

PERBANDINGAN ANTARA METODE CART (*CLASSIFICATION AND REGRESSION TREE*) DAN REGRESI LOGISTIK (*LOGISTIC REGRESSION*) DALAM MENGLASIFIKASIKAN PASIEN PENDERITA DBD (DEMAM BERDARAH DENGUE)

R. Lestawati¹, Rais², dan I. T. Utami³

^{1,2,3} Program Studi Matematika Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tadulako

Jalan Sukarno-Hatta Km. 9 Palu 94118, Indonesia

rialestawati.asman@gmail.com, rais76_untad@yahoo.co.id, triutami.iut@gmail.com

ABSTRACT

Classification is one of statistical methods in grouping the data compiled systematically. The classification of an object can be done by two approaches, namely classification methods parametric and non-parametric methods. Non-parametric methods is used in this study is the method of CART to be compared to the classification result of the logistic regression as one of a parametric method. From accuracy classification table of CART method to classify the status of DHF patient into category of severe and non-severe exactly 76.3%, whereas the percentage of truth logistic regression was 76.7%, CART method to classify the status of DHF patient into categories of severe and non-severe exactly 76.3%, CART method yielded 4 significant variables that hepatomegaly, epitaksis, melena and diarrhea as well as the classification is divided into several segmens into a more accurate whereas the logistic regression produces only 1 significant variables that hepatomegaly.

Keywords : CART , DHF, Logistic Regression.

ABSTRAK

Pengklasifikasian merupakan salah satu metode statistik dalam pengelompokkan suatu data yang disusun secara sistematis. Pengklasifikasian suatu objek dapat dilakukan dengan dua pendekatan metode klasifikasi yaitu metode parametrik dan non parametrik. Metode non parametrik yang digunakan dalam penelitian ini adalah metode CART yang akan dibandingkan dengan hasil pengklasifikasian dari regresi logistik yang merupakan metode parametrik. Dari table ketepatan hasil klasifikasi diketahui bahwa metode CART mengklasifikasikan status pasien penderita DBD kedalam kategori parah dan tidak parah dengan tepat sebesar 76.3%, sedangkan pada regresi logistik presentase kebenarannya adalah 76.7%. Metode CART menghasilkan 4 variabel yang signifikan yaitu hepatomegali, epitaksis, melena dan diare sedangkan pada regresi logistik menghasilkan 1 variabel yang signifikan yaitu hepatomegali.

Kata Kunci : CART, DBD, Regresi Logistik.

I. PENDAHULUAN

1.1. Latar Belakang

Pengklasifikasian merupakan salah satu metode statistik dalam pengelompokan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari, dalam segala bidang kehidupan. Masalah klasifikasi ini muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat diidentifikasi secara langsung. Pengklasifikasian suatu objek dapat dilakukan dengan dua pendekatan metode klasifikasi yaitu metode parametrik dan non parametrik.

Metode parametrik yang digunakan dalam penelitian ini adalah regresi logistik yang akan dibandingkan dengan hasil pengklasifikasian dari metode CART yang merupakan metode non parametrik. Regresi logistik digunakan untuk mengetahui hubungan antara variable independen dan variable dependen. Kelebihan dari metode regresi logistik ini adalah pelanggaran beberapa asumsi yang harus ada pada regresi linear biasa seperti asumsi kenormalan dan homokedastisitas tidak harus ada pada regresi logistik.

CART (*Classification and Regression Tree*) merupakan metodologi statistika nonparametrik berdasarkan kaidah pohon keputusan, baik untuk peubah respon kategorik maupun kontinu. Metode pohon klasifikasi memiliki kelebihan antara lain struktur datanya dapat dilihat secara visual, dan selain memberikan klasifikasi namun juga memberikan estimasi probabilitas kesalahan pengklasifikasian. Perbandingan antar jenis metode baik parametrik dan nonparametrik dalam suatu penelitian, telah banyak dilakukan oleh peneliti-peneliti sebelumnya contohnya penelitian yang berjudul Ketepatan Masa Studi Mahasiswa UIN Sunan Kalijaga (Muflihan H, 2013), penelitian yang berjudul Klasifikasi Dengan Metode CHAID Dan Penerapannya Pada Klasifikasi Alumni FMIPA UNY (Husein Permana, 2011). Namun, pada penelitian ini peneliti tertarik untuk membandingkan metode CART dan metode regresi logistik dengan menggunakan data pasien penderita DBD yang diambil dari RSUD Anutapura Palu.

1.2. Batasan Masalah

Batasan Ruang lingkup dalam penelitian ini terbatas pada :

1. Faktor-faktor gejala penyakit DBD yakni diantaranya epitaksis, hepatomegali, mual, melenas, penurunan nafsu makan, menggigil, diare, sakit kepala, konjungtiva, muntah dan puspura akan diklasifikasikan dengan menggunakan CART dan regresi logistik.
2. Data yang digunakan adalah data penderita DBD yang bersumber dari RSUD Anutapura Palu.

II. METODE PENELITIAN

Penelitian dilakukan sesuai prosedur dibawah ini:

- a. Memulai melakukan pengumpulan data.
- b. Melakukan pengkategorian/pengkodean data sesuai tipe dan kategori datanya.
- c. Kemudian data di input kedalam SPSS 20.0
- d. Kemudian di buat model persamaan regresi logistik biner dan diterapkan metode regresi logistik biner untuk di analisa hubungan-hubungannya dan melihat signifikansinya.
- e. Langkah selanjutnya, adalah uji kelayakan model regresi digunakan *Uji Hosmer* dan *Lemeshow* dengan nilai *chi-square*. Pengujian ini akan melihat nilai *chi-square* pada tingkat signifikansi 5%. Jika nilai signifikansinya kurang dari 0,05 maka model cocok dengan data. Untuk melihat besarnya pengaruh variabel-variabel bebas terhadap variabel tak bebas maka digunakan Uji Wald atau Uji Koefisien Regresi, jika nilai signifikansi Wald lebih kecil dari 0,05 maka variabel bebas berpengaruh terhadap variabel tak bebas.
- f. Hasil dan Pembahasan.
- g. Menyimpulkan hasil penelitian.
- h. Selesai.

III. HASIL DAN PEMBAHASAN

3.1. Deskripsi Data

Pada penelitian ini menggunakan dua analisis yaitu analisis univariat dan analisis bivariat. Pengklasifikasian suatu objek sangat umum dilakukan dalam berbagai bidang. Metode klasifikasi memungkinkan peneliti untuk mengklasifikasikan pengamatan baru, yaitu menetapkan objek baru masuk ke dalam kelompok tertentu berdasarkan nilai atribut-atributnya (Salazar *et al.* 2012). Analisis hasil yang dilakukan bertujuan untuk membandingkan hasil ketelitian pengklasifikasian dengan metode CART dan regresi logistik berdasarkan data awal yakni data yang diperoleh dari data rekam medik pasien penderita DBD di RSUD Anutapura Palu pada bulan januari tahun 2012 sampai dengan bulan juni tahun 2014. Analisis univariat dilakukan untuk memperoleh gambaran setiap variabel, distribusi frekuensi berbagai variabel yang diteliti baik variabel dependen maupun variabel independen. Variabel bebas yang digunakan dalam penelitian ini ada 11 sedangkan variabel terikat hanya 1 dan merupakan variabel kategorik. Terpenuhinya asumsi-asumsi pada metode parametrik akan menghasilkan data yang dapat diklasifikasikan dengan baik (Johnson & Wichern, 2007).

3.1.1. Analisis Univariat

Analisis univariat dilakukan untuk memperoleh gambaran setiap variabel, distribusi frekuensi berbagai variabel yang diteliti. Dengan melihat distribusi frekuensi dapat diketahui deskripsi masing-masing variabel dalam penelitian.

Tabel 1 : Distribusi Frekuensi Variabel

Variabel	Frekuensi (N)		Total N	Persentase (%)		Total %
	0	1		0	1	
Epitaksis	154	305	459	33,55	66,45	100,00
Hepatomegali	120	339	459	26,14	73,86	100,00
Mual	320	139	459	69,72	30,28	100,00
Melena	113	346	459	24,62	75,38	100,00
Penurunan Nafsu Makan	206	253	459	44,88	55,12	100,00
Menggigil	111	348	459	24,18	75,82	100,00
Diare	297	162	459	64,71	35,29	100,00
Sakit Kepala	330	129	459	71,89	28,11	100,00
Konjungtiva	331	128	459	72,11	27,89	100,00
Muntah	326	133	459	71,02	28,98	100,00
Puspora	56	403	459	12,20	87,80	100,00

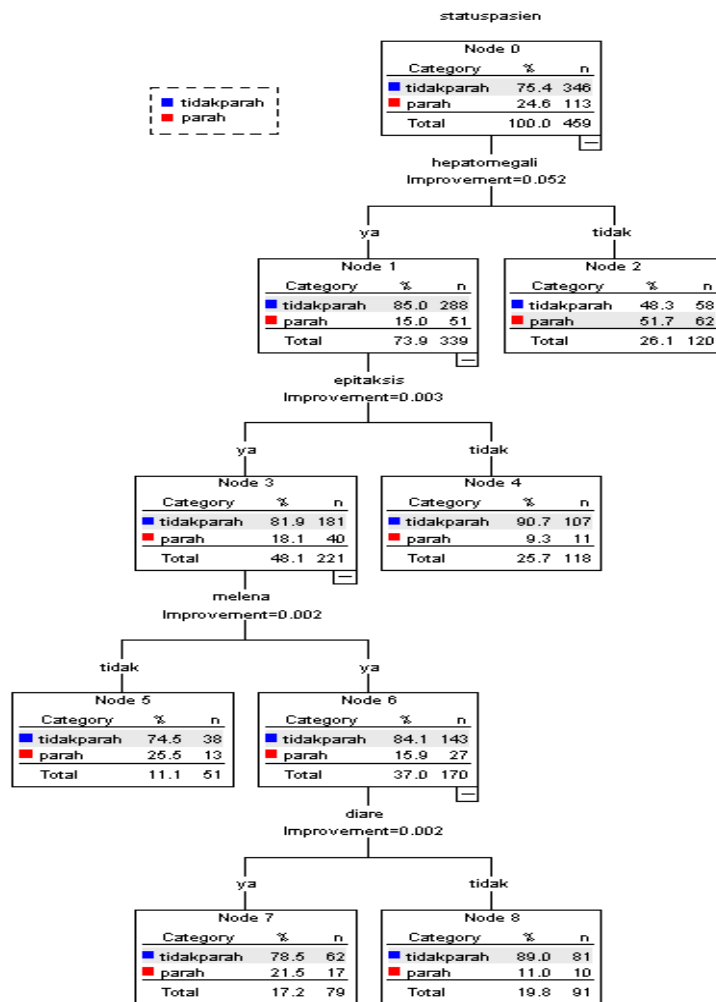
3.1.2. Analisis Bivariat

Untuk mengetahui ada tidaknya hubungan antara variabel bebas dengan variabel terikat dapat digunakan uji *chi square* melalui uji independensi antara variabel bebas dengan variabel terikat. Uji independensi dihitung berdasarkan tabel bantu yaitu tabel kontingensi 2 x 2. Dari uji *chi square* tersebut diperoleh nilai *chi square* hitung. Kriteria pengambilan keputusan atau penolakan H_0 diperoleh berdasarkan perbandingan nilai *chi square* hitung dengan nilai *chi square* tabel. Selain itu juga bisa dengan membandingkan nilai probabilitas dengan taraf nyata. Jika nilai *chi square* hitung lebih besar dari nilai *chi square* tabel atau nilai probabilitas yang lebih kecil dari taraf nyata maka hipotesis nol ditolak. Hal ini berarti bahwa terdapat hubungan antara variabel bebas dengan variabel terikat. Berdasarkan hasil tabel kontingensi di atas antara variabel Y (status pasien) dengan variabel X (gejala) $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$, dan X_{11} serta hasil pengujian independensi antara variabel bebas dan variabel terikat

3.1.3. Analisis CART

Pada penelitian ini digunakan CART yang merupakan salah satu metode eksplorasi non parametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk peubah respon kategorik maupun kontinu (Breiman *et al.* 1993). Proses pembentukan model pohon menggunakan metode CART akan menghasilkan sebuah model pohon yang memiliki *terminal node* yang paling banyak dan juga memiliki nilai resiko yang besar, untuk itu diperlukan adanya pengoptimalan model pohon yang

merupakan suatu model yang memiliki nilai resiko paling kecil. Di dalam model pohon optimal tersebut terdapat aturan-aturan klasifikasi yang dapat menjelaskan mengenai aturan pemberian keputusan RSUD Anutapura Palu terhadap pasien penderita DBD, apakah pasien penderita DBD tersebut parah atau tidak. Pohon klasifikasi yang berukuran besar akan memberikan nilai penaksir pengganti paling kecil, sehingga pohon ini cenderung dipilih untuk menaksir nilai dari variabel terikat. Tetapi ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi, karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimal yang berukuran sederhana tetapi memberikan nilai penaksir pengganti cukup kecil. Oleh karena variabel terikat yang digunakan dalam penelitian ini adalah variabel kategorik maka CART akan menghasilkan model *tree* berupa *classification tree* dengan aturan pemisahan yang dipakai adalah *Twoing splitting rule*.



Gambar 1 : Pohon Optimal

Dengan mengikuti alur pada gambar di atas yang merupakan model pohon optimal maka aturan klasifikasi dapat diklasifikasikan. Aturan klasifikasi beserta kemungkinan parah atau tidaknya pasien yang menderita DBD akan dibahas dari klasifikasi berdasarkan *terminal node* 0 sampai *terminal node* 8.

1. Pada Gambar 1 dapat dilihat bahwa pada simpul 0 menginformasikan bahwa dari 459 pasien penderita DBD di RSUD Anutapura yang statusnya tidak parah sebanyak 346 pasien (74.5%) dan sisanya 113 pasien (24.6%) statusnya parah. Peubah penjas yang paling berpengaruh terhadap status pasien adalah gejala hepatomegali. Hal ini dapat dilihat dari letaknya yang berada pada simpul awal dengan nilai penurunan tingkat keheterogenan sebesar 0.052.
2. Berdasarkan simpul 2, terlihat bahwa pasien penderita DBD yang tidak mengalami gejala hepatomegali sebanyak 120 pasien dengan kecenderungan pasien yang tidak parah sebanyak 58 pasien (48.3%) dan pasien yang parah sebanyak 62 pasien (51.7%). Secara eksploratif pohon klasifikasi (Gambar 1), memberikan gambaran bahwa peubah penjas kedua yang berpengaruh adalah gejala epitaksis dengan nilai penurunan tingkat keheterogenan sebesar 0.003.
3. Pada simpul 4 terlihat bahwa pasien penderita DBD yang tidak mengalami gejala epitaksis dan mengalami gejala hepatomegali sebagian besar statusnya tidak parah sebanyak 107 pasien (90.7%) dan pasien yang parah sebanyak 11 pasien (9.3%). Pohon klasifikasi (Gambar 1), memberikan gambaran bahwa peubah penjas ketiga yang berpengaruh adalah gejala melena dengan nilai penurunan tingkat keheterogenan sebesar 0.002.
4. Pada simpul 5 terlihat bahwa pasien penderita DBD yang tidak mengalami gejala melena, mengalami epitaksis dan mengalami hepatomegali sebagian besar statusnya tidak parah sebanyak 38 pasien (74.5%) dan pasien yang parah sebanyak 13 pasien (25.5%). Gambar 2 memberikan gambaran bahwa peubah penjas keempat yang berpengaruh adalah diare dengan nilai penurunan tingkat keheterogenan sebesar 0.002.
5. Pada simpul 7 terlihat bahwa pasien penderita DBD yang mengalami gejala diare, mengalami gejala melena, mengalami gejala epitaksis serta mengalami pula hepatomegali sebagian besar statusnya tidak parah sebanyak 62 pasien (78.5%) dan pasien yang parah sebanyak 17 pasien (21.5%).
6. Pada simpul 8 terlihat bahwa pasien penderita DBD yang tidak mengalami gejala diare, mengalami gejala melena, mengalami gejala epitaksis serta mengalami pula hepatomegali sebagian besar statusnya tidak parah sebanyak 81 pasien (81%) dan pasien yang parah sebanyak 10 pasien (11%).

Dari Tabel 2 berikut diketahui bahwa model CART yang terbentuk mengklasifikasikan sebuah kejadian kedalam kategori parah dan tidak parah dengan tepat sebesar 76,3%.

Tabel 2 : Ketepatan Klasifikasi

Teramati	Prediksi		
	Tidakparah	Parah	Persentase Ketepatan
Tidak parah	288	58	83.2%
Parah	51	62	54.9%
Persentase Keseluruhan	73.9%	26.1%	76.3%

3.1.4. Analisis Regresi Logistik

Pengujian statistik dengan menggunakan regresi logistik sebanyak 459 orang pasien yang diteliti dalam penelitian ini. Tujuan dari penelitian ini adalah untuk mengetahui status pasien yang mengidap penyakit DBD di RSUD Undata Palu. Model regresi logistik digunakan dalam penelitian ini karena variabel dependen bersifat dikotomi, yaitu status pasien, dimana hanya ada dua kriteria yaitu parah dan tidak parah, parah diberi kode 1 dan tidak parah diberi kode 0.

Hasil dari pengujian regresi logistik biner menunjukkan bahwa model regresi logistik biner yang memasukkan konstanta saja pada penelitian ini memiliki tingkat akurasi sebesar 76,7%. Berikut adalah tabel keakuratan dengan model yang hanya memasukkan konstanta saja :

Tabel 3 : Ketepatan Klasifikasi

Teramati	Prediksi		
	Tidakparah	Parah	Persentase Ketepatan
Tidakparah	309	37	89.3%
Parah	70	43	38.1%
Persentase Keseluruhan			76.7%

Setelah dilakukannya pengujian kelayakan model regresi dan diperoleh bahwa model regresi logistik biner dapat digunakan pada penelitian ini, maka dilanjutkan dengan dilakukannya pengujian hipotesis menggunakan regresi logistik biner. Berikut adalah hasil pengujian hipotesis secara parsial dan hasil uji hipotesis secara bersama-sama (simultan).

a. Hasil pengujian hipotesis secara parsial

Dalam pengujian hipotesis dengan menggunakan regresi logistik biner cukup dengan melihat tabel *Variabel in the Equation*, pada kolom signifikansi dibandingkan dengan tingkat signifikansi yang digunakan dalam model regresi logistik adalah $\alpha = 5\%$. Jika tingkat signifikansi lebih kecil dari taraf nyata 5% maka hipotesis dinyatakan terbukti. Pengujian hipotesis bertujuan untuk melihat hubungan secara parsial masing-masing variabel independen terhadap variabel dependen.

b. Hasil Pengujian Hipotesis Secara Simultan

Hasil pengujian hipotesis secara simultan adalah untuk mengetahui hubungan antara epitaksis, hepatomegali, mual, melena, penurunan nafsu makan, muntah, diare, sakit kepala, konjungtiva, muntah, puspura secara bersama-sama terhadap status pasien penderita DBD.

Tabel 4 : Nilai *Odds Ratio*

Variabel	<i>Odds Ratio</i>
Epitaksis (X1)	1,638
Hepatomagali (X2)	0,166
Mual (X3)	0,616
Melena (X4)	0,649
Penurunan Nafsu Makan (X5)	1,315
Menggigil (X6)	0,958
Diare (X7)	0,586
Sakit Kepala (X8)	0,579
Konjungtiva (X9)	0,547
Muntah (X10)	0,626
Puspura (X11)	0,878

Interpretasi Tabel 4 adalah pasien yang mengalami epitaksis memiliki peluang sebesar 1,638 kali lebih besar untuk terkena penyakit DBD dibanding pasien yang tidak mengalami epitaksis. Pasien yang mengalami hepatomegali memiliki peluang sebesar 0,166 kali lebih besar untuk terkena penyakit DBD. Pasien yang mengalami rasa mual memiliki peluang sebesar 0,616 kali lebih kecil untuk terkena penyakit DBD. Pasien yang mengalami melena memiliki peluang sebesar 0,649 kali lebih besar untuk terkena penyakit DBD dibanding pasien yang tidak mengalami melena. Pasien yang mengalami penurunan nafsu makan memiliki peluang sebesar 1,315 kali lebih besar untuk terkena penyakit DBD dibanding pasien yang nafsu makannya tidak mengalami penurunan. Pasien yang mengalami perasaan menggigil memiliki peluang sebesar 0,958 kali lebih besar

untuk terkena penyakit DBD dibanding pasien yang tidak mengalami hal tersebut. Demikian pula dengan pasien yang mengalami puspura yang memiliki peluang sebesar 0,878 kali lebih besar untuk terkena penyakit DBD. Sedangkan untuk pasien yang mengalami diare, sakit kepala, konjungtiva dan muntah memiliki peluang yang lebih kecil masing-masing 0,586; 0,579; 0,547; 0,626 kali untuk terkena penyakit DBD dibandingkan dengan pasien yang tidak mengalami diare, sakit kepala, konjungtiva dan muntah.

3.1.5. Perbandingan Metode CART dan Regresi Logistik

Berdasarkan hasil pengklasifikasian dengan metode CART diperoleh ketepatan hasil klasifikasi sebesar 76,3%. Sedangkan hasil pengklasifikasian dengan regresi logistik diperoleh ketepatan hasil klasifikasi sebesar 76,7%. Perbandingan hasil dari kedua analisis tersebut menunjukkan bahwa ketepatan hasil klasifikasi dengan metode CART (76,3%) memiliki nilai ketepatan prediksi yang lebih kecil dari analisis regresi logistik biner (76,7%). Oleh karena itu regresi logistik lebih baik dalam memprediksi peubah respon dilihat dari besarnya hasil klasifikasi dalam kasus status pasien penderita DBD di RSUD Anutapura Palu.

IV. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya, didapatkan beberapa kesimpulan sebagai berikut :

1. Metode CART mendapatkan ketepatan hasil klasifikasi sebesar 76,3% sedangkan metode regresi logistik biner mendapatkan ketepatan hasil klasifikasi sebesar 76,7%. Oleh sebab itu dapat disimpulkan bahwa metode klasifikasi dengan regresi logistik lebih akurat dibandingkan metode CART.
2. Metode CART menghasilkan 4 variabel yang signifikan yaitu hepatomegali, epitaksis, melena dan diare sedangkan metode regresi logistik biner hanya menghasilkan 1 variabel yang signifikan yaitu hepatomegali.

DAFTAR PUSTAKA

- [1] Breimen L, Friedman JH, Olshen RA, Stone CJ, *Classification and Regression Trees*. Newyork, 1993, Chapman and Hall
- [2] Johnson RA, Wichern DW, *Applied Multivariate Statistical Analysis Sixth Edition*. New Jersey. 2007, Prentice Hall International Inc.
- [3] Muflihun H, *Ketepatan Masa Studi Mahasiswa UIN Sunan Kalijaga*, 2013.
- [4] Permana H, *Klasifikasi Dengan Metode CHAID dan Penerapannya Pada Klasifikasi Alumni FMIPA UNY*, 2011.
- [5] Salazar DA, Velez JI, Salazar JC, *Comparison Between SVM and Logistic Regression: Which One is Better to Discriminate?* Revista Colombiana de Estadística, 2012, 35 (2): 223-237.