

Penggunaan Metode *Classification and Regression Trees* (CART) untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya

Siti Holis Sumartini dan Santi Wulan Purnami

Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: santi_wp@statistika.its.ac.id

Abstrak — Kanker serviks merupakan kanker yang menyerang area bawah rahim. Pengobatan kanker serviks tergantung pada besarnya ukuran dan stadium kanker. Kasus rekurensi sering terjadi walaupun pengobatan awal telah dilakukan. Salah satu rumah sakit yang menyediakan fasilitas terapi untuk pasien kanker serviks adalah RSUD dr. Soetomo Surabaya. Permasalahannya adalah bagaimana klasifikasi pasien kanker serviks di RSUD dr. Soetomo Surabaya yang rekuren dan tidak rekuren berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dan mengetahui faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks. Rekurensi yang dimaksudkan di penelitian ini adalah kembalinya pasien kanker serviks ke RSUD dr. Soetomo karena penyakit yang sama. Data yang digunakan merupakan data sekunder, yang diperoleh dari rekam medis pasien kanker serviks di RSUD dr. Soetomo Surabaya pada tahun 2014 dengan jumlah data sebanyak 810 pasien. Berdasarkan analisis yang telah dilakukan, pasien yang rekuren lebih banyak dibandingkan pasien yang tidak rekuren dengan persentase sebesar 57,78 persen untuk yang rekuren. Klasifikasi CART menghasilkan bahwa variabel yang paling berpengaruh terhadap rekurensi kanker serviks adalah variabel jenis pengobatan yang dijalani oleh pasien, selain itu variabel usia, status anemia dan status penyakit penyerta juga berpengaruh terhadap rekurensi kanker serviks. Ketepatan klasifikasi yang diperoleh untuk data prediksi sebesar 69,14 persen.

Kata Kunci —Klasifikasi, CART, kanker serviks, rekurensi

I. PENDAHULUAN

KANKER serviks merupakan salah satu penyebab utama kematian dibanyak negara terutama negara berkembang [1]. Kanker serviks merupakan kanker yang menyerang area bawah rahim [2]. Penyebab utama kanker serviks adalah HPV (*Human Papilloma Virus*). Virus ini ditemukan pada 77% perempuan yang terjangkit HIV positif dan sisa nya kepada wanita siapa saja [3]. Penyebab banyak kematian pada kaum wanita adalah HPV tipe 16 dan 18[4].

Pengobatan kanker serviks tergantung pada besarnya ukuran kanker dan stadium kanker [5]. Namun, pengobatan pada stadium awal tidak menjamin dapat menyembuhkan keanker serviks. Kasus rekurensi sering terjadi setelah pengobatan awal dilakukan. Rekurensi yang dimaksudkan dalam penelitian ini adalah kembalinya pasien ke RSUD dr. Soetomo Surabaya setelah melakukan perawatan sebelumnya.

RSUD (Rumah Sakit Umum Daerah) dr. Soetomo Surabaya merupakan rumah sakit yang menjadi pusat rujukan terutama untuk kawasan Indonesia Timur. Selain itu, rumah sakit ini merupakan salah satu rumah sakit yang

menyediakan layanan kesehatan yang menyediakan fasilitas terapi untuk pasien kanker serviks [6]. Poli Onkologi Satu Atap RSUD dr. Soetomo Surabaya tahun 2013 menunjukkan bahwa kanker serviks menempati urutan kedua terbanyak di RSUD dr. Soetomo Surabaya dan mempunyai kecenderungan meningkat dari tahun ke tahun[7].

Terdapat beberapa faktor resiko yang diduga berpengaruh terhadap rekurensi kanker serviks, yaitu usia, status anemia, stadium, jenis pengobatan yang dijalani, status komplikasi dan status penyakit penyerta. Permasalahannya adalah bagaimana klasifikasi pasien kanker serviks di RSUD dr. Soetomo Surabaya yang rekuren dan tidak rekuren berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks dan mengetahui faktor-faktor yang berpengaruh terhadap rekurensi kanker serviks. Terdapat berbagai macam metode klasifikasi, yaitu CART (*Classification And Regression Trees*), CHAID (*Chi-Square Automatic Interaction Detection*), SVM dan metode klasifikasi lainnya. Dalam penelitian ini, metode yang digunakan adalah klasifikasi CART karena dapat mengklasifikasikan data baru secara efisien.

Metode CART (*Classification and Regression Trees*) merupakan suatu metode statistika nonparametrik yang dapat menggambarkan hubungan antara variabel respon (dependen) dengan satu atau lebih variabel prediktor (variabel independen). Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian[8].

Penelitian menggunakan analisis CART yang dilakukan oleh Nuriyah (2013) menghasilkan bahwa hasil ketepatan klasifikasi CART lebih baik dibandingkan metode CHAID [9]. Sedangkan penelitian mengenai kanker serviks yang dilakukan oleh Novitasari (2014) menghasilkan bahwa jenis kanker dan stadium kanker serviks berpengaruh terhadap rekurensi kanker serviks [10].

Data yang diambil merupakan data kedatangan kedua pasien ke RSUD dr. Soetomo Surabaya. Selain itu, data yang digunakan merupakan data kanker serviks tahun 2014 yang melakukan rawat inap.

II. TINJAUAN PUSTAKA

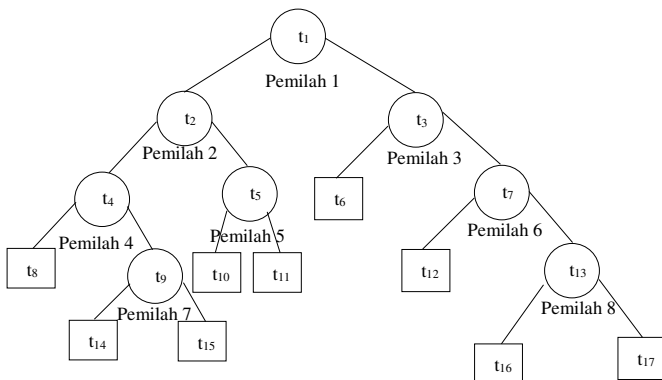
A. *Classification And Regression Trees* (CART)

Classification and Regression Trees (CART) merupakan salah satu metode atau algoritma dari teknik pohon keputusan (*decision tree*). Metode yang dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard A. Olshen, dan Charles J. Stone ini merupakan teknik klasifikasi dengan

menggunakan algoritma penyekatan rekursif secara biner (*binary recursive partitioning*) [11].

Ilustrasi dari struktur pohon klasifikasi ditunjukkan pada Gambar 1. Simpul awal disebut *parent node* dinotasikan t_1 , simpul dalam dinotasikan dengan t_2, t_3, t_4, t_7, t_9 dan t_{10} , serta simpul akhir (*terminal nodes*) dinotasikan dengan $t_5, t_6, t_8, t_{11}, t_{12}, t_{13}, t_{14}$, dan t_{15} dimana setelahnya tidak ada lagi pemilahan.

CART akan menghasilkan pohon klasifikasi jika variabel respon mempunyai skala kategorik dan akan menghasilkan pohon regresi jika variabel respon berupa data kontinu [6]. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian.



Sumber: Breiman L., Friedman J.H., Olshen R.A., & Stone C.J. (1993). *Classification and Regression Trees (CART)*
Gambar 1. Struktur Pohon Klasifikasi

Algoritma CART melalui tiga tahapan, yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi dan penentuan pohon klasifikasi optimum.

1. Pembentukan pohon klasifikasi

Tahap inidiawali dengan menentukan variabel dan *threshold* untuk dijadikan pemilah tiap simpul. Tahapan pembentukan pohon klasifikasi terdiri dari:

a. Pemilahan Pemilah

Data yang digunakan merupakan sampel data *learning*. Himpunan bagian yang dihasilkan dari proses pemilahan harus lebih homogen dibandingkan pemilahan sebelumnya. Fungsi keheterogenan yang digunakan adalah Indeks Gini karena akan selalu memisahkan kelas dengan anggota paling besar/kelas terpenting dalam simpul terlebih dahulu. Fungsi Indeks Gini ditunjukkan pada persamaan (1).

$$i(t) = \sum_{i,j=1} p(j|t)p(i|t), i \neq j \tag{1}$$

Dengan $p(j|t)$ adalah proporsi kelas j pada simpul t dan $p(i|t)$ adalah proporsi kelas i pada simpul t .

Pemilahan yang terpilih akan membentuk suatu himpunan kelas yang disebut simpul. Simpul tersebut akan melakukan pemilahan secara rekursif sampai diperoleh *terminal nodes*. Tahapan berikutnya adalah menentukan kriteria *goodness of split* untuk mengevaluasi pemilah dari pemilah s pada simpul t dengan rumus:

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \tag{2}$$

Pemilah yang menghasilkan $\phi(s, t)$ lebih tinggi merupakan pemilah terbaik karena mampu mereduksi heterogenitas lebih tinggi.

b. Penentuan Simpul Terminal

Pengembangan pohon akan berhenti apabila pada simpul terdapat pengamatan berjumlah kurang dari atau sama dengan 5 ($n \leq 5$) [5]. Selain itu, proses pembentukan pohon juga akan berhenti apabila sudah mencapai batasan jumlah level yang telah ditentukan atau tingkat kedalaman (*depth*) dalam pohon maksimal.

c. Penandaan Label Kelas

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

$$p(j_0 | t) = \max_j \frac{N_j(t)}{N(t)} \tag{3}$$

Label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t yang paling kecil sebesar $r(t) = 1 - \max_j p(j|t)$.

2. Pemangkasan pohon klasifikasi

Pohon yang dibentuk dengan aturan pemilah dan kriteria *goodness of split* berukuran sangat besar karena penghentian pohon berdasarkan banyaknya amatan pada simpul terminal atau besarnya tingkat kehomogenan. Ukuran pohon yang besar dapat dapat memunculkan adanya *overfitting*, akan tetapi apabila pengamatan pohon dibatasi dengan ketepatan batas tertentu, maka dapat terjadi kasus *underfitting*. ukuran pohon yang layak dapat dilakukan dengan pemangkasan pohon dengan ukuran *cost complexity minimum* [5].

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \tag{4}$$

$R_\alpha(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan *cost penalty* bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. Selanjutnya, dilakukan pencarian pohon bagian $T(\alpha) < T_{max}$ yang meminimumkan $R_\alpha(t)$ yaitu

$$R_\alpha(T(\alpha)) = \min_{T < T_{max}} R_\alpha(T) \tag{5}$$

3. Penentuan pohon klasifikasi optimum

Penduga pengganti yang sering digunakan apabila pengamatan yang ada tidak cukup besar adalah *Cross Validation Estimate*. Pengamatan dalam L dibagi secara random menjadi V bagian yang saling lepas dengan ukuran kurang lebih sama besar untuk setiap kelas. Pohon $T^{(v)}$ dibentuk dari sampel *learning* ke- v dengan $v=1,2,\dots,V$. dimisalkan $d^{(v)}(x)$ adalah hasil pengklasifikasian, maka penduga sampel uji untuk $R(T_i^{(v)})$ adalah sebagai berikut.

$$R(T_i^{(v)}) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_v} X(d^{(v)}(x_n) \neq j_n) \tag{6}$$

Dengan $N_v \cong N/V$ adalah jumlah pengamatan dalam L_v .

B. Ukuran Ketepatan Klasifikasi

Ketepatan klasifikasi digunakan untuk mengetahui apakah data diklasifikasikan dengan benar atau tidak [12]. Beberapa cara yang umum digunakan untuk mengukur ketepatan klasifikasi adalah melalui perhitungan *APER*, *sensitivity*, *specivicity*, dan akurasi. Tabel untuk menghitung ketepatan klasifikasi ditunjukkan pada Tabel 1.

Tabel 1. Ketepatan Klasifikasi

Observasi	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

$$APER = \frac{n_{12} + n_{21}}{N} \tag{7}$$

$$Sensitivity = \frac{n_{11}}{N_1} \tag{8}$$

$$Specificity = \frac{n_{22}}{N_2} \tag{9}$$

$$Akurasi(1 - APER) = \frac{n_{11} + n_{22}}{N} \tag{10}$$

Cross Validation merupakan salah satu teknik untuk menilai/menvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu. Data yang digunakan dalam proses pembangunan model disebut data *training*, sedangkan data yang akan digunakan untuk menvalidasi model disebut data *testing* [13].

C. Gambaran Mengenai Kanker Serviks

Kanker serviks menyerang daerah leher rahim yang disebabkan infeksi virus HPV yang tidak sembuh dalam waktu lama[4]. Penyebab banyak kematian pada kaum wanita adalah virus HPV 16 dan 18, virus ini sangat mudah berpindah dan menyebar, tidak hanya melalui cairan, tetapi juga berpindah melalui sentuhan kulit [14].

Kasus rekurensi kanker serviks sering terjadi setelah dilakukan pengobatan kanker serviks. Rekurensi yang dimaksudkan dalam penelitian ini adalah kembalinya pasien ke RSUD dr. Soetomo Surabaya setelah melakukan pengobatan sebelumnya. Faktor-faktor penyebab rekurensi kanker serviks diantaranya anemia[15], usia[16][17], komplikasi[18], jenis dan stadium kanker[19][20] dan jenis pengobatan[17].

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Variabel dependen yang digunakan adalah status rekurensi kanker serviks. Rekurensi yang dimaksudkan adalah kembalinya pasien ke RSUD dr. Soetomo karena penyakit yang sama, terdiri dari.

Y = 1 jika tidak rekuren

Y = 2 jika rekuren

Sedangkan variabel independen yang digunakan adalah

Tabel 2. Variabel Independen

Variabel	Nama Variabel	Kategori	Skala
X ₁	Usia Pasien	-	Rasio
X ₂	Status Anemia	1 : Tidak ada 2 : Ada	Nominal
X ₃	Stadium kanker serviks	0 : Stadium 0 1 : Stadium I 2 : Stadium II 3 : Stadium III 4 : Stadium IV	Ordinal
X ₄	Jenis pengobatan yang dijalani pasien	1 : Transfusi PRC 2 : Kemoterapi 3 : Kemoterapi & Tranfusi PRC 4 : Operasi	Nominal
X ₅	Status Komplikasi	1 : Tidak ada 2 : Ada	Nominal
X ₆	Status Penyakit Penyerta	1 : Bukan Penyakit Penyerta 2 : Penyakit Penyerta	Nominal

B. Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini adalah.

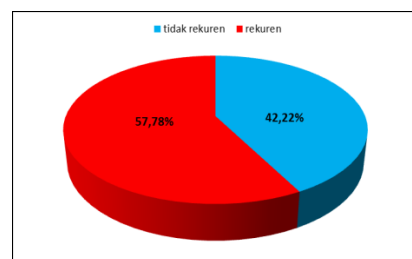
1. Mendeskripsikan pasien kanker serviks dengan melakukan analisis statistika deskriptif.
2. Membagi data menjadi dua bagian dengan metode *5-fold cross validation*. Selanjutnya membandingkan hasil ketepatan klasifikasi dari setiap *fold* untuk data *learning* dan *testing*. Kombinasi data yang dipilih adalah kombinasi data yang memiliki tingkat akurasi tertinggi dan jumlah *terminal nodes* yang tidak terlalu banyak.
3. Melakukan metode klasifikasi CART melalui tahapan:
 - a. Membentuk pohon klasifikasi maksimal menggunakan data *learning* dengan langkah-langkah berikut.
 1. Melakukan pemilihan pemilah dari variabel prediktor berdasarkan aturan pemilahan indeks gini, kemudian hasil pemilahan dievaluasi dengan menggunakan kriteria *gooness of split*.
 2. Menentukan *terminal nodes*.
 3. Melakukan penandaan label kelas pada *terminal nodes* berdasarkan aturan jumlah terbanyak dari tiap kelas yang ada pada variabel respon.
 - b. Memangkas pohon (*pruning*) berdasarkan kriteria *cost complexity minimum*.
 - c. Memilih pohon klasifikasi yang optimal dengan *test sample estimates*.
 - d. Menghitung ketepatan klasifikasi pohon dengan Akurasi, *sensitivity* dan *specificity*.

IV. ANALISIS DAN PEMBAHASAN

A. Deskripsi Karakteristik Pasien

Gambaran karakteristik pasien kanker serviks di RSUD dr. Soetomo Surabaya akan diberikan berdasarkan keenam faktor yang berpengaruh terhadap rekurensi kanker serviks. Rekurensi yang dimaksudkan merupakan kembalinya pasien ke RSUD dr. Soetomo setelah melakukan perawatan sebelumnya.

Karakteristik mengenai status rekurensi kanker serviks di RSUD dr. Soetomo Surabaya disajikan pada Gambar 2.



Gambar 2. Deskripsi Status Rekurensi Kanker Serviks

Gambar 2 menjelaskan bahwa dari 810 pasien, 57,78 persen atau sebanyak 468 pasien rekuren kanker serviks, sedangkan sisanya sebanyak 342 pasien tidak rekuren kanker serviks.

Karakteristik pasien berdasarkan usia disajikan pada Tabel 3.

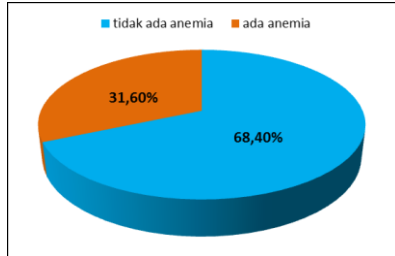
Tabel 3. Karakteristik Pasien Berdasarkan Usia (Tahun)

Status Rekurensi	Mean	Varians	Min	Median	Max
Tidak Rekuren	50,620	86,008	27	50	79
Rekuren	49,346	63,777	27	49	72

Tabel 3 menunjukkan bahwa rata-rata usia pasien yang tidak rekuren dan pasien rekuren tidak jauh berbeda, yaitu

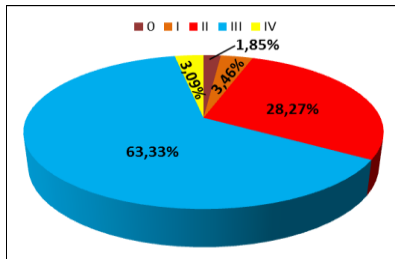
50,62 untuk pasien yang tidak rekuren dan 49,346 pada pasien rekuren, begitu juga dengan median dan usia minimumnya. Penjelasan lebih lengkap mengenai karakteristik usia pasien dapat dilihat di Tabel 3.

Selanjutnya, deskripsi faktor status anemia, stadium kanker serviks, jenis terapi yang dijalani pasien, status komplikasi dan status penyakit penyerta disajikan dengan *pie chart* berikut.



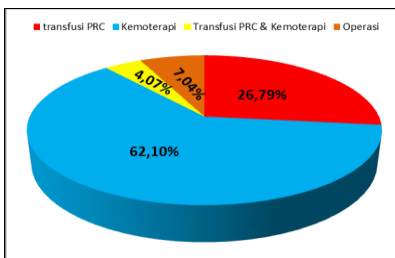
Gambar 3. Deskripsi Status Anemia

Gambar 3 menunjukkan bahwa mayoritas pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 tidak mengalami anemia, ditunjukkan oleh persentase sebesar 68,40 persen atau sebanyak 554 pasien.



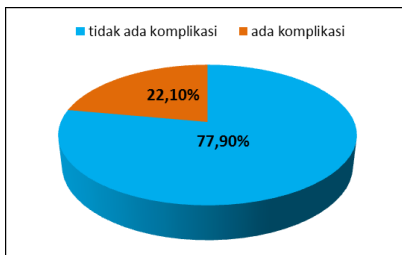
Gambar 4. Deskripsi Stadium Kanker Serviks

Berdasarkan faktor stadium pada Gambar 4, pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2014 cenderung menderita kanker serviks stadium III, ditunjukkan oleh persentase sebesar 63,33 persen atau sebanyak 513 pasien.



Gambar 5. Deskripsi Jenis Pengobatan yang Dijalani oleh Pasien

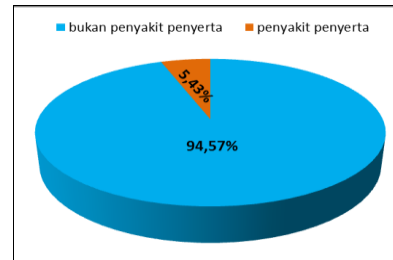
Gambar 5 menunjukkan bahwa mayoritas pasien menjalani pengobatan kemoterapi, yang ditunjukkan oleh persentase sebesar 62,10 persen atau sebanyak 503 pasien.



Gambar 6. Deskripsi Status Komplikasi

Gambar 6 menjelaskan bahwa mayoritas pasien tidak mengalami komplikasi, yang ditunjukkan oleh persentase sebesar 77,90 persen atau sebanyak 631 pasien.

Gambar 7 menjelaskan bahwa mayoritas penyakit kanker serviks yang diderita oleh pasien bukan merupakan penyakit penyerta, melainkan sebagai penyakit utama, yang ditunjukkan oleh persentase sebesar 94,57 persen atau sebanyak 766 pasien.

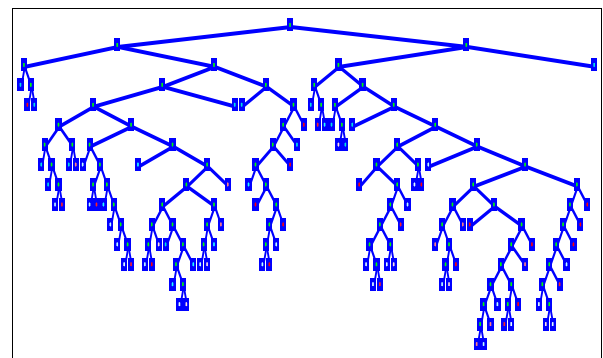


Gambar 7. Deskripsi Status Penyakit Penyerta

B. Analisis CART untuk Klasifikasi Rekurensi Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya.

Data pengamatan yang digunakan untuk mengklasifikasikan pasien kanker serviks sebesar 810 pengamatan. Data tersebut dibagi menjadi dua bagian yaitu data *learning* dan data *testing*. Pembagian data menjadi dua bagian ini bertujuan untuk mendapatkan suatu model yang optimal. Pembagian data *learning* dan *testing* ini menggunakan *cross validation*. Pada penelitian ini menggunakan *5-fold cross validation*.

Langkah pertama dalam analisis klasifikasi CART adalah pembentukan pohon klasifikasi maksimal dengan menentukan variabel pemilah dan nilai variabel (*threshold*) dari beberapa kemungkinan pemilah. Dari berbagai kemungkinan pemilah tersebut, kemudian menghitung Indeks Gini. Selanjutnya hasil perhitungan indeks gini digunakan untuk menentukan *goodness of split* dari masing-masing pemilah. Pemilah yang terpilih adalah variabel pemilah dan *threshold* yang memiliki nilai *goodness of split* tertinggi. Konstruksi pohon maksimal yang terbentuk disajikan pada Gambar 7.



Gambar 7. Konstruksi Pohon Maksimal

Gambar 7 menjelaskan bahwa hasil pohon klasifikasi maksimal terbentuk berukuran besar, yaitu dengan jumlah *terminal nodes* sebanyak 84 dan tingkat kedalaman sebesar 17.

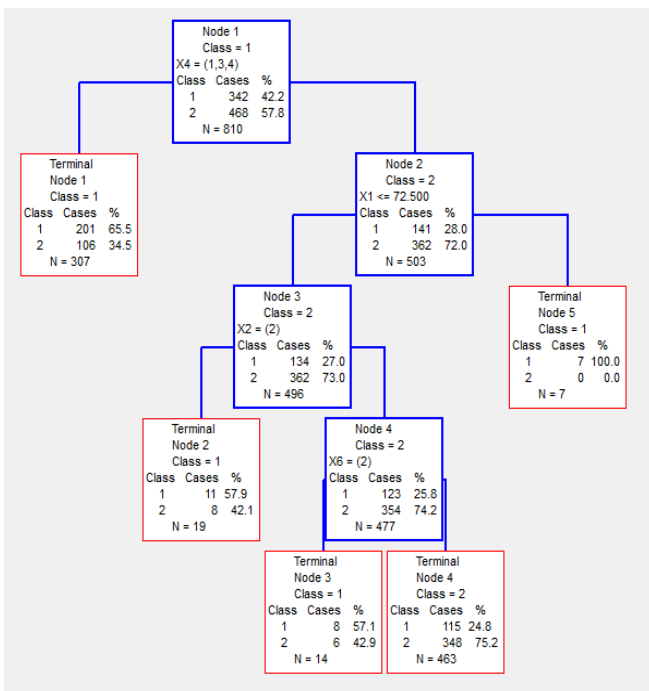
Karena pohon yang terbentuk berukuran besar, maka langkah selanjutnya adalah pemangkasan pohon maksimal. Pemangkasan dilakukan dengan memilih pemangkasan dengan nilai *relative cost* minimum, seperti yang disajikan pada Tabel 4.

Tabel 4. Urutan Pembentukan Pohon Klasifikasi

Tree Number	Terminal Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity Parameter
1	84	0,743 ± 0,034	0,454	0,000
19*	5	0,629 ± 0,033	0,593	0,003
20	4	0,653 ± 0,033	0,603	0,005
21	3	0,653 ± 0,033	0,618	0,008
22	2	0,657 ± 0,033	0,639	0,010

*Pohon Klasifikasi Optimal

Setelah dilakukan pemangkasan, langkah selanjutnya adalah memilih pohon klasifikasi optimal. Tabel 5 menunjukkan bahwa pohon klasifikasi yang terpilih adalah ukuran pohon sebesar 19 dengan jumlah *terminal nodes* sebanyak 5. Pohon klasifikasi optimum yang terbentuk disajikan pada Gambar 8.



Gambar 8. Pohon Klasifikasi Optimum

Gambar 8 menunjukkan bahwa variabel jenis pengobatan yang dijalani oleh pasien kanker serviks (X_4) merupakan variabel pemilah utama dan paling menentukan klasifikasi pasien kanker serviks. Selanjutnya variabel yang juga menentukan klasifikasi pasien kanker serviks adalah variabel usia (X_1), status anemia (X_2) dan status penyakit penyerta (X_6). Karakteristik simpul terminal yang mengklasifikasikan data menjadi salah satu dari kedua status rekurensi dijelaskan sebagai berikut.

1. Simpul terminal 1 terdiri dari 307 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang menjalani pengobatan transfusi PRC, kombinasi kemoterapi dan transfusi PRC, dan operasi.
2. Simpul terminal 2 terdiri dari 19 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.

3. Simpul terminal 3 terdiri dari 14 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang penyakit kanker serviks yang dideritanya merupakan penyakit penyerta, tidak memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.
4. Simpul terminal 4 terdiri dari 463 pasien kanker serviks yang diprediksi sebagai pasien yang rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang penyakit kanker serviks yang dideritanya bukan merupakan penyakit penyerta, tidak memiliki anemia, berusia kurang dari atau sama dengan 72,5 tahun dan menjalani pengobatan kemoterapi.
5. Simpul terminal 5 terdiri dari 7 pasien kanker serviks yang diprediksi sebagai pasien yang tidak rekuren kanker serviks. Karakteristik pasien pada simpul ini adalah pasien kanker serviks yang berusia lebih dari 72,5 tahun dan menjalani pengobatan kemoterapi.

Karakteristik pasien kanker serviks berdasarkan kelima simpul terminal adalah sebagai berikut.

Tabel 5. Karakteristik Pasien Kanker Serviks di RSUD dr. Soetomo

Tidak Rekuren		Rekuren	
• Pasien yang menjalani pengobatan transfusi PRC, kemoterapi & transfusi PRC dan operasi	• Pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun, tidak ada anemia dan kanker serviks yang dideritanya bukan penyakit penyerta		
• Pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun dan ada anemia			
• Pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun, tidak ada anemia dan kanker serviks yang dideritanya merupakan penyakit penyerta			
• Pasien yang menjalani pengobatan kemoterapi dan berusia > 72.5 tahun			

Langkah selanjutnya adalah menghitung ketepatan klasifikasi dari pohon CART yang diperoleh. Hasil ketepatan klasifikasi untuk data *learning* ditunjukkan pada Tabel 6.

Tabel 6. Evaluasi Performansi pada Data *Learning*

Fold	Akurasi	Sensitivity	Specificity
1	76,85%	74,73%	78,40%
2	78,40%	72,53%	82,67%
3	77,62%	72,66%	81,10%
4	79,01%	73,72%	82,89%
5	76,85%	71,17%	81,02%
Rata-rata	77,75%	72,96%	81,22%

Tabel 6 menunjukkan bahwa model yang terbaik ditunjukkan oleh *fold 4* karena memiliki nilai akurasi paling tinggi, yaitu sebesar 79,01 persen. Rata-rata akurasi untuk data *learning* sebesar 77,75 persen. Selain itu, rata-rata nilai

sensitivitas dan spesifisitas masing-masing sebesar 72,96 persen dan 81,22 persen.

Sedangkan ketepatan klasifikasi untuk data prediksi ditunjukkan pada Tabel 7.

Tabel 7. Ketepatan Klasifikasi Untuk Data Prediksi

Observasi	Prediksi		Total
	Tidak rekuren	Rekuren	
Tidak rekuren	222	120	342
Rekuren	130	338	468
Total	352	458	810

Perhitungan akurasi, *sensitivity* dan *specificity* untuk data *testing* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{222 + 338}{810} \right) \times 100\% = 69,14\%$$

$$\text{Sensitivity} = \left(\frac{222}{342} \right) \times 100\% = 64,91\%$$

$$\text{Specificity} = \left(\frac{338}{468} \right) \times 100\% = 72,22\%$$

Dari perhitungan yang dilakukan, nilai akurasi untuk data prediksi sebesar 69,14 persen, maka dapat dikatakan bahwa pohon optimal yang terbentuk sudah baik dan sesuai jika digunakan untuk mengklasifikasikan data baru. Sedangkan nilai *sensitivity* dan *specificity* masing-masing adalah sebesar 64,91 persen dan 72,22 persen.

V. KESIMPULAN / RINGKASAN

Karakteristik pasien kanker serviks berdasarkan status rekurensi kanker serviks diperoleh bahwa mayoritas pasien mengalami rekuren kanker serviks. Rata-rata usia pasien yang tidak rekuren adalah 50,62 tahun dan pada pasien yang rekuren sebesar 49,32 tahun. Selain itu, mayoritas pasien menderita stadium III, menjalani pengobatan kemoterapi, tidak ada komplikasi dan bukan merupakan penyakit penyerta. Klasifikasi CART menghasilkan rata-rata nilai akurasi untuk data learning sebesar 77,75 persen dan 69,14 persen untuk data prediksi. Jumlah *terminal nodes* yang terbentuk sebanyak 5, yaitu 4 *terminal nodes* diklasifikasikan sebagai pasien yang tidak rekuren kanker serviks, dan 1 *terminal nodes* diklasifikasikan sebagai pasien yang rekuren kanker serviks. Karakteristik pasien berdasarkan faktor-faktor yang mempengaruhi rekurensi kanker serviks adalah.

- Karakteristik pasien yang tidak rekuren adalah pasien yang menjalani pengobatan transfusi PRC, kemoterapi & transfusi PRC dan operasi; pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun dan ada anemia; pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun, tidak ada anemia dan kanker serviks yang dideritanya merupakan penyakit penyerta; dan pasien yang menjalani pengobatan kemoterapi dan berusia > 72.5 tahun
- Karakteristik pasien yang rekuren adalah pasien yang menjalani pengobatan kemoterapi yang berusia ≤ 72.5 tahun, tidak ada anemia dan kanker serviks yang dideritanya bukan penyakit penyerta.

DAFTAR PUSTAKA

- [1] World Health Organization. (2010). *World Cancer Report*. WHO Press
- [2] Nurlaila, I dan Hadi, Miftachul. (2014). *Kanker: Pertumbuhan Terapi dan Nonmedis*. Diperoleh dari: <http://www.nano.lipi.go.id/utama.cgi?artikel&1187593839>. Diakses tanggal 27 Maret 2015
- [3] Almarogi, S. (2015). *Gejala dan Penyebab Kanker Serviks*. Diperoleh dari : <http://www.infosehat.id/gejala-penyebab-kanker-serviks/>. Diakses tanggal 26 Maret 2015
- [4] Meilinda, T. (2014). *Kanker serviks*. Diperoleh dari: <http://mahasiswa.ung.ac.id/811412083/home/2014/12/9/makalah-kanker-serviks.html>. Diakses tanggal: 27 Maret 2015
- [5] Chamim. (2006). *Buku Acuan Nasional Onkologi*. In: M Farid Aziz, Adrijoyo, Abdul Bari Saifuddin, editors. Penentuan stadium klinik dan pembedahan kanker ginekologi. Jakarta : Yayasan Bina Pustaka Sarwono Prawirohadjo. p. 173 - 181
- [6] Mauliddah, N. F. (2014). *Kualitas Hidup Berdasarkan Karakteristik Pasien Kanker Serviks yang telah Melakukan Terapi di RSUD Dr. Soetomo Surabaya*. Skripsi Fakultas Kesehatan Masyarakat Universeitas Jember
- [7] Dewi, C. C. (2013). *Pengaruh Terapi Menulis Terhadap Penurunan Tingkat Kecemasan Pasien Kanker Serviks Stadium Lanjut di POSA RSUD Dr. Soetomo Surabaya*. Skripsi Program Studi Pendidikan Bidan Universitas Airlangga
- [8] Breiman L., Friedman J.H., Olshen R.A., & Stone C.J. (1993). *Classification And Regression Trees*. New York: Chapman And Hall
- [9] Nuriyah. (2013). *Perbandingan Metode Chi-Square Automatic Interaction Detection (CHAID) dan Classification And Regression Trees (CART) dalam Menentukan Klasifikasi Alumini UIN Sunan Kalijaga Berdasarkan Masa Studi*. Skripsi Program Sarjana Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga
- [10] Novitasari, D.A. (2014). *Analisis Survival pada Data Rekurensi dengan Menggunakan Counting Process Approach dan Model PWP-GT Study Kasus: Data Kanker Serviks di Rumah Sakit Dr. Soetomo Surabaya*. Tesis Program Magister Jurusan Statistika Institut Teknologi Sepuluh Nopember
- [11] Lewis, M.D dan Roger, J. (2000). *An Introduction to Classification and Regression Trees (CART) Analysis*. Annual Meeting of Society For Academic Emergency. California, UCLA Medical Center
- [12] Agresti, A. (2002). *Categorical Data Analysis*. John Wiley and Sons, Inc., New York
- [13] Schneider, J., 1997. *Carnegie Mellon University : School of Computer Science*. [Online] Available at: <http://www.cs.cmu.edu/~schneide/tut5/node42.html>
- [14] Yamato, K., Fen, J., Kobuchi, H., Nasu, Y., Yamada, T., Nishihara, T., et al. (2006). Induction of Cell Death in Human Papillomavirus 18-Positive Cervical Cancer Cells by E6 SIRNA. *Cancer Gene Therapy*. 13: 234-241.
- [15] Dunst, J., Kuhnt, T., Strauss, H., Krause, U., Pelz, T., Koelbl, H., et al. (2003). Anemia in Cervical Cancers : Impact on Survival, Pattern of Relapse, and Association with Hypoxia and Angiogenesis. *International Journal of Radiation*. 56. 778 – 787
- [16] Loka, A., Tsukuma, H., Ajiki, W., Oshima, A. (2005). Influence Of Age On Cervical Cancer Survival in Japan. *Journal Oxford JJCO*. 35. 464 – 469
- [17] Lee, Y-Y., Kim, T-J., Kim, J-Y., Choi, C, H., Do, I-G., Song, S, Y., et al (2013). Genetic Profiling to Predict Recurrence of Early Cervical Cancer. *Journal Gynecologic Oncology*. 13. 650 – 654
- [18] Matsuura, Y., Kawagoe, T., Toki, N., Tanaka, M., Kashimura, M. (2006). Long-standing Complications after Treatment for Cancer of the Uterine Cervix – Clinical Significance of Medical Examination at 5 years after Treatment. *International Journal of Gynecological Cancer*. 16. 294 – 297
- [19] Choi, M., Fuller, C, D., Thomas, C, R., Wang, S, J. (2008). Conditional Survival In Ovarian Cancer : Result From The SEER Dataset 1988 – 2001. *Journal Gynecologic Oncology*. 109. 203 – 209
- [20] Novitasari, D.A. (2014). *Analisis Survival pada Data Rekurensi dengan Menggunakan Counting Process Approach dan Model PWP-GT Study Kasus: Data Kanker Serviks di Rumah Sakit DR. Soetomo Surabaya*. Tesis Program Magister Jurusan Statistika Institut Teknologi Sepuluh Nopember