IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Content-Based Filtering untuk Sistem Rekomendasi Produk E-Commerce

Fiki Pratama[1, a)], Sabib Prastio [2, b)], Erni Seniwati [3, c)], Ninik Tri Hartanti [4, d)]
Sudarmanto [5, e)]

[1,2,3,4)] *Program Studi Sistem Informasi,*
[5)]*Program Studi Rekayasa Perangkat Lunak Aplikasi*
Fakultas Ilmu Komputer
Universitas Amikom Yogyakarta, Jl. Padjajaran, Yogyakarta 55283, Indonesia
Universitas Teknologi Digital Indonesia, Jl. Raya Janti Karang Jambe, Yogyakarta 55198, Indonesia

Author Emails
[c)] Corresponding author: erni.s@amikom.ac.id
[a)]fikipratama@students.amikom.ac.id
[b)]sabibprastio@students.amikom.ac.id
[d)]ninik.t@amikom.ac.id
[e)]darmanto@utdi.ac.id

Abstract. *The rapid growth of e-commerce platforms has increased the demand for recommendation systems that can effectively assist users in discovering relevant products. Although recommendation systems have been extensively studied, empirically validated implementations using Indonesian e-commerce product datasets remain limited. This study aims to develop a content-based filtering e-commerce product recommendation system by leveraging content similarity among products. The dataset used in this research is the Tokopedia Products 2025 dataset obtained from Kaggle, which contains product information such as product names, categories, and descriptions. Text feature representation is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method to extract significant characteristics from each product. Subsequently, product similarity is calculated using Cosine Similarity to generate the most relevant product recommendations based on user preferences. The system is implemented using the Python programming language with the support of the pandas, NumPy, and scikit-learn libraries. Evaluation results using various query sample sizes and K values show that, for a representative sample of 50 query products, the system achieves a Precision@10 of 88.40%, Recall@10 of 46.81%, and F1-Score@10 of 50.88%, with the optimal F1-Score of 66.79% achieved at K = 30. These results indicate that the combination of TF-IDF and Cosine Similarity is capable of producing relevant product recommendations with good accuracy, demonstrating its feasibility as a content-based recommendation solution for e-commerce platforms.*

Keywords:
Confusion Matrix, Content-Based Filtering, Cosine Similarity, E-Commerce, Recommendation System, TF-IDF

Abstraksi. Perkembangan pesat platform e-commerce mendorong kebutuhan akan sistem rekomendasi yang mampu membantu pengguna menemukan produk yang relevan secara efektif. Meskipun sistem rekomendasi telah banyak dikembangkan, implementasi yang teruji secara empiris pada dataset produk e-commerce Indonesia masih terbatas. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi produk e-commerce berbasis *content-based filtering* dengan memanfaatkan kemiripan konten antar produk. Dataset yang digunakan bersumber dari Tokopedia Products 2025 yang diperoleh melalui Kaggle, berisi informasi produk seperti nama, kategori, dan deskripsi. Representasi fitur teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengekstraksi karakteristik penting dari setiap produk. Selanjutnya, tingkat kemiripan antar produk dihitung menggunakan *Cosine Similarity* guna menghasilkan rekomendasi produk yang paling relevan berdasarkan preferensi pengguna. Implementasi

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

dilakukan menggunakan bahasa pemrograman Python dengan pustaka *pandas*, *NumPy*, dan *scikit-learn.* Hasil evaluasi menggunakan berbagai ukuran sampel query dan nilai K yang terdiri dari sample size *50* query yang paling representatif, sistem mencapai *Precision@10* sebesar *88.40%*, *Recall@10* sebesar *46.81%*, dan *F1-Score@10* sebesar *50.88%*, dengan nilai *F1-Score* optimal dicapai pada *K=30* sebesar *66.79%*. Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan Cosine Similarity mampu memberikan rekomendasi produk yang relevan dengan tingkat akurasi yang baik, sehingga metode ini layak diterapkan sebagai solusi sistem rekomendasi pada platform e-commerce berbasis konten

Kata Kunci:
Confusion Matrix, Content-Based Filtering, Cosine Similarity. E-Commerce, Sistem Rekomendasi, TF-IDF

PENDAHULUAN

Perkembangan platform digital dan e-commerce telah mendorong peningkatan jumlah produk dan informasi yang tersedia bagi pengguna. Kondisi ini memberikan kemudahan dalam proses pencarian produk, namun sekaligus menimbulkan permasalahan *information overload*, yaitu kesulitan pengguna dalam menentukan pilihan produk yang sesuai dengan kebutuhan dan preferensinya. Untuk mengatasi permasalahan tersebut, sistem rekomendasi banyak dikembangkan sebagai solusi yang mampu membantu pengguna dalam proses pengambilan keputusan secara lebih efektif.

Berbagai penelitian sebelumnya menunjukkan bahwa sistem rekomendasi berbasis *content-based filtering* merupakan pendekatan yang efektif dalam memberikan rekomendasi berdasarkan karakteristik item. Pebrianti *et al.* [1] mengombinasikan *content-based filtering* dengan algoritma Apriori pada sistem belanja cerdas, sementara Sulami *et al.* [2] menerapkan *content-based filtering* untuk pemilihan produk skincare. Pendekatan serupa juga diterapkan pada berbagai domain lain, seperti rekomendasi resep makanan [3], berita berbasis RSS [4], pariwisata [5][6], film [7], dan restoran [8]. Hasil penelitian-penelitian tersebut menunjukkan bahwa *content-based filtering* mampu memberikan rekomendasi yang relevan tanpa bergantung pada data pengguna lain.

Dalam konteks representasi data berbasis teks, beberapa penelitian menerapkan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) yang dikombinasikan dengan *Cosine Similarity* untuk mengukur tingkat kemiripan antar item. Pendekatan ini terbukti efektif pada sistem rekomendasi film berbasis sinopsis [7] serta berbagai sistem rekomendasi berbasis deskripsi teks lainnya. Dibandingkan dengan pendekatan *collaborative filtering* [9], *content-based filtering* memiliki keunggulan dalam mengatasi permasalahan *cold-start* serta lebih sesuai untuk dataset yang tidak memiliki riwayat interaksi pengguna yang lengkap. Beberapa penelitian terkini bahkan telah mulai mengeksplorasi pendekatan berbasis *deep learning* untuk meningkatkan performa sistem rekomendasi [10], namun pendekatan tersebut memiliki kompleksitas komputasi yang lebih tinggi dan membutuhkan sumber daya yang besar.

Berdasarkan celah penelitian tersebut, permasalahan yang dikaji dalam penelitian ini adalah bagaimana menerapkan metode *content-based filtering* menggunakan TF-IDF dan *Cosine Similarity* pada dataset produk e-commerce lokal berskala besar serta bagaimana kinerja sistem rekomendasi yang dihasilkan apabila dievaluasi menggunakan pendekatan *confusion matrix*. Penelitian ini bertujuan untuk membangun sistem rekomendasi produk e-commerce berbasis konten dengan memanfaatkan TF-IDF sebagai metode representasi fitur dan *Cosine Similarity* sebagai metode pengukuran kemiripan antar produk, serta mengevaluasi kinerjanya menggunakan metrik akurasi, presisi, *recall*, dan F1-score.

Kontribusi (novelty) dari penelitian ini terletak pada tiga aspek utama. Pertama, penelitian ini menggunakan dataset Tokopedia Products 2025 yang merepresentasikan data e-commerce Indonesia yang aktual dan berskala besar, sehingga memperluas kajian *content-based filtering* pada konteks lokal. Kedua, penelitian ini menerapkan skema evaluasi berbasis *confusion matrix* pada sistem rekomendasi berbasis konten, yang masih relatif jarang digunakan dalam penelitian-penelitian sebelumnya. Ketiga, penelitian ini menyajikan analisis empiris kinerja sistem rekomendasi berbasis TF-IDF dan *Cosine Similarity* pada data e-commerce nyata, sehingga dapat menjadi referensi metodologis bagi penelitian dan pengembangan sistem rekomendasi di bidang ilmu komputer dan sistem informasi.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

TINJAUAN PUSTAKA

Sistem rekomendasi merupakan salah satu bidang penelitian yang berkembang pesat dalam ilmu komputer, khususnya dalam konteks pengolahan data dan sistem pendukung keputusan. Sistem ini bertujuan untuk memberikan rekomendasi item yang relevan kepada pengguna berdasarkan data dan karakteristik tertentu. Secara umum, pendekatan sistem rekomendasi dapat diklasifikasikan menjadi *content-based filtering*, *collaborative filtering*, dan pendekatan hibrida yang mengombinasikan keduanya.

Pendekatan *content-based filtering* merekomendasikan item berdasarkan kesamaan karakteristik atau konten antar item. Beberapa penelitian menunjukkan bahwa pendekatan ini efektif dalam berbagai domain. Pebrianti *et al.* [1] mengombinasikan *content-based filtering* dengan algoritma Apriori dalam sistem belanja cerdas, sedangkan Sulami *et al.* [2] menerapkan metode ini untuk rekomendasi produk skincare. Pendekatan serupa juga digunakan dalam sistem rekomendasi resep makanan [3], berita berbasis RSS [4], pariwisata [5][6], film [7], dan restoran [8]. Hasil penelitian-penelitian tersebut menunjukkan bahwa *content-based filtering* mampu memberikan rekomendasi yang relevan tanpa bergantung pada preferensi pengguna lain.

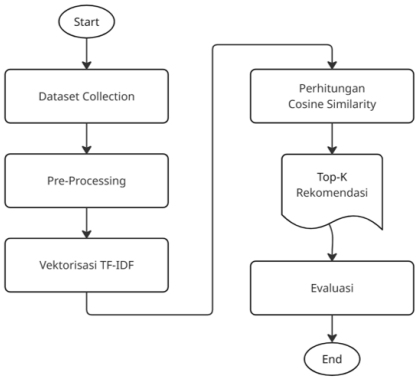
Dalam sistem rekomendasi berbasis teks, proses kombinasi TF-IDF dengan *Cosine Similarity* telah terbukti efektif dalam mengukur tingkat kemiripan antar item, sebagaimana diterapkan pada sistem rekomendasi film berbasis sinopsis [7] serta berbagai sistem rekomendasi berbasis deskripsi teks lainnya.

Selain *content-based filtering*, pendekatan *collaborative filtering* juga banyak digunakan dalam sistem rekomendasi. Pendekatan ini memanfaatkan pola perilaku dan preferensi pengguna lain untuk menghasilkan rekomendasi, sebagaimana diterapkan pada sistem rekomendasi tempat wisata [9]. Perkembangan terkini dalam sistem rekomendasi juga mengarah pada pemanfaatan metode pembelajaran mesin dan *deep learning* untuk meningkatkan kinerja rekomendasi. Maulana dan Setiawan [10] menerapkan algoritma *deep feed-forward neural network* berbasis *content-based filtering* dan menunjukkan peningkatan performa dibandingkan metode konvensional.

Berdasarkan kajian pustaka tersebut, dapat disimpulkan bahwa *content-based filtering* dengan TF-IDF dan *Cosine Similarity* merupakan pendekatan yang telah banyak digunakan dan terbukti efektif dalam berbagai domain. Pada penelitian ini difokuskan untuk mengisi celah tersebut dengan menyajikan analisis empiris kinerja sistem rekomendasi produk e-commerce berbasis konten.

METODE PENELITIAN

Pada penelitian ini memiliki alur yang dapat dilihat pada Gambar 1.



GAMBAR 1. Alur Penelitian

Dataset

Metode penelitian diawali dengan pencarian dataset yang relevan dari dataset publik di platform *Kaggle*. Tahap dataset collection ini bertujuan untuk mengumpulkan data produk e-commerce yang akan digunakan sebagai objek penelitian. Data yang dikumpulkan berisi informasi produk yang merepresentasikan karakteristik konten setiap item dan menjadi dasar utama dalam proses analisis serta pengembangan sistem rekomendasi.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Pre-processing

Tahap *Pre-Processing* dilakukan untuk membersihkan dan mempersiapkan data sebelum diproses lebih lanjut. Proses ini mencakup seleksi atribut yang relevan, penghapusan data yang tidak diperlukan, serta normalisasi data teks agar memiliki format yang konsisten dan siap untuk tahap vektorisasi.

Vektorisasi TF-IDF

Tahap Vektorisasi *TF-IDF* bertujuan untuk mengubah data teks menjadi bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (*TF-IDF*). Proses ini menghasilkan representasi vektor yang menggambarkan tingkat kepentingan setiap kata dalam suatu produk terhadap keseluruhan dataset. Pada tahap ini akam menggunakan 3 persamaan yaitu Term-Frequency (TF), Inverse Document Frequency (IDF), dan TF-IDF.

$$\mathrm{TF}(t,d) = n_{t,d} \quad 1)$$

$$\mathrm{IDF}(t) = \log_{10}\left(\frac{N}{df_t}\right) + 1 \quad 2)$$

$$\text{TF-IDF}(t,d) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t) \quad 3)$$

Perhitungan Cosine Similarity

Tahap Perhitungan *Cosine Similarity* dilakukan untuk mengukur tingkat kemiripan antar produk berdasarkan vektor *TF-IDF* yang telah dihasilkan. Nilai kemiripan ini digunakan untuk menentukan seberapa dekat karakteristik konten antar produk dalam ruang vektor. Pada tahapan ini akan menggunakan persamaan 4, yaitu:

$$\left|\left|\mathrm{D1}\right|\right| = \sqrt{\mathrm{D1}_1^2 + \mathrm{D1}_2^2 + \cdots + \mathrm{D1}_n^2} \quad 4)$$

Rekomendasi Top-K Produk

Tahap *Top-K Rekomendasi* merupakan proses pemilihan sejumlah *Top-K* item produk dengan nilai kemiripan tertinggi terhadap produk acuan, Tahap ini merupakan sebuah hasil implementasi sistem rekomendasi produk e-commerce dalam bentuk prototype website. Produk-produk tersebut kemudian ditetapkan/ditampilkan dalam interface ui website sebagai hasil rekomendasi yang paling relevan bagi pengguna berdasarkan perhitungan nilai kemiripan dari metode cosine similarity.

Evaluasi

Tahap Evaluasi bertujuan untuk menilai kinerja sistem rekomendasi yang telah dibangun. Proses ini dilakukan dengan menganalisis kualitas dan relevansi rekomendasi yang dihasilkan berdasarkan pada metode evaluasi *confusion matrix*.

HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset publik yang bersumber dari platform Kaggle dengan judul *"Tokopedia Products 2025"* yang dapat diakses melalui url https://www.kaggle.com/datasets/nsmlehq/tokopedia-products-2025/. Dataset ini berisi informasi produk e-commerce yang diambil dari platform Tokopedia, salah satu marketplace terbesar di Indonesia. Total dataset terdiri dari *1.200* record produk yang tersimpan dalam format csv.

Dataset memiliki struktur yang terdiri dari *36 atribut* yang merepresentasikan berbagai karakteristik produk. Atribut-atribut tersebut meliputi informasi dasar produk seperti *id (ID)*, *name (Nama Produk)*, *description (Deskripsi)*, *price (Harga)*, *condition (Kondisi)*, dan *status*. Informasi kategori produk mencakup *category_id (ID Kategori)*, *category_name (Kategori)*, dan *category_url (URL Kategori)*. Informasi toko penjual terdiri dari *shop_id (ID Toko)*, *shop_name (Nama Toko)*, *shop_city (Kota Toko)*, *shop_url (URL Shop)*, dan *shop_reputation (Reputasi Toko)*. Atribut visual produk mencakup *thumbnail* dan *images* yang berisi url gambar produk. Selain itu, dataset juga menyediakan informasi statistik performa produk seperti *sold_count*, *view_count*, *review_count*, *talk_count*, dan *rating* yang merepresentasikan popularitas dan tingkat kepercayaan konsumen terhadap produk. Informasi diskon tersedia melalui atribut *discount_percent*, *price_text*, dan *price_slash_text*. Atribut tambahan seperti *min_order*, *max_order*, *weight*,

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

dan *weight_unit* memberikan informasi teknis terkait pemesanan dan pengiriman produk. Berikut pada Gambar 2 adalah contoh beberapa data yang ada dalam dataset.

TABEL 1. Struktur Dataset

ID	Nama Produk	Deskripsi	Harga	Berat	Kondisi	Status	Kategori	URL Kategori	Sold Count	View Count	Rating
110296338456	ONLINE EXCLUSIVE VIVO Y100	Vivo Y100 Smartphone	3.077.000	0.6 kg	Baru	Aktif	Handphone	tokopedia.com	17rb+	988	4.9
110429538485	Paket Bundle Your Coffee	Kopi & Tumbler	117.000	0.5 kg	Baru	Aktif	Kopi	tokopedia.com	1rb+	398	4.9
110264351	Sweaty Bronze Dry T-Shirt	Kaos Olahraga	110.000	0.2 kg	Baru	Aktif	Pakaian	tokopedia.com	3rb+	684	4.9

Pada penelitian ini, atribut yang digunakan sebagai fitur utama untuk sistem rekomendasi adalah *name*, *description*, *category_name*, dan *shop_name*. Keempat atribut ini dipilih karena mengandung informasi tekstual yang relevan untuk menganalisis karakteristik konten produk dan mengidentifikasi kesamaan antar produk menggunakan metode *Content-Based Filtering*.

1. Pre-processing

Pada penelitian ini, proses pre-processing dilakukan melalui 5 tahapan berurutan yang diimplementasikan menggunakan library *Natural.js* dan *Stopword*. Proses tersebut meliputi:

A. Feature Selection

Tahap pertama adalah pemilihan fitur yang relevan dari dataset. Dari total *36* atribut yang tersedia dalam dataset, dipilih 4 atribut utama yang mengandung informasi tekstual paling relevan untuk analisis konten produk. Atribut-atribut tersebut adalah *name* yang merepresentasikan nama produk, *description* yang berisi deskripsi detail produk, *category_name* yang menunjukkan kategori produk, dan *shop_name* yang mengidentifikasi nama toko penjual. Keempat atribut ini digabungkan menjadi satu teks yang merepresentasikan karakteristik konten produk.

B. Cleaning & Normalization

Tahap cleaning dan normalisasi dilakukan untuk membersihkan data dari elemen-elemen yang tidak diperlukan dalam analisis teks. Proses ini dilakukan secara berurutan melalui beberapa tahapan. Pertama, dilakukan penghapusan *html tags* untuk menghilangkan elemen markup html. Kedua, penghapusan emoji dilakukan dengan mendeteksi dan menghilangkan karakter emoji unicode yang tidak relevan untuk analisis konten tekstual. Ketiga, penghapusan simbol dan karakter seperti tanda baca dan simbol, hanya menyisakan huruf, angka, dan spasi. Keempat, normalisasi spasi dilakukan dengan mengganti multiple spaces menjadi single space untuk konsistensi format. Kelima, case folding diterapkan dengan mengkonversi seluruh teks menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi dalam pencocokan term dan menghindari duplikasi term yang sama dengan kapitalisasi berbeda.

C. Tokenization

Pada penelitian ini, tokenisasi dilakukan untuk memecah teks menjadi unit-unit kata individual. Tokenisasi ini memisahkan teks berdasarkan delimiter seperti spasi dan tanda baca dengan pemisahan kata. Hasil dari proses tokenisasi adalah sebuah *array of strings* yang berisi kata-kata individual yang telah terpisah yang siap untuk tahap preprocessing selanjutnya.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

D. Stopword Removal

Stopword removal adalah proses penghapusan kata-kata umum seperti *"yang"*, *"dan"*, *"di"*, *"the"*, *"a"*, *"is"* yang tidak memberikan nilai diskriminatif. Pada penelitian ini, penghapusan stopword dilakukan menggunakan kombinasi daftar stopword bahasa indonesia dan bahasa inggris mengingat dataset produk *e-commerce* mengandung istilah dari kedua bahasa. Setiap token dibandingkan dengan kumpulan stopword yang telah didefinisikan dan dihilangkan jika termasuk di dalamnya. Hasil akhir adalah array token bermakna dengan nilai informasi tinggi untuk analisis konten dan mengurangi dimensi vektor.

E. Stemming

Stemming merupakan tahap akhir dari pre-processing dalam penelitian ini yang mentransformasi kata menjadi bentuk dasar dengan menghilangkan imbuhan bahasa Indonesia seperti *"me-"*, *"ber-"*, *"di-"*, *"-kan"*, *"-an"*, dan kombinasinya. Proses ini memastikan variasi bentuk kata yang secara semantik sama dipetakan ke satu term konsisten. Sebagai contoh, *"membeli"*, *"dibeli"*, *"pembelian"*, dan *"belian"* akan dikonversi menjadi *"beli"*. Hasil akhir dari seluruh tahapan pre-processing adalah array of string tokens yang telah dibersihkan, dinormalisasi, dan siap untuk vektorisasi *TF-IDF*.

Vektorisasi TF-IDF

Pada bagian ini akan diuraikan contoh perhitungan dari algoritma *TF-IDF* dan kemudian pada tahapan berikutnya proses *Cosine Similarity* melalui perhitungan manual menggunakan data sampel produk e-commerce. Tahapan cara kerja model dan algoritma ini dilakukan secara sistematis mulai dari tahap pemilihan data sampel, tahap preprocessing data, vektorisasi menggunakan *TF-IDF*, hingga pengukuran kemiripan menggunakan *Cosine Similarity* untuk menghasilkan rekomendasi produk yang relevan.

Data sampel yang digunakan dalam perhitungan ini terdiri dari 5 item produk e-commerce yang dipilih dari dataset *Tokopedia Products 2025*. Informasi mengenai attribut yang digunakan dalam data sampel dapat dilihat pada TABEL 2.

TABEL 2. Data Sampel

	Nama	***Deskripsi***	***Kategori***	***Toko***
D1	Teko Listrik Stainless 2L Pemanas Air Kettle Listrik Electric Kettle	Electric Kettle - Teko Listrik\n\nModel : Electric Kettle / Teko Listrik\nKapasitas : 1,8L\nDaya : 500 Watt\nVoltase : 220-240V\n\nKelengkapan :\nTeko x 1\nInduksi x 1\nManual x 1\n\nElectric Kettle - Teko Listrik yang sangat berguna untuk Rumah Tangga Anda memudahkan dan dengan waktu yang cepat untuk memasak air \nDengan Tenaga Listrik sangat memudahkan anda untuk memperoleh air panas saat dibutuhkan tanpa kompor\nDilengkapi Overheat protector sangat aman digunakan\n	Teko Listrik	AchillesCollection
D2	Tolak Angin Madu Sido Muncul Box isi 12 sachet Tolakangin	Tolak Angin Madu Sido Muncul Box isi 12 Sachet adalah produk herbal yang mengandung bahan-bahan alami seperti ekstrak jahe, daun mint, dan madu, yang dirancang untuk membantu menjaga kesehatan tubuh dan meningkatkan daya tahan tubuh. Dengan tambahan madu, produk ini memiliki rasa yang lebih manis dan menyegarkan, membuatnya lebih mudah diminum. Setiap box berisi 12 sachet yang praktis, sehingga memudahkan Anda untuk mengonsumsinya kapan saja, baik di rumah, di kantor, atau saat bepergian.\n	Multivitamin	Tobaking Murah Online
D3	Kipas Angin Sekai HDO 615-S	Type HDO 615-S\n- Kipas angin high velocity 15 cm (6")\n- Baling baling dari Plastik\n- 2 pengatur kecepatan\n- Hembusan angin kencang\n- Hemat listrik\n- Tidak berisik\n- Dilengkapi sekring pengaman\n- Garansi 3 tahun\n- Daya 20 W, 220 VAC, 50 Hz\n	Kipas Angin Listrik	amac electronic store
D4	Kopi LIONG BULAN Plus Gula 28 GR (Isi 20 Sachet)	HALAL \n\nIsi : 20 Sachet (20 x 28 Gr) Kopi Liong Bulan Plus Gula\n\nTerbuat dari 100% Biji kopi murni dan gula tanpa campuran dan pengawet.\n\nTHE LOCAL LEGEND\n\nKopi LIONG BULAN merupakan merek kopi asal Bogor yang sudah hadir sejak tahun 1945.\nCita RASA yang UNIK serta AROMA nya yang KHAS membuat Kopi LIONG BULAN menjadi cerita dari mulut ke mulut secara turun temurun.\n\nBarang selalu fresh, produksi terbaru dari	Kopi Kemasan	Kopi Liong Bulan

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

		pabrik.\nExp Date / Kadaluarsa : 3 bulan sejak tanggal pemesanan (Tercantum di kode produksi di belakang)\n		
D5	saringan kopi teh portable V60 / coffee filter stainless double mesh	saringan kopi portable double mesh bahan terbuat dari stainless 304 dengan pinggiran ABS\ntidak perlu menggunakan kertas lagi\nBrand: vesper\ndimensi\ntinggi 7,5cm\ndiameter atas 6,5cm\ndiamater atas dengan kuping 12cm\npremium coffee filter\n	Saringan Masak	Vesperindo _NEW

Dari data sampel yang terdapat pada TABEL 2 dilakukan tahap pre-processing yaitu tahap feature selection, cleaning & normalization, tokenization, stopword removal, dan stemming yang kemudian menghasilkan kumpulan token yang bersih, konsisten, dan relevan untuk analisis konten produk. Hasil pre-processing dapat dilihat pada TABEL 3.

TABEL 3. Token Data Sampel

	Tokens
D1	*[listrik, teko, kettle, 1, electric, air, x, panas, lengkap, mudah, 220, 500, stainless, 2l, model, kapasitas, 8l, daya, watt, voltase, 240v, induksi, manual, guna, rumah, tangga, cepat, masak, tenaga, oleh, butuh, kompor, overheat, protector, aman, achillescollection]*
D2	*[madu, 12, box, isi, sachet, tolak, angin, sido, muncul, produk, bahan, tubuh, mudah, tolakangin, herbal, kandung, alami, ekstrak, jahe, daun, mint, rancang, bantu, jaga, sehat, tingkat, tahan, tambah, milik, manis, segar, buat, minum, praktis, konsumsi, baik, kantor, pergi, multivitamin, tobaking, murah, online, daya, rumah]*
D3	*[kipas, angin, 615, hdo, s, baling, listrik, 2, 3, 6, 15, 20, 50, seka, type, high, velocity, cm, plastik, atur, hembus, kencang, hemat, berisik, sekring, kam, garansi, w, vac, hz, amac, electronic, store, 220, cepat, lengkap, daya]*
D4	*[kopi, liong, gula, 20, 28, plus, gr, mulut, turun, produksi, isi, sachet, 100, 1945, halal, biji, murni, campur, awet, local, legend, merek, bogor, hadir, cita, unik, aroma, nya, khas, cerita, barang, fresh, baru, pabrik, exp, date, kadaluarsa, tanggal, mesan, cantum, kode, kemas, 3, x, buat]*
D5	*[saring, portable, coffee, filter, double, mesh, 5cm, kopi, stainless, 7, 304, teh, v60, pinggir, abs, kertas, brand, vesper, dimensi, diameter, diamater, kuping, 12cm, premium, vesperindo, new, 6, bahan, masak, buat]*

Setelah melewati tahapan preprocessing, selanjutnya dilakukan konversi/ekstraksi fitur menggunakan algoritma *TF-IDF* untuk mengkonversi token-token teks menjadi representasi vektor numerik. Proses ini dimulai dengan menghitung frekuensi kemunculan setiap term dalam dokumen (*Term Frequency*/*TF*), kemudian mengalikannya dengan nilai kebalikan frekuensi dokumen (*Inverse Document Frequency*/*IDF*) untuk menghasilkan bobot *TF-IDF*. Bobot *TF-IDF* ini mencerminkan tingkat kepentingan setiap term dalam merepresentasikan karakteristik unik suatu produk terhadap keseluruhan dataset.

Langkah pertama dalam *TF-IDF* adalah dengan mengukur *Term Frequency* (*TF*), yakni mengukur seberapa sering suatu *term* muncul dalam dokumen tertentu menggunakan rumus pada persamaan 1. Dari rumus tersebut, berikut adalah contoh perhitungan *TF* untuk term "*daya*".

Term *daya* muncul 1 kali dalam *D1*:

$$\text{TF}(daya, D1) = \mathbf{1}$$

Term *daya* muncul 1 kali dalam *D2*:

$$\text{TF}(daya, D2) = \mathbf{1}$$

Lakukan proses perhitungan yang sama untuk memperoleh TF(*daya*, *D3*), TF(*daya*, *D4*), dan TF(*daya*, *D5*). Selanjutnya adalah perhitungan *Inverse Document Frequency* (*IDF*) yang menggunakan persamaan 2.

$$\text{IDF}(daya) = \log_{10}\left(\frac{5}{3}\right) + 1$$
$$\text{IDF}(daya) = \log_{10}(1.667) + 1$$
$$\text{IDF}(daya) = 0.2218 + 1 = \mathbf{1.2218}$$

Kemudian menghitung *TF-IDF* dengan menggunakan persamaan 3 maka perhitungan *TF-IDF* term *daya* pada dokumen *D1*, *D2*, *D3*, *D4*, dan *D5* adalah sebagai berikut:

$$\text{TF-IDF}(daya, D1) = 1 \times 1.2218 = 1.2218$$

Lakukan proses perhitungan yang sama untuk memperoleh TF-IDF(*daya*, *D2*) sampai TF-IDF(*daya*, *D5*).

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Dari hasil perhitungan *TF-IDF* semua term pada dokumen *D1*, *D2*, *D3*, *D4*, dan *D5*. Hasil perhitungan *TF-IDF* untuk seluruh term yang muncul dalam kelima dokumen sampel disajikan pada TABEL 4.

TABEL 4. Vector Hasil Perhitungan TF-IDF

	Vector TF-IDF
D1	*[10.576, 11.498, 9.581, 7.665, 7.665, 5.749, 4.532, 3.833, 3.022, 3.022, 1.511, 1.916, 1.511, 1.916, 1.916, 1.916, 1.916, 1.223, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.511, 1.916, 1.511, 1.511, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 0]*
D2	*[0, 0, 0, 0, 0, 0, 0, 0, 0, 3.022, 0, 0, 0, 0, 0, 0, 0, 1.223, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7.665, 5.749, 5.749, 4.532, 4.532, 3.833, 3.022, 3.833, 3.833, 3.833, 3.022, 3.833, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.223, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]*
D3	*[3.022, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 1.511, 0, 0, 0, 0, 0, 0, 1.223, 0, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6.043, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.749, 3.833, 3.833, 3.833, 3.833, 1.916, 1.511, 1.511, 1.916, 1.511, 1.916, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]*
D4	*[0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.022, 3.022, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.223, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 4.532, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 12.087, 9.581, 5.749, 3.833, 3.833, 3.833, 3.833, 3.833, 3.833, 1.916, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]*
D5	*[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.022, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.223, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.511, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3.022, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5.749, 3.833, 3.833, 3.833, 3.833, 3.833, 3.833, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916, 1.916]*

Cosine Similarity

Setelah menghitung *TF-IDF*, selanjutnya ialah menghitung similaritas antar sudut kosinus vector *TF-IDF* menggunakan algoritma *Cosine Similarity*. Namun sebelum itu, disini pada penelitian ini kami hanya menggunakan beberapa data vector hasil perhitungan *TF-IDF* sebagai sampel perhitungan agar mempermudah ilustrasi perhitungan *Cosine Similarity*. Data vector hasil perhitungan TF-IDF yang kami gunakan dapat dilihat pada TABEL 5.

TABEL 5. Matrix Bobot Sampel TF-IDF

Term	*D1*	*D2*	*D3*	*D4*	*D5*
listrik	10.5758	0	3.0217	0	0
buat	0	1.2231	0	1.2231	1.2231
panas	3.8326	0	0	0	0
lengkap	3.0217	0	1.5108	0	0
mudah	3.0217	3.0217	0	0	0
sachet	0	4.5325	0	3.0217	0
aman	1.9163	0	0	0	0
stainless	1.5108	0	0	0	3.0217
model	1.9163	0	0	0	0
kapasitas	1.9163	0	0	0	0
daya	1.2231	1.2231	1.2231	0	0

Pada contoh perhitungan berikut ini akan menghitung similaritas antar kedua dokumen menggunakan *Cosine Similarity* untuk dokumen *D1* dengan dokumen *D3*. Langkah awal ialah menghitung *Dot Product D1·D3*, Dot product dihitung dengan mengalikan nilai *TF-IDF* setiap term yang bersinggungan pada kedua dokumen, kemudian menjumlahkan seluruh hasil perkalian.

$$D1 \cdot D3 = (D1_1 \times D3_1) + (D1_2 \times D3_2) + \cdots + (D1_n \times D3_n)$$

$$\begin{aligned} \mathrm{D1} \cdot \mathrm{D3} = {} & (10.5758 \times 3.0217) + (0 \times 0) + (3.8326 \times 0) + (3.0217 \times 1.5108) + \\ & (3.0217 \times 0) + (0 \times 0) + (1.9163 \times 0) + (1.5108 \times 0) + \\ & (1.9163 \times 0) + (1.9163 \times 0) + (1.2231 \times 1.2231) \end{aligned}$$

$$\mathrm{D1} \cdot \mathrm{D3} = 31.9620 + 0 + 0 + 4.5644 + 0 + 0 + 0 + 0 + 0 + 0 + 1.4960$$

$$\mathrm{D1} \cdot \mathrm{D3} = \mathbf{38.0224}$$

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Kemudian dilanjutkan dengan menghitung magnitude vektor |D1| yang dimana dihitung dengan akar kuadrat dari jumlah kuadrat setiap elemen dengan menggunakan persamaan 4.

$$||D1|| = \sqrt{10.5758^2 + 0^2 + 3.8326^2 + 3.0217^2 + 3.0217^2 + 0^2 + 1.9163^2 + 1.5108^2 + 1.9163^2 + 1.9163^2 + 1.2231^2}$$

$$||D1|| = \sqrt{111.8475 + 0 + 14.6888 + 9.1307 + 9.1307 + 0 + 3.6722 + 2.2825 + 3.6722 + 3.6722 + 1.4960}$$

$$||D1|| = \sqrt{159.5928} = \mathbf{12.6332}$$

Lalu dilanjutkan dengan menghitung magnitude vektor |D3| yang dimana dihitung dengan cara yang sama seperti menghitung magnitude vektor |D1|:

$$||D3|| = \sqrt{3.0217^2 + 0^2 + 0^2 + 1.5108^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 1.2231^2}$$

$$||D3|| = \sqrt{9.1307 + 0 + 0 + 2.2825 + 0 + 0 + 0 + 0 + 0 + 0 + 1.4960} = \sqrt{12.9092} = \mathbf{3.5930}$$

Setelah mendapatkan nilai dot product dan magnitude kedua vektor, dapat dilanjutkan dengan menghitung *Cosine Similarity*:

$$\text{Cos}(D1,D3) = \frac{D1 \cdot D3}{||D1|| \times ||D3||}$$

$$\text{Cos}(D1,D3) = \frac{38.0224}{12.6332 \times 3.5930}$$

$$\text{Cos}(D1,D3) = \mathbf{0.8377}$$

Dari hasil perhitungan nilai *Cosine Similarity* diatas antara produk *D1* dan produk *D3* dapat disimpulkan bahwa kemiripannya adalah sebesar **0.8377** atau **83.77%** similar/mirip. Nilai tersebut mengindikasikan bahwa kedua produk memiliki tingkat kemiripan yang sangat tinggi *>80%*. Hal ini dapat dijelaskan karena keduanya merupakan produk elektronik yang menggunakan listrik, sehingga berbagi term-term yang sama seperti "*listrik*", "*lengkap*", dan "*daya*" yang berkontribusi signifikan terhadap nilai similarity.

TABEL 6 merupakan hasil lengkap dari nilai *Cosine Similarity* antar semua dokumen dalam sampel yang digunakan.

TABEL 6. Matrix Similarity

	D1	***D2***	***D3***	***D4***	***D5***
D1	1.0000	0.1472	0.8377	0.0000	0.1201
D2	0.1472	1.0000	0.1030	0.8154	0.0000
D3	0.8377	0.1030	1.0000	0.0000	0.0000
D4	0.0000	0.8154	0.0000	1.0000	0.0000
D5	0.1201	0.0000	0.0000	0.0000	1.0000

Rekomendasi Top-K Product

Tahapan ini dilakukan untuk memvalidasi efektivitas algoritma *Content-Based Filtering* yang di implementasikan dengan variasi parameter *Top-K* dengan menggunakan dataset beberapa item produk. Evaluasi menggunakan *50* query produk yang disampling dari berbagai kategori dengan similarity threshold *0.85* untuk menentukan relevansi. Gambar 2 dibawah ini menunjukkan hasil evaluasi komprehensif sistem rekomendasi pada semua variasi sample size dan nilai K.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

SUMMARY EVALUASI SEMUA SAMPLE SIZE

	SampleSize	K	Precision	Recall	F1-Score	MAP	AvgSim
0	5	10	100.00	36.30	50.44	36.30	2.607
1	5	15	97.33	51.37	63.46	51.37	2.181
2	5	20	93.00	61.82	70.45	61.82	1.811
3	5	30	88.67	82.73	82.09	82.73	1.623
4	5	50	67.60	100.00	78.83	100.00	1.289
5	5	100	33.80	100.00	49.32	100.00	0.928
6	10	10	78.00	50.65	41.84	50.65	2.419
7	10	15	75.33	60.98	47.93	60.98	2.183
8	10	20	71.50	67.97	50.85	67.97	1.887
9	10	30	63.33	76.14	51.87	76.14	1.583
10	10	50	54.80	86.90	52.28	86.90	1.255
11	10	100	37.30	99.35	45.62	99.35	0.896
12	20	10	97.50	42.73	53.42	42.73	2.784
13	20	15	94.00	57.85	64.29	57.85	2.356
14	20	20	87.50	66.67	67.84	66.67	2.067
15	20	30	79.00	81.33	72.15	81.33	1.700
16	20	50	68.40	92.58	65.72	92.58	1.319
17	20	100	36.00	97.86	47.18	97.86	0.933
18	30	10	98.00	50.71	60.08	50.71	2.973
19	30	15	90.89	65.44	67.46	65.44	2.500
20	30	20	81.67	72.13	67.07	72.13	2.201
21	30	30	69.11	80.51	64.24	80.51	1.818
22	30	50	54.87	91.88	58.87	91.88	1.394
23	30	100	34.70	97.82	44.10	97.82	0.948
24	40	10	92.75	49.46	55.20	49.46	2.538
25	40	15	87.50	63.10	62.94	63.10	2.165
26	40	20	81.13	71.59	65.35	71.59	1.913
27	40	30	69.58	81.56	64.53	81.56	1.585
28	40	50	54.80	91.20	59.86	91.20	1.247
29	40	100	34.48	100.00	45.41	100.00	0.889
30	50	10	88.40	46.81	50.88	46.81	2.457
31	50	15	85.33	60.91	60.54	60.91	2.085
32	50	20	80.70	71.12	65.18	71.12	1.839
33	50	30	70.67	83.72	66.79	83.72	1.529
34	50	50	51.56	98.65	57.17	98.65	1.202
35	50	100	31.36	95.96	41.23	95.96	0.862

GAMBAR 2. Hasil Top-K

Berdasarkan hasil evaluasi pada berbagai sample size dan nilai K yang terlihat seperti pada Gambar 2, sistem rekomendasi yang diimplementasikan menunjukkan performa yang konsisten dan robust untuk produk e-commerce. Untuk implementasi praktis, nilai *K=20* hingga *K=30* direkomendasikan sebagai ukuran optimal yang memberikan keseimbangan antara kualitas rekomendasi (precision *70.67%-87.50%*) dan coverage produk relevan (recall *71.12%-83.72%*) pada sample size representatif. Precision tinggi pada *K=10* (rentang *88.40%-98.00%* untuk sample size *20-50*) mengindikasikan bahwa pengguna akan mendapatkan rekomendasi yang sangat relevan pada posisi teratas, meningkatkan user experience dan potensi konversi pembelian. Recall yang mencapai *90.65%-92.58%* pada *K=50* menunjukkan bahwa sistem mampu mengidentifikasi hampir seluruh produk relevan, memberikan variasi pilihan yang cukup bagi pengguna tanpa mengorbankan relevansi secara signifikan. Konsistensi performa pada berbagai sample size memvalidasi robustness algoritma *Content-Based Filtering* dengan *TF-IDF* dan *Cosine Similarity* untuk kategori produk item e-commerce pada penelitian ini.

Evaluasi

Evaluasi sistem rekomendasi dilakukan menggunakan *metric calculation*/*confussion matrix* standar yang umum digunakan dalam penelitian sistem rekomendasi yaitu *Precision@K*, *Recall@K*, dan *F1-Score@K*. Proses evaluasi diimplementasikan dengan menggunakan bahasa pemrograman Python yang instruksi evaluasi tertampil pada Gambar 3. Pada Gambar 3 terlihat pengaturan nilai K yang digunakan dalam evaluasi adalah 10, 15, 20, 30, 50, dan 100 dengan menggunakan treshold sebesar 0,05.

evaluation.ts

```ts
import { getProducts, type ProductWithRelations } from "@src/lib/data";
import { recommender } from "./lib/recommender";

// --- Parameter Evaluasi ---
const K_LIST = [10, 15, 20, 30, 50, 100];
const THRESHOLD = 0.05;
const SAMPLE_SIZES = [5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100];
const VERBOSE = false;

// --- Utilitas ---

function precision(recommended: number[], relevant: number[]): number {
  const hit = recommended.filter((id) => relevant.includes(id)).length;
  return recommended.length === 0 ? 0 : hit / recommended.length;
}

function recall(recommended: number[], relevant: number[]): number {
  const hit = recommended.filter((id) => relevant.includes(id)).length;
  return relevant.length === 0 ? 0 : hit / relevant.length;
}

function f1(p: number, r: number): number {
  return p + r === 0 ? 0 : (2 * p * r) / (p + r);
}

function mean(arr: number[]): number {
  return arr.length === 0 ? 0 : arr.reduce((a, b) => a + b, 0) / arr.length;
}

function meanAveragePrecision(recommended: number[], relevant: number[]): number {
  let hit = 0,
    sum = 0;
  recommended.forEach((id, idx) => {
    if (relevant.includes(id)) {
      hit++;
      sum += hit / (idx + 1);
    }
  });
  return relevant.length === 0 ? 0 : sum / relevant.length;
}
```

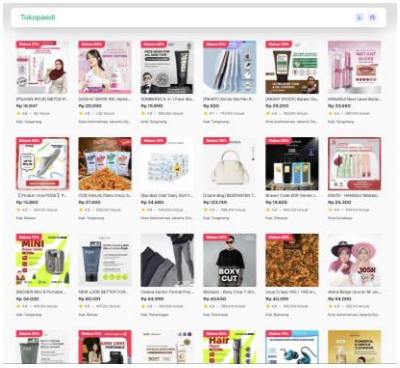
GAMBAR 3. Evaluation File

Proses evaluasi dilakukan dengan menjalankan script evaluasi yang mengiterasi seluruh query produk dalam ground truth, menghitung rekomendasi untuk setiap query, dan mengakumulasi metrik evaluasi. Evaluasi dilakukan untuk berbagai nilai K (*5*, *10*, *15*, *20*, *30*, ..., *100*) untuk menganalisis performa sistem pada berbagai ukuran rekomendasi.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

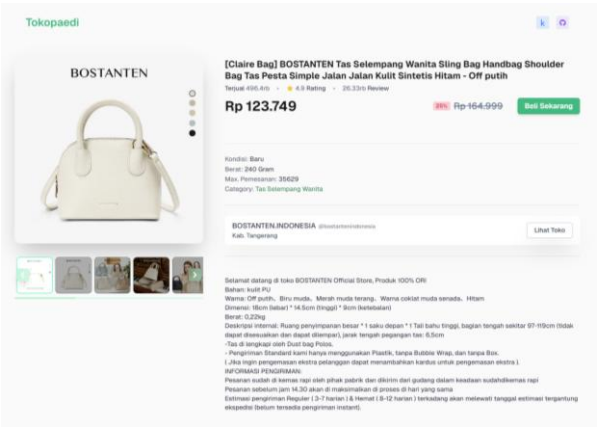
Hasil Implementasi

Hasil implementasi dari penelitian ini ialah berupa halaman website yang terdiri dari homepage, toko, product detail serta rekomendasinya. Gambar 4 menampilkan halaman homepage website yang berisi daftar produk atau berbagai macam produk. Produk-produk yang tampil pada Gambar 4 dapat dipilih oleh pengguna dengan bebas.



GAMBAR 4. Halaman Homepage

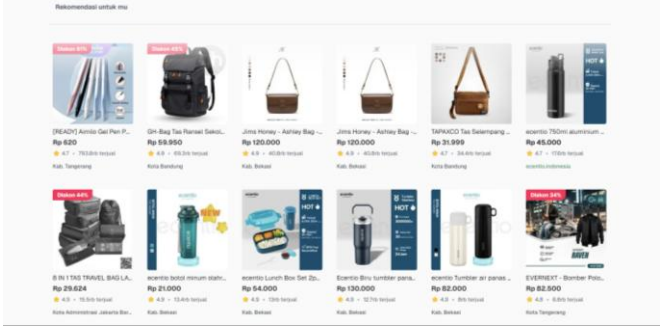
Ketika seorang pengguna memilih salah satu produk yang ditampilkan pada halaman homepage (Gambar 4), maka akan tampil halaman detail produk yang dipilih. Misal pengguna memilih produk Claire Bag dengan merek BOSTANTEN maka detail produk akan tampil seperti terlihat pada Gambar 5.



GAMBAR 5. Detail Produk

Pada Gambar 5, jika pengguna scrolling kebawah maka akan menampilkan hasil rekomendasi produk berdasarkan tingkat kemiripan produk terhadap produk referensinya (dalam case/kasus/contoh ini adalah jenis *Claire Bag*), mulai dari tingkat kemiripan yang tertinggi ke yang terendah.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR



GAMBAR 6. Halaman Detail Produk Section Rekomendasi Produk

Gambar 6 merupakan hasil rekomendasi dari hasil aksi yang dilakukan pada Gambar 5. Pada Gambar 5 dipilih produk berjenis Claire Bag dengan merek BOSTANTEN kemudian hasil produk lain yang direkomendasikan kepada user dapat terlihat pada Gambar 6 yang urutan produk rekomendasi nya telah diurutkan dari yang memiliki nilai kemiripan tertinggi sampai ke rendah.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem rekomendasi produk e-commerce menggunakan metode *Content-Based Filtering* dengan pendekatan *TF-IDF* dan *Cosine Similarity* pada dataset produk *Tokopedia* yang terdiri dari *1.199* record. Pada hasil evaluasi menggunakan berbagai ukuran sampel query dan nilai K yang terdiri dari sample size *50* query yang paling representatif, sistem mencapai *Precision@10* sebesar *88.40%*, *Recall@10* sebesar *46.81%*, dan *F1-Score@10* sebesar *50.88%*, dengan nilai *F1-Score* optimal dicapai pada *K=30* sebesar *66.79%*. Analisis variasi sample size memvalidasi stabilitas algoritma, dimana sample size yang lebih besar (*20-50* query) menghasilkan metrik evaluasi yang lebih representatif dengan variance rendah, dengan precision tetap berada di atas *70%* pada *K=30* untuk semua sample size representatif dan recall mencapai *90.65%-95.96%* pada *K=50-100*, membuktikan bahwa algoritma *Cosine Similarity* berhasil mengurutkan produk berdasarkan tingkat kemiripan dengan akurat. Berdasarkan hasil evaluasi, nilai *K=20* hingga *K=30* direkomendasikan sebagai ukuran optimal untuk implementasi praktis pada aplikasi e-commerce, memberikan keseimbangan antara kualitas rekomendasi (precision *70.67%-80.70%*) dan coverage produk relevan (recall *71.12%-83.72%*).

TINJAUAN PUSTAKA

1. D. Pebrianti, D. Ahmad, L. Bayuaji, L. Wijayanti, and M. Mulyadi, "Using Content-Based Filtering and Apriori for Recommendation Systems in a Smart Shopping System," *Indonesian Journal of Computing, Engineering, and Design (IJoCED)*, vol. 6, no. 1, pp. 58–70, Apr. 2024, doi: 10.35806/ijoced.v6i1.393.
2. A. Sulami, V. Atina, and N. Nurmalitasari, "Penerapan Metode Content Based Filtering dalam Sistem Rekomendasi Pemilihan Produk Skincare," *Satuan Tulisan Riset dan Inovasi Teknologi (STRING)*, vol. 9, no. 2, p. 172, Dec. 2024, doi: 10.30998/string.v9i2.24066.
3. N. P. T. Widiantari, I. M. A. D. Suarjaya, and N. K. D. Rusjayanthi, "Food Recipe Recommendation System with Content-Based Filtering and Collaborative Filtering Methods," *Jurnal Dan Penelitian Teknik Informatika (Sinkron)*, vol. 9, no. 3, pp. 1167–1776, Aug. 2025, doi: 10.33395/sinkron.v9i3.14778.
4. I. A. W. Nandita, I. M. A. Dwi Suarjaya, and I. P. A. Bayupati, "News Recommendation System Using Content-Based Filtering through RSS Customization Service," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, pp. 1858–1867, Aug. 2025, doi: 10.30871/jaic.v9i4.9807.
5. C. Salsabilla and D. W. Utomo, "Pekalongan Regency Tourism Recommendation System with Content based Filtering," *Jurnal Sistem Informasi (SISTEMASI)*, vol. 14, no. 1, p. 262, Jan. 2025, doi: 10.32520/stmsi.v14i1.4839.

6. I. Yulfihani and M. Zakariyah, "Optimization of Tourism Destination Recommendations in Batang Regency Using Content-Based Filtering," *Journal of Applied Informatics and Computing (JAIC)*, vol. 8, no. 2, pp. 499–508, Nov. 2024, doi: 10.30871/jaic.v8i2.8618.
7. A. H. J. P. Juni Permana and Agung Toto 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/ijoict.v9i2.747.
8. I. Ilhamsyah, S. Rahmayuda, D. M. Midyanti, and S. Martha, "Pemodelan Sistem Rekomendasi Restoran berdasarkan Preferensi Pengguna dengan Pendekatan Content-Based Filtering," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 10, no. 1, p. 154, Apr. 2024, doi: 10.26418/jp.v10i1.74008.
9. S. R. Cholil, N. A. Rizki, and T. F. Hanifah, "Sistem Rekomendasi Tempat Wisata Di Kota Semarang Menggunakan Metode Collaborative Filtering," *Jurnal Informatika dan Komputer (JIKO)*, vol. 7, no. 1, p. 118, Feb. 2023, doi: 10.26798/jiko.v7i1.727.
10. F. Maulana and E. B. Setiawan, "Performance of Deep Feed-Forward Neural Network Algorithm Based on Content-Based Filtering Approach," *Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi (INTENSIF)*, vol. 8, no. 2, pp. 278–294, Aug. 2024, doi: 10.29407/intensif.v8i2.22904.