

Diagnosis of the type of delivery of pregnant women at Semen Padang Hospital Using the C4.5 Method

Rama Novialdi, Dony Permana*, Dodi Vionanda, Fadhilah Fitri

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Indonesia

*Corresponding author: donypermana@fmipa.unp.ac.id

Submitted : 06 November 2023

Revised : 31 Januari 2024

Accepted : 20 Februari 2024

ABSTRACT

The health of the mother and fetus is very important, but there are many challenges and risks associated with pregnancy and childbirth. According to WHO, in 2020 there were 287,000 cases of women dying during pregnancy and childbirth. Causative factors that affect the type of delivery include the age of pregnant women, MGG, systole, diastole, and pulse. One method that can be used to group the types of childbirth of pregnant women is classification. C4.5 is one of the methods used in forming decision trees to produce decisions. The purpose of C4.5 is to obtain attributes that will be the main criteria in the classification. Based on optimal tree results, the attribute that is the main criterion in classifying the type of delivery of pregnant women who give birth by caesar section and normal delivery at Semen Padang Hospital is MGG. Determination of classification results using confusion matrix resulted in an accuracy value of 74%, sensitivity of 80% to classify the type of delivery of pregnant women who gave birth caesar, and specificity of 66.67% to classify the type of delivery of pregnant women who gave birth normally.

Keywords: WHO, classification, C4.5



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Kehamilan merupakan kondisi seorang ibu sedang mempunyai janin didalam rahimnya dan diperlukan perawatan khusus untuk ibu dan janinnya (Hikmatulloh, 2019). Kesehatan ibu dan janin yang baru dilahirkan adalah sebuah hal yang penting untuk menentukan kesehatan generasi berikutnya, namun terdapat tantangan dan risiko yang banyak terkait dengan kehamilan dan persalinan. Perempuan yang meninggal selama masa kehamilan dan pada saat persalinan merupakan salah satu hal yang menyebabkan tingginya angka kematian ibu, karena berdasarkan *World Health Organization* pada tahun 2020 terdapat 287.000 kasus angka perempuan meninggal selama masa kehamilan dan persalinan tahun 2020.

Adapun beberapa faktor yang mempengaruhi jenis persalinan yaitu, menurut Detiana (2010) usia ibu untuk hamil jika usia untuk hamil terlalu muda atau terlalu tua maka akan terjadi komplikasi. Kedua minggu kehamilan atau MGG, pada saat usia kehamilan masih muda umumnya lebih rentan terjadinya keguguran yang disebabkan adanya masalah pada janin. Ketiga sistole dan diastole, jika sistole dan diastole tidak terkontrol maka dapat menghambat perkembangan janin, dan keempat nadi, denyut nadi seorang ibu hamil akan meningkat jika dibandingkan dengan seorang ibu yang tidak hamil. Dengan adanya keberagaman karakteristik seperti periode kehamilan dan keluhan yang terjadi pada saat masa kehamilan. Dalam hal ini sangat dibutuhkan kinerja pengolahan algoritma untuk menentukan pengklasifikasian tersebut, oleh sebab itu metode yang digunakan dalam penelitian ini adalah metode klasifikasi karena mampu mengklasifikasikan antara kehamilan yang melahirkan secara normal dengan kehamilan melahirkan secara yang caesar.

Ada beberapa macam metode dalam klasifikasi, yaitu C4.5, CART (*Classification And Regression Trees*), ID3, analisis regresi logistik, CHAID (*Chi-Square Automatic Detection*), SVM dan lain sebagainya. Metode klasifikasi yang dapat mengelompokkan sekumpulan data dalam bentuk kelas yang bersifat kategori adalah C4.5. Kelebihan dari C4.5 yaitu mampu mengatasi data dengan tipe data numerik. Cara kerja algoritma C4.5 adalah dengan membentuk pohon keputusan dan membentuk aturan-aturan berdasarkan bentuk pohon keputusan yang telah didapatkan. Konsep dalam pemilihan atribut dalam C4.5 dengan melihat nilai *gain ratio* pada tiap atribut, atribut yang dijadikan akar merupakan atribut dengan nilai *gain ratio* tertinggi.

Penelitian ini dilakukan untuk mengetahui atribut manakah yang akan menjadi kriteria utama dalam mendiagnosa jenis persalinan ibu hamil di Semen Padang Hospital, yang mana itu akan mempermudah ibu hamil dalam menentukan jenis persalinan dimasa mendatang.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang digunakan adalah data rekam medis kehamilan yang diperoleh dari Semen Padang Hospital tahun 2022. Variabel target (Y) yang akan digunakan dalam penelitian ini yaitu sesar dan normal dengan lima atribut yaitu usia ibu (X1), usia kehamilan (X2), sistole yaitu tekanan darah pada saat jantung memompa darah ke dalam pembuluh nadi (X3), diastole yaitu tekanan darah pada saat jantung mengembang dan menyedot darah kembali (X4) dan denyut nadi (X5).

B. Teknik Analisis Data

Adapun teknik analisis data yang akan digunakan dalam penelitian adalah C4.5 dengan langkah-langkah sebagai berikut.

1. Melakukan analisis statistik deskriptif untuk setiap variabel penelitian
Analisis statistik deskriptif bertujuan untuk merangkum informasi yang terdapat pada data dan menggambarkan karakteristik data dengan lebih jelas.
2. Membagi data *training* dan *testing*
Dalam penelitian ini menggunakan perbandingan 75% untuk data *training* dan 25% untuk data *testing*. Data *training* digunakan untuk proses pembelajaran, sedangkan data *testing* digunakan untuk melakukan prediksi.
3. Membuat pohon keputusan dengan langkah-langkah sebagai berikut.

- a. Menghitung *entropy* pada setiap atribut menggunakan persamaan (1)

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i \log_2 (p_i) \quad (1)$$

Keterangan:

p_i : proporsi setiap kategori

S : jumlah data

n : jumlah partisi S

- b. Menghitung nilai *information gain* menggunakan persamaan (2)

$$\text{Gain (A)} = \text{Entropy(S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy (S}_i) \quad (2)$$

Keterangan:

$|S_i|$: proporsi setiap kategori terhadap jumlah data

$|S|$: jumlah data

A : jumlah atribut

- c. Menghitung *split info* menggunakan persamaan (3)

$$\text{Split information}_A(S) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan:

S_i : banyaknya sampel untuk atribut A dengan kelas ke- i

- d. Menghitung *gain ratio* menggunakan persamaan (4)

$$\text{Gain Ratio (A)} = \frac{\text{Gain Information (A)}}{\text{Split information}_A(S)} \quad (4)$$

- e. Menentukan atribut utama dengan nilai *gain ratio* tertinggi sebagai akar.
- f. Mengulang tahapan a sampai d untuk setiap cabang hingga semua atribut mempunyai daun keputusan
- g. Evaluasi hasil klasifikasi dengan menggunakan *confusion matrix*

Tabel 1. *Confusion Matrix*

Classification		Predicted Class	
		Class = YES	Class = NO
Aktual Class	Class = YES	True Positif (TP)	False Positif (FP)
	Class = NO	False Negatif (FN)	True Negatif (TN)

(Sumber : Gorunescu, F. 2011)

Keterangan :

1. *True Positif* (TP) adalah data yang sebenarnya positif dan diklasifikasikan sebagai data positif oleh system.
2. *True Negatif* (TN) adalah data yang sebenarnya negatif dan diklasifikasikan sebagai data negatif oleh system.
3. *False Positif* (FP) adalah data yang sebenarnya negatif dan diklasifikasikan positif oleh system.
4. *False Negatif* (FN) adalah data yang sebenarnya positif dan diklasifikasikan sebagai data negatif oleh system.

Terdapat berbagai metode untuk menghitung ketepatan klasifikasi seperti *sensitivity*, *specificity*, dan akurasi (Rokarch dan Maimon, 2015).

$$Sensitivity (\%) = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$Specificity (\%) = \frac{TN}{FP+FN} \times 100\% \quad (6)$$

$$Accuracy (\%) = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (7)$$

III. HASIL DAN PEMBAHASAN

A. Statistik Deskriptif

Tabel 2 menyajikan karakteristik ibu hamil di Semen Padang Hospital.

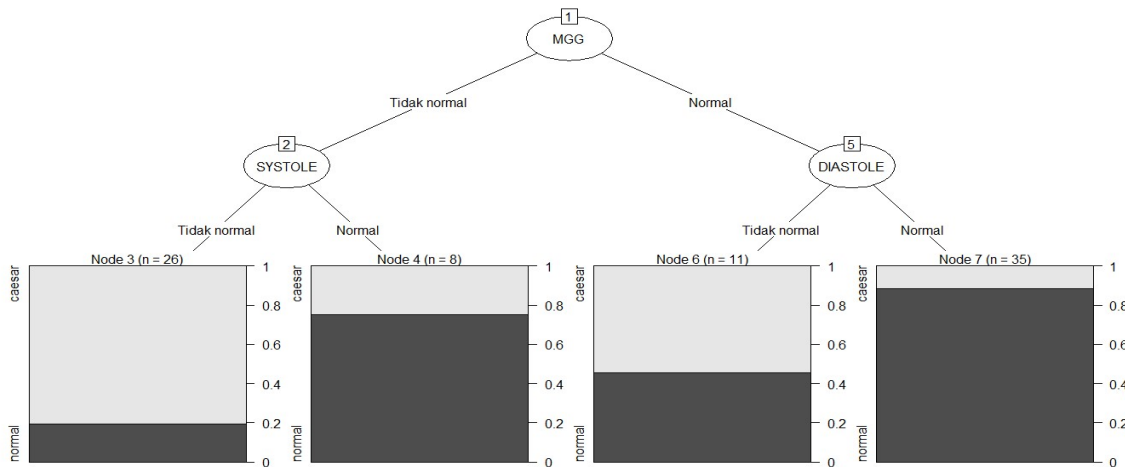
Tabel 2. Statistik Deskriptif

Variabel	Kategori	Jumlah
Usia ibu hamil (X1)	>29 tahun	34
	≤29 tahun	73
Usia Kehamilan (X2)	Tidak Normal	50
	Normal	57
Sistole (X3)	Tidak Normal	67
	Normal	40
Diastole (X4)	Tidak Normal	32
	Normal	75
Denyut Nadi (X5)	Tidak Normal	48
	Normal	59

Berdasarkan Tabel. 2 dapat dilihat bahwa dari 107 ibu hamil terdapat 54% atau 58 ibu hamil yang melahirkan secara normal. Ibu hamil banyak melahirkan pada saat berusia kurang dari 29 tahun yaitu sebanyak 73 dari 107 ibu hamil. Kebanyakan usia kehamilan seorang ibu hamil tersebut normal yaitu sebanyak 57 dari 107 ibu hamil. Tekanan darah pada ibu hamil terbagi 2 yaitu sistole dan diastole. sistole adalah tekanan darah pada saat jantung memompa darah ke dalam pembuluh nadi, sedangkan diastole adalah tekanan darah pada saat jantung mengembang dan darah menyedot darah kembali. Paling banyak ibu hamil mendapatkan nilai sistole yang tidak normal yaitu sebanyak 67 dari 107 ibu hamil, sedangkan pada diastole terdapat diastole yang normal lebih tinggi yaitu sebanyak 75 dari 107 ibu hamil. Umumnya denyut nadi ibu hamil ini kebanyakan normal yaitu sebanyak 59 dari 107 ibu hamil.

B. Algoritma C4.5

C4.5 digunakan untuk mengklasifikasi 107 data rekam ibu hamil di Semen Padang Hospital Tahun 2022. Pada penelitian ini terdapat 80 data *training* dan 27 data *testing*. Software yang digunakan menganalisis algoritma C4.5 ini yaitu Rstudio. *Dataset* dimasukkan ke Rstudio lalu algoritma C4.5 akan memulai mencari nilai sesuai dengan rumus persamaan 1 sampai persamaan 4. Nilai yang memiliki *gain ratio* tertinggi akan dipilih sebagai simpul akar. Perhitungan akan terus berlanjut sampai semua simpul memiliki keputusan. Setelah mendapatkan keputusan dengan menggunakan data *training* lalu algoritma akan langsung menguji keputusan tersebut terhadap data *testing* dan akan didapatkan hasil keputusan dan prediksi. Hasil dari analisis tersebut akan dilakukan evaluasi menggunakan metode *confusion matrix* seperti yang telah dijelaskan dalam Tabel 1. Gambar 1 menyajikan hasil pohon keputusan menggunakan algoritma C4.5



Gambar 1. Hasil pohon keputusan C4.5

Hasil pohon keputusan yang telah didapatkan menjelaskan bahwa atribut MGG yang menjadi atribut penentu dalam mendiagnosa jenis persalinan ibu hamil. Pada pohon keputusan diatas terdapat satu simpul akar sebagai kriteria utama, dua *internal node* dan empat *node* daun atau keputusan. Atribut yang ditempatkan pada simpul akar yaitu yang memiliki nilai *gain ratio* tertinggi seperti MGG berkategori normal. Pada atribut sistole dan diastole yang menunjukkan simpul internal. Pada simpul daun adalah diagnosa jenis persalinan caesar dan normal..

Tahap selanjutnya yaitu melakukan evaluasi ketepatan hasil klasifikasi menggunakan data *testing* sebanyak 27 data. Tabel 3 menyajikan hasil *confusion matrix*.

Tabel 3. Uji Ketetapan Hasil Klasifikasi

Classification		Predicted Class	
		Caesar	Normal
Aktual Class	Caesar	12	3
	Normal	4	8

Berdasarkan Tabel 3. Dapat dilihat bahwa jumlah caesar yang juga diprediksi caesar (*sensitivity*) sebanyak 12 ibu hamil. Sedangkan ibu hamil yang melahirkan secara normal diprediksi normal (*specificity*) sebanyak 8 ibu hamil. Nilai ketepatan model yang didapatkan dengan menggunakan persamaan (5), (6), dan (7) adalah sebagai berikut.

$$Sensitivity (\%) = \frac{12}{12+3} \times 100\% = 80 \%$$

$$\text{Specificity (\%)} = \frac{8}{4+8} \times 100\% = 66,67 \%$$

$$\text{Accuracy (\%)} = \frac{12+8}{12+3+4+8} \times 100\% = 74 \%$$

Berdasarkan hasil yang telah diperoleh, didapatkan nilai akurasi sebesar 74 %. Hal ini berarti bahwa pohon keputusan yang didapatkan mampu bekerja secara optimal sebesar 74%. Nilai *sensitivity* untuk mengukur ketepatan klasifikasi ibu hamil yang melahirkan secara caesar diperoleh sebesar 80%. Sedangkan nilai *specificity* yang didapatkan berguna untuk mengukur ketepatan hasil klasifikasi ibu hamil yang melahirkan secara normal yaitu sebesar 66,67 %

IV. KESIMPULAN

Berdasarkan hasil klasifikasi dengan menggunakan algoritma C4.5 pada data ibu hamil di Semen Padang Hospital telah didapatkan atribut yang kriteria utama dalam mendiagnosa jenis persalinan ibu hamil adalah MGG dan faktor lain dalam mendiagnosa jenis persalinan ibu hamil yaitu sistole dan diastole. Ketetapan hasil klasifikasi dengan menggunakan *confusion matrix* menghasilkan nilai akurasi sebanyak 74 %, *sensitivity* sebanyak 80 % untuk mengklasifikasikan ibu hamil yang melahirkan secara caesar. Dan *specificity* sebanyak 66,67 % untuk mengklasifikasikan ibu hamil yang melahirkan secara normal. Saran untuk peneliti selanjutnya yaitu menambahkan variabel respon untuk lebih bervariasi dan menggunakan lebih banyak data dalam mengklasifikasikan jenis persalinan ibu hamil berdasarkan faktor penyebabnya. Hal itu bertujuan agar klasifikasi yang dilakukan berdasarkan pertimbangan yang lebih matang.

DAFTAR PUSTAKA

- Detiana, Prilia. (2010). *Hamil Aman dan Nyaman di Atas 30 Tahun*. Media Pressindo : Yogyakarta.
- Gorunescu, Florin. (2011). *Data Mining: Concepts, Models and techniques*. Verlag Berlin Heidelberg : Springer
- Hikmatulloh. (2019). Penerapan Iterative Dichotomizer Three (ID3) Dalam Mendiagnosa Kesehatan Kehamilan. *Kumpulan Jurnal Ilmu Komputer, Vol.6 No.2* , 116-127.
- Manuaba, I. (2012). *Asuhan Kebidanan Pada Masa Kehamilan*. Jakarta : EGC.
- Rahmawati. (2019). *Perbandingan Klasifikasi Menggunakan Metode Backpropagation Dan Metode Learning Vector Quantization*. Skripsi. Universitas Brawijaya.
- Rokarch, L., dan Maimon, O. (2015). *Data Mining With Decision Trees Theory And Application 2nd Ed*. Singapore : World Scientific
- World Health Organization. (2023). *Maternal Mortality*. WHO
- Zhang, Pei. (2020). Construction of decision tree based on C4.5 algorithm for online voltage stability assessment. *International Journal of Electrical Power & Energy Systems, Volume.118*