

# INDANA

*by* Ppj Unhasy

---

**Submission date:** 08-Dec-2019 03:48PM

(UTC+0700)**Submission ID:** 1195923200

**File name:** Full\_coba\_Indana.pdf (335.35K)

**Word count:** 2683

**Character count:** 17090

# 1 ANALISIS FAKTOR PREDIKSI DIAGNOSIS TINGKAT KEPARAHAN PENYAKIT JANTUNG (*HEART DISEASE*) MENGGUNAKAN METODE STEPWISE BINARY LOGISTIC REGRESSION

Indana Lazulfa<sup>1\*</sup>, Reza Augusta J.F.<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari  
Jalan Irian Jaya No 55 Jombang 61471, Telp. (0321)861719  
Email : [indanazulfa@gmail.com](mailto:indanazulfa@gmail.com)

\*Korespondensi penulis

## Abstrak

Penyakit jantung atau "*heart disease*", yang juga dikenal dengan istilah penyakit kardiovaskuler adalah berbagai kondisi yang terjadi penutupan dan pesumbatan pembuluh darah yang dapat menyebabkan serangan jantung, "nyeri dada" (*angina*), atau stroke. Penyakit jantung terjadi kepada siapa saja dalam segala umur, jenis kelamin, pekerjaan, dan lifestyle. Bahkan lebih parahnya, penyakit jantung belum bisa disembuhkan dengan mudah. Oleh karena itu, pengobatan dan pengawasan intensif dibutuhkan selama masa hidup. Apabila dalam rentang pengobatan ini gagal, penderita harus menjalani rangkaian operasi yang relatif mahal dan rumit. Telah disebutkan oleh WHO bahwa penyakit jantung ini penyakit pembunuh orang nomor satu di dunia yang telah merenggut relatif banyak nyawa di berbagai belahan dunia. Tingkat kematian yang diakibatkan *heart disease* ini cukup tinggi sekitar 12,8%. Pengurangan factor risiko menjadi focus utama untuk upaya pencegahan. Peran Teknologi Informasi dapat diupayakan dengan metode penggalian data ("*data mining*") untuk mempersingkat waktu dan juga keakurasian serta pemilihan factor pendeteksi dini dari penyakit jantung. Metode "stepwise binary logistic regression" mempunyai keunggulan untuk meningkatkan dan mengurangi "variabel independent" menyesuaikan pada "tingkat signifikansi" dari model yang terbentuk. Berdasarkan hasil analisis, lima variabel dengan proporsi tertinggi adalah tipe "nyeri dada" (*cp*), "kolesterol" (*chol*), "depresi" (*oldpeak*), "jumlah arteri" (*ca*), "tingkat kerusakan/defect" (*thal*). Sehingga dengan demikian "akurasi dan kecepatan" proses dari diagnosa tingkat keparahan "penyakit jantung" dapat diketahui melalui metode ini.

**Kata kunci:** model prediksi, data mining, penyakit jantung, regresi logistik biner

## Abstract

*Heart disease*, we also known as *cardiovascular*, a condition in which blocking of blood vessel(s). It can cause "heart attacks", chest pain, or stroke. Today, every people any age, gender, occupation, and lifestyle can be occurred by this disease. Besides, "heart disease" can't be cured. The careful treatment and care are required to this condition throughout life. An expensive and complex surgery is needed if those treatment fails. Based WHO information late years, it said that the most killer disease in world is "heart disease". Many people lives in different parts of the world had been threaten. The death rate from heart disease is quite high around 12,8 percents. Reducing risk factors is one of prevented step. The role of Information Technology can be realized with "data mining technics" to shorten the accuration rate and running time. The "stepwise binary logistic regression method" has the advantage of add and subtract many "independent variables" according to the level of significance of the formed model. Based on the result analysis, there are five variables with the highest proportion that should be watchful. Chest pain type (*cp*), "cholesterol" (*chol*), "depression" (*oldpeak*), number of "major vessel" (*ca*) and rate of defect (*thal*). Through this method and model, we can know the accuracy and processing speed of the diagnosis of the severity of "heart disease".

**Keyword:** prediction model, data mining, "heart disease", binary logistic regression

## PENDAHULUAN

Penyakit jantung atau *heart disease*, dikenal dengan istilah penyakit kardiovaskuler adalah situasi/kejadian penutupan dan pesumbatan pembuluh darah yang mengakibatkan serangan jantung, "nyeri dada" (*angina*), atau stroke. Siapapun orangnya di segala usia, jenis kelamin, pekerjaan, dan gaya hidup

dapat terjangkau penyakit ini. Di sisi lain, penyakit jantung ini tidak bisa disembuhkan secara total. Pengobatan dan pengawasan hati-hati dibutuhkan dalam kondisi seperti ini. Serangkaian operasi bedah yang cukup mahal dan kompleks harus dijalani apabila pengobatan ini gagal. Telah disebutkan oleh WHO bahwa “penyakit jantung” merupakan penyakit pembunuh nomor satu. Serta telah merenggut banyak nyawa di berbagai belahan dunia. Tingkat kematian akibat penyakit ini cukup tinggi sekitar 12.8 persen.

Menurut Asosiasi Jantung Amerika (AHA), satu dari tiga pria mengidap penyakit jantung serta lebih dari 48 persen pria meninggal karena disease jantung. Kaum pria lebih berisiko tinggi terkena penyakit jantung dibandingkan kaum wanita karena lebih dari 60 persen pria di Indonesia merupakan perokok aktif. Aktivitas merokok tersebut menjadi salah satu faktor pemicu penyumbatan pembuluh darah, sebagai pertanda awal jenis penyakit jantung tertentu. Diantara banyaknya penyakit jantung yang biasa dikenal antara lain gagal jantung, “arteri koroner”, “aritmia”, “angina”, dan berbagai infeksi, serta kelainan lahir terkait jantung.

Penyakit jantung yang dalam istilah medis juga disebut “jantung koroner” adalah situasi/kejadian pembuluh darah/arteri mengalami suatu malfungsi atau kerusakan (*defect*). Penyebab penyakit ini salah satunya akibat dari tumpukan kolesterol pada pembuluh darah serta proses peradangan. Pada tahun 2014 diperkirakan angka kematian akibat penyakit jantung koroner di Asia Tenggara mencapai 1,8 juta oleh Federasi Jantung Dunia. Mayoritas penderita penyakit ini di Indonesia tercatat berusia 55-64 tahun. Angka kematian juga cukup tinggi, yakni sekitar 45 persen. Oleh karena penyakit jantung ini memiliki karakteristik khusus yakni tidak dapat disembuhkan secara total, maka diperlukan keberlanjutan untuk pencegahan, pengobatan, dan pemantauan berkala sepanjang hidup. Dengan demikian penyakit jantung dapat dicegah tingkat keparahannya apabila terpantau sejak dini.

Bidang Teknologi Informasi ikut berperan dalam upaya pendeteksian dini salah satunya yaitu sebuah aplikasi Spot it Yourself (Nawawi, 2013). Aplikasi ini memberi informasi mengenai cara untuk mendeteksi kanker payudara sejak dini. Teknologi “penggalian data” untuk mempersingkat waktu dan faktor pendeteksian dini penyakit jantung juga merupakan bentuk peran TI lainnya. Untuk mengakses secara bebas/free data terkait faktor-faktor tingkat keparahan penyakit jantung dapat mengunjungi situs UCI Machine Learning. Dengan adanya data-data ini, maka peluang riset untuk akurasi dan deteksi dini masih terbuka sangat luas.

Dalam hal pemilihan variable yang dilakukan untuk preprocessing ada beberapa teknik regresi dalam statistika yang dapat digunakan. Studi ini menggunakan data “Heart Disease Cleveland” dengan regresi. Teknik regresi ini mengambil teknik “stepwise regression” yang dapat berfungsi mengenali variabel-variabel independen yang akan mempengaruhi akurasi dari variabel dependen. Hal tersebut termasuk dalam kelebihan “stepwise regression”.

Berdasarkan data yang tersedia di UCI repository, terdapat banyak variable yang nantinya akan dikenali sebagai variable independent yang lebih dari satu. Sehingga dengan alasan ini digunakan teknik regresi berganda. Factor penting lain yaitu karakteristik dari variabel dependen. Ada beberapa teknik “penggalian data klasifikasi” yang memiliki klas kategori. Kelas kategori/ kategorikkal disini maksudnya hanya terdapat dua kemungkinan, seperti ya – tidak, sukses – gagal, sakit – sehat, normal – cacat, dll. Kasus variable dua kelas juga dimiliki oleh Data “Heart Disease Cleveland”. Kelas yang dimaksud dalam data adalah tipe low risk dengan parameter diameter pesumbatan pembuluh darah kurang dari 50persen. Dan tipe high risk dengan parameter diameter pesumbatan pembuluh lebih dari 50persen. Jadi pada studi kali ini menggunakan tipe regresi jenis binary.

#### DATA “HEART DISEASE CLEVELAND”

Data yang didapatkan dari situs UCI repository adalah data Cleveland. Data tersebut terdiri dari 76 atribut lalu direduksi menjadi 11 atribut. Data penyakit jantung berasal dari “Cleveland Clinic Foundation” yang diadopsi oleh “Hungarian Institute of Cardiology” di kota Budapest. Berdasarkan 11 atribut penentu level keparahan penyakit jantung tersebut di level high risk atau low risk. Di samping sebelas atribut penentu/indikator penyakit, di dalam data ini ada juga nomor identifikasi pasien dan hasil diagnosa kanker. Kesebelas atribut tersebut adalah tipe “sakit dada” (“*chest pain*” type), tekanan darah “sistol” (“*resting*

blood" pressure), "kolesterol" (serum "cholesterol"), "gula darah puasa" ("fasting blood sugar"), hasil "elektrokardiografik" ("resting electrocardiographic" results), maksimum "denyut jantung" yang dapat dicapai ("maximum heart rate achieved"), "exercise induced angina" (exang), "depresi yang diinduksi oleh latihan relatif" ("oldpeak"), "slope of the peak exercise" ("slope"), "jumlah pembuluh darah utama" (number of "major vessel"), tipe "kerusakan pembuluh darah" ("thal"). Terdapat lebih kurang 303 baris datap pasien terdiagnosis penyakit jantung baik dari level high risk maupun low risk.

#### "STEPWISE BINARY LINEAR REGRESSION"

Kesebelas atribut diolah dan dianalisis dengan metode "regresi binary logistik". "Regresi logistik" adalah salah satu jenis regresi yang dipakai dalam memprediksi variabel-variabel dependent yang mempunyai ciri kategorik (dua kemungkinan). Data independennya berupa data numerikal. "Regresi logistik" dipakai untuk menentukan seberapa besar impact dari setiap variabel independent kepada variabel dependent. "Regresi Logistik Binary" atau binomial adalah jenis regresi yang dipakai di saat variabel dependent bersifat dikotomi atau dua kelas.

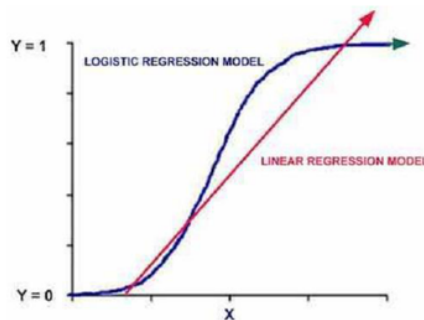
Proses dari "regresi logistik binary" adalah dengan memperkirakan atau mengestimasi kesamaan atau kemiripan terbesar dari data-data setelah perubahan variabel-variabel dependent menjadi "variabel logit". Variabel logit atau  $\text{logit}[\theta(x)]$  adalah bentuk natural log dari upper and lower bound apakah suatu variabel dependent dapat dipetakan ke suatu nilai tertentu atau tidak. Perhitungan  $\theta$  dan  $\text{logit}[\theta(x)]$  sebagai berikut

$$\theta = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Keterangan  $\alpha$  konstanta,  $\beta$  koefisien variabel independent (Walpole, 2012). Persamaan lain dari regresi logistik berikut

$$\text{logit} [\theta(x)] = \log \left[ \frac{\theta(x)}{1 - \theta(x)} \right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Regresi linear berbeda dengan "binary logistic" dapat dilihat melalui grafik yang muncul diantara keduanya seperti Gambar 1.



Gambar 1. Kurva model "regresi linear" dan model "regresi logistik"

Regresi "logistik stepwise" bisa masuk ke dalam berbagai regresi bermacam variabel seperti dalam regresi "logistik binary", multinomial maupun "regresi linear" sederhana. Kelebihan dari metode "stepwise" ini yaitu dapat menambah dan/atau mengurangi variabel independent serta menyesuaikan dengan tingkat signifikansi dari model yang terbentuk. Dan tentu saja untuk melakukan hal ini, memerlukan

sejumlah iterasi yang akan mengulangi beberapa proses dan analisa pada setiap model baru yang terbentuk karena penambahan dan/atau pengurangan variabel dependen.

Terdapat beberapa uji untuk menilai seberapa bagus persamaan regresi logistik binary yang terbentuk, diantaranya “maximum likelihood” dengan likelihood rasionya, “hosmer-lemeshow goodness of fit” dan “Wald”. Uji Wald merupakan salah satu cara menguji signifikansi dari sebagian variabel independent di suatu model statistik. Variabel dependent akan bernilai 0 atau 1. Jika suatu hasil uji “Wald” diperoleh angka signifikansinya kurang dari tingkat kepercayaan atau bernilai 0, maka variabel tersebut dimasukkan ke dalam model regresi logistik. Uji “Wald” menggunakan “uji Z”.

## METODE PENELITIAN

Langkah-langkah yang dilakukan secara garis besar sama dengan step by step dalam “data mining”. Terdapat tiga langkah utama seperti praproses, proses dan postproses. Langkah praproses terdiri dari: proses pelabelan variabel dan pengaturan skala pengukuran untuk masing-masing variabel. Dari proses pelabelan variabel diperoleh sebelas label yaitu seperti yang ditunjukkan oleh Tabel 1. Setelah itu, semua variabel terpilih dari proses sebelumnya diberikan skala pengukuran, yang mencakup skala numerikal baik interval maupun rasio akan masuk dalam skala ini. Sedangkan untuk skala nominal berasal dari identification number yang bersifat unik.

Tabel 1: Pelabelan, arti dan singkatan

Label	Makna label	Singkatan
Identification number	Nomor identifikasi	noId
Chest pain type	Tipe sakit dada	cp
Resting blood pressure (in mm Hg)	Tekanan darah batas bawah	restbps
Serum cholestorol (in mg/dl)	Kolesterol	chol
Fasting blood sugar	Gula darah puasa	fbs
Resting electrocardiographic results	Hasil elektrokardiografi	restecg
Maximum heart rate achieved	Maksimum denyut jantung	thalach
Exercise induced angina	Latihan yang menginduksi nyeri dada	exang
ST depression induced by relative to rest	Depresi yang diinduksi oleh latihan berat	oldpeak
The slope of the peak exercise ST segment	Kemiringan puncak ST segment	slope
Number of major vessels	Jumlah pembuluh darah utama	ca
Defect type	Tipe kerusakan pembuluh darah	thal
Diagnosa	Diagnosis penyakit jantung	num

Langkah kedua setelah praproses yaitu proses utama. Akan dilakukan “penggalan data” dengan tujuan mengenali diagnosis tingkat keparahan penyakit jantung berdasarkan data sebelumnya. Tingkatan keparahan penyakit ini terbagi menjadi dua kategori yakni risiko ringan (*low risk*) dan risiko tinggi (*high risk*). Untuk penyakit jantung yang masih di level bawah dan diameter pembuluh darah arteri yang mengalami penyumbatan kurang dari 50 persen termasuk ke dalam penyakit jantung ringan (*low risk*). Apabila terdapat kondisi sebaliknya dari ciri-ciri *low risk*, maka dikategorikan dalam *high risk*. Untuk yang



telah masuk ke dalam tingkatan high risk, dapat menyebabkan akibat bahaya lainnya seperti serangan jantung,, gagal jantung,,s troke dan kematian. Hal tersebut mungkin saja terjadi karena pembuluh darah yang terlampau tersumbat dan diameter yang jauh dari ukuran normal. Dalam proses utama ini dilakukan penggalian data dengan “regresi stepwise logistik binary”. Adanya faktor stepwise memungkinkan terjadi proses eliminasi variabel prediktor secara otomatis.

Setelah dilakukan langkah kedua dengan regresi binary stepwise, ada langkah terakhir atau postproses. Dari langkah terakhir ini diperoleh hasil analisis dari “regresi binary logistic”. Selain itu, di langkah postproses ini juga melihat variabel manakah yang secara konsisten atau terus menerus muncul di setiap persamaan. Hal ini menandakan variabel prediktor paling kuat adalah tidak tereliminasi hingga step akhir.

## HASIL DAN PEMBAHASAN

Seperti yang dijelaskan di bagian sebelumnya bahwa “model regresi logistik” berguna untuk mendeskripsikan dan menjelaskan hubungan antara variabel respon biner dengan satu atau beberapa variabel prediktor. Setelah langkah praproses, kemudian akan memasuki langkah kedua yakni estimasi model diagnosa penyakit jantung menggunakan “regresi binary logistik”. Pada langkah ini digunakan software SPSS dan matlab.

Untuk mengestimasi konstanta dari model prediksi tingkat risiko penyakit jantung yang terbentuk nantinya akan dipengaruhi oleh beberapa faktor. Dari hasil regresi binary stepwise didapatkan sebelas faktor sebagai variabel prediktor. Berdasarkan pelabelan pada Tabel 1 yaitu tipe “sakit dada” (“*chest pain type*”), tekanan darah “sistol” (“*resting blood pressure*”), “kolesterol” (serum “*cholesterol*”), “gula darah puasa” (“*fasting blood sugar*”), hasil “elektrokardiografik” (“*resting electrocardiographic results*”), maksimum denyut jantung yang dapat dicapai (“*maximum heart rate achieved*”), “exercise induced angina” (“*exang*”), depresi yang diinduksi oleh latihan relatif (“*oldpeak*”), “slope of the peak exercise” (*slope*), jumlah “pembuluh darah utama” (“*number of major vessel*”), tipe “kerusakan pembuluh darah” (“*thal*”). Sebelas variabel ini melakukan prediksi diagnosa tingkat risiko atau keparahan *heart disease*. Sehingga variabel respon di dalam “model regresi binary logistik” ini adalah variabel diagnosa (num).

Iteration	-2 Log likelihood	Coefficients	
		Constant	
Step 0	1	417.982	-.165
	2	417.982	-.165

a. Constant is included in the model.  
b. Initial -2 Log Likelihood: 417,982  
c. Estimation terminated at iteration number 2 because parameter estimates changed by less than ,001.

Gambar 2: Nilai dari  $L_0$

Diperoleh besar nilai dari  $L_0$  dan  $L_1$  yakni 417,982 dan 219,371. Sehingga

$$-2(L_0 - L_1) = 417,982 - 219,371 = 198,611$$

Dengan taraf signifikansi  $\alpha = 0,05$  dan  $d_f = k = 11$ , dengan k adalah jumlah variabel prediktore, diperoleh nilai chi kuadrat  $\chi^2_{(p)}$  dari tabel distribusi *chi-kuadrat* sebesar 19,6751. Karena nilai dari  $-2(L_0 - L_1) > \chi^2_{(p)}$  maka dapat disimpulkan bahwa secara bersama-sama (simultan), variabel prediktor berpengaruh signifikan terhadap variabel diagnosa (num).

		Iteration History <sup>a,b,c,d</sup>							
Iteration	Step	-2 Log likelihood	Coefficients						
			Constant	cp	trestbps	chol	fbs	restecg	thala
Step 1	1	243.581	-3.157	.321	.006	.000	-.252	.143	-.C
	2	222.372	-5.169	.462	.012	.001	-.432	.230	-.C
	3	219.470	-6.302	.535	.016	.001	-.542	.278	-.C
	4	219.371	-6.554	.552	.017	.001	-.567	.289	-.C
	5	219.371	-6.564	.553	.017	.001	-.568	.289	-.C
	6	219.371	-6.564	.553	.017	.001	-.568	.289	-.C

- a. Method: Enter
- b. Constant is included in the model.
- c. Initial -2 Log Likelihood: 417,982
- d. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Gambar 3. Nilai L<sub>1</sub>

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	219.371 <sup>a</sup>	.481	.643

- a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Gambar 4: Koefisien determinasi

Hasil dari tiruan ukuran R<sup>2</sup> pada multiple linear regression yang didasarkan pada teknik estimasi likelihood dapat dilihat pada Gambar 4 Cox & Snell's R Square. Tetapi memiliki nilai maksimum kurang dari 1 sehingga relatif sulit di interpretasikan. Opsi lainyadengan menggunakan Nagelkerke's R Square sebagai modifikasi dari koefisien Cox and Snell dan untuk membuat nilanya berada di selang 0 sampai 1. Dengan kata lain nilai R square pada Nagelkerke's R Square bisa mewakili nilai R<sup>2</sup> pada multiple linear regression.

Pada Gambar 4 model summary dan koefisien determinasi , diperoleh nilai Nagelkerke's R Square sebesar 0,643. Berarti hal ini mengatakan bahwa variabilitas variabel independen (cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, num secara simultan) dapat menjelaskan variabilitas variabel dependent diagnosa (num). Penjelasan yang diberikan oleh sebelas variabel independent tersebut adalah sebesar 64,3 persen. Sedangkan variabilitas variabel lain diluar kesebelas variabel independent yang diteliti menjelaskan diagnosa sisanya sebesar 35,7%.

Model "regresi binary logistic" diambil dengan memperhatikan tabel Variables in the Equation (Gambar 5). Dari tabel "variables in the equation" dapat dilihat nilai taksiran model regresi, dan didapatkan model "regresi binary logistik" sebagai berikut

$$\begin{aligned}
 \log[\pi(x)] &= \text{logit} \left[ \frac{\pi(x)}{1 - \pi(x)} \right] \\
 &= -6,564 + 0,553 \text{ cp} + 0,017 \text{ trestbps} + 0,001 \text{ chol} - 0,568 \text{ fbs} \\
 &\quad + 0,289 \text{ restecg} - 0,013 \text{ thalach} + 1,006 \text{ exang} + 0,451 \text{ oldpeak} \\
 &\quad + 0,427 \text{ slope} + 1,159 \text{ ca} + 0,431 \text{ thal}
 \end{aligned}$$

atau

$$\frac{\pi(x)}{1 - \pi(x)} = e^{(-6,564 + 0,553 \text{ cp} + 0,017 \text{ trestbps} + 0,001 \text{ chol} - 0,568 \text{ fbs} + 0,289 \text{ restecg} - 0,013 \text{ thalach} + 1,006 \text{ exang} + 0,451 \text{ oldpeak} + 0,427 \text{ slope} + 1,159 \text{ ca} + 0,431 \text{ thal}}$$

Dengan e adalah konstanta bernilai 2,718. Model persamaan “regresi logistik” diatas belum dapat secara langsung diartikan dari besar nilai koefisiennya seperti dalam regresi linier biasa atau sederhana. Namun diharuskan untuk melihat nilai Exp(B) atau nilai eksponen dari koefisien persamaan regresi yang terbentuk untuk dapat menginterpretasiikan .

Variables in the Equation							95,0% C.I. for EXP(B)		
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step	cp	.553	.187	8.752	1	.003	1.739	1.205	2.508
1 <sup>a</sup>	trestbps	.017	.010	2.830	1	.093	1.017	.997	1.037
	chol	.001	.003	.061	1	.805	1.001	.994	1.007
	fbs	-.568	.515	1.217	1	.270	.567	.207	1.554
	restecg	.289	.178	2.648	1	.104	1.335	.943	1.892
	thalach	-.013	.009	2.158	1	.142	.987	.970	1.004
	exang	1.006	.393	6.548	1	.010	2.736	1.266	5.913
	oldpeak	.451	.229	3.880	1	.049	1.570	1.002	2.458
	slope	.427	.341	1.571	1	.210	1.532	.786	2.987
	ca	1.159	.241	23.046	1	.000	3.186	1.985	5.114
	thal	.431	.089	23.352	1	.000	1.540	1.292	1.834
	Constant	-6.564	2.332	7.919	1	.005	.001		

a. Variable(s) entered on step 1: cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal.

**Gambar 5:** “Variabel independent”

5

Berdasar dari tolakukur nilai-nilai dari variabel prediktor yang sudah diketahui, dapat dilakukan prediksi diagnosa tingkat keparahan heart disease ini. Sebagai contoh, diketahui nilai-nilai variabel prediktor. Cp sebesar 2,1; trestbps sebesar 145; chol sebesar 221; fbs sebesar 0; restecg sebesar 2; thalach sebesar 108; exang sebesar 1; oldpeak 2,3; slope 1; ca 0; dan thal 3. Kemudian nilai-nilai tersebut disubstitusikan ke model yang sudah terbentuk :

$$\frac{\pi(x)}{1 - \pi(x)} = e^{(-6,564 + 0,553 (2,1) + 0,017 (145) + 0,001 (221) - 0,568 (0) + 0,289 (2) - 0,013 (108) + 1,006 (1) + 0,451 (2,3) + 0,427 (1) + 1,159 (0) + 0,431 (3)}$$

$$= e^{0,2206} = 1,247$$

1

Dari hasil perhitungan ke dalam persamaan tersebut, didapatkan hasil bernilai 1.247. Jika nilai setara dengan 1 seperti itu, maka dikatakan penderita tersebut dan dengan data besaran nilai variabel prediktor seperti diatas didiagnosis menderita penyakit jantung parah/ high risk. Berarti dikatakan bahwa diamati dari pembuluh darah utama atau “arteri” mengalami penyempitan dan penyumbatan di sebagian besar wilayah dan lebih dari 50 persen dari semua “arteri”. Apabila nilai perhitungan mempunyai nilai kurang dari 1, maka tingkat keparahan penyakitnya masih ringan/low risk. Dengan demikian, model regresi diatas dapat memprediksi sebuah diagnosis tingkat keparahan penyakit jantung.



1

## KESIMPULAN DAN SARAN

Berdasarkan hasil “model regresi binary logistik” yang terbentuk, ada sebanyak sebelas variabel yang paling berpengaruh terhadap penentuan tingkat keparahan penyakit jantung (*heart disease*). Yang bertindak sebagai variabel dependent adalah diagnosa(num). Sedangkan variabel prediktornya yaitu oldpeak, chol, restecg, exang, trestbps, fbs, thalach, cp, thal, slope, dan ca. Beragam tipe data dan kuantitas data yang dipunyai membuat model yang terbentuk menjadi lebih kompleks dengan hasil yang bisa berbeda-beda. Jumlah maksimum iterasi juga menjadi faktor penting penentu model terbaik.

Masalah terbuka untuk kasus penelitian dan topik yang mirip dengan penelitian ini adalah penentuan parameter inputan misalnya jumlah iterasi yang digunakan. Penelitian berikutnya bisa menggunakan hasil ini sebagai referensi untuk memakai metode “data mining” lainnya, pembentukan model regresi baru dan prediksi tingkat penyakit yang lainnya. Kurang tingginya tingkat akurasi juga masih menjadi masalah lanjutan untuk bisa dilakukan penelitian dengan parameter dan penyelesaian berbeda.

# INDANA

## ORIGINALITY REPORT

19%

SIMILARITY INDEX

44%

INTERNET SOURCES

1%

PUBLICATIONS

%

STUDENT PAPERS

## PRIMARY SOURCES

1

[ejournal.unhasy.ac.id](http://ejournal.unhasy.ac.id)

Internet Source

19%

2

[kuattanpaefeksampingdanketergantungan.blogspot.com](http://kuattanpaefeksampingdanketergantungan.blogspot.com)

Internet Source

1%

3

[www.ahlinyaobatambeien.apotek45.com](http://www.ahlinyaobatambeien.apotek45.com)

Internet Source

<1%

4

[media.neliti.com](http://media.neliti.com)

Internet Source

<1%

5

[menrvalab.com](http://menrvalab.com)

Internet Source

<1%

6

[blogkesehatan.net](http://blogkesehatan.net)

Internet Source

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography Off