

IMPLEMENTASI CONTENT-BASED FILTERING PADA PERANCANGAN SISTEM REKOMENDASI OBAT BERBASIS GEJALA PENYAKIT

Yogi Putra Auditama^{1*}, Agustina Purwatiningsih², Dwi Hartanti³

¹Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

^{1*}220103170@mhs.udb.ac.id

²Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

²agustinapurwa@udb.ac.id

³Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

³dwihartanti@udb.ac.id

Abstrak— Banyaknya informasi kesehatan di internet sering kali membuat masyarakat kesulitan memilih obat yang tepat untuk keluhan ringan, sehingga memicu risiko kesalahan swamedikasi. Penelitian ini bertujuan merancang sistem rekomendasi obat berbasis web menggunakan metode Content-Based Filtering. Proses pencocokan teks dilakukan melalui pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) dan kalkulasi Cosine Similarity untuk mengukur kemiripan antara input gejala pengguna dengan deskripsi obat. Pengujian performa sistem dilakukan secara formal melalui lima skenario variasi input gejala keluhan ringan menggunakan Confusion Matrix. Hasil evaluasi menunjukkan sistem memperoleh rata-rata Precision sebesar 65%, Recall sebesar 100%, dan F1-Score sebesar 77%. Nilai Recall sebesar 100% menunjukkan bahwa seluruh obat yang relevan pada data uji berhasil direkomendasikan oleh sistem, sedangkan nilai Precision sebesar 65% menunjukkan masih terdapat beberapa rekomendasi yang kurang relevan. Secara keseluruhan, sistem mampu menghasilkan rekomendasi obat yang relevan untuk gejala penyakit ringan berdasarkan kemiripan deskripsi obat.

Kata kunci— Sistem Rekomendasi Obat, Content-Based Filtering, TF-IDF, Cosine Similarity.

Abstract— The increasing amount of health information available on the internet often makes it difficult for people to select appropriate medicines for mild health complaints, thereby increasing the risk of inappropriate self-medication. This study aims to develop a web-based medicine recommendation system using the Content-Based Filtering method. The text matching process employs Term Frequency–Inverse Document Frequency (TF-IDF) weighting and Cosine Similarity to measure the similarity between user-input symptoms and medicine descriptions. System performance was evaluated using the Confusion Matrix based on twenty symptom-input scenarios representing common mild health complaints. The evaluation results showed that the proposed system achieved an average Precision of 65%, Recall of 100%, and F1-Score of 77%. A Recall of 100% indicates that all relevant medicines in the test data were successfully retrieved by the system, while a Precision of 65% indicates that some recommended medicines were not fully relevant to the input symptoms. Overall, the proposed approach is capable of providing relevant medicine recommendations for common mild symptoms based on the similarity between user symptoms and medicine descriptions.

Keywords— Drug Recommendation System, Content-Based Filtering, TF-IDF, Cosine Similarity.

I. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan yang besar dalam berbagai bidang, termasuk bidang kesehatan dan pelayanan apotek. Saat ini masyarakat semakin mudah memperoleh informasi mengenai penyakit maupun obat melalui internet dan aplikasi digital. Namun, banyaknya informasi yang tersedia berpotensi menyebabkan masyarakat kesulitan dalam memilih obat yang relevan sesuai gejala yang dialami. Kurangnya kesadaran masyarakat terhadap penggunaan obat yang sesuai indikasi dapat menyebabkan kesalahan dalam memilih obat sehingga pengobatan menjadi kurang efektif [1].

Apotek merupakan tempat pelayanan kefarmasian yang menyediakan beragam jenis

obat untuk membantu penanganan keluhan kondisi medis ringan seperti demam, batuk, pilek, sakit kepala, mual, sakit lambung, dan alergi. Dalam praktiknya, masyarakat sering membeli obat berdasarkan pengalaman pribadi, saran orang lain, atau informasi dari media sosial tanpa mengetahui apakah obat tersebut benar-benar sesuai dengan gejala yang dirasakan. Kondisi ini menunjukkan perlunya suatu sistem pencarian pintar yang dapat memberikan informasi ke masyarakat dalam mencari rekomendasi obat yang sesuai, aman, dan mudah dipahami [1]. Beberapa pendekatan komputasi sebelumnya telah menerapkan metode *Forward Chaining* dalam membangun sistem untuk merekomendasikan obat berlandaskan gejala penyakit menular umum [2]. Namun, sistem

pakar berbasis aturan (*rule-based*) memiliki keterbatasan dalam fleksibilitas pengolahan teks deskripsi obat yang dinamis [2].

Sebagai solusi alternatif, teknologi sistem rekomendasi dimanfaatkan untuk membantu pengguna dalam mencari informasi secara personal sesuai dengan kebutuhan mereka [3], [4]. Sistem rekomendasi telah banyak diimplementasikan pada berbagai bidang karena kemampuannya dalam memberikan saran berbasis karakteristik data tertentu, seperti pada sistem yang merekomendasikan buku yang menerapkan metode *K-Nearest Neighbor* (*K-NN*) [5], [6] maupun aplikasi kursus online berbasis web [7]. Pada domain kesehatan, sistem rekomendasi dapat dioptimalkan untuk memberikan saran obat berdasarkan keluhan gejala yang dimasukkan oleh pengguna dengan cara mencocokkannya ke dalam basis data [1].

Metode yang optimal untuk mengolah data tekstual adalah *Content-Based Filtering* (*CBF*). Metode ini bekerja dengan cara menganalisis dan mencocokkan kemiripan konten antara data masukan pengguna dengan fitur atau deskripsi yang melekat pada setiap dokumen item [8], [9], [10]. Dalam pengujian komparatif, metode *CBF* sering kali dibandingkan dengan metode *Collaborative Filtering* guna mengukur efisiensinya dengan parameter evaluasi *Precision*, *Recall*, dan *F1-Score* [11]. Di samping itu, metode berbasis konten ini juga sukses dikombinasikan dengan algoritma lain seperti *Multilevel Feedback Queue* pada aplikasi pemesanan makanan [12], serta algoritma *Haversine Formula* untuk rekomendasi hotel [5].

Pada sistem rekomendasi berbasis teks, proses pencocokan konten dapat dioptimalkan secara presisi menggunakan metode untuk memberikan bobot antar kata/kalimat menggunakan *Term Frequency-Inverse Document Frequency* (*TF-IDF*) dan pengukuran kemiripan tiap vektor memanfaatkan *Cosine Similarity* [13]. Metode *TF-IDF* diterapkan untuk menilai setiap kata berdasarkan tingkat kepentingannya dalam dokumen, sementara *Cosine Similarity* bertugas menghitung sudut kemiripan vektor antar-dokumen tersebut [14]. Kombinasi kedua metode ini terbukti sangat kokoh untuk mengukur tingkat

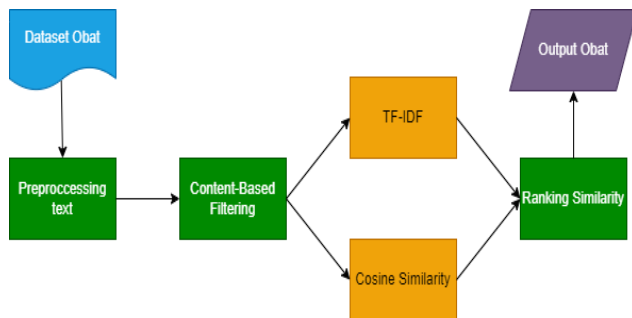
kedekatan nilai vektor *similarity* pada artikel berita maupun deskripsi produk teks lainnya [15].

Meskipun penelitian sebelumnya telah menggunakan metode *Forward Chaining* pada sistem rekomendasi obat, metode tersebut masih bergantung pada aturan (*rule-based*) yang telah ditentukan sehingga kurang fleksibel dalam menangani variasi data teks yang dinamis. Selain itu, penelitian sebelumnya belum menerapkan pembobotan kata menggunakan *TF-IDF*. Oleh karena itu, penelitian ini mengusulkan metode *Content-Based Filtering* (*CBF*) yang dikombinasikan dengan *TF-IDF* dan *Cosine Similarity* untuk menghasilkan rekomendasi obat yang lebih adaptif dan relevan.

Tujuan dari penelitian ini adalah untuk merancang dan membangun sistem rekomendasi obat berbasis web yang dikembangkan dengan metode *CBF* dengan representasi teks berbasis *TF-IDF* dan pengukuran kemiripan antara 2 vektor dengan *Cosine Similarity*. Diharapkan sistem ini dapat membantu masyarakat dalam memperoleh rekomendasi obat secara cepat, mudah, dan praktis berdasarkan keluhan gejala ringan sehari-hari tanpa harus mencari secara manual. Selain memberikan rekomendasi obat, sistem ini juga menyediakan informasi tambahan seperti kategori obat, deskripsi singkat, tingkat persentase kecocokan hasil rekomendasi, serta estimasi harga obat guna meningkatkan efektivitas pencarian informasi kesehatan di masyarakat.

II. METODOLOGI PENELITIAN

Pada penelitian ini memiliki tahap-tahapan yang terdiri dari pengumpulan dan pembersihan data obat, pemberian bobot antar kata dengan *TF-IDF* dan perhitungan sudut antar vektor dengan *cosine similarity* kemudian menggunakan ranking similarity untuk menghasilkan rekomendasi obat yang sesuai berdasarkan gejala pengguna. Adapun alur dari tahapan penelitian disajikan pada gambar 1.



Gambar 1. Tahapan Sistem Rekomendasi

A. Dataset Obat

Pengumpulan data obat dilakukan dengan meninjau katalog obat dan informasi terkait, meliputi nama obat, indikasi, kategori, deskripsi, serta informasi pendukung lainnya. Data yang diperoleh kemudian didokumentasikan dan disimpan ke dalam basis data untuk digunakan dalam proses rekomendasi. Dataset yang digunakan pada penelitian ini terdiri dari 100 data obat yang dikelompokkan ke dalam 21 kategori penyakit/jenis obat. Sumber data diperoleh dari katalog obat apotek mandjur.

Contoh dataset obat yang diimplementasikan disajikan pada Tabel 1.

Tabel 1. Dataset obat, deskripsi, dan kategori

Nama Obat	Deskripsi	Kategori
OBH Combi	Batuk berdahak flu	Batuk
Paratusin	Batuk flu pilek	Batuk
Flucadex	Flu demam batuk pilek	Flu
Antasida	Maag mual lambung	Lambung
Cetirizine	Alergi gatal bersin	Alergi gatal bersin

B. Preprocessing Text

Tahapan sistem rekomendasi obat berbasis CBF dimulai dari proses input gejala oleh pengguna, kemudian dilakukan *preprocessing text* untuk membersihkan data. Setelah itu proses pembobotan tiap kata menggunakan metode *TF-IDF* dan Perhitungan tingkat kemiripan antar vektor dilakukan menggunakan metode *Cosine Similarity*. Hasil *similarity* kemudian diurutkan sehingga sistem dapat menampilkan rekomendasi obat dengan tingkat kecocokan tertinggi. Tahap *preprocessing text* dilakukan untuk membersihkan dan menormalisasi data teks

sebelum dilakukan proses pembobotan menggunakan metode *TF-IDF*. Sistem dikembangkan menggunakan bahasa pemrograman *Python* dengan memanfaatkan beberapa fungsi bawaan dan *library* pendukung. Tahapan *preprocessing* yang digunakan meliputi:

1. Case Folding

Tahap *case folding* bertujuan menormalisasi teks dengan mengonversi huruf dari inputan gejala seluruh huruf menjadi huruf kecil (*lowercase*) menggunakan fungsi bawaan *Python* *lower()*.

Contoh: Batuk Pilek → batuk pilek

2. Tokenizing

Pada tahap ini, kalimat atau teks diuraikan menjadi kumpulan kata tunggal yang merepresentasikan setiap elemen kata dalam dokumen menggunakan fungsi *explode()* dan *str_word_count()*. Contoh: batuk pilek demam → [batuk, pilek, demam]

3. Stopword Removal

Tahap *Stopword removal* digunakan untuk menghilangkan kata yang tidak memiliki makna penting seperti “dan”, “yang”, “di”, dan “ke” dengan fungsi *array_filter()*.

4. Stemming

Tahap *stemming* digunakan untuk mengubah kata menjadi bentuk dasar menggunakan *library* Sastrawi.

Contoh: berdahak → dahak

C. Content-Based Filtering

Content-Based Filtering (CBF) adalah metode yang bekerja dalam suatu sistem rekomendasi dengan cara memanfaatkan keunikan, fitur, dan atribut sebuah item untuk disesuaikan dengan profil dan kebutuhan pengguna. Dalam penerapannya metode ini berfokus pada analisis konten teks dari deskripsi produk atau dokumen.

Pada penelitian ini, metode CBF digunakan untuk menjembatani dan mencocokkan secara literal kata kunci gejala penyakit ringan yang dimasukkan oleh pengguna dengan data teks deskripsi (seperti indikasi, komposisi, dan

kategori) yang melekat pada setiap dokumen obat di dalam basis data. Langkah-langkah komputasi dalam algoritma *CBF* pada sistem ini secara umum dibagi menjadi tiga tahapan utama:

D. Metode TF-IDF

Metode *Term Frequency-Inverse Document Frequency* (*TF-IDF*) digunakan untuk memberikan bobot pada setiap kata dalam dokumen berdasarkan tingkat kepentingannya. *TF* mengukur frekuensi kemunculan kata pada dokumen, sedangkan *IDF* mengurangi bobot kata yang sering muncul pada seluruh dokumen.

Rumus *TF-IDF* ditunjukkan pada Persamaan (1).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Keterangan:

$TF(t, d)$: Banyaknya kemunculan *term* pada dokumen tertentu.

$IDF(t)$: Nilai yang merepresentasikan tingkat kemunculan *term* dalam keseluruhan dokumen.

Sedangkan rumus *IDF* ditunjukan pada persamaan (2).

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

Keterangan:

N : Total dari seluruh dokumen.

$Df(t)$: Banyaknya dokumen yang memuat seluruh dokumen *term t*.

E. Cosine Similarity

Metode *Cosine Similarity* digunakan untuk mengukur tingkat kesamaan antara gejala pengguna dan data deskripsi obat yang tersimpan dalam basis data. Metode ini menghitung sudut kosinus antar dua vektor sehingga mampu menentukan tingkat kedekatan antar dokumen.

Rumus *Cosine Similarity* ditunjukkan pada Persamaan (3).

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (3)$$

Keterangan:

A = Vektor gejala pengguna yang dibobot menggunakan metode *TF-IDF*.

B = Representasi vektor *TF-IDF* dari deskripsi obat.

$A \cdot B$ = Dot produk antar vektor.

$\|A\| \cdot \|B\|$ = Panjang vektor.

Nilai similarity memiliki rentang dari nilai 0 sampai dengan 1. Apabila nilai mendekati 1, tingkat kesesuaian antara gejala dan obat semakin tinggi.

F. Ranking Similarity

Ranking Similarity merupakan tahap pengurutan hasil rekomendasi berdasarkan nilai kemiripan (*similarity score*) yang diperoleh dari perhitungan *Cosine Similarity*. Pada tahap ini, seluruh data obat yang telah dibandingkan dengan gejala pengguna akan diurutkan dari nilai kemiripan tertinggi hingga terendah. Semakin besar angka similarity yang didapat suatu obat, maka semakin besar tingkat kecocokan antara deskripsi obat dengan gejala yang dimasukkan pengguna. Oleh karena itu, obat dengan nilai *similarity* tertinggi ditempatkan pada peringkat pertama dan direkomendasikan sebagai pilihan utama bagi pengguna, sedangkan obat dengan nilai similarity yang lebih rendah akan berada pada peringkat berikutnya sesuai tingkat relevansinya terhadap gejala penyakit yang dicari.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Hasil penelitian pada sistem rekomendasi obat berbasis gejala penyakit berupa aplikasi web yang mampu memberikan rekomendasi obat berdasarkan gejala yang dimasukkan oleh pengguna. sistem rekomendasi obat berbasis web yang dikembangkan dengan metode *CBF* dengan representasi teks berbasis *TF-IDF* dan pengukuran kemiripan dengan *Cosine Similarity* yang berfungsi mengidentifikasi tingkat kesamaan tiap gejala yang dimasukkan pengguna dan data obat dalam basis data. Penjelasan konteks penelitian mengikuti dokumen yang telah disusun sebelumnya.

Sistem memiliki dua fitur utama yaitu pencarian berdasarkan gejala penyakit dan

penelitian berdasarkan nama obat. Pengguna dapat memasukkan gejala seperti batuk, pilek, demam, sakit kepala, mual, maag, alergi, dan gejala ringan lainnya. Setelah input dimasukkan, sistem melakukan *preprocessing* teks sebelum dilakukan proses pembobotan kata. Contoh input gejala pada penelitian: “batuk pilek demam”.

Untuk menjelaskan proses pembobotan *TF-IDF* digunakan contoh *query* pengguna berupa “batuk pilek demam”. Setelah melalui tahap *preprocessing*, diperoleh tiga kata yaitu batuk, pilek, dan demam.

Tahap pertama dilakukan perhitungan TF, yaitu jumlah kemunculan kata pada dokumen yang dijabarkan pada tabel 2

Tabel 2. Nilai *TF* pada *Query* "batuk pilek demam"

Term	Frekuensi	TF
batuk	1	1
Pilek	1	1
demam	1	1

Selanjutnya dilakukan perhitungan *IDF* berdasarkan jumlah dokumen yang mengandung kata tertentu pada keseluruhan data obat. Berikut merupakan contoh perhitungan menggunakan 10 dataset agar proses komputasi lebih mudah dipahami.

Jumlah seluruh dokumen (n): 10

Df(t) batuk : 4

Df(t) pilek : 5

Df(t) demam : 4

Maka

Untuk kata batuk:

$$IDF(batuk) = \log \log \left(\frac{10}{4} \right) = 0,398$$

Untuk kata pilek:

$$IDF(pilek) = \log \log \left(\frac{10}{5} \right) = 0,301$$

Untuk kata demam:

$$IDF(demam) = \log \log \left(\frac{10}{4} \right) = 0,398$$

Tahap berikutnya dilakukan perhitungan TF-IDF menggunakan Persamaan:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

Untuk kata batuk

$$TF - IDF = 1 \times 0,398 = 0,398$$

Untuk kata pilek

$$TF - IDF = 1 \times 0,301 = 0,301$$

Untuk kata demam

$$TF - IDF = 1 \times 0,398 = 0,398$$

Sehingga diperoleh vector *TF-IDF*:

[0.398, 0.301, 0.398]

Untuk nilai vektor yang dihasilkan dari query “batuk pilek demam” menggunakan bahasa pemrograman *Python* dengan library *scikit-learn* melalui *TfidfVectorizer()* yaitu:

[0.597, 0.562, 0.573]

Kemudian contoh dataset obat yang dijadikan acuan ditunjukkan pada Tabel 3.

Tabel 3. Dataset obat

Nama Obat	Deskripsi	Vektor TF-IDF
Paraflu	Pilek demam flu	[0.574, 0.585, 0.574]
Paratusin	Batuk pilek flu	[0.6, 0.565, 0.565]
Flucadex	Flu demam batuk pilek	[0.52, 0.499, 0.49, 0.49]

Setelah proses pembobotan selesai, dilakukan perhitungan menggunakan metode *Cosine Similarity* untuk mengetahui tingkat kesamaan antara input gejala pengguna dan dataset obat. Rumus *Cosine Similarity* ditunjukkan pada Persamaan (4).

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} \quad (4)$$

Berikut kalkulasi *similarity* antara input gejala dan obat “Flucadex”:

$$\frac{(0.597 \times 0.52) + (0.562 \times 0.499) + (0.573 \times 0.49) + (0 \times 0.49)}{\sqrt{(0.597^2 + 0.562^2 + 0.573^2 + 0^2)} \times \sqrt{0.52^2 + 0.499^2 + 0.49^2 + 0.49^2}}$$

Dot Product A.B

$$\begin{aligned} &(0.597 \times 0.52) + (0.562 \times 0.499) + (0.573 \times 0.49) + (0 \times 0.49) \\ &0.31044 + 0.280438 + 0.28077 + 0 \\ &= 0.871648 \end{aligned}$$

Panjang vektor A

$$\begin{aligned} |A| &= \sqrt{0.597^2 + 0.562^2 + 0.573^2 + 0^2} \\ &= 1.0003 \end{aligned}$$

Panjang vektor B

$$\begin{aligned} |B| &= \sqrt{0.52^2 + 0.499^2 + 0.49^2 + 0.49^2} \\ &= 0.9998 \end{aligned}$$

Cosine Similarity

$$\text{Cos}(A, B) = \frac{0.871648}{1.0003 \times 0.9998}$$

Hasil kalkulasi diperoleh nilai *similarity*:

Cosine Similarity=0,8716

Nilai *similarity* kemudian diurutkan dari yang tertinggi sehingga diperoleh output rekomendasi obat seperti pada Tabel 4.

Tabel 4. Output rekomendasi obat

Nama Obat	Hasil Similarity	Persentase
Flucadex	0,8716	87.16%
Paratusin	0,676	67.6%
Paraflu	0,6573	65.73%

Berdasarkan hasil pengujian, Flucadex memperoleh nilai *similarity* sebesar 87,16%, lebih tinggi dibandingkan Paratusin. Hal ini disebabkan karena deskripsi Flucadex mengandung kata yang lebih relevan terhadap gejala pengguna yaitu batuk, pilek, dan demam. Pada proses *TF-IDF*, kata-kata tersebut memiliki bobot yang lebih tinggi sehingga menghasilkan nilai *Cosine Similarity* yang lebih besar. Sebaliknya, pada Paratusin terdapat lebih sedikit kata yang sesuai dengan gejala pengguna sehingga tingkat kemiripan yang diperoleh lebih rendah.

Selain menampilkan rekomendasi obat, sistem juga menyediakan informasi tambahan seperti kategori obat, bentuk obat, dan estimasi harga obat sehingga memudahkan pengguna dalam memahami output rekomendasi yang diberikan. Tampilan output rekomendasi sistem disajikan pada Gambar 2.



Gambar 2. Tampilan Output Rekomendasi

Dan berikut output ranking rekomendasi obat.

Ranking	Nama Obat	Similarity (%)
Top-1	Flucadex	87.18
Top-2	Paratusin	67.6
Top-3	Ifarsyl	67.4
Top-4	Dicom	67.4
Top-5	Inclarin	67.4
Top-6	Bodrexin Flu dan Batuk	67.4
Top-7	Sedares	67.4
Top-8	Ultraflu	65.73
Top-9	Puscold	65.73
Top-10	Intunal	65.73
Top-11	Paraflu	65.73
Top-12	Stop Cold	65.73

Gambar 3. Ranking rekomendasi obat

B. Pembahasan

Metode *CBF* merupakan metode rekomendasi yang bekerja berdasarkan kesamaan karakteristik atau atribut antar item. Dalam penelitian ini, metode bertujuan untuk mencocokkan gejala penyakit yang dimasukkan pengguna dengan data obat yang tersedia pada sistem. Proses rekomendasi dilakukan dengan memanfaatkan deskripsi gejala dan informasi obat. Dengan demikian, sistem mampu menyarankan obat berdasarkan gejala atau keluhan user.

Berdasarkan hasil pengujian, sistem menghasilkan tingkat kemiripan yang berbeda pada setiap obat. Perbedaan nilai tersebut dipengaruhi oleh jumlah kata yang memiliki kesamaan antara gejala pengguna dengan deskripsi obat. Sebagai contoh, pada masukan gejala batuk pilek demam, obat Flucadex memperoleh nilai *similarity* lebih tinggi dibandingkan paratusin karena deskripsi Flucadex mengandung lebih banyak kata yang relevan terhadap gejala pengguna. Kata seperti batuk, pilek, dan demam memiliki bobot *TF-IDF* yang tinggi sehingga memberikan kontribusi lebih besar pada proses perhitungan *Cosine Similarity*. Sebaliknya, paratusin memiliki jumlah kata yang lebih sedikit atau bobot kata yang lebih rendah sehingga menghasilkan nilai kemiripan yang lebih kecil.

Hasil tersebut menunjukkan bahwa metode *Content-Based Filtering* dengan *TF-IDF* dan *Cosine Similarity* mampu menghasilkan rekomendasi berdasarkan tingkat relevansi isi dokumen, bukan hanya berdasarkan pencocokan kata secara langsung. Semakin banyak kata yang memiliki hubungan dengan gejala pengguna,

maka semakin tinggi nilai kemiripan yang dihasilkan.

Penelitian ini juga menunjukkan bahwa kualitas rekomendasi sangat dipengaruhi oleh isi deskripsi obat. Jika deskripsi obat terlalu singkat atau tidak memuat gejala yang relevan, maka nilai *similarity* yang dihasilkan cenderung lebih rendah meskipun obat tersebut sebenarnya sesuai digunakan.

Sistem rekomendasi obat berbasis gejala penyakit ini menunjukkan performa yang baik dalam mencocokkan input keluhan pengguna dengan dataset obat yang tersedia. Keberhasilan sistem dalam memberikan rekomendasi yang relevan dipengaruhi oleh integrasi yang solid antara tahapan text *preprocessing*, Proses pemberian bobot pada kata memanfaatkan metode *TF-IDF*, serta pengukuran tiap antar vektor dengan metode *Cosine Similarity*.

Untuk mengetahui sejauh mana tingkat akurasi sistem, dilakukan pengujian formal menggunakan parameter evaluasi *Precision*, *Recall*, dan *F1-Score*. Pengujian ini dilakukan dengan menguji 20 skenario variasi input gejala keluhan ringan yang paling sering dialami masyarakat. Hasil rekapitulasi uji dipaparkan pada Tabel 5.

Tabel 5. Data Confusion Matrix

Input Gejala	Target relevan	Reko mend asi	TP	FP	FN
Batuk pilek demam	7	12	7	5	0
Demam sakit kepala	8	12	8	5	0
Maag mual lambun g	6	9	6	3	0
Gatal alergi bersin	5	7	5	2	0
Flu batuk dahak	5	11	5	6	0
Pegal nyeri otot	4	7	4	3	0

Batuk berdaha k tenggor okan	2	4	2	2	0
Sariawa n tenggor okan	2	2	2	0	0
Diare sakit perut	3	3	3	0	0
Masuk angin kembun g	3	3	3	0	0
Pilek hidung tersumb at	3	4	3	1	0
Penamb ah darah anemia	3	4	3	1	0
Jamur kulit gatal	6	7	6	1	0
Sakit kepala migrain	1	3	1	2	0
Iritasi mata	3	4	3	1	0
Nyeri sendi tulang	1	3	1	2	0
Vitamin daya tahan tubuh	2	3	2	1	0
Batuk berdaha k kental	2	5	2	3	0
Cacina n sakit perut	1	3	1	2	0
Flu demam hidung tersumb at	3	4	3	1	0

Berdasarkan hasil pengujian pada Tabel 5, performa sistem rekomendasi obat dievaluasi menggunakan metode *Confusion Matrix* yang terdiri atas komponen *True Positive (TP)*, *False Positive (FP)*, dan *False Negative (FN)*. Pengujian dilakukan terhadap 20 skenario input gejala yang mewakili berbagai penyakit ringan, seperti batuk, flu, demam, gangguan pencernaan, alergi, serta penyakit kulit. Hasil pengujian menunjukkan bahwa sistem mampu menemukan seluruh obat yang relevan terhadap setiap input gejala. Hal ini ditunjukkan oleh nilai *FN* yang bernilai 0 pada seluruh skenario pengujian, sehingga tidak terdapat obat yang seharusnya direkomendasikan tetapi gagal ditampilkan oleh sistem.

Meskipun demikian, pada beberapa skenario masih ditemukan nilai *FP*. Hal ini menunjukkan bahwa sistem masih memberikan rekomendasi obat yang memiliki kemiripan kata dengan gejala pengguna, tetapi tidak seluruhnya merupakan obat yang paling relevan. Kondisi tersebut dipengaruhi oleh metode *Content-Based Filtering* yang memanfaatkan pembobotan *TF-IDF* dan *Cosine Similarity*, sehingga rekomendasi didasarkan pada tingkat kesamaan kata antara gejala yang dimasukkan pengguna dan deskripsi obat. Akibatnya, obat yang memiliki kata kunci serupa masih dapat memperoleh nilai kemiripan yang cukup tinggi meskipun indikasi utamanya berbeda.

Sebagai contoh, pada skenario "flu batuk dahak", sistem menghasilkan 11 rekomendasi obat, dengan 5 obat yang sesuai sebagai *TP* dan 6 obat sebagai *FP*. Hal ini terjadi karena beberapa obat lain memiliki deskripsi yang juga mengandung kata flu, batuk, atau dahak, sehingga tetap memperoleh nilai *Cosine Similarity* di atas ambang batas (*threshold*) yang telah ditentukan. Sebaliknya, pada skenario "sariawan tenggorokan", "diare sakit perut", dan "masuk angin kembung", sistem menghasilkan nilai *FP* = 0, sehingga seluruh rekomendasi yang diberikan sesuai dengan data relevan. Hasil tersebut menunjukkan bahwa semakin spesifik kata kunci gejala yang dimasukkan, semakin tinggi kemampuan sistem dalam menghasilkan rekomendasi yang tepat dan semakin kecil

kemungkinan munculnya rekomendasi yang kurang relevan.

Nilai *TP*, *FP*, dan *FN* yang diperoleh dari hasil pengujian belum dapat menggambarkan performa sistem secara menyeluruh apabila hanya dilihat dalam bentuk data mentah. Oleh karena itu, dilakukan perhitungan lebih lanjut menggunakan parameter evaluasi *Precision*, *Recall*, dan *F1-Score*. Matriks tersebut berfungsi mengevaluasi tingkat akurasi dari output dari sistem rekomendasi, kemampuan sistem dalam mengidentifikasi seluruh obat yang relevan, serta keseimbangan performa sistem secara keseluruhan.

Berikut rumus *Precision* ditunjukkan pada persamaan (3.3).

$$Precision = \frac{TP}{TP+FP} \quad (3.3)$$

Berikut rumus *Recall* ditunjukkan pada persamaan (3.4).

$$Recall = \frac{TP}{TP+FN} \quad (3.4)$$

Berikut rumus *F1-score* ditunjukkan pada persamaan (3.5).

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3.5)$$

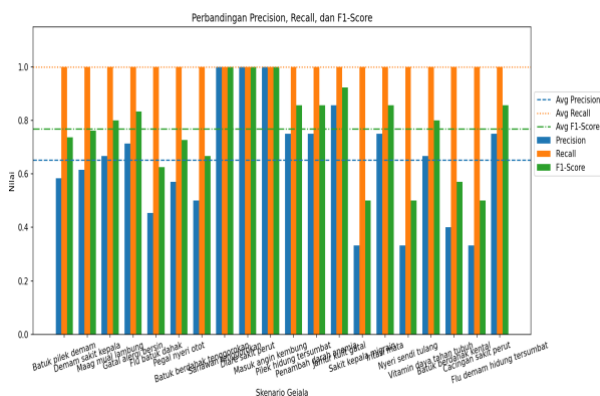
Hasil perhitungan evaluasi tersebut kemudian disajikan pada Tabel 6 sebagai bentuk analisis performa sistem rekomendasi obat berbasis gejala penyakit.

Tabel 6. Hasil Kalkulasi Matrix

input gejala	Precision	Recall	F1-score
Batuk pilek demam	0.58	1.00	0.74
Demam sakit kepala	0.62	1.00	0.76
Maag mual lambung	0.67	1.00	0.80
Gatal alergi bersin	0.71	1.00	0.83
Flu batuk berdahak	0.45	1.00	0.62
Pegal nyeri otot	0.57	1.00	0.73
Batuk berdahak tenggorokan	0.50	1.00	0.67
Sariawan tenggorokan	1.00	1.00	1.00
Diare sakit perut	1.00	1.00	1.00

Masuk angin kembang	1.00	1.00	1.00
Pilek hidung tersumbat	0.75	1.00	0.86
Penambah darah anemia	0.75	1.00	0.86
Jamur kulit gatal	0.86	1.00	0.92
Sakit kepala migrain	0.33	1.00	0.50
Iritasi mata	0.75	1.00	0.86
Nyeri sendi tulang	0.33	1.00	0.50
Vitamin daya tahan tubuh	0.67	1.00	0.80
Batuk berdahak kental	0.40	1.00	0.57
Cacingan sakit perut	0.33	1.00	0.50
Flu demam hidung tersumbat	0.75	1.00	0.86
Rata-rata	0.65	1.00	0.77

Untuk mempermudah analisis dan memetakan perbandingan tren performa dari masing-masing skenario pengujian secara visual, data rekapitulasi pada Tabel 6 tersebut direpresentasikan kembali ke dalam bentuk diagram batang. Perbandingan distribusi nilai *Precision*, *Recall*, dan *F1-Score* pada tiap kombinasi input gejala keluhan ringan disajikan dalam bentuk grafik perbandingan pada Gambar 4.



Gambar 4. Grafik Perbandingan Precision, Recall, dan F1-Score

Berdasarkan visualisasi pada Gambar 4, hasil evaluasi menunjukkan bahwa nilai *Recall* pada seluruh skenario pengujian mencapai 1,00 (100%), dengan rata-rata *Recall* sebesar 1,00. Hasil tersebut menunjukkan bahwa sistem berhasil menemukan seluruh obat yang termasuk dalam data relevan untuk setiap input gejala. Dengan kata lain, selama proses pengujian tidak terdapat obat yang relevan tetapi gagal direkomendasikan oleh sistem ($FN = 0$). Hal ini menunjukkan bahwa metode *Content-Based Filtering* yang dipadukan dengan *TF-IDF* dan *Cosine Similarity* mampu mengidentifikasi seluruh kandidat obat yang sesuai berdasarkan kemiripan deskripsi obat terhadap gejala yang dimasukkan pengguna.

Sementara itu, nilai *Precision* menunjukkan variasi pada setiap skenario pengujian dengan rata-rata sebesar 0,65. Nilai *Precision* tertinggi sebesar 1,00 diperoleh pada skenario "Sariawan tenggorokan", "Diare sakit perut", dan "Masuk angin kembang". Hasil tersebut menunjukkan bahwa seluruh obat yang direkomendasikan pada ketiga skenario tersebut merupakan obat yang relevan tanpa menghasilkan *FP*. Sebaliknya, nilai *Precision* terendah sebesar 0,33 diperoleh pada skenario "Sakit kepala migrain", "Nyeri sendi tulang", dan "Cacingan sakit perut". Nilai tersebut menunjukkan bahwa masih terdapat beberapa obat yang direkomendasikan meskipun hanya memiliki kemiripan sebagian terhadap gejala yang dimasukkan pengguna. Kondisi ini disebabkan oleh adanya kata-kata yang serupa pada deskripsi beberapa obat sehingga menghasilkan nilai *Cosine Similarity* yang masih berada di atas ambang batas (*threshold*).

Variasi nilai *Precision* turut memengaruhi nilai *F1-Score*, karena metrik ini merupakan kombinasi harmonis antara *Precision* dan *Recall*. Berdasarkan hasil pengujian diperoleh rata-rata *F1-Score* sebesar 0,77, yang menunjukkan bahwa sistem memiliki keseimbangan yang baik antara kemampuan menemukan seluruh obat yang relevan (*Recall*) dan kemampuan memberikan rekomendasi yang tepat (*Precision*). Nilai *F1-Score* tertinggi sebesar 1,00 diperoleh pada skenario yang tidak menghasilkan *FP*, sedangkan nilai terendah sebesar 0,50 diperoleh

pada skenario dengan nilai *Precision* yang rendah.

Meskipun demikian, penggunaan *TF-IDF* serta *Cosine Similarity* terbukti efektif dalam membantu proses pencocokan gejala dan deskripsi obat pada basis data. Selain itu, penelitian ini menerapkan *threshold similarity* sebesar 50% sebagai batas minimum rekomendasi karena mampu memberikan keseimbangan antara jumlah rekomendasi dan tingkat relevansinya. Hal ini karena *threshold* yang lebih rendah berpotensi meningkatkan *FP*, sedangkan *threshold* yang lebih tinggi dapat mengurangi jumlah obat yang masih relevan. Oleh karena itu, *threshold* 50% dipilih agar rekomendasi yang ditampilkan tetap relevan tanpa menghilangkan terlalu banyak kandidat obat.

Secara keseluruhan, hasil pengujian dan evaluasi menunjukkan bahwa sistem rekomendasi obat berbasis gejala penyakit yang dirancang berhasil menampilkan hasil rekomendasi yang relevan dan dapat membantu pengguna dalam memperoleh informasi obat secara cepat. Hasil evaluasi menggunakan *Confusion Matrix Precision*, *Recall*, dan *F1-Score* juga menunjukkan bahwa sistem memiliki performa yang baik sebagai dasar pengembangan sistem rekomendasi obat berbasis web pada penelitian selanjutnya.

IV. KESIMPULAN

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem, dapat disimpulkan bahwa penelitian ini berhasil membangun sistem rekomendasi obat berbasis web menggunakan metode *Content-Based Filtering (CBF)* yang dipadukan dengan proses *preprocessing*, pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)*, serta perhitungan kemiripan menggunakan *Cosine Similarity*. Sistem mampu memberikan rekomendasi obat berdasarkan kemiripan antara gejala yang dimasukkan pengguna dengan deskripsi obat yang tersimpan pada basis data.

Hasil evaluasi menggunakan *Confusion Matrix* menunjukkan bahwa sistem memperoleh rata-rata *Precision* sebesar 0,65, *Recall* sebesar 1,00,

dan *F1-Score* sebesar 0,77. Nilai *Recall* sebesar 1,00 menunjukkan bahwa seluruh obat yang relevan berhasil direkomendasikan oleh sistem, sedangkan nilai *Precision* sebesar 0,65 menunjukkan masih terdapat beberapa rekomendasi yang kurang relevan akibat kemiripan kata pada deskripsi obat. Secara keseluruhan, hasil tersebut menunjukkan bahwa metode yang diterapkan mampu menghasilkan rekomendasi obat yang relevan untuk gejala penyakit ringan.

Penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang digunakan masih relatif kecil, sehingga variasi gejala dan deskripsi obat yang dapat diproses oleh sistem masih terbatas. Kedua, dataset hanya mencakup obat bebas (*over-the-counter*) sehingga hasil rekomendasi belum mencakup obat resep maupun obat untuk penyakit tertentu. Ketiga, sistem hanya memberikan rekomendasi berdasarkan gejala penyakit ringan yang direpresentasikan melalui deskripsi obat. Selain itu, sistem belum mempertimbangkan faktor-faktor lain yang memengaruhi pemilihan obat, seperti usia pasien, dosis penggunaan, riwayat alergi, kondisi kehamilan, maupun penyakit penyerta. Oleh karena itu, hasil rekomendasi yang diberikan bersifat sebagai informasi awal dan tidak dimaksudkan untuk menggantikan konsultasi dengan tenaga kesehatan.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Ibu Agustina Purwatiningsih, S.Kom., M.Cs dan Ibu Dwi Hartanti, S.Kom., M.Kom selaku dosen pembimbing yang telah memberikan bimbingan, arahan, masukan, serta dukungan selama proses penyusunan penelitian ini. Segala saran dan ilmu yang diberikan sangat membantu penulis dalam menyelesaikan penelitian hingga tahap akhir. Penulis juga mengucapkan terima kasih atas waktu, perhatian, dan motivasi yang telah diberikan selama proses penelitian berlangsung.

REFERENSI

- [1] I. Irama Permana and M. Miftahudin, "Penerapan Metode Content-Based Filtering untuk Rekomendasi pada Resep Obat Berdasarkan Diagnosa Pasien," 2025.
- [2] S. Aprilia, R. Agustin, M. Marthalena, V. H. Pranatawijaya, and R. Priskila, "Sistem Pakar Rekomendasi Obat Berdasarkan Gejala Penyakit Menular Umum di Masyarakat menggunakan Metode Forward Chaining," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4258.
- [3] N. Latif *et al.*, "Sistem Informasi Geografis untuk Rekomendasi Destinasi Wisata di Kabupaten Enrekang," 2026.
- [4] T. Ridwansyah, B. Subartini, and S. Sylviani, "Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi," *Mathematical Sciences and Applications Journal*, vol. 4, no. 2, pp. 70–77, Apr. 2024, doi: 10.22437/msa.v4i2.32136.
- [5] A. Muliawan, T. Badriyah, and I. Syarif, "Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula," *Technomedia Journal*, vol. 7, no. 2, pp. 231–247, Sep. 2022, doi: 10.33050/tmj.v7i2.1893.
- [6] H. Dharmawan, S. Shofiah Hilabi, I. Kaniawulan, S. Informasi Fakultas Ilmu Komputer Universitas Buana Perjuangan Karawang Jl Ronggo Waluyo Sirnabaya, T. Timur, and J. Barat, "Sistem Rekomendasi Buku Dengan Metode K-Nearest Neighbor (K-NN) Pada Gramedia," 2023.
- [7] Y. Christian and K. Kelvin, "Rancang Bangun Aplikasi Kursus Online Berbasis Web dengan Sistem Rekomendasi Metode Content-Based Filtering," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 23–36, Jan. 2022, doi: 10.36341/rabit.v7i1.2181.
- [8] S. A. Gunarto, E. S. Honggara, and D. D. Purwanto, "Website Sistem Rekomendasi dengan Content Based Filtering pada Produk Perawatan Kulit," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 3, p. 399, Jul. 2023, doi: 10.26418/justin.v11i3.59049.
- [9] J. M. Azri Saputra, L. M. Huizen, and D. B. Arianto, "Sistem Rekomendasi Film pada Platform Streaming Menggunakan Metode Content-Based Filtering," *Jurnal Transformatika*, vol. 22, no. 1, pp. 10–21, Jul. 2024, doi: 10.26623/transformatika.v22i1.7041.
- [10] N. Azizah and A. F. Rozi, "Sistem Rekomendasi Produk Somethinc Menggunakan Metode Content-based Filtering," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 3, pp. 461–468, Jul. 2024, doi: 10.47233/jteksis.v6i3.1411.
- [11] N. Kalimantan Selatan, J. A. Yani NoKM, B. Hanyar, K. Kertak Hanyar, K. Banjar, and U. Nahdlatul Ulama Kalimantan Selatan, "Perbandingan Keakuratan Sistem Rekomendasi Produk Berbasis Content-Based Filtering Dan Collaborative Filtering Pada E-Commerce Shopee Menggunakan Matrik Precision, Recall Dan F1-Score Siti Aminah," 2024. [Online]. Available: <https://repository.unukase.ac.id/>
- [12] A. D. O. Putri, F. Fauziah, and A. Iskandar, "Algoritma Content-Based Filtering dan Algoritma Multilevel Feedback Queue pada Proses Pemesanan Makanan Berbasis Web," *JIP1 (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 1, pp. 266–280, Feb. 2023, doi: 10.29100/jipi.v8i1.3426.
- [13] V. Meida Hersianty, E. Larasati Amalia, D. Puspitasari, and D. Wahyu Wibowo, "Penerapan Algoritma TF-IDF dan Cosine Similarity dalam Sistem Rekomendasi Lowongan Pekerjaan," 2025.
- [14] R. Ardiansyah, M. Ari Bianto, and B. D. Saputra, "Sistem Rekomendasi Buku Perpustakaan Sekolah menggunakan Metode Content-Based Filtering," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 2, pp. 510–518, Oct. 2023, doi: 10.37859/coscitech.v4i2.5131.
- [15] 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, Dec. 2022, doi: 10.47065/bits.v4i3.2511.