

Implementasi Algoritma C5.0 Untuk Memprediksi Kondisi Kelahiran Bayi

Siti Sahara Nasution, Natalia Silalahi, Abdul Karim

Teknik Informatika Universitas Budi Darma, Medan, Indonesia

Email: siti112@gmail.com

Abstrak—Kondisi kelahiran adalah kondisi kelahiran bayi merupakan salah satu masalah utama dalam bidang kesehatan yang saat ini terjadi di Negara Indonesia. Dalam kesehatan, terdapat beberapa indikator dalam kondisi kelahiran yaitu normal dan caesar. Menentukan derajat kondisi kelahiran bayi merupakan cerminan status kelahiran bayi saat ini. Anak sebagai penerus bangsa berkaitan erat dengan pertumbuhan dan perkembangan bayi saat ini. Akan tetapi, pelaksanaan untuk meningkatkan tumbuh kembang bayi sampai saat ini masih belum optimal. Penelitian ini menghasilkan pohon keputusan dari kasus yang diangkat menunjukkan bahwa ada beberapa atribut yang mempengaruhi dalam penentuan kondisi kelahiran. Akan ditampilkan daftar nilai gain dari tiap atribut. Posisi atribut yang memiliki nilai gain paling tinggi yang ditampilkan menunjukkan penentuan atribut mana menjadi penentuan kelayakan.

Kata Kunci: Data Mining; Algoritma C5.0; Pengelompokan Data

Abstract—The condition of birth is the condition of the birth of a baby is one of the main problems in the health sector that is currently happening in Indonesia. In health, there are several indicators of birth conditions, namely normal and caesarean. Determining the degree of condition of the baby's birth is a reflection of the current status of the baby's birth. Children as the nation's successor are closely related to the growth and development of babies today. However, the implementation to improve infant growth and development is still not optimal. This study produces a decision tree of the cases raised showing that there are several attributes that influence the determination of birth conditions. A list of gain values for each attribute will be displayed. The position of the attribute that has the highest gain value displayed indicates the determination of which attribute is the determination of eligibility.

Keywords: Data Mining; C5.0 Algorithm; Data Group

1. PENDAHULUAN

Kondisi kelahiran bayi merupakan salah satu masalah utama dalam bidang kesehatan yang saat ini terjadi di Negara Indonesia. Sampai saat ini banyak program pembangunan kondisi kelahiran di Indonesia yang ditunjukkan guna menanggulangi angka kematian bayi. Dalam kesehatan, terdapat beberapa indikator dalam kondisi kelahiran yaitu normal dan caesar. Menentukan derajat kondisi kelahiran bayi merupakan cerminan status kelahiran bayi saat ini. Anak sebagai penerus bangsa berkaitan erat dengan pertumbuhan dan perkembangan bayi saat ini. Akan tetapi, pelaksanaan untuk meningkatkan tumbuh kembang bayi sampai saat ini masih belum optimal.

Dengan menjaga kondisi janin dalam keadaan yang sehat dan memiliki perkembangan yang normal dengan usianya, salah satunya yang perlu diperhatikan setiap ibu hamil dan menjadi tanggung jawab yang besar. Kondisi janin yang terjaga dengan kondisi kesehatan dalam kandungan kehamilan akan berlangsung dengan normal serta sehat dan selamat baik bayi maupun ibu itu sendiri. Keselamatan bayi dan perkembangan yang normal didalam kandungan salah satu bentuk hal yang diinginkan oleh setiap ibu hamil. Kondisi Bayi yang baru lahir masih sangat sensitive terhadap gangguan kesehatan. Pada tahap awal kelahiran, tumbuh kembang bayi menjadi salah satu fokus penting. Setelah bayi lahir biasanya dilakukan pemeriksaan dengan beberapa organ tubuh dan kondisi kesehatannya, Bayi yang lahir dengan cara normal biasanya memiliki bentuk kepala yang lebih lonjong jika dibandingkan dengan bayi yang operasi Caesar.

Istilah data mining disebut juga *knowledge discovery*. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang sangat membantu dalam pengambilan keputusan. Data mining telah diimplentasikan keberbagai bidang proses untuk mendapatkan informasi yang sangat berguna dari gudang basis data. Data mining sebagai penambang data atau upaya untuk menggali informasi yang berharga dan berguna pada database yang sangat besar. Munculnya data mining didasarkan kenyataan bahwa jumlah data yang tersimpan dalam basis data semakin besar. Misalnya, data diolah untuk memprediksi kondisi kelahiran normal atau caesar. Data yang tersimpan dibasis data dapat dimanfaatkan untuk membuat laporan untuk memprediksi kondisi kelahiran dan sebagainya. Yang pada akhirnya dapat digunakan untuk mengetahui kondisi kelahiran bayi tersebut dan bisa diterapkan untuk memprediksi kelahiran bayi normal ataupun caesar. Salah satu metode yang digunakan dalam implementasi pengetahuan dalam data mining adalah menggunakan metode C5.0 salah satu algoritma klasifikasi data mining yang khususnya diterapkan pada teknik decision tree atau merupakan penyempurnaan algoritma sebelumnya [1]. Penelitian yang digunakan adalah melalui tahapan Knowledge Discovery Database.

Proses untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi adalah proses pembentukan definisi konsep umum yang akan dipelajari. Setelah data dikumpulkan maka dapat untuk memprediksi kondisi kelahiran bayi. Penyajian data untuk memberikan deskripsi mengenai data yang telah dikumpulkan dan memudahkan untuk data dikelompokkan agar bisa diprediksi. Berdasarkan penelitian sebelumnya metode C5.0 memang dapat digunakan mengelompokkan data, sehingga memudahkan pembahasan dan analisisnya yang menunjukkan data secara visual [2].

2. METODOLOGI PENELITIAN

Menurut Hanifah Harsono, implementasi adalah kebijakan politik menyatakan pendapatnya mengenai implementasi atau pelaksana sebagai berikut : Implementasi adalah proses untuk melaksanakan kebijakan tersebut dalam tindakan kebijakan politik dalam membangun kebijakan administrasi dalam rangka meningkatkan program. Menurut Purwanto dan Sulistyastuti, implementasi adalah kegiatan untuk mendistribusikan keluaran bijakan (*to deliver policy output*) yang dilakukan oleh para implementer kepada kelompok sasaran upaya untuk terwujudnya tujuan kebijakan. Tujuan kebijakan diharapkan akan muncul dimana *policy output* dapat menerima dan dimanfaatkan dengan kelompok sasaran sehingga dalam jangka panjang hasil kebijakan akan mampu terwujud. Menurut Purwanto dan Sulistyastuti upaya untuk memahami implementasi akan lebih baik dilakukan dengan menggunakan dua pendekatan, yaitu : pendekatan pertama, memahami implementasi bagian dari proses atau siklus kebijakan. Pendekatan kedua, implementasi kebijakan dilihat dalam suatu studi atau bidang kajian. Implementasi merupakan salah satu dari proses tahapan serangkaian suatu siklus kebijakan[1].

Syukur dan Surmayadi menemukan ada tiga unsur yang penting dalam proses implementasi, yaitu : (1) adanya suatu program atau kebijakan yang sedang dilaksanakan (2) Kelompok sasaran yaitu kelompok masyarakat yang diambil menjadi sasaran yang ditetapkan untuk manfaat dari program, perbaikan dan perubahan (3) penerapan elemen yang baik untuk organisasi atau individu yang bertanggung jawab akan memperoleh pelaksanaan dan pengawasan proses implementasi.

Sebaiknya proses implementasi keseluruhan dapat dievaluasi dengan cara kebijakan mengukur atau membandingkan hasil akhir antara program tersebut dengan tujuan-tujuan kebijakan untuk memberikan pelayanan atau mengatur kelompok sasaran. Kebijakan implementasi yang sederhana yang melibatkan satu yang berfungsi sebagai implementor. Kebijakan implementasi dalam keberhasilan akan ditentukan banyaknya variabel atau faktor yang masing-masing saling berhubungan satu sama yang lain. Adanya kebijakan implementasi yang melaksanakan kepemimpinan untuk pelaksanaan agar pengendalian melakukan pelaksanaan dengan rinci dimulai dari implementasi strategi, organisasi, pengendalian akan berjalan dengan lancar sesuai yang diinginkan.

2.1 Data Mining

Data mining adalah suatu istilah yang diproses menggunakan teknik statistik, menguraikan penemuan pengetahuan didatabase. Data mining adalah proses yang menggunakan teknik kecerdasan buatan, matematika, *machine learning* untuk mengekstraksi dan identifikasi informasi yang manfaat dan pengetahuannya yang terkait dari beberapa database yang sangat besar. Menurut Gartner Group data mining adalah proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan teknik pengenalan pola seperti teknik statistik dan matematika [3].

Data mining diartikan sebagai suatu proses ekstraksi informasi yang berguna dalam potensi data sekumpulan yang merupakan dalam suatu basis data. Data mining mencari informasi dari data dasar yang sangat besar, dapat dianalogikan dengan teknologi ini digunakan untuk prediksi dan sifat dimana data mining mencari informasi memprediksi dalam basis data [3]. *Berry dan linoff* dalam bukunya *data mining technique for marketing, Sales, and Customers Support* mendefenisikan data mining sebagai proses eksplorasi dan analisis secara otomatis terhadap dalam data jumlah besar agar tujuan menemukan aturan dan pola yang melekat dapat berupa korelasi dikontrol, selama konstruksi desain eksperimental. Data Mining yaitu memiliki aktifitas dan bukanlah sebuah algoritma ataupun program. Dalam penggunaan aktifitas data mining maka sering kali digunakan diberbagai teknik disiplin ilmu misalnya statistik ataupun *machine learning*. Data mining sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi dengan seluruh permukaan relasi-relasi pada komplek data yang sangat besar atau pola-pola yang tersembunyi dengan manfaat data dalam volume yang besar dengan menggunakan pendekatan *discovery-based* dimana pola yang cocok dan algoritma lain untuk menentukan relasi kunci dalam data yang dieksplorasi baru pada arsitektur sistem perusahaan. Agar lebih jelasnya memahami data mining, dapat dilihat pada gambar.

2.2 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma klasifikasi data mining yang diterapkan pada teknik *decision tree*. C5.0 merupakan algoritma sebelumnya yang dibentuk oleh *Ross Quinlan* pada tahun 1987. Algoritma ini pemilihan atribut yang diproses menggunakan *information gain*. Pemilihan atribut untuk memecah objek dalam beberapa kelas harus dipilih atribut yang dihasilkan *information gain*. Atribut dengan nilai tertinggi akan dipilih sebagai parent bagi *node* selanjutnya [5].

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i \log_2(p_i) \dots \dots \dots (1)$$

Keterangan:

S = himpunan kasus

S₁ = jumlah sampel

P₁ = proporsi kelas

Untuk mendapatkan informasi nilai subset dari atribut A tersebut maka digunakan formula dibawah ini :

$$E(A) = \sum_{j=1}^y \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \dots \dots \dots (2)$$

Ket : $\frac{S_{1j} + \dots + S_{mj}}{S}$ = jumlah subset j yang dibagi dengan jumlah sampel S

Untuk mendapatkan nilai *gain* selanjutnya digunakan formula dibawah ini :

$$Gain A = I(S_1, S_2, \dots, S_m) - E \dots \dots \dots (3)$$

Keterangan:

A = atribut

S = himpunan kasus

S1 = jumlah sampel

Algoritma C5.0 adalah menyempurnakan dari algoritma yang dibentuk oleh Ross Quinlan pada tahun 1987. Dalam algoritma C5.0 pemilihan atribut yang akan diproses menggunakan ukuran huruf *information gain* dan memiliki ukuran *information gain* digunakan untuk memilih atribut uji pada setiap *node* pada pohon keputusan.

2.3 Pengelompokkan Data

Data adalah kumpulan data atas catatan atau sekumpulan keterangan atau fakta mentah berupa simbol, angka, kata-kata ataupun citra melalui proses pengamatan atau pencarian ke sumber-sumber tertentu. Data merupakan bentuk jamak dari **datum**, berasal dari bahasa latin, yang arti nya "sesuatu yang diberikan". Keterangan suatu deskripsi dari suatu hal yang diperoleh dari dasar hasil pengamatan dan dapat diolah menjadi bentuk yang kompleks seperti informasi, database atau solusi masalah tertentu. Untuk lebih memahami dan apa saja pengelompokkan dalam data, sebagai berikut [5]:

Jenis data Menurut Cara Memperolehnya

- a. Data Primer
Data Primer adalah yang diambil secara langsung dari objek penelitian oleh perorangan atau organisasi.
- b. Data Sekunder
objek penelitian data yang didapat tidak langsung dari penelitian data yang sudah mendapatkan data kumpulan oleh pihak lain dengan cara atau metode baik secara komersial ataupun non komersial.
2. Macam - Macam Data Berdasarkan Sumber Data
 - a. Data internal
Data internal adalah data situasi dan kondisi pada suatu organisasi yang menggambarkan secara internal.
 - b. Data Eksternal
Data eksternal adalah data yang menggambarkan situasi dan kondisi yang ada diluar organisasi.
3. Klasifikasi Data Berdasarkan Jenisnya
 - a. Data Kuantitatif
Data kuantitatif adalah dalam bentuk kata-kata yang mengandung makna atau data yang dipaparkan.
 - a. Data Kualitatif
Data Kualitatif dalam bentuk kata-kata yang mengandung makna atau data yang disajikan
4. Pembagian Jenis Data Berdasarkan Sifat Data
 - a. Data Diskrit
Data Diskrit adalah data yang nilainya adalah bilangan asli.
 - b. Data Kontinyu
Data kontinyu adalah suatu interval tertentu yang berada pada nilai data yang nilainya ada pada nilai satu kenilai yang lain.
5. Jenis-jenis Data Menurut Waktu Pengumpulannya
 - a. Data Cross Section
Data Cross section adalah data yang menunjukkan pada titik waktu.
 - b. Data Times Series/ berkala
Data berkala adalah datanya yang menggambarkan sesuatu dari waktu ke waktu atau periode secara historis.

3. HASIL DAN PEMBAHASAN

Berhubungan dengan klasifikasi memprediksi kelahiran bayi dilakukan analisis data. Proses klasifikasi dalam memprediksi kelahiran bayi terbagi menjadi beberapa kriteria. Dari data kriteria yang telah diperoleh, maka akan ditentukan apakah keputusan yaitu bayi dilahirkan normal dan dilahirkan secar. Data yang menjadi variabel penentu dalam pembentukan pohon keputusan pada penelitian ini adalah Keluhan, Tekanan Darah, Kenaikan Berat Badan Ibu, Tinggi Fundus, Keadaan Fisik, Kondisi Lahiran. Pemilihan variabel penentu tersebut dengan pertimbangan bahwa variabel-variabel yang dipilih merupakan

variabel yang paling mempengaruhi keputusan pengelompokkan data kondisi kelahiran. Masalah sarana penunjang menjadi prioritas utama dalam perbaikan mutu kesehatan bayi diperiode selanjutnya. Masalah yang dihadapi adalah proses penilaian kepada yang hal tertentu sebagai bahan acuan dalam penilaian memprediksi kondisi kelahiran bayi. Setelah melakukan analisa masalah, selanjutnya menentukan variabel input data. Pada algoritma C5.0 akan dilakukan *decision tree*, dengan memilih atribut pada *field-field* data sebagai akar, membuat cabang untuk masing-masing nilai. Membagi kasus dalam cabang, dan mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama. Proses pengambilan keputusan dalam klasifikasi data untuk memprediksi kondisi kelahiran dapat dikelompokkan menjadi beberapa kriteria. Dari data yang telah diperoleh, maka akan ditentukan variabel data yang menjadi variabel penentu keputusan yaitu, menentukan kesehatan ibu sedang mengandung sampai melahirkan. Proses *klasifikasi* data kondisi kelahiran menjadi beberapa kriteria, berdasarkan variabel-variabel yang sudah dipilih dari 8 data sampel, *format* data menjadi seperti tabel 1:

Tabel 1. Kriteria Keluhan

Keluhan	Kategori
Mual, Muntah, Pusing, Badan terasa lemas, Kaki bengkok, Pinggang terasa sakit	Normal
Pendarahan	Tidak Normal

Tabel 2. Kriteria Tekanan Darah

Tekanan Darah	Kategori
140 – 170	Tinggi
100 – 130	Normal
90 – 70	Rendah

Tabel 3. Kriteria Kenaikan Berat Badan Ibu

Kenaikan Berat Badan ibu	Kategori
17 kg – 20 kg	Ideal
20 kg – 26 kg	Tidak Ideal

Tabel 4. Kriteria Tinggi Fundus

Tinggi Fundus	Kategori
Setinggi pusat atau 2 jari di atas pusat	Ideal
$\frac{1}{2}$ pusat-prosessus xifoideus atau $\frac{1}{3}$ di bawah pusat	Tidak Ideal

Tabel 5. Kriteria Keadaan Fisik

Keadaan Fisik	Kategori
Tidak ada keluhan	Sehat
Pembengkakan yang berlebihan	Tidak Sehat

3.1 Penerapan Algoritma C 5.0

Untuk melihat atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* digunakan rumus (1) sedangkan untuk menghitung nilai *Entropy* dapat dilihat pada rumus (2). Dengan menggunakan dua persamaan tersebut maka akan didapat nilai *Entropy* pada *Gain* yang digunakan sebagai akar dalam membuat pohon keputusan. Menghitung jumlah kasus, jumlah untuk keputusan, dan kasus yang dibagi berdasarkan atribut nama, keluhan, tekanan darah, kenaikan berat badan ibu, tinggi fundus, keadaan fisik, dan kondisi kelahiran. Setelah itu lakukan perhitungan *Gain* untuk setiap atribut.

1. $Entropy\ Total = Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$
 $Entropy\ Total$
 $= ((-4/10 * \log_2 (4/10)) + (-6/10 * \log_2 (6/10)))$
 $= 0,970951$
 - a. Nilai *Entropy* Keluhan = Normal
 $Entropy\ Keluhan$
 $Entropy\ atribut\ "Normal"$
 $= ((-1/6 * \log_2 (1/6)) + (-5/6 * \log_2 (5/6)))$
 $= 0,650022$
 - b. *Entropy* Tekanan Darah

$$\begin{aligned} \text{Nilai atribut "Tinggi"} \\ &= ((-0/1 * \log_2 (0/1) + (-1/1 * \log_2 (1/1))) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Normal"} \\ &= ((-1/5 * \log_2 (1/5) + (-4/5 * \log_2 (4/5))) \\ &= 0,464386 \end{aligned}$$

c. *Entropy* Kenaikan Berat Badan Ibu

$$\begin{aligned} \text{Nilai atribut "Ideal"} \\ &= ((-0/3 * \log_2 (0/3) + (-3/3 * \log_2 (3/3))) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Tidak Ideal"} \\ &= ((-1/3 * \log_2 (1/3) + (-2/3 * \log_2 (2/3))) \\ &= 0,528321 \end{aligned}$$

d. *Entropy* Tinggi Fundus

$$\begin{aligned} \text{Nilai atribut "Ideal"} \\ &= ((-1/5 * \log_2 (1/5) + (-4/5 * \log_2 (4/5))) \\ &= 0,464386 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Tidak Ideal"} \\ &= ((-0/1 * \log_2 (0/1) + (-1/1 * \log_2 (1/1))) \\ &= 0 \end{aligned}$$

e. *Entropy* Keadaan Fisik

$$\begin{aligned} \text{Nilai atribut "Sehat"} \\ &= ((-1/6 * \log_2 (1/6) + (-5/6 * \log_2 (5/6))) \\ &= 0,430827 \end{aligned}$$

$$\begin{aligned} \text{Nilai atribut "Tidak Sehat"} \\ &= ((-0/0 * \log_2 (0/0) + (-0/0 * \log_2 (0/0))) \\ &= 0 \end{aligned}$$

Sementara itu, nilai *gain* pada harga dan jaringan dihitung dengan persamaan sebagai berikut:

2. Nilai *Gain*

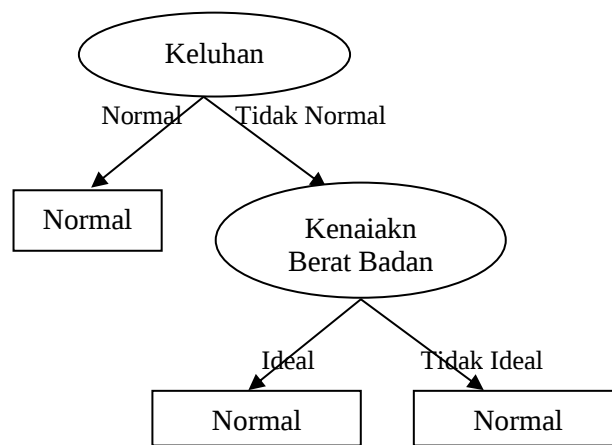
- a. Nilai *Gain* Tekanan Darah
 $= 0,650022 - ((-1/6) * 0) + ((5/6) * 0,464386)$
 $= 0,263033$
- b. Nilai *Gain* Kenaikan Berat Badan Ibu
 $= 0,650022 - ((-3/6) * 0) + ((3/6) * 0,528321)$
 $= 0,3858615$
- c. Nilai *Gain* Tinggi Fundus
 $= 0,650022 - ((-5/6) * 0,464386) + ((1/6) * 0)$
 $= 0,37010$
- d. Nilai *Gain* Keadaan Fisik
 $= 0,650022 - ((-6/6) * 0,430827) + ((0/6) * 0)$
 $= 0,080848$

Tabel 6. Node 1.1

Node	Keterangan	Jumlah kasus (S)	Sum(Secar)	Sum(Normal)	Entropy	Gain
1.1	Keluhan = Normal	6	1	5	0,650022	
	Tekanan Darah					0,263033
	Tinggi	1	0	1	0	
	Normal	5	1	4	0,464386	
	Rendah	0	0	0	0	
	Berat Badan					0,3858615
	Ideal	3	0	3	0	
	Tidak Idela	3	1	2	0,528321	

Node	Keterangan	Jumlah kasus (S)	Sum(Secar)	Sum(Normal)	Entropy	Gain
Tinggi Fundus	Ideal	5	1	4	0,464386	0,37010
	Tidak Idela	1	0	1	0	
Keadaan Fisik	Sehat	6	1	5	0,430827	0,080848
	Tidak Sehat	0	0	0	0	

Tabel di atas menunjukkan bahwasanya kriteria jumlah Kenaikan Berat Badan Ibu memiliki nilai *Gain* yang paling tinggi. Untuk *fase* selanjutnya adalah pembentukan *Tree* (pohon keputusannya). Berikut ini adalah tree dari nilai Entropy dan Gainnya ditunjukkan pada 6:



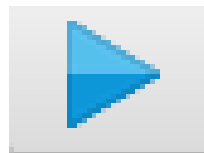
Gambar 1. Hasil Keputusan

Dari hasil pohon keputusan tersebut, didapatkan rule aturan sebagai berikut :

1. Jika Keluhan = Normal maka Kelahiran = Normal
2. Jika Keluhan = Tidak Normal dan Kenaikan Berat Badan Ibu Ideal maka Kelahiran = Normal
3. Jika Keluhan = Tidak Normal dan Kenaikan Berat Badan Ibu Tidak Ideal maka Kelahiran = Caesar

3.2 Hasil Pengujian

Setelah melakukan perancangan pada aplikasi Rapidminer 10 maka pada tahap ini maka dilakukan pengujian melalui aplikasi *Rapidminer 9.3* dengan cara menekan tombol *run or resume the current process* atau dengan menekan *shortcut* tombol *F11* seperti tampilan gambar icon pada gambar dibawah ini :



Gambar 2. Icon run or resume the current process

Hasil akhir atau *output* dari aplikasi *RapidMiner* adalah berupa pohon keputusan atau *decision tree / id3*. Dari pohon keputusan inilah akan dihasilkan *rules* atau aturan-aturan yang dapat membantu dalam mengambil keputusan. Pohon keputusan akhir dapat dilihat pada gambar berikut:



Gambar 3. Pohon Keputusan Hasil Pengujian

Dari hasil pohon keputusan tersebut, didapatkan rule aturan sebagai berikut :

1. Jika Keluhan = Normal maka Kelahiran = Normal
2. Jika Keluhan = Tidak Normal dan Kenaikan Berat Badan Ibu Ideal maka Kelahiran = Normal
3. Jika Keluhan = Tidak Normal dan Kenaikan Berat Badan Ibu Tidak Ideal maka Kelahiran = Caesar

Selanjutnya setelah didapatkan rule dan juga pohon keputusan, kemudian menghitung tingkat akurasi yang dari proses klasifikasi. Adapun tingkat akurasi dapat dilihat pada gambar 4 berikut :

Table View Plct View

accuracy: 80.00%

	true Normal	true Caesar	class precision
pred. Normal	6	2	75.00%
pred. Caesar	0	2	100.00%
class recall	100.00%	50.00%	

Gambar 4. Hasil Tingkat Akurasi

Pada gambar 4 didapatkan bahwa tingkat akurasi dari pohon keputusan sebesar 100%.

4. KESIMPULAN

Dari hasil penelitian yang penulisan lakukan disimpulkan bahwa algoritma C5.0 dan pengujian *RapidMiner* dapat membantu klasifikasi. Proses untuk menghasilkan kombinasi item menggunakan algoritma C5.0 dilakukan dengan menerapkan nilai *Entropy* dan nilai *Gain* tertinggi. Untuk membentuk pohon keputusan. Pengujian yang telah dilakukan dapat disimpulkan bahwa *data mining* menggunakan metode pohon keputusan algoritma C5.0 untuk membentuk pohon keputusan dalam mengklasifikasi memprediksi kondisi kelahiran menggunakan aplikasi *RapidMiner* dan mendapatkan hasil yang maksimal.

REFERENCES

- [1] Jihan Ratnasari, Ibnu Asror, S.T, M. T, Dr. Moch Arif Bijaksana, Ir. M. Tech, Analisis Implementasi Sistem OLAP dan Klasifikasi Ketepatan Waktu Lulus dan Undur Diri Mahasiswa S1 Tektik Informatika Universitas Telkom Menggunakan Decision Tree C5.0, e-Proceeding of Engineering: Vol.5, No.2 Agustus 2018, ISSN : 2355-9365
- [2] Kurniawan I, Saputra R A. Penerapan Algoritma C5.0 Pada Sistem Pendukung Keputusan Kelayakan Penerimaan Beras Masyarakat Miskin. JURNAL INFORMATIKA, Vol.4 No.2 September 2017, pp. 236-240
- [3] Eko Prasetyo. Data Mining Mengelolah Data Menjadi Informasi Menggunakan Matlab. Andi. 2014. Yogyakarta.
- [4] Larose D, T., Data Mining Methods and Models, 2006, Jhon Wiley & Sons, Inc. Hoboken New Jersey.
- [5] Yogi Yusuf W. Membandingkan performasi algoritma decision tree C5.0 CART dan CHAID untuk kasus memprediksi status resiko kredit disuatu bank.2007
- [6] Irmawati, Bayi dan Balita Sehat & Cerdas, PT Elex Media Komputindo,2015, Jakarta.
- [7] D. A. C, D. A. Baskoro, L. Ambarwati and I. W. S Wicaksana, Belajar Data Mining Dengan Rapid Miner, 2013.