

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

Indana Lazulfa

Program Studi Teknik Informatika, Fakultas Teknologi Informasi
Universitas Hasyim Asy'ari Tebuireng Jombang
indanazulf@gmail.com

Reza Augusta J.F.

Program Studi Sistem Informasi, Fakultas Teknologi Informasi
Universitas Hasyim Asy'ari Tebuireng Jombang
indanazulf@gmail.com

Abstrak

Penyakit jantung atau *heart disease*, yang juga dikenal dengan istilah penyakit kardiovaskuler adalah berbagai kondisi dimana terjadi penyempitan atau penyumbatan pembuluh darah yang dapat menyebabkan serangan jantung, nyeri dada (*angina*), atau stroke. Penyakit jantung dapat terjadi pada siapapun di segala usia, jenis kelamin, pekerjaan, dan gaya hidup. Selain itu, penyakit jantung tidak bisa disembuhkan. Kondisi ini membutuhkan pengobatan dan pemantauan hati-hati sepanjang masa hidup. Ketika pengobatan ini gagal, penderita diharuskan menjalani operasi bedah yang cukup mahal dan rumit. WHO menyebutkan penyakit jantung merupakan penyakit pembunuh orang di dunia nomor 1, yang tentu saja telah merenggut banyak nyawa di berbagai belahan dunia. Tingkat kematian akibat penyakit jantung ini cukup tinggi sekitar 12,8%. Hal ini dapat dicegah dengan mengurangi faktor risiko. Peran Teknologi Informasi dapat diwujudkan dengan teknik penggalian data (*data mining*) untuk mempersingkat waktu, akurasi dan pemilihan faktor pendeteksian dini penyakit jantung. Metode *stepwise binary logistic regression* memiliki keunggulan untuk menambah dan mengurangi variabel independen sesuai dengan tingkat signifikansi dari model yang terbentuk. Berdasarkan hasil analisis, lima variabel dengan bobot tertinggi yang harus lebih diwaspadai adalah tipe nyeri dada (*cp*), kolesterol (*chol*), depresi (*oldpeak*), jumlah arteri (*ca*), tingkat kerusakan/defect (*thal*). Sehingga akurasi dan kecepatan pemrosesan dari diagnosis tingkat keparahan penyakit jantung dapat diketahui melalui metode ini.

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

Abstract

Heart disease, also known as cardiovascular, is a condition in which narrowing or blocking of blood vessels can cause heart attacks, chest pain (angina), or stroke. Heart disease can occur to anyone at any age, gender, occupation, and lifestyle. In addition, heart disease can not be cured. This condition requires careful treatment and care throughout life. When this treatment fails, the patient is required to undergo a surgery that is quite expensive and complicated. WHO says heart disease is the most killer disease in world, that has claimed many lives in different parts of the world. The death rate from heart disease is quite high around 12.8%. This can be prevented by reducing risk factors. The role of Information Technology can be realized with data mining techniques to shorten the time, accuracy and selection of early detection factors of heart disease. The stepwise binary logistic regression method has the advantage of adding and subtracting independent variables according to the level of significance of the model that had been formed. Based on the analysis, the five variables with the highest weights that should be more watchful are chest pain type (cp), cholesterol (chol), depression (oldpeak), number of major vessel (ca) and rate of defect (thal). So the accuracy and processing speed of the diagnosis of the severity of heart disease can be known through this model.

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

PENDAHULUAN

Penyakit jantung Penyakit jantung atau *heart disease*, yang juga dikenal dengan istilah penyakit kardiovaskuler adalah berbagai kondisi dimana terjadi penyempitan atau penyumbatan pembuluh darah yang dapat menyebabkan serangan jantung, nyeri dada (angina), atau stroke. Penyakit jantung dapat terjadi pada siapapun di segala usia, jenis kelamin, pekerjaan, dan gaya hidup. Selain itu, penyakit jantung tidak bisa disembuhkan. Kondisi ini membutuhkan pengobatan dan pemantauan hati-hati sepanjang masa hidup. Ketika pengobatan ini gagal, penderita diharuskan menjalani operasi bedah yang cukup mahal dan rumit. WHO menyebutkan penyakit jantung merupakan penyakit pembunuh orang di dunia nomor 1, yang tentu saja telah merenggut banyak nyawa di berbagai belahan dunia. Tingkat kematian akibat penyakit jantung ini cukup tinggi sekitar 12,8%. Hal ini dapat dicegah dengan mengurangi faktor risiko.

Menurut Asosiasi Jantung Amerika (AHA), lebih dari 1 pada 3 pria mengidap penyakit jantung dan lebih dari 48% pria meninggal karena kondisi jantung. Pria berisiko tinggi terkena penyakit jantung. Lebih dari 60% pria di Indonesia merokok. Hal tersebut dapat menjadi salah satu faktor pemicu penyempitan pembuluh darah, sebagai pertanda awal jenis penyakit jantung tertentu. Penyakit jantung bisa termasuk gagal jantung, penyakit arteri koroner, aritmia, angina, dan berbagai infeksi, serta kelainan lahir yang terkait dengan jantung.

Penyakit jantung atau dalam istilah medis disebut penyakit jantung koroner adalah kondisi yang terjadi ketika pembuluh darah utama yang menyuplai darah ke jantung mengalami kerusakan (*defect*). Tumpukan kolesterol pada pembuluh darah serta proses peradangan diduga menjadi penyebab penyakit ini. Bahkan penyakit jantung koroner ini termasuk salah satu penyebab kematian tertinggi di dunia. Federasi Jantung Dunia memperkirakan angka kematian akibat penyakit jantung koroner di Asia Tenggara mencapai 1,8 juta kasus pada tahun 2014. Di Indonesia tercatat mayoritas penderita berusia 55-64 tahun. Angka kematian akibat penyakit jantung pun cukup tinggi, yakni sekitar 45 persen dari seluruh angka kematian di Indonesia. Oleh karena penyakit jantung ini memiliki karakteristik khusus yakni tidak dapat disembuhkan, maka diperlukan komitmen untuk mencegah, mengobati, dan memantau berkala sepanjang hidup. Sehingga penyakit jantung dapat dicegah tingkat keparahannya jika terpantau sejak dini.

Salah satu peran Teknologi Informasi dalam pendeteksian dini telah diimplementasikan melalui aplikasi Spot it Yourself (Nawawi, 2013). Aplikasi ini memberikan informasi yang bersifat eksplanatoris mengenai cara untuk mendeteksi kanker payudara sejak dini. Peran Teknologi Informasi lainnya juga dapat dirupakan dengan teknologi penggalian data untuk mempersingkat waktu dan faktor pendeteksian dini penyakit jantung. Data mengenai faktor-faktor diagnosis tingkat keparahan penyakit jantung dapat diakses secara luas dan bebas (tidak berbayar) pada situs UCI Machine Learning. Sehingga peluang riset untuk akurasi dan deteksi dini sangat luas.

Beberapa teknik regresi dalam statistika dapat digunakan untuk pemilihan variabel sebelum pemrosesan klasifikasi. Dalam penelitian ini akan digunakan data Heart Disease Cleveland dengan teknik regresi. Teknik regresi yang diambil disini adalah teknik stepwise regression yang dapat memilih dengan cermat mana saja variabel-variabel independen yang akan mempengaruhi akurasi dari variabel dependen. Ini merupakan kelebihan stepwise regression dibandingkan regresi biasa.

Sesuai dengan data yang disediakan oleh situs UCI, variabel independen dapat dipastikan lebih dari satu. Hal ini menjadi landasan utama untuk menggunakan regresi berganda. Hal yang perlu diperhatikan lagi yakni karakteristik dari variabel dependen. Hampir semua teknik penggalian data klasifikasi selalu memiliki kelas kategorikal dimana terdapat dua kemungkinan. Misalkan yes – no, rata – tidak rata, sakit – sehat, normal – cacat, dll. Pada kasus Heart Disease Cleveland memiliki karakter variabel dua kelas. Kelas tersebut adalah tipe low risk dengan diameter penyempitan pembuluh darah kurang dari 50%. Sedangkan tipe high risk dengan diameter penyempitan pembuluh lebih dari sama dengan 50%. Jadi pada penelitian akan digunakan tipe regresi jenis binary.

DATA HEART DISEASE CLEVELAND

Data yang digunakan dalam penelitian ini adalah data Penyakit Jantung Cleveland Database yang terdiri dari 11 atribut hasil reduksi dari 76 atribut. Data-data tersebut berasal dari lokasi Cleveland Clinic Foundation yang telah diadopsi oleh instansi Hungarian Institute of Cardiology di Budapest. Dalam data ini terdiri dari 11 atribut penentu apakah penyakit jantung tersebut parah (high risk) atau ringan (low risk). Selain 11 atribut, data ini juga dilengkapi dengan nomor identifikasi pasien dan hasil diagnosis kanker. Sebelas 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*). Data ini memiliki 303 baris untuk diagnosis penyakit jantung high risk dan low risk.

STEPWISE BINARY LINEAR REGRESSION

Kesebelas atribut tersebut akan diproses menggunakan regresi binary logistik. Regresi logistik adalah jenis regresi yang dapat digunakan untuk memprediksikan variabel dependen yang memiliki sifat kategorikal. Sedangkan untuk data independen berupa numerik. Regresi logistik dapat digunakan untuk menentukan besarnya dampak dari setiap variabel independen kepada variabel dependen. Regresi Logistik Binary atau binomial adalah bentuk regresi yang digunakan ketika variabel dependen bersifat dikotomi atau dua kelas. Bentuk regresi yang membolehkan lebih dari dua kelas variabel adalah Regresi Logistik Multinomial.

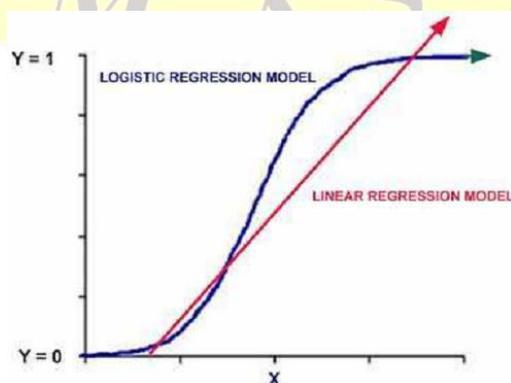
Proses dari regresi logistik binary adalah dengan cara menggunakan estimasi kesamaan maksimal setelah mengubah variabel-variabel dependen menjadi variabel logit. Variabel logit ($\text{logit}[\theta(x)]$) adalah natural log dari batas nilai apakah suatu variabel dependen dapat ditransformasi menjadi suatu nilai tertentu atau tidak. Persamaan θ dan $\text{logit}[\theta(x)]$ tampak dalam persamaan berikut

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

dengan α adalah konstanta dan β adalah koefisien dari variabel independen (Walpole, 2012). Sedangkan persamaan alternatif dari regresi logistik adalah sebagai berikut

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

Gambar 1 menunjukkan perbedaan antara regresi linear dengan regresi binary logistic.



Gambar 1. Perbedaan kurva model regresi linear dan model regresi logistik

Regresi logistik stepwise tersedia baik dalam regresi logistik binary, multinomial maupun regresi linear biasa. Metode stepwise memiliki keunggulan untuk menambah dan mengurangi variabel independen sesuai dengan tingkat signifikansi dari model yang terbentuk. Tentu saja, untuk melakukan hal ini diperlukan sejumlah iterasi yang akan memproses dan menganalisis setiap model baru yang terbentuk karena penambahan atau pengurangan variabel dependen.

Untuk menilai seberapa bagus persamaan regresi logistik binary yang terbentuk, stepwise memiliki beberapa cara untuk menguji. Uji tersebut antara lain maximum likelihood yang dinilai dengan likelihood ratio, hosmer-lemeshow goodness of fit dan Wald. Uji Wald adalah salah satu cara untuk menguji signifikansi dari sebagian variabel independen pada suatu model statistika. Dalam hal regresi logistik binary, variabel dependen akan bernilai 0 atau 1. Apabila suatu hasil uji test Wald signifikansinya kurang dari tingkat kepercayaan atau bernilai 0 maka variabel tersebut akan dimasukkan dalam model regresi logistik. Uji Wald menggunakan uji Z statistik.

METODE PENELITIAN

Langkah-langkah yang dilakukan untuk menyelesaikan penelitian ini untuk mengikuti langkah-langkah dari penggalian data secara umum. Langkah tersebut memiliki tiga langkah utama seperti praproses, proses dan postproses. Langkah praproses terdiri dari dua langkah pendukung, yaitu labelisasi variabel dan mengatur skala pengukuran setiap variabel. Labelisasi variabel menghasilkan sebelas label yaitu seperti yang ditunjukkan oleh Tabel 1. Setelah itu semua variabel akan diberikan skala pengukuran. Sebelas variabel akan diberi skala Scale yang mencakup skala numerikal baik interval maupun rasio akan masuk dalam skala ini. Sedangkan nomor identification number yang bersifat unik dan hasil diagnosis diberikan skala nominal.

Tabel 1. Labelisasi, makna label 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	trestbps
Serum cholestoral (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 segmen	slope
Number of major vessels	Jumlah pembuluh darah utama	ca
Defect type	Tipe kerusakan pembuluh darah	thal
Diagnosa	Diagnosis penyakit jantung	num

Langkah berikutnya yaitu langkah proses utama yaitu penggalan data dengan tujuan untuk diagnosis tingkat keparahan penyakit jantung. Tingkat keparahan terdiri dari 2 skala yaitu ringan dan parah. Ringan (low risk) diartikan sebagai tingkat penyakit jantung yang masih dalam level bawah dan diameter dari pembuluh darah arteri yang menyempit kurang dari 50%. Sedangkan untuk tingkatan parah (high risk) diartikan sebagai diagnosis penyakit jantung yang telah parah dan dapat menyebabkan bahaya yang lainnya seperti serangan jantung, gagal jantung, stroke bahkan kematian. Hal ini dikarenakan diameter pembuluh darah telah jauh menyempit dari kondisi normal sekitar lebih dari 50%. Proses utama memiliki dua proses pendukung yaitu penggalan data dengan regresi stepwise logistik binary. Adanya faktor stepwise memungkinkan terjadi proses eliminasi variabel prediktor secara otomatis.

Langkah terakhir yaitu postproses. Langkah ini bertujuan untuk melihat hasil analisis dari regresi binary logistic. Langkah postproses juga melihat variabel manakah yang konsisten muncul di setiap persamaan regresi stepwise binary logistic.

HASIL DAN PEMBAHASAN

Seperti yang dijelaskan pada bagian tinjauan pustaka regresi logistik, bahwa model regresi logistik digunakan untuk menggambarkan hubungan antara variabel respon biner dengan satu atau beberapa variabel prediktor. Setelah langkah praproses, maka akan diestimasi model diagnosa penyakit jantung menggunakan regresi binary logistik dengan menggunakan software SPSS dan matlab.

Faktor-faktor yang mempengaruhi model prediksi tingkat risiko penyakit jantung nantinya akan digunakan untuk mengestimasi konstanta dari model persamaan yang terbentuk. Sehingga untuk ke depannya dapat dihitung nilai total diagnosa dengan menginputkan beberapa nilai variabel dependen. Faktor-faktor tersebut telah diidentifikasi sebanyak sebelas faktor yang kemudian akan menjadi variabel prediktor, antara lain seperti yang dilabelkan 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*). Kesebelas variabel ini akan memprediksi diagnosa tingkat risiko atau keparahan penyakit jantung, sehingga variabel respon di dalam model regresi binary logistik ini adalah variabel diagnosa (*num*).

		-2 Log likelihood	Coefficients
Iteration			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. Besar L_0

Dari Gambar 2 dan Gambar 3, didapatkan besar nilai dari L_0 dan L_1 yakni masing-masing 417,982 dan 219,371. Sehingga didapatkan hasil

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

Analisis Faktor Prediksi Diagnosis Tingkat Keparahan Penyakit Jantung (*Heart Disease*) Menggunakan Metode Stepwise Binary Logistic Regression

Jika mengambil taraf signifikansi $\alpha = 0,05$ dan degree of freedom (df) = k = 11, dengan k adalah jumlah variabel prediktor, didapatkan nilai $\chi^2_{(p)}$ 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^{a,b,c,d}

Iteration	-2 Log likelihood	Coefficients						
		Constant	cp	trestbps	chol	fbs	restecg	thala
Step 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. Besar L_1

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	219.371 ^a	.481	.643

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

Gambar 4. Model summary dan koefisien determinasi

Pada Gambar 4 disajikan *Cox & Snell's R Square* yang merupakan ukuran tiruan ukuran R^2 pada multiple linear regression yang didasarkan pada teknik estimasi likelihood dengan nilai maksimum kurang dari 1 sehingga sulit untuk diinterpretasikan. *Nagelkerke's R Square* merupakan modifikasi dari koefisien *Cox and Snell* untuk memastikan bahwa nilainya bervariasi dari 0 sampai 1. Nilai Nagelkerke's R Square dapat diinterpretasikan seperti nilai R^2 pada multiple linear regression. Melalui Gambar 4 Model summary diatas, didapatkan nilai *Nagelkerke's R Square* sebesar 0,643. Hal ini mengindikasikan bahwa variabilitas variabel dependen diagnosa (num) dapat dijelaskan oleh variabilitas variabel independen (cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, num secara simultan) adalah sebesar 64,3 %. Sedangkan sisanya sebesar 35,7% dijelaskan oleh variabilitas variabel lain diluar kesebelas variabel independen yang diteliti tersebut.

Melalui tabel *Variables in the Equation* pada Gambar 5 di bawah ini dapat terlihat nilai taksiran koefisien regresi modelnya, sehingga didapatkan model regresi binary logistik sebagai berikut

Analisis Faktor Prediksi Diagnosis Tingkat Keparahan Penyakit Jantung (*Heart Disease*) Menggunakan Metode Stepwise Binary Logistic Regression

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

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 bilangan konstanta bernilai 2,71828. Hasil persamaan regresi logistik diatas tidak bisa secara langsung diinterpretasikan dari nilai koefisiennya seperti dalam regresi linier biasa. Interpretasi dapat dilakukan dengan melihat nilai Exp(B) atau nilai eksponen dari koefisien persamaan regresi yang terbentuk.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I. for EXP(B)	
							Lower	Upper
Step 1 ^a								
cp	.553	.187	8.752	1	.003	1.739	1.205	2.508
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 independen

Melalui persamaan model tersebut kita dapat melakukan prediksi diagnosa tingkat keparahan penyakit jantung (num) berdasarkan nilai-nilai variabel prediktor yang telah diketahui. Misalkan diketahui nilai variabel-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 dalam persamaan model sebagai berikut :

$$\begin{aligned} \frac{\pi(x)}{1 - \pi(x)} &= e^{(-6,564 + 0,553 (2,1) + 0,017 (145) + 0,001 (221) - 0,568 (0) + 0,289 (2) - 0,013 (108) + 1,006 (1) + 0,451 (2,3) + 0,427 (1) + 1,159 (0) + 0,431 (3))} \\ &= e^{0,2206} = 1,247 \end{aligned}$$

Karena hasilnya bernilai 1,247 setara dengan 1, maka penderita dengan data variabel prediktor seperti diatas didiagnosis menderita penyakit jantung parah. Hal tersebut bisa terlihat dari hasil diagnosa bernilai 1 yang berarti pembuluh darah arteri diameternya menyempit di sebagian besar wilayah, lebih dari 50 %. Jika nilainya kurang dari 1, maka diagnosa tingkat keparahan penyakit jantung masih ringan karena nilai dibawah 1 menunjukkan diameter pembuluh darah menyempit di sebagian kecil wilayah, kurang dari 50%. Sehingga dari model diatas dapat diprediksi sebuah diagnosa tingkat keparahan penyakit jantung seseorang dengan menginputkan nilai variabel-variabel prediktor.

KESIMPULAN DAN SARAN

Berdasarkan hasil persamaan regresi binary logistik yang dihasilkan, sebelas variabel yang paling berpengaruh terhadap akurasi penentuan tingkat keparahan penyakit jantung dalam model persamaan regresi dengan konstanta dan variabel prediktor. Diagnosa merupakan variabel dependen yang hasilnya tergantung pada variabel prediktornya (cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, dan thal). Kesimpulan dari penelitian ini akan didapatkan dari berbagai uji coba dan hasil terbaik dari masing-masing iterasi. Beragam tipe data membuat persamaan menjadi lebih kaya dan kompleks. Oleh karena itu, perlu dilakukan uji coba berulang kali untuk mendapatkan akurasi yang tertinggi.

Sebagai saran dari penelitian berikutnya yaitu dapat menggunakan hasil ini sebagai praproses dari algoritma data mining lainnya. Atau bisa digunakan untuk referensi pembuatan model regresi dan peramalan tingkat diagnosa penyakit yang lainnya. Hal ini bertujuan untuk meningkatkan akurasi dan kecepatan pemrosesan dari diagnosis tingkat keparahan penyakit jantung sehingga dapat mencegah lebih dini ataupun meminimalisir risiko karena keterlambatan diagnosa tingkat keparahan penyakit jantung.

DAFTAR PUSTAKA

- Brant, R. 1990. *Journal Biometrics. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression*. Vol. 46, pp.1171 – 1178
- Ghozali, Imam. 2011. *Aplikasi Multivariat dengan Program IBM SPSS 19*, Edisi 5. Semarang : Badan Penerbit Universitas Diponegoro
- Hilbe, J.M. 2014. *Journal International Encyclopedia of Statistical Science. Logistic Regression*. Pp. 755 – 758.
- Kyngas H. and Rissanen M. 2001. *Journal of Clinical Nursing. Blackwell Science,ltd*, Vol 10 p.212
- Nawawi, Qalbinur. *Deteksi Dini Kanker Payudara lewat Aplikasi Ponsel. Okehealth*. <http://health.okezone.com> (Cited : Oktober 5, 2017)
- Ralf, B., Grouven, U. 1998. *Journal Clin Epidemiol. Using Binary Logistic Regression Models for Ordinal Data with Non-proportional Odds*. Vol. 51, No 10, pp. 809 – 816.
- Vinarti, R.A., Anggraeni, W. 2014. *Jurnal Informatika. Identifikasi Faktor Prediksi Diagnosis Tingkat Keganasan Kanker Payudara Menggunakan Metode Regression*. Vol 12, No 2, pp. 70-76
- Walpole, R.E., Myers, R.H., Myers, S.L. (2012). *Probability and Statistics for Engineers and Scientists 9th edition*. Prentice Hall : Pearson
- Yamin, Sofyan, Kurniawan, H. 2014. *SPSS Complete Teknik Analisis Statistik Terlengkap dengan Software SPSS Edisi 2*. Jakarta : Salemba Infotek