

PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN METODE CART PADA FAKTOR-FAKTOR YANG MEMPENGARUHI PENYAKIT JANTUNG KORONER (Studi Kasus di RSUD Dr. M. Yunus Bengkulu)

Wina Ayu Lestari¹, Sigit Nugroho², Fachri Faisal²

¹Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu

²Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

Email: wina.yhrd@gmail.com

Abstrak— Penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi penyakit jantung koroner. Analisis yang dapat digunakan untuk membuat klasifikasi adalah metode regresi logistik biner dan metode CART (*Classification and regression trees*). Data yang digunakan pada penelitian ini berupa data sekunder yang diperoleh dari data rekam medis di Rumah Sakit M. Yunus Bengkulu pada Tahun 2014 dengan variabel respon berskala katagorik. Pengklasifikasian menggunakan regresi logistik biner menghasilkan ketepatan klasifikasi sebesar 84.2% dimana terdapat empat faktor yang berpengaruh yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Sedangkan dengan menggunakan metode CART terdapat tiga faktor yang mempengaruhi yaitu dislipidemia, diabetes melitus dan obesitas dengan ketepatan klasifikasi sebesar 95 %.

Kata Kunci : *PJK, Klasifikasi, Regresi Logistik Biner, CART*

I. PENDAHULUAN

Penyakit jantung (kardiovaskuler) merupakan penyakit yang sangat menakutkan dan mematikan. Jumlah penderita penyakit jantung di seluruh dunia selalu bertambah setiap tahunnya. Penyakit jantung koroner (PJK) merupakan jenis gangguan pada jantung yang paling sering ditemui dan merupakan penyebab utama kematian di negara maju dan berkembang termasuk Indonesia (Zakiyah dkk., 2012). Menurut Soeharto (2004), penyakit jantung koroner adalah kelainan pada jantung yang disebabkan oleh terjadinya penyempitan atau penyumbatan arteri yang mengalirkan darah ke otot jantung.

Melihat permasalahan tersebut, penulis tertarik untuk mengetahui hubungan antara penyakit jantung koroner dan faktor-faktor risikonya dengan klasifikasi. Pada penelitian ini penulis membatasi faktor – faktor risiko yang akan diteliti yaitu umur, jenis kelamin, dislipidemia, tekanan darah tinggi (*hipertensi*), penyakit diabetes melitus dan obesitas. Studi kasus pada penelitian ini adalah di RSUD Dr. M. Yunus Bengkulu. Hal ini dikarenakan RSUD Dr. M. Yunus merupakan salah satu rujukan tempat berobat di Bengkulu.

Analisis yang dapat digunakan untuk membuat klasifikasi adalah metode regresi logistik biner dan CART (*Classification and regression trees*). Kedua analisis tersebut akan menghasilkan ketepatan klasifikasi tersendiri dan selanjutnya akan dilihat analisis mana yang memberikan ketepatan klasifikasi tertinggi (Margasari, 2014).

II. TINJAUAN PUSTAKA

Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terdapat dua atau lebih variabel prediktor yang digunakan dalam regresi saling berkorelasi. Salah satu cara untuk mendeteksi multikolinieritas adalah dengan melihat besarnya nilai *Variance Inflation Factor* (VIF) untuk masing-masing variabel prediktor

$$VIF_j = \frac{1}{1-R_j^2}$$

dimana :

R_j^2 : Koefisien determinasi berganda dari variabel prediktor X_j dengan seluruh variabel prediktor X lainnya.

Jika nilai VIF_j lebih dari atau sama dengan 10, maka ini menunjukkan data mengalami masalah multikolinieritas (Neter

dkk., 1990). Salah satu cara untuk mengatasi multikolinieritas adalah dengan mengeluarkan satu atau beberapa variabel prediktor yang terindikasi memiliki hubungan dengan variabel prediktor lainnya (Gujarati, 2010).

Analisis Regresi Logistik Biner

Regresi logistik biner merupakan salah satu metode regresi logistik yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat *biner* atau dikotomis dengan variabel prediktor (x) yang bersifat dikotomis atau polikotomis (Hosmer dan Lemeshow, 2000). Variabel respon (y) terdiri dari 2 kategori yaitu 0 dan 1, sehingga variabel respon akan mengikuti distribusi Bernoulli untuk setiap observasi tunggal dengan fungsi probabilitas sebagai berikut,

$$f(y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad ; y_i = 0, 1$$

Menurut Hosmer dan Lemeshow (2000) model umum regresi logistik dengan p faktor (p variabel prediktor) adalah

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}}{1 + e^{(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}}$$

Untuk mempermudah pendugaan parameter regresi maka model regresi logistik dapat diuraikan dengan menggunakan transformasi logit dari $\pi(x_i)$ sebagai berikut

$$\ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

Pendugaan parameter dalam model regresi logistik biner menggunakan metode kemungkinan maksimum (*Maximum Likelihood Estimation*), kemudian diselesaikan dengan metode iterasi Newton Raphson. Pengujian terhadap parameter-parameter model regresi logistik biner dilakukan baik secara simultan maupun secara parsial. Menurut Hosmer dan Lemeshow (2000), statistik uji yang digunakan untuk pengujian parameter model secara simultan adalah statistik uji G atau *likelihood ratio test*. Hipotesis dalam uji simultan adalah

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ (tidak ada pengaruh antara variabel prediktor dengan variabel respon)

$H_1 : \text{minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, p$
(ada pengaruh paling sedikit satu variabel prediktor dengan variabel respon)

Statistik uji :

$$G = -2 \ln \frac{\binom{n_1}{n_1} \binom{n_0}{n_0}^{n_0}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}}$$

Kriteria penolakan H_0 yaitu tolak H_0 jika $G > \chi^2(\alpha, v)$ dengan v derajat bebas adalah banyaknya parameter dalam model tanpa β_0 atau nilai $p - value < \alpha$.

Setelah dilakukan pengujian parameter secara simultan, langkah selanjutnya adalah melakukan pengujian parameter secara parsial. Menurut Hosmer dan Lemeshow (2000), pengujian parameter secara parsial menggunakan uji Wald dengan hipotesis sebagai berikut:

$H_0 : \beta_j = 0$ (tidak ada pengaruh antara variabel prediktor dengan variabel respon) (2)

$H_1 : \beta_j \neq 0; j = 1, 2, \dots, p$ (ada pengaruh antara variabel prediktor dengan variabel respon)

Statistik uji :

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (3)$$

Kriteria penolakan H_0 yaitu tolak H_0 jika $W^2 > \chi^2(\alpha, v)$ dengan v merupakan derajat bebas, $v = 1$ atau nilai $p - value < \alpha$.

Uji kesesuaian model digunakan untuk mengetahui apakah ada perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model (Hosmer dan Lemeshow, 2000). Hipotesis yang digunakan adalah

$H_0 : \text{Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)}$

$H_1 : \text{Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)}$

Statistik Uji :

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Intepretasi koefisien untuk model regresi logistik biner dapat dilakukan dengan menggunakan nilai rasio oddsnya. Odds sendiri dapat diartikan sebagai rasio peluang kejadian sukses dengan kejadian tidak sukses dari peubah respon. Rasio odds didefinisikan sebagai :

$$\psi = \exp(\beta_i)$$

Menurut Hosmer dan Lemeshow (2000), salah satu ukuran kebaikan model dalam regresi logistik adalah jika memiliki peluang kesalahan klasifikasi yang minimal.

$$\text{Ketepatan Klasifikasi (\%)} = \left(\frac{n_{00} + n_{11}}{n_0 + n_1} \right) \times 100\%$$

$$\text{APER (\%)} = \left(\frac{n_{01} + n_{10}}{n_0 + n_1} \right) \times 100\%$$

Analisis CART (*Classification and Regression Trees*)

CART menggunakan pendekatan nonparametrik, sehingga tidak ada asumsi harus mengikuti suatu distribusi tertentu, baik untuk variabel respon ataupun variabel prediktor. Selain itu, analisis ini tidak terpengaruh oleh adanya pencilan (*outlier*), kolinieritas, ataupun heterokedastisitas (Yohannes dan Webb, 1999). Menurut Breiman dkk. (1993), CART menghasilkan suatu pohon klasifikasi (*classification trees*) jika variabel responnya katagorik dan menghasilkan pohon regresi (*regression trees*) jika variabel responnya kontinu. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian.

Pembentukan Pohon Klasifikasi

Menurut Breiman dkk. (1993) pembentukan pohon klasifikasi terdiri atas 3 tahapan yaitu sebagai berikut :

A. Pemilahan Pemilah

Pemilihan pemilah menggunakan data *learning sample* (\mathcal{L}) yang masih bersifat heterogen untuk pembentukan pohon klasifikasi. Pemilahan dilakukan untuk mengurangi keheterogenan pada simpul utama dan memaksimumkan ukuran kehomogenan pada setiap simpul dalam. Untuk mengukur tingkat keheterogenan suatu kelas dari suatu simpul pada pohon klasifikasi digunakan *Impurity measure* $i(t)$. Fungsi

impurity measure yang digunakan dalam penelitian ini adalah kaidah *Twoing*. Menurut Zambon dkk. (2006), fungsi kaidah *Twoing* dapat ditulis sebagai berikut:

$$i(t) = \frac{P_L P_R}{4} (\sum_i |(P(i|t_L) - P(i|t_R))|)^2$$

Misalkan terdapat pemilahan s yang memilah simpul t menjadi simpul kiri t_L (dengan proporsi P_L) dan simpul kanan t_R (dengan proporsi P_R) dimana $P_R = 1 - P_L$ dan $t_R \cup t_L = 1$, maka kebaikan dari pemilah (*goodness of split*) didefinisikan sebagai penurunan keheterogenan (*improvement*) (Breiman dkk., 1993) :

$$\Delta i(s, t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R)$$

dimana :

$\Delta i(s, t)$: penurunan tingkat keheterogenan (*improvement*)

Pengembangan pohon dilakukan dengan cara pada simpul utama t_1 , carilah s^* yang memberikan nilai penurunan keheterogenan tertinggi, yaitu :

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1)$$

Maka simpul t_1 dipilah menjadi t_2 dan t_3 menggunakan s^* , dengan cara yang sama dilakukan juga pemilah terbaik pada t_2 dan t_3 secara terpisah dan seterusnya.

B. Penentuan Simpul Terminal

Suatu simpul t akan menjadi simpul terminal atau tidak akan dipilah lagi bila tidak terdapat penurunan keheterogenan yang berarti, hanya terdapat satu pengamatan pada tiap simpul anak atau adanya batasan minimum n dan adanya batasan jumlah level atau tingkatan kedalaman pohon maksimal (Lewis, 2000).

C. Penandaan Label Kelas

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

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

Maka label kelas untuk simpul terminal t adalah j_0 yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul t terkecil sebesar

$$r(t) = 1 - \max_j p(j|t)$$

Resubstitution estimate (proporsi kesalahan) untuk tingkat kesalahan pengklasifikasian keseluruhan dari pohon klasifikasi, yaitu:

$$R(T) = \sum_{t \in T} R(t) = \sum_{t \in T} r(t) p(t)$$

Penentuan Pohon Klasifikasi Optimal

Menurut Breiman dkk. (1993), salah satu cara mendapatkan pohon optimal yaitu dengan pemangkasan (*pruning*). Bagian pohon yang kurang penting dipangkas untuk mendapatkan pohon klasifikasi yang optimal. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah *cost complexity measure*.

Definisi 1. (Breiman dkk., 1993) Diketahui subtree $T < T_{max}$ didefinisikan *complexity* (kompleksitas) dari subtree ini adalah $|\tilde{T}|$, yaitu banyaknya simpul terminal yang dimiliki T , $\alpha \geq 0$ adalah kompleksitas parameter dan *cost complexity measure* (ukuran biaya kompleksitas) $R_\alpha(T)$ maka

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

Proses 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 dua simpul anak dan simpul induk yang memenuhi persamaan (2.31) maka simpul t_R dan t_L dipangkas. Hasilnya adalah pohon T_1 yang memenuhi kriteria $R(T_1) = R(T_{max})$. Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin terjadi.

$$R(t) = R(t_R) + R(t_L)$$

Hasil proses pemangkasan berupa sederet pohon klasifikasi T_k . Pemilihan Pohon klasifikasi optimal berdasarkan pada *subtree* yang mempunyai nilai *cost complexity measure* yang terkecil dan nilai penduga pengganti yang cukup kecil. Nilai penduga pengganti diperoleh dengan menggunakan validasi silang (*cross-validation sample*) sebagai berikut:

$$R^{CV}(T_k) = \frac{1}{N} \sum_i C(i|j) N_{ij}$$

Interpretasi Pohon Klasifikasi Optimal

Interpretasi dari pohon klasifikasi yang telah terbentuk disajikan dalam suatu aturan hubungan (*association rules*) yang disebut dengan *classification rules*. *Classification Rules* adalah aturan pengklasifikasian yang berbentuk *if...then...* (**jika... maka...**) (Breiman dkk, 1993).

Ketepatan Prediksi Pohon Klasifikasi Optimal

Pohon klasifikasi optimal yang dihasilkan oleh data *learning* diuji tingkat keakuratannya dengan memasukkan data *testing* ke dalam pohon klasifikasi optimal, sehingga dihasilkan angka ketepatan klasifikasi. Uji keakuratan pohon klasifikasi optimal dilakukan dengan menghitung proporsi *misclassification*.

$$R(d) = \frac{1}{N} \sum_{n=1}^N X(d(x_n) \neq j_n)$$

Semakin besar $R(d)$ maka tidak akurat pohon klasifikasi yang terbentuk dalam melakukan prediksinya. Ketepatan klasifikasi pohon klasifikasi optimal dapat diperoleh sebagai berikut

$$\text{Ketepatan Klasifikasi (\%)} = 1 - R(d)$$

III. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari data rekam medis pasien penyakit jantung yang menjalani rawat jalan di RSUD Dr. M. Yunus Bengkulu Tahun 2014. Sampel dalam penelitian ini sebanyak 846 orang. variabel respon (Y) dalam penelitian ini adalah penyakit jantung yang terdiri dari dua kategori yaitu PJK dan NONPJK. Sedangkan variabel prediktor yang digunakan adalah jenis kelamin X_1 , umur X_2 , dislipidemia X_3 , hipertensi X_4 , diabetes melitus X_5 , dan obesitas X_6 .

Tahapan –tahapan yang akan dilakukan dalam penelitian ini:

1. Mendiskripsikan variabel - variabel penelitian yang digunakan
2. Analisis Regresi Logistik Biner. Tahapannya adalah sebagai berikut:
 - a. Melakukan pengujian multikolinieritas

- b. Membentuk model logit regresi logistik biner
 - c. Menduga parameter dengan metode maksimum *likelihood* dan iterasi Newton-Raphson.
 - d. Melakukan pengujian parameter secara serentak dengan uji G (*likelihood ratio test*)
 - e. Melakukan pengujian parameter secara parsial dengan uji *Wald*
 - f. Memilih model terbaik dengan uji Hosmer and Lemeshow
 - g. Interpretasi model dengan menggunakan *Odds ratio*
 - h. Menghitung ketepatan klasifikasi model
3. Analisis CART. Tahapannya adalah sebagai berikut:
- a. Membagi data menjadi 2 yaitu data *learning* dan data *testing*
 - b. Pembentukan pohon klasifikasi menggunakan data *learning*.
 - i. Mengukur tingkat keheterogenan untuk setiap variabel prediktor dengan menggunakan kaidah Twoing.
 - ii. Mengukur penurunan tingkat keheterogenan setiap variabel prediktor dengan menggunakan *Goodness of Split*.
 - iii. Memilih variabel dengan nilai *improvement* atau penurunan tingkat keheterogenan tertinggi sebagai kriteria dalam memilih pemilah terbaik yang digunakan untuk memecah sebuah simpul.
 - iv. Penentuan simpul terminal
 - v. Penandaan label kelas pada simpul terminal
 - c. Pembentukan pohon klasifikasi optimal dengan pemangkasan menggunakan *Cost Complexity Measure* dan *Cross Validation*.
 - d. Menguji keakuratan model pohon klasifikasi optimal yang dihasilkan oleh data *learning* dengan memasukkan data *testing* ke dalam model tersebut sehingga dihasilkan angka ketepatan klasifikasi.
 - e. Interpretasi model pohon klasifikasi optimal.

4. Membandingkan hasil dari analisis regresi logistik biner dan analisis CART berdasarkan tingkat ketepatan klasifikasi yang dihasilkan kedua analisis.

IV. HASIL DAN PEMBAHASAN

Deskripsi Data Pasien Penyakit Jantung yang Dirawat di RSUD Dr. M. Yunus Bengkulu

Pasien penyakit jantung yang dirawat di RSUD Dr. M. Yunus Bengkulu pada tahun 2014 ada lah sebanyak 846 orang. Jumlah pasien yang termasuk kategori PJK sebanyak 577 orang (68.2%) dan kategori NONPJK sebanyak 269 orang (31.8%). Pasien penyakit jantung mayoritas berjenis kelamin laki-laki, umur lebih besar dari 45 tahun, menderita dislipidemia, hipertensi, diabetes melitus dan obesitas.

Hasil Analisis Regresi Logistik Biner

Pendugaan model regresi logistik dengan enam variabel prediktor menghasilkan nilai *likelihood ratio* (G^2) = 402.643 dengan nilai p -value = 0.000. Pengambilan keputusan dilakukan dengan membandingkan nilai G^2 dengan kai-kuadrat dari tabel $\chi^2_{(6;0.05)} = 12.592$. Karena $G^2 = 402.643 > \chi^2_{(6;0.05)} = 12.592$ dan p -value = 0.000 < $\alpha = 0.05$ maka H_0 ditolak, artinya minimal ada satu variabel prediktor yang berpengaruh secara simultan terhadap variabel respon. Setelah dilakukan pengujian signifikansi parameter secara simultan akan dilakukan pengujian secara parsial dengan menggunakan uji *Wald*. Pengujian signifikansi parameter secara parsial menghasilkan empat variabel prediktor yang berpengaruh nyata terhadap variabel respon yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Model terbaik diperoleh dengan hanya memasukkan variabel yang signifikan ke dalam langkah analisisnya.

Berdasarkan Tabel 1(Lampiran) kolom *Sig.* diketahui bahwa variabel dislipidemia, diabetes melitus dan obesitas mempunyai nilai p -value = 0.000 < $\alpha = 0.05$ dan variabel hipertensi mempunyai nilai p -value = 0.012 < $\alpha = 0.05$, artinya variabel dislipidemia, hipertensi, diabetes melitus dan

obesitas merupakan variabel yang signifikan berpengaruh terhadap penyakit jantung.

Dengan demikian model regresi logistik layak digunakan untuk memprediksi variabel respon dan model akhir logit dapat dinyatakan sebagai berikut:

$$g(x_i) = -3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}$$

Sehingga diperoleh model terbaik regresi logistik biner sebagai berikut:

$$\pi(x_i) = \frac{\exp[g(x_i)]}{1 + \exp[g(x_i)]} = \frac{\exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}{1 + \exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}$$

Dari Tabel 1 dapat dilihat nilai rasio odds untuk variabel dislipidemia, hipertensi, diabetes melitus dan obesitas pada kolom $\text{Exp}(B)$. Nilai rasio odds untuk variabel dislipidemia adalah 11.962, artinya penderita dislipidemia berisiko menderita PJK 11.962 kali dibandingkan yang bukan penderita dislipidemia. Variabel hipertensi mempunyai nilai rasio odds sebesar 0.592, artinya penderita hipertensi berisiko menderita PJK 0.592 kali dibandingkan yang bukan penderita hipertensi. Nilai rasio odds untuk variabel diabetes melitus adalah 11.421, artinya penderita diabetes melitus dibandingkan yang tidak diabetes melitus berisiko menderita PJK 11.421 kali. Variabel obesitas mempunyai nilai rasio odds sebesar 14.112, artinya penderita obesitas berisiko menderita PJK 14.112 kali dibandingkan yang bukan penderita obesitas.

Berdasarkan Tabel 2 (Lampiran) diketahui bahwa ketepatan klasifikasi yang dihasilkan metode regresi logistik biner adalah 84.2%.

Hasil CART

Pohon klasifikasi maksimal yang dihasilkan mempunyai 28 simpul anak dan 15 simpul terminal dengan 7 kedalaman, dimana seluruh variabel prediktor berpengaruh pada pohon klasifikasi maksimal. Pohon klasifikasi maksimal yang terbentuk mengandung ukuran biaya kompleksitas (*test set relative cost*) sebesar 0.12000 ± 0.06499 atau antara 0.05501 sampai 0.18499 dengan penduga pengganti (*resubstitution relative cost*) sebesar 0.32383

dan kompleksitas parameter 0.000000. Pohon klasifikasi maksimal menghasilkan nilai *relative cost* menggambarkan struktur data yang sangat kompleks, sehingga perlu dilakukan pemangkasan pohon agar diperoleh pohon optimal yang memiliki nilai *relative cost* yang lebih kecil. Pohon optimal yang terbentuk terdiri dari 6 simpul anak dan 4 simpul terminal dengan kedalaman 4 (Gambar 1 pada Lampiran). Pohon klasifikasi ukuran biaya kompleksitas (*test set relative cost*) sebesar 0.10667 ± 0.07539 atau antara 0.03128 sampai 0.18206 dengan penduga pengganti (*resubstitution relative cost*) sebesar 0.33492 dan kompleksitas parameter 0.002371.

Adapun interpretasi hasil untuk pohon klasifikasi optimal sebagai berikut:

1. Jika pasien menderita dislipidemia maka diprediksi akan menderita PJK.
2. Jika pasien tidak menderita dislipidemia dan tidak menderita obesitas maka diprediksi akan menderita NONPJK.
3. Jika pasien tidak menderita dislipidemia tetapi menderita obesitas dan menderita diabetes melitus maka diprediksi akan menderita PJK.
4. Jika pasien tidak menderita dislipidemia tetapi menderita obesitas dan tidak menderita diabetes melitus maka diprediksi akan menderita NONPJK.

Berdasarkan Tabel 3 (Lampiran) diketahui bahwa ketepatan klasifikasi yang dihasilkan metode CART adalah 95%.

V. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Variabel prediktor yang signifikan terhadap variabel respon dari hasil analisis regresi logistik biner yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Model regresi logistik biner yang diperoleh adalah

$$\pi(x_i) = \frac{\exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}{1 + \exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}$$

dimana $x_i = 0$ dan $1 ; i = 0,1$

Ketepatan klasifikasi yang dihasilkan metode regresi logistik biner adalah 84.2%.

2. Pohon klasifikasi yang terbentuk menghasilkan pohon optimal yang terdiri dari 6 simpul anak dan 4 simpul terminal dengan kedalaman 4. Variabel prediktor yang masuk ke dalam pohon klasifikasi yaitu dislipidemia, diabetes melitus dan obesitas. Variabel yang menjadi pemilah utama adalah variabel dislipidemia. Metode CART memiliki nilai ketepatan klasifikasi sebesar 95%.
3. Perbandingan hasil dari kedua analisis menunjukkan bahwa pada kasus ini metode CART lebih baik dari metode regresi logistik biner, karena metode CART memiliki ketepatan klasifikasi sebesar 95% sedangkan metode regresi logistik biner menghasilkan ketepatan klasifikasi sebesar 84.2%

Saran

Penelitian tentang PJK ini hanya menggunakan enam variabel prediktor, yaitu variabel jenis kelamin, umur, dislipidemia, hipertensi, diabetes melitus dan obesitas. Penelitian selanjutnya hendaknya menggunakan variabel prediktor yang lebih lengkap, misalnya dengan menambahkan variabel keturunan, stress dan merokok.

DAFTAR PUSTAKA

- Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1993. *Cassification and Regression Tree*. Chapman And Hall. New York.
- Gujarati, D. 2010. *Ekonometri Dasar. Selemba Empat*. Jakarta.
- Hosmer, D. W. and S. Lemeshow. 2000. *Applied Logistic Regression Second Edition*. John Willey & Son, Inc : New York.
- Lewis, R. J. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.

- Margasari, A. 2014. Penerapan Metode CART (Classification and Regression Trees) dan Analisis Regresi Logistik Biner pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya. *Jurnal Mahasiswa Statistik*, Vol. 2, No. 4 (2014). FMIPA Universitas Brawijaya

<http://statistik.studentjournal.ub.ac.id/index.php/statistik/article/view/151>.

Diakses tanggal 16 Februari 2014.

- Neter, J., W. Wasserman and M. H. Kutner. 1990. *Applied Linear Statistical Models*. Third Edition. Richard D. Irwin, Inc, Homewood, Illinois 60430.
- Sezgin, O., E. Buyukbeci, A. I. Cekic, F. Y. Ozkurt, P. Taylan and G. H. Weber. 2006. *CMARS and GAM & CPQ, Modern Optimization Methods Applied to International Credit Default Prediction*. The Middle East Technical University. Turkey.
- Soeharto, I. 2004. *Serangan Jantung dan Stroke Hubungannya dengan Lemak dan Kolesterol Edisi Kedua*. Penerbit PT. Gramedia Pustaka Utama. Jakarta.
- Yohannes, Y. and P. Webb. 1999. *Classification and Regression Trees, CARTTM : A User Manual For Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. International Food Policy Research Institute. Washington, U. S. A.
- Zakiyah, W., H. Permadi dan S. Rahardjo. 2012. Implementasi Regresi Logistik Biner pada Penentuan Faktor-faktor yang Mempengaruhi Penyakit Jantung. *Jurnal Online*, Vol. 1, No.3 (2012). Universitas Negeri Malang. <http://jurnal-online.um.ac.id/data/artikel/artikel202375F42424FC943D2F28AE7B51DA32.pdf> . Diakses tanggal 14 Februari 2015.
- Zambon, M., R. Lawrence, A. Bunn and S. Powell. 2006. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. *Photogrammetric Engineering & Remote Sensing* Vol. 72, N0. 1 : 25-30.

LAMPIRAN

Tabel 1. Output Model Regresi Logistik Biner dengan Variabel Prediktor yang Signifikan

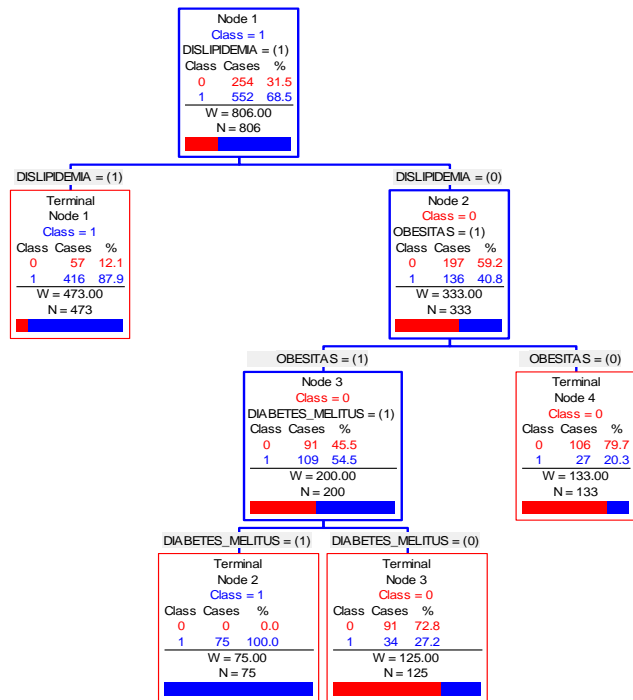
	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)		
							Lower	Upper	
Step 1 ^a	Dislipidemia(1)	2.482	.207	144.347	1	.000	11.962	7.979	17.931
	Hipertensi(1)	-.523	.208	6.319	1	.012	.592	.394	.891
	Diabetes_Melitus(1)	2.435	.260	87.938	1	.000	11.421	6.865	19.000
	Obesitas(1)	2.647	.262	102.214	1	.000	14.112	8.448	23.576
	Constant	-3.163	.317	99.406	1	.000	.042		

a. Variable(s) entered on step 1: Dislipidemia, Hipertensi, Diabetes_Melitus, Obesitas.

Tabel 2. Klasifikasi Analisis Regresi Logistik Biner

	Observed	Predicted		
		Penyakit_Jantung		Percentage Correct
		NONPJK	PJK	
Step 1	Penyakit_Jantung NONPJK	214	55	79.6
	PJK	79	498	86.3
	Overall Percentage			84.2

a. The cut value is .500



Gambar 1. Pohon Klasifikasi Optimal

Tabel 3. Ketepatan Klasifikasi Data *Testing* pada Pohon Optimal

Actual Class	Predicted		Total Class	Percent Correct
	NONPJK	PJK		
NONPJK	14	1	15	93.33%
PJK	1	24	25	96.00%
Overall Percent Correct				95%