

Implementasi Pendekatan *Hybrid Filtering* pada Sistem Rekomendasi Berita Sepak Bola berbasis *Website*

Implementation of a Hybrid Filtering Approach in a Website-based Football News Recommendation System

¹Krisna Haridarma Putra*, ²Arif Nur Rohman, ³Norhikmah

^{1,2,3}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
^{1,2,3}Jl. Ring Road Utara, Ngringin, Condongcatut, Kec. Depok, Sleman, Daerah Istimewa Yogyakarta, Indonesia

*e-mail: Krissnadarma@students.amikom.ac.id

(received: 14 November 2025, revised: 7 December 2025, accepted: 8 December 2025)

Abstrak

Peningkatan jumlah berita sepak bola pada portal digital menyebabkan pengguna mengalami kesulitan menemukan informasi yang sesuai dengan minat mereka. Penelitian ini mengembangkan sistem rekomendasi berita berbasis web menggunakan metode *Content-Based Filtering* dan *Collaborative Filtering* yang digabungkan melalui pendekatan *Hybrid Filtering* pada level fitur. Pendekatan hybrid ini menjadi kebaruan utama penelitian karena tidak menggunakan penggabungan skor seperti pada penelitian sebelumnya, sehingga lebih ringan, sederhana, dan tetap cocok diterapkan pada dataset kecil serta interaksi pengguna yang terbatas. Sistem memanfaatkan *Term Overlap Matching* untuk menghitung kemiripan judul berita dan *Cosine Similarity* untuk mengukur kesamaan preferensi pengguna berdasarkan data bookmark. Hasil evaluasi menunjukkan bahwa metode *Content-Based Filtering* memiliki performa terbaik dengan Precision sebesar 0,60, Recall 0,75, dan F1-Score 0,67, sedangkan metode *Collaborative Filtering* belum efektif dengan Precision 0, Recall 0, dan F1-Score 0 akibat data sparsity pada interaksi pengguna. Secara keseluruhan, pendekatan hybrid berbasis fitur mampu memberikan rekomendasi yang relevan baik dari sisi konten maupun preferensi, meskipun akurasi sistem masih didominasi oleh *Content-Based Filtering*. Temuan ini menunjukkan bahwa model hybrid sederhana yang diusulkan dapat menjadi solusi efektif untuk platform berita olahraga berskala kecil, serta berpotensi ditingkatkan melalui penambahan data, peningkatan interaksi pengguna, dan penerapan teknik NLP yang lebih kaya.

Kata kunci: berita sepak bola, collaborative filtering, content-based filtering, hybrid filtering, sistem rekomendasi

Abstract

The rapid growth of football news on digital portals has made it increasingly difficult for users to find information that matches their interests. This study develops a web-based news recommendation system by combining *Content-Based Filtering* and *Collaborative Filtering* through a feature-level *Hybrid Filtering* approach. The proposed hybrid approach constitutes the main novelty of this research, as it does not rely on score aggregation methods commonly used in previous studies, making it lighter, simpler, and more suitable for small datasets and limited user interactions. The system employs *Term Overlap Matching* to measure the similarity between news titles and *Cosine Similarity* to assess user preference similarity based on bookmark data. The evaluation results show that *Content-Based Filtering* achieves the best performance, with a Precision of 0.60, Recall of 0.75, and an F1-score of 0.67, while *Collaborative Filtering* performs poorly due to data sparsity in user interactions, resulting in a Precision of 0, Recall of 0, and an F1-score of 0. Overall, the feature-based hybrid approach is able to provide relevant recommendations from both content and preference perspectives, although system accuracy is still predominantly driven by *Content-Based Filtering*. These findings indicate that the proposed simple hybrid model can serve as an effective solution for small-scale sports news platforms and has the potential to be further improved through increased data availability, enhanced user interaction, and the adoption of more advanced NLP techniques.

Keywords: collaborative filtering, content-based filtering, football news, hybrid filtering, recommendation system

<http://sistemasi.ftik.unisi.ac.id>

1 Pendahuluan

Perkembangan teknologi informasi telah mengubah cara masyarakat mengakses berita olahraga, termasuk sepak bola, yang diperbarui secara cepat setiap harinya. Banyaknya konten dari berbagai portal membuat pengguna kesulitan menemukan informasi yang sesuai minat mereka, sehingga memunculkan fenomena *information overload* ketika mencari berita yang relevan dengan minat mereka [1]. Kondisi ini menuntut adanya sistem yang mampu menyajikan berita yang relevan sesuai minat pengguna. Sistem rekomendasi menjadi salah satu solusi penting untuk membantu pengguna menemukan informasi secara lebih efisien tanpa harus menelusuri seluruh konten yang tersedia. Dua pendekatan yang paling umum digunakan adalah *Content-Based Filtering* dan *Collaborative Filtering* [2]. *Content-based* bekerja dengan membandingkan karakteristik konten, biasanya menggunakan teks seperti kata kunci pada judul berita atau isi. Pendekatan ini unggul karena tidak membutuhkan data pengguna lain, mudah diterapkan pada dataset kecil, dan memberikan rekomendasi yang konsisten. Sementara itu, *collaborative filtering* merekomendasikan berita berdasarkan kesamaan perilaku antar pengguna dengan preferensi yang serupa, tetapi sangat bergantung pada banyaknya interaksi pengguna [3], [4].

Sejumlah penelitian sebelumnya telah mencoba menggabungkan kedua pendekatan tersebut melalui skema *hybrid*, umumnya dalam bentuk penggabungan bobot atau skor rekomendasi [5]. Namun, pendekatan berbasis pembobotan sering menambah kompleksitas, membutuhkan parameter tambahan, dan kurang sesuai jika data terbatas. Berdasarkan analisis tersebut, penelitian ini memilih penerapan Hybrid Filtering pada level fitur (*feature-level hybrid*), di mana *Content-Based Filtering* dan *Collaborative Filtering* dijalankan secara terpisah sesuai fungsi masing-masing fitur dalam sistem. Pendekatan ini lebih ringan, tidak memerlukan proses penggabungan skor, dan tetap dapat memanfaatkan keunggulan kedua metode untuk menghasilkan rekomendasi yang relevan pada konteks yang berbeda.

Berbagai penelitian telah menerapkan *Content-Based Filtering*, *Collaborative Filtering*, maupun *hybrid filtering* pada domain e-commerce, rekomendasi wisata, maupun perpustakaan digital [6]. Namun, belum banyak penelitian yang menerapkan *hybrid filtering* pada domain berita sepak bola yang bersifat sangat dinamis, di mana konten diperbarui dengan cepat dan interaksi pengguna pada platform lokal biasanya masih terbatas. Selain itu, sebagian besar penelitian sebelumnya menggunakan dataset statis dan tidak memanfaatkan interaksi nyata pengguna seperti bookmark. Oleh karena itu, terdapat *research gap* berupa kebutuhan model rekomendasi hybrid yang sederhana, ringan, dan tetap efektif pada sistem berita olahraga dengan jumlah data dan interaksi pengguna yang terbatas.

Penelitian ini berkontribusi dengan mengembangkan sistem rekomendasi berita sepak bola berbasis web menggunakan pendekatan hybrid pada level fitur, di mana *Content-Based Filtering* digunakan untuk menghasilkan rekomendasi berdasarkan kemiripan judul berita, sementara *Collaborative filtering* digunakan untuk personalisasi berdasarkan pola bookmark pengguna. Pendekatan ini diharapkan dapat menjawab keterbatasan masing-masing metode dan meningkatkan relevansi rekomendasi pada lingkungan data kecil serta konteks berita olahraga yang cepat berubah [7], [8].

2 Tinjauan Literatur

Penelitian terkait sistem rekomendasi telah berkembang secara signifikan, terutama dalam penerapan dua metode utama, yaitu *Content-Based Filtering* dan *Collaborative Filtering*. Beberapa studi menunjukkan bahwa *Content-Based Filtering* efektif dalam memetakan kesamaan konten melalui analisis teks. Anando dan Arif, misalnya, menekankan bahwa preprocessing seperti case folding, tokenizing, dan pembersihan teks sangat memengaruhi akurasi kemiripan antarartikel [5]. Namun, pendekatan mereka masih terbatas pada analisis berbasis fitur konten tanpa mempertimbangkan preferensi aktual pengguna, sehingga rekomendasi yang dihasilkan sangat bergantung pada struktur kata dan rentan menghasilkan kesamaan yang bersifat dangkal.

Salsabilla dan Danang juga menerapkan *Content-Based Filtering* untuk sistem rekomendasi wisata dengan menggunakan *TF-IDF* dan cosine similarity [9]. Studi tersebut menunjukkan bahwa *Content-Based Filtering* mampu menangkap relevansi antaritem berdasarkan kedekatan kata kunci. Namun, penelitian tersebut turut menyoroti kelemahan utama *Content-Based Filtering*, yakni

<http://sistemasi.ftik.unisi.ac.id>

kecenderungan menghasilkan rekomendasi yang terlalu mirip (*overspecialization*) sehingga variasi rekomendasi menjadi terbatas. Kondisi ini dapat menjadi permasalahan ketika sistem membutuhkan keberagaman konten, misalnya pada berita sepak bola yang datanya dinamis dan sangat bervariasi.

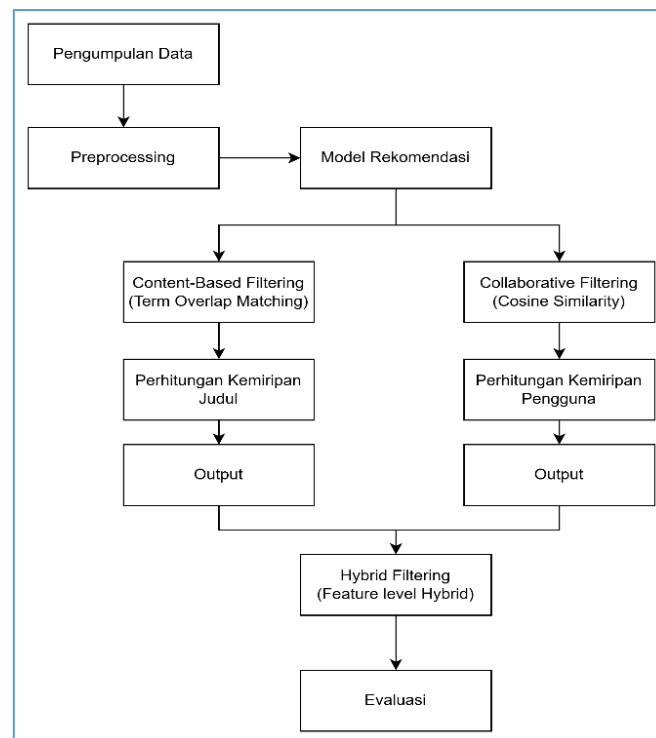
Di sisi lain, penelitian mengenai *Collaborative Filtering* erfokus pada pola kesamaan perilaku antar pengguna. Waskito menunjukkan bahwa *Collaborative Filtering* mampu meningkatkan tingkat personalisasi rekomendasi pada sistem e-commerce melalui analisis kedekatan antar pengguna [10]. Meskipun demikian, Meskipun demikian, beberapa penelitian lain, termasuk studi oleh Siti Aminah mengungkapkan bahwa *Collaborative Filtering* sangat rentan terhadap masalah *data sparsity* dan *cold start* [3]. Ketergantungan yang tinggi terhadap interaksi pengguna membuat performa *Collaborative Filtering* menurun pada dataset kecil, kondisi yang relevan dengan penelitian ini mengingat interaksi pengguna dalam bentuk bookmark masih terbatas.

Upaya untuk mengatasi kekurangan kedua metode tersebut mendorong munculnya berbagai penelitian yang menerapkan metode *Hybrid Filtering*. Kartikasari menemukan bahwa *hybrid filtering* mampu menghasilkan rekomendasi yang lebih stabil karena memadukan kekuatan *Content-Based Filtering* dalam ketepatan konten dan keunggulan *Collaborative Filtering* dalam variasi rekomendasi [11]. Yusrida dan Wardhani juga menunjukkan bahwa metode *hybrid mixed* dapat meningkatkan kualitas rekomendasi pada sistem berbasis web [12]. Namun demikian, sebagian besar pendekatan hybrid pada penelitian sebelumnya masih mengandalkan mekanisme penggabungan bobot atau skor rekomendasi. Mekanisme ini cukup kompleks, memerlukan parameter tambahan, dan kurang optimal ketika data serta interaksi pengguna terbatas. Selain itu, sebagian besar penelitian terdahulu berfokus pada domain e-commerce, perpustakaan digital, dan pariwisata, sementara penerapan *hybrid filtering* pada domain berita, khususnya berita sepak bola yang bersifat sangat dinamis, masih jarang ditemukan.

Berdasarkan analisis kritis tersebut, terlihat adanya celah penelitian yang penting untuk diisi. Pertama, diperlukan pendekatan hybrid yang lebih sederhana dan efisien, terutama untuk sistem dengan dataset kecil dan rendahnya intensitas interaksi pengguna. Kedua, terdapat kebutuhan untuk menerapkan hybrid filtering pada domain berita sepak bola yang memiliki karakteristik unik berupa pembaruan konten yang cepat dan pola ketertarikan pengguna yang sangat spesifik. Untuk menjawab celah tersebut, penelitian ini mengusulkan penerapan *Hybrid Filtering* pada level fitur (*feature-level hybrid*), di mana *Content-Based Filtering* dan *Collaborative Filtering* dijalankan secara terpisah sesuai fungsinya tanpa penggabungan bobot atau skor. Pendekatan ini lebih ringan, tetap efektif untuk dataset kecil, dan mampu memanfaatkan kekuatan masing-masing metode dalam konteks berbeda, sehingga memberikan kontribusi nyata terhadap pengembangan sistem rekomendasi berita olahraga berbasis web.

3 Metode Penelitian

Penelitian ini menggunakan perangkat lunak berbasis web yang dikembangkan dengan bahasa pemrograman PHP dan framework CodeIgniter, serta menggunakan MySQL sebagai sistem manajemen basis data. Sistem dibangun dan diuji menggunakan server lokal (localhost). Selain itu, Visual Studio Code digunakan sebagai *text editor* utama, dan browser dimanfaatkan untuk pengujian antarmuka sistem. Dataset berita disiapkan secara manual oleh admin dan dimasukkan langsung ke dalam database. Metode penelitian ini terdiri atas empat tahapan utama sebagaimana ditunjukkan pada Gambar 1. Setiap tahapan dilakukan secara terstruktur agar sistem rekomendasi berita sepak bola yang dikembangkan dapat berfungsi sesuai dengan tujuan penelitian [13].



Gambar 1 Tahapan metode penelitian

Gambar 1 menunjukkan alur tahapan penelitian ini terdiri dari empat langkah utama, yaitu pengumpulan data, preprocessing, penerapan algoritma rekomendasi *Content-Based Filtering* dan *Collaborative Filtering*, serta evaluasi system [11].

3.1 Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh kumpulan berita sepak bola yang akan digunakan sebagai dataset dalam sistem rekomendasi. Dalam penelitian ini, data dimasukkan secara manual melalui dashboard admin yang dikembangkan dalam sistem, di mana admin menyalin berita dari portal resmi olahraga daring seperti Bola.net dan Detik Sport. Pemilihan metode input manual ini bertujuan untuk memastikan bahwa setiap berita yang dimasukkan telah melalui proses seleksi dan tidak melanggar hak cipta konten. Pendekatan ini juga memudahkan pengendalian kualitas data, terutama dalam menjaga konsistensi format dan struktur teks berita. Menurut Rafael Verdhassrullah, pengumpulan data manual lebih efektif dibandingkan metode *scraping* otomatis karena mampu menjamin keakuratan isi dan mencegah duplikasi data [8]. Daftar berita yang digunakan ditunjukkan pada Tabel 1.

3.2 Preprocessing

Tahapan preprocessing data merupakan proses penting yang berfungsi untuk mengubah data berita menjadi data yang siap diolah sebagai objek penelitian. Langkah-langkah ini mengacu pada pendekatan umum yang digunakan dalam sistem rekomendasi berbasis teks, yang menerapkan case folding, tokenizing, dan stemming untuk meningkatkan kualitas representasi data teks [5]. Menurut Yusrida Muflihah, *preprocessing* yang optimal berpengaruh signifikan terhadap peningkatan akurasi hasil rekomendasi karena membantu mengurangi *noise* dalam data teks [12]. Hasil akhir preprocessing berupa daftar token unik yang telah distandarkan, sehingga dapat digunakan sebagai dasar perhitungan kemiripan pada metode *Content-Based Filtering* maupun untuk kebutuhan analisis lain dalam sistem rekomendasi.

3.3 Sistem Rekomendasi

Tahap ini merupakan inti dari penelitian karena menentukan bagaimana rekomendasi berita dihasilkan. Sistem rekomendasi yang dikembangkan menggunakan dua metode utama, yaitu *Content-Based Filtering* dan *Collaborative Filtering*, yang kemudian digabungkan melalui pendekatan *Hybrid Filtering* pada level fitur. Pemilihan metode didasarkan pada karakteristik dataset yang kecil, struktur judul berita yang pendek, serta interaksi pengguna yang terbatas pada fitur bookmark [7], [13].

3.3.1 Content-Based Filtering

Fitur ini digunakan ketika pengguna membuka atau membaca suatu berita. Sistem menghitung kemiripan antarjudul menggunakan metode Term Overlap Matching, kemudian menampilkan daftar berita lain yang memiliki tingkat kemiripan tertinggi. Pendekatan ini juga memiliki kompleksitas komputasi rendah sehingga cocok diterapkan pada sistem rekomendasi berskala kecil.

$$\text{Similarity}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Kemiripan dihitung menggunakan Persamaan (1), yang mengukur proporsi kata yang muncul pada kedua judul. Semakin besar nilai persamaan tersebut, semakin tinggi tingkat kesamaan konten antarberita. Pendekatan ini efektif untuk menemukan berita yang memiliki kemiripan tinggi, terutama pada judul berita yang relatif pendek [5].

3.3.2 Collaborative Filtering

Fitur ini menghasilkan rekomendasi berdasarkan pola perilaku pengguna melalui data bookmark. Sistem menghitung kemiripan antar pengguna. Metode *Cosine Similarity* dipilih karena mampu mengukur kesamaan preferensi pengguna secara proporsional, meskipun jumlah berita yang dibookmark tiap pengguna berbeda. *Cosine Similarity* efektif pada data biner seperti bookmark dan telah banyak digunakan pada penelitian rekomendasi dengan kondisi data yang *sparse* [3]. Pendekatan ini memungkinkan sistem menemukan pengguna lain dengan pola minat serupa, kemudian menghasilkan rekomendasi yang lebih personal.

$$\text{Similarity}(A, B) = \frac{|I_A \cap I_B|}{\sqrt{|I_A| \times |I_B|}} \quad (2)$$

Nilai kemiripan antar pengguna dihitung menggunakan Persamaan (2), yang membandingkan jumlah berita yang sama-sama dibookmark dengan tingkat aktivitas masing-masing pengguna.[2].

3.3.3 Hybrid Filtering

Kedua metode diintegrasikan melalui pendekatan Hybrid Filtering pada level fitur. Pendekatan ini dipilih karena lebih sesuai untuk dataset kecil dan tidak memerlukan penggabungan skor atau pembobotan seperti pada hybrid berbasis model. Pada fitur baca berita, sistem menampilkan rekomendasi yang mirip secara konten melalui *Content-Based Filtering*, sedangkan pada halaman utama sistem menampilkan rekomendasi berbasis perilaku pengguna melalui *Collaborative Filtering*. Pendekatan ini membuat kedua metode bekerja optimal dalam konteksnya masing-masing tanpa menambah kompleksitas sistem, sekaligus memungkinkan rekomendasi yang lebih relevan dan variatif. Kedua metode dijalankan secara terpisah (*independent execution*) [1], namun hasil rekomendasinya dikombinasikan dalam tampilan sistem. Model hybrid seperti ini memungkinkan sistem memberikan rekomendasi yang tidak hanya serupa dengan berita yang sedang dibaca, tetapi juga sesuai dengan preferensi pengguna secara historis.

3.4 Evaluasi

Evaluasi dilakukan melalui pengujian fungsional serta pengujian performa sistem menggunakan metrik Precision, Recall, dan F1-Score. Pengujian fungsional bertujuan untuk memastikan semua fitur sistem berfungsi sebagaimana mestinya mulai dari fitur registrasi, login, pencarian, *bookmark*, komentar, hingga proses rekomendasi [1]. Sementara itu, pengujian dengan metrik *Precision*, *Recall*, dan *F1-Score* digunakan untuk menilai tingkat akurasi hasil rekomendasi berdasarkan data interaksi pengguna [11].

Ketiga metrik ini digunakan secara luas dalam penelitian sistem rekomendasi untuk mengevaluasi keseimbangan antara ketepatan dan kelengkapan hasil rekomendasi. Perhitungan *precision*, *recall*, dan *f1-score* dapat dihitung menggunakan persamaan (3), (4), (5):

3.4.1 Precision

$$\text{Precision} = \frac{\text{Jumlah berita relevan yang direkomendasikan}}{\text{Total berita yang direkomendasikan}} \quad (3)$$

Nilai *Precision* mengukur seberapa tepat sistem dalam memberikan rekomendasi yang relevan.

3.4.2 Recall

$$\text{Recall} = \frac{\text{Jumlah berita relevan yang direkomendasikan}}{\text{Total berita relevan yang ada}} \quad (4)$$

Recall mengukur kemampuan sistem menemukan berita yang relevan dari seluruh berita yang tersedia.

3.4.3 F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

F1-Score digunakan untuk menilai keseimbangan antara *precision* dan *recall*.

Menurut Rafael Verdhasrullah dan Sukma Kartikasari, metrik ini paling tepat digunakan dalam sistem rekomendasi berbasis teks karena dapat menilai performa sistem secara menyeluruh terhadap variasi minat pengguna [8], [11].

4 Hasil dan Pembahasan

Bagian ini membahas hasil implementasi sistem rekomendasi berita sepak bola berbasis website yang menerapkan algoritma *Content-Based Filtering* dan *Collaborative Filtering*. Pembahasan mencakup hasil setiap tahapan penelitian, analisis kemiripan berita, pola preferensi pengguna, serta evaluasi efektivitas dan efisiensi metode yang digunakan.

4.1 Pengumpulan Data

Tahap pengumpulan data ini dilakukan dengan memasukkan kumpulan berita sepak bola secara manual ke dalam basis data sistem. Pendekatan ini juga digunakan dalam penelitian Sukma Kartikasari karena dapat menjaga validitas konten dan mencegah duplikasi [11]. Sumber berita diperoleh dari portal resmi olahraga daring seperti *Bola.net* dan *Detik Sport* yang menyediakan informasi aktual dan kredibel terkait pertandingan dan klub sepak bola seperti yang ditunjukkan tabel 1.

Tabel 1 Data berita

idBerita	Judul
39	Udinese vs Ac Milan: Brace Pulisic bawa Rossoneri menang telak
47	Atletico Madrid vs Real Madrid: Julian Alvarez hajar el real
55	Psm vs PSIM jogja: Laskar Mataram menjaga unbeaten saat away
56	Persis solo vs persija Jakarta: macan kemayoran bantai tuan rumah
66	Liverpool vs Manchester United: Yang Nangis Warga Anfield
72	Real Madrid vs Barcelona: Los blancos kokoh di puncak
77	Malut vs PSIM: Laskar Mataram berpesta di Ternate

Tabel 1 menampilkan daftar berita yang digunakan sebagai dataset utama dalam penelitian ini, mencakup beberapa pertandingan dari liga domestik maupun internasional. Data yang dikumpulkan menjadi dasar bagi tahapan selanjutnya seperti *preprocessing* dan penerapan algoritma rekomendasi.

4.2 Preprocessing

Tahapan *preprocessing* menghasilkan representasi teks yang lebih terstruktur melalui proses *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil dari setiap tahapan telah ditampilkan pada tabel 2-tabel 5. Representasi kata yang telah dinormalisasi ini menjadi dasar perhitungan kemiripan pada metode *Content-Based Filtering*. Proses *preprocessing* terbukti mengurangi noise dan meningkatkan konsistensi token yang digunakan dalam perhitungan similarity. Menurut Anando, Arif Nur dan Rohman, Cholil, tahapan ini penting dalam sistem rekomendasi berbasis teks karena memengaruhi akurasi perhitungan kemiripan antarberita [5], [6].

4.2.1 Case Folding

Tahap *case folding* dilakukan untuk menyeragamkan teks dengan mengubah semua huruf menjadi huruf kecil. Proses ini bertujuan agar sistem dapat membaca setiap kata tanpa membedakan

huruf kapital atau tidak [2]. Hasil penerapan proses ini dapat dilihat pada Tabel 2, di mana seluruh judul berita telah dikonversi menjadi huruf kecil.

Tabel 2 Hasil case folding

idBerita	Judul
39	udinese vs ac milan: brace pulisic bawa rossoneri menang telak
47	atletico madrid vs real madrid: julian alvarez hajar el real
55	psm vs psim jogja: laskar mataram menjaga unbeaten saat away
56	persis solo vs persija jakarta: macan kemayoran bantai tuan rumah
66	liverpool vs manchester united: yang nangis warga anfield
72	real madrid vs barcelona: los blancos kokoh di puncak
77	malut vs psim jogja: laskar mataram berpesta di ternate

4.2.2 Tokenizing

Proses *tokenizing* memecah teks menjadi kata-kata tunggal (*token*) agar sistem dapat menganalisis kata secara individual. Teknik ini banyak digunakan dalam sistem rekomendasi berbasis teks untuk menemukan kemiripan antar dokumen [12]. Hasil tokenizing pada judul berita dapat dilihat pada Tabel 3, di mana setiap judul telah dipisahkan menjadi token-token yang menyusun kalimat aslinya.

Tabel 3 Hasil tokenizing

idBerita	Judul
39	udinese, vs, ac, milan, brace, pulisic, bawa, rossoneri, menang, telak
47	atletico, madrid, vs, real, madrid, julian, alvarez, hajar, el, real
55	psm, vs, psim, jogja, laskar, mataram, menjaga, unbeaten, saat, away
56	persis, solo, vs, persija, jakarta, macan, kemayoran, bantai, tuan, rumah
66	liverpool, vs, manchester, united, yang, nangis, warga, anfield
72	real, madrid, vs, barcelona, los, blancos, kokoh, di puncak
77	malut, vs, psim, jogja, laskar, mataram, berpesta, di, ternate

4.2.3 Stopword Removal

Tahap *stopword removal* digunakan untuk menghapus kata umum yang tidak memiliki makna penting seperti “di”, “yang”, dan “vs”. Pendekatan serupa diterapkan oleh Lily Apriliyani pada sistem rekomendasi buku untuk meningkatkan kualitas analisis kata kunci dan mengurangi *noise* pada data [7]. Hasil penerapan proses ini dapat dilihat pada Tabel 4, di mana kata-kata yang tidak relevan telah dihilangkan sehingga hanya menyisakan token yang dianggap penting untuk analisis.

Tabel 4 Hasil stopwords removal

idBerita	Judul
39	udinese, ac, milan, brace, pulisic, bawa, rossoneri, menang, telak
47	atletico, madrid, real, madrid, julian, alvarez, hajar, el, real
55	psm, psim, jogja, laskar, mataram, menjaga, unbeaten, away
56	persis, solo, persija, jakarta, macan, kemayoran, bantai
66	liverpool, manchester, united, nangis, warga, anfield
72	real, madrid, barcelona, blancos, kokoh, puncak
77	malut, psim, jogja, laskar, mataram, berpesta, ternate

4.2.4 Stemming

Stemming dilakukan untuk mengembalikan kata ke bentuk dasarnya, misalnya “menjaga” menjadi “jaga” dan “berpesta” menjadi “pesta”. Proses ini membantu sistem mengenali kata yang memiliki akar semantik sama meskipun berbeda bentuk [3]. Hasil *stemming* yang diterapkan pada judul berita ditunjukkan pada tabel 5, yang menjadi keluaran akhir dari seluruh tahap preprocessing. Representasi teks yang telah distem ini kemudian digunakan untuk membangun model perhitungan kemiripan secara lebih efisien dan akurat [12].

Tabel 5 Hasil stemming

idBerita	Judul
39	udinese, ac, milan, brace, pulisic, bawa, rossoneri, menang, telak
47	atletico, madrid, real, madrid, julian, alvarez, hajar, el, real
55	psm, psim, jogja, laskar, mataram, menjaga, unbeaten, away
56	persis, solo, persija, jakarta, macan, kemayoran, bantai
66	liverpool, manchester, united, nangis, warga, anfield
72	real, madrid, barcelona, blancos, kokoh, puncak
77	malut, psim, jogja, laskar, mataram, berpesta, ternate

4.3 Sistem Rekomendasi

Pendekatan Hybrid Filtering pada penelitian ini terinspirasi dari kombinasi dua metode utama, yakni *Content-Based Filtering* dan *Collaborative Filtering* yang telah banyak diterapkan dalam berbagai domain aplikasi. Studi oleh Arif dan Rohmad menunjukkan bahwa penerapan metode *Content-Based Filtering* dengan pembobotan *TF-IDF* dan *Cosine Similarity* mampu meningkatkan relevansi rekomendasi artikel berita [14]. Sementara itu, penelitian oleh Ani dan Susi menggabungkan kedua pendekatan tersebut untuk mengatasi masalah *cold start* dan *data sparsity*, serta membuktikan efektivitas metode hibrida melalui perhitungan *precision*, *recall*, dan *F1-score* [15].

4.3.1 Content-Based Filtering

Metode *Content-Based Filtering* menampilkan berita yang memiliki tingkat kemiripan konten tertinggi dengan berita yang sedang dibaca pengguna. Perhitungan kemiripan menggunakan Term Overlap Matching pada persamaan (1) menghasilkan nilai similarity yang menunjukkan seberapa besar kesamaan kata unik antarjudul.

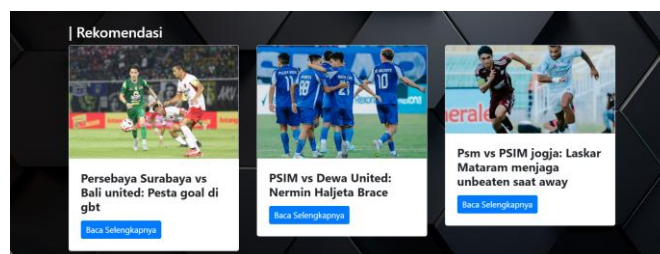
Tabel 6 Hasil perhitungan

Langkah	Hasil
idBerita 55 (A)	psm, psim, jogja, laskar, mataram, jaga, unbeaten, away
idBerita 77 (B)	malut, psim, jogja, laskar, mataram, pesta, ternate
$(A \cap B)$	psim, jogja, laskar, mataram
$(A \cup B)$	psm, psim, jogja, laskar, mataram, jaga, unbeaten, away, pesta, ternate

Berikut perhitungan mengacu pada persamaan (1):

$$Similarity(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{4}{10} = 0,4$$

Contoh perhitungan pada Tabel 6 menunjukkan bahwa judul berita ID 55 dan ID 77 memiliki similarity sebesar 0,4. Nilai ini muncul karena kedua berita membahas tim dan konteks pertandingan yang serupa (PSIM Jogja). Hal ini terlihat pada Gambar 2, di mana sistem merekomendasikan berita lain dengan kata kunci yang sama seperti “psim”, “jogja”, atau “laskar mataram”. Menurut Rafael Verdhasrullah dan Yusrida Muflihah, pendekatan ini efektif untuk mengenali kesamaan tematik antar konten dengan representasi teks yang terbatas, seperti judul berita[8], [12]. Gambar 2 menunjukkan bahwa sistem menghasilkan rekomendasi yang relevan dan adaptif terhadap preferensi pengguna.



Gambar 2 Tampilan halaman baca berita

Gambar 2 menampilkan halaman baca berita yang digunakan pengguna untuk melihat isi berita secara lengkap. Halaman ini memuat judul, gambar, dan isi berita, serta menyediakan tombol bookmark untuk menyimpan berita yang disukai. Fitur bookmark ini menjadi sumber data preferensi pengguna dalam proses rekomendasi.

4.3.2 Collaborative Filtering

Metode *Collaborative Filtering* menghasilkan rekomendasi berdasarkan pola bookmark pengguna. Data interaksi pada Tabel 7 menunjukkan bahwa pengguna dengan ID berbeda memiliki beberapa berita yang sama-sama dibookmark, sehingga menghasilkan nilai similarity antar pengguna menggunakan persamaan (2). Perhitungan dilakukan menggunakan user-item matrix yang berisi interaksi pengguna terhadap berita. Semakin banyak berita yang sama-sama dibookmark oleh dua pengguna, semakin tinggi nilai kemiripannya.

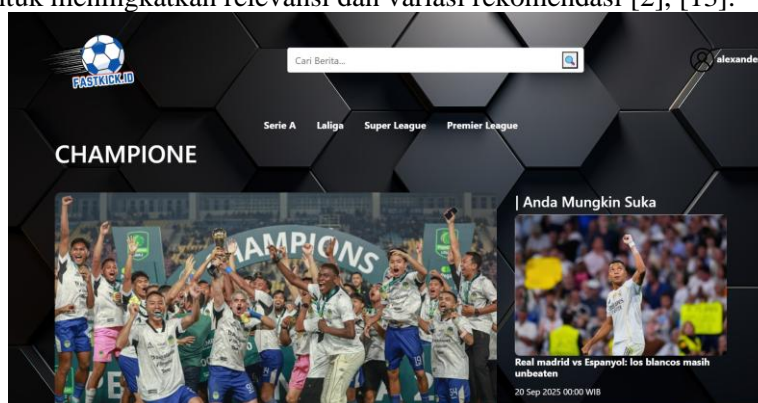
Tabel 7 Interaksi pengguna

idBerita	idUser 9 (A)	idUser 18 (A)
39	Tidak	Tidak
47	Ya	Tidak
55	Ya	Ya
56	Tidak	Ya
66	Ya	Tidak
72	Tidak	Tidak
77	Ya	Ya

Berdasarkan data interaksi, pengguna A memiliki daftar bookmark $I_A = \{47, 55, 66, 77\}$, sedangkan pengguna B memiliki $I_B = \{55, 56, 77\}$. Berita yang dibookmark oleh keduanya adalah $I_A \cap I_B = \{55, 77\}$. Dengan nilai $|I_A| = 4$, $|I_B| = 3$, dan $|I_A \cap I_B| = 2$, perhitungan Cosine Similarity mengacu pada persamaan (2):

$$\text{Similarity}(A, B) = \frac{2}{\sqrt{4 \times 3}} = \frac{2}{\sqrt{12}} = \frac{2}{3.464} = 0.577$$

Nilai similarity sebesar 0.577 menunjukkan bahwa pengguna A dan B memiliki tingkat kemiripan preferensi yang cukup baik. Artinya, ada sebagian berita yang sama-sama mereka minati. Pendekatan ini sejalan dengan konsep *Collaborative Filtering* yang memanfaatkan kesamaan perilaku antar pengguna untuk meningkatkan relevansi dan variasi rekomendasi [2], [13].



Gambar 3 Tampilan halaman utama

Gambar 3 menampilkan bagian rekomendasi yang berada di sisi kanan halaman dengan label “Anda Mungkin Suka”, di mana sistem menampilkan daftar berita yang dibookmark oleh pengguna

lain dengan minat serupa. Hal ini selaras dengan temuan Yudha dan Munir yang memperoleh tingkat kepuasan pengguna tinggi dengan sistem *collaborative filtering* pada domain akomodasi konser [16].

4.3.3 Analisis Hybrid Filtering

Pendekatan Hybrid Filtering pada level fitur menggabungkan metode *Content-Based Filtering* dan *Collaborative Filtering* untuk meningkatkan kualitas rekomendasi dalam kondisi data yang terbatas. Dari sisi konten, kemiripan antarjudul sangat dipengaruhi oleh keberadaan kata-kata kunci yang konsisten dalam konteks klub atau kompetisi sepak bola, sehingga judul yang memiliki istilah serupa cenderung dianggap mirip meskipun pertandingan berbeda. Hal ini menjelaskan mengapa *Content-Based Filtering* mampu memberikan rekomendasi yang relevan ketika pengguna membaca berita dengan kata kunci yang berulang pada berbagai artikel. Sementara itu, dari sisi preferensi pengguna, pola bookmark mencerminkan minat aktual sehingga *Collaborative Filtering* dapat menangkap ketertarikan terhadap klub atau kompetisi tertentu yang tidak selalu terlihat melalui analisis judul. Namun, jumlah interaksi yang rendah menyebabkan matriks pengguna item menjadi sangat sparse, sehingga *Collaborative Filtering* belum dapat bekerja optimal pada dataset kecil. Meskipun demikian, menjalankan kedua metode secara paralel tetap memberikan dua perspektif rekomendasi yang saling melengkapi: *Content-Based Filtering* menghasilkan kesamaan konten yang stabil, sedangkan *Collaborative Filtering* memberikan nilai personalisasi tambahan berdasarkan perilaku pengguna. Dengan demikian, hybrid filtering pada level fitur menjadi strategi yang efisien dan adaptif untuk menghasilkan rekomendasi yang relevan baik secara konten maupun preferensi dalam lingkungan data terbatas.

4.4 Evaluasi

Evaluasi dilakukan untuk mengukur akurasi rekomendasi menggunakan *Precision*, *Recall*, dan *F1-Score*. Pengujian menggunakan data interaksi pengguna idUser 18 pada dua metode: *Content-Based Filtering* dan *Collaborative Filtering*. Data yang digunakan dalam proses evaluasi ditunjukkan pada Tabel 8, yang berisi daftar berita yang dibookmark oleh idUser 18 serta hasil rekomendasi dari metode *Content-Based Filtering* dan *Collaborative Filtering*. Hal ini sesuai dengan penelitian sebelumnya yang menyatakan bahwa kedua metode saling melengkapi: *Content-Based* unggul pada kesamaan konten, sedangkan *Collaborative* unggul pada variasi dan relevansi yang sesuai minat pengguna [11]. Sementara itu, penelitian oleh Ani dan Susi menggabungkan kedua pendekatan tersebut untuk mengatasi masalah *cold start* dan *data sparsity*, serta membuktikan efektivitas metode hibrida melalui perhitungan *precision*, *recall*, dan *F1-score* [15].

Tabel 8 Data pengujian

Jenis Data	idBerita
Berita yang dibookmark oleh idUser 18	55, 66, 77
Content-Based Filtering	55, 66, 77, 80, 81
Collaborative Filtering	44, 52, 53, 56

4.4.1 Perhitungan Content-Based Filtering

Pada metode *Content-Based Filtering*, sistem memberikan lima rekomendasi kepada pengguna idUser 18, yaitu (55, 66, 77, 80, 81). Dari daftar tersebut, tiga berita (55, 66, 77) termasuk dalam berita yang benar-benar relevan.

Mengacu pada Persamaan (3):

$$Precision = \frac{3}{5} = 0.60$$

Mengacu pada Persamaan (4):

$$Recall = \frac{3}{4} = 0.75$$

Mengacu pada Persamaan (5):

$$F1 = 2 \times \frac{0.60 \times 0.75}{0.60 + 0.75} = 0.67 = 2 \times \frac{0.45}{1.35} = 0.67$$

4.4.2 Perhitungan Collaborative Filtering

Pada metode *Collaborative Filtering* sistem merekomendasikan empat berita kepada pengguna idUser 18, yaitu (44, 52, 53, 56). Dari hasil tersebut, tiga berita (55, 56, 77) berita relevan.

Mengacu pada Persamaan (3):

$$Precision = \frac{0}{4} = 0$$

Mengacu pada Persamaan (4):

$$Recall = \frac{0}{3} = 0$$

Mengacu pada Persamaan (5):

$$F1 = 2 \times \frac{0 \times 0}{0+0} = 0$$

Hasil pengujian *Content-Based Filtering* menunjukkan nilai *Precision* sebesar 0,60, *Recall* sebesar 0,75, dan *F1-Score* sebesar 0,67. Nilai tersebut menunjukkan bahwa metode ini cukup efektif dalam mengenali berita dengan topik serupa, misalnya berita tentang klub PSIM Jogja yang memiliki kesamaan istilah pada judul berita. Namun, pendekatan ini memiliki keterbatasan dalam keberagaman rekomendasi karena hanya berfokus pada kemiripan konten. Sementara itu, hasil pengujian *Collaborative Filtering* menunjukkan performa yang lebih rendah dengan *Precision* sebesar 0, *Recall* sebesar 0, dan *F1-Score* sebesar 0. Hasil lengkap pengujian akurasi untuk kedua metode dapat dilihat pada Tabel 9 yang merangkum nilai *Precision*, *Recall*, dan *F1-Score*.

Tabel 9 Hasil evaluasi idUser 18

Metode	Precision	Recall	F1-Score
Content-Based Filtering	0.60	0.75	0.67
Collaborative Filtering	0	0	0

Menurut Sukma Kartikasari, kombinasi kedua pendekatan dalam *Hybrid Filtering* mampu meningkatkan *F1-Score* secara signifikan karena masing-masing metode saling melengkapi: *Content-Based Filtering* unggul dalam ketepatan konten, sedangkan *Collaborative Filtering* unggul dalam keluasan dan variasi rekomendasi [11].

4.4.3 Efektivitas dan Efisiensi Metode

Hasil evaluasi menunjukkan bahwa *Content-Based Filtering* lebih efektif dan efisien dibandingkan *Collaborative Filtering* pada dataset ini, ditandai dengan nilai Precision, Recall, dan F1-score yang lebih tinggi. *Content-Based Filtering* mampu bekerja baik karena hanya bergantung pada kemiripan kata pada judul berita sehingga prosesnya ringan dan tidak membutuhkan data interaksi pengguna. Sebaliknya, *Collaborative Filtering* belum optimal karena interaksi pengguna masih sangat terbatas, sehingga pola kesamaan preferensi sulit terbentuk dan menyebabkan rekomendasi yang kurang relevan. Secara kritis, kondisi ini menunjukkan bahwa efektivitas kedua metode sangat dipengaruhi oleh karakteristik data: *Content-Based Filtering* cenderung stabil tetapi dapat menghasilkan rekomendasi repetitif, sedangkan *Collaborative Filtering* lebih personal tetapi tidak efisien pada dataset kecil. Pendekatan hybrid pada level fitur yang digunakan pada sistem ini menjadi solusi yang tepat untuk mengatasi kelemahan masing-masing metode tanpa menambah kompleksitas perhitungan.

4.4.4 Analisis Kritis Terhadap Hasil

Hasil penelitian ini menunjukkan bahwa *Content-Based Filtering* dan *Collaborative Filtering* memiliki keunggulan serta keterbatasan yang bergantung pada karakteristik data. *Content-Based Filtering* efektif untuk teks pendek dan mampu memberikan rekomendasi stabil, tetapi rentan menghasilkan rekomendasi berulang (*overspecialization*). Sebaliknya, *Collaborative Filtering* berpotensi memberikan rekomendasi yang lebih personal, namun performanya sangat dipengaruhi oleh kelengkapan data interaksi pengguna, sehingga kurang optimal pada kondisi *sparse data*. Pendekatan Hybrid Filtering pada level fitur terbukti relevan untuk dataset kecil karena memungkinkan kedua metode berfungsi dalam konteks terbaiknya tanpa menambah kompleksitas, meskipun kontribusi utamanya masih didominasi oleh *Content-Based Filtering* akibat keterbatasan *Collaborative Filtering*. Secara kritis, hybrid berbasis fitur efektif pada lingkungan data terbatas, namun kualitas rekomendasi dapat meningkat signifikan jika interaksi pengguna bertambah sehingga *Collaborative Filtering* dapat bekerja optimal.

5 Kesimpulan

Penelitian ini berhasil mengimplementasikan *Content-Based Filtering* dan *Collaborative Filtering* dalam pendekatan *Hybrid Filtering* berbasis fitur untuk sistem rekomendasi berita sepak bola, di mana *Content-Based Filtering* menunjukkan performa lebih baik pada dataset kecil sementara *Collaborative Filtering* belum optimal akibat keterbatasan interaksi pengguna. Kontribusi utama penelitian ini terletak pada penerapan hybrid sederhana yang efisien tanpa penggabungan skor, pemanfaatan *Term Overlap Matching* yang sesuai untuk judul berita pendek, serta penggunaan data bookmark sebagai representasi preferensi aktual pengguna. Implikasi penelitian ini menunjukkan bahwa pendekatan hybrid berbasis fitur efektif diterapkan pada platform berita olahraga dengan sumber daya terbatas karena mampu memberikan rekomendasi relevan baik dari sisi konten maupun preferensi pengguna. Namun, penelitian ini masih dibatasi oleh ukuran dataset yang kecil, variasi berita yang terbatas, dan kondisi *data sparsity* pada interaksi pengguna. Untuk pengembangan selanjutnya, diperlukan penambahan jumlah data dan interaksi pengguna, penggunaan teknik NLP yang lebih kaya seperti TF-IDF, word embedding, atau model transformer, serta eksplorasi hybrid pada level skor atau model ketika dataset sudah memadai.

Referensi

- [1] K. S. Yanisa Putri, I Made Agus Dwi Suarjaya, dan Wayan Oger Vihikan, "Sistem Rekomendasi *Skincare* menggunakan Metode *Content Based Filtering* dan *Collaborative Filtering*," *Decode: Jurnal Pendidikan Teknologi Informasi*, Vol. 4, No. 3, hlm. 764–774, 2024, DOI: 10.51454/decode.v4i3.601.
- [2] R. Martin Permana, A. Id Hadiana, dan P. Nurul Sabrina, "Rekomendasi Pemilihan Sepeda Motor menggunakan Metode *Content Based Filtering* dan *Item Based Colaborative Filtering*," *Jurnal Teknik Informatika Unis*, Vol. 12, No. 2, hlm. 207–217, 2024, DOI: 10.33592/jutis.v12i2.5149.
- [3] S. Aminah, "Perbandingan Keakuratan Sistem Rekomendasi Produk berbasis *Content-Based Filtering* dan *Collaborative Filtering* pada *e-commerce Shopee* menggunakan Matrik *Precision*, *Recall* dan *F1-Score*," 2024, pp. 1–18.
- [4] M. Alfin, T. Putra, dan R. Alit, "Sistem Rekomendasi Destinasi Wisata di Kabupaten Mojokerto dengan Metode *Hybrid Filtering* berbasis *Website*," *Journal of Informatics and Computer Science*, Vol. 07, hlm. 57–64, 2025.
- [5] A. M. R. Haz dan A. N. Rohman, "Sistem Rekomendasi Berita dengan Metode *Content-Based Filtering*," *JIKA (Jurnal Informatika)*, Vol. 9, No. 2, hlm. 143, 2025, DOI: 10.31000/jika.v9i2.13247.
- [6] A. S. Muarif dan E. Winarno, "Sistem Rekomendasi Tempat Parkir di Kota Lama Semarang menggunakan *Collaborative Filtering*," *Jurnal Ilmiah Universitas Batanghari Jambi*, Vol. 22, No. 2, hlm. 906, 2022, DOI: 10.33087/jiubj.v22i2.2066.
- [7] L. Aprilyani, N. Ransi, R. A. Saputra, dan J. T. Informatika, "Sistem Rekomendasi dan Peminjaman Buku menggunakan *Algoritma Hybrid Based Filtering Book Recommendation and Lending System using Hybrid-based Filtering Algorithm*," *Jurnal Sisfotenika*, Vol. 14, No. 2, hlm. 140–151, 2024.
- [8] A. M. H. S. Muh. Alvahrefi Bakhtiar, Josef S.B tuda, "Jurnal Inovasi Global," *Jurnal Inovasi Global*, Vol. 2, No. 3, hlm. 543–551, 2024.
- [9] C. Salsabilla dan D. W. Utomo, "Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan *Content Based Filtering*," *Sistemasi: Jurnal Sistem Informasi*, Vol. 14, No. 1, hlm. 262–270, 2025.
- [10] M. R. Waskito, A. D. Rahajoe, dan A. L. Nurlaili, "Implementasi Metode *Collaborative Filtering* menggunakan *Algoritma Cosine Similarity* dan *Jaccard Similarity* pada Sistem *E-Commerce*," *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 12, No. 3S1, hlm. 4307–4316, 2024, DOI: 10.23960/jitet.v12i3s1.5315.
- [11] U. A. Surabaya, U. A. Surabaya, dan U. A. Surabaya, "Pengembangan Sistem Rekomendasi Buku menggunakan Metode *Hybrid Filtering* dalam Layanan Perpustakaan Modern Sukma Kartikasari Shiefti Dyah Alyusi Siti Muzaroh A . PENDAHULUAN Perpustakaan modern dihadapkan pada tantangan besar dalam mengelola dan menyajika," Vol. 7, No. 1, 2025.

- [12] Y. Muflihah dan M. W. Wardhani, “Implementasi *Hybrid Filtering* dalam Sistem Rekomendasi berbasis *Web* untuk Persewaan Tenda Pesta,” *Jurnal Rekayasa Sistem Informasi dan Teknologi*, Vol. 1, No. 4, hlm. 361–372, 2024, DOI: 10.59407/jrsit.v1i4.814.
- [13] K. R. Putra dan M. A. Rachman, “Perbandingan Metode *Content-based*, *Collaborative* dan *Hybrid Filtering* pada Sistem Rekomendasi Lagu,” *MIND Journal*, Vol. 9, No. 2, hlm. 179–193, 2024, DOI: 10.26760/mindjournal.v9i2.179-193.
- [14] A. A. Huda, R. Fajarudin, dan A. Hadinegoro, “Sistem Rekomendasi *Content-based Filtering* menggunakan *TF-IDF Vector Similarity* untuk Rekomendasi Artikel Berita,” *Building of Informatics, Technology and Science (BITS)*, Vol. 4, No. 3, hlm. 1679–1686, 2022, DOI: 10.47065/bits.v4i3.2511.
- [15] A. Rachmaniar, S. Widayati, K. Rokoyah, *et al.*, “Sistem Rekomendasi Produk *E-Commerce* menggunakan *Collaborative Filtering* dan *Content-Based Filtering*,” *Journal of Information System, Informatics and Computing*, Vol. 9, No. 1, pp. 40–54, 2025, DOI: 10.52362/jisicom.v9i1.1904.
- [16] A. A. P. Yudha, Munir, dan Ani Anisyyah, “Perancangan Sistem Rekomendasi Akomodasi pada *Event* Konser dengan Metode *Hybrid Filtering*,” *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, Vol. 4, No. 2, hlm. 631–641, 2025, DOI: 10.62712/juktisi.v4i2.493.
- [17] Hanafi dkk., “*Improvement of E-Commerce Recommender System using Hybridization of Bert, Matrix Factorization and Attention Mechanism*,” *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 5, hlm. 725–740, 2024, DOI: 10.22266/ijies2024.1031.55.