

# Pendekatan Metode *Classification and Regression Tree* untuk Diagnosis Tingkat Keganasan Kanker pada Pasien Kanker Tiroid

Sri Hartati Selviani Handayani dan Santi Wulan Purnami

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: santi\_wp@statistika.its.ac.id

**Abstrak**—Kanker Tiroid merupakan urutan kesembilan dari insiden kanker di Indonesia. Kanker Tiroid merupakan penyakit yang tidak menular yang menyerang pada bagian depan leher sedikit dibawah laring yang berbentuk kupu-kupu. Kelenjar tiroid sering kali membatasi kemampuan menyerap yodium dan membatasi kemampuan menghasilkan hormon tiroid, tetapi kadang menghasilkan cukup banyak hormon tiroid yang mengakibatkan hipertiroidisme. Sehingga perlu dilakukan pemeriksaan agar diketahui kanker tersebut jinak atau ganas, dan agar tidak terjadi kesalahan dalam penanganannya. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi timbulnya kanker tiroid. Metode *Classification and Regression Tree* merupakan analisis klasifikasi dimana pada variable respon dapat berskala katagorik dan kontinu. Data yang digunakan pada penelitian ini berupa data sekunder pasien yang diperoleh dari data rekam medis di rumah sakit “X” Surabaya pada tahun 2011 dan 2012 dengan variable respon berskala katagorik. Kombinasi learning dan testing yang digunakan adalah kombinasi/kelompok 1 dengan data learning sebanyak 65 data dan testing sebanyak 17 data. Hasil pengklasifikasian pada metode CART dari diagnosis tingkat keganasan pada kanker tiroid dengan faktor yang paling berpengaruh yaitu hasil pemeriksaan klinis.

**Kata Kunci**—CART, Kanker Tiroid

## I. PENDAHULUAN

Kanker tiroid merupakan kelenjar tiroid yang berada pada bagian depan leher sedikit di bawah laring berbentuk kupu-kupu. Kelenjar tiroid merupakan bagian tubuh yang jarang mengalami keganasan, terjadi 0,85% dan 2,5% dari seluruh keganasan kanker tiroid pada laki-laki dan perempuan dengan perbandingan 1:3 [1]. Dengan kata lain kanker tiroid lebih sering terjadi pada perempuan. Umumnya, kanker tiroid paling sering muncul pada usia 20-50 tahun, namun kanker ini dapat terjadi pada semua usia [2].

Di Indonesia kanker kelenjar tiroid termasuk urutan kesembilan dari insiden kanker yang ada di Indonesia. Kelenjar tiroid menghasilkan hormon tiroksin yang mengatur tekanan darah, laju metabolisme dalam tubuh, mengatur denyut jantung, suhu tubuh, berat badan dan mengontrol tingkat kalsium dalam tubuh [3].

Terdapat dua klasifikasi tipe kanker tiroid berdasarkan tingkat keganasannya yaitu jinak dan ganas. Kanker tiroid yang jinak tidak akan mengganggu aktivitas maupun fungsi bagian tubuh lainnya, sedangkan pada kanker tiroid yang ganas sel kanker tersebut akan mengganggu fungsi tubuh dan akan menyebar keseluruh tubuh secara sporadik sehingga dapat menyebabkan kematian. Kanker tiroid yang ganas

bernodul tunggal, terkadang keras, tidak rata, dan fungsinya kurang baik jika dibandingkan dengan fungsi jaringan tiroid di sekitarnya.

Diagnosis pada kanker tiroid sangat perlu untuk mengetahui kanker tersebut ganas atau jinak, sehingga tidak salah dalam pemberian terapi atau pengobatan terhadap kanker tersebut. Pada penelitian ini diketahui klasifikasi pasien kanker tiroid berdasarkan tingkat keganasannya (jinak atau ganas) yang menyebabkan timbulnya kanker tiroid yang berkembang di rumah sakit “X” Surabaya.

Penelitian kanker tiroid sebelumnya pernah dilakukan oleh Sumartini (2013) dengan menggunakan metode Logistik Biner didapatkan bahwa variabel yang signifikan terhadap hasil patologi kanker tiroid yaitu usia pasien dan riwayat penyakit keluarga pasien [4]. Selain itu kasus kanker tiroid pernah diteliti oleh Istichana (2013) dengan menggunakan metode Chi-Squared Automatic Interaction Detection Analysis di dapatkan bahwa variabel yang signifikan terhadap hasil patologi kanker tiroid yaitu hasil pemeriksaan USG [5]. Permasalahan yang akan dipecahkan dalam penelitian ini adalah bagaimana klasifikasi dan faktor yang memberi kontribusi tertinggi terhadap klasifikasi pasien kanker tiroid berdasarkan tingkat keganasan kanker tiroid dengan pendekatan metode *Classification and Regression Tree*. Tujuan yang ingin dicapai dalam penelitian ini adalah ingin mengetahui klasifikasi, beserta faktor-faktor yang mempengaruhinya.

## II. TINJAUAN PUSTAKA

### A. *Classification and Regression Tree* (CART)

CART adalah metode statistik nonparametrik yang digunakan untuk menggambarkan hubungan antara variabel respon (tidak bebas) dengan satu atau lebih variabel prediktor (bebas) yang di kembangkan untuk topik analisis klasifikasi, baik untuk variabel respon yang katagorik maupun kontinu. Pada CART model pohon yang dihasilkan bergantung pada skala variabel respon, jika variabel respon data berbentuk kontinu maka model pohon yang dihasilkan adalah *regression trees* (pohon regresi) sedangkan bila variabel respon mempunyai skala kategorik maka pohon yang dihasilkan adalah *classification trees* (pohon klasifikasi) [6]. Tujuan metode CART yaitu untuk mendapatkan suatu kelompok data yang akurat sebagai penciri suatu pengklasifikasian.

Pengklasifikasian pada pohon klasifikasi ada 4 komponen yaitu variabel respon, variabel prediktor, data *learning* dan data *testing* [7]. Data *learning* adalah data yang digunakan

untuk verifikasi model yaitu kemampuan model dalam mengakomodasi keragaman data. Data *testing* untuk validasi model yaitu untuk mengetahui seberapa tepat model yang sudah dibentuk.

Analisis pohon klasifikasi merupakan salah satu metode klasifikasi *Binary Recursive Partitioning* [6]. Penyekatan dimulai dengan dua anak gugus berdasarkan peubah bebas yang dianggap paling signifikan. Kemudian dilakukan penyekatan/*Rekursive* yaitu proses penyekatan biner yang dapat dilakukan secara berulang-ulang, simpul induk dapat dipilah menjadi dua simpul anak dan simpul anak dapat dipilah menjadi dua simpul anak yang lain dan seterusnya. Proses pemilahan dimulai dari simpul utama (*root node*) yang berisi data yang akan dipilah. Pemilahan dilakukan untuk memilah data menjadi 2 kelompok, yaitu kelompok yang masuk simpul kiri dan yang masuk simpul kanan. Pemilahan dilakukan pada tiap simpul sampai didapatkan suatu simpul terminal/akhir. Variabel yang memilah pada simpul utama adalah variabel terpenting dalam menduga kelas dari amatan.

Langkah-langkah pada metode CART yaitu sebagai berikut

1. Pembentukan Pohon Klasifikasi

Pembentukan pohon klasifikasi ada beberapa tahap yang memerlukan *learning sample L*, berikut tahapnya.

a. Pemilihan Pemilahan (*Classifier*)

Pemilihan pemilah dilihat dari pohon atau variable responnya. Pemilahan dilakukan untuk mengurangi keheterogenan pada simpul utama atau agar mendapatkan pemilah yang optimal. Untuk mengukur tingkat keheterogenan suatu kelas dari suatu simpul pada pohon klasifikasi digunakan *Impurity measure i(t)*. Pada pemilihan pemilahan data belum ada ketentuan untuk pemilah yang terbaik maka unsur pemilah yang di gunakan yaitu Indeks *Gini* karena sangat mudah dan sesuai untuk diterapkan dalam berbagai kasus dan mempunyai perhitungan yang sangat sederhana dan cepat dengan fungsi keheterogenan Indeks *Gini* yaitu:

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

(1)

dengan *i(t)* adalah fungsi keheterogenitas indeks gini,

*p(i|t)* adalah proporsi kelas *i* pada simpul *t*,

*p(j|t)* adalah proporsi kelas *j* pada simpul *t*.

Evaluasi pemilahan digunakan *Goodness of Split O(s,t)* yaitu suatu penurunan keheterogenan dengan rumus

$$O(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

(2)

dimana:

*O(s,t)* = Kriteria *Goodness-of-split*

*i(t)* = Fungsi keheterogenan indeks Gini

*p\_L* = Proporsi pengamatan menuju simpul kiri

*p\_R* = Proporsi pengamatan menuju simpul kanan

*i(t\_L)* = Fungsi keheterogenan pada simpul anak kiri

*i(t\_R)* = Fungsi keheterogenan pada simpul anak kanan

Pemilahan yang menghasilkan *O(s,t)* lebih tinggi merupakan pemilah terbaik karena mereduksi heterogenitas lebih tinggi. Pengembangan pohon dilakukan dengan pencarian pemilah yang mungkin pada simpul *t1* yang kemudian akan dipilah menjadi *t2* dan *t3*, dan seterusnya.

b. Penentuan Simpul Terminal

Simpul *t* akan menjadi simpul terminal atau tidak akan dipilah lagi bila tidak terdapat penurunan keheterogenan dan

adanya batasan minimum *n* misalnya hanya terdapat satu pengamatan pada tiap simpul anak.

c. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah anggota kelas terbanyak, yaitu :

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

dimana:

*p(j|t)* = proporsi kelas *j* pada simpul *t*

*N\_j(t)* = jumlah pengamatan kelas *j* pada simpul *t*

*N(t)* = jumlah pengamatan pada simpul *t*

Maka label kelas untuk simpul terminal *t* adalah *j0* yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul *t* paling kecil sebesar  $r(t) = 1 - \max p(j|t)$ .

2. Pemangkasan Pohon Klasifikasi

Pemangkasan pohon klasifikasi dilakukan dengan memangkas pohon yang kurang penting sehingga didapatkan pohon yang optimal. Untuk mendapatkan ukuran pohon yang layak digunakan *Cost complexity minimum* [6]. Untuk sembarang pohon *t* yang merupakan sub pohon dari pohon terbesar  $T_{max}(T < T_{max})$ . Dimana  $\alpha$  adalah bilangan real yang disebut *complexity parameter*, Nilai  $\alpha \geq 0$ . Besarnya resubstitution estimate pohon *T* pada kompleksitas  $\alpha$  yaitu :

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

dimana:

$R_\alpha(T)$  = *resubstitution* suatu pohon *T* pada kompleksitas  $\alpha$

$R(T)$  = *resubstitution Estimate* (proporsi kesalahan pada sub pohon),

$\alpha$  = kompleksitas parameter (*complexity parameter*) pada penambahan satu simpul akhir pada pohon *T*

$|T|$  = ukuran banyaknya simpul terminal pohon *T* (*complexity*)

$R_\alpha(T)$  merupakan kombinasi linear dari *Resubstitution Estimate* dan nilai kompleksitas pohon yang dibentuk. *Cost complexity* menentukan suatu pohon bagian  $T(\alpha)$  yang meminimumkan  $R_\alpha(T)$  pada seluruh pohon bagian untuk setiap nilai  $\alpha$ . Nilai kompleksitas  $\alpha$  akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya pencarian pohon bagian  $\max(T(\alpha) < T$  yang dapat meminimumkan  $R_\alpha(T)$  yaitu:

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

Pemangkasan pohon klasifikasi dimulai dengan mengambil  $t_R$  yang merupakan simpul anak kanan dan  $t_L$  yang merupakan simpul anak kiri dari  $T_{max}$  yang dihasilkan dari simpul induk *t*. Jika diperoleh 2 simpul anak dan simpul induk yang memenuhi persamaan  $R(t) = R(t_R) + R(t_L)$ , maka simpul anak  $t_R$  dan  $t_L$  pemangkasan di hentikan.

3. Penentuan Pohon Klasifikasi Optimal

Pada ukuran pohon klasifikasi yang sangat besar akan memberikan nilai penduga yang sangat kecil, maka pohon ukuran ini sering dipilih untuk menduga respon. Tetapi ukuran pohon yang besar akan menyebabkan nilai *complexity* yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai pengganti yang cukup kecil.

B. Ketepatan Hasil Klasifikasi

Ketepatan klasifikasi dapat di ketahui dengan menghitung Total Tingkat Akurasi, yang merupakan ketepatan klasifikasi

observasi oleh fungsi klasifikasi. *Sensitivity* menggambarkan

Tabel 1.  
Hasil Klasifikasi

Observasi	Prediksi		Total Prediksi	<i>Sensitivity</i>	<i>Specificity</i>	Total Tingkat Akurasi
	Negatif	Positif				
Negatif	$n_{11}$	$n_{12}$	$N_1$	$\frac{n_{11}}{N_1}$	$\frac{n_{22}}{N_2}$	$\frac{n_{11} + n_{22}}{N}$
Positif	$n_{21}$	$n_{22}$	$N_2$			
Total	$N_1$	$N_2$	$N$			

Keterangan:

- $n_{11}$  = Jumlah observasi yang benar diprediksi negatif
- $n_{12}$  = Jumlah observasi yang salah diprediksi negatif
- $n_{21}$  = Jumlah observasi yang salah diprediksi positif
- $n_{22}$  = Jumlah observasi yang benar diprediksi positif
- $N_1$  = Jumlah observasi negatif
- $N_2$  = Jumlah observasi positif
- $N$  = Jumlah observasi

Tabel 2.  
Variabel Penelitian

No	Variabel	Skala	Keterangan
1	Tipe Kanker (Y)	Nominal	0 : Jinak 1 : Ganas
2	Jenis Kelamin (X1)	Nominal	0 : Laki-laki 1 : Perempuan
3	Usia (X2)	Nominal	0 : < 42 Tahun 1 : $\geq$ 42 Tahun
4	Hasil Pemeriksaan USG (X3)	Nominal	0 : Struma 1 : Ca thyroid 2 : Kista
5	Hasil Pemeriksaan Klinis (X4)	Nominal	0 : Struma 1 : Ca thyroid 2 : Kista
6	Riwayat Penyakit Gondok (X5)	Nominal	0 : Ada Riwayat 1 : Tidak Ada Riwayat
7	Riwayat Penyakit Keluarga (X6)	Nominal	0 : Ada Riwayat 1 : Tidak Ada Riwayat

akurasi pada sampel kelas negatif sedangkan *specificity* menggambarkan akurasi pada sampel kelas positif. Langkah perhitungannya dapat dilihat pada tabel 1.

### III. METODOLOGI PENELITIAN

#### A. Sumber Data dan Variabel Penelitian

Data sekunder yang digunakan dalam penelitian ini adalah data pasien yang melakukan pemeriksaan klinis dan patologi kanker tiroid yang diperoleh dari data rekam medis di rumah sakit "X" Surabaya pada tahun 2011 dan 2012.

Variabel yang digunakan dalam penelitian ini adalah variabel respon dan predictor pada tabel 2.

#### B. Langkah Analisis

Langkah-langkah yang dilakukan dalam penelitian ini yaitu:

1. Melakukan analisis deskriptif yaitu dengan menggunakan *pie chart* dan tabulasi silang untuk mengetahui karakteristik pasien hasil patologi kanker tiroid.



Gambar 1. Hasil Patologi

2. Melakukan pemisahan data penelitian menjadi 2 (dua) bagian yaitu data *learning* dan data *testing*.
3. Kombinasi data yang dicobakan pada penelitian ini yaitu dengan perbandingan antara *learning* dan *testing* yaitu data di bagi menjadi 5 kelompok. Dimana banyak data pada kelompok ke-*i* di jadikan sebagai data *testing* dan banyak data pada data *learning* untuk setiap kelompok ke-*i* merupakan penjumlahan data dari data *testing* kelompok lainnya dimana  $i = 1,2,3,4,5$ .
4. Membentuk pohon klasifikasi yaitu dengan pemilahan pemilih terbaik *n* penentuan simpul terminal. Simpul terminal diperoleh ketika tidak memungkinkan lagi melakukan pemilahan pada suatu simpul atau tidak terdapat penurunan keheterogenan yang berarti. Penandaan label kelas dari simpul terminal ditentukan berdasarkan aturan jumlah terbanyak.
5. Pemangkasan pohon klasifikasi yang maksimal sampai diperoleh ukuran pohon klasifikasi yang paling kecil dengan kriteria kompleksitas kesalahan (*cost complexity*) yang minimum.
6. Memilih pohon terbaik dengan melakukan validasi model dengan memasukkan data *testing* pada pohon klasifikasi optimal.

### IV. ANALISIS DAN PEMBAHASAN

#### A. Karakteristik Pasien di Rumah Sakit "X" Surabaya

Hasil patologi kanker tiroid pada rumah sakit "X" Surabaya pada tahun 2011 dan 2012 ada dua tipe yaitu jinak dan ganas yang menunjukkan bahwa hasil patologi kanker tiroid pada pasien yang didiagnosis terkena kanker tiroid jinak sejumlah 67 persen dan pada pasien yang didiagnosis terkena kanker tiroid ganas sejumlah 33 persen, dengan karakteristik pasien pada hasil patologi kanker tiroid yang signifikan yaitu hasil pemeriksaan USG dan hasil pemeriksaan klinis mayoritas hasil pemeriksaan struma, dan mayoritas tidak memiliki riwayat penyakit keluarga pada pasien.

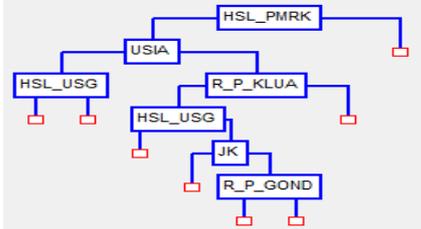
#### B. Klasifikasi Pasien Hasil Patologi Kanker Tiroid dengan Metode CART

Pembagian data *learning* dan data *testing* tidak memiliki ketentuan yang khusus, tetapi data *learning* lebih banyak dari data *testing*. Proporsi yang akan dicobakan pada penelitian ini dengan perbandingan antara *learning* dan *testing* yaitu data di bagi menjadi 5 kelompok. Dimana banyak data pada kelompok ke-*i* di jadikan sebagai data *testing* dan banyak data pada data *learning* untuk setiap kelompok ke-*i* merupakan penjumlahan data dari data *testing* kelompok lainnya dimana  $i = 1,2,3,4,5$ . Ketepatan klasifikasi pada data *learning* dan *testing* dijadikan dasar karena menggambarkan klasifikasi yang terbentuk.

Ketepatan klasifikasi data *learning* dan data *testing* yang paling tinggi yaitu pada kelompok 1 dengan ketepatan

Tabel 3.  
Perbandingan Ketepatan Klasifikasi antar Kombinasi Data

Kelompok Data	Kombinasi Data		Ketepatan Klasifikasi	
	Learning	Testing	Learning	Testing
1	65	17	0,846	0,941
2	65	17	0,862	0,882
3	66	16	0,909	0,813
4	66	16	0,848	0,688
5	66	16	0,909	0,875



Gambar 2. Pohon Klasifikasi Maksimal yang Terbentuk  
Tabel 4.

Skor Variabel Penting Pada Pohon Klasifikasi Maksimal	
Variable	Skore
Hasil Pemeriksaan Klinis ( $X_4$ )	100%
Hasil Pemeriksaan USG ( $X_3$ )	86,63%
Riwayat Penyakit Keluarga ( $X_6$ )	24,71%
Usia ( $X_2$ )	16,46%
Jenis Kelamin ( $X_1$ )	5,71%
Riwayat Penyakit Gondok ( $X_5$ )	4,23%

klasifikasi data *learning* sebesar 0,846 dan *testing* sebesar 0,941 sehingga kombinasi data ini yang akan digunakan pada analisis selanjutnya.

1) *Pembentukan Pohon Klasifikasi Maksimal*

Pohon klasifikasi maksimal merupakan pohon klasifikasi dengan jumlah simpul terminal terbanyak. Metode pemilihan pemilah pada penelitian ini dengan menggunakan Indeks Gini. Berikut gambar Pohon Klasifikasi Maksimal.

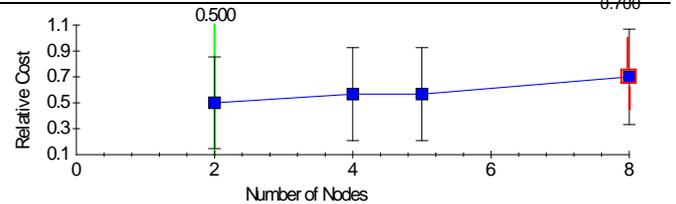
Pada pohon klasifikasi maksimal diatas maka terdapat 8 simpul terminal dengan 7 kedalaman., dimana seluruh variable berpengaruh pada pohon klasifikasi Maksimal.

Dari Gambar 1 dan tabel diatas dapat di lihat bahwa pemilah utama pada simpul induk adalah variabel hasil pemeriksaan klinis yang berada pada urutan pertama dengan skor 100 persen. Hal ini menunjukkan bahwa variabel hasil pemeriksaan klinis merupakan pemilah yang memiliki peranan utama dalam pembentukan pohon maksimal dan Pohon Klasifikasi maksimal menghasilkan nilai *relative cost* lebih besar dibandingkan *relative cost* pada pohon klasifikasi optimal. Maka perlu dilakukan pemangkasan pada pohon klasifikasi maksimal agar didapatkan nilai *relative cost* yang paling kecil. Garis merah menunjukkan nilai *relative cost* pada pohon maksimal sebesar 0,700 dan garis hijau menunjukkan nilai *relative cost minimum* pada pohon optimal sebesar 0,500.

Pohon klasifikasi maksimal adalah pohon klasifikasi dimana faktor-faktor yang mempengaruhi seluruhnya masuk ke dalam pohon klasifikasi. Pemangkasan dilakukan dengan melihat nilai *Test Set Relative Cost* yang terkecil yaitu pada pohon ke-4 dengan simpul ke-2 merupakan pohon optimal dengan uji kesalahan relatif  $0,500 \pm 0,354$  dimana biaya

Tabel 5.  
Ketepatan Hasil Klasifikasi pada Pohon Maksimal

Observasi	Prediksi			Total Prediksi	<i>Sensitivity</i>	<i>Specificity</i>	Total Tingkat Akurasi
	0	1					
Data Learning	0	36	4	40	90%	92%	90,76%
	1	2	23	25			
Data Testing	0	12	3	15	80%	50%	76,47%
	1	1	1	2			



Gambar 3. Plot Relative Cost

Tabel 6.  
Urutan Pohon Simpul Akhir

Tree Number	Terminal Nodes	Test Set Relative Cost	Resubstitution Relative Cost	Complexity
1	8	$0,700 \pm 0,368$	0,178	0,000
2	5	$0,567 \pm 0,359$	0,224	0,008
3	4	$0,567 \pm 0,359$	0,264	0,020
4**	2	$0,500 \pm 0,354$	0,400	0,034
5	1	$1,000 \pm 0,000$	1,000	0,300

\*\* : Pohon Klasifikasi Optimal

pengganti relatif sebesar 0,400 dan kompleksitas parameter sebesar 0,034.

2) *Pohon Klasifikasi Optimal*

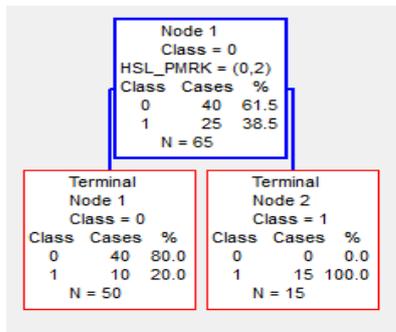
Pohon klasifikasi optimal merupakan pohon yang dihasilkan dari pemangkasan pohon klasifikasi maksimal. Pemilah yang terpilih akan membentuk suatu himpunan kelas yang disebut simpul utama (simpul 1). Di setiap pemilah akan menghasilkan 2 kelas yang dinamakan pada penelitian ini yaitu kelas 0 (jinak) dan kelas 1 (ganas) yang juga dinamakan sebagai simpul anak begitu seterusnya sampai simpul anak tersebut tidak dapat dipilah lagi dan berubah menjadi simpul terminal.

Klasifikasi pohon optimal yang terbentuk pada penelitian ini diperoleh 2 simpul terminal dengan 2 kedalaman. Berikut gambar pohon klasifikasi optimal yang dihasilkan pada pemangkasan pohon maksimal.

Dari 2 simpul terminal yang dihasilkan dalam klasifikasi pohon optimal 1 simpul terminal yang diprediksi masuk sebagai kategori pasien penderita kanker tiroid yang jinak dan 1 simpul terminal yang diprediksi masuk sebagai kategori pasien penderita kanker tiroid yang ganas.

Simpul terminal 1: Pada simpul terminal ini terdapat 50 pasien penderita kanker tiroid yang jinak dengan karakteristik pasien dimana hasil pemeriksaan klinis struma dan kista.

Simpul terminal 2: Pada simpul terminal ini terdapat 15 pasien penderita kanker tiroid yang ganas dengan karakteristik pasien hasil pemeriksaan klinisnya *ca thyroid*. Tabel Ketepatan hasil klasifikasi pada pohon optimal diatas memberikan informasi bahwa sebanyak 40 pasien yang didiagnosis kanker tiroid yang jinak seluruh datanya tepat dalam pengklasifikasian, sedangkan untuk pasien penderita kanker tiroid yang ganas ada 10 orang pasien salah dalam



Gambar 4. Pohon Klasifikasi Optimal

Tabel 7. Ketepatan Klasifikasi pada Pohon Optimal

Observasi	Prediksi			Misklasifikasi/ Kesalahan
	Jinak	Ganas	Total	
Jinak	40,00	0,00	40,00	0%
Ganas	10,00	15,00	25,00	40%
Tingkat Akurasi Total %				84,6%

Tabel 8.

Ketepatan Klasifikasi pada Pohon Optimal dengan menggunakan Data *Testing*

Observasi	Prediksi			Misklasifikasi/ Kesalahan
	Jinak	Ganas	Total	
Jinak	15,00	0,00	15,00	0%
Ganas	1,00	1,00	2,00	50%
Tingkat Akurasi Total %				94,1%

Tabel 9.

Ukuran Akurasi Klasifikasi

Kelompok Data	Evaluasi		Ketepatan Klasifikasi	
	<i>Sensitivity</i>	<i>Specificity</i>	<i>Learning</i>	<i>Testing</i>
1	1	0,6	0,846	0,941
2	1	0,571	0,862	0,882
3	0,96	0,8	0,909	0,813
4	1	0,474	0,848	0,688
5	0,98	0,783	0,909	0,875
Rata-rata	0,988	0,646	0,875	0,839

pengklasifikasiannya dan diperoleh total akurasi dari data *learning* sejumlah 84,6 persen. Maka dapat disimpulkan untuk nilai total akurasi pada data *learning* sejumlah 84,6 persen, yang artinya model klasifikasi optimal yang terbentuk baik untuk menggambarkan model klasifikasi.

### 3) Validasi Pohon Klasifikasi

Pada validasi pohon klasifikasi ini data yang diambil yaitu data *testing* sebanyak 17 data pasien yang dimasukkan ke dalam model pohon klasifikasi optimal yang telah terbentuk sebelumnya dari data *learning*. Data *testing* yang dimasukkan sebesar 17 data dari 82 data dengan pengambilan secara random pada pengelompokan yang telah dilakukan. Berikut tabel hasil ketepatan klasifikasi data *testing* pada pohon optimal.

Tabel ketepatan klasifikasi diatas memberikan informasi bahwa sebanyak 17 orang pasien yang menderita kanker tiroid pada data *testing*, dengan ketepatan klasifikasi pada pohon optimal yang dihasilkan sejumlah 94,1 persen, yang artinya model pohon yang telah terbentuk memiliki ketepatan hasil prediksi sejumlah 94,1 persen. Dari ketepatan klasifikasi

yang dihasilkan oleh data *testing* tersebut maka model klasifikasi optimal yang dihasilkan sudah baik atau layak dalam model merupakan variabel yang paling dominan diantara variabel lainnya terhadap pengklasifikasian.

Pada data *learning*, hasil klasifikasi pada pohon maksimal menunjukkan bahwa nilai *sensitivity* sejumlah 90 persen yang artinya dari 40 pasien yang mengidap kanker tiroid jinak, yang benar pengklasifikasiannya sebanyak 36 orang dan 4 orang pasien diduga salah dalam pengklasifikasiannya. Sedangkan nilai *specificity* sejumlah 92 persen karena dari 25 pasien yang mengidap kanker ganas, 23 orang pasien yang tepat dalam pengklasifikasiannya, dan diperoleh total akurasi dari data *learning* sejumlah 90,76 persen.

Pada data *testing*, hasil klasifikasi pada pohon maksimal menunjukkan bahwa nilai *sensitivity* sejumlah 80 persen yang artinya dari 15 pasien yang mengidap kanker tiroid jinak, yang benar pengklasifikasiannya sebanyak 12 orang dan 3 orang pasien diduga salah dalam pengklasifikasiannya. Sedangkan nilai *specificity* sejumlah 50 persen karena dari 2 pasien yang mengidap kanker ganas, 1 orang pasien yang tepat dalam pengklasifikasiannya, dan diperoleh total akurasi dari data *testing* sejumlah 76,47 persen.

### 4) Pemangkasan Pohon Klasifikasi Maksimal

Pemangkasan pohon klasifikasi dilakukan apabila pohon klasifikasi yang terbentuk berukuran sangat besar dan kompleks dalam penggambaran struktur data sehingga akan diperoleh ukuran pohon yang layak dan berdasarkan *cost Complexity minimum*.

## C. Evaluasi Ketepatan Klasifikasi

### 1) Perbandingan antar Kombinasi Data

Pada klasifikasi CART akan dicoba beberapa kombinasi data. Perbandingan yang akan dicobakan pada penelitian ini yaitu dengan perbandingan antara *learning* dan *testing* dengan data di bagi menjadi 5 kelompok dan hasilnya pada tabel 9.

Dari 5 kelompok kombinasi yang dicobakan diatas dapat dilihat bahwa ketepatan klasifikasi data *learning* dan data *testing* yang paling tinggi yaitu pada kelompok 1 dengan tingkat akurasi data *learning* sebesar 0,846 dan *testing* sebesar 0,941 sehingga kombinasi data kelompok 5 yang digunakan pada analisis dan rata-rata nilai akurasi dari 5 kelompok data diatas maka pada data *learning* sebesar 0,875 dan *testing* yaitu sebesar 0,839.

### 2) Hasil Penelitian Sebelumnya

Penelitian kanker tiroid sebelumnya pernah dilakukan oleh Sumartini (2013) dengan menggunakan metode Logistik Biner didapatkan bahwa variabel yang signifikan terhadap hasil patologi kanker tiroid yaitu usia pasien dan riwayat penyakit keluarga pasien. Selain itu kasus kanker tiroid pernah diteliti oleh Istichana (2013) dengan menggunakan metode *Chi-Squared Automatic Interaction Detection Analysis* di dapatkan bahwa variabel yang paling signifikan terhadap hasil patologi kanker tiroid yaitu hasil pemeriksaan USG.

Dari hasil penelitian kanker tiroid di rumah sakit "X" Surabaya dengan mengklasifikasikan faktor-faktor yang mempengaruhi hasil patologi kanker tiroid pada tahun 2011 dan 2012 dengan data, jumlah data dan variabel-variabel yang mempengaruhi sama, maka peneliti ingin mengetahui ketepatan klasifikasi tertinggi dari ketiga metode tersebut.

Tabel 11.  
Perbandingan Ketepatan Klasifikasi antar Metode

0	Variabel yang Signifikan/Berpengaruh	Metode		
		Logistik Biner	CHAID	CART
1	Hasil Pemeriksaan Klinis			✓
2	Hasil Pemeriksaan USG		✓	
3	Riwayat Penyakit Keluarga	✓		
4	Usia	✓		
5	Jenis Kelamin			
6	Riwayat Penyakit Gondok			
Ketepatan Klasifikasi		86,6%	86,6%	94,1%

Berikut tabel perbandingan antar metode Logistik Biner, CHAID (*Chi-Squared Automatic Interaction Detection Analysis*), dan CART (*Classification and Regression Tree*).

Dari ketiga metode klasifikasi diatas menunjukkan bahwa faktor-faktor yang berpegaruh terhadap hasil patologi kanker tiroid yaitu hasil pemeriksaan klinis, hasil pemeriksaan USG, riwayat penyakit keluarga dan usia. Ketepatan klasifikasi paling tinggi pada metode CART yaitu sejumlah 94,1 persen, yang artinya metode terbaik dalam pengklasifikasian pasien yang menderita kanker tiroid di rumah sakit "X" Surabaya yaitu dengan menggunakan metode CART (*Classification and Regression Tree*).

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Dari hasil analisis dan pembahasan mengenai klasifikasi pasien kanker tiroid di Rumah Sakit "X" Surabaya dan berdasarkan faktor-faktor yang mempengaruhi, dengan pendekatan CART diperoleh kesimpulan sebagai berikut :

1. Hasil penelitian menunjukkan bahwa hasil patologi kanker tiroid pada pasien yang didiagnosis terkena kanker tiroid jinak sejumlah 67 persen dan pada pasien yang didiagnosis terkena kanker tiroid ganas sejumlah 33 persen, dengan karakteristik pasien pada hasil patologi kanker tiroid yang signifikan yaitu hasil pemeriksaan USG dan hasil pemeriksaan klinis mayoritas hasil pemeriksaan struma, dan mayoritas tidak memiliki riwayat penyakit keluarga pada pasien.

2. Pada Model CART dengan menggunakan pengelompokan data secara random yaitu pada kelompok 1 dengan banyak data *learning* sebanyak 65 data dan *testing* 17 data yang menghasilkan ketepatan klasifikasi terbaik/tertinggi dibandingkan dari kelompok lainnya. Klasifikasi dari diagnosis tingkat keganasan pada kanker tiroid dengan faktor yang paling berpengaruh terhadap kanker tiroid yaitu pada hasil pemeriksaan klinis. Hasil klasifikasi pada metode CART pada pasien hasil patologi dapat disimpulkan bahwa pasien yang mengalami kanker tiroid sangat berhubungan erat dengan hasil pemeriksaan klinis yaitu pasien hasil patologi yang beresiko menderita kanker tiroid jinak hasil pemeriksaan klinisnya struma dan kista. Sedangkan pasien yang beresiko menderita kanker tiroid ganas yaitu *ca thyroid*. Dengan ketepatan klasifikasi data *learning* pada pasien kanker tiroid sejumlah 84,6 persen, maka dapat disimpulkan bahwa model klasifikasi optimal yang terbentuk baik untuk menggambarkan model klasifikasi pohon dan hasil validasi pada data testing sejumlah 94,1 persen, hal ini menyimpulkan

bahwa pohon klasifikasi yang diklasifikasikan pada data baru layak dalam model.

3. Faktor yang memberi kontribusi tertinggi pada klasifikasi pasien kanker tiroid dengan menggunakan metode CART yaitu hasil pemeriksaan klinis, sedangkan pada ketiga metode yang telah diteliti sebelumnya faktor-faktor yang berpengaruh pada klasifikasi pasien kanker tiroid yaitu hasil pemeriksaan klinis, hasil pemeriksaan USG, riwayat penyakit keluarga dan usia.

### B. Saran

Hasil pemeriksaan klinis merupakan variable yang sangat berpengaruh yang menyebabkan timbulnya kanker tiroid, oleh karena itu disarankan bagi pihak rumah sakit atau bagi orang yang pernah melakukan pemeriksaan klinis pada kelenjar tiroid sebaiknya memeriksakan dirinya sedini mungkin dan lebih mendalam karena bila dibiarkan akan membawa dampak yang buruk bagi kesehatan.

Pada penelitian selanjutnya sebaiknya melakukan konsultasi kembali dengan dokter ahli dalam melakukan riset, dan memperbanyak refrensi dari penelitian-penelitian sebelumnya agar memperoleh variable prediktor baru, seperti adanya riwayat pasien terkena radiasi, adanya riwayat pasien perokok atau adanya riwayat keluarga pasien perokok yang menyebabkan timbulnya kanker tiroid.

## DAFTAR PUSTAKA

- [1] Djokomoeljanto, (2009). Indonesian Society of Endocrinology, Indonesian Study Group of Thyroidology, Indonesian Society of Endocrinology Semarang Branch, The 2<sup>nd</sup> Thyroidology Update 2009. Semarang: Badan Penerbit Undip (Di unduh tanggal 22 September 2013).
- [2] Utama, Y. D. C (2012). Nilai diagnostik Karakteristik klinis di Bandingkan dengan Biopsi patologi Anatomi dalam mendiagnosis Karsinoma Tiroid, Skripsi, fakultas Kedokteran Universitas diponegoro, Semarang. [http://eprints.undip.ac.id/37824/1/Yohanes\\_Dona\\_G2\\_A008200\\_Lap\\_KTI.pdf](http://eprints.undip.ac.id/37824/1/Yohanes_Dona_G2_A008200_Lap_KTI.pdf) (Di unduh tanggal 14 Oktober 2013).
- [3] Laseduw, J.(2012) Kanker Tiroid. <http://kankertiroid.wordpress.com/2013/06/21/penyebab/>. (Diunduh tanggal 18 September 2013).
- [4] Sumartini, S. H. (2013). Analisis Regresi Biner Pada Faktor yang Mempengaruhi Hasil Patologi kanker Tiroid. Tugas Akhir Diploma III Jurusan Statistika, Surabaya, Institut Teknologi Sepuluh November.
- [5] Istichana, Y. Y. (2013). Penggunaan Metode CHAID Untuk Klasifikasi Pasien Hasil Patologi kanker Tiroid. Tugas Akhir Diploma III Jurusan Statistika, Surabaya, Institut Teknologi Sepuluh November.
- [6] Breiman L., Friedman J.H Olshen R.A & Stone C.J. (1993). *Classification And Regression Tree*. New York, NY: Chapman And Hall
- [7] Lewis, M.D dan Roger, J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*.
- [8] American Cancer Society. 2012. [Online]. Available:(<http://www.thyroid.cancer.int/> ,diakses tanggal 8 Oktober 2013).
- [9] Hamdani, W dan Sampepajung, D. Thyroid cancer. The diagnose and the management. Department of Surgery Medical Faculty, Hasanuddin University, Makassar <http://med.unhas.ac.id/jurnal/attachments/article/74/TYROID%20Ca-William%20Hamdani.pdf> (Diunduh tanggal 23 September 2013).
- [10] Ramdani, Y. (2013). Klasifikasi Hasil Pap Smear Test Kanker Serviks berdasarkan Faktor Resiko. Tugas Akhir Sarjana Jurusan Statistika, Surabaya, Institut Teknologi Sepuluh November.
- [11] Subekti, I., Aru, W., Sudoyo., B. S., Idrus, A., Marcellius, S., Sesiati, S. (2010). *Karsinoma Tiroid. Buku Ajaran Ilmu Penyakit Dalam. Edisi 5. Jakarta : Interna Publishing. P. 2031-37*
- [12] Walpole, (1995). *Pengantar Statistika*. Gramedia Pustaka Utama: Jakarta.
- [13] Agresti, A. (2002). *Categorikal Data Analysis*. Jhon Wiley and Sons, Inc., New York.