

Implementasi Content-Based Filtering, TF-IDF, dan Cosine Similarity pada Sistem Rekomendasi Artikel Berita Media Jelajah Nusantara

Ikmaliah Putri¹, Wawan Kurniawan²

^{1,2}Teknik Informatika, Universitas Dian Nusantara, Jakarta, Indonesia

Email: ¹411221100@mahasiswa.undira.ac.id, ²wawan.kurniawan@undira.ac.id

Email Penulis Korespondensi: 411221100@mahasiswa.undira.ac.id,

Info Artikel	Abstrak
Kata Kunci: Content-Based Filtering TF-IDF Cosine Similarity Sistem Rekomendasi Portal Berita	Portal berita digital saat ini menghadapi tantangan dalam menyajikan artikel yang sesuai dengan minat pengguna, seiring bertambahnya jumlah artikel yang terbit setiap harinya. Sebagian besar portal berita masih menampilkan artikel berdasarkan tingkat popularitas atau urutan waktu publikasi, sehingga rekomendasi yang dihasilkan belum mencerminkan preferensi individual masing-masing pembaca. Penelitian ini bertujuan membangun sistem rekomendasi artikel berita pada portal Media Jelajah Nusantara melalui penerapan metode <i>Content-Based Filtering</i> yang dipadukan pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF) dan perhitungan <i>Cosine Similarity</i>. Analisis kebutuhan sistem ditempuh menggunakan pendekatan kualitatif deskriptif. Profil minat pengguna disusun berdasarkan riwayat interaksi terhadap artikel, kemudian dijadikan acuan dalam menghasilkan rekomendasi berdasarkan tingkat kemiripan antara profil pengguna dan konten artikel. Performa sistem diukur menggunakan metrik <i>Precision@10</i>, <i>Recall@10</i>, F1-Score, <i>Mean Average Precision</i> (MAP), dan <i>Normalized Discounted Cumulative Gain</i> (NDCG@10). Hasil evaluasi menunjukkan bahwa <i>Content-Based Filtering</i> memperoleh <i>Precision@10</i> sebesar 0,1250, <i>Recall@10</i> sebesar 0,0286, F1-Score sebesar 0,0462, MAP sebesar 0,0151, dan NDCG@10 sebesar 0,1373, sementara metode Popularitas (<i>Baseline</i>) memperoleh nilai yang lebih tinggi pada seluruh metrik dalam dataset dan skenario pengujian yang digunakan, sehingga hasil ini perlu dipandang sebagai temuan empiris pada konteks pengujian tersebut, bukan sebagai bukti keunggulan mutlak salah satu metode. Kebaruan penelitian terletak pada integrasi pembentukan profil minat dari data operasional Media Jelajah Nusantara, mekanisme rekomendasi berbasis konten, serta evaluasi komparatif terhadap metode Popularitas dalam satu aplikasi yang sama. Temuan ini menegaskan bahwa <i>Content-Based Filtering</i> dapat diterapkan untuk menghasilkan rekomendasi personal berdasarkan profil minat pengguna, meskipun performa evaluasinya tetap dipengaruhi oleh karakteristik data serta mekanisme pembentukan profil yang digunakan.
	Abstract
Keywords: Content-Based Filtering TF-IDF Cosine Similarity Recommender System News Portal	Digital news portals currently face challenges in presenting articles that align with user interests, given the growing number of articles published daily. Most news portals still display articles based on popularity levels or publication order, so the resulting recommendations do not yet reflect each reader's individual preferences. This study aims to build a news article recommendation system for the Media Jelajah Nusantara portal by applying the Content-Based Filtering method combined with Term Frequency-Inverse Document Frequency (TF-IDF) weighting and Cosine Similarity calculation. A descriptive qualitative approach was used to analyze system requirements. User interest profiles were constructed based on interaction history with articles and served as the basis for generating recommendations according to the similarity level between user profiles and article content. System performance was measured using <i>Precision@10</i>, <i>Recall@10</i>, F1-Score, <i>Mean Average Precision</i> (MAP), and <i>Normalized Discounted Cumulative Gain</i> (NDCG@10) metrics. Evaluation results show that Content-Based Filtering achieved a <i>Precision@10</i> of 0.1250, <i>Recall@10</i> of 0.0286, F1-Score of 0.0462, MAP of 0.0151, and NDCG@10 of 0.1373, while the Popularity (Baseline) method obtained higher values across all metrics within the dataset and testing scenario used, indicating that these results should be regarded as empirical findings specific to this testing context rather than absolute evidence of one method's superiority. The novelty of this research lies in integrating user interest profiling based on operational data from Media Jelajah Nusantara, a content-based recommendation mechanism, and comparative evaluation against the Popularity method within a single application. These findings confirm that Content-Based Filtering can be applied to generate personalized recommendations based on user interest profiles, although its evaluation performance remains influenced by data characteristics and the profiling mechanism employed.



1. PENDAHULUAN

Perkembangan pesat teknologi informasi telah mengubah pola konsumsi berita masyarakat, dari yang semula bergantung pada media cetak menjadi beralih ke media digital yang dapat diakses tanpa batasan waktu melalui berbagai perangkat berbasis internet. Pergeseran ini mendorong berkembangnya portal berita digital sebagai kanal informasi utama yang unggul dalam kecepatan penyajian, pembaruan konten yang berkelanjutan, serta jangkauan audiens yang lebih luas dibandingkan media konvensional. Mengacu pada laporan Profil Internet Indonesia 2025 yang dipublikasikan APJII, populasi pengguna internet di Indonesia pada semester pertama 2025 tercatat telah melampaui 229 juta jiwa, naik dari sekitar 221 juta jiwa pada tahun sebelumnya [1]. Peningkatan jumlah pengguna digital ini turut memicu bertambahnya volume konten berita yang diterbitkan setiap harinya, sehingga potensi information overload bagi pembaca portal berita menjadi semakin nyata dan penting untuk dikaji, yaitu kondisi ketika volume informasi yang beredar melampaui kapasitas pengguna dalam menyeleksi artikel yang sesuai dengan kebutuhan dan preferensinya [2], [3]. Kondisi ini berdampak pada menurunnya efisiensi pencarian informasi, sebab pengguna dituntut mencurahkan lebih banyak waktu untuk menyaring artikel yang relevan di tengah melimpahnya informasi yang tersedia [4].

Persoalan tersebut mendorong berkembangnya kajian mengenai news recommender system sebagai upaya memperbaiki relevansi informasi bagi pengguna. Berbeda dari sistem penyajian berita konvensional yang umumnya hanya mengandalkan kategori atau kronologi waktu terbit, sistem rekomendasi dirancang untuk mengenali karakteristik dan preferensi pengguna, sehingga artikel yang ditampilkan dapat lebih selaras dengan minat masing-masing individu. Penelitian [2], [3] menegaskan bahwa personalisasi menjadi unsur penting dalam pengembangan portal berita masa kini, mengingat setiap individu memiliki pola konsumsi informasi yang berbeda-beda, sehingga sistem rekomendasi tidak sekadar membantu pengguna menemukan konten relevan, tetapi juga berkontribusi terhadap peningkatan keterlibatan pengguna, durasi membaca, serta kepuasan terhadap platform berita yang digunakan [2], [4].

Sejumlah penelitian terdahulu telah mengeksplorasi berbagai pendekatan dalam merancang sistem rekomendasi artikel berita. Penelitian [5] membangun sistem rekomendasi dengan Content-Based Filtering yang dipadukan pembobotan TF-IDF dan *Vector Similarity*, dan menemukan bahwa representasi kata melalui TF-IDF berkontribusi terhadap akurasi pencocokan artikel berbasis konten. Penelitian [6] menerapkan kombinasi TF-IDF dan *Cosine Similarity* pada Microsoft News Dataset, dengan hasil menunjukkan bahwa perpaduan kedua metode tersebut mampu menghasilkan rekomendasi berita yang relevan, dengan capaian hit rate sebesar 80,77%. Temuan serupa dilaporkan oleh peneliti [7] yang mengimplementasikan Content-Based Filtering pada sistem rekomendasi berita berbasis web, dan mendapati bahwa pendekatan berbasis konten menghasilkan rekomendasi lebih tepat sasaran dibandingkan penyajian artikel secara acak. Sementara itu, penelitian [8] memadukan *Part-of-Speech Tagging*, TF-IDF, dan *Cosine Similarity* untuk memperkaya representasi dokumen sebelum tahap pengukuran kemiripan, dan terbukti meningkatkan kualitas rekomendasi karena sistem lebih mampu menangkap informasi penting dari setiap artikel. Kombinasi TF-IDF dan *Cosine Similarity* juga telah diterapkan pada berbagai riset berbasis teks lainnya, seperti penelusuran dokumen [9]. Selain pada ranah berita, pendekatan ini juga diaplikasikan pada domain lain: Penelitian [10] mengembangkan sistem rekomendasi produk *e-commerce* dan menunjukkan bahwa perpaduan TF-IDF dan *Cosine Similarity* mampu memperbaiki relevansi rekomendasi melalui representasi fitur teks dan pengukuran kemiripan antarproduk, sementara pendekatan serupa turut diterapkan oleh penelitian [11] pada sistem rekomendasi destinasi wisata. Variasi pendekatan ini menandakan bahwa riset sistem rekomendasi berbasis konten terus berkembang guna meningkatkan kualitas rekomendasi yang dihasilkan.

Meskipun demikian, kajian-kajian terdahulu tersebut masih menyisakan sejumlah keterbatasan yang dapat dirumuskan sebagai berikut. Pertama, dari sisi objek data: penelitian [5], [6], [7] sebagian besar memanfaatkan dataset publik atau data uji coba (seperti Microsoft News Dataset), sehingga belum sepenuhnya mencerminkan kondisi operasional portal berita yang sesungguhnya; penelitian ini sebaliknya memanfaatkan data operasional nyata dari portal Media Jelajah Nusantara. Kedua, dari sisi representasi minat pengguna: Penelitian [2], [4] mengungkap bahwa proses pembentukan profil minat dari riwayat interaksi pengguna belum dijabarkan secara mendalam, sementara penelitian ini menguraikan proses tersebut secara eksplisit dengan mempertimbangkan bobot pada setiap jenis interaksi. Ketiga, dari sisi skema evaluasi: mayoritas penelitian terdahulu hanya melaporkan skor metrik metode yang diusulkan tanpa membandingkannya dengan metode pembandingan (*baseline*) non-personal, sehingga besarnya manfaat personalisasi belum dapat diukur secara relatif; penelitian ini menutup celah tersebut dengan menghadirkan skema evaluasi komparatif terhadap metode Popularitas (*Baseline*) dalam aplikasi yang sama. Penelitian [3] turut menegaskan bahwa salah satu

tantangan utama *news recommender system* adalah merepresentasikan preferensi pengguna yang mampu beradaptasi terhadap perubahan minat secara dinamis, agar rekomendasi tetap relevan sepanjang waktu.

Mengacu pada celah tersebut, penelitian ini menerapkan Content-Based Filtering dengan TF-IDF dan *Cosine Similarity* pada portal berita Media Jelajah Nusantara secara langsung pada lingkungan operasional, dengan membangun profil minat berdasarkan riwayat interaksi pengguna terhadap artikel. Setiap jenis interaksi diberi bobot yang berbeda, kemudian informasi judul dan deskripsi artikel dari riwayat interaksi tersebut dimanfaatkan dalam pembentukan profil minat pengguna, yang selanjutnya menjadi dasar penghitungan kemiripan melalui *Cosine Similarity* serta pemeringkatan artikel untuk menghasilkan rekomendasi yang lebih sesuai preferensi masing-masing pengguna.

Kontribusi metodologis penelitian ini tidak sebatas pada penerapan *Content-Based Filtering*, TF-IDF, dan *Cosine Similarity* pada Media Jelajah Nusantara, melainkan juga pada tersusunnya kerangka evaluasi yang menyatukan tiga tahapan dalam satu aplikasi, yaitu: (1) pembentukan profil minat dari riwayat interaksi pengguna, (2) mekanisme rekomendasi berbasis konten, dan (3) evaluasi performa rekomendasi secara komparatif terhadap metode Popularitas (Baseline). Kerangka evaluasi tiga tahap ini berpotensi diterapkan pada portal berita atau aplikasi berbasis teks lain dengan karakteristik serupa, sehingga kontribusi penelitian tidak hanya terletak pada implementasi sistem pada satu studi kasus, tetapi juga pada mekanisme evaluasi rekomendasi yang dihasilkan.

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan *Content-Based Filtering* menggunakan TF-IDF dan *Cosine Similarity* untuk menghasilkan rekomendasi artikel berita berbasis minat pengguna pada platform Media Jelajah Nusantara, melalui pembentukan profil minat pengguna dari riwayat interaksi, pembobotan dokumen dengan TF-IDF, pengukuran kemiripan melalui *Cosine Similarity*, hingga penyajian rekomendasi artikel yang relevan bagi setiap pengguna. Hasil penelitian ini diharapkan berkontribusi terhadap pengembangan sistem rekomendasi berita berbasis konten sekaligus menjadi acuan bagi penelitian selanjutnya pada domain berbasis teks

2. METODOLOGI PENELITIAN

2.1 Jenis Penelitian

Penelitian ini mengacu pada pendekatan kualitatif guna mengidentifikasi kebutuhan sistem rekomendasi artikel berita pada portal Media Jelajah Nusantara. Proses pengembangan sistem diawali dengan identifikasi permasalahan serta analisis kebutuhan melalui observasi, wawancara, dan studi dokumentasi. Hasil analisis tersebut menjadi landasan dalam perancangan aplikasi, sehingga sistem yang dikembangkan dapat selaras dengan kebutuhan pengguna sesungguhnya [12].

2.2 Objek Penelitian

Objek penelitian ini adalah portal berita Media Jelajah Nusantara yang menyajikan artikel dari beragam kategori berita. Dataset yang digunakan mencakup artikel beserta metadatanya, meliputi judul, isi artikel, kategori, ringkasan, tanggal publikasi, serta riwayat interaksi pengguna terhadap masing-masing artikel [4]. Dalam proses rekomendasi, informasi judul dan deskripsi artikel dimanfaatkan sebagai representasi konten, sementara riwayat interaksi pengguna dijadikan dasar dalam pembentukan profil minat sebagai landasan proses rekomendasi berbasis konten [2], [3]. Karakteristik dataset penelitian ini secara kuantitatif dirangkum pada Tabel 1.

Tabel 1. Karakteristik Dataset Penelitian

Karakteristik	Nilai
Artikel aktif	90
Pengguna yang dianalisis	4
Total interaksi	617 terdiri atas 504 riwayat baca, 73 <i>like</i> , 11 <i>bookmark</i> , 11 komentar, dan 18 <i>feedback</i>
Rata-rata interaksi per pengguna	154,25
Jumlah kategori	8
Periode artikel	11 Juni–28 Juli 2026
Periode riwayat baca	29 Juni–31 Juli 2026

Data telah melalui tahap pembersihan berupa penghapusan data duplikat dan data *orphan*. Artikel yang digunakan dibatasi pada artikel berstatus aktif, sedangkan pengguna yang dilibatkan dalam evaluasi disyaratkan memiliki minimal lima riwayat baca.

2.3 Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui observasi terhadap proses pengelolaan dan penyajian artikel pada Media Jelajah Nusantara, wawancara bersama administrator dan pengelola portal guna mengidentifikasi kebutuhan sistem, serta

studi dokumentasi terhadap artikel, metadata, dan data interaksi pengguna. Data hasil observasi dan wawancara dianalisis secara deskriptif, didukung analisis SWOT untuk memetakan faktor internal dan eksternal dalam pengembangan sistem [13].

2.4 Implementasi Metode Content-Based Filtering

Metode *Content-Based Filtering* diterapkan sebagai landasan utama dalam pembangunan sistem rekomendasi artikel berita pada portal ini [14]. Metode tersebut menghasilkan rekomendasi dengan mempertimbangkan karakteristik konten yang berkaitan dengan preferensi pengguna [15], sehingga artikel yang direkomendasikan memiliki kesesuaian dengan profil minat pengguna [5], [7]. Pembobotan kata dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), sedangkan tingkat kemiripan antara profil pengguna dan artikel dihitung menggunakan metode *Cosine Similarity* [6], [16]. Penerapan kedua metode tersebut ditempuh melalui tahapan berikut:

- Sistem menyusun profil minat berdasarkan riwayat interaksi dan bobot pada setiap jenis interaksi
- Teks judul dan deskripsi artikel diproses melalui *case folding*, *tokenizing*, *stopword removal*, dan *stemming*
- Sistem menghitung bobot kata dengan TF-IDF untuk membentuk representasi profil pengguna dan artikel
- Cosine Similarity* diterapkan untuk mengukur tingkat kemiripan antara profil pengguna dan artikel kandidat
- Artikel diurutkan berdasarkan nilai kemiripan tertinggi dan ditampilkan sebagai rekomendasi. [2].

2.5 Konfigurasi Preprocessing dan Parameter TF-IDF

Tahap preprocessing dan perhitungan TF-IDF diimplementasikan secara langsung menggunakan bahasa pemrograman PHP tanpa memanfaatkan pustaka eksternal. *Case folding* dilakukan dengan mengubah seluruh teks menjadi huruf kecil, dilanjutkan dengan penghapusan tag HTML dan karakter yang tidak diperlukan. *Tokenizing* ditempuh berdasarkan spasi, sementara *stopword removal* menerapkan daftar kata Bahasa Indonesia dan Inggris yang telah didefinisikan dalam sistem. Proses *stemming* menggunakan fungsi pemotongan sufiks Bahasa Indonesia yang dibangun secara manual. Nilai TF dihitung berdasarkan frekuensi kemunculan token terhadap jumlah keseluruhan token dalam dokumen, sedangkan IDF dihitung menggunakan rumus $IDF(t) = \ln\left(\frac{N}{df(t)+1}\right) + 1$. Bobot akhir setiap token diperoleh dari hasil perkalian TF dan IDF, dengan representasi kata berbentuk unigram. Pada profil produksi, bobot token dari seluruh interaksi pengguna dijumlahkan, dan token dengan bobot kurang dari atau sama dengan nol dihapus sebelum proses rekomendasi dijalankan.

2.6 Rancangan Eksperimen dan Teknik Pengujian

Evaluasi sistem menerapkan skema *hold-out* dengan proporsi 80:20 yang disusun secara kronologis pada setiap pengguna, yaitu 80% riwayat baca paling awal digunakan sebagai data latih, sedangkan 20% riwayat terbaru dijadikan *ground truth* data uji. Empat pengguna dengan minimal lima riwayat baca dievaluasi menggunakan $K=10$. Pada modul evaluasi, profil pengguna dibentuk dari artikel data latih dengan bobot yang seragam, berbeda dengan profil produksi yang memanfaatkan berbagai jenis interaksi berbobot. Metode Popularitas (*Baseline*) digunakan sebagai pembandingan non-personal dengan skor dihitung dari $Skor\ Popularitas = (like \times 2) + (komentar \times 3) + (kunjungan \times 1)$. Performa kedua metode diukur menggunakan metrik Precision@10, Recall@10, F1-Score, *Mean Average Precision* (MAP), dan *Normalized Discounted Cumulative Gain* (NDCG@10) [2], [3].

3. HASIL DAN PEMBAHASAN

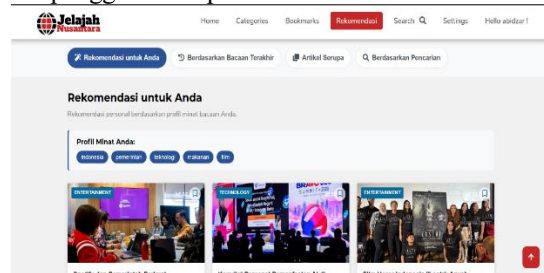
3.1 Hasil Penelitian

a. Implementasi Sistem

Pengembangan aplikasi web rekomendasi artikel berita pada portal Media Jelajah Nusantara telah berhasil diwujudkan dengan sejumlah fungsi utama, meliputi pembentukan profil minat pengguna berdasarkan riwayat interaksi pengguna terhadap artikel, penyajian rekomendasi artikel melalui penerapan metode *Content-Based Filtering* yang dipadukan pembobotan TF-IDF dan *Cosine Similarity*, serta halaman evaluasi guna mengukur performa sistem rekomendasi yang dibangun. Ketiga fungsi tersebut saling terhubung dalam satu alur kerja aplikasi, mulai dari pengambilan data interaksi pengguna, pengolahan konten artikel, hingga penyajian hasil rekomendasi dan pengukuran performanya.

Halaman rekomendasi menjadi bagian inti dari implementasi penelitian ini, sebab halaman tersebut menampilkan artikel-artikel yang telah disesuaikan dengan profil minat masing-masing pengguna. Sistem menyusun profil minat berdasarkan riwayat interaksi pengguna terhadap artikel, dengan mempertimbangkan bobot pada setiap jenis interaksi yang dilakukan, seperti aktivitas membaca, memberi *like*, menyimpan (*bookmark*), maupun memberikan komentar.

Informasi judul dan deskripsi artikel dari riwayat interaksi tersebut dimanfaatkan sebagai representasi teks dalam pembentukan profil, kemudian sistem menghasilkan daftar artikel berdasarkan tingkat kemiripan terhadap profil pengguna yang telah terbentuk. Selain menampilkan rekomendasi, sistem turut menyajikan kata kunci yang merepresentasikan minat pengguna, sehingga karakteristik profil yang digunakan dalam proses rekomendasi dapat diketahui secara lebih transparan oleh pengguna maupun administrator sistem.



Gambar 1. Halaman Rekomendasi Artikel

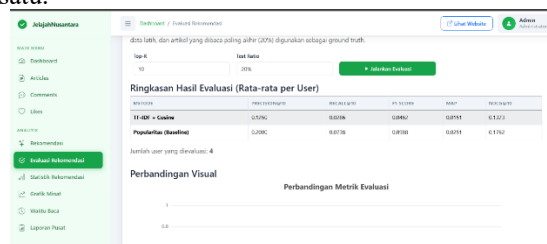
Untuk memperjelas alur kerja sistem secara visual, tahapan pembentukan rekomendasi mulai dari riwayat interaksi pengguna hingga penyajian Top-N rekomendasi digambarkan pada Gambar 2.



Gambar 2. Diagram Alur Proses *Content-Based Filtering* (TF-IDF + *Cosine Similarity*)

Diagram tersebut menampilkan urutan proses secara menyeluruh, dimulai dari pengumpulan riwayat interaksi, pembentukan profil minat, pembobotan dokumen menggunakan TF-IDF, penghitungan kemiripan dengan Cosine Similarity, hingga pemerinkatan dan penyajian artikel yang direkomendasikan kepada pengguna.

Selain halaman rekomendasi, sistem turut menyediakan halaman evaluasi yang berfungsi menampilkan hasil pengukuran performa metode rekomendasi yang diterapkan. Pada halaman ini, disajikan perbandingan hasil evaluasi antara metode Content-Based Filtering yang menggunakan TF-IDF dan Cosine Similarity dengan metode Popularitas (Baseline), berdasarkan sejumlah metrik meliputi Precision@10, Recall@10, F1-Score, MAP, dan NDCG@10. Selain berupa tabel ringkasan, sistem juga menampilkan visualisasi data hasil evaluasi dalam bentuk grafik, sehingga memudahkan administrator dalam menganalisis performa metode rekomendasi secara lebih menyeluruh tanpa perlu membaca angka mentah satu per satu.



Gambar 3. Halaman Evaluasi Sistem

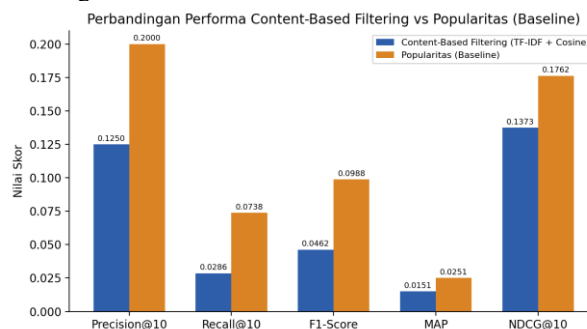
b. Hasil Evaluasi Sistem

Evaluasi dilaksanakan guna mengukur kualitas rekomendasi yang dihasilkan oleh metode *Content-Based Filtering* dengan TF-IDF dan *Cosine Similarity*. Pengujian menerapkan skema *hold-out* dengan proporsi 80:20 secara kronologis pada empat pengguna, dengan riwayat baca paling awal digunakan sebagai data latih dan riwayat baca terbaru sebagai data uji atau *ground truth*. Skema kronologis ini dipilih karena lebih mencerminkan kondisi penggunaan nyata, yaitu sistem memprediksi artikel yang akan dibaca pengguna di masa mendatang berdasarkan pola bacaan yang telah terjadi sebelumnya. Hasil rekomendasi selanjutnya dibandingkan dengan metode Popularitas (*Baseline*) menggunakan metrik Precision@10, Recall@10, F1-Score, *Mean Average Precision* (MAP), dan *Normalized Discounted Cumulative Gain* (NDCG@10). Ringkasan hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Sistem Rekomendasi

Metode	Precision@10	Recall@10	F1-Score	MAP	NDCG@10
Content-Based Filtering (TF-IDF + Cosine Similarity)	0,1250	0,0286	0,0462	0,0151	0,1373
Popularitas (Baseline)	0,2000	0,0738	0,0988	0,0251	0,1762

Perbandingan kelima metrik evaluasi antara kedua metode divisualisasikan pada Gambar 4 untuk memudahkan interpretasi selisih performa secara visual, sehingga pola perbedaan antarmetrik dapat diamati secara lebih langsung dibandingkan hanya membaca tabel angka.



Gambar 4. Grafik Perbandingan Metrik Evaluasi Content-Based Filtering vs Popularitas (Baseline)

Mengacu pada Tabel 2, metode *Content-Based Filtering* terbukti mampu menghasilkan rekomendasi artikel yang didasarkan pada karakteristik konten yang sebelumnya pernah diakses oleh pengguna. Nilai-nilai evaluasi yang diperoleh mengindikasikan bahwa sistem tetap dapat menyajikan rekomendasi yang relevan, kendati pada dataset penelitian ini metode Popularitas (Baseline) memperlihatkan capaian nilai yang lebih unggul pada seluruh metrik evaluasi yang digunakan. Secara kuantitatif, selisih Precision@10 antara kedua metode tercatat sebesar 0,0750, setara dengan 37,5% dari skor Popularitas, sementara selisih pada MAP sebesar 0,0100 dan pada NDCG@10 sebesar 0,0389. Pola ini menunjukkan bahwa kesenjangan terbesar terjadi pada metrik yang peka terhadap cakupan hasil (Recall) dan urutan relevansi awal (Precision, MAP), sedangkan kesenjangan pada NDCG@10 relatif lebih kecil dibandingkan metrik lainnya. Hal ini mengindikasikan bahwa artikel yang direkomendasikan oleh Content-Based Filtering, meskipun jumlah yang benar-benar relevan lebih sedikit dibandingkan metode Popularitas, cenderung tetap ditempatkan pada posisi peringkat yang cukup baik ketika artikel tersebut memang relevan bagi pengguna.

Kesenjangan hasil tersebut mengindikasikan bahwa performa sistem rekomendasi dipengaruhi oleh tiga faktor utama yang saling berkaitan. Pertama, karakteristik data yang digunakan, khususnya jumlah interaksi pengguna yang tercatat masih relatif sedikit, sehingga vektor profil minat menjadi jarang terisi (sparse) dan representasinya belum sepenuhnya menggambarkan preferensi pengguna secara utuh. Kondisi ini wajar terjadi pada tahap awal penerapan sistem rekomendasi berbasis konten, ketika riwayat interaksi pengguna belum terkumpul dalam jumlah besar. Kedua, kualitas representasi konten artikel dalam proses pembentukan profil pengguna, mengingat metode berbasis TF-IDF sangat bergantung pada kekayaan kosakata yang tumpang tindih (vocabulary overlap) antarartikel, maka semakin sedikit kesamaan istilah antara artikel yang pernah dibaca dan artikel kandidat, semakin rendah pula nilai kemiripan yang dihasilkan meskipun secara topikal keduanya sebenarnya berkaitan. Ketiga, karakteristik metrik evaluasi *top-N* yang menghitung *Precision* dan *Recall* berdasarkan kecocokan eksak dengan *ground truth*, sehingga metode Popularitas yang cenderung merekomendasikan artikel dengan jumlah pembaca tertinggi memiliki peluang lebih besar untuk mencakup artikel yang memang paling banyak dibaca, dan karenanya lebih sering muncul sebagai *ground truth*, dibandingkan metode personalisasi pada kondisi jumlah data yang masih terbatas.

Temuan evaluasi ini selanjutnya menjadi landasan pada bagian pembahasan untuk mengkaji lebih lanjut faktor-faktor yang memengaruhi performa metode *Content-Based Filtering*, sekaligus membandingkannya dengan hasil penelitian-penelitian terdahulu yang menerapkan pendekatan serupa pada domain yang berbeda.

3.2 Pembahasan

a. Hubungan Hasil Dengan Tujuan Penelitian

Penelitian ini bertujuan mengembangkan sistem rekomendasi artikel berita berbasis *Content-Based Filtering* yang mampu menyajikan rekomendasi sesuai minat pengguna pada portal Media Jelajah Nusantara. Hasil implementasi

menunjukkan bahwa sistem berhasil membentuk profil minat berdasarkan riwayat interaksi pengguna, sekaligus menghasilkan rekomendasi artikel berdasarkan kemiripan antara profil pengguna dan konten artikel. Dengan demikian, tujuan utama penelitian dalam membangun sistem rekomendasi berbasis konten telah tercapai, meskipun capaian nilai evaluasinya masih perlu ditingkatkan pada penelitian selanjutnya.

Penerapan TF-IDF yang dipadukan dengan *Cosine Similarity* memungkinkan sistem mengukur tingkat kemiripan antara representasi profil minat pengguna dan konten artikel kandidat secara numerik. Pendekatan ini memungkinkan rekomendasi disesuaikan dengan karakteristik minat masing-masing pengguna secara individual, berbeda dari metode Popularitas yang tidak menyusun profil individual dan hanya mengandalkan tingkat keterlibatan pembaca secara agregat. Mekanisme tersebut selaras dengan prinsip *Content-Based Filtering* yang memanfaatkan karakteristik konten dan preferensi pengguna dalam menghasilkan rekomendasi [5], [7].

b. Faktor Performa dan Keterbatasan Inheren Metode

Performa *Content-Based Filtering* yang lebih rendah dibandingkan Popularitas (*Baseline*) perlu dipahami berdasarkan karakteristik skenario evaluasi yang digunakan, bukan semata-mata sebagai kelemahan metode itu sendiri. Evaluasi dilakukan terhadap empat pengguna dengan pemisahan riwayat baca secara kronologis, yaitu 80% data awal sebagai data latih dan 20% data terbaru sebagai *ground truth*. Pada kondisi tersebut, profil evaluasi hanya dibentuk dari artikel pada data latih dengan bobot yang seragam untuk seluruh jenis interaksi. Kondisi ini berbeda dari sistem produksi yang memanfaatkan aktivitas membaca, *like*, *bookmark*, komentar, dan *feedback* dengan bobot yang berbeda-beda sesuai tingkat kepentingannya masing-masing. Perbedaan mekanisme antara skema evaluasi dan sistem produksi tersebut menjadi salah satu keterbatasan evaluasi, karena hasil pengukuran yang dilaporkan belum sepenuhnya merepresentasikan seluruh informasi interaksi yang sesungguhnya digunakan pada sistem produksi.

Di samping itu, *Content-Based Filtering* secara alami memiliki kecenderungan menghasilkan rekomendasi yang dekat dengan karakteristik konten yang telah membentuk profil pengguna sebelumnya. Kondisi ini berpotensi membatasi keberagaman rekomendasi yang dihasilkan, terutama ketika preferensi pengguna berubah seiring waktu, atau ketika artikel yang relevan pada data uji memiliki karakteristik konten yang cukup berbeda dari artikel-artikel pada data latih, sebagaimana tantangan personalisasi berita yang juga dibahas oleh [2], [3]. Selain itu, penelitian ini juga belum melakukan uji signifikansi statistik terhadap perbedaan performa kedua metode, mengingat jumlah pengguna yang dievaluasi hanya empat orang sehingga generalisasi hasil masih terbatas. Oleh karena itu, hasil perbandingan pada penelitian ini perlu dipahami sebagai temuan empiris pada dataset dan skenario pengujian yang digunakan, bukan sebagai bukti umum bahwa metode Popularitas secara keseluruhan lebih unggul dibandingkan *Content-Based Filtering* pada kondisi data yang lebih besar sekalipun.

c. Perbandingan Dengan Penelitian Terdahulu

Tabel 3. Perbandingan Penelitian dengan Penelitian Terdahulu

Penelitian	Metode	Fokus Penelitian	Perbedaan dengan Penelitian Ini
Huda dkk	Content-Based Filtering, TF-IDF, <i>Vector Similarity</i>	Mengembangkan sistem rekomendasi artikel berita berdasarkan kemiripan konten.	Penelitian ini menggunakan <i>Cosine Similarity</i> sebagai metode pengukuran kemiripan serta mengevaluasi sistem menggunakan <i>Precision@10</i> , <i>Recall@10</i> , <i>F1-Score</i> , MAP, dan <i>NDCG@10</i> .
Yunanda dkk	TF-IDF, <i>Cosine Similarity</i>	Mengembangkan sistem rekomendasi berita menggunakan Microsoft News Dataset.	Penelitian ini menggunakan data operasional portal Media Jelajah Nusantara dan membentuk profil minat pengguna berdasarkan riwayat interaksi pengguna terhadap artikel.
Haz dkk	Content-Based Filtering	Mengembangkan sistem rekomendasi berita berbasis web.	Penelitian ini menambahkan evaluasi performa serta perbandingan dengan metode Popularitas (Baseline) .
Wiratama dkk	POS Tagging, TF-IDF, <i>Cosine Similarity</i>	Meningkatkan representasi dokumen sebelum proses rekomendasi.	Penelitian ini tidak menggunakan POS Tagging, tetapi memanfaatkan riwayat interaksi pengguna sebagai dasar pembentukan profil minat.

Merujuk pada Tabel 3, penelitian ini menunjukkan kesamaan dengan penelitian terdahulu dalam hal penerapan *Content-Based Filtering* untuk menghasilkan rekomendasi berbasis konten. Perbedaannya terletak pada penerapan sistem secara langsung pada portal Media Jelajah Nusantara, dengan memanfaatkan data operasional dan riwayat interaksi pengguna dalam pembentukan profil minat, bukan sekadar dataset uji coba sebagaimana pada sebagian besar penelitian terdahulu. Penelitian ini juga mengevaluasi metode menggunakan *Precision@10*, *Recall@10*, *F1-Score*, MAP, dan

NDCG@10, serta membandingkan hasilnya dengan metode Popularitas (Baseline) sebagai pembanding non-personal. Perbandingan tersebut memberikan gambaran yang lebih jelas mengenai performa pendekatan personal dan non-personal pada dataset operasional yang digunakan.

Kontribusi penelitian ini terletak pada integrasi pembentukan profil minat pengguna, mekanisme rekomendasi berbasis Content-Based Filtering, dan evaluasi komparatif terhadap metode Popularitas dalam satu kesatuan aplikasi. Kebaruan penelitian ini terletak pada penerapan kerangka tersebut menggunakan data operasional Media Jelajah Nusantara, sehingga performa pendekatan berbasis profil pengguna dapat dibandingkan secara langsung dengan pendekatan non-personal pada lingkungan portal berita yang menjadi objek penelitian, bukan pada data simulasi. Hasil evaluasi menunjukkan bahwa personalisasi tidak secara otomatis menghasilkan nilai metrik yang lebih tinggi, karena performanya turut dipengaruhi oleh karakteristik data serta mekanisme pembentukan profil pengguna yang digunakan. Secara praktis, temuan ini menegaskan pentingnya mempertimbangkan ketersediaan dan kelengkapan riwayat pengguna ketika hendak menerapkan sistem rekomendasi pada portal berita berskala kecil hingga menengah.

Penelitian selanjutnya disarankan menggunakan jumlah pengguna dan data interaksi yang lebih besar agar profil minat yang terbentuk lebih representatif, menyelaraskan mekanisme pembentukan profil antara sistem produksi dan skema evaluasi agar hasil pengukuran lebih mencerminkan kondisi sesungguhnya, serta mengembangkan pendekatan *Hybrid Recommender System*. Pendekatan *hybrid* dapat dipertimbangkan mengingat penelitian terkini menunjukkan potensinya dalam menggabungkan informasi dari beberapa pendekatan rekomendasi sekaligus, sehingga dapat saling menutupi kelemahan masing-masing metode ketika diterapkan secara tunggal [17], [18].

Secara keseluruhan, hasil penelitian ini memperlihatkan bahwa penerapan *Content-Based Filtering* pada portal berita berskala kecil dengan jumlah data interaksi yang masih terbatas tetap dapat menghasilkan rekomendasi yang relevan dan bersifat personal, meskipun belum memperoleh nilai yang lebih tinggi dibandingkan metode Popularitas pada seluruh metrik evaluasi. Temuan ini memberikan gambaran realistis mengenai tantangan penerapan sistem rekomendasi berbasis konten pada tahap awal operasional sebuah portal berita, sekaligus membuka ruang bagi pengembangan lebih lanjut, baik dari sisi jumlah data, mekanisme pembobotan interaksi, maupun kombinasi dengan pendekatan rekomendasi lain yang bersifat kolaboratif.

4. KESIMPULAN

Penelitian ini berhasil menghasilkan sistem rekomendasi artikel berita pada portal Media Jelajah Nusantara melalui penerapan metode *Content-Based Filtering* yang dipadukan dengan pembobotan TF-IDF dan *Cosine Similarity*. Sistem yang dibangun mampu menyusun profil minat berdasarkan riwayat interaksi pengguna, sekaligus menghasilkan rekomendasi artikel yang bersifat personal berdasarkan tingkat kemiripan antara profil pengguna dan konten artikel. Hasil implementasi menunjukkan bahwa seluruh fungsi utama sistem dapat berjalan sebagaimana mestinya, mulai dari pembentukan profil pengguna, mekanisme rekomendasi, hingga proses evaluasi menggunakan Precision@10, Recall@10, F1-Score, Mean Average Precision (MAP), dan Normalized Discounted Cumulative Gain (NDCG@10).

Kontribusi ilmiah penelitian ini dapat dirangkum ke dalam tiga poin utama. Pertama, penelitian ini menerapkan Content-Based Filtering secara langsung pada data operasional portal berita nyata, bukan pada dataset publik sebagaimana lazim dilakukan pada penelitian-penelitian sebelumnya. Kedua, penelitian ini mengintegrasikan tiga tahapan dalam satu kesatuan aplikasi, yaitu pembentukan profil minat dari riwayat interaksi pengguna, mekanisme rekomendasi berbasis konten, dan evaluasi performa secara komparatif terhadap metode Popularitas (*Baseline*). Ketiga, kerangka tersebut memberikan gambaran empiris mengenai perbandingan performa pendekatan personal dan non-personal pada lingkungan portal berita berskala kecil, sekaligus memperlihatkan bahwa personalisasi tidak selalu menghasilkan nilai metrik yang lebih tinggi, karena performanya turut dipengaruhi oleh karakteristik data dan mekanisme pembentukan profil yang digunakan. Secara praktis, temuan ini menegaskan pentingnya mempertimbangkan ketersediaan riwayat interaksi pengguna sebelum menerapkan sistem rekomendasi berbasis konten pada portal berita dengan skala pengguna yang masih terbatas.

Penelitian ini memiliki sejumlah keterbatasan, yaitu jumlah pengguna yang dievaluasi hanya sebanyak empat orang, adanya perbedaan mekanisme pembentukan profil antara sistem produksi dan modul evaluasi, serta belum dilakukannya uji signifikansi statistik terhadap perbedaan performa kedua metode yang dibandingkan. Oleh karena itu, penelitian selanjutnya disarankan untuk memanfaatkan jumlah pengguna dan data interaksi yang lebih besar, menyelaraskan mekanisme pembentukan profil antara sistem produksi dan skema evaluasi, melakukan uji signifikansi statistik guna memperkuat validitas perbandingan, serta mengembangkan pendekatan *Hybrid Recommender System* dengan mengombinasikan *Content-Based Filtering* dan *Collaborative Filtering*.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Media Jelajah Nusantara atas izin yang diberikan untuk melaksanakan penelitian ini, serta kepada Universitas Dian Nusantara sebagai institusi tempat penulis menempuh studi dan menyelesaikan penelitian ini.

REFERENCES

- [1] A. P. J. I. I. (APJII), "Profil Internet Indonesia 2025," 2025. [Online]. Available: <https://www.apjii.or.id/>
- [2] X. Meng, H. Huo, X. Zhang, W. Wang, and J. Zhu, "A Survey of Personalized News Recommendation," *Data Sci. Eng.*, vol. 8, no. 4, pp. 396–416, 2023, doi: 10.1007/s41019-023-00228-5.
- [3] S. Raza and C. Ding, *News recommender system: a review of recent progress, challenges, and opportunities*, vol. 55, no. 1. Springer Netherlands, 2022. doi: 10.1007/s10462-021-10043-x.
- [4] C. Wu, F. Wu, Y. Huang, and X. Xie, "Personalized News Recommendation: Methods and Challenges," *ACM Trans. Inf. Syst.*, vol. 41, no. 1, 2021, doi: 10.1145/3530257.
- [5] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.
- [6] G. Yunanda, D. Nurjanah, and S. Meliana, "Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 277–284, 2022, doi: 10.47065/bits.v4i1.1670.
- [7] A. M. R. Haz and A. N. Rohman, "Sistem Rekomendasi Berita Dengan Metode Content-Based Filtering," *JIKA (Jurnal Inform.)*, vol. 9, no. 2, p. 143, 2025, doi: 10.31000/jika.v9i2.13247.
- [8] A. K. Wiratama and Z. K. A. Baizal, "Optimizing News Recommendations: Utilizing POS-Tagging and Content-Based Methods to Enhance Personalization in News Recommendations," *J. Inf. Syst. Res.*, vol. 6, no. 1, pp. 117–124, 2024, doi: 10.47065/josh.v6i1.5761.
- [9] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks," *Sintesia*, vol. 1, pp. 81–88, 2022.
- [10] S. I. Adam and W. G. Mokodaser, "Rekomendasi Produk E-Commerce Menggunakan Content-Based Filtering Berbasis Cosine Similarity," *J. Sist. Inf. dan Tek. Komput.*, vol. 10, no. 2, pp. 418–424, 2025.
- [11] S. Oyadila, D. Abdullah, and A. Razi, "Implementasi Content-Based Filtering Dengan Tf-Idf Dan Cosine Similarity Untuk Sistem Rekomendasi Destinasi Wisata Di Aceh Tengah," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1329–1339, 2025, doi: 10.36341/rabit.v10i2.6532.
- [12] A. A. Prawanti, I. Kumala, M. N. Fadillah, R. Y. S. Damanik, and R. N. Lubis, "Desain, Jenis, dan Metode Dalam Penelitian Kualitatif," *Tarb. J. Ilmu Pendidik. dan Pengajaran*, vol. 4, no. 1, pp. 384–388, 2025, doi: 10.64464/tarbiyah.v4i1.186.
- [13] A. Wardana and N. A. Wiyani, "Analisis Swot Terhadap Implementasi Pembelajaran Tematik Terpadu di MIN 3 Purbalingga," *Qalam J. Ilmu ...*, vol. 11, no. 1, pp. 1–8, 2022, doi: 10.33506/jq.v11i1.1593.
- [14] E. W. Pratomo, E. Utami, and R. Artikel, "Analisis dan implementasi Content-Based Filtering dengan Cosine Similarity untuk sistem rekomendasi tugas akhir mahasiswa," *AITI J. Teknol. Inf.*, vol. 23, no. Juni, pp. 319–333, 2026, doi: 10.24246/aiti.v23.i2.319-333.
- [15] L. Rosidah and P. Dellia, "Library Book Recommendation System Using Content-Based Filtering," *Internet Things Artif. Intell. J.*, vol. 4, no. 1, pp. 42–65, 2024, doi: 10.31763/iota.v4i1.693.
- [16] Z. F. Meillano, D. Richasdy, and H. Hasmawati, "Web-Based Movie Recommendation System Using Content-Based Filtering and Cosine Similarity," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, p. 157, 2023, doi: 10.24014/ijaidm.v7i1.27390.
- [17] O. Remadnia, F. Maazouzi, and D. Chefrour, "Hybrid Book Recommendation System Using Collaborative Filtering and Embedding Based Deep Learning," *Inform.*, vol. 49, no. 8, pp. 189–204, 2025, doi: 10.31449/inf.v49i8.6950.
- [18] A. Sami, W. El Adrousy, S. Sarhan, and S. Elmougy, "A deep learning based hybrid recommendation model for internet users," *Sci. Rep.*, vol. 14, no. 1, pp. 1–20, 2024, doi: 10.1038/s41598-024-79011-z.