PENINGKATAN SISTEM REKOMENDASI LAYANAN KESEHATAN MENGGUNAKAN FILTER KOLABORATIF BERBASIS PENGGUNA DENGAN IMPUTASI K-NEAREST NEIGHBORS UNTUK MENGATASI KEKURANGAN DATA

Mochammad Robby Sugara, Ahsanun Naseh Khudori, M. Syauqi Haris
Informatika, Institut Teknologi, Sains, dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW, Malang
Jl. S. Supriadi No.22, Kota Malang, Indonesia
Srobby109@gmail.com

ABSTRAK

Di era digital, aplikasi pemesanan layanan kesehatan telah menyederhanakan akses ke layanan medis secara signifikan, namun mereka menghadapi tantangan dalam memberikan rekomendasi yang dipersonalisasi karena masalah kelangkaan data. Penelitian ini bertujuan meningkatkan akurasi rekomendasi layanan kesehatan dengan mengatasi kelangkaan data melalui integrasi *User-Based Collaborative Filtering* (UBCF) dikombinasikan dengan metode imputasi *K-Nearest Neighbors* (KNN). Metodologi yang digunakan meliputi pembuatan dataset simulasi dengan berbagai tingkat kelangkaan data, melakukan *Preprocessing* data melibatkan normalisasi dan imputasi berbasis KNN, menetapkan bobot atribut melalui *Analytic Hierarchy Process* (AHP), dan mengimplementasikan UBCF menggunakan *Weighted Pearson Correlation*. Kinerja sistem rekomendasi dievaluasi secara kuantitatif menggunakan *metrik Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), divalidasi melalui prosedur validasi silang yang ekstensif. Hasil menunjukkan peningkatan signifikan dalam stabilitas prediksi akibat imputasi KNN, ditunjukkan oleh penurunan RMSE sebesar 7,89% (dari 0,633 menjadi 0,583). Meskipun MAE sedikit meningkat dari 0,502 menjadi 0,511, *trade-off* kecil ini diimbangi oleh peningkatan keandalan rekomendasi secara keseluruhan. Penelitian ini menyimpulkan bahwa integrasi imputasi data KNN dengan UBCF secara substansial meningkatkan kinerja dan konsistensi sistem rekomendasi dalam kondisi kelangkaan data. Pendekatan ini memberikan implikasi praktis bagi platform layanan kesehatan, memungkinkan rekomendasi yang lebih akurat dan berpusat pada pengguna, sehingga meningkatkan kepuasan dan pengalaman pengguna secara signifikan

Kata kunci : *scarsity data, user-based collaborative filtering, sistem rekomendasi, K-Nearest Neighbors.*

1. PENDAHULUAN

Perkembangan teknologi informasi di era digital telah memberikan dampak signifikan di berbagai bidang kehidupan, khususnya sektor layanan kesehatan. Adanya aplikasi layanan kesehatan berbasis online, seperti platform untuk reservasi janji temu dokter dan layanan medis lainnya, semakin mendapatkan perhatian luas [1].

Aplikasi pemesanan layanan kesehatan, seperti portal untuk menjadwalkan konsultasi dengan dokter atau memilih penawaran perawatan kesehatan, mendapatkan daya tarik[2]. Platform inovatif ini merampingkan perjalanan bagi individu yang mencari perawatan Kesehatan, memungkinkan mereka mengakses layanan dengan kecepatan dan kesederhanaan yang luar biasa[3].

Namun, dengan semakin banyaknya pilihan layanan dan variasi preferensi pengguna, muncul tantangan besar dalam menyediakan rekomendasi layanan yang sesuai kebutuhan masing-masing pengguna secara efektif[4].

Pengguna sering kesulitan dengan tantangan untuk menemukan layanan kesehatan yang sesuai dengan preferensi dan kondisi kesehatan mereka[5].

Dengan tidak adanya sistem rekomendasi yang bagus, pengguna bisa kesulitan dengan banyaknya pilihan, yang dapat mengurangi pengalaman dan kepuasan mereka secara keseluruhan saat menggunakan aplikasi[6].

Kurangnya saran yang disesuaikan dalam rekomendasi ini dapat memicu kebutuhan solusi yang ditingkatkan lebih baik untuk memberikan saran layanan yang lebih sesuai dengan kebutuhan pengguna[7].

Memilih pendekatan yang tepat sangat penting karena dapat mempengaruhi pada hasil yang ditampilkan. Salah satu solusi yang dapat menjawab permasalahan ini adalah dengan menggunakan algoritma *collaborative filtering (CF)*[8].

Sistem rekomendasi menjadi solusi strategis untuk menghadapi masalah ini. Metode rekomendasi yang paling umum digunakan adalah *Collaborative Filtering* (CF), yang mampu menghasilkan rekomendasi berdasarkan pola interaksi antar pengguna (*User-Based* CF) atau antar item (Item-Based CF). Metode ini memiliki keunggulan karena tidak bergantung pada atribut konten spesifik item, sehingga sangat fleksibel untuk berbagai jenis layanan. Algoritma CF merupakan pendekatan untuk menghasilkan rekomendasi berdasarkan kesamaan pengguna dengan pengguna lain (*user-based)* atau berdasarkan kesamaan antara item yang pernah pengguna gunakan atau pengguna lihat (*item-based)*[9].

Kelebihan dari algoritma *user-based collaborative filtering* (UBCF) yaitu tidak memerlukan informasi konten item sehingga cocok untuk rekomendasi berbagai jenis konten tanpa bergantung pada deskripsi atau fitur spesifik item, rekomendasi yang dihasilkan biasanya lebih sesuai dengan preferensi pengguna karena UBCF memanfaatkan data dari pengguna lain yang memiliki pola preferensi serupa[9].

Metode *CF* telah banyak digunakan di berbagai sektor, termasuk *e-commerce*, layanan *streaming*, dan aplikasi hiburan lainnya, untuk menyajikan rekomendasi yang relevan bagi pengguna[10].

Metode ini bekerja dengan memanfaatkan data historis perilaku pengguna untuk memprediksi kecenderungan mereka terhadap item tertentu. Namun, salah satu kelemahan utama metode CF, terutama *User-Based* CF (UBCF), adalah terjadinya kelangkaan data (data *scarsity*), yang mengakibatkan penurunan akurasi rekomendasi secara signifikan[11].

Data *scarsity* terjadi ketika interaksi pengguna-item tidak lengkap atau sangat sedikit, menyebabkan kesulitan dalam menemukan pola kesamaan antar pengguna atau antar item. Penelitian sebelumnya terkait rekomendasi layanan kesehatan berbasis CF belum sepenuhnya mengatasi secara mendalam permasalahan ini, khususnya belum ada integrasi efektif antara UBCF dengan metode imputasi data untuk menangani data sparsity dalam konteks layanan kesehatan.

sBerdasarkan penelitian sebelumnya dapat disimpulkan bahwa penggunaan algoritma CF menunjukkan keberhasilan metode ini di berbagai sektor seperti alat kesehatan, layanan apotek, pariwisata, hiburan, dan makanan. Algoritma CF efektif dalam memberikan rekomendasi yang akurat, relevan, dan mempermudah pengguna dalam pengambilan keputusan. Tingkat kepuasan pengguna yang tinggi serta hasil pengujian yang baik membuktikan bahwa CF adalah metode yang adaptif dan dapat diimplementasikan secara luas untuk memenuhi kebutuhan masyarakat. Namun, penelitian tersebut belum sepenuhnya mempertimbangkan permasalahan data *scarsity*. [12]

Masalah data *scarsity* menjadi kendala besar dalam sistem rekomendasi karena menyebabkan penurunan akurasi prediksi dan menyulitkan proses pencocokan antar pengguna. Beberapa pendekatan telah diusulkan untuk mengatasi masalah ini, salah satunya adalah dengan imputasi data menggunakan algoritma *K-Nearest Neighbors* (KNN). Metode ini mampu mengisi nilai-nilai yang hilang dengan memanfaatkan kemiripan antar pengguna atau item. Meskipun pendekatan ini telah diterapkan dalam beberapa kasus, penelitian mengenai efektivitas integrasi UBCF dengan KNN dalam konteks layanan kesehatan masih terbatas, khususnya dalam skenario dengan tingkat *scarsity* tinggi.

Berdasarkan kesenjangan literatur tersebut, penelitian ini secara khusus bertujuan untuk mengatasi permasalahan data sparsity dalam sistem rekomendasi aplikasi pemesanan layanan kesehatan dengan menerapkan algoritma UBCF yang dioptimalkan menggunakan metode imputasi data *K-Nearest Neighbors* (KNN). Berbeda dengan studi sebelumnya, novelty dari penelitian ini terletak pada penggunaan pendekatan integratif antara UBCF dan imputasi KNN secara simultan untuk meningkatkan stabilitas dan akurasi rekomendasi layanan kesehatan, khususnya dalam kondisi data yang sangat terbatas. Kontribusi ilmiah penelitian ini adalah memberikan pendekatan komprehensif dan sistematis dalam mengatasi permasalahan data sparsity secara langsung melalui simulasi *dataset dummy*, validasi ekstensif dengan teknik *cross-validation*, serta analisis statistik yang ketat untuk menguji signifikansi perbaikan akurasi rekomendasi. Diharapkan penelitian ini mampu memberikan panduan praktis dan teoritis untuk pengembangan lebih lanjut sistem rekomendasi layanan kesehatan yang lebih akurat, andal, dan sesuai kebutuhan pengguna.

2. TINJAUAN PUSTAKA

Dalam menyusun sistem rekomendasi layanan kesehatan yang efektif dan adaptif, penting untuk memahami terlebih dahulu landasan konseptual dan teknis dari berbagai komponen yang mendukung pengembangannya. Penelitian ini dibangun di atas sejumlah teori dan pendekatan yang telah digunakan secara luas dalam domain sistem rekomendasi, termasuk algoritma Collaborative Filtering, permasalahan sparsity data, serta teknik optimasi seperti imputasi dan pembobotan atribut. Oleh karena itu, pada bagian ini akan dijelaskan beberapa konsep utama yang menjadi dasar penelitian, dimulai dari pemahaman mengenai sistem rekomendasi secara umum.

2.1. Sistem Rekomendasi

Sistem rekomendasi adalah sistem yang dirancang untuk menyarankan item yang relevan kepada pengguna berdasarkan preferensi mereka. Sistem ini telah diterapkan secara luas di berbagai domain seperti e-commerce, media sosial, pendidikan, dan kesehatan[13].

Tujuannya adalah membantu pengguna menavigasi informasi yang melimpah dengan memberikan saran yang personal dan relevan.

2.2. Collaborative Filtering

Collaborative Filtering (CF) merupakan pendekatan yang banyak digunakan dalam sistem rekomendasi. CF bekerja dengan menganalisis pola interaksi antara pengguna dan item, tanpa membutuhkan informasi eksplisit tentang atribut item tersebut [14].

CF terbagi menjadi dua kategori utama, yaitu User-Based CF dan Item-Based CF. Kedua metode ini bekerja berdasarkan asumsi bahwa pengguna yang

memiliki preferensi serupa di masa lalu cenderung menyukai item yang sama di masa depan.

2.3. User-Based Collaborative Filtering (UBCF)

User-Based Collaborative Filtering (UBCF) merupakan metode CF yang merekomendasikan item berdasarkan kesamaan antar pengguna[15].

Kelebihan UBCF adalah kemampuannya untuk menghasilkan rekomendasi yang bersifat personal, karena didasarkan pada pengalaman nyata dari pengguna lain dengan preferensi yang serupa. Namun, kelemahannya muncul ketika terdapat sedikit data interaksi antar pengguna, sehingga menghambat perhitungan kesamaan yang akurat.

2.4. Permasalahan Data Sparsity

Data sparsity atau kelangkaan data adalah kondisi di mana matriks pengguna-item memiliki banyak nilai kosong (missing values). Hal ini umum terjadi dalam sistem rekomendasi skala besar, di mana sebagian besar pengguna hanya berinteraksi dengan sebagian kecil item. Masalah ini menyebabkan sistem kesulitan menemukan kemiripan yang akurat antar pengguna atau antar item, sehingga menurunkan akurasi rekomendasi[16].

2.5. Metode Imputasi Data dengan K-Nearest Neighbors (KNN)

Untuk mengatasi data sparsity, salah satu metode yang digunakan adalah imputasi data menggunakan K-Nearest Neighbors (KNN). Metode ini mengisi nilai yang hilang dengan menggunakan informasi dari tetangga terdekat yang memiliki karakteristik atau preferensi serupa . Imputasi menggunakan KNN mampu meningkatkan densitas matriks pengguna-item sehingga perhitungan kesamaan antar pengguna menjadi lebih akurat. Beberapa penelitian menunjukkan bahwa pendekatan ini secara signifikan dapat meningkatkan performa sistem rekomendasi [17].

2.6. Analytic Hierarchy Process (AHP)

Analytic Hierarchy Process (AHP) adalah metode pengambilan keputusan multikriteria yang digunakan untuk menentukan bobot atribut berdasarkan tingkat kepentingannya. Dalam konteks sistem rekomendasi, AHP membantu menyusun prioritas fitur (seperti preferensi spesialisasi, gender, biaya, dan rating) berdasarkan perbandingan berpasangan, sehingga hasil rekomendasi menjadi lebih relevan dengan kebutuhan pengguna[18].

2.7. Evaluasi Sistem Rekomendasi

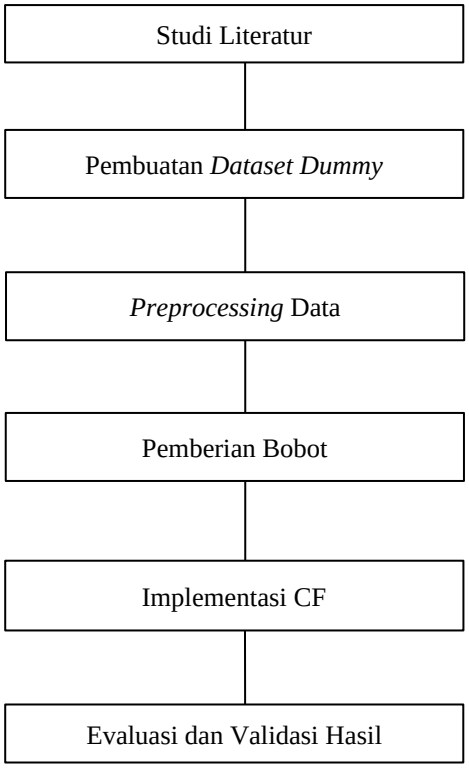
Evaluasi sistem rekomendasi umumnya dilakukan menggunakan dua metrik utama, yaitu Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). MAE mengukur rata-rata kesalahan absolut antara prediksi dan nilai aktual, sedangkan RMSE memberikan penalti yang lebih besar terhadap prediksi yang jauh dari nilai aktual[11]. Kedua metrik ini banyak digunakan karena mampu memberikan gambaran akurasi sistem secara kuantitatif

3. METODE PENELITIAN

Metode penelitian adalah suatu cara atau pendekatan sistematis yang digunakan untuk mengumpulkan, menganalisis, dan menginterpretasikan data guna menjawab pertanyaan penelitian atau mencapai tujuan tertentu[14].

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode simulasi berbasis data *dummy*. Tujuannya adalah untuk mengembangkan dan mengevaluasi sistem rekomendasi layanan kesehatan berbasis *User-Based Collaborative Filtering* (UBCF) yang dioptimalkan dengan teknik imputasi data *scarsity* menggunakan *K-Nearest Neighbors* (KNN).

Tahap-tahap penelitian ini ditunjukkan pada Gambar 1. Tahap penelitian dimulai dari tahap studi literatur, pembuatan *dataset dummy*, *preprocessing* data, pemberian bobot, implementasi *collaborative filtering*, evaluasi dan validasi hasil.



Gambar 1. Tahap-tahap penelitian

3.1. Studi Literatur

Pada tahap studi literatur bertujuan untuk mengumpulkan informasi dari berbagai sumber ilmiah yang relevan. Sumber informasi bisa didapatkan dari jurnal, buku, artikel, dan sumber referensi lainnya. Tujuannya adalah untuk memahami metode yang sudah ada dan mencari solusi terbaik yang dapat diterapkan dalam studi. Studi literatur penting untuk memastikan bahwa penelitian baru dibangun di atas

pengetahuan yang ada dan tidak mengulang pekerjaan yang sudah dilakukan, serta membantu dalam merumuskan hipotesis yang lebih tepat dan metodologi yang lebih efektif.

3.2. Pembuatan *Dataset Dummy*

Penggunaan data *dummy* dipilih karena data *dummy* bersifat sangat fleksibel, memungkinkan peneliti untuk menyesuaikan data dengan kebutuhan spesifik penelitian, termasuk mengontrol variabel-variabel tertentu yang relevan. Dalam penelitian ini, teknik pengambilan sampel tidak dilakukan terhadap data nyata atau populasi pengguna sebenarnya, melainkan melalui simulasi data (Data *Dummy*). Oleh karena itu, pendekatan sampling dilakukan dengan cara merancang skenario data secara sintetik menggunakan parameter-parameter tertentu. Peneliti secara sengaja menentukan proporsi data yang lengkap dan data yang kosong untuk mensimulasikan kondisi data *scarcity*, sehingga proses evaluasi sistem rekomendasi dapat dilakukan secara terkontrol dan terukur.

Data yang digunakan dalam penelitian ini tidak berasal dari sistem atau pengguna nyata, tetapi dikembangkan secara artifisial untuk merepresentasikan interaksi pengguna dengan item. Pada data *dummy*, berbagai skenario dapat dibuat untuk menguji batasan metode yang digunakan dengan lebih netral untuk analisis. Hal ini mempermudah proses eksplorasi dan pengujian awal algoritma sebelum diterapkan pada data riil. berdasarkan literatur, penggunaan data *dummy* diperbolehkan dalam tahap pengembangan untuk simulasi dan validasi algoritma, seperti yang disebutkan dalam penelitian tandon, 2020[19].

Untuk memastikan bahwa data *dummy* yang digunakan dapat secara efektif mensimulasikan kondisi yang diinginkan, perlu diperhatikan beberapa atribut penting yang harus dimasukkan dalam data tersebut. Pada Tabel 1 dibawah ini adalah atribut dari data yang akan digunakan pada penelitian ini.

Tabel 1. Perhitungan Evaluasi dengan Imputasi

Atribut	Deskripsi	Tipe data
Pasien_ID	Id Unik Pasien	Nominal
Dokter_ID	Id Unik Dokter	Nominal
Preferensi Spesialis	Spesialis Dokter	Nominal
Preferensi Gender	Gender Dokter	Nominal
Biaya Konsultasi	Biaya Konsultasi	Rasio
Rating	Rating Dokter	Ordinal

3.3. Preprocessing Data

Pada tahap ini, dilakukan preprocessing data untuk mempersiapkan data yang digunakan dalam penelitian. Langkah pertama adalah normalisasi data, yang bertujuan untuk menyelaraskan nilai pada atribut tertentu ke dalam rentang yang seragam. Proses normalisasi sangat penting untuk memastikan bahwa skala nilai dalam dataset tidak terlalu besar atau terlalu kecil, sehingga dapat mengoptimalkan kinerja algoritma *K-Nearest Neighbors* (KNN) yang digunakan dalam sistem rekomendasi. Normalisasi membantu model bekerja lebih efisien dengan menghindari dominasi fitur tertentu akibat skala yang berbeda.

Setelah normalisasi, langkah kedua adalah imputasi data *scarsity* menggunakan metode *K-Nearest Neighbors* (KNN) untuk menangani missing values pada data sparse. Metode ini menggantikan nilai yang hilang dengan cara mencari k tetangga terdekat yang memiliki nilai serupa dan menghitung nilai imputasi berdasarkan rata-rata atau nilai mayoritas dari tetangga tersebut. Imputasi ini bertujuan untuk mengatasi masalah data *scarsity*, yaitu kondisi di mana sebagian besar nilai dalam *matriks* pengguna-item kosong atau tidak memiliki rating. Masalah ini sering terjadi dalam sistem rekomendasi, terutama pada dataset besar dengan banyak pengguna dan item.

Dalam konteks CF, metode KNN dapat digunakan untuk mengestimasi rating yang hilang dengan memanfaatkan pola kesamaan antar pengguna atau antar item. Sebelum data digunakan untuk analisis, nilai yang hilang perlu diisi agar model dapat bekerja dengan baik. Untuk atribut biaya konsultasi, nilai kosong diisi dengan memanfaatkan KNN untuk mencari nilai yang paling relevan berdasarkan kesamaan dengan data lain dalam dataset. Dengan langkah preprocessing ini, sistem rekomendasi dapat memberikan prediksi yang lebih akurat dan meningkatkan kualitas hasil rekomendasi.

3.4. Pemberian Bobot

Pada tahap ini, bobot diberikan kepada atribut-atribut yang digunakan dalam sistem rekomendasi berdasarkan tingkat kepentingannya untuk memastikan relevansi dan akurasi rekomendasi yang dihasilkan. Atribut preferensi spesialis diberikan bobot yang lebih tinggi karena langsung mencerminkan kebutuhan pasien dengan keahlian dokter. Penentuan bobot dilakukan dengan pendekatan heuristik menggunakan algoritma *Analytic Hierarchy Process* (AHP), yang melibatkan perbandingan berpasangan antar atribut untuk menentukan prioritas relatif[18]. Hasil perbandingan ini diolah menjadi *matriks* perbandingan, dan bobot dihitung berdasarkan nilai *eigen* dari *matriks* tersebut. Pendekatan ini memastikan bahwa penetapan bobot dilakukan secara logis dan terstruktur sehingga sistem rekomendasi dapat menghasilkan hasil yang relevan dan sesuai dengan kebutuhan pengguna.

3.5. Implementasi *Collaborative Filtering*

Pada tahap ini, algoritma UBCF diimplementasikan untuk mengembangkan sistem rekomendasi yang memberikan saran dokter berdasarkan pola kesamaan preferensi antar pengguna. Pendekatan ini bertujuan untuk mengidentifikasi

pengguna lain yang memiliki pola interaksi serupa, sehingga rekomendasi dapat lebih relevan dan sesuai dengan preferensi pengguna target. Implementasi dimulai dengan membangun *matriks* pengguna-dokter yang berisi informasi tentang interaksi antara pengguna dan dokter. *Matriks* ini mencakup data seperti rating yang diberikan oleh pengguna terhadap dokter. Jika terdapat nilai yang kosong, data tersebut sebelumnya telah diisi melalui proses imputasi pada tahap *preprocessing*.

Selanjutnya, dilakukan perhitungan kesamaan antar pengguna untuk menentukan tingkat kemiripan preferensi. Perhitungan kesamaan ini menggunakan metode seperti *Weighted Pearson Correlation* (WPC), yang mengevaluasi hubungan antara pola interaksi pengguna berdasarkan rating mereka terhadap dokter dengan mempertimbangkan bobot yang dihasilkan dari perhitungan AHP. Hasil dari perhitungan ini digunakan untuk menemukan pengguna lain yang memiliki preferensi serupa dengan pengguna target. Berdasarkan hasil perhitungan kesamaan, sistem merekomendasikan dokter yang sering dipilih atau diberikan rating tinggi oleh pengguna dengan kesamaan preferensi tinggi. Berikut ini merupakan persamaan dari *Weighted Pearson Correlation* pada rumus 1.

$$r_w = \frac{\sum w_i(x_i - \underline{x}_w)(y_i - \underline{y}_w)}{\sqrt{\sum w_i(x_i - \underline{x}_w)^2}\sqrt{\sum w_i(y_i - \underline{y}_w)^2}} \quad (1)$$

Dengan w_i adalah bobot untuk pasangan data ke- i, x_i, y_i adalah nilai dari pasangan data ke- i, $\underline{x}_w$ adalah rata-rata berbobot dari x, $\underline{y}_w$ adalah rata-rata berbobot dari y.

3.6. Evaluasi dan Validasi Hasil

Setelah algoritma rekomendasi diterapkan, evaluasi dilakukan untuk menilai performa model menggunakan beberapa *matriks*. Salah satu matrik yang digunakan adalah *Mean Absolute Error* (MAE), yang menghitung rata-rata kesalahan absolut rekomendasi yang diberikan oleh algoritma UBCF. *Matriks* ini digunakan untuk mengukur perbedaan antara nilai prediksi yang dihasilkan oleh sistem dengan nilai sebenarnya yang diberikan oleh pengguna. Persamaan MAE diberikan pada rumus 2.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |yi - \tilde{y}i| \quad (2)$$

Dengan yi adalah nilai aktual pada data ke- i, $\hat{y}i$ adalah nilai prediksi pada data ke- i, $|\ yi - \hat{y}i\ |$ adalah selisih absolut antara nilai aktual dan prediksi untuk setiap data, $\frac{1}{n}$ adalah resiprokal dari jumlah data yang dipakai untuk menormalkan jumlah kesalahan agar menjadi rata-rata. Jadi, indeks i dijumlahkan dari 1 hingga n, yaitu total jumlah sampel dalam *dataset* yang digunakan untuk evaluasi.

Selanjutnya, *Root Mean Squared Error* (RMSE) adalah ukuran yang digunakan untuk mengevaluasi seberapa baik model prediktif dalam merepresentasikan data aktual [7]. Berikut ini merupakan persamaan dari RMSE pada rumus 3.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \tilde{y}i)^2} \quad (3)$$

Dengan n adalah jumlah total data, yi adalah nilai aktual pada data ke- i, $\tilde{y}i$ adalah nilai prediksi pada data ke- i, $(yi - \tilde{y}i)^2$ adalah selisih antara nilai aktual dan prediksi yang dikuadratkan, $\frac{1}{n}$ digunakan untuk menghitung rata-rata dari jumlah kesalahan kuadrat.

Selain evaluasi dengan *matriks* di atas, sistem rekomendasi diuji dengan skenario data *scarsity* untuk memahami dampak kelangkaan data terhadap performa sistem. Sistem ini membandingkan hasil performa dengan dan tanpa proses imputasi data untuk melihat sejauh mana pengisian nilai yang hilang dapat meningkatkan akurasi prediksi. Evaluasi juga dilakukan menggunakan teknik 5-*fold cross-validation*, di mana *dataset* dibagi menjadi 5 bagian (*folds*) yang digunakan secara bergantian sebagai data uji dan data latih. Pada setiap iterasi, 4-*fold* digunakan sebagai data latih dan 1 *fold* sebagai data uji, sehingga model dievaluasi sebanyak 5 kali dengan kombinasi data yang berbeda. Setiap iterasi menghasilkan nilai evaluasi berupa MAE dan RMSE yang kemudian dirata-rata untuk mendapatkan performa keseluruhan model. Teknik ini digunakan untuk mengurangi kemungkinan *overfitting* dan memastikan model memiliki performa yang stabil serta tidak bergantung pada pembagian data tertentu. Dengan jumlah *fold* yang lebih banyak, evaluasi menjadi lebih menyeluruh, karena setiap bagian data lebih sering digunakan untuk pelatihan dan pengujian, sehingga hasil evaluasi lebih akurat dan mencerminkan kemampuan generalisasi model dalam kondisi data yang bervariasi.

Dari semua metode yang digunakan dapat disimpulkan bahwa UBCF dipilih karena mampu memberikan rekomendasi yang relevan tanpa memerlukan informasi konten item, cocok untuk layanan kesehatan yang item-nya kompleks. KNN untuk imputasi dipilih karena kesederhanaannya, efektivitas pada data scarsity, dan kemampuannya menjaga distribusi data asli. AHP digunakan karena mampu memberikan dasar logis dalam penentuan bobot atribut berdasarkan pertimbangan multikriteria. MAE dan RMSE dipilih karena merupakan standar evaluasi akurasi prediksi dalam sistem rekomendasi.

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi berbasis algoritma UBCF dengan mempertimbangkan faktor *scarsity* data. Hasil dari implementasi metodologi yang dirancang disajikan dalam beberapa bagian yang menjelaskan pembuatan

dataset dummy, *preprocessing* data, pemberian bobot atribut, implementasi algoritma UBCF, serta evaluasi dan validasi hasil. Pembahasan dimulai dengan pemaparan mengenai *dataset dummy* yang digunakan.

4.1. Hasil Pembuatan *Dataset* Dummy

Pembuatan *dataset dummy* dilakukan untuk mensimulasikan kondisi data yang sesuai dengan kebutuhan penelitian. *Dataset* ini dirancang dengan atribut utama seperti Pasien_ID, Dokter_ID, Spesialisasi Dokter, Preferensi *Gender*, Biaya Konsultasi, dan Rating. *Dataset dummy* ini dirancang sedemikian rupa sehingga mencakup 2000 data untuk memberikan cakupan yang cukup dalam pengujian algoritma. *Dataset* ini juga memiliki fleksibilitas untuk mengatur tingkat *scarsity* sesuai skenario penelitian. Contoh *dataset dummy* sebelum dilakukan proses imputasi ditampilkan pada Tabel 2, yang memperlihatkan data *scarsity* dalam atribut tertentu. Hasilnya adalah *dataset* dengan berbagai tingkat *scarsity* yang memungkinkan pengujian algoritma secara komprehensif.

Tabel 2. Contoh dataset sebelum imputasi

pasienid	dokterid	Spesialis	Gender	Biaya	Rating
P001	D001	mata	Pria	20000	3
P002	D002	-	Wanita	50000	5
P003	D003	THT	Pria	-	3
P004	D004	mata	-	20000	-

Dataset ini menggambarkan interaksi antara pasien dan dokter dalam konteks sistem rekomendasi layanan kesehatan. Setiap baris mewakili data pasien, sementara kolom-kolom berisi informasi seperti dokterid, spesialisasi, jenis kelamin, biaya layanan, dan rating yang diberikan oleh pasien terhadap layanan tersebut. *Dataset* ini sengaja dirancang dengan nilai-nilai kosong atau tidak lengkap (*missing values*) untuk mensimulasikan kondisi data *sparsity* yang umum terjadi pada sistem rekomendasi nyata.

4.2. Hasil Preprocessing Data

Pada tahap *preprocessing*, dilakukan proses normalisasi data dilakukan untuk menyamakan skala antar atribut, termasuk atribut kategorikal seperti preferensi spesialisasi dan preferensi *gender*, yang terlebih dahulu dikonversi menjadi data numerik menggunakan teknik *one-hot encoding*. Normalisasi dilakukan menggunakan metode *Min-Max Scaling* untuk meningkatkan akurasi perhitungan kesamaan. Kemudian, dilakukan imputasi data *scarsity* menggunakan KNN. Hasil *preprocessing* yang menunjukkan *dataset* setelah dilakukan imputasi dapat dilihat pada Tabel 3.

Tabel 3. Contoh dataset setelah imputasi

pasienid	dokterid	Spesialis	Gender	Biaya	Rating
P001	D001	mata	Pria	20000	3
P002	D002	mata	Wanita	50000	5
P003	D003	THT	Pria	20000	3
P004	D004	mata	Pria	20000	3,6

4.3. Hasil Pemberian Bobot

Bobot atribut ditentukan menggunakan metode *Analytic Hierarchy Process* (AHP). Berdasarkan hasil *matriks* perbandingan, atribut Preferensi Spesialis mendapat bobot tertinggi sebesar 0,48 diikuti oleh rating sebesar 0,22, Preferensi gender sebesar 0,15 dan biaya konsultasi sebesar 0,08. Detail bobot masing-masing atribut dapat dilihat pada Tabel 4, yang mencerminkan prioritas kebutuhan pasien dalam memilih dokter. Pemberian bobot yang dilakukan telah diuji konsistensinya melalui penghitungan rasio konsistensi, dan hasil menunjukkan bahwa nilai tersebut berada dalam batas yang dapat diterima. Dengan demikian, bobot yang dihasilkan dapat dianggap valid dan sesuai dengan prinsip AHP.

Tabel 4. Bobot Atribut

Atribut	Bobot
Preferensi Spesialis	0,48
Rating	0,22
Preferensi Gender	0,15
Biaya Konsultasi	0,08

Hasil perhitungan menunjukkan bahwa Preferensi Spesialis memiliki bobot tertinggi sebesar 0,48, diikuti oleh Rating (0,22), Preferensi Gender (0,15), dan Biaya Konsultasi (0,08). Ini menunjukkan bahwa pengguna lebih mengutamakan spesialisasi dokter dibandingkan faktor lainnya. Perhitungan rasio konsistensi juga menunjukkan nilai yang masih dalam batas wajar (< 0,1), sehingga bobot yang diperoleh dianggap valid dan dapat digunakan dalam sistem rekomendasi.

4.4. Hasil Implementasi Collaborative filtering

Algoritma UBCF berhasil diimplementasikan. *Matriks* pasien-dokter yang telah diisi melalui imputasi data *scarsity* menjadi dasar untuk perhitungan kesamaan antar pengguna. Hasil perhitungan kesamaan menggunakan *Weighted Pearson Correlation* ditampilkan pada Tabel 5, yang menunjukkan bahwa pengguna dengan preferensi serupa dapat diidentifikasi dengan akurasi yang baik.

Tabel 5. Perhitungan Korelasi

Pasien_ID1	Pasien_ID2	Korelasi
P001	P004	0,998
P002	P003	0,575

Sistem rekomendasi dokter yang memiliki rating tinggi oleh pengguna dengan kesamaan preferensi tinggi. Kesimpulan dari sistem rekomendasi ini dapat dilihat pada Tabel 6, yang menyajikan dokter yang direkomendasikan untuk setiap pasien berdasarkan kesamaan preferensi.

Tabel 6. Kesimpulan Rekomendasi

Pasien_ID	Rekomendasi
P001	D004
P002	D003

4.5. Hasil Evaluasi dan Validasi

Evaluasi menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) menunjukkan bahwa algoritma memiliki performa yang stabil, terutama setelah dilakukan proses imputasi pada data yang bersifat *scarse*. Untuk memahami dampak imputasi terhadap akurasi prediksi, evaluasi dilakukan dalam dua skenario: dengan imputasi dan tanpa imputasi. Hasil evaluasi dengan proses imputasi disajikan pada Tabel 7, yang menunjukkan bahwa meskipun nilai MAE mengalami sedikit peningkatan, RMSE justru menurun, yang mengindikasikan bahwa model dengan imputasi menghasilkan prediksi yang lebih konsisten dan tidak memiliki *error* yang terlalu ekstrem.

Tabel 7. Perhitungan Evaluasi dengan Imputasi

Evaluasi	MAE	RMSE
1	0,508672	0,579031
2	0,510320	0,583474
3	0,507323	0,583579
4	0,511054	0,585414
5	0,518178	0,587166
Rata-rata	0,511110	0,583733

Sebagai perbandingan, evaluasi juga dilakukan dalam skenario tanpa proses imputasi, seperti ditampilkan pada Tabel 8. Terlihat bahwa meskipun MAE sedikit lebih kecil dibandingkan dengan skenario imputasi, RMSE lebih tinggi, yang menunjukkan adanya variasi *error* yang lebih besar dalam prediksi.

Tabel 8. Perhitungan Evaluasi tanpa Imputasi

Evaluasi	MAE	RMSE
1	0,503571	0,635871
2	0,501013	0,629600
3	0,491519	0,627834
4	0,499269	0,631284
5	0,516224	0,641313
Rata-rata	0,502319	0,633180

Dari hasil di atas, dapat disimpulkan bahwa imputasi membantu meningkatkan kestabilan prediksi, dengan penurunan nilai RMSE, yang menunjukkan bahwa model lebih mampu menghasilkan prediksi yang tidak terlalu menyimpang secara ekstrem. Dengan kata lain, meskipun MAE sedikit lebih tinggi, proses imputasi membantu model menjadi lebih konsisten dan andal dalam memberikan rekomendasi. Selain itu, model tetap mampu bekerja dengan baik dalam kondisi data yang memiliki keterbatasan interaksi.

Tabel 9. Perhitungan *Evaluation Results with K-Fold*

Evaluasi	MAE	RMSE
1	0,535015	0,602637
2	0,517619	0,590950
3	0,487942	0,558901
4	0,497445	0,573096
5	0,505081	0,570635
6	0,508738	0,574765
7	0,508826	0,588713
8	0,535508	0,611174
9	0,527692	0,591713
10	0,491346	0,572846
11	0,503293	0,568925
12	0,495478	0,567377
13	0,495493	0,575823
14	0,504063	0,575398
15	0,498208	0,573934
16	0,531884	0,609865
17	0,518417	0,597879
18	0,496494	0,567495
19	0,506754	0,581590
20	0,540453	0,616577
21	0,506943	0,586772
22	0,496771	0,575128
23	0,486273	0,557677
24	0,532473	0,599386
25	0,506840	0,573687
26	0,505215	0,575767
27	0,523508	0,590227
28	0,543850	0,613302
29	0,514531	0,584482
30	0,511389	0,579107
Rata-rata	0,511118	0,583528

Tabel 9. Perhitungan *Evaluation Results with K-Fold*

Evaluasi	MAE	RMSE
1	0,558217	0,677092
2	0,512567	0,647345
3	0,470543	0,605882
4	0,489615	0,621344
5	0,490062	0,627699
6	0,497934	0,632028
7	0,487291	0,622328
8	0,536913	0,659123
9	0,538734	0,659993
10	0,463088	0,596846
11	0,498800	0,624398
12	0,486735	0,617497
13	0,461408	0,605251
14	0,486848	0,623117
15	0,470459	0,605201
16	0,531599	0,658124
17	0,522508	0,660831
18	0,469575	0,605170
19	0,491319	0,620781
20	0,567808	0,690192
21	0,485459	0,626078
22	0,473909	0,609028
23	0,453016	0,584435
24	0,534677	0,662738
25	0,491979	0,621989
26	0,490387	0,619732
27	0,533969	0,653909
28	0,556099	0,679584
29	0,505319	0,625657
30	0,513187	0,637189
Rata-rata	0,502334	0,632686

Tabel 9 dan 10 menyajikan evaluasi yang lebih rinci mengenai kinerja model menggunakan validasi silang k-lipat (k=30) untuk memberikan penilaian yang kuat terhadap efektivitas algoritma *User-Based Collaborative Filtering* (*UBCF*), baik dengan (Tabel

9) maupun tanpa (Tabel 10) imputasi data yang hilang. Hasilnya, yang dirata-ratakan selama 30 lipatan, mengungkapkan *Mean Absolute Error* (*MAE*) sebesar 0,511 untuk data yang diimputasi dan 0,502 untuk data yang tidak diimputasi, dengan nilai *Root Mean Squared Error* (*RMSE*) yang sesuai sebesar 0,583 dan 0,633. Meskipun nilai *MAE* dan *RMSE* ini menunjukkan tingkat kesalahan yang perlu diperhatikan, penting untuk menganalisisnya dalam konteks tantangan yang melekat pada *dataset dummy* yang digunakan. *Dataset* ini dicirikan oleh kelangkaan data yang tinggi akibat pengenalan data yang hilang secara sengaja, yang memperumit prediksi akurat karena algoritma dipaksa membuat estimasi dengan data interaksi pengguna-item yang terbatas. Selain itu, *dataset* ini memiliki keragaman fitur yang terbatas, dengan hanya sejumlah kecil atribut seperti spesialisasi dokter, jenis kelamin, biaya, dan peringkat, yang dapat membatasi kemampuan model untuk membedakan pola preferensi pengguna yang bernuansa. Terakhir, penggunaan *dataset dummy*, meskipun memungkinkan eksperimen terkontrol, mungkin tidak sepenuhnya mencerminkan kompleksitas dan variabilitas data preferensi layanan kesehatan di dunia nyata. Oleh karena itu, kesalahan yang diamati harus diinterpretasikan dengan mempertimbangkan batasan-batasan ini, serta dalam kaitannya dengan skala peringkat yang digunakan. Penting untuk dicatat bahwa penurunan *RMSE* dari 0,633 dalam skenario tanpa imputasi menjadi 0,583 dengan imputasi menunjukkan bahwa metode imputasi *KNN* berkontribusi pada stabilisasi prediksi dengan mengurangi dampak kesalahan prediksi yang ekstrem. Namun, besarnya kesalahan secara keseluruhan menyoroti perlunya penyempurnaan model lebih lanjut.

Untuk menilai secara ketat signifikansi statistik dari perbedaan kinerja model, kami melakukan serangkaian uji coba acak dan uji t sampel berpasangan, seperti yang disarankan oleh reviewer. Dataset dibagi secara acak menjadi set pelatihan dan pengujian sebanyak 30 kali, dan nilai MAE dan RMSE dicatat untuk setiap iterasi. Uji t sampel berpasangan dilakukan pada 30 pasang nilai MAE dan RMSE, menghasilkan hasil sebagai berikut: Uji t untuk MAE menghasilkan statistik t = 2,021 dengan derajat bebas (df) = 29 dan nilai p = 0,053. Uji t untuk RMSE menghasilkan statistik t = 9,894 dengan derajat bebas (df) = 29 dan nilai p < 0,001. Hasil ini menunjukkan bahwa tidak ada perbedaan yang signifikan secara statistik dalam MAE (p = 0,053 > 0,05), namun terdapat perbedaan yang signifikan secara statistik dalam RMSE (p < 0,001), yang mengkonfirmasi bahwa imputasi secara konsisten mengurangi variabilitas kesalahan. Analisis ini memberikan dasar statistik yang lebih kuat untuk kesimpulan kami dan menjawab kekhawatiran tentang potensi pengaruh kebetulan acak. Penelitian di masa depan harus fokus pada penanganan keterbatasan dataset dan eksplorasi teknik pemodelan yang lebih canggih untuk meningkatkan akurasi dan penerapan model dalam sistem rekomendasi layanan kesehatan di dunia nyata.

5. KESIMPULAN DAN SARAN

User-Based Collaborative Filtering (UBCF) dan imputasi data menggunakan *K-Nearest Neighbors* (KNN). Eksperimen pada *dataset dummy* (2.000 interaksi pasien-dokter) menunjukkan bahwa metode imputasi KNN berhasil menurunkan *Root Mean Squared Error* (RMSE) sebesar 7,89 % (dari 0,633 menjadi 0,583), meski *Mean Absolute Error* (MAE) meningkat sedikit dari 0,502 menjadi 0,511. Uji statistik *paired-sample t-test* mengonfirmasi bahwa penurunan RMSE bersifat signifikan (p < 0,001), menandakan prediksi yang lebih konsisten dan minim *outlier*.

DAFTAR PUSTAKA

[1] K. Ramadahni, "Penerapan Teknologi Blockchain dalam Sistem Manajemen Kesehatan Elektronik," *J. Sos. Teknol.*, vol. 4, no. 2, pp. 116–119, 2024, doi: 10.59188/jurnalsostech.v4i2.1097.

[2] E. Sumantri and Y. Maulana, "Penerimaan Teknologi Kesehatan Masyarakat ALODOKTER Menggunakan Metode Technology Acceptance Model (TAM)," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 7, no. 1, pp. 227–236, 2024, doi: 10.31539/intecoms.v7i1.7173.

[3] N. Lisnarini, J. R. Suminar, and Y. Setianti, "Keunggulan dan Hambatan Komunikasi dalam Layanan Kesehatan Mental pada Aplikasi Telemedicine Halodoc," *PsikobuletinBuletin Ilm. Psikol.*, vol. 4, no. 3, p. 176, 2023, doi: 10.24014/pib.v4i3.25231.

[4] A. Rizky Mangunsong, V. Sihombing, and I. Rasyid Munthe, "Pengembangan Sistem Rekomendasi Produk Berdasarkan Pola Pembelian dengan Pendekatan Algoritma Apriori," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 82–86, 2024, doi: 10.55338/jikomsi.v7i1.2718.

[5] M. Sediono and S. Kusumadewi, "Analisis User Acceptance Pada Aplikasi Layanan Kesehatan Online di Jawa Tengah dan Daerah Istimewa Yogyakarta," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 203–218, 2022, doi: 10.35957/jatisi.v9i1.1463.

[6] R. Gaputra and A. Sidiq Purnomo, "Sistem Rekomendasi Dompet Digital Menggunakan Metode Simple Additive Weighting (Saw)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1775–1782, 2024, doi: 10.36040/jati.v8i2.9154.

[7] N. Fitroh, "Peran Teknologi Disruptif dalam Transformasi Perbankan dan Keuangan Islam," *MUSYARAKAH J. Sharia Econ.*, vol. 3, no. 1, pp. 76–90, 2023, [Online]. Available: http://journal.umpo.ac.id/index.php/musyarakah.

[8] A. Riyan and D. Ramayanti, "A Model-based

Music Recommender System using Collaborative Filtering Technique," *Int. J. Comput. Appl.*, vol. 181, no. 36, pp. 1–4, 2019, doi: 10.5120/ijca2019918320.

[9] S. R. Anggraeni, D. Purwitasari, C. Fatichah, and Y. Yustiawan, "Rekomendasi Produk Berbasis Collaborative Filtering Menggunakan Factorization Machine Graph Convolutional Networks," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 5, no. 2, pp. 1–11, 2023, doi: 10.28926/ilkomnika.v5i2.556.

[10] M. R. A. Zayyad and A. Kurnawardhani, "Penerapan Metode Deep Learning pada Sistem Rekomendasi Film," *Automata*, vol. 2, no. 1, pp. 206–210, 2021.

[11] I. W. Jepriana and S. Hanief, "Metode Item-Based Collaborative Filtering Untuk Model Sistem Rekomendasi Konsentrasi Penjurusan Di Stmik Stikom Bali," *J. Teknol. Inf. dan Komput.*, vol. 6, no. 1, pp. 20–29, 2020, doi: 10.36002/jutik.v6i1.1000.

[12] S. Devarajan and A. C. Fisher, "Exploration and scarcity.," *J. Polit. Econ.*, vol. 90, no. 6, pp. 1279–1290, 1982, doi: 10.1086/261121.

[13] H. Februariyanti, A. Dwi Laksono, J. Sasongko Wibowo, and M. Siswo Utomo, "Implementasi Metode Collaborative Filtering Untuk Sistem Rekomendasi Penjualan Pada Toko Mebel," *Khatulistiwa Inform.*, vol. 9, no. 1, pp. 43–45, 2021.

[14] E. Anugerah Rahayu Kasim, N. Ransi, and J. Teknik Informatika, "Sistem Rekomendasi Produk UMKM Menggunakan Algoritma User-Based Collaborative Filtering Berbasis Website Website-Based MSME Product Recommendation System Using User-Based Collaborative Filtering Algorithm," vol. 14, no. 2, pp. 152–162, 2024, [Online]. Available: https://stmikpontianak.org/ojs/index.php/sisfotenika

[15] M. I. Wardah and S. D. Putra, "Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen IMDB," *J. Manajamen Inform. Jayakarta*, vol. 2, no. 3, p. 243, 2022, doi: 10.52362/jmijayakarta.v2i3.868.

[16] Q. Y. Shambour, M. M. Al-Zyoud, A. H. Hussein, and Q. M. Kharma, "A doctor recommender system based on collaborative and content filtering," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 884–893, 2023, doi: 10.11591/ijece.v13i1.pp884-893.

[17] E. Fernando, P. Mudjiraharjo, and M. Aswin, "Implementasi Pendekatan Collaborative Filtering Dan K-Means Clustering Pada Sistem Rekomendasi Mata Kuliah," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 84–91, 2022, doi: 10.33387/jiko.v5i2.4559.

[18] A. L. Fernandes, A. T. Devega, A. Suryadi, R. illya Badri, N. Busra, and N. Aini, "Seleksi Mahasiswa Teknik Informatika untuk KPM, KP, dan TA Menggunakan Metode AHP," *JR J. Responsive Tek. Inform.*, vol. 7, no. 02, pp. 81–92, 2023, doi: 10.36352/jr.v7i02.751.

[19] S. Mondal, A. Basu, and N. Mukherjee, "Building a trust-based doctor recommendation system on top of multilayer graph database," *J. Biomed. Inform.*, vol. 110, no. April, p. 103549, 2020, doi: 10.1016/j.jbi.2020.103549.