

# Algoritma C4.5 pada Registrasi Pasien

**DEWI ROSMALA, RIZANDI NUGRO LIBRANTO**

Jurusan Teknik Informatika, Fakultas Teknologi Industri,  
Institut Teknologi Nasional Bandung

Email: [d\\_rossmala@itenas.ac.id](mailto:d_rossmala@itenas.ac.id)

## **ABSTRAK**

*Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Klinik dan Rumah Bersalin Melong Asih yang bertempat di Cimahi, bertugas melayani pemeriksaan umum, konsultasi masalah kandungan serta pemeriksaan ibu hamil. Proses registrasi yang panjang dapat menimbulkan antrian yang menyebabkan ketidaknyamanan pasien. Melihat kondisi tersebut diperlukan suatu sistem registrasi pada pasien untuk memprediksi waktu antrian di Klinik dan Rumah Bersalin Melong Asih. Data yang dimanfaatkan adalah data Rekam Medis Pasien (yang digunakan sebagai data training dan data testing). Data keluhan tersebut diproses menggunakan decision tree sehingga dapat menghasilkan suatu keputusan waktu penanganan pasien yang diterjemahkan ke dalam bahasa linguistik yaitu, cepat, sedang, dan lama. Partisi data 90:10 merupakan partisi terbaik karena memiliki nilai precision, recall, dan accuracy yang paling tinggi daripada partisi lainnya.*

**Kata kunci :** Registrasi Pasien, Klasifikasi, Prediksi, Algoritma C.45, Decision Tree

## **ABSTRACT**

*C4.5 algorithm is an algorithm that is used to form the decision tree. Clinic and Maternity Home Melong Asih housed in Cimahi, duty to serve the public examination, consultation and examination of the problems womb of pregnant women. Lengthy registration process can lead to queues that cause patient discomfort. Seeing these conditions we need a registration system for patients to predict the queuing time at the Clinic and Maternity Hospital Melong Asih. The data will be used here is data Patient Medical Record (which will be used as training data and data testing). The complaint data is processed using a decision tree that can produce a decision when handling patients who translated into linguistics, namely, fast, medium, and long. 90:10 is the best partition because it has the precision, recall, and the highest accuracy than other partitions.*

**Keywords :** Registration Patient, Classification, Prediction, C4.5 Algorithm, Decision Tree

## 1. PENDAHULUAN

Klinik dan Rumah Bersalin Melong Asih yang bertempat di Cimahi, bertugas melayani pemeriksaan umum, konsultasi masalah kandungan serta pemeriksaan ibu hamil. Proses registrasi yang panjang dapat menimbulkan masalah ketidaknyamanan pada pasien. Melihat kondisi tersebut diperlukan suatu sistem registrasi pasien untuk memprediksi waktu antrian di Klinik dan Rumah Bersalin Melong Asih. Data yang dimanfaatkan adalah data Rekam Medis Pasien (yang digunakan sebagai *data training* dan *data testing*). Dengan memanfaatkan data pasien, Algoritma C4.5 digunakan untuk memprediksi waktu pemeriksaan pasien yang digunakan dalam registrasi.

Dibutuhkan suatu teknik klasifikasi yang merupakan salah satu teknik untuk menganalisis data pasien tersebut. Algoritma C4.5 merupakan salah satu algoritma *Decision Tree* yang dapat digunakan untuk memprediksi suatu kejadian, algoritma ini merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J. Ross Quinlan<sup>[1]</sup>. Pemanfaatan Algoritma C4.5 pada sistem digunakan pada fitur registrasi pasien untuk proses penentuan atau prediksi waktu penanganan pasien, metode ini menentukan waktu penanganan dengan melihat keluhan yang dialami oleh pasien. Data keluhan tersebut diproses menggunakan *Decision Tree* sehingga dapat menghasilkan suatu keputusan waktu penanganan pasien yang diterjemahkan ke dalam bahasa linguistik yaitu, cepat, sedang, dan lama. Pasien dapat melakukan pendaftaran secara *online* dengan mengisi data pendaftaran dan keluhan yang dialami, kemudian pasien dapat melihat nomer antrian dan waktu pemeriksaan pada *website*, sehingga pasien tersebut dapat menunggu di rumah sampai dengan waktu periksa yang telah ditentukan.

## 2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu, studi pustaka terhadap sumber-sumber yang mendukung penelitian ini, analisis Algoritma C4.5 dalam penentuan registrasi pemeriksaan pasien, perancangan terhadap perangkat lunak (*website*) yang dibangun berdasarkan hasil yang diperoleh tahap analisis, implementasi terhadap perangkat lunak yang dibangun berdasarkan hasil yang diperoleh pada tahap perancangan, kemudian pengujian dan evaluasi terhadap hasil yang diperoleh berdasarkan kasus yang diteliti.

### 2.1 Pengumpulan Data

Data primer adalah data yang diperoleh langsung dari sumbernya dengan melakukan wawancara dan survei langsung di tempat penelitian. Data primer yang dibutuhkan pada penelitian ini seperti data pasien, data rekam medis pasien untuk menentukan *attribut* dalam pembuatan sistem, alur kerja pada Klinik dan Rumah Bersalin untuk menentukan fitur-fitur yang dibangun pada sistem, serta data set yang digunakan sebagai data *training* dan data *testing* dalam penerapan Algoritma C4.5 untuk menentukan waktu pemeriksaan pasien. Data set yang dibutuhkan dalam penelitian ini memiliki beberapa *attribut* dan kelas. *Attribut* yang digunakan terdiri umur, kontrol kehamilan, dan tujuh belas keluhan pasien yang sering terjadi pada klinik bersalin tempat penelitian. Sedangkan data sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data sekunder pada penelitian ini adalah : buku-buku, jurnal tentang Algoritma *Decision Tree* C4.5.

### 2.2 Sumber Data

Dari data yang ada, kolom yang diambil sebagai *attribut* atau variabel keputusan adalah kolom waktu. *Attribut-attribut* yang digunakan sebagai variabel penentu dalam pembentukan pohon keputusan adalah Kontrol Hamil Keluhan, Sakit Kepala, Batuk/Demam, Mual, Muntah,

Mulas, Sakit Perut/Pinggang, Sering Cape, Telat/Haid, Keputihan, Pendarahan, Keluar Cairan Ketuban, Keluar Lendir/Flek, Suntik KB, Pasang/Kontrol IUD, Ingin Hamil, Gerakan Bayi. Data set yang dibutuhkan dalam penelitian ini memiliki beberapa *attribut* dan kelas. *Attribut* yang digunakan terdiri umur, kontrol kehamilan, dan tujuh belas keluhan pasien yang sering terjadi pada klinik bersalin tempat penelitian. Berikut merupakan *attribut* dan nilai dari *attribut* yang digunakan sebagai data set pada penelitian ini.

**Tabel 1. *Attribut* dan Nilai *Attribut***

No	<i>Attribut</i>	Nilai
1	Umur	Numeris
2	Kontrol Hamil dan Keluhan	Ya atau Tidak
3	Sakit Kepala	Ya atau Tidak
4	Batuk / Demam	Ya atau Tidak
5	Mual	Ya atau Tidak
6	Muntah	Ya atau Tidak
7	Mulas	Ya atau Tidak
8	Sakit Perut / Pinggang	Ya atau Tidak
9	Sering Cape	Ya atau Tidak
10	Telat / Nyeri Haid	Ya atau Tidak
11	Keputihan	Ya atau Tidak
12	Pendarahan	Ya atau Tidak
13	Keluar Cairan Ketuban	Ya atau Tidak
14	Keluar Lendir / Flek	Ya atau Tidak
15	Suntik Kb	Ya atau Tidak
16	Pasang / Kontrol Iud	Ya atau Tidak
17	Ingin Hamil	Ya atau Tidak
18	Gerakan Bayi	Ya atau Tidak
19	Usia Kehamilan	Ya atau Tidak

Sumber : (Hasil Analisis,2016)

Kelas pada data set yang digunakan pada penelitian ini berupa kelas waktu yang merupakan lama waktu dari pemeriksaan pasien terdiri dari tiga kategori yaitu, cepat, sedang, dan lama. Penentuan kategori tersebut merupakan hasil konversi dari bentuk waktu dalam satuan menit. Jika pemeriksaan pasien kurang dari sepuluh menit maka waktu cepat, jika pemeriksaan antara sepuluh dan dua puluh menit maka waktu sedang, dan jika pemeriksaan lebih dari dua puluh menit maka waktu lama.

**Tabel 2. Waktu Tunggu**

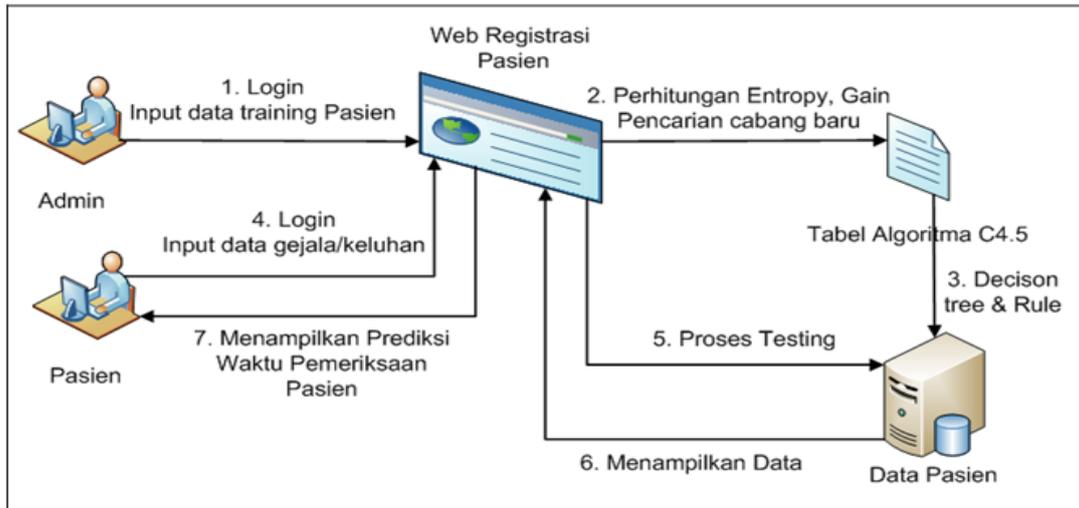
Waktu	Cepat	Sedang	Lama
Menit	<= 10	11 --- 20	>= 21

Sumber : Depkes RI, 2007

### 3. ANALISIS DAN PEMBAHASAN

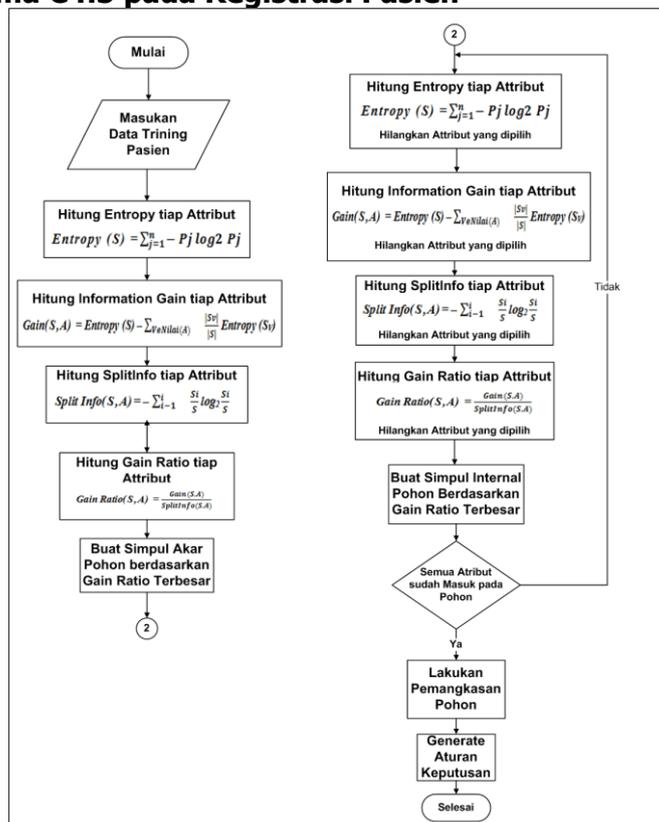
#### 3.1 Sistem Registrasi Pasien

*Website* registrasi pasien memiliki beberapa fitur yang diharapkan dapat membantu Klinik dan Rumah Bersalin Melong Asih. Proses aplikasi registrasi pasien dapat dilihat pada Gambar 1. Sistem Registrasi Pasien.



Gambar 1. Sistem Registrasi Pasien

### 3.2 Proses Algoritma C4.5 pada Registrasi Pasien



Gambar 2. Proses Algoritma C4.5 pada Registrasi Pasien

Gambar 2 merupakan proses sistem dari aplikasi *website* registrasi pasien, berikut adalah penjelasan dari setiap tahapan proses yang dilakukan sesuai dengan penomoran yang telah dilakukan : [3]

1. Admin melakukan *login* pada *website* untuk memasukkan data *training* pasien. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.

**Tabel 3. Data Training Pasien**

Nama	Tujuh Belas Atribut Gejala Keluhan	Kelas Waktu Registrasi	Jumlah
N1	Kontrol Hamil dan Keluhan, Sakit Kepala, Batuk/ Demam, Mual, Muntah, Mulas, Sakit Perut, Sering Cape, Telat/ Nyeri Haid, Keputihan, Pendarahan, Keluar Cairan Ketuban, Keluar Lendir/Flek, Suntik Kb, Pasang/ Kontrol Iud, Ingin Hamil, Gerakan Bayi	Lama	12 Kasus
N2		Sedang	16 Kasus
N3		Cepat	72 Kasus
—			
N100			
<b>Jumlah Record Keseluruhan</b>			100

Sumber : Data Rekam Medis Klinik Dan Rumah Bersalin Melong Asih

- Perhitungan dilakukan berdasarkan kasus data pasien untuk menentukan waktu pemeriksaan dilakukan dengan mengimplementasikan Algoritma C4.5 terlihat pada tabel 3. Perhitungan dilakukan berdasarkan 17 gejala keluhan yang dialami oleh pasien. Dengan nilai atribut "Ya atau Tidak". Keluhan pasien ini dijadikan *atribut* awal dalam implementasi Algoritma C4.5 dan *atribut* tujuannya adalah waktu pemeriksaan dengan nilai dari *atribut* waktu pemeriksaan adalah "Cepat", "Sedang", dan "Lama". Cepat artinya kurang dari sepuluh menit, Sedang artinya antara sepuluh dan dua puluh menit, serta Lama artinya lebih dari dua puluh menit.

**Tabel 4. Probabilitas Kelas**

Kelas	Jumlah	Probabilitas Kelas
Lama	12	12/100
Sedang	16	16/100
Cepat	72	72/100

Sumber : Data Training Pasien

Jumlah *record* keseluruhan = 100

Distribusi kelas :

Kelas Lama = 12

Kelas Sedang = 16

Kelas Cepat = 72

- Menghitung akar dari pohon. Akar diambil dari *atribut* yang terpilih, dengan cara menghitung nilai *Gain* dari masing-masing *atribut*, nilai *Gain* yang paling tinggi yang menjadi akar pertama. Sebelum menghitung nilai *Gain* dari *atribut*, hitung dahulu nilai Entropy seperti pada Gambar 2. Proses Algoritma C4.5 pada proses 1. Untuk menghitung nilai *Entropy* digunakan rumus :

$$Entropy(S) = \sum_{j=1}^n - P_j \log_2 P_j \tag{1}$$

Keterangan :

S = himpunan (dataset) kasus.

n = banyaknya partisi S.

P<sub>j</sub> = probabilitas yang di dapat dari kelas dibagi total kasus.

Perhitungan *Entropy* kelas :

$$Entropy(S) = \left(-\frac{12}{100} \left(\log_2 \left(\frac{12}{100}\right)\right)\right) + \left(-\frac{16}{100} \left(\log_2 \left(\frac{16}{100}\right)\right)\right) + \left(-\frac{72}{100} \left(\log_2 \left(\frac{72}{100}\right)\right)\right) = \mathbf{0.7542}$$

Menghitung informasi tiap *atribut* :

Distribusi kelas *atribut* Kontrol Hamil & Keluhan

**Tabel 5. Distribusi Kelas *Attribut Kontrol Hamil & Keluhan***

<i>Outcomes</i>	Lama	Sedang	Cepat	Jumlah <i>Records</i>
Tidak	12	16	50	78
Ya	0	0	22	22
Jumlah	2	2	11	100

- 2.2 Perhitungan *Entropi* dan *Information Gain* untuk *attribut* Kontrol Hamil & Keluhan, sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{v \in Nilai(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

$$Entropy(S_v) = (-\frac{12}{78} (\log_2(\frac{12}{78})) + (-\frac{16}{78} (\log_2(\frac{16}{78})) + (-\frac{50}{78} (\log_2(\frac{50}{78})))$$

$$= \mathbf{0,8637}$$

$$Information\ Gain(S,A) = Entropy(S) - ((\frac{78}{100}) * \mathbf{0,8637})$$

$$= 0,7542 - 0,6736$$

$$= \mathbf{0,0805}$$

- 2.3 Sedangkan perhitungan *Split Info* dan *Gain Ratio* untuk *attribut* Hamil & Keluhan sebagai berikut:

$$Split\ Info(S,A) = -\sum_{i=1}^i \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

$$Split\ Info(S,A) = -\frac{78}{100} (\log_2(\frac{78}{100})) = \mathbf{0,7602}$$

$$Gain\ Ratio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (4)$$

$$Gain\ Ratio(S,A) = \frac{0,0805}{0,7602} = \mathbf{0,1059}$$

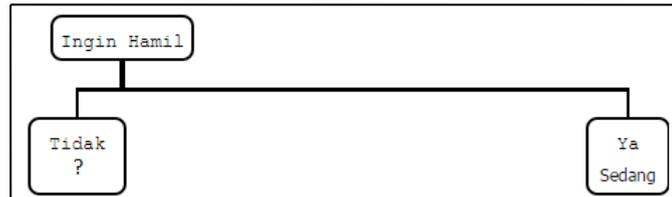
Hasil perhitungan *Information Gain* dan *Gain Ratio* dapat dilihat pada Tabel 6.

**Tabel 6. Hasil Perhitungan *Information Gain* dan *Gain Ratio* Iterasi 1**

Class	<i>Entropy(S)</i>	<i>Gain(S,A)</i>	<i>SplitInfo(S,A)</i>	<i>GainRatio(S, A)</i>
kontrol_hamil_keluhan	0.8637	0.0805	0.7602	0.1059
mual	0.8207	0.0484	0.5842	0.0828
mulas	0.6371	0.1391	0.4022	0.3458
pendarahan	0.6841	0.0323	0.5842	0.0553
keluar_cairan_keluhan	0.7309	0.0269	0.1944	0.1384
keluar_lendir_flek	0.6939	0.0881	0.2423	0.3636
suntik_kb	0.6935	0.0954	0.2864	0.3331
Pasang_kontrol_iud	0.7166	0.0439	0.3274	0.1341
<b>Ingin_hamil</b>	<b>0.6621</b>	<b>0.1384</b>	<b>0.3659</b>	<b>0.3782</b>
gerakan_bayi	0.4257	0.2154	0.9044	0.2382

- 2.4 Setelah didapatkan nilai *Gain Ratio* tiap-tiap *attribut*, dipilih *attribut* dengan nilai paling besar. Dari perhitungan pada Tabel 6, didapat *Gain Ratio* terbesar pada *attribut* **ingin\_hamil** yaitu **0.3782**. Maka *attribut* **ingin\_hamil** yang di jadikan *node*. Kemudian bagi *node* berdasarkan *attribut* yang terpilih. *Decision Tree* yang terbentuk ditunjukkan pada Gambar 3.

ingin\_hamil = Ya (Lama = 0, Sedang = 7, Cepat = 0) : Sedang  
 ingin\_hamil = Tidak (Lama = 12, Sedang = 9, Cepat = 72) : ?



Gambar 3. *Decision tree* iterasi 1

2.5 Pengecekan *node* dilakukan untuk melihat apakah *node* sudah murni atau belum. Pemilihan *attribut* yang baik adalah *attribut* yang memungkinkan untuk mendapatkan pohon keputusan yang paling kecil ukurannya atau *attribut* yang bisa memisahkan obyek menurut kelasnya. Jika *node* belum murni, maka proses membagi *node* dilakukan kembali secara berulang hingga semua *node* yang terbentuk menjadi murni dan sudah tidak ada lagi *node* yang dapat dibagi. Untuk mencari *attribut* yang menjadi *node*, dilakukan kembali menghitung informasi tiap *attribut* yang belum menjadi *node*. Seperti pada Gambar 2. Proses Algoritma C4.5 pada proses 2.

Contoh perhitungan iterasi 2 yaitu untuk pembagian *node*.

Tabel 7. Probabilitas Kelas Iterasi 2

Kelas	Jumlah	Probabilitas Kelas
Lama	12	12/93
Sedang	9	16/93
Cepat	72	72/93

Jumlah record keseluruhan = 93

Distribusi kelas :

Kelas Lama = 12  
 Kelas Sedang = 9  
 Kelas Cepat = 72

Perhitungan *Entropi* kelas iterasi 2 :

$$Entropy(S) = \left(-\frac{12}{93} \left(\log_2 \left(\frac{12}{93}\right)\right)\right) + \left(-\frac{9}{93} \left(\log_2 \left(\frac{9}{93}\right)\right)\right) + \left(-\frac{72}{93} \left(\log_2 \left(\frac{72}{93}\right)\right)\right) = \mathbf{0,6621}$$

Menghitung informasi tiap *attribut*.

Distribusi kelas *attribut* Kontrol Hamil & Keluhan Iterasi 2

Tabel 8. Distribusi Kelas *Attribut* Kontrol Hamil & Keluhan Iterasi 2

Outcomes	Lama	Sedang	Cepat	Jumlah Records
Tidak	12	9	50	71
Ya	0	0	22	22
Jumlah	12	9	72	93

Perhitungan *Entropi* dan *Information Gain* untuk *attribut* Kontrol Hamil & Keluhan Iterasi 2 :

$$Entropy(S_v) = \left(-\frac{12}{71} \left(\log_2 \left(\frac{12}{71}\right)\right)\right) + \left(-\frac{9}{71} \left(\log_2 \left(\frac{9}{71}\right)\right)\right) + \left(-\frac{50}{71} \left(\log_2 \left(\frac{50}{71}\right)\right)\right) = \mathbf{0,7783}$$

$$Information\ Gain(S,A) = Entropy(S) - \left(\frac{71}{93} * 0.7783\right)$$

$$= 0,6621 - 0,5941 = \mathbf{0,0679}$$

Sedangkan perhitungan *Split Info* dan *Gain Ratio* untuk *attribut* Hamil & Keluhan Iterasi 2 :

$$Split\ Info(S,A) = -\frac{71}{93} (\log_2 \left(\frac{71}{93}\right)) = \mathbf{0,7893}$$

$$Gain\ Ratio(S,A) = \frac{0,0679}{0,7893} = \mathbf{0,086}$$

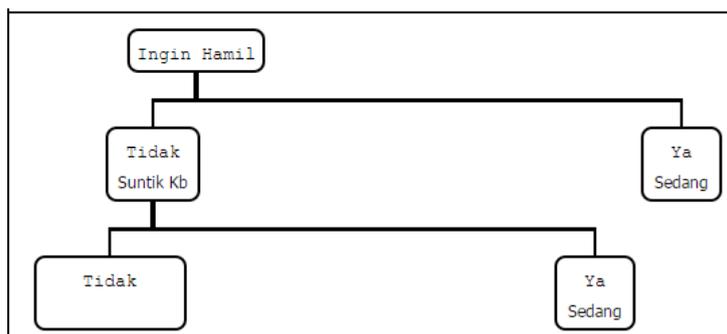
2.6 Perhitungan diulangi pada masing-masing *attribut* yang belum menjadi *node*. Hasil perhitungan *Information Gain* dan *Gain Ratio* tiap *attribut* pada iterasi kedua dapat dilihat pada Tabel 9.

**Tabel 9. Hasil perhitungan *information-gain* dan *gain-ratio* tiap *attribut* pada iterasi 2**

Class	Entropy(S)	Gain(S,A)	SplitInfo(S,A)	GainRatio(S,A)
kontrol_hamil_keluhan	0.7783	0.0679	0.7893	0.086
mual	0.7315	0.0407	0.6112	0.0666
mulas	0.5335	0.1433	0.423	0.3388
pendarahan	0.5674	0.0364	0.6112	0.0596
keluar_cairan_ketuban	0.6362	0.0267	0.2056	0.1299
keluar_lendir_flek	0.5961	0.0916	0.2559	0.358
<b>suntik_kb</b>	<b>0.5544</b>	<b>0.1375</b>	<b>0.3022</b>	<b>0.455</b>
Pasang_kontrol_iud	0.5889	0.0717	0.3451	0.2078
Ingin_hamil	0.6621	0	0	0
gerakan_bayi	0.4257	0.1432	0.8398	0.1705

3. Setelah didapatkan nilai *Gain Ratio* tiap *attribut*, dengan nilai paling besar. Maka dari perhitungan Tabel 9, didapat *Gain Ratio* terbesar pada *attribut* **suntik\_kb** yaitu **0,455**. *Attribut* **suntik\_kb** yang di jadikan *node*. Kemudian dibentuk *node* berdasarkan *attribut* yang terpilih. *Decision Tree* yang terbentuk pada iterasi 2 ditunjukkan pada Gambar 3.

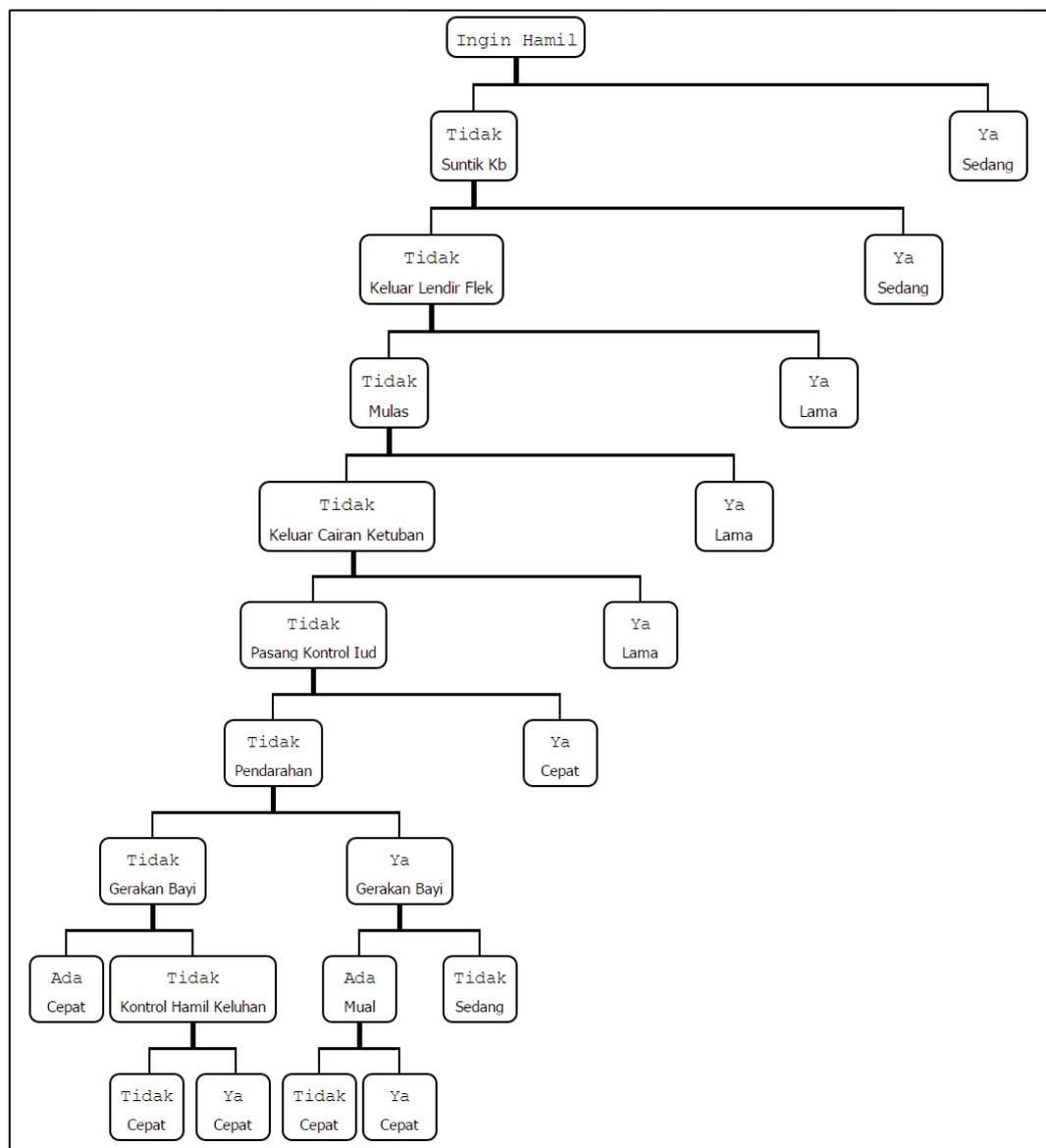
suntik\_kb = Ya (Lama = 0, Sedang = 5, Cepat = 0) : Sedang  
 ingin\_hamil = Tidak (Lama = 12, Sedang = 9, Cepat = 72) : ?



**Gambar 4. Decision tree iterasi 2**

1.1 Pengecekan dilakukan untuk melihat apakah semua *attribut* sudah terbentuk pada pohon keputusan. Jika belum, maka diulangi proses 2, jika sudah maka dilanjutkan pada proses berikutnya. Seperti pada iterasi 1, *node* dicek apakah *node* sudah murni atau belum. Jika *node* sudah murni maka *Decision Tree* terbentuk. Jika *node* belum murni maka proses diulangi sampai *node* murni dan tidak ada lagi *node* yang dapat dibagi. *Decision tree* yang terbentuk dari beberapa iterasi setelahnya dapat dilihat pada Gambar 5.

- 1.2 Pemangkasan pohon dilakukan untuk menghilangkan cabang-cabang yang tidak perlu.
- 1.3 Aturan keputusan dibentuk mengikuti pohon yang telah dibentuk sebelumnya.



Gambar 5. *Decision tree* (hasil akhir)

2. *Login* pada *website* dilakukan pasien untuk melakukan proses registrasi dengan mengisi nama, umur, serta gejala serta keluhan yang dialami.
3. Pencocokan hasil data *training* dari proses Algoritma C4.5 dengan data *testing*.
4. Data hasil perhitungan Algoritma C4.5 digunakan untuk proses pembentukan keputusan.
5. Pasien mendapatkan waktu pemeriksaan.

### 3.4 Pengukuran Kinerja

Berdasarkan penelitian yang sudah dilakukan untuk mengetahui *Precision*, *Recall*, serta *Accuracy* [3] telah diujikan dari sistem yang sudah dibuat dengan membandingkan dari 70 data *training* dan 30 data *testing* dengan proses Algoritma C4.5 dan dicocokkan dengan data asli dapat dilihat dari persamaan (5), (6), dan (7) dan pada Tabel 10.

$$Precision = \left( \frac{d}{b+d} \right) \times 100\% \quad (5)$$

$$Recall = \left( \frac{d}{c+d} \right) \times 100\% \quad (6)$$

$$Accuracy = \left( \frac{a+d}{total\ sample} \right) \times 100\% \quad (7)$$

**Tabel 10. Tabel Penilaian C4.5**

Tabel Penilaian C4.5	Diidentifikasi Cepat oleh C4.5	Diidentifikasi Lama oleh C4.5
<b>Keputusan Asli: Lama = 4</b>	a = 19	b = 2
<b>Keputusan Asli: Cepat = 22</b>	c = 1	d = 3

$$a = 19 \quad b = 2 \\ c = 1 \quad d = 3$$

1. *Precision* Lama Banding Cepat

$$Precision = \left( \frac{3}{2+3} \right) * 100\% = \mathbf{60\%}$$

2. *Recall* Lama Banding Cepat

$$Recall = \left( \frac{3}{1+3} \right) * 100\% = \mathbf{75\%}$$

3. *Accuracy* Lama Banding Cepat

$$Accuracy = \left( \frac{19+3}{4+22} \right) * 100\% = \mathbf{84.62\%}$$

Berdasarkan hasil perhitungan diatas, kesalahan prediksi tertinggi berada pada antara keputusan Lama dengan Cepat yaitu mencapai *Precision* 60 %, *Recall* 75 %, *Accuracy* 84.62 %.

#### 4. KESIMPULAN

Berdasarkan penelitian mengenai Algoritma C4.5 pada registrasi pasien didapat kesimpulan yang diperoleh setelah melakukan penelitian sebagai berikut :

1. Fitur registrasi pasien yang dibangun dengan menerapkan Algoritma C4.5 menggunakan tujuh belas *attribut* yaitu Kontrol Hamil, Sakit Kepala, dan lima belas jenis keluhan yang sering terjadi. Penerapan Algoritma C4.5 dapat memprediksi waktu pemeriksaan pasien sehingga pasien tidak menunggu lama di Klinik dan Rumah Bersalin tersebut untuk mendapatkan giliran periksa.
2. Tingkat akurasi dari implementasi Algoritma klasifikasi C4.5 dipengaruhi oleh beberapa hal seperti jenis, jumlah, isi dataset dan jumlah partisi data set. Partisi data 90:10 merupakan partisi terbaik karena memiliki nilai *precision*, *recall*, dan *accuracy* yang paling tinggi daripada partisi lainnya. Dari hasil uji coba partisi data set semakin besar jumlah partisi data maka dapat menghasilkan jumlah akurasi yang semakin tinggi pula.

## DAFTAR RUJUKAN

- [1] Al Fatta, Hanif., 2007, *Analisis Dan Perancangan Sistem Informasi*, Andi Yogyakarta.
- [2] C. Vercellis., 2009, *Bussiness Intelligence : Data Mining and Optimization for Decision Making*, John Wiley & Sons, Ltd.
- [3] F. Gorunescu., 2011, *Data Mining Concept Model Technique*, Springer India.
- [4] Han, J., Kamber, M., 2006, *Data Mining Concepts and Techniques*, Morgan Kaufman Pub San Fransisco.
- [5] Kusrini dan Luthfi, E.T., 2009, *Algoritma Data Mining*, Andi Offset Yogyakarta.
- [6] Larose, D. T., 2005, *Discovering Knowledge in Data*, John Willey & Sons, Inc New Jersey.
- [7] Perpustakaan.Depkes., 2008, *Peraturan Menteri Kesehatan Republik Indonesia. Nomor 269/Menkes/Per/Iii/2008. Tentang. Rekam Medis.*
- [8] Santosa, Budi., 2007, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu Yogyakarta.
- [9] Tan, P. N., Steinbach, M., & Kumar, V., 2005. *DATA MINING*. New York: Addison Wesley.
- [10] Witten, I. H., Frank, E., Hall, M. A., 2011, *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed), Elsevier USA.