

# Penerapan Metode *Content-Based Filtering* untuk Rekomendasi pada Resep Obat Berdasarkan Diagnosa Pasien

Idham Irama Permana<sup>1\*</sup>, Muhamad Miftahudin<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia

e-mail: idham16ip@gmail.com

\*Corresponding Author

## ABSTRACT

Health is an essential factor in improving the quality of human life. Medication errors frequently occurring in healthcare services can result in serious harm to patients. Therefore, a computational approach is needed to assist healthcare professionals in providing medication prescription recommendations. This study aims to develop a prescription recommendation system using the Content-Based Filtering method to reduce the risk of medication errors by matching patient diagnosis data with relevant drug attributes. Content-Based Filtering is a recommendation method that focuses on the similarity of attributes between items, in this case, diagnoses and prescriptions. The system utilizes the descriptions or features available from patient diagnoses to match them with prescription references. The system employs the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm, which is a method to convert text into numerical representations. This is crucial for calculating the relevance between diagnoses and prescriptions. After calculating TF-IDF, each diagnosis and prescription is represented as a vector. Cosine Similarity is used to measure the angle between these vectors, determining the degree of similarity between patient diagnoses and references. This research produced a prototype system based on the Laravel web framework for the frontend and Python Flask for the backend. The system is capable of accurately providing prescription recommendations based on patient diagnoses. System trials demonstrated an 85% increase in prescription recommendation accuracy, thereby supporting the effectiveness of healthcare services, particularly in community health centers (Puskesmas) and hospitals. This prescription recommendation system proves effective in meeting patient needs, as evidenced by evaluations based on Precision, Recall, and F1 Score metrics. With all these metrics, the defined thresholds were met (80% for Precision and Recall, and 75% for F1 Score).

**Keywords:** Medication Errors, Content-Based Filtering Method, prescription recommendations, Term Frequency-Inverse Document Frequency (TF-IDF) algorithm, Cosine Similarity Algorithm, Evaluation Matrix, Python, Health Information System

## ABSTRAK

Kesehatan merupakan faktor esensial dalam meningkatkan kualitas hidup manusia. Kesalahan medikasi yang sering terjadi dalam pelayanan kesehatan dapat menyebabkan kerugian serius bagi pasien. Oleh karena itu, diperlukan pendekatan komputasi untuk mempermudah tenaga kesehatan dalam memberikan rekomendasi resep obat. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi resep obat berbasis metode Content-Based Filtering untuk mengurangi risiko kesalahan medikasi dengan mencocokkan data diagnosa pasien dengan atribut obat yang relevan. Content-Based Filtering merupakan metode rekomendasi yang berfokus pada kesamaan atribut antar item, dalam hal ini diagnosis dan resep obat. Sistem memanfaatkan deskripsi atau fitur yang tersedia dari diagnosis pasien untuk mencocokkannya dengan referensi resep obat. Sistem ini memanfaatkan algoritma Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode untuk mengubah teks menjadi representasi numerik. Ini penting untuk menghitung relevansi diagnosis dan resep obat, setelah TF-IDF dihitung, setiap diagnosis dan resep obat direpresentasikan sebagai vektor. Cosine Similarity digunakan untuk mengukur sudut antara dua vektor tersebut, sehingga menentukan tingkat kesamaan antara diagnosis pasien dan referensi. Penelitian ini menghasilkan prototipe sistem berbasis web framework Laravel sebagai frontend menggunakan Python Flask sebagai backend yang mampu memberikan rekomendasi obat secara akurat berdasarkan diagnosa pasien. Uji coba sistem menunjukkan peningkatan akurasi rekomendasi obat sebesar 85%, sehingga mendukung efektivitas layanan kesehatan terutama di Puskesmas dan rumah sakit. Sistem rekomendasi resep obat ini menunjukkan efektif dengan kebutuhan pasien, Hal ini dapat dilihat berdasarkan berdasarkan evaluasi Precision, Recall, dan F1 Score. Dengan semua metrik tersebut threshold yang telah ditentukan (80% untuk Precision dan Recall, 75% untuk F1 Score).

**Kata Kunci:** Kessalahan Medikasi, Metode Content-Based Filtering, Term Frequency-Inverse Document Frequency (TF-IDF), Algoritma Cosine Similarity, Evaluation Matix, Python, Sistem Informasi Kesehatan

## A. PENDAHULUAN

### 1. Latar Belakang

Kesehatan merupakan hal yang sangat penting serta krusial bagi semua manusia, karena tanpa kesehatan yang baik maka setiap manusia akan sulit dalam melaksanakan aktivitas sehari-hari. Oleh karena itu, banyak masyarakat yang kini sadar akan kesehatan dan akan mengobati jika terjadi masalah dalam kesehatannya. Dalam perkembangan ilmu teknologi yang semakin pesat, masyarakat kini dapat mencari tahu tentang Kesehatan melalui Website akan tetapi tidak semua pencarian informasi itu sesuai dengan masalah kesehatan karena harus ada pemeriksaan lebih lanjut oleh dokter yang sesuai dengan masalah kesehatannya.

Kesehatan merupakan salah satu faktor kualitas hidup yang mencerminkan pada pemenuhan kebutuhan dasar manusia. Peningkatan bidang kesehatan sangat penting untuk diperhatikan karena bidang ini sangat erat kaitannya dengan pembangunan, khususnya pembangunan yang menyangkut sumber daya manusia. Tanpa adanya kondisi yang sehat maka kualitas sumber daya manusia yang tinggi sulit untuk tercapai. Usaha-usaha meningkatkan kesehatan penduduk Indonesia lebih banyak dikerahkan pada pelayanan kesehatan yang merupakan penanganan orang sakit, atau lebih tepat disebut sebagai pengobatan. Pembangunan kesehatan diarahkan untuk meningkatkan mutu pelayanan kesehatan yang dilaksanakan melalui peningkatan mutu pelayanan rumah sakit, puskesmas, puskesmas pembantu dan lembaga pelayanan kesehatan lainnya (Depkes RI, 1997).

*Medication Error* (Kesalahan Medikasi) merupakan kesalahan yang dilakukan oleh tenaga kesehatan dalam merawat pasien yang sebenarnya dapat dicegah dan beresiko terjadi di berbagai rangkaian layanan kesehatan, termasuk perawatan di rumah sakit. Kesalahan obat dapat terjadi di setiap rangkaian layanan obat dimulai dari tahap *prescribing*, *transcribing* dan *dispensing*. Kesalahan medis khususnya dalam pelayanan pemberian obat merupakan peristiwa yang berbahaya bagi pasien dan dapat dicegah dikarenakan adanya kesalahan dalam pelayanan obat dimulai dari salah satu atau seluruh tahapan pelayanan obat baik *prescribing*, *transcribing* dan *dispensing*. Timbulnya kesalahan tersebut, dapat terjadi dimanapun baik fasilitas kesehatan swasta maupun pemerintah.

Menurut (Resnick & Varian, 1997, hlm. 14) mengemukakan bahwa *Content-Based Filtering* atau *Content-Based Recommender System* merupakan salah satu bentuk pendekatan dari sistem rekomendasi yang bekerja dengan menyamakan atau mencocokkan item yang disukai *user* dengan item lainnya yang mempunyai kesamaan atau kemiripan berdasarkan atribut pada objek atau item-nya dari proses pencocokan kemiripan item yang disukai *user* berupa pencocokan atribut profil user dimana preferensi dan minat user disimpan selanjutnya dicocokkan dengan atribut objek atau item untuk memberikan rekomendasi item baru yang mungkin menarik bagi pengguna.

*TF-IDF* merupakan implementasi dari statistik numerik yang menunjukkan relevansi kata kunci dengan beberapa dokumen tertentu, dengan menyediakan kata kunci yang tersedia, beberapa dokumen tertentu dapat diidentifikasi atau dikategorikan sesuai dengan relevansinya (Qaiser & Ali, 2018). *TF-IDF* merupakan gabungan dari dua kata yang berbeda, yaitu *TF (Term Frequency)* dan *IDF (Inverse Document Frequency)*. *Term Frequency* digunakan untuk mengukur berapa kali suatu term muncul dalam suatu dokumen. Misalnya dalam sebuah dokumen "A7" yang berisi 1000 kata dan terdapat kata "Informasi" dalam dokumen tersebut sebanyak 10 kali. Perlu diketahui bahwa panjang total dokumen dapat bervariasi dari sangat kecil hingga besar, sehingga ada kemungkinan bahwa istilah apapun dapat muncul lebih sering dalam dokumen yang lebih besar dibandingkan dengan dokumen kecil (Qaiser & Ali, 2018).

## 2. Permasalahan

Kejadian *medication error* dapat terjadi pada setiap tahapan pelayanan pemberian obat, baik tahap peresepan (*prescribing process*), penerjemahan resep (*transcribing process*), peracikan hingga penyiapan obat (*dispensing process*), dan pemberian obat pada pasien (*administration process*). Penulisan resep yang baik dan benar, dapat memudahkan petugas farmasi menerjemahkan dan menyiapkan obat.

Kesalahan pengobatan (*medication error*) dapat terjadi di rumah sakit, Puskesmas, Klinik atau instansi kesehatan manapun, salah satunya di Puskesmas. Data kejadian kesalahan medikasi atau *medication error* di Puskesmas dapat dilihat pada Tabel 1.

Uraian permasalahan diatas dapat dikemukakan proses rekomendasi resep obat belum efektif. Adanya peningkatan *medication error*, bila tidak dilakukan perbaikan dapat berisiko menurunnya mutu pelayanan, yang berujung pada rendahnya *patient safety* dan tingkat kepercayaan pasien terhadap Puskesmas.

Dari uraian diatas, dapat ditarik ke dalam poin-poin indikator masalah di antaranya:

- a. Sistem peresepan elektronik belum efektif karena ketidaksesuaian pencatatan resep obat, gangguan teknis dan kesulitan akses pada sistem program.
- b. Beban kerja yang tinggi, gangguan lingkungan seperti interupsi, dan keterbatasan jumlah staf farmasi menyebabkan tidak terlaksananya prosedur standar, seperti *double-checking*.

Tabel 1. Pertanyaan Kuesioner Permasalahan

Presentase total kejadian kesalahan medikasi	Presentase total jenis kesalahan medikasi yang sering terjadi	Presentase total faktor penyebab kesalahan medikasi
Sering terjadi 87,5%	Kesalahan dalam pemberian dosis 31,3%	Beban kerja yang tinggi 57,1%
Kadang-kadang 12,5	Resep yang salah 25%	Sistem pencatatan yang buruk 35,7%

Sumber: Kuisisioner Google Form  
 (<https://forms.gle/A3r4aLSywDftfjy5>)

Permasalahan yang diangkat pada penelitian ini yaitu tulisan dokter yang sulit dibaca menjadi hambatan bagi apoteker maupun tenaga kesehatan untuk menafsirkan resep dengan benar. Hal ini terjadi terutama karena dokter

sering terburu-buru atau terganggu oleh panggilan mendesak saat menuliskan resep. Penggunaan sistem yang peneliti akan diterapkan diharapkan dapat membuat proses kerja menjadi lebih efektif dan efisien terutama banyaknya pasien yang mengeluh karena kurang akuratnya obat yang diresepkan oleh dokter sehingga sangat merugikan masyarakat.

### 3. Tujuan

Memberikan informasi terkait sistem rekomendasi resep obat berdasarkan diagnosa pasien yang lebih tepat dan mengukur tingkat keefektifan sistem rekomendasi pada resep obat berdasarkan diagnosa pasien menggunakan metode *Content-Based Filtering*.

## B. METODE

Metode *Content-Based Filtering* digunakan untuk merekomendasikan item (dalam hal ini resep obat) berdasarkan kemiripan konten antara input pengguna (diagnosis yang diberikan) dengan data yang sudah ada dalam sistem. Dalam konteks ini, *Content-Based Filtering (CBF)* merupakan teknik rekomendasi berbasis konten yang bertujuan untuk memberikan rekomendasi berdasarkan karakteristik dari data yang ada. Dalam kasus ini, metode *CBF* digunakan untuk menganalisis diagnosis pasien dan memberikan rekomendasi resep obat yang sesuai.

Cara Kerja *Content-Based Filtering* meliputi:

- 1) Data diagnosis pasien yang diberikan sebagai teks
- 2) Diagnosis pasien dan dataset diagnosis yang ada diubah menjadi vektor numerik menggunakan metode *TF-IDF*
- 3) Kemiripan antara input diagnosis pasien dengan diagnosis dalam dataset dihitung menggunakan *Cosine Similarity*.
- 4) Diagnosis dengan kemiripan di atas ambang batas (*threshold*) dipilih
- 5) Obat terkait diagnosis yang terpilih direkomendasikan kepada pasien

*TF-IDF* adalah metode penghitungan skor untuk setiap istilah (term) dalam suatu dokumen relatif terhadap seluruh dokumen dalam dataset. Pada kasus ini, istilah-istilah yang dianalisis adalah kata-kata dalam deskripsi diagnosis penyakit. Diagnosis input diubah menjadi vektor numerik menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)*. Vektor ini merepresentasikan diagnosis dalam bentuk angka, dengan masing-masing angka menunjukkan pentingnya suatu istilah (kata) dalam diagnosis tersebut. Pada log yang terlihat, *TF-IDF* menghasilkan vektor yang kebanyakan terdiri dari nol, yang menunjukkan bahwa kata-kata dalam diagnosis input sangat jarang atau tidak relevan dengan kata-kata lain dalam dataset.

*Term Frequency (TF)*: Mengukur frekuensi kemunculan suatu kata dalam satu dokumen (dalam kasus ini, diagnosis pasien). Rumus *TF-IDF*:

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

*Term Frequency (TF)*: Mengukur frekuensi kemunculan suatu kata dalam sebuah dokumen:

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dokumen } d}{\text{Total kata dalam dokumen } d}$$

*Inverse Document Frequency (IDF)*: Mengukur seberapa umum atau jarang suatu kata dalam kumpulan dokumen:

$$IDF(t) = \log \left( \frac{N}{1 + DF(t)} \right)$$

dimana:

$N$  = Total jumlah dokumen.

$DF(t)$  = Jumlah dokumen yang mengandung kata  $t$ .

*Cosine Similarity* digunakan untuk menghitung tingkat kemiripan antara dua vektor berdasarkan sudut antar-vektor tersebut.

Rumus *Cosine Similarity*:

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

dimana:

$A \cdot B$  = *Dot Product* antara dua vektor  $A$  dan  $B$ .

$\|A\|$  = Norma atau panjang vektor  $A$ .

$\|B\|$  = Norma atau panjang vektor  $B$ .

## C. HASIL DAN PEMBAHASAN

### 1. Hasil

Implementasi dan perhitungan langkah-langkah diawali dengan meng-input diagnosis.

"Demam tinggi disertai sakit kepala"

Selanjutnya pada langkah representasi data (*TF-IDF*)

Dataset diagnosis (contoh):

Tabel 2. Tabel Dataset Diagnosis

Diagnosis	TF-IDF Vektor
Demam tinggi	[0.9, 0.1, 0.0]
Sakit kepala	[0.2, 0.8, 0.1]
Demam berdarah	[0.6, 0.0, 0.9]

**TF-IDF input:**

$$TF - IDF(\text{"Demam tinggi disertai sakit kepala"}) = [0.8, 0.3, 0.1]$$

**Penghitungan Cosine Similarity**

Hitung Cosine Similarity antara input diagnosis dan dataset diagnosis:

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

Contoh perhitungan dengan diagnosis "Demam tinggi":

$$\text{Cosine Similarity (A, B)} = \frac{(0,8 \times 0,9) + (0,3 \times 0,1) + (0,1 \times 0,0)}{\sqrt{(0,8^2 + 0,3^2 + 0,1^2)} \times \sqrt{(0,9^2 + 0,1^2 + 0,0^2)}}$$

Hasil:

$$\text{Cosine Similarity} = 0,85$$

**Seleksi Diagnosis**

Diagnosis dengan nilai *Cosine Similarity* di atas *threshold* 0,85 dipilih:

1. "Demam tinggi" (0.85)
2. "Sakit kepala" (0.57)

**Rekomendasi Obat**

Setelah diagnosis relevan ditemukan, resep obat direkomendasikan:

Tabel 3. Tabel Resep Obat Direkomendasi

Diagnosis	Resep Obat
Demam tinggi	Paracetamol
Sakit kepala	Ibuprofen

Rekomendasi akhir:

"Paracetamol, Ibuprofen"

Jadi, Metode *TF-IDF* merepresentasikan teks diagnosis menjadi vektor numerik, sementara *Cosine Similarity* menentukan tingkat kemiripan antara input dan dataset diagnosis. *Threshold Cosine Similarity* memastikan hanya diagnosis relevan yang dipilih. Rekomendasi obat diambil berdasarkan diagnosis relevan dalam dataset, memastikan hasil yang akurat dan relevan. Metode *Content-Based Filtering* berfokus pada pencocokan atribut deskriptif antara diagnosa pasien dan deskripsi obat. Dengan memanfaatkan algoritma *TF-IDF* dan *Cosine Similarity*, sistem dapat memberikan rekomendasi obat yang relevan, akurat, dan efisien, sehingga mengurangi risiko kesalahan medikasi. Metode *Content-Based Filtering*, dengan dukungan algoritma *TF-IDF* dan *Cosine Similarity* juga merupakan pendekatan yang sesuai untuk meningkatkan akurasi dan efisiensi dalam rekomendasi resep obat berdasarkan diagnosa pasien. Metode ini memberikan dasar yang kuat untuk pengembangan prototipe sistem berbasis web dengan fokus pada solusi teknologi informasi di bidang kesehatan. Untuk diagnosis input diubah menjadi vektor *TF-IDF* berdasarkan frekuensi kemunculan kata, selanjutnya kemiripan antara diagnosis input dan diagnosis lain dihitung menggunakan *Cosine Similarity* dan tahap terakhir diagnosis dengan kemiripan tertinggi dipilih, dan obat

yang terkait dengan diagnosis tersebut direkomendasikan. Dengan proses ini, sistem dapat memberikan rekomendasi obat berdasarkan kemiripan teks diagnosis pasien dengan diagnosis lain dalam dataset.

## 2. Pembahasan

Hasil pengujian menggunakan *Evaluation Matrix* dilakukan digunakan untuk serangkaian pengukuran yang digunakan untuk menilai kinerja model rekomendasi atau model klasifikasi. Dalam konteks sistem rekomendasi resep obat berbasis Content-Based Filtering, tiga metrik evaluasi utama yang sering digunakan adalah Precision, Recall, dan F1 Score. Uji hasil efektivitas dari sebuah model klasifikasi dapat diukur menggunakan metrik dari evaluation matrix. Berikut merupakan data pada tabel 4. yang berisi data keseluruhan pasien:

Tabel 4. Data Keseluruhan Pasien

No	Nama	Gender	Umur	Diagnosa	Resep Obat
1	Ari Wibowo	Laki-laki	48	Antidiarrea (Opsional & Harus dengan Hati-hati)	Loperamide, Bismuth subsalicylate
2	Kartika Putri Saraswati	Perempuan	42	Kanker Payudara Triple-Negatif (Tidak Respon Terhadap Estrogen, Progesteron, atau HER2)	Carboplatin
3	Lestari Ananda	Perempuan	27	Gagal Jantung	Enalapril, Lisinopril, Ramipril
4	Pratama Putra	Laki-laki	28	Malaria berat/komplikasi	Artesunate intravena/intramuskular, Quinidine intravena
5	Rahayu Setyawati	Perempuan	23	Strongyloides stercoralis (Infeksi cacing benang)	Ivermectin, Albendazole
6	Pratama Putra	Laki-laki	28	Tifus	Ciprofloxacin, Azithromycin, Ceftriaxone
7	Hendra Saputra	Laki-laki	37	Anemia Defisiensi Vitamin B12	Vitamin B12 injeksi, Vitamin B12 pil oral
8	Yudi Hartono	Laki-laki	44	Antidiarrea (Opsional & Harus dengan Hati-hati)	Loperamide, Bismuth subsalicylate
9	Hendra Lesmana	Laki-laki	23	Diare Antibiotik (hanya jika infeksi bakteri terkonfirmasi)	Ciprofloxacin, Azithromycin, Metronidazole
10	Jajang Mulyana	Laki-laki	66	Influenza	Oseltamivir (Tamiflu), Zanamivir (Relenza), Baloxavir marboxil
11	Budi Santoso	Laki-laki	45	Bronkitis Kronis (berhubungan dengan Penyakit Paru Obstruktif Kronik atau PPOK)	albuterol/ipratropium, budesonide/fluticasone, Azitromisin
12	Fitri Ayu	Perempuan	26	Skizofrenia	Risperidone (Risperdal), Olanzapine (Zyprexa), Quetiapine (Seroquel)
13	Sari Mukti Hasanah	Perempuan	43	Anemia Defisiensi Asam Folat (Vitamin B9)	Suplemen asam folat
14	Fadel Alfarisi	Laki-laki	24	Kanker Payudara Stadium Awal (Stadium 0-1)	Lumpektomi atau Mastektomi, Terapi radiasi, Tamoxifen
15	Ahmad Rivaldy	Laki-laki	24	Anemia Defisiensi Vitamin B12	Vitamin B12 injeksi, Vitamin B12 pil oral

No	Nama	Gender	Umur	Diagnosa	Resep Obat
16	Farhan Afliansyah	Laki-laki	43	Tukak Lambung	Omeprazole (Prilosec), Lansoprazole (Prevacid), Pantoprazole (Protonix)
17	Muhammad Husein	Laki-laki	19	Bronkitis Bakteri	penisilin, azithromycin
18	Sri Utami	Perempuan	39	Anemia Defisiensi Asam Folat (Vitamin B9)	Suplemen asam folat
19	Jajang Mulyana	Laki-laki	66	Antidiarrea (Opsional & Harus dengan Hati-hati)	Loperamide, Bismuth subsalicylate
20	Joko Prasetyo	Laki-laki	50	Gagal Jantung	Enalapril, Lisinopril, Ramipril
21	Mukti Ranuwidjaya	Laki-laki	32	Malaria Plasmodium falciparum tanpa komplikasi	Artemether-Lumefantrine (Coartem), Dihydroartemisinin- Piperaquine
22	Sissy Nasution	Perempuan	17	Demam Berdarah Dengue (DBD)	Istirahat total, Paracetamol, Oralit/cairan elektrolit
23	Andi Wijaya	Laki-laki	28	Cacing Gelang (Ascaris lumbricoides)	Albendazole, Mebendazole, Pyrantel pamoate
24	Atep Permana	Laki-laki	29	Cacing Pita (Taenia spp.)	Praziquantel, Niclosamide
25	Arga Falih	Laki-laki	23	Cacing Tambang (Ancylostoma duodenale/Necator americanus)	Albendazole
27	Putri Kurniasari	Perempuan	42	Kanker Serviks Pra- Kanker atau Stadium Awal (Stadium 0 dan IA)	Krioterapi atau terapi laser, LEEP (Loop Electrosurgical Excision Procedure) atau konisasi (eksisi kerucut), Lumpektomi atau Mastektomi
28	Rina Mulyani	Perempuan	52	Faringitis Bakteri (akibat Streptococcus)	Cefuroxime/Azithromycin, Penisilin V/Amoksisilin
29	Sari Mukti Hasanah	Perempuan	43	Bronkitis Akut	Paracetamol/Ibuprofen, pseudoefedrin, loratadine atau cetirizine
30	Sissy Nasution	Perempuan	17	Cacing Kremi (Enterobius vermicularis)	Albendazole, Mebendazole, Pyrantel pamoate
31	Siti Badriah	Perempuan	33	Faringitis Akibat Virus	Paracetamol/Ibuprofen
32	Fitri Ayu	Perempuan	26	Tifus	Ciprofloxacin, Azithromycin, Ceftriaxone
33	Khairunnisa Rizky	Perempuan	31	Asma Persisten Sedang hingga Berat	Budesonide/Fluticasone, Budesonide-Formoterol
34	Putri Salsabila	Perempuan	24	Pneumonia Aspirasi (disebabkan masuknya makanan, minuman, atau cairan tubuh ke paru-paru)	Clindamycin/Piperacillin- tazobactam, Terapi Oksigen, Drainase Postural
35	Larasati Sanusi	Perempuan	33	Malaria Plasmodium falciparum tanpa komplikasi	Artemether-Lumefantrine (Coartem), Dihydroartemisinin- Piperaquine

No	Nama	Gender	Umur	Diagnosa	Resep Obat
36	Maya Sari	Perempuan	40	Cacing Gelang (Ascaris lumbricoides)	Albendazole, Mebendazole, Pyrantel pamoate
37	Atep Permana	Laki-laki	29	Malaria Plasmodium vivax/Plasmodium ovale	Chloroquine, Primaquine
38	Siti Aminah	Perempuan	34	Gagal Jantung	Enalapril, Lisinopril, Ramipril
39	Muhammad Husein	Laki-laki	19	Anemia Defisiensi Vitamin B12	Vitamin B12 injeksi, Vitamin B12 pil oral
40	Kartika Putri Saraswati	Perempuan	42	Bronkitis Kronis (berhubungan dengan Penyakit Paru Obstruktif Kronik atau PPOK)	albuterol/ipratropium, Azitromisin
41	Widia Pratama Putra	Laki-laki	22	Malaria Plasmodium vivax/Plasmodium ovale	Chloroquine
42	Atep Permana	Laki-laki	29	Malaria berat/komplikasi	Artesunate intravena/intramuskular, Quinidine intravena
43	Azril Gymnastiar	Laki-laki	20	Anemia Defisiensi Zat Besi	ferrous sulfate, ferrous gluconate, ferrous fumarate
44	Ari Wibowo	Laki-laki	48	Malaria pada kehamilan	Quinine, Clindamycin, Artemether-Lumefantrine
45	Arga Falih	Laki-laki	23	Pneumonia Jamur	Amfoterisin B
46	Putri Kurniasari	Perempuan	42	Kanker Payudara Stadium 4 (Metastatik atau Kanker yang Menyebar ke Organ Lain)	paclitaxel, capecitabine
47	Saepul	Laki-laki	48	Skizofrenia	Risperidone (Risperdal), Quetiapine (Seroquel)

Pada Metode *Content-Based Filtering* merupakan metode rekomendasi yang berfokus pada kesamaan atribut antar item, dalam hal ini diagnosis dan resep obat. Sistem memanfaatkan deskripsi atau fitur yang tersedia dari diagnosis pasien untuk mencocokkannya dengan referensi resep obat. langkah-langkah dalam *Content-Based Filtering* yaitu:

- a. Persiapan data, pada Dataset pasien berisi informasi tentang pasien, seperti nama, gender, diagnosis dan resep obat yang diberikan. Selanjutnya dataset referensi berisi pasangan diagnosis dan resep obat yang umum digunakan.
- b. *Preprocessing*, semua teks diubah menjadi huruf kecil untuk menghindari *mismatch* akibat perbedaan huruf besar dan huruf kecil. dan pada diagnosa serta resep obat dipecah menjadi format token untuk mempermudah analisis.
- c. Representasi Diagnosis dan Resep Obat, pada diagnosis pasien direpresentasikan sebagai vektor fitur berdasarkan kata-kata yang muncul. dan selanjutnya pada setiap resep obat juga direpresentasikan dalam bentuk vektor yang sama, menggunakan pendekatan *TF-IDF*.
- d. Pencocokan dan Pemberian Rekomendasi, Dengan menggunakan *Cosine Similarity*, diagnosis pasien dibandingkan dengan diagnosis pada referensi untuk menentukan kemiripan dan pada resep obat dengan kemiripan tertinggi direkomendasikan kepada pasien.

Selanjutnya pada metode *TF-IDF* (Term Frequency-Inverse Document Frequency) adalah metode untuk mengubah teks menjadi representasi numerik. Ini penting untuk menghitung relevansi diagnosis dan resep obat. Langkah-langkah perhitungan pada metode *TF-IDF* yaitu:

- a. *Term Frequency (TF)* untuk mengukur seberapa sering kata *t* muncul dalam dokumen *d*, dan rumusnya:

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Total kata dalam dokumen } d}$$

- b. *Inverse Document Frequency (IDF)* yaitu untuk mengukur seberapa unik suatu kata di seluruh dokumen dimana  $N$  adalah jumlah total dokumen, dan rumusnya:

$$IDF(t) = \log \frac{N}{1 + \text{Jumlah dokumen yang mengandung } t}$$

- c. *TF-IDF Score*, yaitu menggabungkan *TF* dan *IDF* untuk mendapatkan skor relevansi setiap kata dalam dokumen, dan rumusnya:

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

Setelah *TF-IDF* dihitung, setiap diagnosis dan resep obat direpresentasikan sebagai vektor. *Cosine Similarity* digunakan untuk mengukur sudut antara dua vektor tersebut, sehingga menentukan tingkat kesamaan antara diagnosis pasien dan referensi. Berikut adalah rumus dari *Cosine Similarity*.

$$Cosine\ Similarity(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

yang artinya Nilai *Cosine Similarity* berkisar antara 0 hingga 1.

nilai 1 yang dimana Diagnosis dan resep obat sangat mirip, dan nilai 0 yang dimana tidak ada kesamaan antara diagnosis dan resep obat.

Berikut adalah contoh perhitungannya pada Representasi *TF-IDF* dan *Cosine Similarity*.

*Diagnosis Pasien: "diabetes kronis"*

*Resep Referensi: "metformin, insulin"*

- a. Representasi *TF-IDF*:

Vektor untuk diagnosis: [0,8, 0,1, 0,3]

Vektor untuk resep referensi: [0,7, 0,2, 0,4]

- b. *Cosine Similarity*:

$$\begin{aligned} Cosine\ Similarity &= \frac{(0,8 \times 0,7) + (0,1 \times 0,2) + (0,3 \times 0,4)}{\sqrt{(0,8^2 + 0,1^2 + 0,3^2)} \times \sqrt{0,7^2 + 0,2^2 + 0,4^2}} \\ &= \frac{0,56 + 0,02 + 0,12}{\sqrt{0,64 + 0,01 + 0,09} \times \sqrt{0,49 + 0,04 + 0,16}} \\ &= \frac{0,7}{\sqrt{0,74} \times \sqrt{0,69}} \\ &= 0,95 \end{aligned}$$

Hasilnya adalah **0.95**, menunjukkan tingkat kesamaan yang tinggi.

Untuk menghitung dengan *Evaluation Matrix* berdasarkan data keseluruhan pasien, proses tersebut menggunakan *python* pada *Cursor Software* dengan beberapa *library* seperti *seaborn* untuk menghitung dari *Evaluation Matrix* dan *Seaborn* untuk mempermudah dalam memvisualisasikan *Evaluation Matrix*. Dengan langkah-langkah sebagai berikut.

```

from sklearn.metrics import precision_score, recall_score, f1_score
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the data
pasiens_df = pd.read_csv('pasiens.csv')
resep_obat_df = pd.read_csv('data_resep_obat_new.csv')

# Standardize column names for merging purposes
pasiens_df.columns = ['id', 'nama', 'gender', 'umur', 'diagnosa', 'resep_obat']
resep_obat_df.columns = ['diagnosa', 'resep_obat']

# Preprocess data to enable comparison of "diagnosa" and "resep_obat"
pasiens_df['diagnosa'] = pasien_df['diagnosa'].str.lower()
pasiens_df['resep_obat'] = pasien_df['resep_obat'].str.lower()
resep_obat_df['diagnosa'] = resep_obat_df['diagnosa'].str.lower()
resep_obat_df['resep_obat'] = resep_obat_df['resep_obat'].str.lower()

# Define thresholds
precision_threshold = 0.8
recall_threshold = 0.8
f1_threshold = 0.75

# Calculate precision, recall, and F1 scores
y_true = []
y_pred = []

for _, patient_row in pasien_df.iterrows():
    # Obtain all potential matching prescriptions for the given diagnosis
    relevant_prescriptions = resep_obat_df[resep_obat_df['diagnosa'] == patient_row['diagnosa']]['resep_obat'].values
    recommended_prescriptions = patient_row['resep_obat'].split(',')

    # Determine if each recommended prescription is relevant (1) or not (0)
    y_true.extend([1 if prescription in relevant_prescriptions else 0 for prescription in recommended_prescriptions])
    y_pred.extend([1] * len(recommended_prescriptions))

# Calculate evaluation metrics
precision = precision_score(y_true, y_pred)
recall = recall_score(y_true, y_pred)
f1 = f1_score(y_true, y_pred)

# Filter metrics based on thresholds
metrics_data = {
    'Metric': [],
    'Score': []
}

if precision >= precision_threshold:
    metrics_data['Metric'].append('Precision')
    metrics_data['Score'].append(precision)

if recall >= recall_threshold:
    metrics_data['Metric'].append('Recall')
    metrics_data['Score'].append(recall)

if f1 >= f1_threshold:
    metrics_data['Metric'].append('F1 Score')
    metrics_data['Score'].append(f1)

# Create DataFrame for visualization
metrics_df = pd.DataFrame(metrics_data)

# Plotting the evaluation metrics using Seaborn if they meet thresholds
if not metrics_df.empty:
    plt.figure(figsize=(8, 6))
    sns.barplot(data=metrics_df, x='Metric', y='Score', palette='viridis')
    plt.title("Evaluation Metrics for Prescription Recommendation System (Threshold Applied)")
    plt.ylabel("Score")
    plt.ylim(0, 1)

    # Add percentage annotations on each bar
    for index, row in metrics_df.iterrows():
        plt.text(x=index, y=row['Score'] + 0.02, s=f"{row['Score']:.2%}", ha='center', fontsize=12, fontweight='bold')

    plt.show()
else:
    print("No metrics met the defined thresholds.")
    
```

Gambar 1. Langkah-langkah Skrip Kode Uji Hasil Evaluation Matrix

Berdasarkan *Script Python* terdapat rumus yang menghitung metrik evaluasi untuk sistem rekomendasi:

a. *Precision*:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

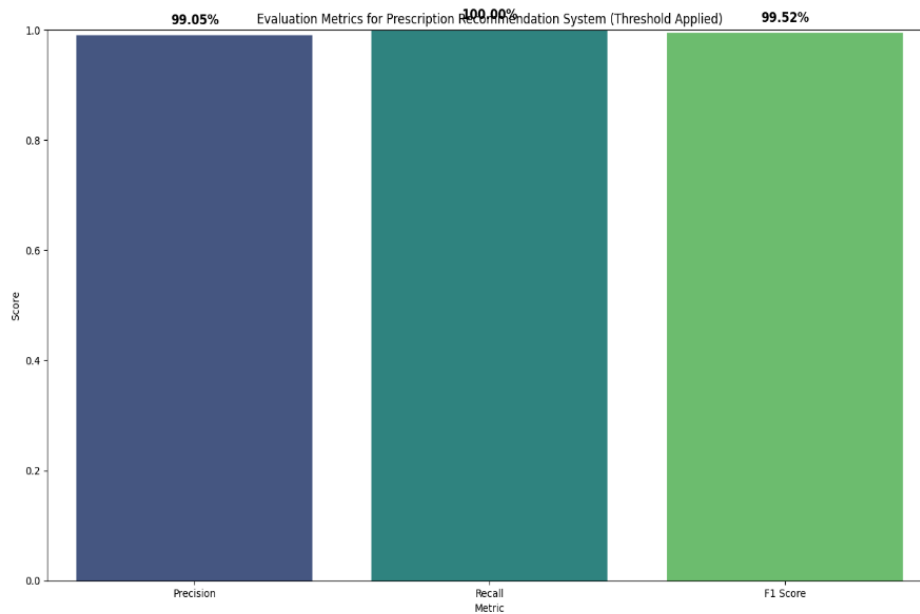
b. *Recall*:

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

c. *F1 Score*:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Setelah seluruh kodingan dijalankan, maka didapatkan hasil Evaluation Matrix sebagai berikut:



Gambar 2. Hasil Dari Evaluation Matrix

Pada gambar tersebut dapat diperoleh 3 metrik utama yaitu:

- Precision:** 99,05%
- Recal:** 100,00%
- F1-Score:** 99,52%

Dari ketiga matrik tersebut memenuhi threshold yang ditentukan sebelumnya yaitu Threshold Precision: 0,8 (80%)

Threshold Recall: 0,8 (80%)

Threshold F1-Score: 0,75 (75%)

Sistem rekomendasi resep obat ini memiliki kinerja yang sangat baik berdasarkan evaluasi Precision, Recall, dan F1 Score. Dengan semua metrik di atas threshold yang telah ditentukan (**80%** untuk Precision dan Recall, **75%** untuk F1 Score), dapat disimpulkan bahwa sistem memiliki akurasi yang tinggi (Precision tinggi), sehingga resep yang diberikan **efektif** untuk memberikan manfaat kepada pasien.

#### D. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang dilakukan pada Penerapan Metode *Content-Based Filtering* Untuk Rekomendasi Pada Resep Obat Berdasarkan Diagnosa Pasien, dapat diuraikan beberapa kesimpulan, antara lain:

- Metode *Content-Based Filtering* ini mampu menentukan resep obat berdasarkan diagnosa pasien yang tepat untuk digunakan oleh Apoteker atau Dokter dan untuk hasil keseluruhan dari pengguna rata-rata persentase kelayakan berkisar 73,75%
- Menghasilkan prototype berupa sistem rekomendasi yang dirancang dapat menerapkan metode *Content-Based Filtering* dengan *TF-IDF* dan *Cosine Similarity* dalam merekomendasikan resep obat untuk pasien berdasarkan kesamaan diagnosa mereka dengan pasien lain yang pernah diberi resep obat yang sama. Dimana penerapan metode *Content-Based Filtering* ini digunakan untuk rekomendasi resep obat berdasarkan diagnosa pasien yang mana dapat membantu dokter dalam memberikan perawatan yang lebih tepat serta personalisasi bagi pasien dan untuk hasil keseluruhan dari para ahli rata-rata persentase kelayakan berkisar 85%.
- Sistem rekomendasi resep obat ini **efektif** dengan kebutuhan pasien, Hal ini dapat dilihat berdasarkan berdasarkan evaluasi Precision, Recall, dan F1 Score. Dengan semua metrik tersebut threshold yang telah ditentukan (**80%** untuk Precision dan Recall, **75%** untuk F1 Score).

#### E. DAFTAR PUSTAKA

- [1] Aprianto, A. (2022). TA: Penerapan Algoritma Content-Based Filtering untuk Rekomendasi Destinasi Wisata pada Aplikasi Picnicker. Universitas Dinamika.
- [2] (Depkes RI, 1997;7. (1997). Depkes Ri Kegunaan Dan Tujuan Rekam Medis,1997;7. jakarta.
- [3] AZHZHORIF, R. M. H. (2022). Implementation of Content-based Health Recommender System in Medical Knowledge on Android. Universitas Mercu Buana Jakarta.

- [4] Badriyah, T., Fernando, R., & Syarif, I. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. 8–9.
- [5] Chunhua Ju, S. Z. (2021). Doctor Recommendation Model Based on Ontology Characteristics and Disease Text Mining Perspective. *BioMed Research International*, 2-10.
- [6] Dewa Ayu Putri Diah Pramesti, I. W. (2022). Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game. *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, 2-5.
- [7] Dwi Ayu Nur Safitri, R. H. (2021). Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori. *Seminar Nasional Inovasi Teknologi* , 2-6.
- [8] Ganney, P. S. (2022). Web Programming1. Dalam *Introduction to Bioinformatics and Clinical Scientific Computing*. <https://doi.org/10.1201/9781003316244-11>
- [9] Ju, C., & Zhang, S. (2021). Doctor recommendation model based on ontology characteristics and disease text mining perspective. *BioMed Research International*, 2021, 1–12.
- [10] L Tim, J. (2023). *Python Programming for Beginners - A Step By Step Direction for Beginners to Learn Python Coding Well and Fast-Independently*.
- [11] Mariani Widia Putri, A. M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *Journal of Information Technology and Computer Science*, 2-7.
- [12] Muhammad Rifqi Ramdhani, R. K. (2022). Pengembangan Sistem Rekomendasi Barbershop di Kota Malang yang Menerapkan Protokol Kesehatan Covid-19. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2-10.
- [13] Resnick, P., & Varian, H. R. (1997). Recommender Systems. Dalam *Communications of the ACM* (Vol. 40, Nomor 3). <https://doi.org/10.1145/245108.245121>