

IMPLEMENTASI ALGORITMA C4.5 UNTUK MENGLASIFIKASI KATEGORI POLI PASIEN BERDASARKAN DATA REKAM MEDIS PADA RUMAH SAKIT GIGI DAN MULUT USUHistori Mei Selamat Gulo¹, Tonni Limbong²^{1,2}Universitas Katolik Santo ThomasEmail: dodyg741@gmail.com

Abstrak: Pada era perkembangan teknologi dan informasi, data pada sistem rekam medis dapat dimanfaatkan sebagai sumber informasi untuk mendukung pengambilan keputusan dalam bidang kesehatan. Rumah Sakit Gigi dan Mulut Universitas Sumatera Utara (RSGM USU) memiliki data rekam medis yang jumlahnya terus meningkat setiap tahunnya. Namun, pemanfaatan data rekam medis di RSGM USU masih terbatas hanya pada kebutuhan administrasi dan belum dimanfaatkan secara optimal sebagai sumber informasi dalam mendukung pengambilan keputusan. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam mengklasifikasikan kategori poli pasien berdasarkan data rekam medis pasien. Penelitian ini menggunakan sebanyak 330 data rekam medis yang mencakup atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit. Atribut tersebut digunakan sebagai variabel prediktor, sedangkan kategori poli digunakan sebagai variabel target. Penelitian ini menerapkan pendekatan data mining dengan tahapan Knowledge Discovery in Database (KDD). Dalam proses pembentukan model, penelitian ini menggunakan 264 data latih dan 66 data uji. Hasil penelitian menunjukkan bahwa algoritma C4.5 memperoleh tingkat akurasi sebesar 87,88% dalam mengklasifikasikan kategori poli pasien. Hasil tersebut menunjukkan bahwa algoritma C4.5 memiliki kemampuan yang baik dalam mengenali pola hubungan antar atribut rekam medis dan berpotensi digunakan sebagai alat bantu pendukung keputusan dalam menentukan kategori poli pasien di RSGM USU.

Kata kunci: Data Mining, Algoritma C4.5, Rekam Medis, Klasifikasi.

Abstract: In the era of technological and information advancement, data from medical record systems can serve as a valuable information source to support healthcare decision-making. The University of Sumatera Utara Dental and Oral Hospital (RSGM USU) possesses a volume of medical record data that grows annually; however, the utilization of this data remains largely confined to administrative purposes and has not been fully leveraged to support decision-making processes. This study aims to apply the C4.5 algorithm to classify patient clinic categories based on medical record data. The study utilized 330 medical records, incorporating attributes such as age, gender, chief complaint, and medical history. These attributes served as predictor variables, while the clinic category acted as the target variable. A data mining approach was employed, following the Knowledge Discovery in Databases (KDD) process. The model development phase utilized 264 training records and 66 testing records. The results demonstrate that the C4.5 algorithm achieved an accuracy rate of 87.88% in classifying patient

clinic categories. These findings indicate that the C4.5 algorithm effectively identifies patterns and relationships among medical record attributes, suggesting its potential as a decision-support tool for determining patient clinic categories at RSGM USU.

Keywords: Data Mining, C4.5 Algorithm, Medical Records, Classification.

PENDAHULUAN

Rumah Sakit Gigi dan Mulut Universitas Sumatera Utara (RSGM USU) merupakan salah satu fasilitas pelayanan kesehatan di bidang gigi dan mulut yang menghasilkan data rekam medis pasien dalam jumlah yang terus bertambah setiap tahunnya. Data rekam medis tersebut memuat berbagai informasi penting, seperti usia, jenis kelamin, keluhan utama, dan riwayat penyakit pasien. Informasi tersebut memiliki potensi besar untuk dimanfaatkan sebagai sumber pengetahuan dan pendukung pengambilan keputusan dalam bidang pelayanan kesehatan. Namun, pemanfaatan data rekam medis di RSGM USU selama ini masih terbatas pada kebutuhan administrasi sehingga belum dimanfaatkan secara optimal untuk menghasilkan informasi yang mendukung proses pelayanan kesehatan.

Berdasarkan bidang pelayanan di RSGM USU, penelitian ini difokuskan pada enam kategori poli, yaitu Poli Konservasi Gigi, Poli Periodonsia, Poli Ortodonti, Poli Bedah Mulut dan Maksilofasial, Poli Prostodontia, serta Poli Penyakit Mulut (Oral Medicine). Setiap kategori poli memiliki karakteristik keluhan pasien yang berbeda-beda sehingga proses penentuan kategori poli secara manual berpotensi menimbulkan kesalahan dalam pengelompokan pasien. Oleh karena itu, diperlukan suatu pendekatan yang mampu membantu proses klasifikasi kategori poli secara cepat, tepat, dan terstruktur.

Salah satu pendekatan yang dapat digunakan untuk mengolah data rekam medis adalah data mining melalui teknik klasifikasi (classification).

Klasifikasi merupakan proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik yang dimiliki. Dalam penelitian ini, proses klasifikasi dilakukan menggunakan algoritma C4.5, yaitu algoritma berbasis pohon keputusan (decision tree) yang mampu menghasilkan aturan klasifikasi dalam bentuk IF-THEN sehingga mudah dipahami dan diinterpretasikan. Algoritma C4.5 juga memiliki kemampuan yang baik dalam menangani data kategorik dan menghasilkan model klasifikasi dengan tingkat akurasi yang tinggi pada berbagai kasus di bidang kesehatan.

Beberapa penelitian terdahulu menunjukkan bahwa algoritma C4.5 memiliki performa

yang baik dalam menyelesaikan permasalahan klasifikasi di bidang kesehatan. Penelitian oleh Santoso dkk. [10] berhasil menerapkan algoritma C4.5 untuk mengklasifikasikan potensi penyakit jantung dan menghasilkan model prediksi yang mudah dipahami. Penelitian oleh Sepharni dkk. [11] menunjukkan bahwa algoritma C4.5 mampu digunakan untuk klasifikasi penyakit jantung dengan tingkat akurasi yang baik. Selanjutnya, Muzakki dkk. [12] menerapkan algoritma C4.5 untuk prediksi penyakit jantung menggunakan RapidMiner dan menunjukkan bahwa algoritma tersebut mampu menghasilkan model klasifikasi yang efektif. Selain itu, Prakoso dan Kurniawati [13] mengembangkan model klasifikasi penyakit jantung menggunakan Decision Tree C4.5 dan menunjukkan bahwa atribut-atribut kesehatan mampu membentuk model klasifikasi yang baik. Hasil penelitian-penelitian tersebut menunjukkan bahwa algoritma C4.5 memiliki kemampuan yang baik dalam mengolah data kesehatan dan menghasilkan model klasifikasi yang mudah dipahami.

Berdasarkan penelitian-penelitian sebelumnya, penerapan algoritma C4.5 untuk mengklasifikasikan kategori poli berdasarkan data rekam medis pasien pada RSGM USU masih sangat terbatas. Sebagian besar penelitian terdahulu berfokus pada klasifikasi penyakit, sedangkan penelitian mengenai klasifikasi kategori poli berdasarkan data rekam medis pasien gigi dan mulut masih belum banyak dilakukan. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam mengklasifikasikan kategori poli pasien berdasarkan atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit menggunakan 330 data rekam medis pasien rawat jalan. Hasil penelitian diharapkan dapat memberikan model klasifikasi yang mampu membantu proses penentuan kategori poli pasien secara lebih cepat, tepat, dan akurat serta mendukung pengambilan keputusan dalam pelayanan kesehatan di RSGM USU.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan metode klasifikasi (classification). Algoritma yang digunakan dalam proses klasifikasi adalah algoritma C4.5 yang membentuk model berupa pohon keputusan (decision tree) berdasarkan nilai Gain Ratio tertinggi. Data mining merupakan proses menggali informasi dan pola yang tersembunyi dari sekumpulan data berukuran besar sehingga menghasilkan pengetahuan yang bermanfaat untuk mendukung pengambilan keputusan. Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining menggunakan algoritma C4.5.

1. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari data rekam medis pasien rawat jalan di Rumah Sakit Gigi dan Mulut Universitas Sumatera Utara (RSGM USU). Dataset yang digunakan berjumlah 330 data rekam medis pasien yang terdiri atas informasi usia, jenis kelamin, keluhan utama, riwayat penyakit, dan kategori poli. Data tersebut digunakan sebagai sumber data dalam proses klasifikasi kategori poli menggunakan algoritma C4.5.

2. Variabel Penelitian

Variabel penelitian merupakan atribut yang digunakan dalam proses klasifikasi kategori poli pasien berdasarkan data rekam medis. Pada penelitian ini, variabel yang digunakan terdiri atas empat variabel prediktor, yaitu usia, jenis kelamin, keluhan utama, dan riwayat penyakit, serta satu variabel target, yaitu kategori poli yang akan diprediksi menggunakan algoritma C4.5.

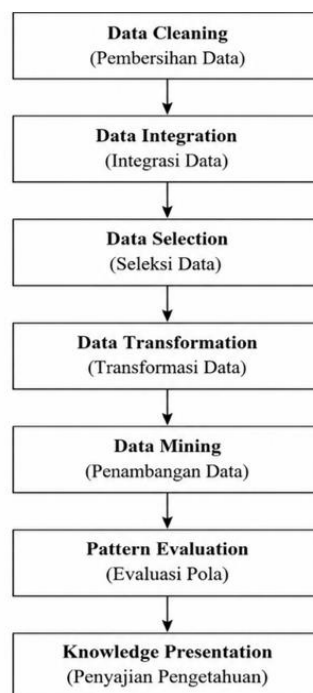
Tabel 1. Variabel Penelitian

No	Variabel	Jenis	Keterangan
1	Usia	Prediktor	Digunakan untuk mengetahui kelompok umur pasien.
2	Jenis Kelamin	Prediktor	Menunjukkan jenis kelamin pasien, yaitu laki-laki atau perempuan.
3	Keluhan Utama	Prediktor	Menunjukkan keluhan utama yang dialami pasien.
4	Riwayat Penyakit	Prediktor	Menunjukkan ada atau tidaknya riwayat penyakit pasien.
5	Kategori Poli	Target	Menunjukkan kategori poli tujuan yang akan diprediksi.

3. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan serangkaian proses untuk menemukan pengetahuan dari data yang meliputi tahapan data cleaning, data selection, data

transformation, data mining, pattern evaluation, dan knowledge presentation. Dalam penelitian ini, metode KDD digunakan sebagai tahapan pengolahan data sebelum penerapan algoritma C4.5, seperti yang telah disajikan pada gambar dibawah ini.



Gambar 1. Tahapan Penelitian Menggunakan Metode KDD

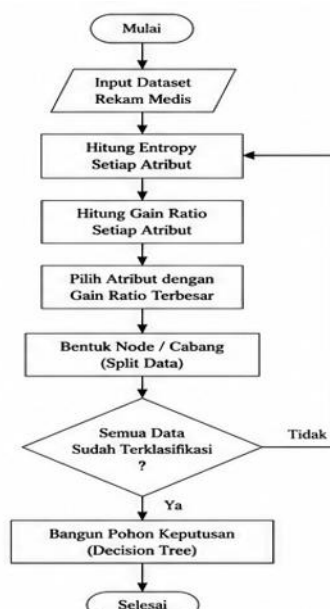
Tahapan-tahapan yang digunakan pada penelitian ini adalah sebagai berikut:

1. Data Cleaning, yaitu proses pembersihan data dari data yang tidak lengkap, terdapat duplikasi, dan tidak konsisten.
2. Data Integration, yaitu proses menggabungkan data rekam medis pasien dari berbagai sumber atau tabel terpisah ke dalam satu kesatuan dataset yang padu.
3. Data Selection, yaitu pemilihan atribut yang digunakan dalam penelitian, meliputi usia, jenis kelamin, keluhan utama, riwayat penyakit, dan kategori poli.
4. Data Transformation, yaitu proses mengubah data ke dalam format yang sesuai untuk proses data mining.
5. Data Mining, yaitu penerapan algoritma C4.5 untuk membentuk model klasifikasi kategori poli pasien.
6. Pattern Evaluation, yaitu proses evaluasi model menggunakan pengukuran akurasi untuk mengetahui kinerja model klasifikasi.

7. Knowledge Presentation, yaitu penyajian hasil penelitian dalam bentuk pohon keputusan (decision tree) dan aturan klasifikasi (IF-THEN).

4. Algoritma C4.5

Algoritma C4.5 merupakan algoritma klasifikasi berbasis pohon keputusan (decision tree) yang dikembangkan oleh Quinlan sebagai penyempurnaan dari algoritma ID3. Algoritma ini bekerja dengan memilih atribut terbaik berdasarkan nilai Gain Ratio tertinggi untuk membentuk node pada pohon Keputusan, dibawah ini telah disajikan pada gambar beberapa tahapan yang akan dilakukan dalam proses algoritma C4.5.



Gambar 2. Tahapan algoritma C4.5

Pada proses pembentukan pohon keputusan dilakukan melalui beberapa tahap perhitungan, yaitu perhitungan entropy, information gain, split information, dan gain ratio seperti rumus berikut ini.

1. Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian atau keberagaman data dalam suatu himpunan data.

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

Keterangan:

- a. S = himpunan kasus.
- b. p_i = proporsi data pada kelas ke- i .
- c. n = jumlah kelas.

Nilai entropy yang semakin besar menunjukkan bahwa data semakin heterogen, sedangkan nilai entropy yang mendekati nol menunjukkan bahwa data semakin homogen.

2. Information Gain

Information Gain digunakan untuk mengetahui kemampuan suatu atribut dalam memisahkan data ke dalam kelas yang berbeda.

$$Gain(S, A) = Entropy(S) - \sum_{v=1}^n \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan:

- a. S = himpunan kasus.
- b. A = atribut yang diuji.
- c. S_v = subset data pada nilai atribut ke- v .
- d. $|S_v|$ = jumlah data pada subset.
- e. $|S|$ = jumlah seluruh data.

Semakin besar nilai gain yang diperoleh, maka semakin baik atribut tersebut dalam membedakan kelas data.

3. Split Information

Split Information digunakan untuk mengukur potensi pemisahan data yang dihasilkan oleh suatu atribut.

$$SplitInfo(S, A) = - \sum_{v=1}^n \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

Keterangan:

- a. S = himpunan kasus.
- b. A = atribut yang diuji.
- c. S_v = subset data pada nilai atribut ke- v .

4. Gain Ratio

Gain Ratio digunakan untuk menentukan atribut terbaik yang akan dipilih sebagai node dalam pohon keputusan.

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

Keterangan:

- a. $Gain(S, A)$ = nilai information gain dari atribut A .
- b. $SplitInfo(S, A)$ = nilai split information dari atribut A .

Atribut yang memiliki nilai Gain Ratio tertinggi akan dipilih sebagai node akar (root node) pada pembentukan pohon keputusan.

Pada penelitian ini, perhitungan entropy, information gain, split information, dan gain ratio dilakukan terhadap atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit. Hasil perhitungan menunjukkan bahwa atribut Keluhan Utama memiliki nilai Gain Ratio tertinggi sehingga dipilih sebagai node akar dalam pembentukan pohon keputusan menggunakan algoritma C4.5.

HASIL DAN PEMBAHASAN

Hasil

1. Deskripsi Dataset

Penelitian ini menggunakan data sekunder berupa data rekam medis pasien rawat jalan di Rumah Sakit Gigi dan Mulut Universitas Sumatera Utara (RSGM USU). Jumlah data yang digunakan dalam penelitian ini sebanyak 330 data rekam medis pasien yang diperoleh dari data kunjungan pasien selama periode penelitian. Data rekam medis tersebut terdiri dari beberapa atribut yang digunakan sebagai variabel prediktor dan variabel target dalam proses klasifikasi.

Variabel prediktor yang digunakan meliputi usia, jenis kelamin, keluhan utama, dan riwayat penyakit, sedangkan kategori poli digunakan sebagai variabel target atau kelas yang

akan diprediksi. Pemilihan atribut tersebut didasarkan pada relevansinya terhadap proses penentuan kategori poli pasien di RSGM USU.

Sebelum dilakukan proses klasifikasi menggunakan algoritma C4.5, data terlebih dahulu melalui tahapan pembersihan, seleksi, dan transformasi sesuai dengan metode Knowledge Discovery in Database (KDD). Setelah proses preprocessing selesai, dataset kemudian dibagi menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data).

Pembagian dataset menggunakan perbandingan 80:20, di mana sebanyak 264 data digunakan sebagai data latih untuk membentuk model klasifikasi, sedangkan 66 data digunakan sebagai data uji untuk mengevaluasi performa model yang dihasilkan. Pembagian data tersebut bertujuan agar model yang dibangun dapat mempelajari pola dari data latih dan kemudian diuji menggunakan data yang belum pernah dipelajari sebelumnya sehingga tingkat akurasi model dapat diukur secara objektif.

2. Perhitungan Algoritma C4.5

Untuk memperlihatkan proses pembentukan pohon keputusan menggunakan algoritma C4.5, penelitian ini menggunakan 20 data sampel sebagai representasi dari keseluruhan dataset yang berjumlah 330 data rekam medis pasien. Perhitungan manual dilakukan dengan menghitung nilai Entropy, Information Gain, Split Information, dan Gain Ratio pada setiap atribut yang digunakan dalam penelitian, yaitu usia, jenis kelamin, keluhan utama, dan riwayat penyakit.

Tabel 2. Sampel Data Penelitian

No	Usia	J.K	Keluhan	Riwayat	Target
1	Dewasa	L	K1 Gigi berlubang dan nyeri gigi	Tidak Ada	T1 Konservasi Gigi
2	Remaja	P	K4 Sisa akar atau gigi impaksi	Tidak Ada	T4 Bedah Mulut dan Maksilofasial
3	Lansia	P	K2 Gusi bengkak dan gusi berdarah	Ada	T2 Periodonsia

No	Usia	J.K	Keluhan	Riwayat	Target
4	Dewasa	L	K5 kehilangan gigi	Tidak Ada	T5 Prostodonsia
5	Anak	L	K1 Gigi berlubang dan nyeri gigi	Tidak Ada	T1 Konservasi Gigi
6	Dewasa	P	K4 Sisa akar atau gigi impaksi	Ada	T4 Bedah Mulut dan Maksilofasial
7	Remaja	L	K3 Susunan gigi tidak rapi	Tidak Ada	T3 Ortodonti
8	Lansia	P	K6 Sariawan atau luka pada mulut	Tidak Ada	T6 Penyakit Mulut
9	Dewasa	L	K6 Sariawan atau luka pada mulut	Tidak Ada	T6 Penyakit Mulut
10	Anak	P	K1 Gigi berlubang dan nyeri gigi	Tidak Ada	T1 Konservasi Gigi
11	Dewasa	P	K2 Gusi bengkak dan gusi berdarah	Ada	T2 Periodonsia
12	Remaja	L	K3 Susunan gigi tidak rapi	Tidak Ada	T3 Ortodonti
13	Lansia	L	K4 Sisa akar atau gigi impaksi	Ada	T4 Bedah Mulut dan Maksilofasial
14	Dewasa	P	K1 Gigi berlubang dan nyeri gigi	Tidak Ada	T1 Konservasi Gigi

No	Usia	J.K	Keluhan	Riwayat	Target
15	Anak	L	K2 Gusi bengkak dan gusi berdarah	Tidak Ada	T2 Periodonsia
16	Dewasa	L	K1 Gigi berlubang dan nyeri gigi	Ada	T2 Periodonsia
17	Lansia	P	K1 Gigi berlubang dan nyeri gigi	Ada	T4 Bedah Mulut dan Maksilofasial
18	Remaja	L	K2 Gusi bengkak dan gusi berdarah	Tidak Ada	T2 Periodonsia
19	Dewasa	P	K4 Sisa akar atau gigi impaksi	Tidak Ada	T1 Konservasi Gigi
20	Lansia	L	K1 Gigi berlubang dan nyeri gigi	Tidak Ada	T1 Konservasi Gigi

Tahap pertama adalah menghitung nilai entropy dari keseluruhan data. Nilai entropy digunakan untuk mengukur tingkat ketidakpastian atau keberagaman data terhadap kategori poli pasien. Setelah nilai entropy diperoleh, dilakukan perhitungan Information Gain untuk mengetahui seberapa besar kemampuan masing-masing atribut dalam memisahkan data ke dalam kelas yang berbeda.

Selanjutnya dilakukan perhitungan Split Information dan Gain Ratio untuk menentukan atribut terbaik yang akan digunakan sebagai node pada pohon keputusan. Atribut yang memiliki nilai Gain Ratio terbesar dipilih sebagai node akar (root node) karena dianggap paling mampu membedakan kategori poli pasien. Hasil perhitungan Gain Ratio dari masing-masing atribut dapat dilihat pada tabel berikut.

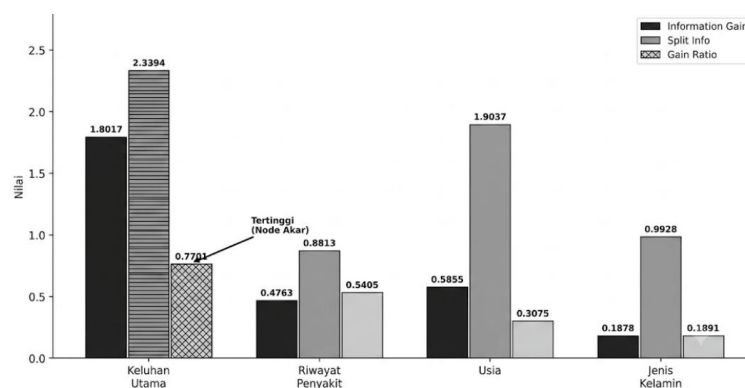
Tabel 3. Rekapitulasi Gain Ratio

Atribut	Gain	Split Info	Gain Ratio
Keluhan Utama	1,8017	2,3394	0,7701

Riwayat Penyakit	0,4763	0,8813	0,5405
Usia	0,5855	1,9037	0,3075
Jenis Kelamin	0,1878	0,9928	0,1891

Berdasarkan hasil perhitungan tersebut, atribut Keluhan Utama memiliki nilai Gain Ratio tertinggi yaitu sebesar 0,7701, sehingga dipilih sebagai node akar dalam pembentukan pohon keputusan. Setelah node akar ditentukan, proses perhitungan dilanjutkan secara berulang pada setiap cabang hingga seluruh data berhasil diklasifikasikan ke dalam kategori poli yang sesuai.

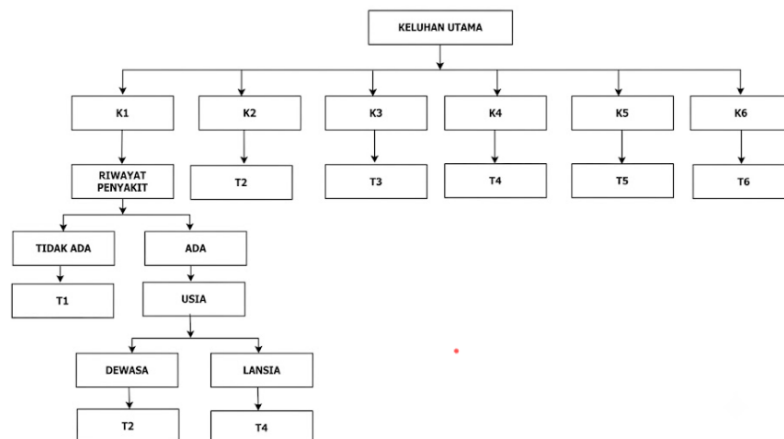
Proses perhitungan manual ini menunjukkan bahwa algoritma C4.5 mampu mengidentifikasi pola hubungan antar atribut rekam medis pasien dan menghasilkan aturan klasifikasi yang mudah dipahami dalam bentuk IF-THEN seperti yang telah disajikan pada gambar berikut.



Gambar 3. Perbandingan setiap atribut

3. Pohon Keputusan

Pohon keputusan yang terbentuk terdiri atas tiga level. Sebagian besar kategori keluhan dapat langsung dipetakan ke kategori poli tertentu, sedangkan pada kategori K1 diperlukan atribut tambahan berupa Riwayat Penyakit dan Usia,, seperti yang telah disajikan pada gambar berikut ini.



Gambar 4. Pohon keputusan

Aturan IF-THEN

1. IF Keluhan = K1 AND Riwayat Penyakit = Tidak Ada THEN T1.
2. IF Keluhan = K1 AND Riwayat Penyakit = Ada AND Usia = Remaja THEN T1.
3. IF Keluhan = K1 AND Riwayat Penyakit = Ada AND Usia = Dewasa THEN T2.
4. IF Keluhan = K1 AND Riwayat Penyakit = Ada AND Usia = Lansia THEN T4.
5. IF Keluhan = K2 THEN T2
6. IF Keluhan = K3 THEN T3
7. IF Keluhan = K4 THEN T4
8. IF Keluhan = K5 THEN T5.
9. IF Keluhan = K6 THEN T6.

4. Pengujian Model Sistem

Pengujian model dilakukan untuk mengetahui tingkat kemampuan algoritma C4.5 dalam mengklasifikasikan kategori poli berdasarkan data rekam medis pasien. Pada penelitian ini digunakan sebanyak 330 data rekam medis, yang dibagi menjadi 264 data latih (80%) dan 66 data uji (20%). Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum pernah digunakan pada proses pelatihan.

Perhitungan Akurasi

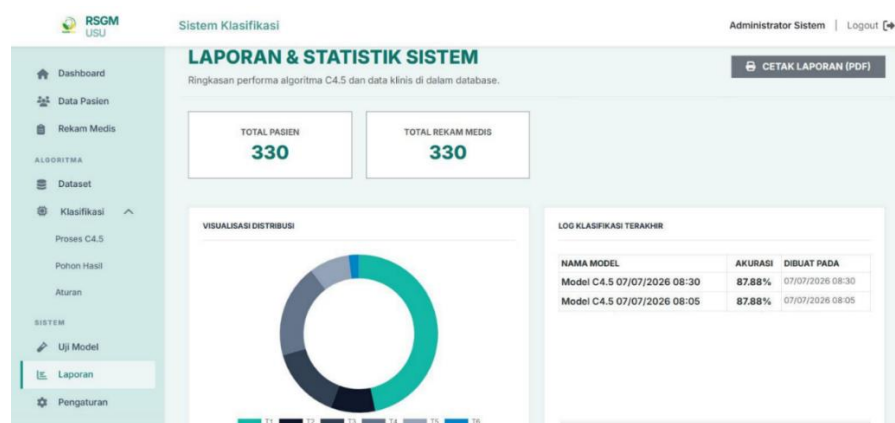
Rumus akurasi yang digunakan adalah:

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data Uji}} \times 100\%$$

Substitusi:

$$Accuracy = \frac{58}{66} \times 100\% = 87,88\%$$

Berdasarkan hasil pengujian, algoritma C4.5 menghasilkan 58 prediksi yang sesuai dari total 66 data uji, sehingga diperoleh tingkat akurasi sebesar 87,88%. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi kategori poli dengan tingkat ketepatan yang baik berdasarkan atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit seperti yang disajikan pada gambar berikut ini.



Gambar 5. Laporan statistik dan hasil akurasi

Setelah sistem selesai dibangun, dilakukan implementasi antarmuka untuk menguji proses klasifikasi serta menampilkan hasil pengolahan data. Implementasi tersebut meliputi halaman uji model dan halaman laporan sistem.

Gambar 6. Uji model pada sistem klasifikasi C.4.5

Halaman Uji Model digunakan untuk melakukan proses klasifikasi data pasien menggunakan algoritma C4.5. Pengguna memasukkan atribut yang terdiri dari usia pasien, jenis kelamin, keluhan utama, dan riwayat penyakit sebagai data masukan. Selanjutnya, sistem memproses data berdasarkan aturan klasifikasi yang telah dibentuk dan menghasilkan rekomendasi kategori tindakan (T1–T6). Pada halaman ini juga ditampilkan tingkat akurasi model sebesar 87%, yang menunjukkan bahwa model memiliki kemampuan yang baik dalam melakukan klasifikasi berdasarkan data rekam medis pasien.

Nilai akurasi tersebut menunjukkan bahwa algoritma C4.5 memiliki kemampuan yang baik dalam mengenali pola hubungan antara atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit terhadap kategori poli pasien.

Pembahasan

Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan kategori poli berdasarkan data rekam medis pasien dengan tingkat akurasi sebesar 87,88%. Nilai akurasi tersebut menunjukkan bahwa model yang dibangun memiliki performa yang baik dalam mengenali pola hubungan antara atribut usia, jenis kelamin, keluhan utama, dan riwayat penyakit terhadap kategori poli pasien.

Berdasarkan hasil perhitungan Gain Ratio, atribut Keluhan Utama memiliki nilai tertinggi sebesar 0,7701, sehingga dipilih sebagai node akar (*root node*) dalam pembentukan pohon keputusan. Hal ini menunjukkan bahwa keluhan utama merupakan faktor yang paling berpengaruh dalam menentukan kategori poli pasien di RSGM USU. Sebagian besar jenis keluhan dapat langsung dipetakan ke kategori poli tertentu, sedangkan beberapa kategori memerlukan atribut tambahan berupa riwayat penyakit dan usia.

Pohon keputusan yang dihasilkan memiliki struktur yang sederhana dan menghasilkan aturan klasifikasi dalam bentuk IF-THEN yang mudah dipahami. Aturan tersebut dapat digunakan sebagai alat bantu dalam menentukan kategori poli pasien secara lebih cepat dan terstruktur sehingga dapat mendukung efektivitas pelayanan kesehatan.

Secara keseluruhan, hasil penelitian membuktikan bahwa algoritma C4.5 dapat diterapkan pada data rekam medis pasien dan menghasilkan model klasifikasi yang cukup baik dengan tingkat akurasi sebesar 87,88%. Akurasi tersebut masih dapat ditingkatkan pada penelitian selanjutnya dengan menambah jumlah data, memperkaya atribut penelitian, dan membandingkan performanya dengan algoritma klasifikasi lainnya.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma C4.5 berhasil diterapkan untuk mengklasifikasikan kategori poli pasien berdasarkan data rekam medis di RSGM USU. Model klasifikasi yang dibangun menggunakan 330 data rekam medis menghasilkan tingkat akurasi sebesar 87,88%. Hasil penelitian menunjukkan bahwa atribut keluhan utama merupakan atribut yang paling berpengaruh dalam pembentukan pohon keputusan. Dengan demikian, algoritma C4.5 dapat digunakan sebagai pendukung pengambilan keputusan dalam menentukan kategori poli pasien secara lebih cepat dan terstruktur.

DAFTAR PUSTAKA

- Kementerian Kesehatan Republik Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 269/MENKES/PER/III/2008 tentang Rekam Medis. Jakarta, Indonesia: Kementerian Kesehatan RI, 2008.
- Konsil Kedokteran Indonesia, Pedoman Rekam Medis Kedokteran Gigi. Jakarta, Indonesia: KKI, 2012.
- S. C. Budi, Manajemen Unit Kerja Rekam Medis. Yogyakarta, Indonesia: Quantum Sinergis Media, 2011.
- U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.
- J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.
- D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. New York, NY, USA: John Wiley & Sons, 2005.
- P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Pearson Education, 2006.
- I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
- J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1993.
- M. Santoso, B. Al-Akbar, H. Nurjaya, S. A. Ramadhan, N. A. Rizky, and A. Fadillah, "Klasifikasi Potensi Penyakit Jantung Menggunakan Algoritma C4.5," *Jurnal INSAN (Journal of Information Systems Management Innovation)*, vol. 3, no. 2, pp. 96–103,

2023.

- A. Sepharni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 7, no. 2, pp. 117–126, 2022.
- F. Muzakki, I. Ubaydillah, N. R. Assyiami, and S. Soleha, "Penerapan Algoritma C4.5 untuk Prediksi Penyakit Jantung Menggunakan RapidMiner," *Jurnal Komputasi dan Analisis*, vol. 2, pp. 71–79, 2024.
- A. Prakoso and Kurniawati, "Analisis Model Klasifikasi Penyakit Jantung dengan Decision Tree C4.5 Berdasarkan Pendekatan Sentralisasi dan Federated Learning pada Dataset UCI," *Journal of System Information Technology and Electronics Engineering (J-SITEE)*, vol. 5, no. 1, pp. 10–21, 2025.
- R. Bellazzi and B. Zupan, "Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines," *International Journal of Medical Informatics*, vol. 77, no. 2, pp. 81–97, 2008.
- A. Kumar, S. Singh, and V. Sharma, "Application of Data Mining Techniques in Dental Health Care: A Review," *International Journal of Computer Applications*, vol. 117, no. 15, pp. 32–38, 2015.
- A. Wibowo and R. Amalia, "Klasifikasi penyakit pasien rawat jalan menggunakan algoritma C4.5 di RSUD X," *Jurnal Informatika Medis*, vol. 4, no. 1, pp. 33–41, 2020.
- Dewa Made Wedagama and I. G. A. A. Hartini, "Perawatan Dalam Bidang Konservasi Kedokteran Gigi Dengan Menggunakan Terapi Ozon (O3)," *Interdental Jurnal Kedokteran Gigi (IJKG)*, vol. 13, no. 2, 2017.
- C. C. P. Yudiantna, E. Karyadi, and A. Y. Anwaristi, "PENATALAKSANAAN KURETASE GINGIVA PADA KASUS CHRONIC INFLAMMATION GINGIVAL ENLARGEMENT," *JKG (Jurnal Ilmu Kedokteran Gigi)*, pp. 57–66, 2025.
- D. Desneli and A. Muryani, "Penatalaksanaan white spot lesion setelah perawatan ortodontik dengan teknik resin infiltration Management of white spot lesion after orthodontic treatment with resin infiltration technique," *Jurnal Kedokteran Gigi Universitas Padjadjaran*, vol. 31, no. 1, pp. 15–21, 2019.
- K. B. Mulut, M. R. D. S. K. Gigi, B. B. Mulut, & F. K. G. Maksilofasial, "STUDI KASUS." ADAM HUSEIN, Rawatan Prostodontik Menggantikan Kehilangan Gigi Manusia: Harapan dan Cabaran. Penerbit USM.

- Ana Mariyam Puspitasari, Dian Eka Ratnawati, and Agus Wahyu Widodo, "Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 802-810, 2018.
- Latifah Mulyana, "Penggunaan Anestesi Umum dalam Bidang Kedokteran Gigi Anak: Literature Review," *Jurnal Kesehatan Gigi dan Mulut (JKGM)*, vol. 7, no. 1, pp. 81-88, 2025.
- N. Musfira, F. Septina, and T. Agnizarridlo, "Peran Pendidikan Klinik terhadap Pencapaian Kompetensi Radiologi Kedokteran Gigi pada Mahasiswa Profesi: Sebuah Tinjauan Literatur: Indonesia," *E-Prodenta Journal of Dentistry*, vol. 6, no. 1, pp. 596-610, 2022.
- Ng Poi Wong, et al., "Perbandingan Algoritma C4.5 dan Classification and Regression Tree (CART) Dalam Menyeleksi Calon Karyawan," *Jurnal SIFO Mikroskil*, vol. 20, no. 1, pp. 11-18, 2019.
- .