

Prediksi Kunjungan Pasien Menggunakan Algoritma C5.0 di RS Regina Maris

Patient Visit Prediction Using the C5.0 Algorithm at Regina Maris Hospital

Muhammad Sowban Adilla*¹, Rakhmat Kurniawan R², Raissa Amanda Putri³
^{1,2,3} Universitas Islam Negeri Sumatera Utara Medan
¹sowban16@gmail.com, ²rakhmat.kr@uinsu.ac.id, ³raissa.ap@uinsu.ac.id

Received: April 04, 2026 | Revised: April 11, 2026 | Accepted: April 29, 2026

Abstrak

Peningkatan kunjungan pasien menuntut kesiapan fasilitas dan sumber daya rumah sakit, khususnya dalam manajemen obat dan layanan medis. Perencanaan yang tidak tepat berpotensi menghambat pelayanan, sehingga sistem prediksi kunjungan yang akurat sangat diperlukan. Penelitian ini bertujuan memprediksi tingkat kunjungan pasien di RS Regina Maris Medan guna mengoptimalkan perencanaan layanan dan meningkatkan kepuasan pasien. Metode *data mining* dengan algoritma klasifikasi C5.0 diterapkan pada data rekam kunjungan pasien. Data tersebut diolah untuk menghasilkan model prediksi berbasis pohon keputusan dan aturan (*rule*). Berdasarkan hasil evaluasi menggunakan *confusion matrix*, model C5.0 menghasilkan tingkat akurasi sebesar 60% dan nilai *F1-score* sebesar 71%. Kesimpulannya, model klasifikasi C5.0 terbukti andal dalam mengenali pola kunjungan. Kontribusi utama dari penelitian ini adalah penyediaan alat bantu rekomendasi berbasis data yang objektif bagi pihak manajemen rumah sakit untuk meningkatkan efisiensi tata kelola obat dan fasilitas pendukung.

Kata kunci: Rumah Sakit, Kesehatan, *Data Mining*, Algoritma C5.0, Web

Abstract

The increase in patient visits demands hospital readiness in providing facilities and resources, particularly in medication management and medical services. Inaccurate planning can hinder service delivery, making an accurate patient visit prediction system highly necessary. This study aims to predict the patient visit rate at Regina Maris Hospital Medan to optimize service planning and improve patient satisfaction. A data mining method employing the C5.0 classification algorithm was applied to historical patient visit records. The data was processed to generate a prediction model based on decision trees and rules. Based on the confusion matrix evaluation, the proposed C5.0 model achieved an accuracy of 60% and an F1-score of 71%. In conclusion, the classification model has proven reliable in recognizing visit patterns. The main contribution of this research is providing an objective, data-driven recommendation tool for hospital management to improve the efficiency of medicine inventory and support facilities management.

Keywords: Hospital, Health, *Data Mining*, C5.0 Algorithm, Web

1. PENDAHULUAN

Perkembangan sistem informasi kesehatan modern memungkinkan rumah sakit melakukan pengolahan data pasien secara real-time untuk mendukung pengambilan keputusan yang lebih efektif dan akurat. Pemanfaatan data kesehatan berbasis machine learning mampu membantu rumah sakit dalam melakukan prediksi kebutuhan layanan, pengelolaan sumber daya

medis, dan peningkatan kualitas pelayanan pasien [1]. Seiring dengan pertumbuhan jumlah penduduk dan perkembangan kondisi ekonomi yang semakin baik, tingkat kesadaran masyarakat terhadap pentingnya kesehatan juga mengalami peningkatan. Kondisi ini berdampak pada bertambahnya jumlah kunjungan pasien ke rumah sakit. Peningkatan tersebut perlu diimbangi dengan kesiapan rumah sakit dalam menyediakan sarana dan prasarana penunjang pelayanan [2]. Oleh karena itu, rumah sakit memerlukan sistem prediksi berbasis data yang mampu membantu proses *forecasting* jumlah kunjungan pasien secara lebih akurat guna mendukung kesiapan fasilitas, pengelolaan sumber daya medis, serta optimalisasi kualitas pelayanan kesehatan. Pemanfaatan *machine learning* dalam analisis data kesehatan terbukti mampu meningkatkan efektivitas pengambilan keputusan operasional rumah sakit secara *real-time* [3].

Industri rumah sakit hingga saat ini masih dihadapkan pada berbagai tantangan yang perlu ditangani secara serius. Standar pelayanan rumah sakit harus terus ditingkatkan dan disesuaikan dengan perkembangan ilmu pengetahuan, teknologi, kondisi ekonomi, serta dinamika permasalahan kesehatan [4]. Transformasi digital pada sektor kesehatan menuntut rumah sakit untuk mengembangkan sistem pelayanan yang lebih adaptif, efisien, dan berbasis data. Teknologi artificial intelligence dan machine learning saat ini mulai banyak diterapkan pada Electronic Health Records (EHR) untuk membantu peningkatan akurasi prediksi pasien serta optimalisasi manajemen layanan kesehatan [5].

Penyakit merupakan kondisi yang tidak dapat sepenuhnya dihindari oleh manusia dan dapat menyerang kapan saja. Beberapa jenis penyakit bahkan cenderung muncul atau mewabah pada musim tertentu. Jika pola kemunculan penyakit tersebut dapat diketahui sejak awal, maka upaya pencegahan dapat dilakukan lebih dini. Namun, pada kenyataannya masih banyak masyarakat yang belum memahami atau menyadari pola penyebaran penyakit di masa mendatang [6]. Oleh karena itu, diperlukan suatu proses peramalan untuk membantu memprediksi potensi wabah penyakit di periode berikutnya. Melalui data rekap penyakit bulanan dari periode sebelumnya, pola penyakit musiman dapat diidentifikasi. Hasil pemilihan data ini memungkinkan prediksi beberapa jenis penyakit dengan potensi kemunculan yang tinggi pada periode berikutnya. Informasi ini dapat digunakan oleh penyedia layanan kesehatan, seperti rumah sakit, untuk memberikan edukasi dini tentang upaya pencegahan dan pengelolaan penyakit. Selain itu, rumah sakit juga dapat memperkirakan jumlah kunjungan pasien di masa mendatang. Dengan prediksi ini, kebutuhan akan obat-obatan dasar untuk periode berikutnya dapat disesuaikan dengan perkiraan jumlah pasien, sehingga meminimalkan risiko kekurangan obat.

Algoritma C5.0 merupakan pengembangan lebih lanjut dari algoritma C4.5, yang sebelumnya juga merupakan penyempurnaan dari algoritma ID3 [7]. Sebagai salah satu algoritma klasifikasi, C5.0 dinilai sangat cocok digunakan untuk mengolah dataset berukuran besar [8]. Dalam penerapannya, C5.0 menggunakan ukuran *gain ratio* untuk menentukan atribut yang akan diproses. *Gain ratio* tersebut berfungsi dalam pemilihan atribut uji pada setiap node dalam struktur pohon keputusan. Atribut yang memiliki nilai *gain ratio* tertinggi akan dipilih sebagai node induk (*parent*) untuk pembentukan node berikutnya [9]. Hasil akhir dari algoritma C5.0 dapat berupa model berbasis pohon (*tree*) maupun *rule based model*. Algoritma C5.0 merupakan pengembangan *decision tree* yang memiliki kemampuan pruning lebih baik dibandingkan ID3 dan C4.5 sehingga mampu menghasilkan struktur pohon keputusan yang lebih sederhana dan efisien. Penelitian terbaru menunjukkan bahwa algoritma C5.0 tetap relevan digunakan pada klasifikasi data kesehatan karena memiliki interpretabilitas tinggi dan proses komputasi yang relatif cepat dibandingkan beberapa algoritma klasifikasi lainnya [10].

Berbagai penelitian sebelumnya telah menerapkan algoritma C5.0 pada domain kesehatan dengan hasil performa yang cukup baik. Penelitian oleh [7] menunjukkan bahwa algoritma C5.0 mampu menghasilkan akurasi sebesar 95% dalam prediksi penyakit stroke menggunakan data medis pasien. Selain itu, penelitian [8] memperoleh akurasi 97,92% pada klasifikasi status gizi

balita menggunakan pendekatan *preprocessing* dan *resampling* data. Penelitian lain oleh [11] juga membuktikan bahwa algoritma C5.0 efektif diterapkan pada klasifikasi citra CT-Scan Covid-19 dengan tingkat akurasi mencapai 87%.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada klasifikasi penyakit atau diagnosis medis, sedangkan penerapan algoritma C5.0 dalam prediksi kunjungan pasien berbasis evaluasi kepuasan layanan rumah sakit masih relatif terbatas. Selain itu, penelitian sebelumnya umumnya hanya berorientasi pada performa klasifikasi tanpa mengintegrasikan aspek *usability* sistem berbasis web yang digunakan oleh pengguna akhir di lingkungan rumah sakit. Dengan demikian, masih terdapat research gap terkait implementasi algoritma C5.0 yang tidak hanya berfungsi sebagai model klasifikasi, tetapi juga sebagai sistem pendukung keputusan berbasis *feedback* pasien yang dapat digunakan langsung oleh manajemen rumah sakit.

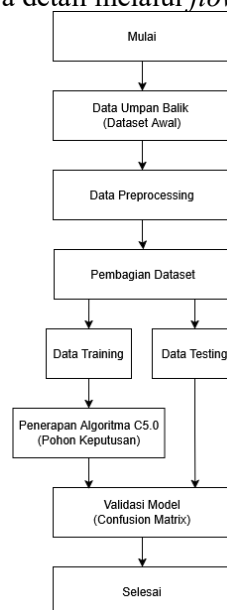
Novelty penelitian ini terletak pada integrasi metode klasifikasi C5.0 dengan sistem *feedback* pasien berbasis web untuk memprediksi potensi kunjungan ulang pasien berdasarkan variabel pelayanan, harga, fasilitas, dan loyalitas pasien. Penelitian ini tidak hanya menghasilkan model decision tree, tetapi juga mengimplementasikan sistem secara langsung dalam bentuk *dashboard monitoring* rumah sakit yang dilengkapi evaluasi *usability* menggunakan metode *System Usability Scale* (SUS). Pendekatan ini memberikan kontribusi praktis dalam membantu pengambilan keputusan manajemen rumah sakit terkait optimalisasi layanan dan perencanaan fasilitas kesehatan

2. METODE PENELITIAN

Dalam penelitian ini, penulis menggunakan algoritma C5.0 sebagai metode utama untuk memprediksi tingkat kunjungan pasien di RS Regina Maris Medan.

2.1 Alur Penelitian dan Preprocessing Data

Alur proses pengolahan data serta pembentukan model keputusan dilakukan secara berurutan. Untuk memperjelas kesiapan data sebelum masuk ke tahap *modeling*, diterapkan prosedur *preprocessing* data kuesioner umpan balik (*feedback*) pasien. Tahapan-tahapan sistematis tersebut digambarkan secara detail melalui *flowchart* pada Gambar 1.



Gambar 1. *Flowchart Preprocessing Algoritma C5.0*

Adapun tahapan *preprocessing* data yang diterapkan meliputi:

1. *Data Cleaning*: Menyeleksi data kuesioner dari rekam umpan balik pasien untuk memastikan tidak terdapat data kosong (missing values) atau baris yang terduplikasi secara tidak sengaja saat pengisian instrumen digital.
2. *Data Integration*: Menggabungkan data tanggapan instrumen dari kategori Pelayanan, Harga, Fasilitas, dan Loyalitas ke dalam satu basis data tabular terpadu.
3. *Data Transformation*: Mengubah format respon mentah yang dikumpulkan dari kuesioner pasien menjadi label kategoris yang seragam untuk mesin klasifikasi, yaitu Sangat Puas (SP), Puas (P), dan Tidak Puas (TP).

2.2 Spesifikasi Dataset

Dataset yang diolah dalam penelitian ini diperoleh langsung dari sebaran formulir kuesioner umpan balik (*customer feedback*) pasien di RS Regina Maris Medan. Jumlah total dataset yang digunakan secara rinci adalah 370 baris entri data.

Berdasarkan hasil observasi dan wawancara dengan manajemen rumah sakit, ditetapkan 4 kriteria utama yang bertindak sebagai atribut, di mana kriteria Loyalitas (Kunjungan Berulang) bertindak sebagai atribut target atau class. Total akumulasi nilai sub-kriteria di dalam *dataset* ini mencapai 1.850 sampel respon untuk kelompok pelayanan, 740 sampel respon untuk kelompok harga, dan 370 sampel respon untuk kelompok loyalitas. Struktur *dataset* disajikan pada Tabel 1.

Tabel 1. Data Spesifikasi Kriteria, Sub-Kriteria, dan Jumlah Data

No	Kriteria	Sub Kriteria	Penilaian	Jumlah Sampel Data
1	Pelayanan	Customer Care	Sangat Puas, Puas Dan Tidak Puas	370 Baris
		Igd	Sangat Puas, Puas Dan Tidak Puas	
		Perawat	Sangat Puas, Puas Dan Tidak Puas	
		Dokter	Sangat Puas, Puas Dan Tidak Puas	
		Keamanan	Sangat Puas, Puas Dan Tidak Puas	
2	Harga	Discharge Pasien	Sangat Puas, Puas Dan Tidak Puas	370 Baris
3	Fasilitas	Farmasi, Lab Dan Radiologi	Sangat Puas, Puas Dan Tidak Puas	370 Baris
		Menu Makanan	Sangat Puas, Puas Dan Tidak Puas	
4	Loyalitas	Kunjungan Berulang	Sangat Puas, Puas Dan Tidak Puas	370 Baris

2.3 Pembagian Data (Data Splitting)

Untuk menguji keandalan model klasifikasi yang dibentuk serta mendeteksi potensi masalah *overfitting*, total *dataset* berukuran 370 entri data ini dipisahkan ke dalam dua bagian data independen:

1. *Data Training*: Digunakan sebagai basis data latih oleh algoritma C5.0 untuk mengekstrak pola, melakukan perhitungan tingkat kecacatan atribut, dan menyusun aturan keputusan (*rule-based*).
2. *Data Testing*: Digunakan sebagai basis data uji untuk memvalidasi kualitas prediksi dari pohon keputusan yang telah selesai dibangun.

2.4 Perhitungan Algoritma C5.0 (Entropy dan Information Gain)

Algoritma C5.0 membentuk pohon keputusan dengan memilih pembelahan (*split*) atribut yang menghasilkan penurunan nilai ketidakmurnian (*impurity*) paling optimal. Atribut yang memiliki perolehan nilai Information Gain tertinggi dipilih menjadi simpul akar (root node) awal. Rumus matematis yang digunakan dijabarkan sebagai berikut:

1. Perhitungan Entropy

Entropy mengukur tingkat keragaman, kecacatan, atau ketidakmurnian dari suatu kumpulan data sampel S . Formula matematis untuk menghitung nilai *Entropy* dinyatakan dalam Persamaan (1) [12]:

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (1)$$

Keterangan:

S = Himpunan kasus atau sampel data kuesioner.

n = Jumlah partisi atau label kelas pada kriteria target (SP, P, TP).

p_i = Proporsi atau rasio probabilitas jumlah sampel untuk kelas ke- i terhadap total keseluruhan sampel data.

2. Perhitungan Information Gain

Setelah tingkat ketidakmurnian awal dihitung, dilakukan perhitungan nilai *Information Gain* pada setiap atribut A untuk mengetahui seberapa efektif atribut tersebut dalam mereduksi nilai *Entropy* total. Rumus menghitung nilai *Gain* adalah sebagai berikut [13]:

$$Gain(S, A) = Entropy(S) - \sum \left(\frac{|S_j|}{|S|} \right) \cdot Entropy(S_j) \quad (2)$$

Keterangan:

S = Himpunan sampel kasus awal.

A = Atribut kriteria yang sedang diuji kualitas pembelahannya.

m = Jumlah partisi nilai respon yang ada pada atribut A .

$|S_j|$ = Jumlah sampel data pada partisi nilai ke- j .

$|S|$ = Jumlah total sampel data dalam himpunan S .

$Entropy(S_j)$ = Nilai Entropy dari sub-dataset hasil partisi ke- j .

Atribut yang memiliki nilai *Gain* tertinggi akan diletakkan pada tingkatan simpul tertinggi untuk mengarahkan aturan prediksi.

2.5 Teknik Validasi Model

Validasi performa hasil klasifikasi diuji secara matematis menggunakan pengujian matriks kecacauan (*Confusion Matrix*). Model yang telah diekstrak aturan keputusannya divalidasi ketepatannya menggunakan data uji berdasarkan empat komponen utama: *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN). Indikator evaluasi pengujian dirumuskan sebagai berikut:

1. Accuracy (Akurasi): Persentase total ketepatan prediksi benar yang dihasilkan model.

$$Accuracy = ((TP + TN) / (TP + FP + FN + TN)) \times 100\%$$

2. Precision (Presisi): Mengukur rasio ketepatan antara informasi yang diproyeksikan positif oleh model dengan kondisi riil aktual lapangan.

$$Precision = (TP / (TP + FP)) \times 100\%$$

3. *Recall* (Sensitivitas): Mencerminkan efektivitas model dalam mengenali dan mendeteksi kembali seluruh entitas informasi kelas positif yang benar.
 $Recall = (TP / (TP + FN)) \times 100\%$
4. F1-Score: Nilai rata-rata harmonis gabungan antara nilai presisi dan recall untuk menyeimbangkan performa klasifikasi data.
 $F1-Score = (2 \times Recall \times Precision) / (Recall + Precision) \times 100\%$

2.6 Perangkat Lunak dan Tools yang Digunakan

Peralatan digunakan dengan tujuan untuk mengidentifikasi kebutuhan sistem sehingga proses perancangan menjadi lebih mudah. Peralatan tersebut mencakup kebutuhan perangkat keras maupun perangkat lunak. Kebutuhan-kebutuhan dibawah ini merupakan kebutuhan minimal dari sistem:

Perangkat keras yang digunakan:

1. Prosesor AMD A8-7410 APU with AMD Radeon R5 Grapics 2.20GHz
2. 64-bit Operating System
3. RAM 4GB
4. 500GB HDD

Perangkat lunak yang digunakan:

1. Sistem Operasi Windows 10 64-bit
2. XAMPP x64-8.1.6
3. MYSQL
4. Visual Studio Code
5. Google Chrome

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma C5.0

Algoritma C5.0 dihitung untuk menentukan pohon keputusan dari dataset yang diambil dari RS Regina Maris Medan. Langkah-langkah perhitungan dalam membangun pohon keputusan (*decision tree*) menggunakan Algoritma C5.0 diuraikan dalam prosedur berikut:

1. Identifikasi jumlah total entitas data dan proporsinya untuk setiap variabel yang digunakan dalam model. Rincian perhitungan data untuk setiap atribut ditunjukkan pada Tabel 2 di bawah ini:

Tabel 2. Data Formulir Umpan Balik Pasien

Dataset	Sub Dataset	SP	P	TP	Total Data
Pelayanan	Customer Care	242	107	21	370
	Igd	195	130	45	370
	Perawat	156	164	50	370
	Dokter	265	95	10	370
	Keamanan	310	50	10	370
Total		1168	546	136	1850
Harga	Discharge	130	120	120	370
Fasilitas	Farmasi, Lab Dan Radiologi	200	128	42	370
	Menu Makanan	297	53	20	370
Total		497	181	62	740
Loyalitas	Kunjungan Berulang	127	120	123	370

2. Menghitung nilai *entropy*

Rumus yang digunakan sebagai acuan dalam menghitung nilai *entropy* dijelaskan dalam persamaan 1.

$$= \left(- \left(\frac{242}{370} \right) \times \log_2 \left(\frac{242}{370} \right) \right) + \left(- \left(\frac{107}{370} \right) \times \log_2 \left(\frac{107}{370} \right) \right) + \left(- \left(\frac{21}{370} \right) \times \log_2 \left(\frac{21}{370} \right) \right)$$

$$= 1,153163793$$

3. Menghitung Nilai Gain

Nilai Gain dapat dihitung menggunakan persamaan 2:

Gain (Pelayanan)

$$= ((1,215333647) - ((242/1168)*1,531638) - ((195/1168)*1.3868663) -$$

$$((156/1168)*1,4358299) - ((265/1168)*0.9893127) - ((310/1168)*0.7448662))$$

$$= 0,130941171$$

Berdasarkan analisis dataset sebanyak 370 entri, proses penentuan pohon keputusan menggunakan algoritma C5.0 menghasilkan parameter *entropy* dan *gain* yang dirinci dalam Tabel 3:

Tabel 3. Entropi dan Gain Data yang Diproses

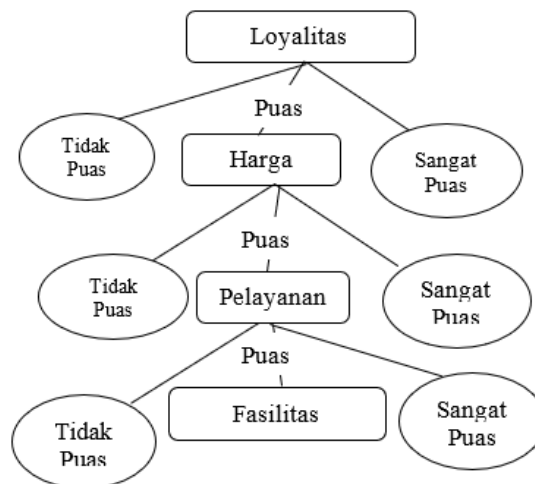
Dataset	Sub Dataset	SP	P	TP	Total Data	Entropy	Gain
Pelayanan	Customer Care	242	107	21	370	1,153163793	
	Igd	195	130	45	370	1,386866282	
	Perawat	156	164	50	370	1,435829881	
	Dokter	265	95	10	370	0,989312733	
	Keamanan	310	50	10	370	0,744866174	
	Total	1168	546	136	1850	1,215333647	0,130941171
Harga	Discharge	130	120	120	370	1,583917789	1,027406133
Fasilitas	Farmasi, Lab Dan Radiologi	200	128	42	370	1,365844786	
	Menu Makanan	297	53	20	370	0,883621664	
	Total	497	181	62	740	1,182311202	0,104635968
Loyalty	Repeat Visits	127	120	123	370	1,584573053	1,040679059

Hasil perhitungan *entropy* dan *information gain* menunjukkan bahwa atribut Loyalitas memiliki nilai gain tertinggi sebesar 1,0406 dibandingkan atribut lainnya. Nilai tersebut mengindikasikan bahwa loyalitas pasien merupakan variabel paling informatif dalam menentukan pola prediksi kunjungan ulang pasien. Secara konseptual, hal ini menunjukkan bahwa pengalaman kunjungan sebelumnya sangat memengaruhi keputusan pasien untuk kembali menggunakan layanan rumah sakit. Temuan ini sejalan dengan teori kualitas layanan kesehatan yang menyatakan bahwa kepuasan pasien memiliki hubungan langsung terhadap repeat visit intention atau loyalitas pasien.

Selain variabel loyalitas, atribut Harga juga menunjukkan pengaruh yang cukup besar dengan nilai gain sebesar 1,0274. Hal ini menunjukkan bahwa persepsi biaya pelayanan masih menjadi faktor penting dalam pengambilan keputusan pasien. Sementara itu, atribut Pelayanan dan Fasilitas memiliki nilai gain yang lebih rendah, yang mengindikasikan bahwa kedua variabel tersebut belum mampu memisahkan kelas data secara optimal pada dataset yang digunakan.

Struktur pohon keputusan yang dihasilkan memperlihatkan bahwa algoritma C5.0 cenderung memilih atribut dengan kemampuan reduksi entropy tertinggi pada setiap node.

Pemilihan atribut Loyalitas sebagai root node menunjukkan bahwa variabel tersebut memiliki kemampuan paling baik dalam mengurangi ketidakpastian klasifikasi dibandingkan atribut lainnya. Dengan demikian, model yang dihasilkan tidak hanya berfungsi sebagai alat prediksi, tetapi juga dapat digunakan sebagai dasar evaluasi kebijakan layanan rumah sakit, khususnya dalam mempertahankan loyalitas pasien. Struktur akhir dari model prediksi ini divisualisasikan sepenuhnya pada Gambar 2.



Gambar 2. Pohon Keputusan

Berdasarkan struktur pohon keputusan yang terbentuk, atribut Loyalitas menjadi akar utama (*root node*) dalam proses klasifikasi. Hal ini menunjukkan bahwa keputusan pasien untuk melakukan kunjungan ulang dipengaruhi secara dominan oleh tingkat kepuasan sebelumnya. Ketika pasien memberikan penilaian “Sangat Puas”, model secara langsung mengklasifikasikan pasien ke dalam kategori berpotensi melakukan kunjungan kembali. Sebaliknya, apabila pasien memberikan penilaian “Tidak Puas”, maka probabilitas pasien untuk kembali berkunjung menjadi rendah.

Pada kondisi ketika pasien memberikan jawaban “Puas”, sistem memerlukan atribut tambahan seperti Harga dan Fasilitas untuk menentukan klasifikasi akhir. Kondisi ini menunjukkan bahwa kelas “Puas” memiliki karakteristik data yang lebih ambigu dibandingkan kelas lainnya, sehingga diperlukan proses pemisahan lanjutan pada node berikutnya. Dengan demikian, pohon keputusan yang dihasilkan tidak hanya menunjukkan hubungan antar atribut, tetapi juga menggambarkan pola perilaku pasien terhadap kualitas layanan rumah sakit.

3.2 Confusion Matrix

Guna menguji validitas dan tingkat presisi dari *rule* yang telah dikonstruksi, dilakukan tahap evaluasi menggunakan mekanisme *confusion matrix*. Instrumen ini berperan krusial dalam mengukur efektivitas model melalui parameter akurasi, presisi, serta *recall*. Adapun rincian data hasil pengujian performa tersebut disajikan secara sistematis pada Tabel 4 di bawah ini.

Table 4. Data Confusion Matrix

N=370	Aktual	
	Positif (1)	Negatif (0)
Prediksi : Positif (1)	TP	FP
	180	67
Prediksi : Negatif (0)	FN	TN
	80	43

Nilai yang dihasilkan oleh metode *confusion matrix* memiliki bentuk estimasi sebagai berikut:

1. *Accuracy*

Menjelaskan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$\begin{aligned} \textit{Accuracy} &= (TP+TN) / (TP+FP+FN+TN) \\ &= (180+43) / (180+67+80+43) \\ &= 0,602702703 \\ &= 0,602702703 * 100 \% = 60 \% \end{aligned}$$

2. *Precision* mewakili rasio akurasi antara informasi yang diproyeksikan oleh model.

$$\begin{aligned} \textit{Precision} &= (TP) / (TP + FP) \\ &= 180 / (180 + 67) \\ &= 0,728744939 \\ &= 0.73 * 100\% = 73\% \end{aligned}$$

3. *Recall* atau *sensitivity*: mencerminkan efektivitas model dalam mendeteksi ulang setiap entitas informasi.

$$\begin{aligned} \textit{Recall} &= TP / (TP + FN) \\ &= 180 / (180+80) \\ &= 0,692307692 \\ &= 0.69 * 100\% = 69\% \end{aligned}$$

4. F1 Score

$$\begin{aligned} \textit{F-1 Score} &= (2 * \textit{Recall} * \textit{Precision}) / (\textit{Recall} + \textit{Precision}) \\ &= (2 * 0.69 * 0.73) / (0.69 + 0.73) \\ &= 0,710059172 * 100\% = 71\% \end{aligned}$$

Berdasarkan hasil pengujian *confusion matrix*, model menghasilkan tingkat *accuracy* sebesar 60,27%, *precision* sebesar 73%, *recall* sebesar 69%, dan F1-score sebesar 71%. Nilai *precision* yang relatif tinggi menunjukkan bahwa model cukup baik dalam memprediksi pasien yang benar-benar berpotensi melakukan kunjungan ulang. Artinya, sebagian besar prediksi positif yang dihasilkan model sesuai dengan kondisi aktual data pasien.

Namun demikian, tingkat *accuracy* yang hanya mencapai 60% menunjukkan bahwa model masih memiliki keterbatasan dalam melakukan klasifikasi secara menyeluruh terhadap seluruh data pengujian. Kondisi ini dapat disebabkan oleh distribusi data yang tidak seimbang (*imbalanced dataset*), terutama pada kategori tingkat kepuasan pasien. Ketidakseimbangan data menyebabkan model lebih dominan mengenali kelas mayoritas dibandingkan kelas minoritas sehingga meningkatkan kemungkinan terjadinya kesalahan klasifikasi.

Selain itu, nilai *recall* sebesar 69% menunjukkan bahwa model belum sepenuhnya mampu mendeteksi seluruh pasien yang benar-benar memiliki potensi melakukan kunjungan ulang. Masih terdapat sejumlah data *False Negative* (FN) sebanyak 80 data, yang berarti pasien yang sebenarnya berpotensi kembali justru diprediksi tidak kembali oleh sistem. Kondisi ini menjadi indikator bahwa model masih mengalami kesulitan dalam mengenali pola tertentu pada data pasien.

Nilai F1-score sebesar 71% menunjukkan keseimbangan performa antara *precision* dan *recall*. Akan tetapi, nilai tersebut tidak dapat diartikan sebagai tingkat akurasi keseluruhan sistem, melainkan sebagai indikator harmonisasi kemampuan model dalam mengurangi kesalahan prediksi positif dan negatif secara bersamaan. Dengan demikian, F1-score lebih tepat digunakan untuk mengevaluasi kestabilan model klasifikasi pada dataset yang memiliki distribusi kelas tidak seimbang.

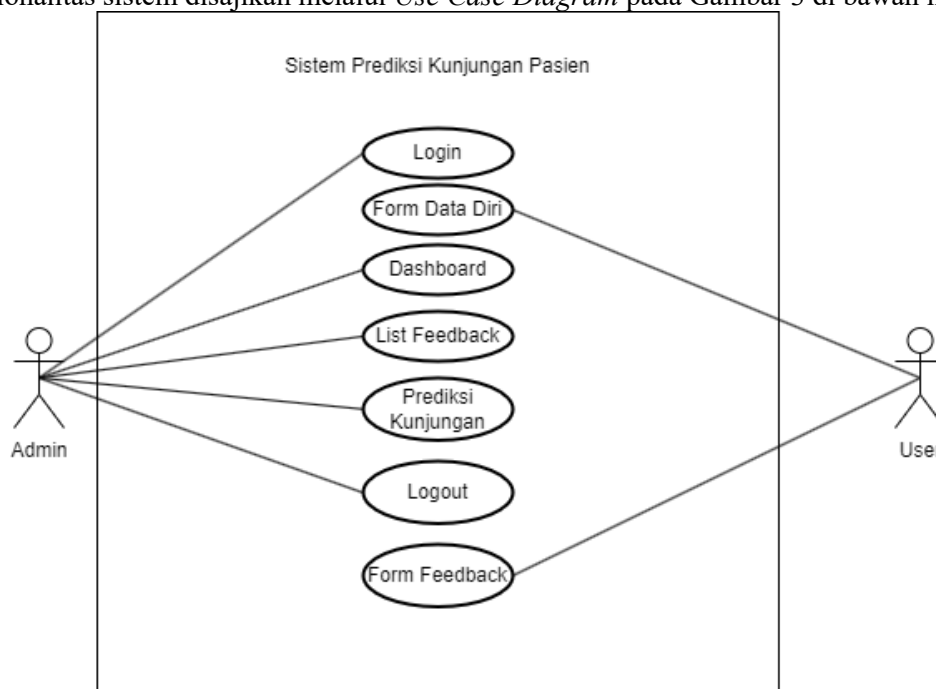
Jika dibandingkan dengan beberapa penelitian terdahulu yang menggunakan algoritma C5.0 pada domain klasifikasi pendidikan dan kesehatan, performa model pada penelitian ini masih berada pada kategori cukup baik, namun belum optimal. Faktor utama yang memengaruhi performa model adalah jumlah dataset yang relatif terbatas, variasi atribut yang belum kompleks, serta belum diterapkannya teknik optimasi data seperti *feature selection* dan *balancing data* menggunakan metode SMOTE.

Untuk meningkatkan performa klasifikasi pada penelitian selanjutnya, model C5.0 dapat dibandingkan dengan algoritma lain seperti Random Forest, Naive Bayes, Support Vector Machine (SVM), maupun XGBoost. Algoritma *ensemble* seperti Random Forest dan XGBoost berpotensi menghasilkan tingkat akurasi yang lebih tinggi karena memiliki kemampuan generalisasi yang lebih baik dalam menangani data kompleks dan noisy.

3.3 Desain

1. Use Case Diagram

Use Case Diagram digunakan untuk menunjukkan interaksi antara pengguna dan entitas eksternal lainnya dengan sistem yang sedang dikembangkan [14]. Langkah-langkah dalam interaksi sistem dengan pengguna. Arsitektur sistem yang dikembangkan melibatkan pengguna sebagai aktor utama dalam operasinya. Secara visual, interaksi fungsionalitas sistem disajikan melalui *Use Case Diagram* pada Gambar 3 di bawah ini:

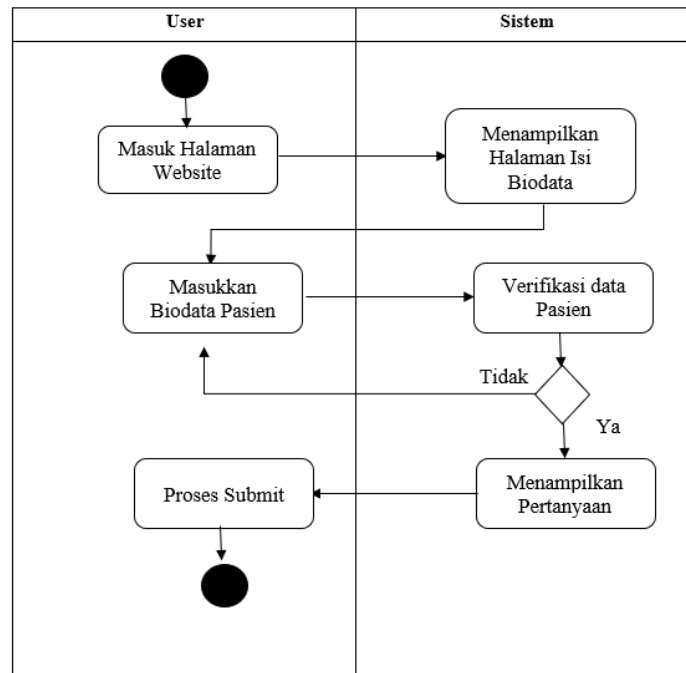


Gambar 3. *Use Case Diagram*

Representasi diagram tersebut mengidentifikasi dua entitas utama, yakni Admin dan User. Bagi Admin, akses ke dalam panel *dashboard* hanya dapat dilakukan setelah melewati proses autentikasi, di dashboard akan ada menampilkan prediksi kunjungan pasien dan grafik, menu *list feedback* yang merupakan detail dari jawaban para pasien yang telah mengisi *feedback*, dan menu *logout*. Sedangkan pada user, pertama *user* akan dihadapkan dengan tampilan isi data diri setelah berhasil isi data diri, kemudian *user* akan diperlihatkan tampilan *form* yg nantinya harus diisi sampai selesai.

2. Activity Diagram Menu User

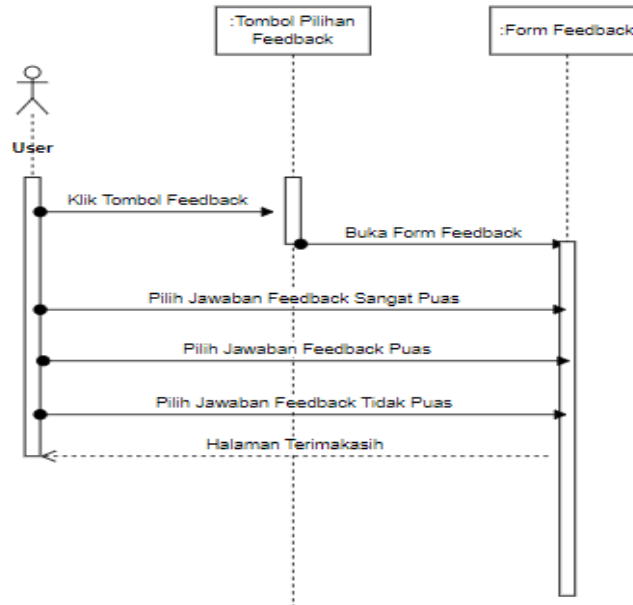
Activity diagram menggambarkan workflow(aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak [15]. *Activity Menu User* ini dimulai dengan *user* memilih ingin mengisi *form feedback* sebagai rawat jalan atau rawat inap, kemudian *user* mengisi data diri sesuai *form* yang ada. Setelah itu *user* menjawab pertanyaan seluruh kategori dengan cara memilih jawaban sangat puas, puas dan tidak puas. *Activity Menu User* seperti pada Gambar 4 di bawah ini.



Gambar 4. Activity Diagram Menu User

3. Sequence diagram

Sequence diagram menunjukkan interaksi antar objek baik di dalam maupun di sekitar sistem, seperti pengguna, tampilan, dan lainnya, dalam bentuk pesan yang digambarkan berdasarkan waktu. Diagram ini memiliki dua dimensi: dimensi vertikal yang mewakili waktu dan dimensi horizontal yang menunjukkan objek yang terlibat [16]. Gambar 5 adalah *Sequence diagram* yang menggambarkan proses pengguna dalam mengisi formulir umpan balik.

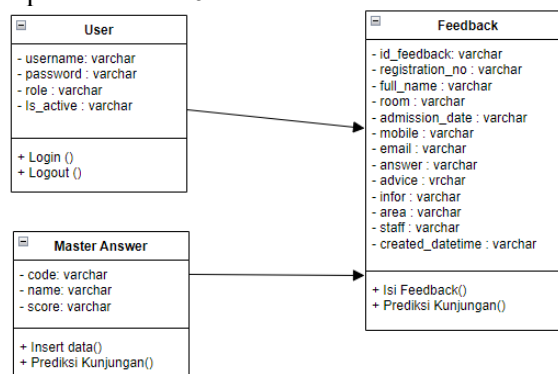


Gambar 5. Diagram *Sequence User*

Gambar 5 di atas menunjukkan proses seorang *user* untuk mengisi *form feedback*. Untuk mengisi *form feedback* yang terjadi ditahap awal *user* akan memilih tombol *feedback* terlebih dahulu. Setelah itu akan masuk pada halaman buka *form*. Pada *form feedback* *user* dapat memilih jawaban sangat puas, puas atau tidak puas. Lalu ketika selesai mengisi *user* melihat langsung halaman terimakasih. Penelitian ini juga memiliki *diagram sequence* admin, yang akan menggambarkan proses admin dalam melihat hasil dari *feedback* pasien. Pada *diagram sequence* terlihat bahwa admin akan dimulai dengan *login* terlebih dahulu, kemudian jika *login* berhasil maka akan tampil halaman *dashboard*. Pada *dashboard* tersebut akan ditampilkan hasil prediksi kunjungan pasien.

4. *Class Diagram*

Arsitektur sistem secara visual dipetakan melalui *class diagram*, yang mengidentifikasi komponen-komponen kelas sebagai fondasi pembangunan sistem. Di dalam diagram ini, setiap entitas dilengkapi dengan atribut untuk mendeskripsikan karakteristik variabelnya, serta metode yang merepresentasikan fungsi operasional kelas tersebut [17]. Representasi detail dari struktur ini dapat dilihat pada Gambar 6.



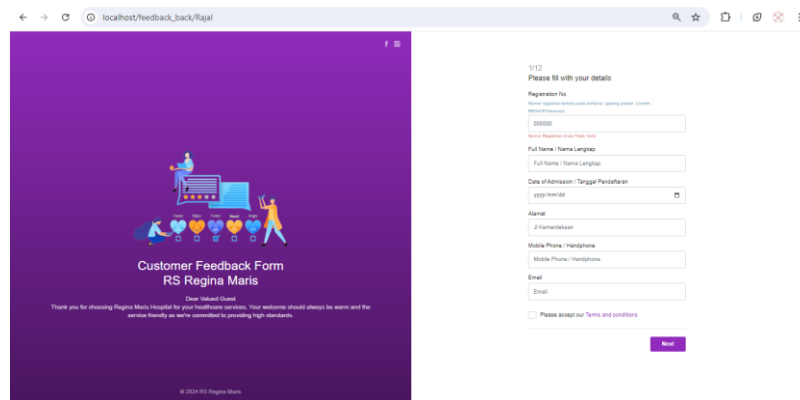
Gambar 6. *Class Diagram*

Konstruksi *class diagram* pada sistem ini mengintegrasikan beberapa entitas utama, di antaranya adalah kelas *user*, *feedback*, dan *master answer*. Masing-masing kelas dibekali dengan atribut dan prosedur (*method*) spesifik yang berperan dalam menjalankan mekanisme kerja sistem.

3.4 Implementasi

1. Halaman Pengisian Biodata Pasien

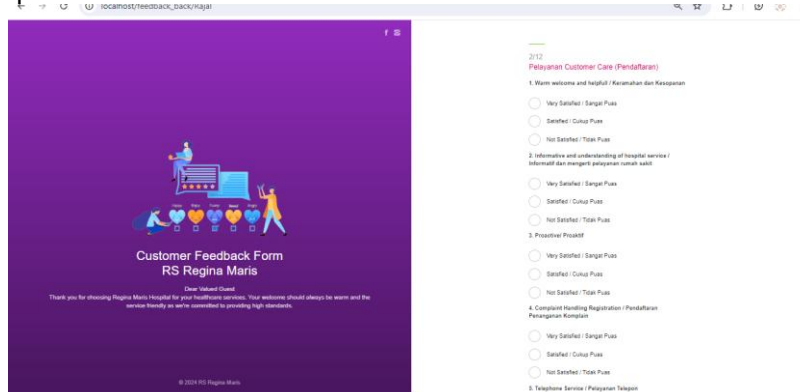
Setelah tampilan awal dan user memilih salah satu antara pasien rawat inap atau rawat jalan maka setelah itu pasien disuguhkan dengan tampilan untuk isi biodata pasien terlebih dahulu yang nantinya sistem akan cek apakah pasien yang ada di biodata tersebut benar ada di RS Regina Maris Medan. Berikut tampilan untuk user isi biodata seperti dibawah ini pada Gambar 7.



Gambar 7. Halaman Pengisian Biodata Pasien

2. Halaman Isi Feedback

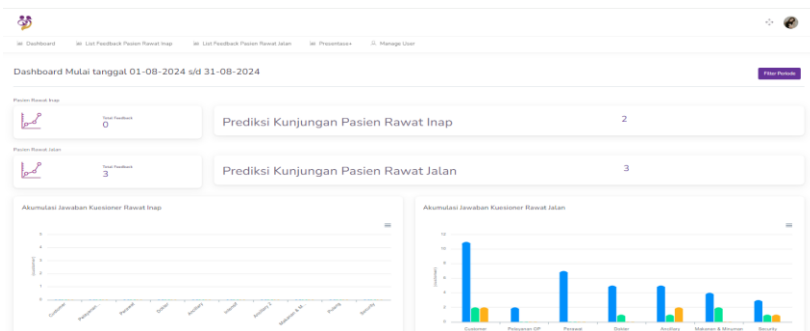
Halaman ini muncul setelah data pasien berhasil diverifikasi. Pasien (*user*) kemudian akan diminta untuk menjawab pertanyaan berdasarkan perasaan mereka dan apa yang telah mereka terima dari rumah sakit. Pertanyaan akan terus ditampilkan dan akan diakhiri dengan tombol kirim. Dengan mengklik tombol submit, formulir *feedback* telah selesai diisi. Tampilannya dapat dilihat pada Gambar 8 di bawah ini.



Gambar 8. Halaman Isi Feedback

3. Halaman *Dashboard* Admin

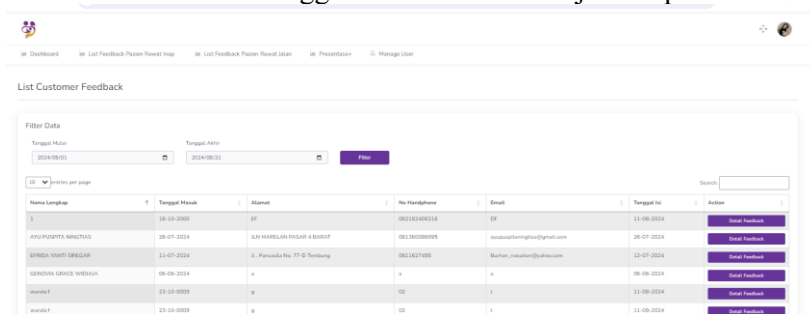
Halaman ini muncul setelah admin berhasil *login*. Halaman ini menampilkan banyak informasi, seperti grafik hasil *feedback* yang diisi oleh *user*, prediksi kunjungan pasien, dan daftar pasien yang telah memberikan *feedback*, beserta jawaban terperinci. Halaman dashboard ini dapat dilihat pada Gambar 9 di bawah ini.



Gambar 9. Halaman Dashboard Admin

4. Halaman Tampilan List Feedback Pasien (Admin)

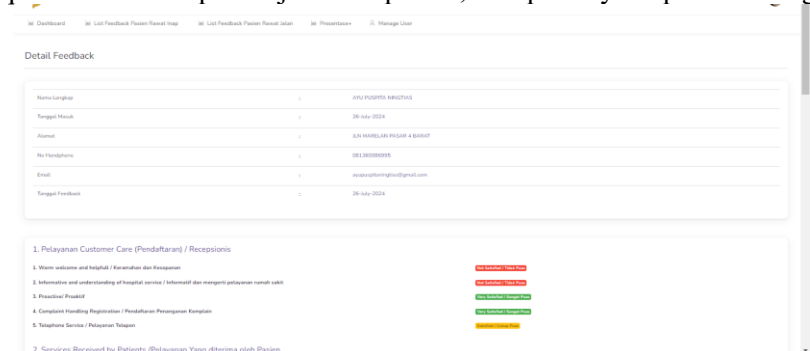
Halaman ini menampilkan daftar pasien (pengguna) yang telah memberikan feedback. Admin dapat memfilter data berdasarkan tanggal. Halaman ini ditunjukkan pada Gambar 10.



Gambar 10. Halaman Tampilan List Feedback Pasien (Admin)

5. Halaman Tampilan Detail Feedback

Halaman ini muncul setelah admin mengklik detail *feedback* pada halaman list *feedback* pasien. Tampilan ini menampilkan jawaban pasien, dari pertanyaan pertama hingga terakhir.



Gambar 11. Halaman Tampilan Detail Feedback

3.5 Evaluasi Usability Sistem

Selain evaluasi performa klasifikasi, penelitian ini juga melakukan pengujian usability sistem menggunakan metode System Usability Scale (SUS). Pengujian dilakukan terhadap 10 responden sebagai pengguna sistem. Instrumen SUS digunakan untuk mengukur tingkat kemudahan penggunaan, efisiensi antarmuka, serta kenyamanan pengguna saat mengoperasikan aplikasi prediksi kunjungan pasien.

Berdasarkan hasil pengujian, sistem memperoleh skor rata-rata SUS sebesar 76,5. Berdasarkan interpretasi standar SUS, skor tersebut termasuk dalam kategori *Acceptable* dengan *grade Good*. Hasil ini menunjukkan bahwa sistem memiliki tingkat *usability* yang baik dan

mudah dipahami oleh pengguna, baik dari sisi navigasi menu, tampilan antarmuka, maupun proses penginputan data feedback pasien.

Nilai *usability* yang baik menunjukkan bahwa implementasi sistem tidak hanya berfokus pada performa algoritma klasifikasi, tetapi juga memperhatikan aspek pengalaman pengguna (*user experience*). Dengan demikian, sistem dapat diterapkan secara lebih efektif dalam mendukung pengambilan keputusan manajemen rumah sakit terkait prediksi kunjungan pasien dan evaluasi kualitas layanan.

4. KESIMPULAN

Implementasi algoritma C5.0 dalam sistem prediksi kunjungan pasien berbasis web di RS Regina Maris Medan menghasilkan tingkat Akurasi Global sebesar 60,27%, Precision 73,00%, Recall 69,23%, dan F1-Score 71,00%. Pengujian pohon keputusan menunjukkan bahwa kriteria Loyalitas (Repeat Visits) menjadi atribut paling dominan dengan nilai *Information Gain* tertinggi sebesar 1,0406 sebagai akar utama (*root node*), diikuti oleh variabel Harga dan Pelayanan yang berpengaruh signifikan terhadap akurasi. Sementara itu, evaluasi antarmuka sistem menggunakan *System Usability Scale* (SUS) menghasilkan skor rata-rata 76,5 (kategori *Acceptable / Good*), yang membuktikan bahwa aplikasi ini sangat layak dan mudah dioperasikan staf manajemen rumah sakit untuk membantu pengoptimalan perencanaan logistik medis.

Meskipun sistem berjalan dengan baik, penelitian ini memiliki keterbatasan pada adanya ketimpangan distribusi data (*imbalanced dataset*) pada kuesioner pasien yang mengakibatkan bias algoritma, sehingga capaian akurasi global tertahan di angka 60,27%. Selain itu, variabel prediktor masih terbatas pada aspek internal tanpa melibatkan faktor eksternal seperti tren epidemiologi penyakit musiman atau demografi pasien. Untuk penelitian selanjutnya, disarankan menerapkan metode *hybrid machine learning* menggunakan pendekatan penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) pada tahap *preprocessing*, serta mengeksplorasi algoritma *ensemble classification* seperti *Random Forest* atau *XGBoost* untuk meningkatkan performa akurasi prediksi.

DAFTAR PUSTAKA

- [1] S. Sivarajkumar, Y. Huang, and Y. Wang, "Fair patient model: Mitigating bias in the patient representation learned from the electronic health records," *J. Biomed. Inform.*, vol. 148, p. 104544, Dec. 2023, doi: 10.1016/j.jbi.2023.104544.
- [2] R. Iskandar, "Prediksi Jumlah Kunjungan Pasien Rawat Jalan Pada Tahun 2024-2028 Dengan Metode Least Square Di RSUD Prasetya Bunda," Politeknik Kesehatan Tasikmalaya, 2024.
- [3] S. A. Alowais *et al.*, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC Med. Educ.*, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.
- [4] M. C. N. Jauhary and A. P. Irawan, "Komparasi Dimensi Servqual Dalam Penilaian Kualitas Layanan Rumah Sakit Swasta Dan Pemerintah Di Semarang," *J. Manaj. BISNIS DAN KEWIRUSAHAAN*, vol. 9, no. 5, 2025, doi: <https://doi.org/10.24912/jmbk.v9i5.35329>.
- [5] A. Khodadadi, N. Ghanbari Bousejin, S. Molaei, V. Kumar Chauhan, T. Zhu, and D. A. Clifton, "Improving Diagnostics with Deep Forest Applied to Electronic Health Records," *Sensors*, vol. 23, no. 14, p. 6571, Jul. 2023, doi: 10.3390/s23146571.
- [6] T. A. Jannah, A. Arifiana, M. W. Rahmawati, and Farida, "Optimalisasi Penyuluhan Pencegahan Tuberkulosis Pada Remaja: Langkah Kecil Untuk Masa Depan Sehat," *JPMS*

- (*Jurnal Pengabd. Masy. Sehati*), vol. 3, no. 2, 2024, doi: <https://doi.org/10.33651/jpms.v3i2.697>.
- [7] F. M. A. Sofyan, A. P. Riyandoro, D. F. Maulana, and J. H. Jaman, "Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 2, p. 619, Jul. 2023, doi: [10.53513/jsk.v6i2.8578](https://doi.org/10.53513/jsk.v6i2.8578).
- [8] N. S. B. Nugrahani and A. Prapanca, "Implementasi Algoritma C5.0 Pada Klasifikasi Status Gizi Balita di Kecamatan Ponorogo," *J. Informatics Comput. Sci.*, vol. 6, no. 04, pp. 1089–1098, Jun. 2025, doi: [10.26740/jinacs.v6n04.p1089-1098](https://doi.org/10.26740/jinacs.v6n04.p1089-1098).
- [9] M. S. Asih and A. Z. Hasibuan, "Penerapan Algoritma C5.0 Dalam Klasifikasi Stunting Pada Anak Dengan Smartphone Android," *J. Unitek*, vol. 17, no. 1, 2024, doi: <https://doi.org/10.52072/unitek.v17i1.850>.
- [10] M. R. Amarta, R. Wahyuni, and Y. Irawan, "Optimasi Algoritma C5.0 untuk Peningkatan Akurasi dalam Klasifikasi Ulasan Masyarakat Terhadap Layanan BPJS Kesehatan," *JEKIN - J. Tek. Inform.*, vol. 5, no. 1, pp. 101–110, Jan. 2025, doi: [10.58794/jekin.v5i1.995](https://doi.org/10.58794/jekin.v5i1.995).
- [11] M. ROFIQ, T. H. Saragih, and D. T. Nugrahadi, "Implementasi Ekstraksi Fitur GLCM dengan Klasifikasi Algoritma C5.0 Pada Data Computerized Tomography Scan Covid-19," *J. Inform. Polinema*, vol. 9, no. 4, pp. 353–362, Aug. 2023, doi: [10.33795/jip.v9i4.1280](https://doi.org/10.33795/jip.v9i4.1280).
- [12] A. Sulistyohati and F. Natsir, "Model Prediksi Keberhasilan Mahasiswa Pada Mata Kuliah Rekayasa Perangkat Lunak Menggunakan Algoritma C5.0," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 6, no. 1, 2025, doi: doi.org/10.31102/jatim.v6i1.3136.
- [13] K. S. Ningsih and I. Zufria, "Penerapan Algoritma C5.0 Untuk Memprediksi Tingkat Kepuasan Siswa Terhadap Kinerja GuruMAN Simalungun," *J. FASILKOM*, vol. 13, no. 3, 2023, doi: <https://doi.org/10.37859/jf.v13i3.6347>.
- [14] A. T. Hidayati, A. E. Widyantoro, and H. J. Ramadhani, "Perancangan Sistem Informasi Wirausaha Mahasiswa (Siwirma) Berbasis Web dengan Unified Modelling Language (UML)," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, pp. 86–107, Nov. 2023, doi: [10.55606/juprit.v2i4.2906](https://doi.org/10.55606/juprit.v2i4.2906).
- [15] M. R. Pramudia, E. Rohaini, and Y. Pratama, "Perancangan Sistem E-Lapor Pada Kantor Desa Lagan Tengah Berbasis Web," *J. Inform. Dan Rekayasa Komput.*, vol. 5, no. 1, 2025, doi: <https://doi.org/10.33998/jakakom.2025.5.1.2177>.
- [16] S. F. Sobrina and R. Gupitha, "Sistem Informasi Jasa Instalasi Dan Service Elektrik Pada Cv Mitra Teknik Alamja," *Global*, vol. 10, no. 1, pp. 43–51, 2023.
- [17] F. Muna, T. Khotimah, and A. Jazuli, "Sistem Administrasi Perpustakaan Desa Kaliputu Berbasis Web," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, 2023, doi: <https://doi.org/10.36040/jati.v7i2.6845>.