

ANALISIS FAKTOR RISIKO PENYAKIT JANTUNG KORONER MENGGUNAKAN METODE REGRESI LOGISTIK DAN CHAID

Astri Atti¹; Bunawan Sunarlim²; Utami Dyah Syafitri³

ABSTRACT

The aims of the article are to investigate risk factor that influences Coronary Heart Disease (CHD) by logistic regression method, and to know structural association in segmentation observation based on CHD patient or not by CHAID method, and comparing the result of classification between them. The result of logistic regression method shows that CHD status are influenced by hypercholesterol, diabetes, Body Mass Index (BMI), and age. CHAID method indicates that hypercholesterol has the strongest association to CHD, with three segmentations or groups for CHAID. The first group is hypercholesterol sufferer and obese, second is hypercholesterol sufferer with normal or thin body and non hypertension, and the last is non hypercholesterol but suffer diabetes mellitus and male. Total misclassification rate for logistic regression is about 32,8% while CHAID method is about 30,7%.

Keywords: logistic regression, chaid, coronary heart disease, misclassification

ABSTRAK

Artikel mempelajari faktor risiko yang mempengaruhi Penyakit Jantung Koroner (PJK) dengan regresi logistik dan mengetahui asosiasi struktural dengan segmentasi pengamatan berdasar pasien PJK atau tidak dengan metode CHAID. Hasil klasifikasi diperbandingkan. Hasil Regresi Logistik diperoleh status PJK dipengaruhi oleh hiperkolesterol, diabetes, indeks berat tubuh, dan umur. Metode CHAID menjelaskan yang asosiasinya paling kuat dengan PJK adalah hiperkolesterol. Ada tiga grup segmentasi berdasar CHAID, grup pertama adalah penderita hiperkolesterol dan kegemukan, kedua penderita hiperkolesterol, berat tubuh normal, dan non hipertensi, dan terakhir tidak hipertensi tetapi menderita diabetes mellitus dan laki-laki. Total salah klasifikasi untuk regresi logistik 32.8 % sedang dengan metode Chaid 30.7 % .

Kata kunci: regresi logistik, chaid, penyakit jantung koroner, salah klasifikasi

¹ Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Nusa Cendana, Kupang

² Departemen Statistika FMIPA IPB, Bogor dan Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Bina Nusantara, Jl. K.H. Syahdan No. 9, Palmerah, Jakarta Barat 11480,

³ Departemen Statistika FMIPA IPB, Bogor

PENDAHULUAN

Menurut Rosnah (1999), salah satu penyakit yang cenderung meningkat di masyarakat modern adalah Penyakit Jantung Koroner (PJK). PJK adalah suatu kelainan yang disebabkan oleh penyempitan pembuluh arteri (Fitriani, 2007). Beberapa faktor penyebab PJK menurut Palilati (2003) diantaranya adalah hipertensi, hiperkolesterol, diabetes mellitus, merokok, obesitas, kurang olahraga, riwayat keluarga, stres, umur, dan bahkan jenis kelamin. Untuk mengungkapkan adanya asosiasi antara peubah respons dan peubah penjas pada data kategorik, dapat digunakan metode regresi logistik dan metode CHAID. Ture et. al., (2006) melakukan penelitian menggunakan metode regresi logistik dan metode CHAID pada kasus “prestasi akademik” yang menyimpulkan bahwa metode CHAID mempunyai kesalahan klasifikasi (*misclassification rate*) yang lebih kecil (32,3%) dibandingkan metode regresi logistik (35,71%). Penelitian lain dalam membandingkan metode regresi logistik dan CHAID juga dilakukan oleh Gonzales (2003) untuk kasus “pendengar stasiun radio” dan diketahui bahwa kedua metode yang digunakan menghasilkan nilai kesalahan klasifikasi yang hampir sama, yaitu untuk metode regresi logistik sebesar 37,92% sedangkan metode CHAID sekitar 38,23%. Dalam penelitian ini akan digunakan metode regresi logistik dan metode CHAID untuk kasus PJK.

Penelitian bertujuan untuk menelusuri faktor risiko yang secara nyata mempengaruhi PJK menggunakan metode regresi logistik. Tujuan lain adalah melihat asosiasi berstruktur dalam pengelompokan pengamatan menurut peubah respons menggunakan metode CHAID serta membandingkan hasil klasifikasi antara kedua metode.

REGRESI LOGISTIK

Analisis regresi logistik adalah analisis yang digunakan untuk melihat hubungan antara peubah respons berupa data kualitatif dengan peubah penjas yang berupa data kualitatif maupun data kuantitatif. Peubah respons dalam regresi logistik dapat dalam bentuk dikotom (biner) maupun polikotom (ordinal atau nominal).

Hasil pengamatan yang memiliki p peubah penjas yang ditunjukkan oleh vektor $\mathbf{x}' = (X_1, X_2, \dots, X_p)$ yang berpasangan dengan peubah respons Y yang bernilai 1 dan 0, dan $y=1$ menyatakan bahwa respons memiliki kriteria yang ditentukan dan $y=0$ tidak memiliki kriteria yang ditentukan merupakan regresi logistik peubah respons biner. Model umum persamaan peluang regresi logistik dengan p peubah penjas sebagai berikut (Hosmer & Lemeshow 2000).

$$\pi(\mathbf{x}) = \frac{\exp[\mathbf{g}(\mathbf{x})]}{1 + \exp[\mathbf{g}(\mathbf{x})]} \quad (1)$$

dengan transformasi logit : $\text{logit} [\pi(\mathbf{x})] = \ln \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \mathbf{g}(\mathbf{x})$

$$\text{dan } \mathbf{g}(\mathbf{x}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2)$$

Pengujian kesesuaian model dilakukan secara simultan menggunakan statistik uji-G dan secara parsial menggunakan statistik uji Wald. Pemilihan peubah penjas menggunakan metode *backward elimination*, diawali dengan semua peubah dimasukkan ke dalam model kemudian peubah akan diuji satu persatu. Jika ditemukan peubah yang tidak nyata pada nilai α yang ditentukan maka peubah tersebut dikeluarkan dari model dan jika nyata maka akan tetap berada dalam model (Gonzales 2003). Menurut Hosmer dan Lemeshow (2000), rasio odds dilambangkan sebagai $\psi = \exp(\beta_i)$ dan diinterpretasikan sebagai kecenderungan $y=1$ pada $x=1$ sebesar ψ kali dibandingkan pada $x=0$.

Selang kepercayaan $(1-\alpha)100\%$ bagi rasio odds adalah :

$$\exp[\hat{\beta}_i \pm Z_{1-\alpha/2} \hat{SE}(\hat{\beta}_i)] \quad (3)$$

sedangkan selang kepercayaan $(1-\alpha)100\%$ bagi $\pi(\mathbf{x})$, yaitu:

$$\frac{e^{\hat{g}(\mathbf{x}) \pm Z_{1-\alpha/2} \hat{SE}[\hat{g}(\mathbf{x})]}}{1 + e^{\hat{g}(\mathbf{x}) \pm Z_{1-\alpha/2} \hat{SE}[\hat{g}(\mathbf{x})]}} \quad (4)$$

dan $\hat{SE}[\hat{g}(\mathbf{x})] = \sqrt{\hat{V}\hat{a}r \hat{g}(\mathbf{x})}$ (5)

dan $\hat{V}\hat{a}r[\hat{g}(\mathbf{x})] = \sum_{j=0}^p x_j^2 \hat{V}\hat{a}r(\hat{\beta}_j) + \sum_{j=0}^p \sum_{k=j+1}^p 2x_j x_k \text{C}\hat{o}v(\hat{\beta}_j, \hat{\beta}_k)$ (6)

METODE CHAID

Metode CHAID (*Chi-square Automatic Interaction Detection*) adalah salah satu tipe dari metode AID (*Automatic Interaction Detection*), yaitu metode yang digunakan untuk menganalisis asosiasi berstruktur antara peubah respons dan peubah penjelas dalam segugus data (Fielding, 1977). CHAID hanya dapat menganalisis data jika peubah respons dan peubah penjelasnya berskala nominal atau ordinal. Metode CHAID merupakan teknik eksplorasi nonparametrik untuk menganalisis sekumpulan data yang berukuran besar dan cukup efisien untuk menduga peubah penjelas yang paling nyata terhadap peubah respons (Du Toit, et. al., 1986).

Metode CHAID, menganalisis suatu gugus data dengan cara memisahkannya menjadi beberapa kelompok secara bertahap (Fielding 1977). Proses pemisahan dilakukan secara iteratif dimulai dari peubah penjelas yang mempunyai asosiasi paling kuat dengan peubah respons yang digambarkan oleh besarnya nilai-p (*p-value*) berdasarkan uji Khi-kuadrat. Dalam proses itu akan dilakukan juga penggabungan kategori dalam satu peubah penjelas yang tidak memiliki asosiasi yang nyata dengan peubah respons. Peubah penjelas kategori yang memenuhi syarat dapat digabung sesuai tipe skala yang ditentukan. Beberapa kategori peubah penjelas berskala nominal dapat digabung sedangkan untuk peubah berskala ordinal, kategori yang dapat digabung adalah yang saling berdekatan (Magidson and Vermunt, 2006).

Tahapan analisis metode CHAID adalah sebagai berikut (Kass 1980). Pertama, untuk setiap peubah penjelas dibuat tabulasi silang antara kategori peubah penjelas dengan kategori peubah respons. Kedua, dari setiap tabulasi silang yang diperoleh, susun subtabel berukuran $2 \times d$ yang mungkin, d adalah banyaknya kategori peubah respons. Dari tabel tersebut cari pasangan kategori peubah penjelas yang memiliki angka uji paling kecil. Jika tidak nyata, gabungkan kedua kategori itu menjadi satu kategori campuran. Jika banyaknya kategori hanya dua dan hasil ujinya tidak nyata maka variabel tersebut tidak perlu dilibatkan lagi dalam model. Ulangi tahap itu sehingga angka uji terkecil subtabel $2 \times d$ pasangan kategori (kategori campuran) peubah penjelas melampaui nilai kritis. Ketiga, untuk setiap kategori campuran yang berisi tiga atau lebih kategori asal, cari pemisahan biner yang memiliki angka uji paling besar. Jika ada buatlah pemisahan tersebut dan kembali ke tahap 2. Keempat, hitung taraf nyata untuk masing-masing tabulasi silang yang baru dan perhatikan diantaranya yang memiliki angka uji paling besar, sebut sebagai tabulasi dengan taraf nyata terbaik. Jika angka itu lebih besar dari nilai kritis, bagilah data menurut kategori tersebut. Kelima, kembali ke tahap 1 untuk melakukan pembagian berdasarkan peubah yang belum terpilih.

KETEPATAN DAN KESALAHAN KLASIFIKASI

Ketepatan klasifikasi (*correct classification*) terbagi atas dua, yaitu *specificity* dan *sensitivity*. *Specificity* atau ketepatan klasifikasi dalam menduga kejadian bahwa respons tidak memiliki kriteria yang ditentukan ($y=0$) dinyatakan sebagai persentase dari dugaan terkoreksi atau nilai dugaan yang sama dengan nilai amatan pada kategori nilai amatan $y=0$. Pengertian yang sama juga berlaku untuk mengevaluasi ketepatan klasifikasi dalam menduga kejadian bahwa respons memiliki kriteria yang ditentukan ($y=1$) atau disebut juga *sensitivity* sedangkan *total correct classification* yaitu ketepatan klasifikasi dalam menduga kejadian secara keseluruhan. Selain ketepatan klasifikasi dapat pula diketahui besarnya kesalahan klasifikasi (*misclassification rate*). Kesalahan klasifikasi dalam menduga kejadian respons terdiri atas kesalahan positif dan kesalahan negatif. Kesalahan positif dinyatakan sebagai persentase besarnya kesalahan ketika respons diduga memiliki kriteria yang ditentukan ($y=1$) tapi amatan sebenarnya bernilai $y=0$ dan sebaliknya kesalahan negatif dinyatakan sebagai persentase besarnya kesalahan ketika respons diduga tidak memiliki kriteria yang ditentukan ($y=0$) namun amatan sebenarnya bernilai $y=1$. *Total misclassification rate* diartikan sebagai kesalahan klasifikasi dalam menduga kejadian secara keseluruhan.

Data dan Metode

Data yang digunakan dalam penelitian ini adalah data sekunder dari hasil penelitian yang dilakukan di Rumah Sakit Umum Dr. Wahidin Sudirohusodo Makassar tahun 2004-2007 oleh Kamar (2004) dan Fitriani (2007) terhadap pasien yang dirawat di rumah sakit tersebut yang datang berobat ke poliklinik penyakit dalam yang dinyatakan menderita penyakit jantung koroner (PJK) dan yang tidak menderita PJK dengan jumlah pasien sebanyak 827 orang. Peubah yang diamati adalah pasien yang menderita penyakit jantung koroner ($y=1$) dan yang tidak ($y=0$) menurut kartu rekam medik pasien.

Peubah penjelasnya sebagai berikut.

Tabel 1 Pengkodean Peubah Penjelas Kategorik

Peubah penjelas	Kategori	D ₁	D ₂	D ₃	D ₄	Pembanding		
Hipertensi	Ya	1				✓		
	Tidak	0						
	Ya	1						
	Tidak	0						
Hiperkolesterol	Ya	1				✓		
	Tidak	0						
	Kurus	1					0	
Normal	0	0						
Umur	Gemuk	0	1			✓		
	< 40	0	0				0	0
	40-49	1	0				0	0
	50-59	0	1				0	0
	60-69	0	0				1	0
Jenis kelamin	> 69	0	0	0	1	✓		
	Laki-laki	1						
Perempuan	0							✓

Keterangan:

D : Peubah *dummy*

✓ : Sebagai pembanding

Metode Analisis

Metode analisis yang dilakukan, pertama mengumpulkan dan menyeleksi (*screening*) data. Dari data yang diperoleh, dilakukan pemeriksaan. Jika terdapat data yang tidak lengkap maka observasi tersebut tidak digunakan dalam pemodelan. Kedua, melakukan eksplorasi data. Dibuat tabulasi silang dengan cara menyajikan frekuensi dan persentase responden untuk setiap peubah penjelas yang diamati. Ketiga, melakukan analisis menggunakan metode regresi logistik dan metode CHAID terhadap semua peubah yang ada serta interpretasi hasil. Keempat, menentukan nilai klasifikasi dari metode regresi logistik dan metode CHAID. Kelima, membandingkan hasil klasifikasi antara metode regresi logistik dan metode CHAID. Pengolahan data untuk regresi logistik menggunakan program SAS 9.0 sedangkan untuk metode CHAID menggunakan program SPSS versi 15.

HASIL DAN PEMBAHASAN

Dari total 827 responden (Tabel 2) menunjukkan bahwa sekitar 52% adalah penderita PJK sedangkan yang bukan penderita PJK sekitar 48%. Penderita hipertensi lebih sedikit (40,87%) dibanding yang tidak menderita hipertensi (59,13%), sebaliknya penderita hiperkolesterol lebih banyak (59,01%) dibandingkan yang tidak hiperkolesterol (40,99%). Sebagian besar responden (70,62%) tidak menderita diabetes.

Persentase terbesar pertama untuk peubah IMT adalah kategori gemuk, yaitu sekitar 65,78%, persentase terbesar kedua adalah normal (32,53%), dan hanya 1,69% yang tergolong kurus. Dari segi umur, sebagian besar responden berumur sekitar 40-69 tahun, sedangkan yang berumur <40 tahun dan >69 tahun masing-masing hanya sekitar 9,07% dan 8,83%. Berdasarkan jenis kelamin responden, frekuensi antara laki-laki dan perempuan relatif seimbang.

Tabel 2 Frekuensi dan Persentase Responden

PEUBAH	Penyakit Jantung Koroner(PJK)		Jumlah	Persentase	
	Ya (1)	Tidak (0)			
Respons	430	397	827	100,00	
Hipertensi	Ya	174	164	338	40,87
	Tidak	256	233	489	59,13
Hiperkolesterol	Ya	327	161	488	59,01
	Tidak	103	236	339	40,99
Diabetes	Ya	139	104	243	29,38
	Tidak	291	293	584	70,62
Indeks Massa Tubuh (IMT)	Kurus	5	9	14	1,69
	Normal	105	164	269	32,53
	Gemuk	320	224	544	65,78
Umur	<40	21	54	75	9,07
	40-49	103	92	195	23,58
	50-59	154	117	271	32,77
	60-69	113	100	213	25,75
	>69	39	34	73	8,83
Jenis Kelamin	Laki-laki	203	211	414	50,06
	Perempuan	227	186	413	49,94

Hasil Metode Regresi Logistik

Model yang dihasilkan dengan melibatkan semua peubah penjelas dapat dilihat pada Tabel 3. Pengujian secara simultan menunjukkan bahwa model nyata pada $\alpha = 5\%$. Hal itu dapat dilihat dari statistik-G sebesar 147,400 dengan nilai-p sebesar 0,000. Berdasarkan uji Wald terlihat bahwa peubah hipertensi, IMT (kurus), dan jenis kelamin tidak nyata pada model sehingga akan dilakukan pereduksian terhadap peubah tersebut. Pereduksian model dilakukan dengan metode *backward elimination*. Berdasarkan Tabel 3 diperoleh tiga tahap pereduksian, yaitu tahap pertama dengan menghilangkan IMT (kurus) yang mempunyai nilai-p terbesar, tahap kedua dengan menghilangkan peubah hipertensi, dan tahap ketiga dengan menghilangkan peubah jenis kelamin. Tahap ketiga adalah tahap terakhir yang merupakan model terbaik yang diperoleh dari hasil pereduksian.

Tabel 3 Hasil Analisis Regresi Logistik Model Penuh

Peubah Penjelas	β_i	SE	Wald	Nilai-p	Rasio Odds	SK 95% untuk Rasio Odds	
Intersep	-2,091	0,330	-6,336	0,000	0,000	-	-
Hipertensi	-0,163	0,159	-1,025	0,306	0,850	0,623	1,160
Hiperkolesterol	1,380	0,159	8,679	0,000*	3,975	2,910	5,430
Diabetes	0,461	0,171	2,696	0,007*	1,586	1,134	2,216
IMT(Gemuk)	0,780	0,168	4,643	0,000*	2,182	1,570	3,033
IMT(Kurus)	-0,014	0,604	-0,023	0,982	0,986	0,302	3,219
Umur(>69)	1,101	0,387	2,845	0,004*	3,007	1,409	6,416
Umur(60-69)	0,953	0,325	2,932	0,003*	2,593	1,371	4,905
Umur(50-59)	0,995	0,315	3,159	0,002*	2,703	1,457	5,016
Umur(40-49)	0,897	0,322	2,786	0,005*	2,453	1,305	4,611
Jenis Kelamin	-0,222	0,154	-1,442	0,150	0,801	0,593	1,084

Statistik-G = 147,400 * Nyata pada $\alpha = 5\%$
 Nilai-p = 0,000

Tabel 4 Hasil Analisis Regresi Logistik Model Backward Elimination (Tahap III)

Peubah Penjelas	β_i	SE	Wald	Nilai-p	Rasio Odds	SK 95% untuk Rasio Odds	
Intersep	-2,203	0,323	-6,820	0,000	0,000	-	-
Hiperkolesterol	1,392	0,158	8,755	0,000*	4,024	2,949	5,490
Diabetes	0,435	0,170	2,559	0,011*	1,544	1,106	2,156
IMT(Gemuk)	0,781	0,165	4,733	0,000*	2,183	1,580	3,017
Umur(>69)	1,023	0,382	2,678	0,007*	2,782	1,315	5,885
Umur(60-69)	0,860	0,319	2,696	0,007*	2,364	1,265	4,418
Umur(50-59)	0,929	0,312	2,978	0,003*	2,531	1,373	4,666
Umur(40-49)	0,839	0,319	2,630	0,009*	2,314	1,238	4,327

Statistik-G = 144,142 * Nyata pada $\alpha = 5\%$
 Nilai-p = 0,000

Dari hasil model reduksi (Tabel 4) diperoleh nilai statistik-G sebesar 144,142 dengan nilai-p sebesar 0,000 yang berarti secara simultan menunjukkan bahwa model nyata pada $\alpha = 5\%$. Hasil uji Wald juga menunjukkan bahwa peubah PJK dipengaruhi oleh faktor hiperkolesterol, diabetes, IMT(gemuk), Umur(>69), Umur(60-69), Umur(50-59) dan Umur(40-49) pada $\alpha = 5\%$.

Berdasarkan nilai selang kepercayaan 95% bagi rasio odds menunjukkan bahwa penderita hiperkolesterol dibandingkan yang tidak hiperkolesterol berisiko paling rendah 2,949 kali dan paling tinggi 5,49 kali untuk menderita PJK. Penderita diabetes berisiko antara 1,106 kali sampai 2,156 kali

untuk menderita PJK dibandingkan yang bukan penderita diabetes. Untuk pasien yang tergolong gemuk, risiko terkena PJK dibandingkan yang tidak gemuk sebesar 1,58 kali sampai 3,017 kali. Pasien yang berumur >69 tahun berisiko antara 1,315 kali sampai 5,885 kali untuk terkena PJK, sedangkan yang berumur 60-69 tahun berisiko antara 1,265 kali sampai 4,418 kali. Untuk umur 50-59 tahun berisiko paling rendah 1,373 kali dan paling tinggi 4,666 kali, dan pasien yang berumur 40-49 tahun berisiko antara 1,238 kali sampai 4,327 kali untuk menderita PJK. Semua kategori umur dibandingkan dengan umur <40 tahun.

Dugaan Peluang Terkena PJK

Untuk menduga peluang seseorang akan terkena PJK, diperlukan matriks kovarian dugaan. Setelah dianalisis dengan persamaan (2), persamaan (4), persamaan (5), dan persamaan (6) maka diperoleh nilai selang kepercayaan (SK) 95% bagi $\pi(\mathbf{x})$. SK tersebut menyatakan nilai dugaan peluang terkena PJK untuk 15 kombinasi kategori berdasarkan peubah penjas yang nyata (Tabel 5).

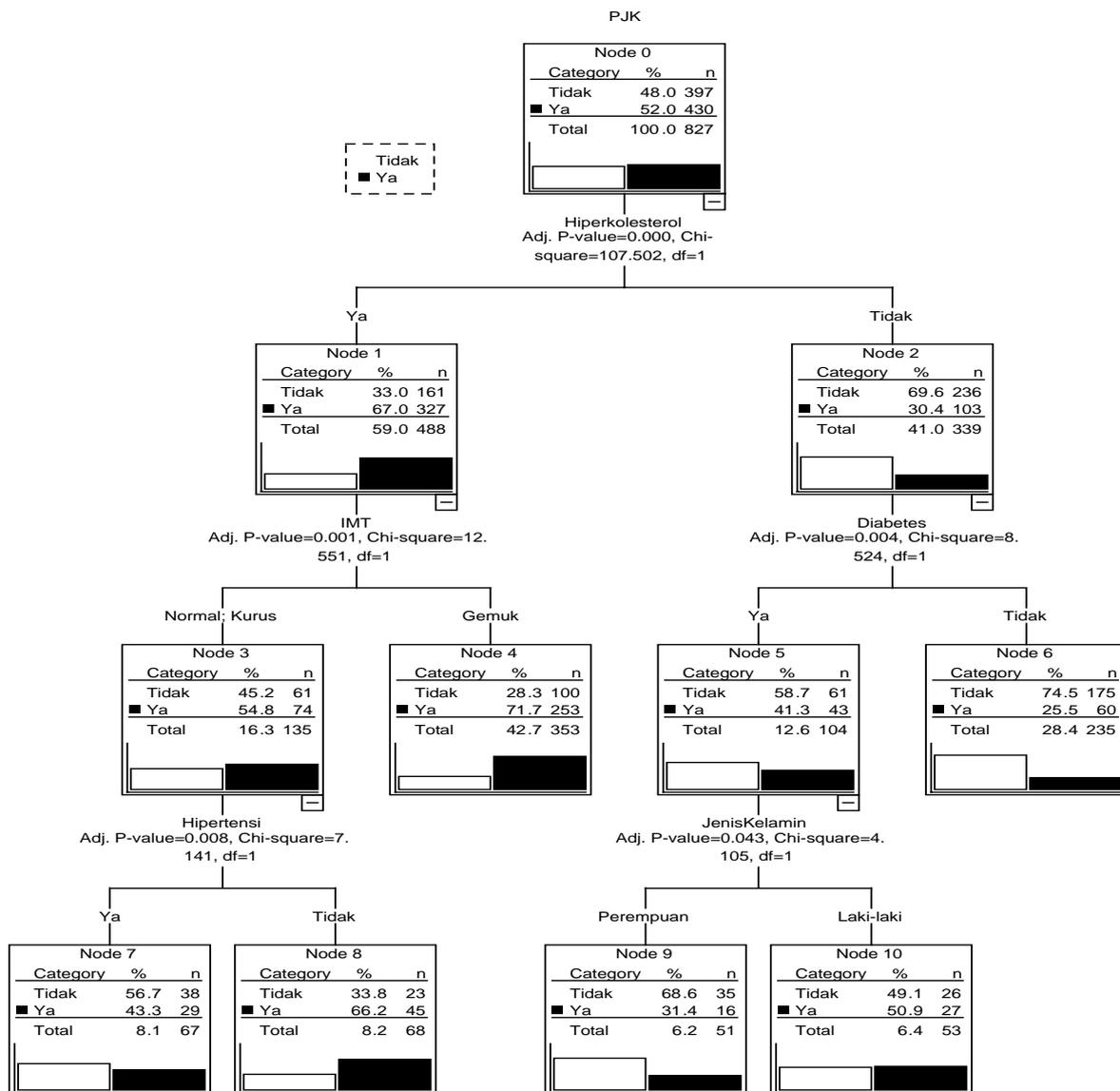
Berdasarkan kasus yang terjadi di RSUD Dr. Wahidin Sudirohusodo Makassar dapat diduga bahwa seseorang yang menderita hiperkolesterol, diabetes, berbadan gemuk, dan berusia di atas 69 tahun berpeluang antara 69,64% sampai 88,35% untuk menderita PJK sedangkan yang tidak menderita hiperkolesterol dan tidak diabetes dengan berat badan yang tergolong kurus atau normal dan mempunyai umur di bawah 40 tahun akan berpeluang sangat kecil untuk menderita PJK, yaitu antara 5,54% sampai 17,22%. Untuk kombinasi kategori yang lain, dapat dilihat pada Tabel 5.

Tabel 5 Dugaan Peluang Terkena PJK Berdasarkan Peubah Penjas yang Nyata

No	Hiperkolesterol	Diabetes	IMT	Umur	$\hat{\pi}(\mathbf{x})$	SK 95% bagi $\pi(\mathbf{x})$	
						Bawah	Atas
1	Ya	Tidak	Gemuk	60-69	0,6964	0,6183	0,7646
2	Ya	Tidak	Tidak gemuk	<40	0,3077	0,1873	0,4615
3	Ya	Ya	Tidak gemuk	<40	0,4071	0,2559	0,5782
4	Ya	Ya	Tidak gemuk	40-49	0,6137	0,4987	0,7174
5	Ya	Ya	Gemuk	>69	0,8066	0,6964	0,8835
6	Ya	Tidak	Gemuk	<40	0,4925	0,3536	0,6325
7	Ya	Ya	Gemuk	60-69	0,7799	0,7004	0,8430
8	Ya	Tidak	Tidak gemuk	>69	0,5528	0,4134	0,6843
9	Tidak	Tidak	Tidak gemuk	40-49	0,2036	0,1428	0,2818
10	Tidak	Ya	Gemuk	40-49	0,4631	0,3551	0,5747
11	Tidak	Ya	Tidak gemuk	60-69	0,2874	0,2075	0,3833
12	Tidak	Tidak	Tidak gemuk	60-69	0,2070	0,1464	0,2843
13	Tidak	Ya	Tidak gemuk	<40	0,1458	0,0814	0,2474
14	Tidak	Ya	Gemuk	>69	0,5090	0,3620	0,6544
15	Tidak	Tidak	Tidak gemuk	<40	0,0995	0,0554	0,1722

Hasil Metode CHAID

Dendogram hasil pemisahan analisis CHAID dapat dilihat pada Gambar 1 dengan nilai α yang digunakan sebesar 5%. Pada tahap pertama pemisahan CHAID, peubah yang mempunyai asosiasi paling kuat dengan PJK adalah peubah hiperkolesterol. Dari 827 responden, sekitar 59% merupakan kelompok penderita hiperkolesterol dan 67% di antaranya adalah penderita PJK sedangkan pada pasien yang tidak hiperkolesterol, hanya 30,4% yang menderita PJK.



Gambar 1 Dendrogram CHAID Status PJK

Pembahasan

Pada penderita hiperkolesterol, peubah IMT berasosiasi dengan PJK. Dari keseluruhan penderita hiperkolesterol, 72,3% termasuk kategori gemuk dan sisanya tergolong normal atau kurus. Pada penderita hiperkolesterol yang gemuk, sekitar 71,7% merupakan penderita PJK sedangkan pada penderita hiperkolesterol dengan berat badan normal atau kurus, penderita PJK sekitar 54,8%.

Selanjutnya, pada penderita hiperkolesterol dengan IMT yang tergolong normal atau kurus, peubah yang mempunyai asosiasi dengan PJK adalah hipertensi. Sekitar 43,3% dari total penderita hipertensi adalah penderita PJK sedangkan penderita PJK pada pasien yang tidak hipertensi sekitar 66,2%.

Pada pasien yang tidak menderita hiperkolesterol, peubah yang berasosiasi dengan PJK adalah diabetes. Pada penderita diabetes, sekitar 41,3% adalah penderita PJK sedangkan pada pasien yang tidak diabetes hanya sekitar 25,5%. Pada pasien yang bukan penderita hiperkolesterol namun menderita diabetes, jenis kelamin berasosiasi dengan PJK. Penderita PJK pada pasien yang diabetes dan berjenis kelamin perempuan adalah sekitar 31,4% sedangkan yang berjenis kelamin laki-laki sekitar 50,9%. Ringkasan dendrogram CHAID dapat dilihat pada Tabel 6.

Tabel 6 Segmentasi CHAID

Penderita PJK	Bukan penderita PJK
<ul style="list-style-type: none"> • Hiperkolesterol, gemuk (Node 4) • Hiperkolesterol, normal atau kurus, tidak hipertensi (Node 8) • Tidak hiperkolesterol, diabetes, laki-laki (Node 10) 	<ul style="list-style-type: none"> • Hiperkolesterol, normal atau kurus, hipertensi (Node 7) • Tidak hiperkolesterol, diabetes, perempuan (Node 9) • Tidak hiperkolesterol, tidak diabetes (Node 6)

Tabel 7 Klasifikasi Metode Regresi Logistik dan CHAID

		Dugaan		Total	(%) <i>Correct classification</i>
		Ya	Tidak		
Logistik	Ya	321	109	430	74,7
	Tidak	162	235	397	59,2
	Total	483	344	827	67,2
	(%) <i>Misclassification rate</i>	33,5	31,7	32,8	
CHAID	Ya	325	105	430	75,6
	Tidak	149	248	397	62,5
	Total	474	353	827	69,3
	(%) <i>Misclassification rate</i>	31,4	29,7	30,7	

Tabel 7 menunjukkan bahwa menggunakan nilai *cutpoint* sebesar 0,5 maka berdasarkan metode regresi logistik diperoleh nilai *sensitivity* sebesar 74,7% dan nilai *specificity* sebesar 59,2% dengan nilai kesalahan positif dan kesalahan negatif masing-masing sebesar 33,5% dan 31,7% sedangkan nilai *total correct classification* adalah sebesar 67,2%.

Pada metode CHAID, nilai *sensitivity* dan *specificity* masing-masing sebesar 75,6% dan 62,5% dengan nilai *total correct classification* sebesar 69,3%, sedangkan nilai kesalahan positif sebesar 31,4% dan nilai kesalahan negatifnya sebesar 29,7%. Nilai *total misclassification rate* untuk metode regresi logistik sebesar 32,8% sedangkan untuk metode CHAID sebesar 30,7%.

PENUTUP

Hasil yang diperoleh dari metode regresi logistik menunjukkan bahwa faktor yang mempengaruhi seseorang menderita PJK adalah hiperkolesterol, diabetes, IMT, dan umur. Berdasarkan nilai rasio odds, diketahui bahwa penderita hiperkolesterol lebih berisiko terkena PJK dibandingkan yang tidak hiperkolesterol, begitu pula pada penderita diabetes dibandingkan yang tidak diabetes. Pasien yang mempunyai IMT gemuk lebih berisiko terkena PJK dibandingkan yang tidak gemuk, dan pasien yang umurnya ≥ 40 tahun juga lebih berisiko terkena PJK dibandingkan yang berumur < 40 .

Dendogram CHAID menunjukkan bahwa peubah yang mempunyai asosiasi paling kuat dengan PJK adalah peubah hiperkolesterol. Berdasarkan hasil segmentasi CHAID, penderita PJK adalah penderita hiperkolesterol yang tergolong gemuk, juga pada penderita hiperkolesterol yang tergolong normal atau kurus dan tidak menderita hipertensi. Penderita PJK yang lain adalah bukan penderita hiperkolesterol namun menderita diabetes dan berjenis kelamin laki-laki. Berdasarkan metode regresi logistik, diperoleh nilai *sensitivity* sebesar 74,7% dan nilai *specificity* sebesar 59,2% sedangkan pada metode CHAID nilai *sensitivity* dan *specificity* masing-masing sebesar 75,6% dan 62,5%. Besarnya nilai *misclassification rate* untuk metode regresi logistik sebesar 32,8% dan metode CHAID sebesar 30,7%.

DAFTAR PUSTAKA

- Du Toit, SHC, Steyn AGW, Steyn RH, and Stumph RH. 1986. *Graphical Exploratory Data Analysis*. New York: Springer-Verlag.
- Fielding A. 1977. *Binary Segmentation: The Automatic Interaction Detector and Related Technique for Exploring Data Structure* (dalam O'Muircheartaigh, C. A., dan C. Payne. 1977. *The Analysis of Survey Data Vol. I. Exploring Data Structure*). London, New York, Sidney, Toronto: John Wiley & Sons.
- Fitriani. 2007. "Faktor Risiko PJK di RSUP Dr. Wahidin Sudirohusodo Makassar." Tesis. Makassar: Program Pascasarjana, UNHAS.
- Gonzales MP. 2003. *A Model for Profiling Radio Station Listeners using Logistic Regression, CART and CHAID for a given data set*. University of the Philippines.
- Hosmer, DW and Lemeshow S. 2000. *Applied Logistic Regression*. New York: John Wiley & Sons.
- Kass, GV. 1980. *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *App. Statist* 29(2):119-127.
- Kamar, N. 2004. "Beberapa Faktor yang Berhubungan dengan Penyakit Jantung Koroner pada Pasien Rawat Inap di RSUP Dr. Wahidin Sudirohusodo Makassar." Skripsi. Makassar: Program Sarjana, UNHAS.
- Magidson, J. and Vermunt JK. 2006. *An Extension of CHAID Tree-based Segmentation Algoritma to Multiple Dependent Variables*. Departement of Methodology and Statistic, Tilburg University. Netherlands.

- Palilati, H. 2003. "Beberapa Faktor Risiko Penyakit Jantung Koroner pada Pasien Rawat Inap di RSUP Dr. Wahidin Sudirohusodo Makassar." Skripsi. Makassar: Program Sarjana, UNHAS.
- Rosnah. 1999. "Tinjauan Kebiasaan Hidup Sehari-hari Sebelum Penderita Dirawat di RSUP Dr. Wahidin Sudirohusodo Makassar." Skripsi Program Sarjana, UNHAS.
- Ture, M., Akturk Z, Kurt I, and Dagdeviren N. 2006. *The effect of Health Status, Nutrition and Some Other Factors on Low School Performance Using Induction Technique*. Trakya Universitesi Tip Fakultesi Dergisi.