

**SISTEM INFORMASI REKOMENDASI BERITA MENGGUNAKAN COSINE
SIMILARITY**

NASKAH PUBLIKASI



diajukan oleh
Tanujaya Chandra Setyawan
22.12.2594

UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

NASKAH PUBLIKASI

**SISTEM INFORMASI REKOMENDASI BERITA MENGGUNAKAN COSINE
SIMILARITY**

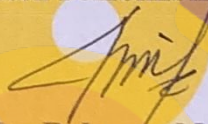
yang dipersiapkan dan disusun oleh

Tanujaya Chandra Setyawan

22.12.2594

Tanggal, 12 Maret 2026

Dosen Pembimbing



Arif Nur Rohman, M.Kom

NIK. 190302684

SISTEM INFORMASI REKOMENDASI BERITA MENGGUNAKAN COSINE SIMILARITY

Tanujaya Chandra Setyawan ¹⁾, Arif Nur Rohman ²⁾

¹⁾ Sistem Informasi Universitas Amikom Yogyakarta

²⁾ Informatika, Universitas Amikom Yogyakarta

email : tanujaya@students.amikom.ac.id¹⁾, arifrahman@amikom.ac.id²⁾

Abstraksi

Perkembangan portal berita digital menyebabkan peningkatan jumlah artikel yang tersedia, sehingga pengguna sering mengalami kesulitan dalam menemukan berita yang relevan dengan minatnya. Penelitian ini bertujuan untuk mengembangkan sistem informasi rekomendasi berita berbasis content-based filtering menggunakan metode TF-IDF dan cosine similarity. Sistem dirancang dan diimplementasikan dalam portal berita berbasis web yang terintegrasi secara end-to-end. Proses rekomendasi dilakukan melalui tahap preprocessing teks, pembobotan kata menggunakan TF-IDF, serta perhitungan tingkat kemiripan dokumen menggunakan cosine similarity. Evaluasi sistem dilakukan menggunakan metrik precision untuk mengukur relevansi rekomendasi yang dihasilkan. Hasil penelitian menunjukkan bahwa sistem mampu menghasilkan rekomendasi berita yang relevan berdasarkan kemiripan konten judul dan deskripsi artikel. Nilai precision yang diperoleh menunjukkan bahwa pendekatan yang digunakan efektif dalam meningkatkan kualitas rekomendasi pada portal berita berbasis web.

Kata Kunci :

Sistem Rekomendasi, Content-Based Filtering, TF-IDF, Cosine Similarity, Portal Berita

Abstract

The rapid growth of digital news portals has significantly increased the number of available articles, making it difficult for users to find news relevant to their interests. This study aims to develop a news recommendation information system based on content-based filtering using TF-IDF and cosine similarity methods. The system is designed and implemented as an end-to-end integrated web-based news portal. The recommendation process includes text preprocessing, term weighting using TF-IDF, and similarity calculation using cosine similarity. System evaluation is conducted using precision to measure the relevance of generated recommendations. The results indicate that the system is capable of producing relevant news recommendations based on the similarity of article titles and descriptions. The obtained precision value demonstrates that the proposed approach effectively improves recommendation quality in a web-based news portal.

Keywords :

Recommendation System, Content-Based Filtering, TF-IDF, Cosine Similarity, News Portal

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat mengakses informasi, khususnya melalui media digital. Portal berita berbasis web menjadi sumber utama informasi karena mampu menyajikan berita secara cepat, real-time, dan mudah diakses kapan saja melalui berbagai perangkat. Namun, meningkatnya jumlah artikel yang dipublikasikan setiap hari memicu kondisi information overload, yaitu kesulitan pengguna dalam menemukan informasi yang benar-benar relevan dengan kebutuhan mereka. Kondisi ini menuntut adanya mekanisme penyaringan informasi yang lebih efektif dan terarah.

Selain itu, banyak portal berita masih bergantung pada Application Programming Interface (API) pihak ketiga sebagai sumber data utama. Ketergantungan ini menimbulkan berbagai risiko, seperti keterbatasan

kontrol terhadap data, perubahan struktur API secara sepihak, hingga potensi gangguan layanan yang dapat memengaruhi keberlanjutan sistem [1]. Di sisi lain, sebagian besar portal berita masih menyajikan artikel berdasarkan urutan waktu atau kategori umum tanpa mempertimbangkan kesesuaian konten dengan minat pengguna, sehingga pengguna harus melakukan pencarian dan penyaringan berita secara manual.

Sejumlah penelitian terdahulu telah mengkaji permasalahan ini dari berbagai sudut pandang. Priskila, Sari, dan Arief mengembangkan portal berita berbasis website, namun belum mengintegrasikan sistem rekomendasi untuk personalisasi konten [2]. Sulistyoko dkk. membahas pemanfaatan API pada portal berita dan menunjukkan kemudahan integrasi data, meski sistem masih bergantung pada API pihak ketiga [1]. Dalam bidang rekomendasi berbasis konten, Permana dan Wibowo menerapkan TF-IDF dan cosine similarity pada

sistem rekomendasi film berdasarkan sinopsis dan membuktikan efektivitasnya, meskipun memiliki keterbatasan performa pada dataset berukuran besar [3]. Nandita, Suarjaya, dan Bayupati mengembangkan sistem rekomendasi berita menggunakan content-based filtering dengan TF-IDF dan cosine similarity pada layanan RSS, dengan hasil rekomendasi yang relevan dan diterima baik oleh pengguna [4]. Haz dan Rohman mengembangkan sistem serupa khusus untuk artikel berita dan menekankan pentingnya tahap preprocessing data dalam meningkatkan kualitas rekomendasi [5]. Sebagai pembanding, Hartatik, Nurhayati, dan Widayani menerapkan collaborative filtering pada sistem rekomendasi wisata kuliner; pendekatan ini terbukti akurat namun memerlukan data interaksi pengguna yang besar sehingga kurang ideal untuk sistem tahap awal [6]. Huda, Fajarudin, dan Hadinegoro mengembangkan sistem rekomendasi artikel berita menggunakan TF-IDF Vector Similarity dan membuktikan kemampuannya dalam memberikan rekomendasi berdasarkan kemiripan konten teks [7].

Berdasarkan kajian tersebut, sebagian besar penelitian terdahulu hanya berfokus pada satu aspek saja, baik pengembangan portal berita tanpa rekomendasi, maupun sistem rekomendasi tanpa integrasi portal secara penuh. Belum ada penelitian yang mengembangkan portal berita sekaligus sistem rekomendasinya secara end-to-end sebagai satu kesatuan sistem yang terintegrasi.

Penelitian ini mengembangkan sistem portal berita berbasis web secara end-to-end dengan arsitektur three-tier, meliputi backend menggunakan Express.js dengan Node.js dan TypeScript, basis data PostgreSQL melalui Supabase, serta frontend menggunakan React.js dengan pendekatan Single Page Application (SPA). Sistem dilengkapi dengan mekanisme rekomendasi berita berbasis konten menggunakan metode TF-IDF dan cosine similarity untuk mengukur kemiripan antar artikel secara otomatis.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan Software Development Life Cycle (SDLC) dengan model waterfall yang terdiri dari lima tahap berurutan: requirements analysis, system design, implementation, testing, dan deployment and maintenance. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

1 Requirements Analysis

Tahap ini diarahkan pada identifikasi permasalahan portal berita digital, terutama kondisi information overload, rendahnya personalisasi konten, dan ketergantungan pada API pihak ketiga. Kegiatan yang dilakukan meliputi studi literatur dan analisis kebutuhan untuk merumuskan fungsi utama serta spesifikasi sistem rekomendasi yang akan dibangun.

2 System Design

Tahap ini mencakup perancangan arsitektur sistem three-tier, desain basis data relasional (ERD dan class diagram), rancangan antarmuka pengguna (wireframe), serta perancangan mekanisme rekomendasi berbasis konten. Pada tahap ini juga disusun flowchart proses representasi dokumen dalam bentuk vektor TF-IDF dan perhitungan kemiripan menggunakan cosine similarity sebagai dasar sistem rekomendasi.

3 Implementation

Tahap ini merupakan realisasi rancangan menjadi sistem yang berjalan secara penuh. Backend dikembangkan dengan Express.js menyediakan REST API untuk operasi CRUD data berita dan autentikasi JWT. Frontend dibangun dengan React.js dan Tailwind CSS sebagai antarmuka pengguna yang responsif. Mekanisme rekomendasi diimplementasikan pada sisi klien (client-side) mencakup preprocessing teks, pembobotan TF-IDF, pembentukan vektor dokumen, serta perhitungan nilai kemiripan antar artikel untuk menghasilkan daftar rekomendasi secara otomatis.

4 Testing

Tahap pengujian memverifikasi kesesuaian sistem dengan kebutuhan yang ditetapkan melalui tiga jenis pengujian: (1) functional testing menggunakan metode black-box untuk memverifikasi seluruh fitur utama; (2) performance testing untuk mengukur response time API dan waktu komputasi algoritma rekomendasi; serta (3) usability testing melibatkan 5 responden untuk mengevaluasi kemudahan penggunaan sistem dari perspektif pengguna akhir.

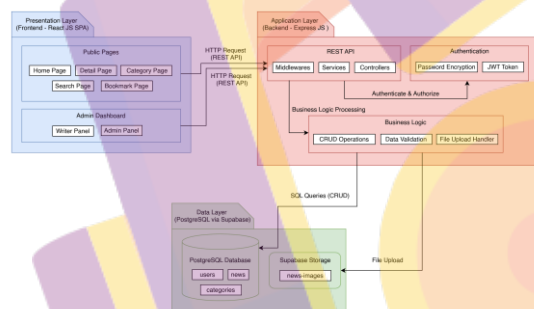
5Deployment and Maintenance

Tahap terakhir adalah penerapan sistem pada lingkungan operasional agar dapat diakses oleh pengguna secara nyata, disertai dokumentasi teknis serta pemeliharaan untuk menjaga stabilitas sistem dan mendukung pengembangan lanjutan.

3. Hasil dan Pembahasan

3.1 Hasil Perancangan Sistem

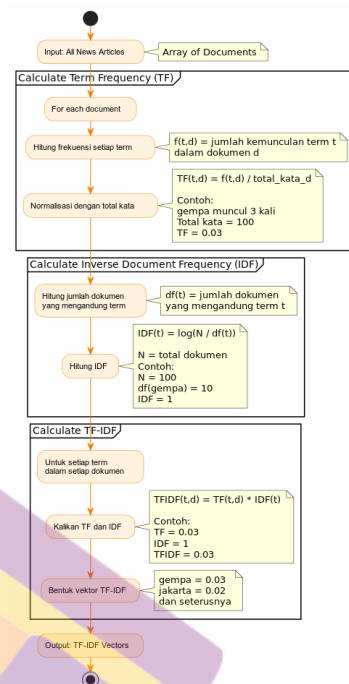
Perancangan sistem menghasilkan arsitektur three-tier yang menggambarkan alur komunikasi antar lapisan sistem portal berita. Diagram arsitektur sistem ditunjukkan pada Gambar 2.



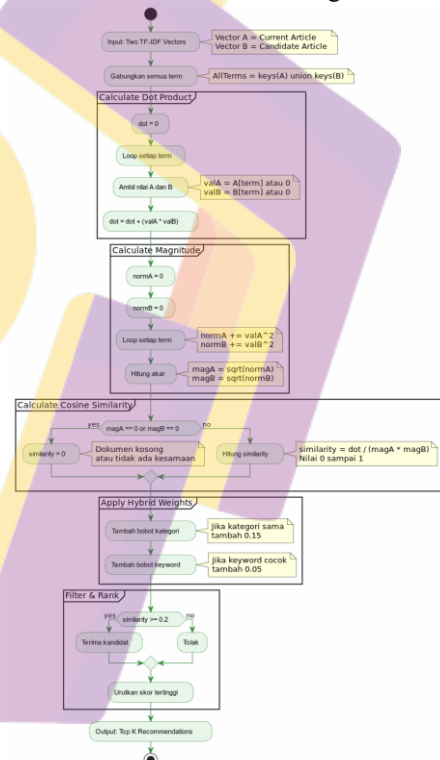
Gambar 2. Diagram Arsitektur Sistem

Arsitektur sistem terdiri dari tiga lapisan utama. Frontend berbasis React.js berjalan di browser pengguna dan berkomunikasi dengan backend melalui HTTP request. Backend berbasis Express.js mengelola logika bisnis dan menyediakan REST API yang terproteksi. Basis data PostgreSQL melalui Supabase menyimpan seluruh data artikel, kategori, dan pengguna secara terstruktur. Mekanisme autentikasi menggunakan JSON Web Token (JWT) diterapkan untuk mengamankan seluruh akses ke endpoint yang memerlukan otorisasi.

Perancangan algoritma rekomendasi divisualisasikan dalam dua flowchart: perhitungan TF-IDF (Gambar 3) dan perhitungan cosine similarity beserta ranking rekomendasi (Gambar 4).



Gambar 3. Flowchart Perhitungan TF-IDF



Gambar 4. Flowchart Cosine Similarity & Ranking. Proses dimulai dengan preprocessing teks artikel (case folding, tokenisasi, stopwords removal, dan stemming). Setiap term kemudian dibobotkan menggunakan TF-IDF. Kemiripan antar artikel dihitung menggunakan cosine similarity berdasarkan vektor TF-IDF, dan hasilnya dikombinasikan dengan category boosting (+0,15) dan keyword matching (+0,05) untuk menghasilkan ranking rekomendasi akhir.

3.2 Implementasi Sistem

Sistem portal berita berhasil dikembangkan secara end-to-end dengan mengintegrasikan komponen

backend, basis data, dan frontend dalam satu kesatuan sistem yang terkoordinasi. Berikut adalah tampilan antarmuka sistem yang berhasil diimplementasikan.



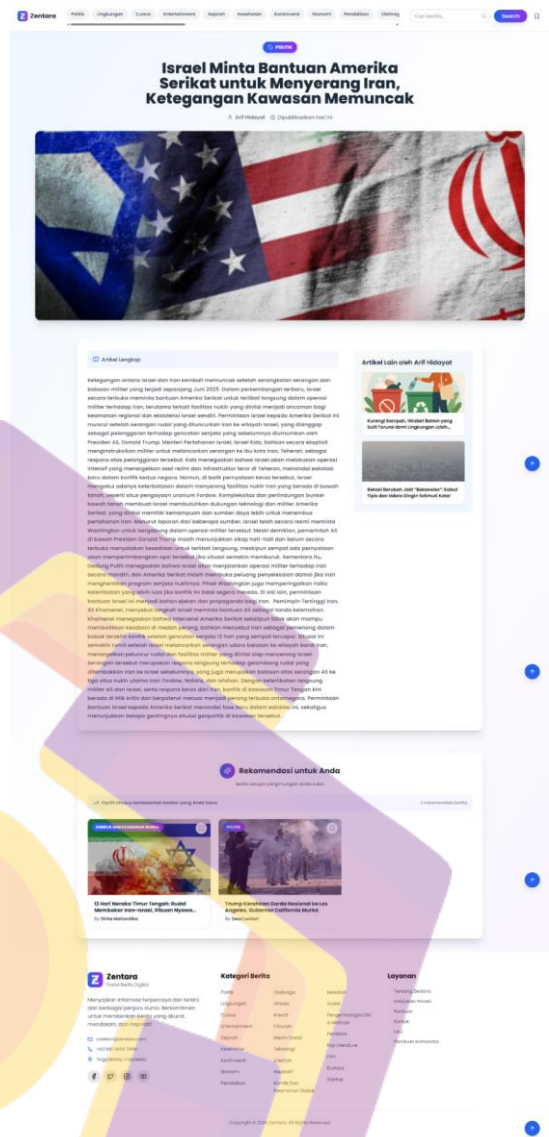
Gambar 5. Tampilan Halaman Beranda (Hero Slider Berita Unggulan)

Halaman beranda menampilkan hero slider berita unggulan, bagian berita terbaru, dan navigasi kategori. Antarmuka dirancang responsif menggunakan Tailwind CSS sehingga dapat diakses dengan baik pada berbagai ukuran layar. Sistem manajemen konten dilengkapi dengan Role-Based Access Control (RBAC): Admin memiliki akses penuh terhadap seluruh artikel, sementara Penulis hanya dapat mengelola artikel miliknya sendiri.



Gambar 6. Tampilan Halaman Pencarian Berita (Grid View Hasil Pencarian)

Halaman pencarian memungkinkan pengguna menyaring artikel berdasarkan kata kunci, kategori, dan rentang tanggal. Hasil pencarian ditampilkan dalam grid view yang responsif. Kata kunci yang dimasukkan pengguna juga turut memengaruhi skor rekomendasi melalui mekanisme keyword matching pada sistem rekomendasi.



Gambar 7. Tampilan Halaman Detail Berita (Desktop)



Gambar 8. Tampilan Section Rekomendasi pada Halaman Detail Berita

Halaman detail berita menampilkan konten artikel secara lengkap beserta metadata (kategori, penulis, dan tanggal publikasi). Di bawah konten artikel, sistem secara otomatis menampilkan tiga artikel rekomendasi yang paling mirip berdasarkan hasil perhitungan TF-IDF dan cosine similarity. Setiap kartu rekomendasi menampilkan judul, kategori, dan thumbnail artikel, sehingga memudahkan pengguna untuk menavigasi ke konten yang relevan.

3.3 Pengujian Sistem

3.3.1 Pengujian Fungsional

Pengujian fungsional dilakukan menggunakan metode black-box testing dengan memfokuskan pengujian pada input dan output sistem tanpa memperhatikan implementasi internal. Hasil pengujian terhadap 8 fitur utama disajikan pada Tabel 1.

Tabel 1. Hasil Functional Testing

No	Fitur yang Diuji	Test Case	Expected Result	Actual Result	Status
1	Login	Kredensial valid dimasukkan	Redirect ke dashboard, token JWT tersimpan	Sesuai	PASS
2	Login	Kredensial tidak valid	Pesan error ditampilkan	Sesuai	PASS
3	Create Article	Admin/Penulis tambah artikel	Artikel tersimpan dan muncul di daftar	Sesuai	PASS
4	Update Article	Penulis ubah artikel miliknya	Data artikel berhasil diperbarui	Sesuai	PASS
5	Delete Article	Admin hapus artikel	Artikel terhapus dari sistem	Sesuai	PASS
6	Bookmark	User klik ikon bookmark	Artikel tersimpan di localStorage	Sesuai	PASS
7	Search	User lakukan pencarian artikel	Artikel relevan ditampilkan	Sesuai	PASS
8	Rekomendasi	User buka halaman detail artikel	3 artikel rekomendasi ditampilkan	Sesuai	PASS

Seluruh 8 fitur utama sistem dinyatakan PASS. Tidak ditemukan critical bug yang mengganggu proses penggunaan sistem. Sistem mampu menangani berbagai skenario input dan menghasilkan output yang sesuai dengan ekspektasi pada setiap kasus uji.

3.3.2 Pengujian Performa

Pengujian performa dilakukan untuk mengukur response time setiap endpoint API dan halaman sistem pada kondisi normal tanpa beban ekstrem. Hasil pengujian disajikan pada Tabel 2.

Tabel 2. Hasil Response Time Testing

Endpoint / Halaman	Rata-rata Response Time	Target	Status
GET /api/publikasi-berita	280 ms	< 500 ms	PASS
POST /api/auth/login	450 ms	< 1000 ms	PASS
POST /api/upload	1,2 s	< 3 s	PASS
POST /api/news	520 ms	< 1000 ms	PASS

Endpoint / Halaman	Rata-rata Response Time	Target	Status
Home Page	1,8 s	< 3 s	PASS
Detail Page (dengan rekomendasi)	2,1 s	< 3 s	PASS
Search Page	1,5 s	< 3 s	PASS
Bookmark Page	1,2 s	< 3 s	PASS

Seluruh endpoint dan halaman memenuhi target response time di bawah 3 detik. Waktu respons tercepat tercatat pada endpoint GET data berita (280 ms), sedangkan waktu terlama pada halaman detail artikel dengan sistem rekomendasi (2,1 detik). Waktu lebih lama ini disebabkan oleh proses komputasi TF-IDF dan cosine similarity yang dilakukan secara client-side. Pada dataset 30 artikel komputasi selesai dalam 35 ms, 50 artikel dalam 65 ms, dan 70 artikel dalam 95 ms.

3.3.3 Pengujian Usability

Pengujian usability melibatkan 5 responden (2 Admin, 2 Penulis, 1 User umum) yang mengevaluasi sistem berdasarkan empat aspek menggunakan skala 1–5. Hasil pengujian disajikan pada Tabel 3.

Tabel 3. Hasil Usability Testing (Skala 1–5)

Aspek	R1	R2	R3	R4	R5	Rata-rata	Kategori
Kemudahan Navigasi	5	4	5	5	4	4,6	Sangat Baik
Clarity	4	5	4	5	4	4,4	Sangat Baik
Responsiveness	5	5	5	4	5	4,8	Sangat Baik
Satisfaction	4	5	4	5	4	4,5	Sangat Baik
Rata-rata Keseluruhan						4,58	Sangat Baik

Sistem memperoleh skor usability rata-rata 4,58 dari 5 dengan kategori Sangat Baik. Aspek responsiveness memperoleh skor tertinggi (4,8) yang menunjukkan bahwa desain responsif yang diterapkan memberikan pengalaman pengguna yang optimal pada berbagai perangkat. Responden juga memberikan feedback positif terkait antarmuka yang bersih, kemudahan fitur bookmark, dan relevansi rekomendasi artikel yang disajikan.

4. Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan pengujian yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1 Sistem portal berita berbasis web berhasil dikembangkan secara end-to-end dengan mengintegrasikan komponen backend (Express.js + Node.js), basis data (PostgreSQL via Supabase), dan frontend (React.js SPA) dalam satu kesatuan sistem tanpa ketergantungan pada API pihak ketiga.

2 Metode TF-IDF dan cosine similarity berhasil diterapkan sebagai mekanisme sistem rekomendasi berita berbasis konten. Sistem mampu

mengidentifikasi kemiripan antar artikel berdasarkan judul dan deskripsi berita, serta menghasilkan rekomendasi secara otomatis tanpa memerlukan data historis interaksi pengguna dalam jumlah besar.

3Pengujian fungsional menunjukkan seluruh 8 fitur utama sistem berjalan dengan hasil PASS, membuktikan bahwa sistem berfungsi sesuai dengan spesifikasi kebutuhan yang telah ditetapkan.

4Pengujian performa memperlihatkan bahwa seluruh endpoint API dan halaman sistem memiliki response time di bawah 3 detik, dengan komputasi algoritma rekomendasi pada client-side yang mampu selesai di bawah 100 ms untuk dataset hingga 70 artikel.

5Pengujian usability menghasilkan skor rata-rata 4,58 dari 5 (kategori Sangat Baik), yang menunjukkan bahwa sistem memiliki antarmuka yang intuitif, informatif, responsif, dan memuaskan bagi pengguna dari berbagai peran.

Daftar Pustaka

- [1] H. W. Sulisty, H. Oktavianto, Z. Arifin, A. E. Wardoyo, and N. Q. Fitriyah, "PEMANFAATAN API PADA APLIKASI PORTAL BERITA SEDERHANA," *JUTIKOMP*, 2024.
- [2] R. Priskila, N. N. K. Sari, and M. R. Arief, "Aplikasi Portal Berita Berbasis Website (Studi Kasus: Lintasberita1.Com)," Sep. 2022.
- [3] A. H. J. Permana and A. T. Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *International Journal on Information and Communication Technology (IJoICT)*, vol. 9, no. 2, pp. 1–14, Dec. 2023, doi: 10.21108/ijoint.v9i2.747.
- [4] I. A. W. Nandita, I. M. A. D. Suarjaya, and I. P. A. Bayupati, "News Recommendation System Using Content-Based Filtering through RSS Customization Service," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] A. M. R. Haz and A. N. Rohman, "JIKA | 143 SISTEM REKOMENDASI BERITA DENGAN METODE CONTENT-BASED FILTERING," Apr. 2025. [Online]. Available: <https://www.kaggle.com/datasets/rizkia14>
- [6] Hartatik, S. D. Nurhayati, and W. Widayani, "Sistem Rekomendasi Wisata Kuliner di Yogyakarta dengan Metode Item-Based Collaborative Filtering," 2021. [Online]. Available: <https://manganenakyog.my.id/>,
- [7] A. A. Huda, R. Fajarudin, and 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, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.