

Klasifikasi Faktor-Faktor Penyebab Hepatitis C dengan Menggunakan Metode *Classification And Regression Tree (CART)*

Fitri Rahmawati

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, fitriirhmawatii006@gmail.com

Try Azisah Nurman

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, try.azisah@uin-alauddin.ac.id

Khalilah Nurfadilah

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, khalilah.nurfadilah@uin-alauddin.ac.id

ABSTRAK, Penelitian ini membahas mengenai klasifikasi faktor-faktor penyebab Hepatitis C. Tujuan penelitian ini adalah untuk mengidentifikasi faktor-faktor yang menyebabkan terjadinya Hepatitis C. Dimana hasil ketepatan klasifikasi yang diperoleh sebesar 94,56%. Hasil penelitian menunjukkan bahwa pasien yang mengidap penyakit Hepatitis C dengan *Aspartate Transaminase* ($< 37u/l$) sebanyak 412 orang, pasien yang mengidap dengan *Aspartate Transaminase* ($\geq 37u/l$) dan *Gamma Glutamyl Transferase* ($< 55u/l, \geq 55u/l$) sebanyak 17 orang, pasien yang mengidap penyakit Hepatitis C dengan *Aspartate Transaminase* ($\geq 37u/l$), *Gamma Glutamyl Transferase* ($< 55u/l$) dan Albumin (Normal) sebanyak 22 orang dan pasien dengan *Aspartate Transaminase* ($\geq 37u/l$), *Gamma Glutamyl Transferase* ($< 55u/l$) dengan Albumin (Rendah, Tinggi) sebanyak 5 orang yang mengidap Hepatitis C.

Kata Kunci: *Hepatitis C, Classification and Regression Tree (CART)*

1. PENDAHULUAN

Hepatitis C merupakan peradangan pada organ hati yang disebabkan adanya infeksi virus Hepatitis C (HCV). Sebagian penderita penyakit ini dapat mengalami penyakit liver kronis hingga mengalami kanker hati. Selama infeksi awal, biasanya orang sering mengalami gejala ringan atau tanpa gejala, sehingga penderita tidak mengetahui bahwa dirinya menderita penyakit hati sampai kondisi penyakitnya sudah kronis.

Menurut laporan global tahun 2017 WHO memperkirakan bahwa 71 juta orang di seluruh dunia hidup dengan penyakit Hepatitis C kronis. Virus bisa bertahan di hati pada sekitar 75% sampai 85% dari mereka yang awalnya terinfeksi. Gejala akan muncul paling lama 6 bulan setelah terpapar virus, penyakit hati dapat menyebabkan masalah kesehatan yang serius, seperti kerusakan hati, sirosis, kanker hati bahkan kematian. Virus ini menular dengan cara paparan darah injeksi yang tidak aman, transfusi

darah yang tidak disaring, penggunaan narkoba suntikan dan praktik seksual dengan orang yang terpapar. Mendiagnosis pasien yang terkena Hepatitis C itu adalah sebuah tantangan karena orang yang terkena penyakit akut umumnya hadir dengan gejala ringan seperti flu yang tidak spesifik sementara transisi dari akut ke kronis adalah sub-klinis.

Perlu dilakukan pengklasifikasian atau pengelompokan data pasien Hepatitis C sesuai faktor-faktor yang berpengaruh terhadap pasien Hepatitis C. Klasifikasi pasien Hepatitis C dengan metode CART berdasarkan hasil uji laboratorium memungkinkan pasien tersebut diklasifikasikan sebagai pasien Hepatitis C yang mengidap atau pasien Hepatitis C yang tidak mengidap.

Menentukan diagnosis penyakit Hepatitis bagi dokter termasuk tugas yang sangat sulit karena banyak faktor yang harus dipertimbangkan dalam prosedur diagnosis penyakit. Seorang dokter dapat mengambil keputusan dengan cara memeriksa hasil laboratorium tes seorang pasien. Dataset Hepatitis C virus mengandung 5 atribut dimana atribut tersebut adalah Albumin, *Alanine Transaminase*, *Aspartate Transaminase* dan *Gamma Glutamyl Transferase*. Maka dari itu, penelitian ini berguna untuk mengklasifikasikan pasien Hepatitis C yang mengidap dan pasien Hepatitis C yang tidak mengidap dan mampu dilakukan proses peengkategorian berdasarkan faktor-faktor penyebab Hepatitis C.

Beberapa cara pengklasifikasian diantaranya yaitu, *Classification and Regression Trees (CART)*, *Analisis Regresi Logistik*, *Chi-Square Automatic Interaction Detectio (CHAID)*, *Support Vector Machine (SVM)*, dan berbagai cara yang lain. Cara yang diterapkan

pada penelitian ini yaitu *Classification and Regression Trees* (CART).

CART merupakan suatu metode nonparametrik, dimana dapat menjadi alternatif karena memiliki beberapa kelebihan yakni, hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat perhitungannya. Proses CART dikenal dengan istilah *binary recursive partition* yang berarti bahwa setiap *parent node* akan mengalami pemecahan tepat menjadi dua *child node* dan setiap *child node* memiliki siklus berulang untuk menjadi *parent node* kembali. CART bertujuan untuk mendapatkan data-data yang sangat akurat sebagai pengklasifikasian. Hubungan antara satu set variabel prediktor dan variabel respon dapat dijelaskan dengan menggunakan metode ini.

2. TINJAUAN PUSTAKA

Data Mining

Data mining adalah istilah yang digunakan untuk menggambarkan penemuan pengetahuan di dalam database, proses ini menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk memperoleh dan mengenali informasi yang bermanfaat sehingga manusia mendapatkan struktur dari sekumpulan data yang mudah dimengerti. Dengan memeriksa dalam sekumpulan besar data yang tersimpan biasanya tidak disadari keberadaan dalam penyimpanannya maka teknik ini dapat menemukan hubungan yang berarti pola dan kecenderungan.[1] Pada proses pemecahan masalah dan pencarian pengetahuan baru dalam data mining terdapat beberapa klasifikasi secara umum yaitu estimasi, asosiasi, klasifikasi, *clustering* dan prediksi.[2]

Proses Data Mining

Data mining adalah rangkaian proses yang terbagi-bagi menjadi tahapan. Beberapa tahapan dalam pemrosesan data mining terbagi menjadi enam bagian yaitu pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data, proses mining [3] dan interpretasi data [4].

Decision Trees

Decision trees adalah teknik yang berguna untuk mengklasifikasikan beberapa set objek.

Pada tahap ini, semua atribut dievaluasi oleh sekumpulan simpul keputusan yang dihubungkan oleh cabang yang dimulai pada simpul akar dan berakhir pada simpul daun. Pohon keputusan adalah jenis diagram alur yang mencoba memperoleh atribut yang benar saat menentukan dan menyelesaikan sebuah kelas yang terbentuk seperti pohon [5].

Sebuah pohon keputusan biner bisa dibangun menggunakan algoritma pohon keputusan, dengan setiap *node* pada pohon membelah menjadi dua cabang. Pohon keputusan non-biner, di mana semua *node* internal juga menyebar menjadi dua *node* atau lebih yang dihasilkan oleh algoritma dari pohon keputusan yang lain.[6]

Classification and Regression Tree (CART)

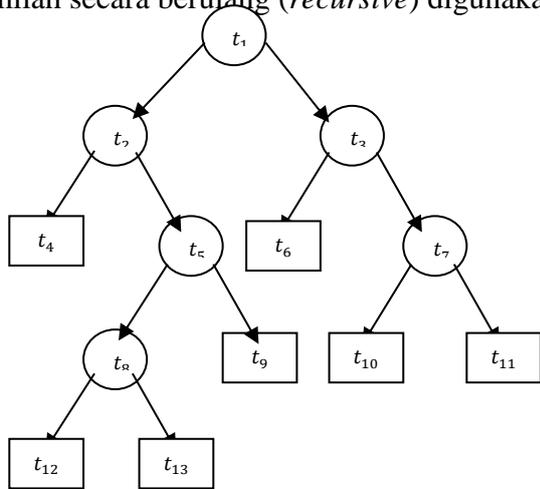
CART adalah salah satu metode statistik nonparametrik, yang berguna dalam pembentukan model pohon keputusan. Metode ini dapat mendeskripsikan keterikatan antara variabel respon (*dependen*) dan variabel prediktor (*independen*) yang lebih dari satu. Apabila variabel respon data bersebelahan, bentuk pohon yang diperoleh akan menjadi *regression tree* (pohon regresi) dan jika variabel respon yang memiliki skala kategori pohon yang diperoleh akan menjadi *classification tree*[7].

Analisis pohon klasifikasi adalah salah satu metode *binary recursive partitioning*. *Binary* berarti suatu kelompok data yang dikumpulkan dalam suatu ruangan dideskripsikan sebagai simpul, dengan syarat bahwa simpul tersebut hanya bisa diklasifikasikan kedalam dua kelompok dalam simpul utama yang disebut sebagai *parent node*. *Recursive* merupakan proses pemisahan biner dengan berulang-ulang. Berarti bahwa semua simpul utama bisa dipecah ke dalam dua simpul anak, dan semua simpul anak bisa membentuk dirinya sendiri sehingga muncul simpul anak yang lain. Sedangkan *partitioning* berarti bahwa prosedur klasifikasi bisa dilakukan dalam beberapa bagian dengan memilih kumpulan data [8].

Algoritma CART

Algoritma CART terdiri dari dua tahap dasar, tahap pertama dari analisis CART adalah

pembuatan pohon klasifikasi. Di mana pemilihan pemilah secara berulang (*recursive*) digunakan



Gambar 2.2 Struktur Pohon Klasifikasi

dalam prosedur pembentukan. Setelah itu, kelas yang diprediksi akan ditandai dan akan dianalisis apakah setiap *node* bisa dipecah kedalam *node* anak lagi atau tidak. Kedua, adalah *pruning* yang juga dikenal sebagai pemangkasan pohon klasifikasi. Proses ini mengurangi kompleksitas rangkaian pohon klasifikasi dengan menghapus *node* yang kebutuhannya meningkat. Dalam penerapan algoritma CART dapat dilakukan langkah-langkah sebagai berikut:

1. Membentuk pohon klasifikasi

Langkah pertama dalam tahap ini adalah menentukan variabel mana dan nilainya yang dapat digunakan sebagai pemilah untuk setiap simpul. Pemilihan pemilah, penentuan simpul terminal dan penandaan label kelas adalah tga langkah dalam proses pembuatan pohon klasifikasi.

a. Pemilihan pemilah

Langkah pemilihan pemilah dijalankan pemilah pada sampel data *learning* (L). Jenis variabel dependen dapat digunakan untuk mengidentifikasi sampel yang diperoleh melalui proses pemilihan pemilah dan menerapkan standar *goodness-of-split*. Hasil dari himpunan bagian dalam proses pemilihan harus lebih homogen dari pada pemilihan sebelumnya. Terlebih dahulu mencari *gain information* pada semua simpul dengan rumus berikut:

$$GI(t) = -\sum_{j=1}^n P(j|t) \log_2 P(j|t) \quad (2.1)$$

di mana:

$GI(t)$ = Gain Information pada *node* t

$P(j|t)$ = Peluang kelas j pada *node* t di mana $j = 1,2,3,4,\dots,n$ untuk $P(j|t) = \frac{n_j(t)}{n(t)}$

$n_j(t)$ = Jumlah proses pengamatan kelas j pada *node* t

$n(t)$ = Jumlah proses pengamatan pada *node* t

Terdapat simpul yang dipilih dari hasil perhitungan tingkat keragaman suatu kelas untuk mendapatkan fungsi pemilah yang maksimal dengan cara pemilihan pada klasifikasi dengan *indeks gini* ($i(t)$). *Indeks gini* yaitu fungsi keragaman yang menghasilkan pohon biner. Definisi *indeks gini* adalah mengukur variasi dalam peluang menghitung nilai atribut target dari 0 hingga 1. Fungsi dari indeks gini terhadap t adalah:

$$i(t) = 1 - \sum_{j=1}^n P^2(j|t) \quad (2.2)$$

di mana:

$i(t)$ = Indeks gini

Pemilihan yang terpilih akan terbentuk menjadi kumpulan kelas yang disebut simpul. Simpul tersebut akan melakukan pemilahan dengan rekursif hingga memperoleh terminal nodes. Selanjutnya yaitu menetapkan kriteria *goodness of split* untuk mengevaluasi pemilah dari pemilah s pada simpul t dengan rumus:

$$\phi(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.3)$$

di mana:

t_R = Simpul kanan

t_L = Simpul kiri

P_R = Peluang banyaknya objek menuju simpul kanan

P_L = Peluang banyaknya objek menuju simpul kiri

Menghitung t_R dan t_L dengan memasukkan peluang peristiwa banyaknya objek, menggunakan rumus berikut:

$$P_R = \frac{\text{bakal simpul kanan}}{\text{data latihan}} \quad (2.4)$$

$$P_L = \frac{\text{bakal simpul kiri}}{\text{data latihan}} \quad (2.5)$$

Pengembangan pohon dilakukan dengan menganalisis semua kemungkinan pemilah pada simpul t, yang kemudian akan ditentukan pemilah s^* yang menghasilkan nilai penurunan keheterogenan tertinggi yaitu:

$$\phi(s^*, t) = \max_{s \in S} \phi(s, t) \quad (2.6)$$

b. Penentuan Simpul Terminal

Perkembangan pohon akan berakhir jika pada simpul memuat pengamatan yang jumlahnya kurang dari atau sama dengan 5. Selain itu, proses pembentukan pohon juga akan berhenti jika telah mendekati batas jumlah level yang sudah dianalisis atau tingkat kedalaman dalam pohon terbaik.

$$R(t) > R(t_R) + R(t_L) \quad (2.7)$$

di mana:

$R(t)$ = Ketidaktepatan klasifikasi sesuai node t

$R(t_R)$ = Ketidaktepatan klasifikasi sesuai node t simpul kanan

$R(t_L)$ = Ketidaktepatan klasifikasi sesuai node t simpul kiri

c. Penentuan label kelas

Penentuan label kelas pada simpul terminal didasarkan pada aturan jumlah terbanyak, yaitu jika

$$P(j|t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.8)$$

di mana:

$P(j|t)$ = Peluang kelas j pada simpul t

$N_j(t)$ = Total pengamatan kelas j pada terminal node t

$N(t)$ = Total pengamatan pada terminal node t

Label kelas pada simpul terminal t adalah j_o yang mendapatkan nilai perkiraan ketidaktepatan klasifikasi pada simpul t yang sangat kecil sebesar $r(t) = 1 - \max_j P(j|t)$.

2. Pemotongan pohon klasifikasi

Dengan menggunakan aturan pemilah dan kriteria *goodness of split* maka pohon terbentuk dengan ukuran yang sangat besar. Hal tersebut dipengaruhi dengan adanya penghentian pohon yang didasarkan pada jumlah pengamatan dari simpul terminal atau tingkat homogenitas. Mengukur pohon dapat dilakukan dengan pemotongan pohon menggunakan *cost complexity minimum*.

$$g_m(t) = \frac{R(t) - R(T_k)}{|\tilde{T}_k| - 1} \quad (2.9)$$

di mana:

$g_m(t)$ = Kompleksitas parameter untuk node t

$R(t)$ = Kesalahan pada *node* t

T_k = Part tree ke-k, untuk $k = 1, 2, 3, 4, \dots, n$

$R(T_k)$ = Ketidaktepatan pengklasifikasian pada pohon T_k dengan $R(T_k) = \sum_m R(t_{mk})$

$R(t_{mk})$ = Ketidaktepatan pengklasifikasian pada simpul t ke m pada pohon ke-k

\tilde{T}_k = Terminal node pada pohon T_k

$|\tilde{T}_k|$ = Banyaknya terminal node pada pohon T_k

Unit yang dihilangkan adalah unit yang nilainya terkecil pada $g_m(t_m)$, yaitu:

$$g_m(t_m) = \min_{t \in t_k} g_m(t) \quad (2.10)$$

Pengklasifikasian simpul (node) t yang mempunyai kekeliruan dapat menggunakan rumus berikut:

$$R(t) = r(t) \cdot P(t) \quad (2.11)$$

di mana:

$r(t)$ = Kesalahan pengklasifikasian node t pada peluang kejadian di mana $r(t) = 1 - \max_j P(j|t)$

$P(t)$ = Banyaknya proporsi setiap objek pada simpul.[9]

Ukuran Ketepatan Klasifikasi

Ketepatan dari klasifikasi merupakan suatu pertimbangan untuk melihat peluang kekeliruan klasifikasi yang didapatkan berdasarkan suatu fungsi klasifikasi. Dalam mengukur ketepatan klasifikasi dapat dilakukan dengan menggunakan metode yang umum seperti dengan perhitungan *APER (apparent error rate)*, *sensitivity*, *spesificity*, *1-APER (total accuracy rate)*. *APER* merupakan peluang observasi yang diperkirakan secara tidak benar. *Sensitivity* menggambarkan akurasi pada sampel kelas i. *Spesificity* menggambarkan akurasi pada kelas j, dan tingkat akurasi total merupakan peluang observasi yang diperkirakan secara benar oleh fungsi klasifikasi [10].

Tabel untuk menghitung ketepatan klasifikasi yang ditunjukkan dalam Tabel 2.1. Adapun dalam menghitung total akurasi, *sensitivity* dan *spesificity* formulasi sebagai berikut:

$$Sensitivity = \frac{n_{11}}{N_1} \quad (2.12)$$

$$Spesificity = \frac{n_{22}}{N_1} \quad (2.13)$$

$$Akurasi = \frac{n_{11} + n_{22}}{N} \quad (2.14)$$

Tabel 2.1 Crosstab Ketepatan Klasifikasi

Kelas Pengamatan Y	Kelas Prediksi Y		Total
	1	2	
1	n_{11}	n_{12}	N_1
2	n_{21}	n_{22}	N_2
Total	N_1	N_2	N

Keterangan:

n_{11} = Jumlah pengamatan dari variabel Y kelas 1 yang tepat prediksi sebagai variabel Y kelas 1
 n_{12} = Jumlah pengamatan dari variabel Y kelas 1 yang salah prediksi sebagai variabel Y kelas 2
 n_{21} = Jumlah pengamatan dari variabel Y kelas 2 yang salah prediksi sebagai variabel Y kelas 1
 n_{22} = Jumlah pengamatan dari variabel Y kelas 2 yang tepat prediksi sebagai variabel Y kelas 2
 N_1 = Jumlah pengamatan dari variabel Y kelas 1
 N_2 = Jumlah pengamatan dari variabel Y kelas 2
 N = Jumlah keseluruhan pengamatan

Hepatitis C

Hepatitis C adalah infeksi parenkim hati akut yang disebabkan oleh virus hepatitis c (HVC). Virus hepatitis c ini dapat menyebabkan sirosis dan penyakit hati kronis dan merupakan indikator umum transplantasi hati di berbagai negara, mulai dari tingkat keparahan penyakit ringan hingga parah yang berlangsung berminggu-minggu [11]. Virus Hepatitis C adalah virus yang ditularkan melalui darah, yang paling sering ditularkan melalui transfusi darah yang tidak disaring, suntikan yang tidak aman, suntikan obat, dan aktivitas seksual dengan orang yang terpapar [12].

3. METODOLOGI

Jenis penelitian yang dipakai dalam penelitian ini adalah penelitian terapan dengan jenis data yang digunakan dalam penelitian ini adalah data sekunder Hepatitis C Virus (HCV data) yang diambil dari *Uci Machine Learning*. Variabel yang digunakan dalam penelitian terdiri dari :

1. Variabel Bebas (Independen)

Variabel bebas dalam penelitian ini adalah sebagai berikut:

Tabel 3.1 Variabel Bebas dalam Penelitian

Variabel	Nama Variabel	Skala	Kategori
X_1	Albumin	Ordinal	1 : Rendah 2 : Normal 3 : Tinggi
X_2	Alanine Transaminase	Nominal	1 : < 47u/l 2 : ≥ 47u/l
X_3	Aspartate Transaminase	Nominal	1 : < 37u/l 2 : ≥ 37u/l
X_4	Gamma Glutamyl Transfere	Nominal	1 : < 55u/l 2 : ≥ 55u/l

2. Variabel Terikat (Dependen)

Variabel terikat dalam penelitian ini adalah sebagai berikut:

$$Y = \begin{cases} 1 = \text{Pasien yang mengidap hepatitis c} \\ 0 = \text{Pasien yang tidak mengidap hepatitis c} \end{cases}$$

4. PEMBAHASAN

Deskripsi Data

Data penelitian ini diambil dari *Uci Machine Learning*.

Tabel 4.1 Jumlah pasien hepatitis c berdasarkan Albumin

		Status Pasien Hepatitis				Jumlah	Total
		Pasien Hepatitis c yang mengidap		Pasien hepatitis c yang tidak mengidap			
		N	%	N	%		
Albumin	Rendah	17	3,25%	26	4,97%	43	523
	Normal	45	8,60%	387	73,99%	432	
	Tinggi	5	0,95%	43	8,22%	48	

Pada Tabel 4.1 dijelaskan bahwa jumlah keseluruhan pasien Hepatitis C yang memiliki Allbumin rendah yang mengidap Hepatitis C adalah sebesar 3,25% (17,n=523) dan untuk pasien Hepatitis C yang memiliki Albumin rendah yang tidak atau pernah mengidap Hepatitis C sebesar 4,97% (26, n=523). Untuk

pasien Hepatitis C yang memiliki Albumin normal yang mengidap Hepatitis C adalah sebesar 8,60% (45, n=523) dan untuk pasien Hepatitis C yang mempunyai Albumin normal yang tidak atau pernah mengidap Hepatitis C sebesar 73,99% (387, n=523). Untuk pasien Hepatitis C yang memiliki Albumin tinggi yang mengidap Hepatitis C adalah sebesar 0,95% (5, n=523) dan untuk pasien Hepatitis C yang memiliki Albumin tinggi yang tidak atau pernah mengidap Hepatitis C sebesar 8,22 (43, n=523).

Tabel 4.2. Jumlah pasien hepatitis c berdasarkan *Alanine Transaminase*

		Status Pasien Hepatitis				Jumlah	Total
		Pasien Hepatitis c yang mengidap		Pasien hepatitis c yang tidak mengidap			
		N	%	N	%		
<i>Alanine Transaminase</i>	< 47u/l	53	10,13%	421	80,49%	474	523
	≥ 47u/l	14	2,67%	35	6,69%	49	

Pada Tabel 4.2 di atas dapat dijelaskan bahwa jumlah pasien Hepatitis C yang memiliki *Alanine Transaminase* < 47u/l yang mengidap Hepatitis C sebesar 10,13% (53, n=523) dan untuk pasien Hepatitis C yang memiliki *Alanine Transaminase* < 47u/l yang tidak atau pernah mengidap Hepatitis C sebesar 80,49% (421, n=523). Untuk pasien Hepatitis C yang memiliki *Alanine Transaminase* ≥ 47u/l yang mengidap Hepatitis C sebesar 2,67% (14, n=523) dan untuk pasien Hepatitis C yang memiliki *Alanine Transaminase* ≥ 47u/l yang tidak atau pernah mengidap Hepatitis C sebesar 6,69% (35, n=523).

Tabel 4.3. Jumlah pasien hepatitis c berdasarkan *Aspartate Transaminase*

		Status Pasien Hepatitis				Jumlah	Total
		Pasien Hepatitis c yang mengidap		Pasien hepatitis c yang tidak mengidap			
		N	%	N	%		
<i>Aspartate Transaminase</i>	< 37u/l	10	1,91%	412	78,77%	422	523
	≥ 37u/l	57	10,89%	44	8,41%	101	

Dilihat pada Tabel 4.3 di atas dapat dijelaskan bahwa jumlah pasien Hepatitis C yang memiliki *Aspartate Transaminase* < 37u/l yang mengidap Hepatitis C sebesar 1,91% (10, n=523) dan untuk pasien Hepatitis C yang memiliki *Aspartate Transaminase* < 37u/l yang tidak atau pernah mengidap Hepatitis C sebesar 78,77% (412, n=523). Untuk pasien Hepatitis C yang memiliki *Aspartate Transaminase* ≥ 37u/l yang mengidap Hepatitis C sebesar 10,89% (57, n=523) dan untuk pasien Hepatitis C yang memiliki *Aspartate Transaminase* ≥ 37u/l yang tidak atau pernah mengidap Hepatitis C sebesar 8,41% (44, n=523).

Tabel 4.4. Jumlah pasien hepatitis c berdasarkan *Gamma Glutamyl Transferse*

		Status Pasien Hepatitis				Jumlah	Total
		Pasien Hepatitis c yang mengidap		Pasien hepatitis c yang tidak mengidap			
		N	%	N	%		
<i>Gamma Glutamyl Transferse</i>	< 55u/l	39	7,45%	407	77,82%	446	523
	≥ 55u/l	28	5,35%	49	9,36%	77	

Pada Tabel 4.4 di atas dapat dijelaskan bahwa jumlah pasien Hepatitis C yang memiliki *Gamma Glutamyl Transferse* < 55u/l yang mengidap Hepatitis C sebesar 7,45% (39, n=523) dan untuk pasien Hepatitis C yang memiliki *Gamma Glutamyl Transferse* < 55u/l yang tidak atau pernah mengidap Hepatitis C sebesar 77,82% (407, n=523). Untuk pasien Hepatitis C yang memiliki *Gamma Glutamyl Transferse* ≥ 55u/l yang mengidap Hepatitis C sebesar 5,35% (28, n=523) dan untuk pasien Hepatitis C yang memiliki *Gamma Glutamyl Transferse* ≥ 55u/l yang tidak atau pernah mengidap Hepatitis C sebesar 9,36% (49, n=523).

Tahapan Menganalisis Data

a. Membagi data menjadi dua bagian data learning dan data testing.

Data hcldata dari *Machine Learning* diolah kembali sesuai pengkategorian. Dalam proses analisisnya, data dibagi menjadi dua data learning 85%, dan data testing 15%. Dimana data learning sebanyak 523 sampel dan data testing sebanyak 92 sampel.

b. Menentukan bentuk pohon klasifikasi maksimal dengan langkah-langkah sebagai berikut:

- 1) Pemilihan pemilah mengikuti aturan pemilihan indeks gini yang kemudian dievaluasi menggunakan kriteria *goodness of split*.

Tabel 4.5. Data Pemilah kadindat simpul kiri dan kanan

No.	Kandidat Simpul Kiri	Kandidat Simpul Kanan
1	Albumin=Rendah	Albumin=Normal,Tinggi
2	Albumin=Normal	Albumin=Rendah,Tinggi
3	Albumin=Tinggi	Albumin=Rendah,Normal
4	Alanine Transaminase(< 47u/l)	Alanine Transaminase(≥ 47u/l)
5	Aspartate Transaminase(< 37u/l)	Aspartate Transaminase(≥ 37u/l)
6	Gamma Glutamyl Transferse(< 55u/l)	Gamma Glutamyl Transferse(≥ 55u/l)

Pada Tabel 4.5 di atas memperlihatkan calon simpul kiri adalah Albumin rendah, Albumin normal, Albumin tinggi, Alanine Transaminase (< 47u/l), Aspartate Transaminase (< 37u/l), dan Gamma Glutamyl Transferse (< 55u/l). Sedangkan calon simpul kanan adalah Albumin normal, tinggi, Albumin rendah, tinggi, Albumin rendah, normal, Alanine Transaminase (≥ 47u/l), Aspartate Transaminase (≥ 37u/l), dan Gamma Glutamyl Transferse (≥ 55u/l). Sehingga parent node pada pohon keputusan yang terbentuk akan dipilih berdasarkan pembagian simpul dari simpul diatas.

2) Gain Information

Langkah awal yaitu menganalisis nilai peluang untuk setiap simpul untuk mempermudah menentukan nilai *gain information*.

Tabel 4.6 Probabilitas Simpul

Simpul	Simpul kiri	Simpul kanan	P_L	P_R	Kelas	$P(j t_L)$	$P(j t_R)$
1	Albumin=Rendah	Albumin=Normal,Tinggi	$\frac{43}{523} = 0,08221$	$\frac{480}{523} = 0,91778$	Hepatitis C	$\frac{17}{43} = 0,39534$	$\frac{50}{480} = 0,10416$

					Tidak Hepatitis C	$\frac{26}{43} = 0,60465$	$\frac{430}{480} = 0,89583$
2	Albumin=Normal	Albumin=Rendah,Tinggi	$\frac{432}{523} = 0,82600$	$\frac{91}{523} = 0,17399$	Hepatitis C	$\frac{45}{432} = 0,10416$	$\frac{22}{91} = 0,24175$
					Tidak Hepatitis C	$\frac{387}{432} = 0,89583$	$\frac{69}{91} = 0,75824$
3	Albumin=Tinggi	Albumin=Rendah,Normal	$\frac{48}{523} = 0,09177$	$\frac{475}{523} = 0,90822$	Hepatitis C	$\frac{5}{48} = 0,10416$	$\frac{62}{475} = 0,13052$
					Tidak Hepatitis C	$\frac{43}{48} = 0,89583$	$\frac{413}{475} = 0,86947$
4	Alanine Transaminase(< 47u/l)	Alanine Transaminase(≥ 47u/l)	$\frac{474}{523} = 0,90630$	$\frac{49}{523} = 0,09369$	Hepatitis C	$\frac{53}{474} = 0,11181$	$\frac{14}{49} = 0,28571$
					Tidak Hepatitis C	$\frac{421}{474} = 0,88818$	$\frac{35}{49} = 0,71428$
5	Aspartate Transaminase(< 37u/l)	Aspartate Transaminase(≥ 37u/l)	$\frac{424}{523} = 0,80688$	$\frac{101}{523} = 0,19311$	Hepatitis C	$\frac{10}{422} = 0,02369$	$\frac{57}{101} = 0,56435$
					Tidak Hepatitis C	$\frac{412}{422} = 0,97630$	$\frac{44}{101} = 0,43564$
6	Gamma Glutamyl Transferse(< 55u/l)	Gamma Glutamyl Transferse(≥ 55u/l)	$\frac{446}{523} = 0,85277$	$\frac{77}{523} = 0,14722$	Hepatitis C	$\frac{39}{446} = 0,08744$	$\frac{28}{77} = 0,36363$
					Tidak Hepatitis C	$\frac{407}{446} = 0,91255$	$\frac{49}{77} = 0,63636$

Pada Tabel 4.6 di atas Nilai dari *gain information* yang digunakan untuk mendeteksi atribut-atribut yang paling banyak memiliki informasi berdasarkan kelas tertentu.

Untuk menghitung nilai *gain information* pada simpul lain digunakan cara yang sama, sehingga menghasilkan Tabel 4.7 berikut:

Tabel 4.7 Perhitungan *gain information*

Simpul	P_L	P_R	Kelas	$GI(P_L)$	$GI(P_R)$
1	Albumin=Rendah	Albumin=Normal,Tinggi	Hepatitis C	0,96816	0,48205
			Tidak Hepatitis C		
2	Albumin=Normal	Albumin=Rendah,Tinggi	Hepatitis C	0,48205	0,79794
			Tidak Hepatitis C		
3	Albumin=Tinggi	Albumin=Rendah,Normal	Hepatitis C	0,48205	0,55887
			Tidak Hepatitis C		
4	Alanine Transaminase(< 47u/l)	Alanine Transaminase(≥ 47u/l)	Hepatitis C	0,50536	0,86312
			Tidak Hepatitis C		
5	Aspartate Transaminase	Aspartate Transaminase	Hepatitis C	0,16169	0,98802

	nase(< 37u/l)	nase(≥ 37u/l)	Tidak Hepatitis C		
6	Gamma Glutamyl Transfers e(< 55u/l)	Gamma Glutamyl Transfers e(> 55u/l)	Hepatitis C	0,42787	0,94566
			Tidak Hepatitis C		

Pada Tabel 4.7 di atas didapatkan bahwa atribut yang menyediakan informasi paling banyak memiliki nilai *gain information* tertinggi adalah atribut *Aspartate Transaminase* ($\geq 37u/l$) yaitu sebesar 0,98802 dan aribut lainnya pun memiliki nilai yang tidak jauh berbeda.

3) Indeks Gini

Tabel 4.8 Rincian Indeks Gini

Simpul ₁	P_L	P_R	$i(t)$
1	Albumin=rendah	Albumin=normal, tinggi	0,07546
2	Albumin=normal	Albumin=rendah, tinggi	0,28745
3	Albumin=tinggi	Albumin=rendah, normal	0,16671
4	Alanine Transaminase(<47u/l)	Alanine Transaminase(<47 u/l)	0,14132
5	Aspartate Transaminase(<37u/l)	Aspartate Transaminase(<37 u/l)	0,31165
6	Gamma Glutamyl Transferse(<55u/l)	Gamma Glutamyl Transferse(<55u/l)	0,25110

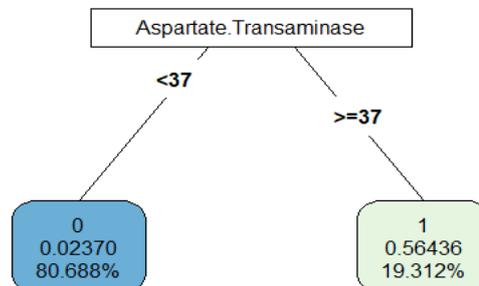
4) Goodness of split

Tabel 4.9 Goodness of split

Simpul	$\emptyset(s, t)$	Kriteria Kebaikan
1	-0,13512	6
2	0,06952	2
3	-0,05655	5
4	0,01817	4
5	0,17937	1
6	0,04687	3

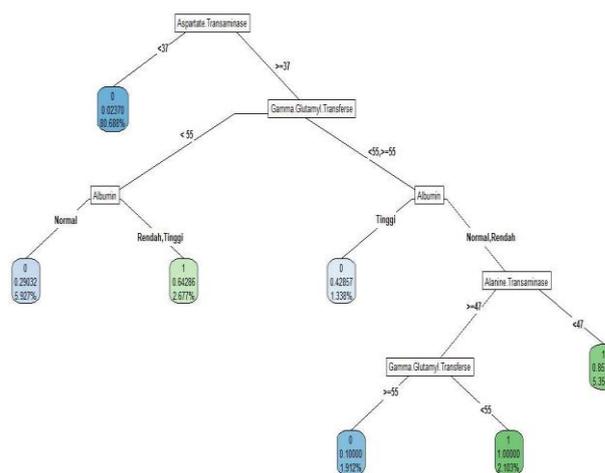
Berdasarkan pada tabel di atas yang cocok dengan nilai tertinggi pada calon simpul adalah kandidat simpul ke-5 yaitu *Aspartate Transaminase* yang berkisar 0,17937 dimana kandidat simpul ke-5 akan menjadi *parent*

node yaitu variabel *Aspartate Transaminase*. Simpul ke-5 ini akan terbagi menjadi cabang kiri adalah atribut *Aspartate Transaminase* ($< 37u/l$) dan cabang kanan adalah atribut *Aspartate Transaminase* ($\geq 37u/l$).



Gambar 4.1 Rincian Indeks Gini

Berdasarkan gambar di atas bahwa *parent node Aspartate Transaminase* ($< 37u/l$) yang berada pada node ke-2 dan simpul akar atau *parent node Aspartate Transaminase* ($\geq 37u/l$) yang berada pada node ke-3. Jumlah pasien Hepatitis C yang mempunyai *Aspartate Transaminase* ($< 37u/l$) sebesar 80,688%, sedangkan untuk jumlah pasien Hepatitis C yang memiliki *Aspartate Transaminase* ($\geq 37u/l$) sebesar 19,312% sama dengan yang ditunjukkan pada Tabel 4.9. Sehingga terbentuk pohon keputusan maksimal pada gambar dibawah ini



Gambar 4.2 Bentuk Pohon Keputusan Maksimal

Berdasarkan gambar 4.2 bahwa simpul akar atau simpul 1 yaitu *Aspartate Transaminase* dimana terdiri atas *Aspartate Transaminase* ($< 37u/l$) di simpul ke-2 dan *Aspartate Transaminase* ($\geq 37u/l$) simpul ke-3.

Setelah terjadi proses pembagian maka simpul ke-2 menjadi simpul terminal. Begitupula simpul ke-3 bercabang menjadi simpul ke-6 yaitu *Gamma Glutamyl Transfere* ($< 55u/l$) dan simpul ke-7 *Gamma Glutamyl Transfere* ($< 55u/l$ dan $\geq 55u/l$). Kemudian simpul ke-6 bercabang menjadi simpul ke-12 Albumin (Normal) dan simpul ke-13 Albumin (Rendah, Tinggi) yang keduanya menjadi simpul terminal. Pada simpul ke-7 tersebar menjadi simpul ke-14 Albumin (Tinggi) yang menjadi simpul terminal dan simpul ke-15 Albumin (Normal, Rendah). Kemudian simpul ke-15 menyebar menjadi simpul ke-30 *Alanine Transaminase* ($\geq 47u/l$) dan simpul ke-31 *Alanine Transaminase* ($< 47u/l$) menjadi simpul terminal. Kemudian simpul ke-30 menyebar menjadi simpul ke-60 *Gamma Glutamyl Transfere* ($\geq 55u/l$) dan simpul ke-61 *Gamma Glutamyl Transfere* ($< 55u/l$) yang keduanya menjadi simpul terminal. Jadi, Gambar 4.2 dapat diuraikan dalam tabel berikut:

Tabel 4.10 Simpul dalam Pohon Keputusan

Simpul	Nama	Keterangan
1	<i>Aspartate Transaminase</i>	Non simpul terminal (root node)
2	<i>Aspartate Transaminase</i> ($< 37u/l$)	Simpul terminal
3	<i>Aspartate Transaminase</i> ($\geq 37u/l$)	Non simpul terminal
4	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$)	Non simpul terminal
5	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$ dan $\geq 55u/l$)	Non Simpul terminal
6	Albumin(Normal)	Simpul terminal
7	Albumin(Rendah,Tinggi)	Simpul terminal
8	Albumin(Tinggi)	Simpul terminal
9	Albumin(Normal,Rendah)	Non simpul terminal
10	<i>Alanine Transaminase</i> ($\geq 47u/l$)	Non simpul terminal
11	<i>Alanine Transaminase</i> (< 47)	Simpul terminal
12	<i>Gamma Glutamyl Transfere</i> ($\geq 55u/l$)	Simpul Terminal
13	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$)	Simpul Terminal

Berdasarkan Tabel 4.10 bahwa *terminal node* merupakan simpul terakhir yang tidak menghadapi penurunan keheterogenan berdasarkan kriteria *goodness of split*. Simpul terminal dari pohon keputusan yang pada pada Gambar 4.2 yaitu simpul 2,6,7,8,11,12,13 dan non simpul terminal yaitu simpul 1,3,4,5,9,10.

5) Pelabelan Kelas

Tabel 4.11 Pelabelan Kelas

Simpul	Nama	Label Kelas
1	<i>Aspartate Transaminase</i>	1
2	<i>Aspartate Transaminase</i> ($< 37u/l$)	0
3	<i>Aspartate Transaminase</i> ($\geq 37u/l$)	1
4	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$)	0
5	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$ dan $\geq 55u/l$)	1
6	Albumin(Normal)	0
7	Albumin(Rendah,Tinggi)	1
8	Albumin(Tinggi)	0
9	Albumin (Normal,Rendah)	1
10	<i>Alanine Transaminase</i> ($\geq 47u/l$)	0
11	<i>Alanine Transaminase</i> ($< 47u/l$)	1
12	<i>Gamma Glutamyl Transfere</i> ($\geq 55u/l$)	0
13	<i>Gamma Glutamyl Transfere</i> ($< 55u/l$)	1

Dari Tabel 4.11, jelas bahwa label kelas pasien Hepatitis C yang mengidap dengan urutan simpul 1,3,5,7,9,11,13 dan untuk simpul pasien Hepatitis C yang tidak mengidap dengan urutan simpul 2,4,6,8,10,12.

6) Pemangkasan pohon klasifikasi berdasarkan *Standar Cost Complexity Minimum*

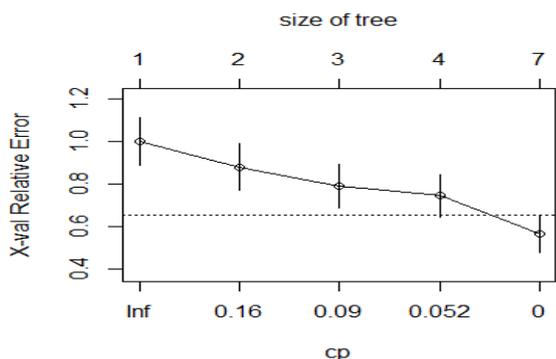
Tabel 4.12 Nilai *Complexity parameter* pohon T

No	Complexity parameter
1	0.194030
2	0.134328
3	0.059701
4	0.044776

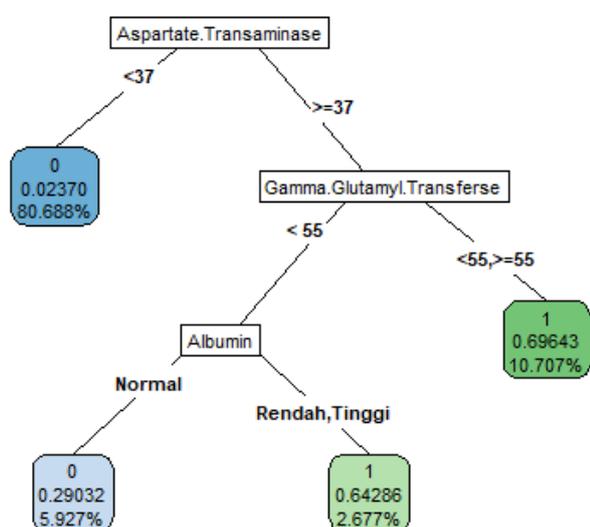
Berdasarkan tabel 4.12 dapat dilihat pada lampiran 4, nilai *complexity parameter* untuk setiap *split* secara berurutan adalah 0,194030, 0,134328, 0,059701, 0,044776 sehingga didapatkan nilai *complexity parameter minimum* berkisar 0,044776. Berikut plot hasil dari nilai *complexity parameter*.

Ditunjukkan pada gambar 4.3, plot *complexity parameter* memberikan nilai yang berhak dipotong yaitu 0,044776 dimana nilainya kurang dari 0,059701 atau nilainya yang paling kecil. sehingga diperoleh struktur pohon

keputusan optimum yang telah dipotong seperti pada gambar 4.4



Gambar 4.3 Plot Complexity Parameter



Gambar 4.4 Bentuk Pohon Keputusan Optimum

Dari gambar 4.4, diperlihatkan parent node 1 yaitu *Aspartate Transaminase* yang berada pada simpul ke-2 untuk *Aspartate Transaminase* ($< 37u/l$) yang menjadi simpul terminal dan simpul ke-3 untuk *Aspartate Transaminase* ($\geq 37u/l$). Setelah terjadi pembagian maka simpul ke-3 bercabang menjadi simpul ke-6 yaitu *Gamma Glutamyl Transferse* ($< 55u/l$) dan simpul ke-7 yaitu *Gamma Glutamyl Transferse* ($< 55u/l$ dan $\geq 55u/l$) yang menjadi simpul terminal. Kemudian pada simpul ke-6 bercabang menjadi simpul ke-12 yaitu Albumin (Normal) dan simpul ke-13 yaitu Albumin (Rendah, Tinggi) yang masing-masing keduanya menjadi simpul terminal.

7) Penafsiran setiap simpul atau node sesuai Pohon Keputusan Optimum

Berdasarkan gambar 4.4 dapat dijelaskan setiap simpul dari pohon tersebut antara lain:

- a. Pada cabang *Aspartate Transaminase* ($\geq 37u/l$) yang memiliki *Gamma Glutamyl Transferse* ($< 55u/l$) dan ($< 55u/l, \geq 55u/l$) pada Gambar 4.4 dapat dideskripsikann pada Tabel 4.13 berikut:

Tabel 4.13 Pasien Hepatitis C berdasarkan cabang *Aspartate Transaminase* dan *Gamma Glutamyl Transferse*

		<i>Gamma Glutamyl Transferse</i>	
		< 55	$< 55, \geq 55$
Observasi	N	45	56
	%	8,60%	10,70%
Hepatitis C	N	27	17
	Kejadian 1 (Yes)	0,6	0,3035
	%	60%	30,35%
Tidak Hepatitis C	N	18	39
	Kejadian 0 (No)	0,4	0,69642
	%	40%	69,64%

Pada Tabel 4.13 terlihat bahwa pasien Hepatitis C yang mempunyai *Aspartate Transaminase* ($\geq 37u/l$) dengan *Gamma Glutamyl Transferse* ($< 55u/l$) berkisar 45 orang (8,60%, $n=523$). Dimana dari jumlah tersebut pasien Hepatitis C yang mengidap adalah sebanyak 27 orang (60%, $n=45$) sedangkan untuk pasien Hepatitis C yang tidak mengidap adalah sekitar 18 orang (40%, $n=45$). Pada pasien Hepatitis C yang memiliki *Aspartate Transaminase* ($\geq 37u/l$) dengan *Gamma Glutamyl Transferse* ($< 55u/l, \geq 55u/l$) sebanyak 56 orang (10,70%, $n=523$). Dimana dari jumlah tersebut pasien Hepatitis C yang mengidap adalah sebanyak 17 orang (30,35%, $n=56$) sedangkan untuk pasien Hepatitis C yang tidak mengidap adalah sebanyak 39 orang (69,64%, $n=56$).

- b. Pada cabang *Aspartate Transaminase* ($\geq 37u/l$) dengan *Gamma Glutamyl Transferse* ($< 55u/l$) yang memiliki Albumin (Normal) dan Albumin (Rendah, Tinggi) pada Gambar 4.4 dapat dideskripsikan di Tabel 4.14 berikut:

Tabel 4.14 Pasien Hepatitis C berdasarkan cabang *Aspartate Transaminase*, *Gamma Glutamyl Transferse* dan Albumin

		Albumin	
		Normal	Rendah,Tinggi
Observasi	N	31	14
	%	5,92%	2,67%
Hepatitis C	N	22	5
	Kejadian 1 (Yes)	0,7096	0,3571
	%	70,96%	35,71%
Tidak Hepatitis C	N	9	9
	Kejadian 0 (No)	0,2903	0,64285
	%	29,03%	64,28%

Pada pasien Hepatitis C yang memiliki *Aspartate Transaminase* ($\geq 37u/l$), *Gamma Glutamyl Transferse* ($< 55u/l$) dengan Albumin (Rendah,Tinggi) sebanyak 14 orang (2,67%, n=523). Dimana dari jumlah tersebut pasien Hepatitis C yang mengidap sebanyak 5 orang (35,71%, n=14) sedangkan pasien Hepatitis C yang tidak mengidap sebanyak 9 orang (64.28%, n=14).

c. Pada cabang *Aspartate Transaminase* ($< 37u/l$) yang ditunjukkan pada Gambar 4.4 adalah dapat ditampilkan di Tabel 4.15 berikut ini:

Tabel 4.15 Pasien Hepatitis C berdasarkan cabang *Aspartate Transaminase*

		<i>Aspartate Transaminase</i>	
		(< 37u/l)	
Observasi	N	422	
	%	80,68%	
Hepatitis C	N	412	
	Kejadian 1 (Yes)	0,9763	
	%	97,63%	
Tidak Hepatitis C	N	10	
	Kejadian 0 (No)	0,0236	
	%	2,36%	

Dari Tabel 4.15 dapat dilihat bahwa Pasien Hepatitis C yang memiliki *Aspartate Transaminase* ($< 37u/l$) sebanyak 422 orang (80,68%, n=523). Dimana pasien Hepatitis C yang mengidap sebanyak 412 orang (97,63%, n=422) sedangkan pasien Hepatitis C yang tidak mengidap sebanyak 10 orang (2,36%, n=422).

8) Menghitung sensitivity, specificity, serta akurasi dalam kebenaran klasifikasi pohon keputusan.

Tabel 4.16 mendeskripsikan bahwa pasien Hepatitis C yang mengidap diperkirakan

sebanyak 7 orang dan pasien Hepatitis C yang tidak mengidap diperkirakan sebanyak 80

Tabel 4.16 Ketepatan Klasifikasi Pada Data Testing

Observasi	Prediksi		Total
	Pasien Hepatitis C yang tidak mengidap	Pasien Hepatitis C yang mengidap	
Pasien Hepatitis C yang tidak mengidap	80	3	83
Pasien Hepatitis C yang mengidap	2	7	9
Total	82	10	92

Orang. Pada perhitungan di atas, nilai akurasi dalam mengukur tingkat ketepatan klasifikasi terhadap data testing sebanyak 94,56%. Jadi pohon keputusan maksimum yang terbentuk bisa mengklasifikasikan data yang baru sebanyak 94,56%. Nilai *Sensitivity* pada pasien Hepatitis C yang tidak mengidap dalam mengukur ketepatan klasifikasi sebanyak 96,38% dan untuk nilai *Specificity* pada pasien Hepatitis C yang mengidap dalam mengukur ketepatan klasifikasi sebanyak 77,77%.

9) Mengetahui faktor-faktor penyebab hepatitis C dan menganalisis jumlah pasien hepatitis c yang mengidap sesuai faktor yang berpengaruh pada pohon keputusan optimum.

Dari pohon keputusan pada Gambar 4.4 diketahui faktor-faktor penyebab hepatitis c dan jumlah pasien hepatitis c dengan kelompok-kelompok sebagai berikut:

a. Kelompok pertama, pasien Hepatitis C yang mempunyai *Aspartate Transaminase* ($< 37u/l$) sebanyak 422 orang (80,68%, n=523). Dimana pasien

Hepatitis C yang mengidap sebanyak 412 (97,63%, n=422).

- b. Kelompok kedua, pasien Hepatitis C yang mempunyai Aspartate Transaminase ($\geq 37u/l$) dengan Gamma Glutamyl Transferase ($< 55u/l, \geq 55u/l$) sebanyak 56 orang (10,70%, n=523). Pasien Hepatitis C yang mengidap sebanyak 17 orang (30,35%, n=56).
- c. Kelompok ketiga, pasien Hepatitis C yang mempunyai Aspartate Transaminase ($\geq 37u/l$), Gamma Glutamyl Transferase ($< 55u/l$) dengan Albumin (Normal) sebanyak 31 orang (5,92%, n=523). Pasien Hepatitis C yang mengidap sebanyak 22 orang (70,96%, n=31).
- d. Kelompok keempat, pasien Hepatitis C yang mempunyai Aspartate Transaminase ($\geq 37u/l$), Gamma Glutamyl Transferase ($< 55u/l$) dengan Albumin (Rendah, Tinggi) sebanyak 14 orang (2,67%, n=523). Dimana pasien Hepatitis C yang mengidap sebanyak 5 orang (35,71%, n=14).

5. KESIMPULAN

Dalam penelitian ini disimpulkan bahwa faktor-faktor penyebab hepatitis c dan jumlah pasien Hepatitis C dalam pengklasifikasian sesuai dengan faktor-faktor yang berpengaruh pada pohon keputusan dengan metode CART dikelompokkan berdasarkan kelompoknya yaitu kelompok pertama pasien Hepatitis C yang mempunyai Aspartate Transaminase ($< 37u/l$) sebanyak 412 orang yang mengidap hepatitis c. Kelompok kedua yaitu pasien Hepatitis C yang mempunyai Aspartate Transaminase ($\geq 37u/l$) dengan Gamma Glutamyl Transferase ($< 55u/l, \geq 55u/l$) sebanyak 17 orang yang mengidap hepatitis c. Kelompok ketiga adalah pasien Hepatitis C yang memiliki Aspartate Transaminase ($\geq 37u/l$), Gamma Glutamyl Transferase ($< 55u/l$) dengan Albumin (Normal) sebanyak 22 orang yang mengidap hepatitis c. Kelompok keempat yaitu pasien Hepatitis C yang mempunyai Aspartate Transaminase ($\geq 37u/l$), Gamma Glutamyl Transferase ($< 55u/l$) dengan Albumin (Rendah, Tinggi) sebanyak 5 orang yang mengidap hepatitis C.

6. DAFTAR PUSTAKA

- [1] Kusriani dan Lutfi Emha Taufiq. *Algoritma Data Mining*. Yogyakarta: CV. ANDI OFFSET, 2009.
- [2] Nurcahyo, Gunandi Widi dan Dicky Nofriansyah. *Algoritma Data Mining dan Pengujian*. Deepublish: CV Budi Utama, 2019.
- [3] Wathoni, Mahbubul. *Prediksi Kecenderungan Konsumen dalam Memilih Jenis Kendaraan (Roda Empat) Berdasarkan Spesifikasi Kendaraan Menggunakan Decision Tree dengan Metode Gini*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah Program Studi Matematika Jurusan MIPA Fakultas Sains dan Teknologi, 2006.
- [4] Eska Juna. *Penerapan Data Mining untuk Prediksi Penjualan Walpaper Menggunakan Algoritma C4.5*. Jurnal Teknologi dan Sistem Informasi. Vol.2, No.2, 2016
- [5] Nafi'iyah, Nur. *Algoritma CART dalam Penentuan Pohon Keputusan Sertifikasi Guru*. Jurnal Spirit, Vol.13, No.2, 2015.
- [6] Han, Jiawei. *Data Mining Concepts and Techniques Third Edition*. USA:Elsevier, 2012.
- [7] Handayani, Siti Hartati Selviani dan Santi Wulan Purnami. *Pendekatan Metode Clasification and Regression Tree untuk Diagnosis Tingkat Keganasan Kanker pada Pasien Kanker Tiroid*. Jurnal Sains dan Seni Pomits Vol.3, No.1, 2014.
- [8] Siahaan, Makhmul Ikhsan. *Pengunaan Metode Classification and Regression Tree (CART) untuk Klasifikasi Tingkat Partisipasi Angkatan Kerja di Kota Medan Tahun 2019*. Skripsi. Medan: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sumatera Utara, 2021.
- [9] Sumartini, Siti Holis dan Santi Wulan Purnami. *Penggunaan Metode Classification and Regression Trees (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya*. Jurnal Sains dan Seni ITS. Vol.4, No.2, 2015.
- [10] Budhiman, Arif. *Evaluasi Ketepatan Klasifikasi Penyakit ISPA (Infeksi Saluran*

Pernapasan Akut) Pada Anak Balita Menggunakan Metode Classification Tree di Puskesmas Cukir Kabupaten Jombang Jawa Timur. Tesis, Surabaya: Fakultas Matematika, Komputasi dan Sains Data Institut Teknologi Sepuluh Nopember, 2018.

- [11] Alfian, Lisa Purwanti. *Hepatitis Virus C (HCV) di RSUP Dr. Wahidin Sudirohusodo Makassar Periode 2017-2019. Skripsi*. Makassar: Fakultas Kedokteran Universitas Hasanuddin, 2020.
- [12] Zuama, Robi Aziz. *Pembelajaran Mesin Untuk Diagnosis Tingkat Kerusakan Hati Akibat Hepatitis C*. Journal Speed-Sentra Penelitian Engineering dan Edukasi. Vol.13, No.3, 2021