering* dengan menggunakan 1.023 buku. Selanjutnya, (Huda et al., 2022) membangun sistem rekomendasi artikel berita menggunakan 22 artikel dan 345 data interaksi dari 20 pengguna. Penelitian tersebut menerapkan tahapan *clean-up*, *normalize*, dan *stopwords*, serta memperoleh nilai *Recall*@5 sebesar 73% dan *Recall*@10 sebesar 80%. Penelitian lainnya oleh (Ridhwanullah et al., 2024)menerapkan *Content-Based Filtering* berbasis TF-IDF dan *Cosine Similarity* pada 162 data buku. Tahap *pre-processing* terdiri atas *tokenization*, *normalization*, dan *stopwords*, sedangkan hasil evaluasi menunjukkan rata-rata *Accuracy* sebesar 92,5%.

Berdasarkan penelitian terdahulu, metode TF-IDF dan *Cosine Similarity* telah berhasil diterapkan pada berbagai jenis data, seperti produk, artikel berita, layanan, maupun buku. Namun, penerapannya pada koleksi buku dengan jumlah data yang lebih besar masih memerlukan pengujian lebih lanjut. Perbedaan penelitian ini dengan penelitian sebelumnya adalah, penelitian ini menggunakan 6.810 data buku dengan atribut judul, deskripsi, penulis, penerbit untuk mengevaluasi efektivitas metode TF-IDF dan *Cosine Similarity* dalam menghasilkan rekomendasi buku yang relevan. Berdasarkan uraian tersebut, penelitian ini bertujuan merancang dan mengimplementasikan sistem rekomendasi buku berbasis *Content-Based Filtering* menggunakan TF-IDF dan *Cosine Similarity* serta mengevaluasi kinerjanya dalam menghasilkan rekomendasi buku yang relevan.

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

II. METODE

Pada Gambar 1, Penelitian ini dilakukan melalui beberapa tahapan yang saling berkaitan untuk menghasilkan sistem rekomendasi buku berbasis *Content-Based Filtering*. Tahapan penelitian dimulai dari pengumpulan dataset buku, dilanjutkan dengan *preprocessing* data, pembobotan teks menggunakan TF-IDF, perhitungan kemiripan menggunakan *Cosine Similarity*, hingga tahap evaluasi sistem menggunakan metrik *Precision* dan *Normalized Discounted Cumulative Gain* (NDCG) dan *Running Time*.



Gambar 1. Alur Penelitian

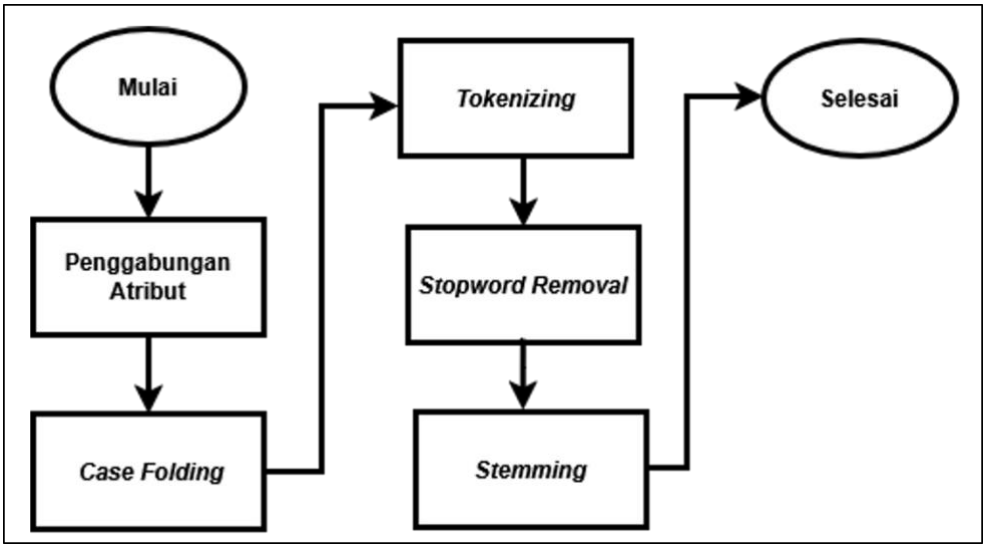
A. *Dataset Penelitian*

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle dengan judul *Books Dataset* dengan 6.810 data buku dengan atribut *title*, *authors*, *categories*, dan *description*. Dataset ini digunakan sebagai sumber utama dalam pembentukan representasi teks buku.

B. *Preprocessing Data*

Preprocessing adalah langkah untuk menyiapkan data agar memiliki format yang tepat agar bisa diproses. Pada penelitian ini, tahapan *preprocessing* terdiri atas penggabungan atribut, *case folding*, *tokenizing*, *filtering*, dan *stemming* yang tertera seperti Gambar 2.

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970



Gambar 2. Tahapan Pre-Processing Data

Sebelum proses *preprocessing* dilakukan sesuai tahapan pada Gambar 2, terlebih dahulu dilakukan pengecekan *missing value* pada setiap atribut dataset. *Missing value* kemudian ditangani menggunakan fungsi *fillna*(' ') dengan mengganti nilai kosong menjadi string kosong. Setelah data bersih, dilakukan beberapa tahapan preprocessing teks. Tahap pertama adalah penggabungan atribut, yaitu atribut *title*, *authors*, *categories*, dan *description* dijadikan satu kolom. Proses ini dilakukan agar seluruh informasi buku dapat diproses sebagai satu dokumen utuh (Reswara et al., 2025). Tahap kedua adalah *case folding*, yaitu proses mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan penulisan antara huruf kapital dan huruf kecil. Tahap ketiga adalah *tokenizing*, yaitu proses memecah teks menjadi kumpulan kata atau token. Selanjutnya adalah *filtering*, yaitu penghapusan kata-kata tidak penting (*stopword removal*) seperti "*and*", "*the*", dan "is". Tahap terakhir adalah *stemming*, yaitu proses mengubah kata ke bentuk dasarnya agar kata dengan makna yang sama dapat diperlakukan sebagai satu kata (Hidayat & Rizqi, 2020).

C. *Split Data*

Pembagian dataset dilakukan menggunakan metode *Hold-Out* dengan dua skenario, yaitu 80:20. Pembagian dilakukan secara acak untuk memperoleh distribusi data yang merata. Pada skenario 80:20, 80% data digunakan sebagai data *training* dan 20% sebagai data *testing*.

D. *Pembobotan TF-IDF*

Metode TF-IDF digunakan untuk mengubah data teks menjadi representasi vektor numerik yang dapat dibandingkan secara matematis (Timur & Rohman, 2025). Proses ini dilakukan dengan menghitung frekuensi kemunculan kata dalam dokumen menggunakan *Term Frequency* (TF), kemudian menentukan tingkat kepentingan kata menggunakan *Inverse Document Frequency* (IDF). TF digunakan untuk menghitung jumlah kemunculan kata pada suatu dokumen,

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

sedangkan IDF berfungsi untuk mengukur seberapa penting suatu kata dengan melihat seluruh kumpulan dokumen. Semakin jarang kata itu muncul di banyak dokumen, semakin tinggi nilai pentingnya. Perhitungan TF-IDF ditunjukkan pada Persamaan (1), Persamaan (2), dan Persamaan (3).

$$TF\ (t,d)\ =\ jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d \tag{1}$$

$$IDF\ = log\ \frac{total\ jumlah\ dokumen\ dalam\ d}{jumlah\ dokumen\ di\ mana\ kata\ muncul} \tag{2}$$

$$W\ =\ TF\ \times\ (IDF\ +\ 1) \tag{3}$$

E. *Cosine Similarity*

Metode ini digunakan untuk menentukan tingkat kemiripan dua vektor berdasarkan sudut kosinus yang terbentuk antar vektor TF-IDF (Syah et al., 2025). Nilai *Cosine Similarity* berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa kedua item memiliki tingkat kemiripan yang tinggi (Altiarika et al., 2025). Perhitungan *Cosine Similarity* dilakukan menggunakan Persamaan (4).

$$Cosine\ similarity = \frac{(v_1.v_2)}{\|v_1\| \times \|v_2\|} \tag{4}$$

F. *Evaluasi Sistem*

Tahap selanjutnya adalah evaluasi sistem rekomendasi menggunakan metrik *Precision*, *Normalized Discounted Cumulative Gain* (NDCG), dan *Running Time*. *Precision* digunakan untuk mengukur ketepatan sistem dalam memberikan rekomendasi. Metrik ini menunjukkan seberapa banyak item yang direkomendasikan benar-benar relevan dibandingkan dengan seluruh item yang direkomendasikan (Halimatussa'diyah & Dini, 2025). Perhitungan *Precision* ditunjukkan pada Persamaan (5).

$$Precision = \frac{|Recommended(u) \cap Testing(u)|}{|Recommended(u)|} \tag{5}$$

Selanjutnya, NDCG digunakan untuk mengevaluasi kualitas urutan rekomendasi berdasarkan tingkat relevansi dan posisi item dalam daftar rekomendasi. Semakin tinggi nilai NDCG, semakin baik kualitas pemeringkatan yang dihasilkan sistem (Widiyaningtyas et al., 2025). Perhitungan NDCG diawali dengan menghitung *Discounted Cumulative Gain* (DCG) dan *Ideal Discounted Cumulative Gain* (IDCG), kemudian nilai NDCG diperoleh dari perbandingan

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

kedua nilai tersebut (Khusna et al., 2020), sebagaimana ditunjukkan pada Persamaan (6), Persamaan (7), dan Persamaan (8).

$$DCG_p = \sum_{i=1}^{p} \frac{rel_i}{\log_2(i+1)} \tag{6}$$

$$IDCG_p = \sum_{i=1}^{REL} \frac{rel_i}{\log_2(i+1)} \tag{7}$$

$$NDCG = \frac{DCG}{IDCG} \tag{8}$$

Selanjutnya *Running Time*, yang digunakan untuk mengukur waktu yang dibutuhkan sistem dalam menyelesaikan proses komputasi hingga menghasilkan rekomendasi. Semakin kecil nilai *Running Time*, semakin efisien kinerja sistem dalam menghasilkan rekomendasi (Saifudin et al., 2025). Efisiensi algoritma dianalisis melalui kompleksitas waktu (*time complexity*), yaitu pertumbuhan waktu eksekusi seiring bertambahnya jumlah data masukan (Arafat et al., 2026). Kompleksitas waktu dianalisis berdasarkan tiga kondisi, yaitu *best case*, *average case*, dan *worst case* (Basir, 2020). Evaluasi dilakukan menggunakan skenario Top-N rekomendasi pada Top-3, Top-5, dan Top-10 rekomendasi untuk mengetahui performa sistem dalam menghasilkan rekomendasi buku yang relevan dan sesuai dengan kebutuhan pengguna.

III. HASIL DAN PEMBAHASAN

Sistem rekomendasi buku dibangun menggunakan metode *Content-Based Filtering* (CBF) dengan pembobotan TF-IDF dan pengukuran kemiripan menggunakan *Cosine Similarity*. Pengujian dilakukan menggunakan buku *Harry Potter and the Sorcerer's (Book 1)* sebagai input untuk menghasilkan daftar buku yang memiliki karakteristik konten serupa. Sebelum dilakukan *hyperparameter tuning*, sistem telah mampu memberikan rekomendasi yang relevan berdasarkan kemiripan konten. Hasil rekomendasi sebelum *hyperparameter tuning* disajikan pada Tabel 1.

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

Tabel 1. Hasil Rekomendasi Sebelum Hyperparameter Tuning

Judul Buku Rekomendasi	***Cosine Similarity***
Harry Potter *and the Chamber of Secrets (Book 2)*	0,429
Harry Potter and the Order of the Phoenix (Book 5)	0,423
Harry Potter *and the Half-Blood Prince (Book 6)*	0,407

Sistem juga dievaluasi untuk mengetahui kualitas rekomendasi yang dihasilkan sebelum dilakukan *hyperparameter tuning*. Hasil evaluasi tersebut ditampilkan pada Tabel 2.

Tabel 2. Hasil Evaluasi Sebelum Hyperparameter Tuning

No	***Matrix Evaluasi***	**TOP-3**	**TOP-5**	**TOP-10**
1.	***Precision***	0,479	0,470	0,460
2.	**NDCG**	0,632	0,656	0,673

Berdasarkan Tabel 1, rekomendasi dengan nilai *Cosine Similarity* tertinggi adalah *Harry Potter and the Chamber of Secrets (Book 2)* dengan nilai 0,429, diikuti *Harry Potter and the Order of the Phoenix (Book 5)* sebesar 0,423 dan *Harry Potter and the Half-Blood Prince (Book 6)* sebesar 0,407. Hasil evaluasi sebelum *hyperparameter tuning* pada Tabel 2 menunjukkan nilai *Precision* sebesar 0,479, 0,470, dan 0,460 pada Top-3, Top-5, dan Top-10, sedangkan nilai NDCG masing-masing sebesar 0,632, 0,656, dan 0,673. Hasil tersebut menunjukkan bahwa sistem telah mampu memberikan rekomendasi berdasarkan kemiripan konten, tetapi masih dapat ditingkatkan melalui optimasi parameter TF-IDF.

Untuk mengatasi hal tersebut dilakukan *hyperparameter tuning*, optimasi ini bertujuan untuk menghasilkan representasi teks yang lebih informatif sehingga hubungan antar dokumen dapat dikenali dengan lebih baik. Setelah dilakukan *hyperparameter tuning*, hasil rekomendasi dan evaluasi menunjukkan peningkatan seperti yang ditunjukkan pada Tabel 3.

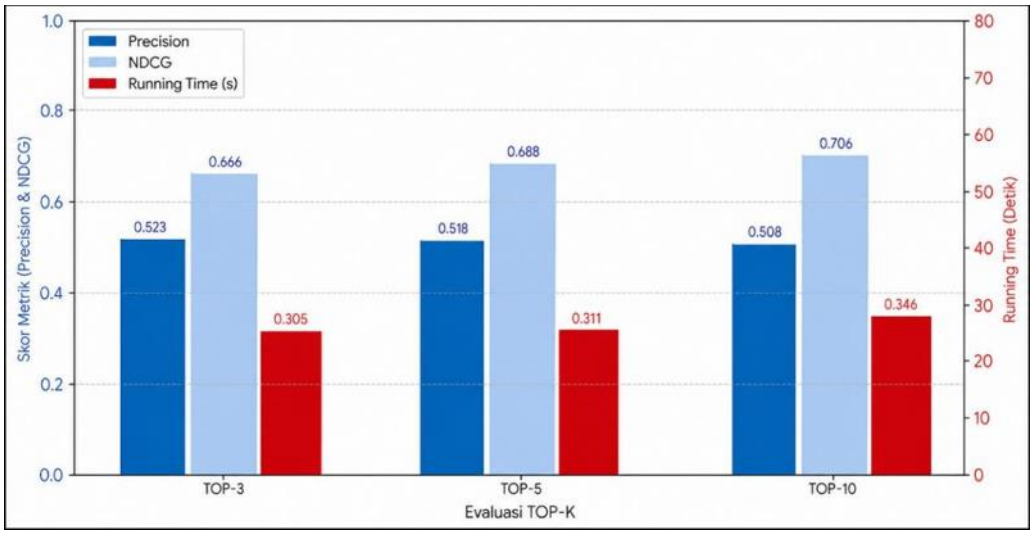
Tabel 3. Hasil Rekomendasi Setelah Hyperparameter Tunning

Judul Buku Rekomendasi	***Cosine Similarity***
Harry Potter and the Chamber of Secrets (*Book* 2)	0,550
Harry Potter *and the Half-Blood Prince (Book 1)*	0,545
Harry Potter and the Order of the Phoenix	0,522

Setelah dilakukan *hyperparameter tuning*, terjadi peningkatan nilai *Cosine Similarity*. Berdasarkan Tabel 3, rekomendasi dengan nilai *Cosine Similarity* tertinggi adalah *Harry Potter and the Chamber of Secrets (Book 2)* sebesar 0,550, diikuti *Harry Potter and the Half-Blood*

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

Prince (Book 1) sebesar 0,545 dan *Harry Potter and the Order of the Phoenix* sebesar 0,522. Hasil tersebut menunjukkan bahwa optimasi parameter TF-IDF dapat meningkatkan kemampuan sistem dalam merepresentasikan karakteristik buku berdasarkan atribut *title, authors, categories*, dan *description*. Selanjutnya, kinerja sistem dievaluasi menggunakan metrik *Precision*, *Normalized Discounted Cumulative Gain* (NDCG), dan *Running Time* pada skenario Top-3, Top-5, dan Top-10, seperti ditunjukkan pada Gambar 3.



Gambar 3. Hasil Evaluasi Setelah Hyperparameter Tuning

Pada Gambar 3, hasil evaluasi menunjukkan bahwa sistem mampu memberikan rekomendasi yang relevan pada berbagai skenario Top-N. Nilai *Precision* yang diperoleh sebesar 0,523 pada Top-3, 0,518 pada Top-5, dan 0,508 pada Top-10. Nilai tersebut menunjukkan bahwa sebagian besar buku yang direkomendasikan memiliki relevansi dengan buku input berdasarkan karakteristik kontennya. Meskipun terjadi penurunan nilai *Precision* seiring bertambahnya jumlah rekomendasi, kondisi tersebut masih dapat terjadi karena semakin banyak buku yang ditampilkan, semakin besar kemungkinan terdapat buku dengan tingkat relevansi yang lebih rendah.

Selain *Precision*, hasil evaluasi NDCG menunjukkan nilai sebesar 0,666 pada Top-3, 0,688 pada Top-5, dan 0,706 pada Top-10. Nilai NDCG yang semakin tinggi pada Top-N yang lebih besar menunjukkan bahwa sistem mampu mempertahankan penempatan buku yang relevan dalam daftar rekomendasi. Dengan demikian, hasil evaluasi menunjukkan bahwa sistem tidak hanya mampu menghasilkan rekomendasi yang relevan, tetapi juga menyusun rekomendasi berdasarkan tingkat relevansinya.

Selain peningkatan performa berdasarkan metrik evaluasi, sistem juga dianalisis dari sisi efisiensi waktu proses. Efisiensi ini dievaluasi menggunakan kompleksitas waktu (*Big* O) untuk menggambarkan perubahan kebutuhan komputasi seiring bertambahnya jumlah dokumen dan term yang diproses. Analisis dilakukan pada dua tahapan utama, yaitu pembobotan TF-IDF dan

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

perhitungan *Cosine Similarity*, dengan mempertimbangkan skenario *best case, average case*, dan *worst case*. Kompleksitas waktu kedua algoritma tersebut disajikan pada Tabel 4.

Tabel 4. Analisis Kompleksitas Waktu Algoritma

Algoritma	*Best Case*	*Average Case*	*Worst Case*
TF-IDF	O (T)	O (D × T)	O (D × T)
Cosine Similarity	O (T)	O (D × T)	O (D^2 × T)

Secara umum pada Tabel 4, TF-IDF memiliki kompleksitas waktu rata-rata sebesar O (D × T), sehingga waktu pemrosesan meningkat secara linier terhadap jumlah dokumen D dan jumlah *term* T. Sementara itu, *Cosine Similarity* memiliki kompleksitas terburuk sebesar O (D^2 × T) karena setiap dokumen harus dibandingkan dengan dokumen lainnya untuk menghitung tingkat kemiripan. Meskipun membutuhkan waktu komputasi lebih besar, penggunaan *Cosine Similarity* memberikan peningkatan kualitas rekomendasi sehingga masih layak diterapkan pada sistem rekomendasi buku.

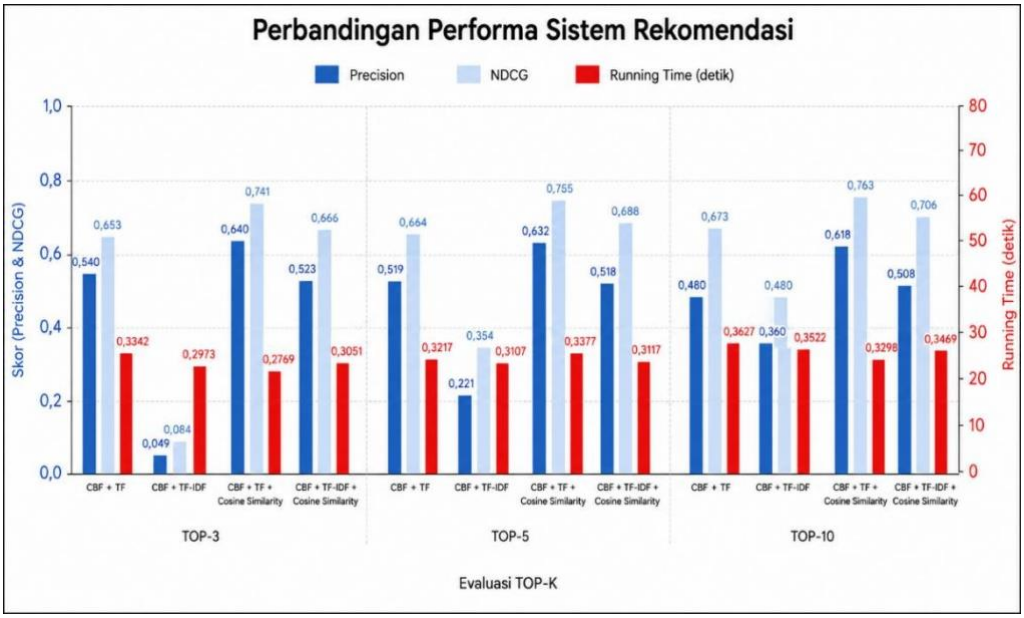
Untuk mengetahui kontribusi setiap komponen dalam sistem rekomendasi, dilakukan ablation study dengan membandingkan empat metode, yaitu CBF+TF, CBF+TF-IDF, CBF+TF+Cosine Similarity, dan CBF+TF-IDF+Cosine Similarity sebagai metode usulan. Hasil pengujian dari keempat metode tersebut disajikan pada Tabel 5.

Tabel 5. Hasil Pengujian Ablation Study

Matriks	CBF+TF	CBF+TF-IDF	CBF+TF+*Cosine Similarity*	CBF+TF-IDF+ *Cosine Similarity*
Top-3 *Precision*	0,540	0,049	0,640	0,532
Top-3 NDCG	0,653	0,084	0,741	0,666
Top-3 *Running Time* (detik)	0,3342	0,2973	0,2769	0,3051
Top-5 *Precision*	0,519	0,221	0,632	0,518
Top-5 NDCG	0,664	0,354	0,755	0,688
Top-5 *Running Time* (detik)	0,3217	0,3107	0,3377	0,3117
Top-10 *Precision*	0,480	0,360	0,618	0,508
Top-10 NDCG	0,673	0,480	0,763	0,706
Top-10 *Running Time* (detik)	0,3627	0,3522	0,3298	0,3469

Untuk memberikan gambaran yang lebih jelas mengenai perbandingan hasil evaluasi tersebut, data pada Tabel 5 juga divisualisasikan dalam bentuk diagram yang ditunjukkan pada Gambar 4.

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970



Gambar 4. Hasil Pengujian Ablation Study

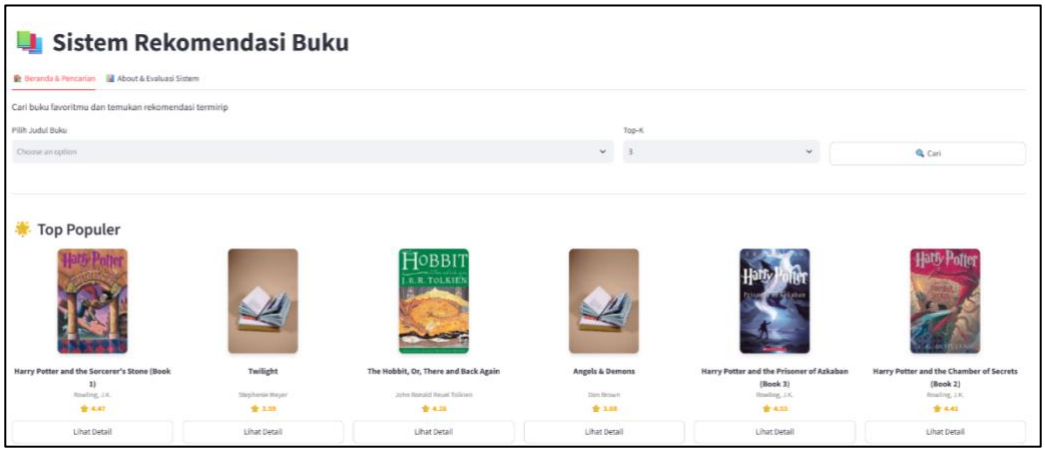
Pada Tabel 5 dan Gambar 4 menunjukkan bahwa CBF+TF menghasilkan performa yang cukup baik karena frekuensi kemunculan kata masih mampu merepresentasikan karakteristik isi buku. Sebaliknya, CBF+TF-IDF menunjukkan performa yang relatif lebih rendah dibandingkan kombinasi metode lainnya. Kondisi tersebut menunjukkan bahwa TF-IDF berperan dalam memberikan bobot berdasarkan tingkat kepentingan kata, tetapi belum dapat menentukan kemiripan antar buku tanpa adanya metode pengukuran kemiripan.

Temuan penelitian ini adalah Penambahan *Cosine Similarity* meningkatkan performa pada metode berbasis TF maupun TF-IDF. Kombinasi CBF+TF+*Cosine Similarity* menghasilkan nilai *Precision* dan NDCG tertinggi secara konsisten pada seluruh skenario Top-N. Hasil tersebut menunjukkan bahwa pengukuran kemiripan antar dokumen berperan penting dalam menentukan relevansi rekomendasi. Meskipun demikian, CBF+TF-IDF+*Cosine Similarity* dipilih sebagai metode utama karena TF-IDF mampu memberikan bobot yang berbeda berdasarkan tingkat kepentingan kata dalam dokumen, kemudian *Cosine Similarity* digunakan untuk mengukur kemiripan antar buku berdasarkan representasi tersebut.

Dari sisi efisiensi, penggunaan *Cosine Similarity* memerlukan proses tambahan untuk menghitung tingkat kemiripan antar buku. Perbedaan waktu komputasi antar metode relatif kecil dan dipengaruhi oleh jumlah data serta proses perhitungan pada masing-masing skenario Top-N. Dengan demikian, penggunaan *Cosine Similarity* dapat meningkatkan kualitas pengukuran kemiripan dengan tambahan waktu komputasi yang masih dapat diterima pada dataset yang digunakan. Hasil penelitian ini sejalan dengan (Huda et al., 2022) dan (Ridhwanullah et al., 2024) yang menunjukkan bahwa pendekatan *Content-Based Filtering* dengan pembobotan TF-IDF dan pengukuran kemiripan dapat digunakan untuk menghasilkan rekomendasi yang relevan.

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

Selain evaluasi kinerja, sistem diimplementasikan dalam bentuk sistem rekomendasi buku berbasis website. Antarmuka sistem menyediakan fitur pencarian dan menampilkan informasi buku berupa judul, penulis, dan sampul sebagai referensi awal. Tampilan utama sistem ditunjukkan pada Gambar 5.



Gambar 5. Halaman Utama Website

Setelah judul buku dimasukkan pada kolom pencarian, sistem berbasis *website* menampilkan hasil pencarian beserta rekomendasi buku yang memiliki tingkat kemiripan tertinggi, sebagaimana ditunjukkan pada Gambar 6. Rekomendasi dihasilkan berdasarkan perhitungan kemiripan menggunakan metode *Content-Based Filtering* berbasis TF-IDF dan *Cosine Similarity*. Buku-buku yang direkomendasikan memiliki karakteristik konten yang serupa dengan buku acuan sehingga dapat menjadi alternatif bacaan yang relevan.



Gambar 6. Tampilan Website Hasil Rekomendasi

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Sistem hanya memanfaatkan informasi tekstual dari atribut buku sehingga belum mempertimbangkan preferensi pengguna maupun riwayat interaksi pengguna. Selain itu, evaluasi dilakukan

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

menggunakan satu dataset dan satu pendekatan rekomendasi, sehingga kemampuan generalisasi sistem pada dataset lain belum diketahui.

IV. KESIMPULAN

Metode Content-Based Filtering berhasil diterapkan pada sistem rekomendasi buku berbasis website dengan memanfaatkan atribut title, authors, categories, dan description melalui pembobotan TF-IDF dan pengukuran kemiripan menggunakan Cosine Similarity. Hasil evaluasi menunjukkan nilai Precision sebesar 0,523 pada Top-3, 0,518 pada Top-5, dan 0,508 pada Top-10, sedangkan nilai NDCG sebesar 0,666, 0,688, dan 0,706. Hasil tersebut menunjukkan bahwa sistem mampu menghasilkan rekomendasi yang relevan dan mengurutkannya berdasarkan tingkat kemiripan. Dari sisi waktu komputasi, Running Time meningkat dari 0,3051 detik pada Top-3 menjadi 0,3117 detik pada Top-5 dan 0,3469 detik pada Top-10, yang menunjukkan bahwa semakin banyak rekomendasi yang diproses, semakin besar waktu yang dibutuhkan sistem. Oleh karena itu, penelitian berikutnya dapat mengembangkan sistem dengan mencoba metode rekomendasi lain atau mengombinasikan Content-Based Filtering dengan metode lainnya dalam pendekatan hybrid untuk meningkatkan hasil rekomendasi. Selain itu, sistem dapat dilengkapi dengan fitur personalisasi pengguna, seperti preferensi kategori buku, genre yang diminati, atau riwayat buku yang telah dibaca, sehingga rekomendasi yang dihasilkan dapat lebih sesuai dengan minat dan kebutuhan pengguna.

DAFTAR PUSTAKA

Alkaff, M., Khatimi, H., & Eriadi, A. (2020). Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering. MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer, 20(1), 193–202. https://doi.org/10.30812/matrik.v20i1.617

Altiarika, E., Hikmawati, A., Fitriana, F., & Sari, W. (2025). SmartTourBabel: AI (Artificial Intelligence) Based Tourism System Development Model to Support Creative Economy and Sustainable Development in Bangka Belitung. Jurnal Teknologi Informatika Dan Komputer, 11(1), 248–260. https://doi.org/10.37012/jtik.v11i1.2555

Arafat, M. F., Sipahutar, N., Perdana, A., Situmorang, R. I., & Sihombing, R. A. H. (2026). Analisis Perbandingan Kompleksitas Waktu dan Ruang pada Algoritma Quick Sort, Merge Sort, dan Heap Sort. MUARA KOMPUTER : Jurnal Ilmiah Ilmu Komputer & Elektronika, 2(2), 16–24. https://doi.org/10.64365/murakom.v2i2.281

Basir, R. R. (2020). Analisis Kompleksitas Ruang dan Waktu Terhadap Laju Pertumbuhan Algoritma Heap Sort, Insertion Sort dan Merge dengan Pemrograman Java. STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 5, 109–118. https://doi.org/10.30998/string.v5i2.6250

Halimatussa'diyah, E., & Dini, S. K. (2025). Implementasi Metode Content-Based Filtering dengan Pendekatan Euclidean Distance dalam Sistem Rekomendasi Produk Skincare. Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab, 10(2), 763–783. https://doi.org/10.36341/rabit.v10i2.6371

ISSN: 2986-0458 (Print) / 2986-044X (Online)
DOI: https://doi.org/10.53624/jsitik.v5i2.970

Hidayat, E. Y., & Rizqi, M. A. (2020). Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier. Jurnal Nasional Teknologi Dan Sistem Informasi, 6(2), 90–99. https://doi.org/10.25077/teknosi.v6i2.2020.90-99

Huda, A. A., Fajarudin, R., & Hadinegoro, A. (2022). Sistem Rekomendasi Content-Based Filtering Menggunakan TF-IDF Vector Similarity untuk Rekomendasi Artikel Berita. Building of Informatics, Technology and Science (BITS), 4(3). https://doi.org/10.47065/bits.v4i3.2511

Nur Khusna, A., Rizkawati, L., & Dahlan Jl Ahmad Yani, A. (2020). Information Retrieval pada Pencarian Menu Balita Menggunakan Dice Coefficient. JEPIN (Jurnal Edukasi Dan Penelitian Informatika), 6. https://doi.org/10.26418/JP.V6I1.37169

Reswara, R., Evanita, & Susanto, A. (2025). Implementasi Content-Based Filtering pada Sistem Rekomendasi Buku Perpustakaan. JATI (Jurnal Mahasiswa Teknik Informatika), 9(2). https://doi.org/10.36040/jati.v9i2.13312

Ridhwanullah, D., Kumarahadi, Y. K., & Raharja, B. D. (2024). Content-Based Filtering pada Sistem Rekomendasi Buku Informatika. Jurnal Ilmiah SINUS, 22(2), 57. https://doi.org/10.30646/sinus.v22i2.840

Saifudin, I., & Widiyaningtyas, T. (2024). Systematic Literature Review on Recommender System: Approach, Problem, Evaluation Techniques, Datasets. IEEE Access, 12, 19827–19847. https://doi.org/10.1109/ACCESS.2024.3359274

Saifudin, I., Widiyaningtyas, T., Ari, I., Zaeni, E., Aminuddin, A., Abdullah, A.-S., & Kuantan, M. (2025). SVD-GoRank: Recommender System Algorithm using SVD and Gower's Ranking. IEEE Access, 13. https://doi.org/10.1109/ACCESS.2024.Doi

Syah, M., Wardani, A., Idhom, M., & Trimono. (2025). Perbandingan Representasi Teks TF-IDF dann Bert Terhadap Akurasi Cosine Similarity dalam Penilaian Otomatis Jawaban Berbasis Teks. Data Sciences Indonesia (DSI), 5(1), 47–59. https://doi.org/10.47709/dsi.v5i1.6021

Timur, A. Y., & Rohman, A. N. (2025). Sistem Rekomendasi Lagu Indonesia Menggunakan Metode Content-Based Filtering dan Cosine Similarity. Jurnal Informatika Dan Teknik Elektro Terapan, 13(1). https://doi.org/10.23960/jitet.v13i1.5949

Widiyaningtyas, T., Saifudin, I., Zaeni, I., Maulana, M., & Caesarendra, W. (2025). Memory-based collaborative filtering based on matrix factorization and Gower's set rank. Journal of King Saud University - Computer and Information Sciences, 37(9). https://doi.org/10.1007/s44443-025-00261-6