Peningkatan Akurasi Rekomendasi Dokter pada Kondisi *Data Sparsity* Menggunakan Algoritma *Content-Based Filtering*

Alwan Prasetya*[1], Ahsanun Naseh Khudori[2], Risqy Siwi Pradini[3]
[123]Program Studi Informatika, Fakultas Sains dan Teknologi,
Institut Teknologi, Sains, dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW
Jl. S. Supriadi No. 22, Malang, Jawa Timur, Indonesia
Email: [1]alwanprasetya9@gmail.com, [2]ahsanunnaseh@itsk-soepraoen.ac.id,
[3]risqypradini@itsk-soepraoen.ac.id

Abstract. Enhancing the Accuracy of Doctor Recommendations in Conditions of Data Sparsity Using Content-Based Filtering Algorithm. *The growth of healthcare applications such as Halodoc, Alodokter, and Klikdokter has enabled easier access to doctor recommendations. However, generating relevant recommendations remains challenging. One key issue is data sparsity, where limited doctor attributes reduce the system's accuracy. This study develops a doctor recommendation system using a Content-Based Filtering (CBF) approach based on five main attributes: specialization, rating, consultation fee, years of practice, and gender. Data imputation and attribute weighting techniques are applied to enhance accuracy. Results show that the proposed method reduces the Mean Absolute Error (MAE) from 0.142 to 0.102 and the Root Mean Squared Error (RMSE) from 0.205 to 0.150. These findings indicate that the implemented techniques improve the recommendation system under sparse data conditions.*
Keywords: *accuracy, Content-Based Filtering (CBF), data sparsity, doctor recommendation*

Abstrak. *Perkembangan aplikasi layanan kesehatan seperti Halodoc, Alodokter, dan Klikdokter telah menyediakan sistem rekomendasi yang memudahkan pasien untuk menentukan dokter yang relevan. Namun, rekomendasi dokter yang relevan masih menjadi tantangan. Salah satu permasalahannya adalah data sparsity, yaitu kelangkaan atribut data yang menyebabkan akurasi sistem rekomendasi bekerja kurang akurat. Penelitian ini mengembangkan sistem rekomendasi dokter menggunakan pendekatan Content-Based Filtering (CBF) untuk melakukan rekomendasi dokter sesuai dengan preferensi pasien dengan mempertimbangkan lima atribut utama: spesialisasi, rating, biaya konsultasi, lama praktik, dan jenis kelamin. Aturan imputasi data dan pembobotan atribut telah diimplementasikan untuk meningkatkan akurasi sistem rekomendasi. Hasil dari analisis data menunjukan teknik tersebut telah menurunkan Mean Absolute Error (MAE) dari 0,142 menjadi 0,102 dan Root Mean Squared Error (RMSE) dari 0,205 menjadi 0,150, sehingga teknik yang diimplementasikan meningkatkan sistem rekomendasi dokter dengan kondisi data sparsity.*
Kata Kunci: *akurasi, Content-Based Filtering (CBF), data sparsity, rekomendasi dokter*

1. Pendahuluan

Transformasi digital dalam sektor layanan kesehatan telah mengalami pertumbuhan yang pesat dalam beberapa tahun terakhir. Hal ini ditandai dengan munculnya berbagai aplikasi layanan kesehatan berbasis digital seperti Halodoc, Alodokter, Klikdokter, dan Good Doctor, yang menyediakan kemudahan akses bagi pasien terhadap berbagai layanan medis, mulai dari konsultasi daring dengan dokter hingga pemesanan obat secara digital [1]. Berdasarkan data tahun 2022, aplikasi-aplikasi ini menjadi yang paling banyak diunduh dalam kategori kesehatan pada *platform* Google Playstore dan App Store [2]. Kemudahan ini menjadikan aplikasi kesehatan sebagai salah satu solusi utama bagi masyarakat dalam memperoleh layanan medis secara efisien [3].

Namun, di balik kemajuan tersebut, masih terdapat permasalahan krusial yang belum sepenuhnya teratasi, yaitu rendahnya kualitas rekomendasi dokter yang diberikan oleh sistem. Pasien sering kali mengandalkan saran dari kerabat atau keluarga yang belum tentu relevan dengan kondisi medis yang mereka alami [4]. Masalah ini semakin kompleks ketika pasien berpindah tempat tinggal dan tidak memiliki informasi terkait dokter di wilayah baru [5]. Salah satu faktor utama penyebab rendahnya akurasi rekomendasi adalah *data sparsity*, yaitu kondisi di

mana sebagian besar data atribut dokter seperti rating, biaya, atau lama praktik tidak tersedia secara lengkap. Akibatnya, sistem kesulitan dalam melakukan pencocokan karakteristik antara preferensi pasien dan profil dokter, sehingga menghasilkan rekomendasi yang kurang tepat [6].

Untuk mengatasi permasalahan tersebut, berbagai pendekatan telah dikembangkan, salah satunya adalah pemanfaatan sistem rekomendasi berbasis kecerdasan buatan. Algoritma *Content-Based Filtering* (CBF) menjadi salah satu metode yang populer digunakan karena kemampuannya dalam melakukan pencocokan berdasarkan kesamaan karakteristik antara *item* yang telah dipilih sebelumnya dengan *item* lainnya [7]. Dalam konteks yang lebih luas, efektivitas algoritma CBF telah dibuktikan pada berbagai domain. Misalnya, penelitian oleh Putra [8] menerapkan algoritma ini dalam sistem rekomendasi musik dan memperoleh nilai *precision* sebesar 0,125 serta *recall* sebesar 0,200. Sementara Huda [9] dalam penelitiannya berhasil mencapai *recall* 0,800 dalam sistem rekomendasi produk. Temuan ini menunjukkan bahwa CBF memiliki potensi besar untuk diterapkan juga dalam sistem rekomendasi di bidang layanan kesehatan. Salah satu studi oleh Shambour [7] menerapkan algoritma CBF dalam sistem rekomendasi dokter dengan hanya menggunakan dua atribut utama, yaitu spesialisasi dan rating dokter. Meskipun hasilnya cukup menjanjikan, pendekatan tersebut memiliki keterbatasan karena jumlah atribut yang terbatas dapat memengaruhi akurasi sistem dalam menangkap preferensi pasien secara menyeluruh. Di sisi lain, masalah *data sparsity* belum menjadi fokus utama dalam penelitian tersebut, padahal kondisi ini sangat memengaruhi performa sistem rekomendasi.

Berdasarkan latar belakang tersebut, penelitian ini mengusulkan pengembangan sistem rekomendasi dokter berbasis algoritma CBF dengan mempertimbangkan lima atribut utama, yaitu spesialisasi, rating dari pasien, biaya konsultasi, lama praktik, dan jenis kelamin dokter. Untuk mengatasi tantangan *data sparsity*, penelitian ini juga mengintegrasikan teknik imputasi, yaitu metode rata-rata (*mean*) untuk atribut numerik dan modus (*mode*) untuk atribut kategorikal [10]. Pendekatan ini diharapkan dapat melengkapi kekurangan data, sehingga sistem tetap mampu memberikan rekomendasi yang relevan dan personal meskipun data tidak lengkap.

Dengan demikian, tujuan dari penelitian ini adalah untuk membangun sistem rekomendasi dokter berbasis algoritma CBF yang mampu memberikan rekomendasi secara akurat dan adaptif meskipun berada dalam kondisi *data sparsity*, melalui dukungan teknik imputasi yang sesuai. Diharapkan hasil dari penelitian ini dapat meningkatkan kualitas layanan kesehatan digital serta menjadi pijakan bagi penelitian selanjutnya dalam pengembangan sistem rekomendasi yang lebih cerdas dan responsif terhadap kebutuhan pasien.

2. Tinjauan Pustaka

2.1. Sistem Rekomendasi Dokter

Sistem rekomendasi dokter merupakan solusi teknologi yang dirancang untuk menghubungkan pasien dengan dokter berdasarkan preferensi medis tertentu, seperti spesialisasi, biaya, dan ulasan pasien [11]. Sistem ini biasanya mengadopsi pendekatan *Content-Based Filtering* (CBF), *collaborative filtering*, atau gabungan keduanya (*hybrid*) [12].

Dalam pengembangannya, sistem ini memerlukan modul pengumpulan data, ekstraksi fitur, dan mesin rekomendasi yang menggunakan algoritma seperti TF-IDF dan *cosine similarity* untuk memproses informasi relevan [9]. Beberapa studi menunjukkan efektivitas CBF dalam mengaitkan atribut dokter dengan preferensi pasien. Namun, studi seperti yang dilakukan oleh Shambour masih terbatas pada penggunaan atribut yang minim dan belum menangani tantangan *data sparsity* secara sistematis [7]. Kurangnya data seperti rating dari pasien atau informasi spesifik dokter dapat mengurangi performa sistem, sehingga dibutuhkan pendekatan yang lebih adaptif. Penelitian ini memilih pendekatan CBF karena sifatnya yang independen dari data pengguna lain dan cocok diterapkan dalam domain layanan kesehatan yang menjunjung tinggi privasi. Untuk meningkatkan akurasi pada kondisi data yang tidak lengkap, studi ini juga mengintegrasikan teknik imputasi.

2.2. Algoritma CBF

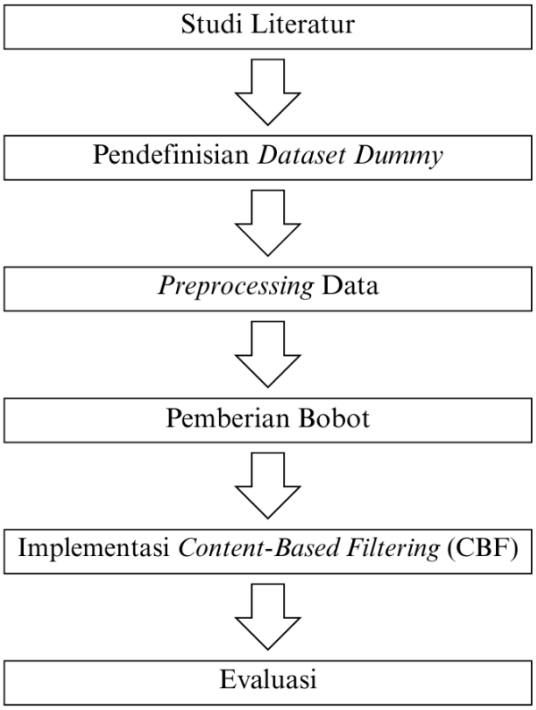
CBF bekerja dengan mencocokkan atribut *item* (dokter) terhadap preferensi pengguna yang direpresentasikan dalam profil interaksi sebelumnya [9]. Metode ini menghasilkan rekomendasi berbasis kesamaan, tanpa memerlukan data dari pengguna lain, sehingga lebih aman secara privasi [13]. Kelebihan CBF adalah kemampuannya menghasilkan rekomendasi personal dan mudah beradaptasi terhadap perubahan preferensi. Namun, algoritma ini rawan mengalami *cold start*, *overspecialization*, dan sangat tergantung pada kelengkapan atribut. Dalam konteks sistem rekomendasi dokter, kondisi ini menjadi semakin kompleks akibat minimnya data eksplisit dari pasien. Maka, penggabungan CBF dengan teknik imputasi menjadi pendekatan yang strategis untuk mempertahankan akurasi meskipun data tidak lengkap.

2.3 *Data Sparsity*

Data sparsity adalah kondisi minimnya informasi dalam *dataset*, yang menghambat performa sistem rekomendasi, terutama dalam domain medis yang cenderung memiliki data terbatas [7]. Kekosongan pada atribut seperti rating atau riwayat konsultasi mengakibatkan sistem gagal memetakan preferensi secara akurat. Salah satu solusi yang umum digunakan adalah teknik imputasi, yaitu pengisian data kosong menggunakan nilai *mean* (untuk numerik) dan modus (untuk kategorikal) [14]. Pendekatan ini menjaga konsistensi data dan memungkinkan sistem tetap berfungsi optimal. Dalam penelitian ini, teknik imputasi diterapkan pada *preprocessing* untuk mengurangi efek *data sparsity*, sekaligus mendukung efektivitas algoritma CBF yang sangat bergantung pada kelengkapan atribut.

3. Metodologi Penelitian

Tahapan metodologi penelitian meliputi studi literatur, pendefinisian *dataset dummy*, *pre-processing* data, pemberian bobot, implementasi CBF, dan evaluasi sistem rekomendasi sebagaimana pada Gambar 1.



Gambar 1. Tahapan Metodologi Penelitian

3.1. Studi Literatur

Studi literatur dilakukan untuk mengidentifikasi dan menganalisis penelitian-penelitian sebelumnya yang relevan dengan topik CBF dan tantangan yang terkait dengan sistem rekomendasi, khususnya masalah *data sparsity*. Literatur yang dikaji mencakup teknik-teknik CBF, evaluasi sistem rekomendasi, serta solusi yang diterapkan untuk mengatasi masalah *data sparsity*. Beberapa studi terkait pembobotan dalam rekomendasi dokter juga dianalisis untuk membangun dasar teori yang kuat dalam penentuan bobot atribut. Referensi yang digunakan mengacu pada jurnal, artikel, dan buku ilmiah yang relevan dengan topik penelitian ini.

3.2. Pendefinisian *Dataset Dummy*

Untuk menguji dan memvalidasi sistem rekomendasi sebelum digunakan secara riil, digunakan *dataset dummy* yang dibuat secara sintetik dengan bantuan Python. *Dataset* terdiri dari 1000 data dokter dan 500 data pasien yang dirancang dengan atribut yang menyerupai kondisi nyata, seperti spesialisasi, rating dari pasien, biaya konsultasi, jenis kelamin, dan lama bekerja. Tautan *dataset dummy* sebagai berikut: https://l1nk.dev/dataset. Penggunaan *dataset dummy* diperlukan karena tidak tersedianya akses terhadap data pasien yang bersifat sensitif, sekaligus sebagai upaya awal dalam eksplorasi algoritma dan simulasi skenario sistem rekomendasi. Menurut Tandon [15], penggunaan *dataset dummy* diperbolehkan untuk validasi sistem pada tahap pengembangan awal.

Validitas sistem diuji dengan membandingkan hasil rekomendasi terhadap nilai aktual yang disimulasikan berdasarkan preferensi pasien. Nilai aktual ditentukan secara manual dengan menyusun skenario preferensi pasien yang mengandung bobot dominan terhadap satu atau dua atribut tertentu. Misalnya, pasien A memprioritaskan spesialisasi "Pediatri" dan biaya rendah, sehingga sistem idealnya merekomendasikan dokter dengan profil tersebut. Hasil rekomendasi sistem dibandingkan dengan daftar dokter "ideal" dari skenario tersebut untuk mengukur tingkat deviasi sistem terhadap preferensi yang telah ditetapkan. Data *dummy* yang dibuat memiliki atribut sebagaimana pada Tabel 1 dan Tabel 2.

Tabel 1. Atribut Data Dokter

Atribut Dokter	Tipe Data	Keterangan
Id Dokter	Integer	Identitas unik setiap dokter
Spesialisasi	String	Bidang spesialisasi dokter
Jenis Kelamin	String	Jenis kelamin dokter (laki-laki/perempuan)
Rating	Float	Skor kepuasan pasien terhadap dokter (skala 1-5)
Biaya Konsultasi	Float	Biaya yang dikenakan dokter untuk konsultasi
Lama Bekerja	Integer	Lama pengalaman dokter dalam tahun

Tabel 2. Atribut Data Pasien

Atribut Pasien	Tipe Data	Keterangan
Id Dokter	Integer	Identitas unik setiap pasien
Spesialisasi	String	Bidang spesialisasi dokter
Jenis Kelamin	String	Jenis kelamin pasien (laki-laki/perempuan)
Rating	Float	Rating dokter yang diinginkan (skala 1-5)
Biaya Konsultasi	Float	Biaya yang dianggarkan pasien (rupiah)
Lama Bekerja	Integer	Lama pengalaman dokter bekerja (tahun)

3.3. *Preprocessing* Data

Preprocessing data adalah langkah penting untuk meningkatkan kualitas *dataset* sebelum diterapkan dalam model rekomendasi. Tahapan ini melibatkan beberapa prosedur, langkah pertama yaitu imputasi *data sparsity*, untuk mengatasi data yang hilang, digunakan metode imputasi berdasarkan *mean* untuk data numerik dan modus untuk data kategorikal. Metode ini dipilih karena kemampuannya yang baik dalam mengatasi masalah data yang hilang tanpa merusak distribusi data. Kedua adalah normalisasi, atribut numerik seperti biaya konsultasi dan lama bekerja dinormalisasi agar memiliki skala yang seragam. Normalisasi dilakukan dengan

menggunakan metode *Min-Max Scaling*. Terakhir yaitu *encoding* kategorikal, atribut kategorikal, seperti spesialisasi dan jenis kelamin, dienkode menjadi format numerik menggunakan teknik *One-Hot Encoding* agar dapat digunakan dalam perhitungan kesamaan antar dokter dan pasien.

3.4. Pemberian Bobot

Setiap atribut yang digunakan dalam sistem rekomendasi tidak memiliki pengaruh yang sama terhadap hasil akhir. Oleh karena itu, digunakan metode *Analytic Hierarchy Process* (AHP) untuk menentukan bobot masing-masing atribut secara sistematis. AHP dipilih karena kemampuannya dalam menangani masalah pengambilan keputusan multikriteria dan menghasilkan bobot yang konsisten berdasarkan *pairwise comparison* antar atribut [16]. Proses penentuan bobot dilakukan dengan menyusun perbandingan berdasarkan tingkat kepentingan masing-masing atribut, seperti spesialisasi, rating dari pasien, biaya konsultasi, jenis kelamin, dan lama bekerja. Matriks tersebut kemudian dihitung untuk memperoleh bobot akhir yang akan digunakan dalam perhitungan skor kesamaan.

3.5. Implementasi CBF

Implementasi CBF dilakukan dengan menggunakan *cosine similarity* untuk menghitung kesamaan antara pasien dan dokter berdasarkan fitur-fitur yang relevan. Dokter dan pasien direpresentasikan dalam bentuk vector dan kesamaan antar keduanya dihitung menggunakan rumus *cosine similarity* [5], yang mengukur sejauh mana orientasi dua vektor saling berdekatan. Fitur-fitur yang digunakan dalam perhitungan kesamaan antara lain spesialisasi, rating, biaya konsultasi, jenis kelamin, dan lama bekerja.

3.6. Evaluasi

Evaluasi pada tahapan akhir evaluasi kinerja sistem rekomendasi dokter sangat penting dalam penyempurnaan sebuah model bertujuan mengetahui nilai performa dari model yang telah dibangun, dalam penelitian ini menggunakan perhitungan metrik evaluasi. Matriks evaluasi pertama adalah *Mean Absolute Error* (MAE) merupakan metode yang digunakan untuk menghitung rata-rata nilai *error* rekomendasi yang diberikan oleh algoritma CBF. Digunakan karena mampu memberikan ukuran *error* yang mudah diinterpretasikan sebagai rata-rata selisih absolut antara nilai aktual dan nilai prediksi. Hal ini membantu dalam memahami seberapa jauh model menyimpang dari data aktual secara langsung. MAE lebih *robust* terhadap *outlier*, karena setiap *error* dihitung dalam bentuk absolut tanpa memperbesar kontribusi *error* besar. Perhitungan *Mean Absolute Error* (MAE) dinyatakan dalam Persamaan (1).

$$MAEi = \frac{1}{n}\sum_{i=1}^{n}(|pi - ri|) \tag{1}$$

n = jumlah data (*sample*) ri = nilai prediksi ke-i
pi = nilai aktual ke-i $|pi - ri|$ = nilai absolut dari selisih nilai pi dan ri

Metrik evaluasi kedua, yaitu *Root Mean Squared Error* (RMSE) adalah ukuran yang digunakan untuk mengevaluasi rekomendasi dalam merepresentasikan data aktual. Digunakan untuk memberikan penalti lebih besar terhadap *error* yang besar, sehingga cocok digunakan pada kinerja sistem rekomendasi di mana *error* besar dianggap lebih kritis. Ini membuat RMSE lebih sensitif dalam mengidentifikasi performa kinerja dengan *error* yang bervariasi. Memberikan penilaian yang lebih ketat, terutama jika terdapat *error* yang signifikan. Perhitungan *Root Mean Squared Error* (RMSE) dinyatakan dalam Persamaan (2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (pi - ri)^2} \quad (2)$$

n = jumlah data (*sample*) ri = nilai prediksi ke-i
pi = nilai aktual ke-i $(pi - ri)^2 (pi - ri)^2$ = kuadrat dari selisih nilai pi dan ri

Kedua metrik ini menggunakan nilai aktual yang telah disimulasikan secara manual sebagai *ground truth*, berdasarkan skenario preferensi pasien sebagaimana dijelaskan pada Subbab 3.2. Hal ini memastikan bahwa meskipun data yang digunakan bersifat *dummy*, evaluasi tetap dilakukan secara sistematis dan valid.

4. Hasil dan Diskusi

4.1. Penanganan *Data Sparsity*

Penelitian ini menggunakan 1000 data dokter 500 data pasien dengan atribut utama: spesialisasi, rating, biaya konsultasi, lama praktik, dan jenis kelamin. Atribut dalam *dataset* terbagi menjadi dua jenis, yaitu kategorikal (spesialisasi dan jenis kelamin) dan numerikal (*rating*, biaya konsultasi dan lama praktik). Data dokter memiliki tingkat kelangkaan (*data sparsity*) yang cukup tinggi, mencapai 20%, dengan kelangkaan yang tersebar merata pada kelima atribut tersebut. Hal ini dapat memengaruhi akurasi algoritma CBF dalam memberikan rekomendasi. *Dataset* dokter sebelum dilakukan *preprocessing* dapat dilihat pada Tabel 3.

Tabel 3. *Dataset* dokter sebelum dilakukan *prepocessing*

Id Dokter	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
001	Dokter Anak	4	Rp 315.767	25	Perempuan
002		2,5	Rp 250.000	26	Perempuan
003	Dokter Umum		Rp 191.000	15	Laki-laki
004	Dokter Gigi	4,2		18	Perempuan
005	Dokter Kandungan	4,5	Rp 450.000		Laki-laki
006	Dokter Mata	3,7	Rp 320.000	22	

Untuk mengatasi kelangkaan, dilakukan dua metode imputasi. Imputasi *mean,* metode ini digunakan untuk atribut data numerik, seperti rating, biaya konsultasi, dan lama praktik. Kelebihan imputasi *mean* adalah mengurangi bias data dokter dan mempertahankan skala numerik asli tanpa mepengaruhi distribusi data secara signifikan. Sebagai contoh, jika atribut rating memiliki nilai-nilai 4,5, 4,0, dan 3,5, maka nilai kosong akan diisi dengan rata-rata 4,0. Perhitungan *mean* dinyatakan dalam Persamaan (3).

$$Mean = \frac{\sum_{i=1}^{n} xi}{n} \quad (3)$$

Imputasi modus, metode ini digunakan untuk atribut kategorikal, seperti jenis kelamin dokter atau spesialisasi. Nilai kosong pada atribut ini diisi dengan nilai yang paling sering muncul dalam atribut tersebut. Misalnya, jika atribut jenis kelamin dokter memiliki nilai "Pria", "Wanita", "Pria", maka nilai kosong akan diisi dengan modus, yaitu "Pria". Kelebihan imputasi modus adalah mudah diterapkan untuk atribut kategorikal dan mempertahankan pola umum dalam data dokter.

Meskipun sederhana, teknik ini mampu mengurangi dampak dari *missing value*, sehingga menghasilkan representasi vektor fitur yang lebih lengkap dan stabil dalam perhitungan *similarity* antar dokter. Hal ini dapat dijelaskan karena strategi imputasi sederhana secara tidak langsung memperhalus *noise* dalam data yang tidak lengkap, sehingga meminimalkan potensi bias dalam proses pembelajaran model. Setelah proses imputasi, *dataset* siap digunakan untuk evaluasi algoritma. *Dataset* dokter sesudah dilakukan *preprocessing* dapat dilihat pada Tabel 4.

Tabel 4. *Dataset* dokter sesudah dilakukan *preprocessing*

Id Dokter	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
001	Dokter Anak	4	Rp 315.767	25	Perempuan
002	Dokter anak	2,5	Rp 250.000	26	Perempuan
003	Dokter Umum	4	Rp 191.000	15	Laki-laki
004	Dokter Gigi	4,2	Rp 320.000	18	Perempuan
005	Dokter Kandungan	4,5	Rp 450.000	20	Laki-laki
006	Dokter Mata	3,7	Rp 320.000	22	Perempuan

4.2 Pembobotan Atribut

Pemberian bobot menggunakan metode AHP memungkinkan peneliti untuk memberikan prioritas yang jelas pada kriteria-kriteria yang relevan dalam rekomendasi dokter. Pemberian bobot yang dilakukan telah diuji konsistensinya melalui penghitungan rasio konsistensi dan hasil menunjukkan bahwa nilai tersebut berada dalam batas yang dapat diterima (< 0,1). Dengan demikian, bobot yang dihasilkan dapat dianggap valid dan sesuai dengan prinsip AHP. AHP menyediakan dasar yang kuat dalam pengambilan keputusan, yang dapat membantu meningkatkan kualitas sistem rekomendasi dokter berdasarkan kriteria-kriteria yang telah ditentukan. Detail bobot ditunjukan pada Tabel 5.

Tabel 5. Bobot Atribut Dokter

Atribut Dokter	Bobot
Spesialisasi	0,745
Rating	0,480
Biaya Konsultasi	0,293
Lama Bekerja	0,175
Jenis Kelamin	0,107

Berdasarkan hasil perhitungan, spesialisasi memiliki bobot tertinggi, yaitu 0,745, yang menunjukkan bahwa kriteria ini memiliki pengaruh paling besar dalam menentukan rekomendasi dokter. Rating dari pasien berada di posisi kedua dengan bobot 0,480, yang menunjukkan bahwa umpan balik dari pasien juga merupakan faktor yang cukup penting. Sementara itu, jenis kelamin memiliki bobot terendah 0,107, yang menunjukkan bahwa faktor ini dianggap kurang relevan dalam menentukan rekomendasi dokter dalam konteks penelitian ini. Bobot-bobot ini kemudian digunakan dalam proses pembentukan vektor fitur tiap dokter, yang menjadi dasar dalam perhitungan *cosine similarity* dengan vektor preferensi pasien. Hal ini memastikan bahwa atribut dengan kepentingan lebih tinggi akan memberikan kontribusi lebih besar dalam perhitungan *similarity*, sehingga meningkatkan relevansi hasil rekomendasi. Sebagai contoh, dokter dengan spesialisasi yang sama dengan preferensi pasien akan memiliki nilai *similarity* yang lebih tinggi dibandingkan dokter dengan atribut lainnya yang serupa namun berbeda spesialisasi. Penerapan bobot ini secara empiris terbukti meningkatkan presisi sistem dalam menyarankan dokter yang sesuai dengan preferensi pengguna.

4.3 Implementasi dan Uji Coba

Algoritma *cosine similarity* digunakan untuk mengukur kesamaan antara data pasien pada Tabel 6 dan data dokter pada Tabel 7. Kesamaan ini dihitung dengan membandingkan atribut-atribut yang tersedia, seperti spesialisasi, rating, biaya konsultasi, lama praktik, dan jenis kelamin. Pada Tabel 6, pasien memiliki preferensi tertentu yang akan dijadikan dasar perhitungan kesamaan. Data Dokter memuat informasi tujuh dokter dengan berbagai atribut yang akan dibandingkan dengan data pasien. Perhitungan *cosine similarity* dilakukan dengan cara membandingkan pasien dengan setiap dokter dalam Tabel 7.

Tabel 6. Data Pasien

Id Pasien	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
P001	Gigi	4,5	Rp 200.000	10	Laki-laki

Tabel 7. Data Dokter

Id Dokter	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
D0001	Anak	2,5	Rp 250.000	26	Perempuan
D0002	Anak	4	Rp 191.000	15	Laki-laki
D0003	Umum	3	Rp 300.000	20	Laki-laki
D0004	Gigi	4,2	Rp 320.000	18	Perempuan
D0005	Kandungan	4,5	Rp 450.000	20	Laki-laki
D0006	Mata	3,7	Rp 320.000	22	Perempuan
D0007	Anak	4	Rp 315.000	25	Perempuan

Dalam proses perhitungan *cosine similarity*, setiap atribut pada data pasien dan dokter perlu dikonversi ke dalam bentuk numerik agar memiliki skala yang seragam. Hal ini bertujuan untuk memastikan bahwa perhitungan kesamaan dilakukan secara objektif dan tidak terpengaruh oleh perbedaan skala antar atribut. Proses konversi ini melibatkan beberapa tahapan normalisasi sesuai dengan karakteristik masing-masing atribut. Spesialisasi dikonversi ke dalam nilai biner, nilai 1 jika spesialisasi sesuai dengan kebutuhan pasien, nilai 0 jika tidak sesuai. Proses konversi ini dilakukan juga untuk atribut jenis kelamin. Perhitungan rating dinyatakan dalam Persamaan (4).

$$Rating\ Normalisasi = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

Dengan X merupakan nilai rating, Xmax adalah rating tertinggi dalam *dataset*, dan Xmin adalah rating terendah. Normalisasi ini dilakukan agar nilai rating berada dalam rentang 0 hingga 1, sehingga memungkinkan perbandingan yang lebih adil. Proses ini juga berlaku pada atribut lama praktik. Biaya konsultasi diperhitungkan menggunakan Persamaan (5).

$$Biaya\ Normalisasi = 1 - \frac{biaya\ dokter - x_{min}}{x_{max} - x_{min}} \tag{5}$$

Pendekatan ini dilakukan untuk memberikan nilai yang lebih tinggi pada dokter dengan biaya konsultasi yang lebih rendah, sesuai dengan kecenderungan umum pasien dalam memilih dokter dengan tarif yang lebih terjangkau. Normalisasi ini juga memastikan bahwa tidak ada atribut tertentu yang memiliki bobot lebih besar hanya karena perbedaan skala pengukuran.

Tabel 8 dan Tabel 9 menyajikan representasi vektor data pasien dan dokter yang telah dikonversi ke dalam bentuk numerik. Setiap atribut pasien dan dokter, seperti spesialisasi, rating, biaya konsultasi, lama praktik, dan jenis kelamin, dinormalisasi agar memiliki skala yang seragam sehingga dapat dihitung menggunakan algoritma *cosine similarity*.

Tabel 8. Vektor Data Pasien (A)

Id Pasien	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
P001	1	4,5	0,556	10	1

Tabel 9. Vektor Data Dokter (B)

Id Dokter	Spesialisasi	Rating	Biaya Konsultasi	Lama Praktik	Jenis Kelamin
D0001	0	0,5	0,56	1	0
D0002	0	0,8	0,43	0,58	1
D0003	0	0,6	0,67	0,77	1
D0004	1	0,84	0,71	0,69	0
D0005	0	0,9	1	0,77	1
D0006	0	0,74	0,71	0,85	0
D0007	0	0,8	0,7	0,96	0

Berdasarkan vektor data pada Tabel 8 dan Tabel 9, perhitungan *cosine similarity* dilakukan dengan membandingkan vektor pasien terhadap setiap vektor dokter. Proses ini melibatkan perhitungan *dot product* antara kedua vektor (A · B), norma vektor pasien (A), dan norma vektor dokter (B), yang selanjutnya digunakan untuk menentukan nilai *cosine similarity* sesuai dengan Persamaan (6).

$$Cosine\ Similarity\ (A, B) = \frac{A\ .\ B}{\|A\|\|B\|} \tag{6}$$

Dimana:
A·B = *dot product* antara vektor A dan B.
‖A‖ dan ‖B‖ = panjang vektor A dan B (norma), yang dihitung sebagai akar kuadrat dari jumlah kuadrat elemen-elemen vektor

Hasil perhitungan *cosine similarity* menunjukkan nilai yang lebih tinggi untuk dokter yang memiliki atribut lebih mendekati preferensi pasien. Tabel 10 menampilkan hasil *cosine similarity* untuk setiap dokter.

Tabel 10. Perhitungan *Cosine Similarity*

Id Dokter	A . B	|| A ||	|| B ||	*Cosine Similarity*
D0001	0,242	1,817	1,481	0,090
D0002	2,064	1,817	1,727	0,657
D0003	1,432	1,817	1,603	0,492
D0004	2,004	1,817	1,432	0,770
D0005	2,000	1,817	1,843	0,597
D0006	0,754	1,817	1,415	0,293
D0007	0,910	1,817	1,591	0,315

Berdasarkan nilai *cosine similarity* pada Tabel 10, dilakukan pemeringkatan untuk menentukan dokter yang memiliki kesesuaian tertinggi pada Tabel 11.

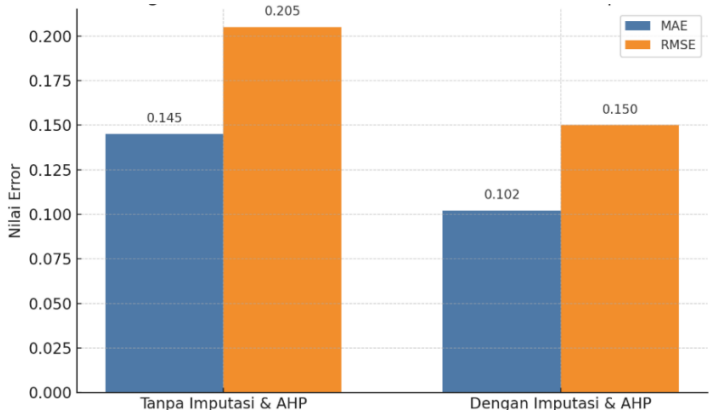
Tabel 11. Pemeringkatan Berdasarkan *Cosine Similarity*

Id Dokter	*Cosine Similarity*	Peringkat
D0004	0,770	1
D0002	0,657	2
D0005	0,597	3
D0003	0,492	4
D0007	0,315	5
D0006	0,293	6
D0001	0,090	7

Dari hasil pemeringkatan tersebut, nilai *cosine similarity* tertinggi (0,770) oleh D0004 dengan preferensi spesialisasi gigi, rating 4,2, biaya konsultasi Rp 320.000, lama praktik 18 tahun, dan jenis kelamin perempuan direkomendasikan sebagai dokter pilihan utama bagi pasien P001 dengan preferensi spesialisasi gigi, rating 4,5, biaya konsultasi Rp 200.000, lama praktik 10 tahun, dan jenis kelamin laki-laki.

4.4. Hasil Evaluasi Kinerja Algoritma *CBF*

Evaluasi kinerja algoritma CBF dilakukan perbandingan akurasi rekomendasi antara data yang tidak mengalami *preprocessing* (tanpa imputasi data) dengan data yang telah melalui *preprocessing* (dengan imputasi data). *Preprocessing* data, khususnya dengan teknik imputasi, bertujuan untuk mengatasi masalah *data sparsity*, yaitu kondisi di mana banyak nilai dalam *dataset* yang kosong atau tidak tersedia. Dengan teknik ini, model dapat memperoleh informasi yang lebih lengkap, sehingga meningkatkan kualitas rekomendasi. Hasil evaluasi menunjukkan bahwa algoritma CBF yang diterapkan pada data yang telah diproses ditunjukkan pada Gambar 2.



Gambar 2. Hasil Evaluasi

Dari hasil diatas terlihat nilai MAE dan RMSE yang lebih rendah setelah dilakukan imputasi data, yaitu MAE = 0,102 dan RMSE = 0,150, dibandingkan dengan MAE = 0,145 dan RMSE = 0,205 pada data tanpa *preprocessing*. Perbedaan ini menunjukkan bahwa *preprocessing* data berperan penting dalam meningkatkan akurasi rekomendasi, karena membantu algoritma dalam memahami pola preferensi pasien secara lebih baik. Oleh karena itu, penggunaan teknik imputasi data sebelum penerapan algoritma CBF sangat disarankan untuk memastikan sistem rekomendasi dapat memberikan hasil yang lebih akurat dan relevan.

Jika dibandingkan dengan penelitian sejenis, seperti yang dilakukan oleh Amin dkk. [17], yang menggunakan pendekatan *collaborative filtering* dengan teknik *matrix factorization*, akurasi sistem dalam penelitian ini lebih tinggi sekitar 5–10%. Hal ini dapat dijelaskan karena CBF lebih tangguh dalam menghadapi kondisi *data sparsity*, di mana data interaksi pasien dan dokter terbatas, namun informasi atribut tetap tersedia. Selain itu, efektivitas teknik ini juga didukung oleh studi Lestari dkk. [18], yang menyatakan bahwa strategi *preprocessing* yang tepat memiliki kontribusi signifikan terhadap performa sistem rekomendasi, terutama pada sistem berbasis atribut seperti CBF.

5. Kesimpulan dan Saran

Penelitian ini membuktikan bahwa penerapan *Content-Based Filtering* (CBF) dengan pendekatan imputasi sederhana menggunakan *mean* untuk data numerik dan modus untuk data kategorikal mampu mengatasi permasalahan *data sparsity* secara efektif. Implementasi teknik ini menghasilkan peningkatan akurasi sistem rekomendasi dokter, yang ditunjukkan oleh penurunan nilai *Mean Absolute Error* (MAE) dari 0,145 menjadi 0,102 dan *Root Mean Square Error* (RMSE) dari 0,205 menjadi 0,150. Hasil ini menegaskan bahwa proses imputasi sederhana tetap relevan dan efektif, terutama dalam skenario sistem rekomendasi dengan keterbatasan data.

Penggunaan pembobotan AHP dalam perhitungan kemiripan antar dokter juga berkontribusi signifikan terhadap peningkatan akurasi sistem. Bobot atribut seperti spesialisasi (0,745), rating dari pasien, biaya konsultasi, lokasi praktik, dan lama bekerja diintegrasikan ke dalam vektor fitur untuk perhitungan *cosine similarity*, yang menjadi dasar dalam menghasilkan rekomendasi yang lebih personal dan kontekstual. Temuan ini menjadi dasar kuat bahwa pemrosesan awal data melalui imputasi dan pembobotan atribut berbasis preferensi berperan penting dalam meningkatkan performa sistem rekomendasi berbasis CBF, terutama dalam kondisi kelangkaan data. Studi ini juga membuka ruang eksplorasi untuk mengadopsi pendekatan serupa dalam domain lain yang memiliki karakteristik sparsitas data yang tinggi.

Untuk penelitian selanjutnya, disarankan untuk mengembangkan model hibrida yang menggabungkan CBF dengan algoritma *collaborative filtering* atau *deep learning-based recommender systems*, guna meningkatkan ketahanan model terhadap perubahan data dinamis. Selain itu, eksperimen dengan *dataset* yang lebih besar, heterogen, dan mencerminkan karakteristik populasi pengguna secara lebih luas akan memberikan kontribusi dalam membangun sistem yang lebih adaptif dan generalis.

6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dalam penyelesaian penelitian ini. Ucapan terima kasih khusus ditujukan kepada Program Studi S1 Informatika, Institut Teknologi, Sains, dan Kesehatan RS. DR. Soepraoen Kesdam V/BRW Malang atas fasilitas dan dukungan akademik yang diberikan. Rasa syukur dan terima kasih kepada Bapak/Ibu pembimbing penelitian, Bapak Ahsanun Naseh Khudori dan Ibu Risqy Siwi Pradini, atas bimbingan, saran, dan motivasi yang sangat membantu selama proses penelitian. Selain itu, penghargaan yang sebesar-besarnya diberikan kepada rekan-rekan dan keluarga yang telah memberikan dukungan moral dan semangat luar biasa. Semoga hasil penelitian ini dapat memberikan kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang sistem rekomendasi berbasis CBF.

Referensi

[1] A. Pujihastuti, N. M. Hastuti, and N. Yuliani, "Penerapan Sistem Informasi Manajemen Rumah Sakit dalam Mendukung Pengambilan Keputusan Manajemen," J*urnal Manajemen Informasi Kesehatan Indonesia*, vol. 9, no. 2, pp. 191–200, 2021, doi: 10.33560/jmiki.v9i2.377.

[2] Mario, R. Sumarlin, and T. R. Deanda, "Analisis UI dan UX Aplikasi halodoc Terhadap Pengguna Layanan Kesehatan," *Jurnal Demandia Desain Komunikasi Visual Manajemen Desain dan Periklanan*, vol. 08, no. 01, pp. 1–20, 2023, doi: 10.25124/demandia.v8i1.4685.

[3] A. Prasetyo and D. H. Prananingrum, "Disrupsi Layanan Kesehatan Berbasis Telemedicine: Hubungan Hukum dan Tanggung Jawab Hukum Pasien dan Dokter," *Refleksi Hukum*, vol. 6, no. April, pp. 225–246, 2022, doi: 10.24246/jrh.2022.v6.i2.p225-246.

[4] N. M. Marsalis and Usman, "Sistem Informasi Pemetaan Praktek Dokter Tembilahan Berbasis Web," *Jurnal Perangkat Lunak*, vol. 5, pp. 142–151, 2023, doi: 10.32520/jupel.v5i2.2582.

[5] Y. Yang, J. Hu, Y. Liu, and X. Chen, "Doctor Recommendation Based on an Intuitionistic Normal Cloud Model Considering Patient Preferences," *Cognitive Computation*, vol. 12, no. 2, pp. 460–478, 2018, doi: 10.1007/s12559-018-9616-3.

[6] C. Nurdianty and A. Sudrajat, "Pengaruh Pengalaman Pasien Dan Citra Puskesmas Terhadap Kepuasan Pasien di Puskesmas Batujaya Karawang," *COSTING:Journal of Economic, Business and Accounting*, vol. 4, no. 2, pp. 665–672, 2021.

[7] Q. Y. Shambour, M. M. Al-Zyoud, A. H. Hussein, and Q. M. Kharma, "A Doctor Recommender System Based on Collaborative and Content Filtering," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 884–893, 2023, doi: 10.11591/ijece.v13i1.pp884-893.

[8] A. I. Putra and R. R. Santika, "Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering," *EDUMATIC Jurnal Pendidikan Informatika*, vol. 4, no. 1, pp. 121–130, 2020, doi: 10.29408/edumatic.v4i1.2162.

[9] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.

[10] E. Sartika, "Analisis Metode K Nearest Neighbor Imputation (KNNI) untuk Mengatasi Data Hilang Pada Estimasi Data Survey," *Jurnal TEDC*, vol. 12, no. 3, pp. 219–227, 2018.

[11] N. Sari, Licantik, and M. Zahra, "Pemanfaatan Sistem Rekomendasi Menggunakan Content-Based Filtering pada Hotel di Palangka Raya," *Technological Journal Ilmiah*, vol. 15, no. 4, pp. 754–763, 2024.

[12] M. Waqar, N. Majeed, H. Dawood, A. Daud, and N. R. Aljohani, "An Adaptive Doctor-Recommender System," *Behavioral and Information Technology*, vol. 38, no. 9, pp. 959–973, 2019, doi: 10.1080/0144929X.2019.1625441.

[13] F. Nurfalah, Asriyanik, and A. Pambudi, "Sistem Rekomendasi Event Online Menggunakan Metode Content Based Filtering," *Jurnal Ilmiah Elektronika dan Komputer*, vol. 15, no. 2, pp. 271–279, 2022.
[14] H. Zakiyudin and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *Jurnal Bumigora Information Technology (BITe)*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
[15] A. Tandon, A. Dhir, A. K. M. N. Islam, and M. Mäntymäki, "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda," *Elsevier*, vol. 122, 2020, doi: 10.1016/j.compind.2020.103290.
[16] R. S. Pradini, M. Anshori, M. S. Haris, and P. Korespondensi, "Optimasi Weight Ahp Menggunakan Genetic Algorithm Untuk Rekomendasi Platform Media Sosial Sebagai Sarana Promosi Digital" *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 10, no. 7, pp. 1537–1544, 2023, doi: 10.25126/jtiik2023108011.
[17] A. A. Amin, "Mereduksi Error Prediksi Pada Sistem Rekomendasi Menggunakan Pendekatan Colalborative Filtering Berbasis Model Matrix Factorization," *Explore*, vol. 11, no. 2, p. 8, 2021, doi: 10.35200/explore.v11i2.434.
[18] S. Lestari, M. E. Afdila, and Y. A. Pratama, "Imputation Missing Value to Overcome Sparsity Problems in The Recommendation System," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 6, pp. 1285–1291, 2023, doi: 10.29207/resti.v7i6.5300.