

Status Proses Persalinan Menggunakan Algoritma C4.5

Mariza Putri Sari Dewi Ningsih^{*1)}, Beta Noranita²⁾

**Jurusan Ilmu Komputer/Informatika, Fakultas Sains dan Matematika
Universitas Diponegoro

¹⁾ smramarizaputri@gmail.com, ²⁾ betanoranita.undip@gmail.com

Abstrak

Setiap orangtua pasti berusaha yang terbaik untuk persalinan akan berjalan lancar dan dapat melahirkan bayi dengan sempurna. Keberadaan penentuan status proses persalinan menjadi penting untuk mempersiapkan dalam hal kesehatan, dan psikologi. Proses persalinan pada seorang ibu dapat ditempuh secara normal, akan tetapi sering kali terdapat beberapa faktor kesehatan yang membuat proses persalinan seorang ibu dilakukan dengan operasi caesar. Aplikasi ini dibangun untuk memberikan saran dalam menentukan proses persalinan yang harus dijalani oleh seorang ibu hamil untuk meningkatkan keselamatan ibu dan bayi. Metode pada aplikasi data mining penentuan status proses persalinan dengan menggunakan algoritma C4.5. Algoritma C4.5 dipilih karena dapat menseleksi atribut dan mudah diinterpretasikan pada aturan yang terbentuk. Aplikasi ini merupakan penggabungan teknologi informasi dan bidang kesehatan untuk membantu para ibu hamil dalam menentukan status proses persalinan normal atau operasi caesar berdasarkan kondisi kesehatan ibu dan janin. Atribut yang digunakan yaitu usia, riwayat penyakit, tekanan darah, urutan kehamilan, jarak kelahiran, riwayat caesar, gawat janin, kelainan letak, berat bayi, dan plasenta provera. Data rekam medis yang digunakan berjumlah 682 record dengan data training berjumlah 545 record dan data testing berjumlah 137 record data. Hasil menunjukkan bahwa algoritma C4.5 mempunyai akurasi prediksi maksimum untuk penentuan status persalinan sebesar 97,08%, precision 96%, recall 88,89%, F-Measure 92,30%, G-mean 0,0894 dan AUC 0,93995.

Kata kunci: Algoritma C4.5, Data Mining, Pohon Keputusan, Status Persalinan

Abstract

Every parent must be make the best effort for childbirth to go well and be able to deliver the baby perfectly. The existence of determination of childbirth process becomes important to prepare in terms of health, and psychology. The process of childbirth in a mother can be taken normally, but there are often of health factors that make the childbirth by cesarean section. This application is built to provide advice in determining of childbirth that a pregnant woman should undergo to improve the safety of mother and baby. Method of data mining application to determine of childbirth process using C4.5 algorithm. The C4.5 algorithm is selected because it can selects atributs and is easy to interpret on rules. This application is a combination of information technology and health field to assist pregnant mother to decided of childbirth as normal or cesarean section based on maternal and fetal health conditions. Atributs used are age, history of disease, blood pressure, gravida, birth spacing, cesarean history, fetal distress, placental abnormality, infant weight, and placenta previa. Medical record data used amounted to 682 records with training data amounted to 545 records and data testing amounted to 137 records data. The results show that the C4.5 algorithm has maximum predictive accuracy for the determination

of childbirth status 97,08%, precision 96%, recall 88,89%, F-Measure 92,30%, G-mean 0,0894 and AUC 0,93995.

Keywords : C4.5 algorithm, data mining, decision tree, determination of childbirth.

1 PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI) Tahun 2012 Angka Kematian Ibu (AKI) di Indonesia masih termasuk tinggi yaitu 395 per 100.000 kelahiran hidup. Upaya penurunan Angka Kematian Bayi (AKB) dan AKI di Indonesia salah satunya melalui Program Perencanaan Persalinan dan Pencegahan Komplikasi (P4K). Program ini membutuhkan peran keluarga dalam upaya deteksi dini dan menghindari resiko kesehatan pada ibu hamil. Pelaksanaan P4K diharapkan mampu membantu keluarga membuat perencanaan persalinan yang baik dan mengambil tindakan yang tepat [1].

Persalinan merupakan kejadian fisiologi yang normal dialami oleh seorang ibu berupa pengeluaran hasil konsepsi yang hidup dalam uterus melalui vagina ke dunia luar yang disebut persalinan normal. Beberapa kasus seperti gawat janin, kelainan letak janin, tekanan darah, usia ibu, kondisi janin, ukuran pinggul, jarak kelahiran terlalu dekat dan ukuran janin dapat meningkatkan resiko kematian pada ibu dan bayi sehingga diperlukan alternatif lain dengan mengeluarkan konsepsi melalui pembuatan sayatan pada dinding uterus melalui dinding perut yang disebut *sectio caesaria*.

Seiring dengan perkembangan teknologi dan informasi yang sangat cepat, pemanfaatan teknologi informasi dapat ditemukan pada berbagai bidang, salah satunya bidang kesehatan. Melalui data rekam medis persalinan dapat dilakukan

proses *data mining* melalui perangkat lunak yang membantu pendukung keputusan penentuan status proses persalinan dalam meningkatkan pelayanan kesehatan serta menurunkan resiko angka kematian ibu dan bayi. Melalui proses *data mining*, diharapkan dapat digali suatu potensi yang lebih dari sekedar informasi data rumah sakit saja tetapi juga dapat menganalisis penentuan status proses persalinan kepada seorang ibu dan bayinya.

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik *decision tree* yang merupakan perbaikan dari algoritma ID3, karena dapat menghilangkan bias pada data. Kelebihan C4.5 adalah dapat mengolah data dalam jumlah besar, menghasilkan aturan – aturan yang mudah direpresentasikan dan tercepat diantara algoritma lainnya dan dapat menangani *missing* data. Pada penelitian ini akan dibahas klasifikasi yang dimaksud adalah mengenai penentuan status proses persalinan pada kehamilan seorang ibu yang diambil dari data persalinan pada suatu pusat pelayanan kesehatan.

Berdasarkan permasalahan diatas, maka dapat dikembangkan sebuah teknik aplikasi *data mining* dalam menentukan status proses persalinan menggunakan algoritma C4.5 dengan metode *decision tree*. Diharapkan dengan adanya perancangan aplikasi ini dapat mendukung pengambilan keputusan status proses persalinan pada kehamilan seorang ibu sekaligus membantu merencanakan persalinan untuk tindakan yang tepat ibu hamil untuk meminimalisir

AKI (Angka Kematian Ibu) dan AKB (Angka Kematian Bayi).

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang yang telah dijelaskan, maka dapat dibuat rumusan masalah yang dibahas dalam penelitian ini adalah bagaimana merancang dan membangun aplikasi *data mining* dalam menentukan status proses persalinan menggunakan algoritma C4.5.

1.3 TUJUAN DAN MANFAAT

Tujuan dari penelitian ini adalah untuk menghasilkan sebuah aplikasi *data mining* untuk menentukan status proses persalinan yang berfungsi secara baik serta dapat digunakan oleh pengguna umum untuk mendukung keputusan perencanaan persalinan.

Manfaat dilaksanakannya penelitian ini adalah:

1. Hasil dari aplikasi dapat digunakan oleh masyarakat umum dan penyedia pelayanan kesehatan untuk menjadi saran bahan pertimbangan dalam perencanaan status proses persalinan kepada ibu hamil.
2. Dapat meminimalisir angka kematian ibu (AKI) dan angka kematian bayi (AKB) setiap tahunnya secara bertahap.

1.4 RUANG LINGKUP

Ruang lingkup pembangunan aplikasi yang akan dikembangkan adalah aplikasi *data mining* dalam menentukan status proses persalinan menggunakan algoritma C4.5.

- 1 Data masukan berupa rekam medis persalinan yang terdapat keterangan tekanan darah, kondisi janin, perkiraan berat janin, posisi bayi, riwayat caesar, riwayat penyakit, usia ibu, jarak kelahiran, urutan kehamilan, *plasenta*

provia dan hasil persalinan yang dilakukan.

2. Dari hasil perhitungan didapatkan data output berupa saran mengenai penentuan status proses persalinan untuk mendukung persiapan persalinan seorang pasien.

2 TINJAUAN PUSTAKA

2.1 PENGERTIAN PERSALINAN

Persalinan / *partus* merupakan suatu kejadian fisiologi yang memungkinkan serangkaian perubahan yang besar pada ibu untuk dapat melahirkan janin melalui jalan lahir. Persalinan juga merupakan suatu proses pengeluaran hasil konsepsi yang dapat hidup, dari dalam *uterus* melalui *vagina* atau jalan lain ke dunia luar. Kejadian berakhir dengan pengeluaran bayi yang cukup bulan disusul dengan pengeluaran *plasenta* dan selaput janin [2].

2.2 JENIS PERSALINAN

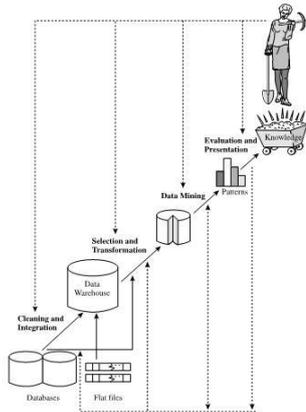
Persalinan normal menurut *World Health Organization (WHO)* 2010 adalah persalinan yang dimulai secara spontan atau dengan kekuatan ibu sendiri dan melalui jalan lahir, beresiko rendah pada tahap awal persalinan dan persentasi belakang kepala pada usia kehamilan 37-42 minggu dengan keadaan ibu maupun bayi dalam kondisi [2].

Proses persalinan bantuan terjadi jika berlangsung dengan bantuan tenaga dari luar misalnya ekstraksi dengan *forceps* atau melakukan operasi *section caesarea*. Operasi caesar adalah operasi pembedahan untuk melahirkan bayi melalui perut ibu.

2.3 DATA MINING

Data mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar [3]. Proses *data mining*

meliputi tugas *preprocessing* seperti ekstraksi data, *data cleaning*, *data fusion*, *data reduction*, *feature construction* serta proses terakhir berupa penemuan model penafsiran yang bersifat cenderung berulang dan interaktif [4].



Gambar 2.1. Tahap Data Mining [3]

Terdapat beberapa macam teknik pada *data mining* salah satunya adalah *classification*. *Classification* adalah proses untuk menemukan model yang menjelaskan kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model tersebut dapat menjadi suatu aturan “jika-maka” berupa *decision tree*. Proses klasifikasi biasanya dibagi menjadi dua fase, yaitu *learning* dan *testing*.

2.4 DECISION TREE

Decision tree merupakan himpunan kemungkinan yang membentuk jika maka (*If – Then*). *Entropy* digunakan untuk menentukan yang manakah *node* yang akan menjadi pemecah data latih berikutnya.

$$Entropy(S) = - \sum_{i=1}^m (p_i * \log_2 p_i) \dots \dots \dots (2.1)$$

Keterangan :
 S : total kasus
 m : jumlah partisi S
 p_i : probabilitas dari S_i terhadap S, dengan S_i adalah jumlah kasus partisi ke-i

Langkah berikut untuk menghitung *gain*, rumus dari pada *gain* adalah sebagai berikut [3]:

$$Entropy_A(S) = \sum_{j=1}^v \frac{|S_j|}{|S|} * Entropy(S_j) \dots \dots \dots (2.2)$$

$$Gain(A) = Entropy(S) - Entropy_A(S) \dots (2.3)$$

Proses perhitungan *info gain* juga dapat ditulis seperti pada persamaan (2.4) berikut ini.

$$Gain(A) = Entropy(S) - \sum_{j=1}^v \frac{|S_j|}{|S|} * Entropy(S_j) \dots (2.4)$$

Keterangan :

- S : total kasus
- A : atribut
- v : jumlah partisi atribut A
- |S_j| : jumlah kasus pada partisi ke-j
- |S| : jumlah kasus dalam S

2.5 ALGORITMA C4.5

Algoritma C4.5 mempunyai prinsip kerja yang sama dengan ID3. Hanya saja dalam algoritma C4.5 proses pemilihan atribut dilakukan dengan menggunakan *gain ratio* dengan rumus yang ditunjukkan pada persamaan 2.5 [3].

$$GainRatio(A, D) = \frac{Gain(A)}{SplitInfo_A(D)} \dots \dots \dots (2.5)$$

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|S_j|}{|S|} * \log_2 \frac{|S_j|}{|S|} \dots \dots \dots (2.6)$$

Keterangan:
 A : atribut
 v : jumlah partisi atribut A
 |S_j| : jumlah kasus pada partisi ke-j
 |S| : jumlah total kasus

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah [5]:

1. Memilih atribut sebagai akar
2. Membuat cabang (simpul internal) untuk masing-masing nilai
3. Membagi kasus dalam cabang
4. Mengulangi proses masing-masing cabang hingga semua kasus pada cabang memiliki kelas yang sama. Atribut yang telah dipilih dan memiliki kelas tidak diikuti lagi dalam perhitungan nilai pemilihan atribut akar.

Setelah proses dijalankan maka akan menghasilkan *decision tree* lalu membentuk suatu *rule*. Melalui *rule* yang terbentuk dapat menentukan aturan keputusan.

2.6 PENGUKURAN KINERJA

Terdapat beberapa istilah yang digunakan pada perhitungan pengukuran kinerja. Proses ini akan mempermudah untuk memahami berbagai langkah.

Tabel 2.1. Confusion Matriks

		Actual Class (Asli)		Total
		T	F	
Predicted Class (Prediksi)	T	TP	FP	P'
	F	FN	TN	N'
Total		P	N	P + N

Precision

$$= \left(\frac{TP}{TP + FP} \right) \dots \dots \dots (2.7)$$

Recall

$$= \left(\frac{TP}{TP + FN} \right)$$

$$= \left(\frac{TP}{P} \right) \dots \dots \dots (2.8)$$

Accuracy

$$= \left(\frac{TP + TN}{TP + TN + FP + FN} \right)$$

$$= \left(\frac{TP + TN}{P + N} \right) \dots \dots \dots (2.9)$$

Menurut (Zhang & Wang, 2011) pada data tidak seimbang, akurasi lebih didominasi oleh ketepatan pada data kelas minoritas. Sehingga metrik yang tepat adalah AUC (*Area Under the ROC Curve*), *F-Measure*, *G-Mean*, akurasi keseluruhan, dan akurasi untuk kelas minoritas. Akurasi kelas minoritas

dapat menggunakan metrik TPrate/recall (sensitivitas). *G-Mean* dan AUC merupakan evaluasi prediktor yang lebih komprehensif dalam konteks ketidakseimbangan [6].

F – Measure

$$= \left(\frac{2 * Recall * Precission}{Recall + Precission} \right) \dots \dots \dots (2.10)$$

G – Mean

$$= \left(\sqrt{Sensitivity * Specificity} \right) \dots \dots \dots (2.11)$$

Sensitivity (TPR)

$$= \left(\frac{TP}{TP + FN} \right) \dots \dots \dots (2.12)$$

Specificity (FPR)

$$= \left(\frac{FP}{TN + FP} \right) \dots \dots \dots (2.13)$$

AUC memberikan ukuran skalar dari kinerja pengklasifikasi dan telah banyak digunakan dalam domain yang tidak seimbang. Ukuran AUC dihitung sebagai luas daerah kurva ROC menggunakan persamaan [7]:

Area Under Curve (AUC)

$$= \left(\frac{1 + TPrate - FPrate}{2} \right) \dots \dots \dots (2.14)$$

Menurut Gorunescu analisis hasil klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok [8].

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

Receiver operating characteristic (ROC) adalah kurva menggambarkan alur dari *true positive rate (TPR)* terhadap *false positive rate (FPR)*. Secara teknis kurva ROC adalah grafik dua dimensi dimana tingkat TP di plot pada sumbu Y dan tingkat FP di plot pada sumbu X.

3 ANALISIS DAN PERANCANGAN

3.1 DESKRIPSI PERANGKAT LUNAK

Aplikasi *data mining* dalam menentukan status proses persalinan menggunakan algoritma C4.5 mengambil studi kasus di Puskesmas Kecamatan Grogol Petamburan Jakarta dan Bidan Lucia Wongso Jakarta.

Aplikasi ini akan digunakan oleh masyarakat umum dan administrator. Admin adalah pengguna yang dapat mengelola dan memasukan data pengujian kedalam *database*. Data yang telah dimasukan oleh admin akan dilanjutkan dengan menggunakan algoritma C4.5. Dari data pengujian ini akan menghasilkan *rule* yang akan digunakan pada penentuan status persalinan. Pengguna lainnya yaitu masyarakat umum dapat memasukan data sesuai dengan rekam medis pasien tersebut kedalam aplikasi, akan dicocokkan dengan *rule* yang telah terbentuk. Hasil akhirnya yaitu pengguna umum tersebut akan mendapatkan saran hasil status proses persalinan sesuai dengan rekam medis yang dimasukan.

3.2 PENGUMPULAN DATA

Data penelitian yang diambil adalah data yang telah menjalani persalinan dengan rentang kehamilan 36 – 41 minggu. Proses persalinan yang diambil adalah dengan kelahiran selamat sebanyak berjumlah 682 data. Berdasarkan 10 atribut yang telah memenuhi kriteria yaitu, usia ibu, riwayat penyakit ibu, tekanan darah, banyak kelahiran, tahun terakhir, riwayat caesar, gawat janin, kelainan letak, berat bayi, dan *plasenta provera*.

Penentuan kategori-kategori didapatkan berdasarkan dari hasil wawancara dengan bidan selaku tenaga kesehatan dan buku kesehatan Survei Demografi dan Kesehatan Indonesia (SDKI) Tahun 2002 - 2003 yang relevan dengan kategori.

Tabel 3.1. Nilai Kondisi Normal Atribut

No.	Nama Atribut	Keterangan Atribut	Satuan	Nilai Normal
1.	Usia ibu	Usia seorang ibu	Tahun	18 - 35
No.	Nama Atribut	Keterangan Atribut	Satuan	Nilai Normal
2.	Jarak Kelahiran	Tahun terakhir ibu melahirkan	Tahun	2
3.	Riwayat Penyakit	Riwayat penyakit berat kelainan jantung, tuberkolosis, kelainan ginjal, kencing manis, anemia, kelainan darah, TBC, dan lain - lain	-	-
4.	Tekanan Darah	Tekanan darah ibu hamil saat pengecekan terakhir	mm Hg	Normal
5.	Banyak Kelahiran	Urutan hamil diakhiri janin lahir dan telah memenuhi syarat kehidupan.	Kelahiran	1 - 3

6.	Riwayat Caesar	Riwayat caesar yang dimiliki ibu saat kehamilan sebelumnya	-	-
7.	Gawat Janin	Kondisi bayi gawat janin jika kekurangan oksigen atau komplikasi lainnya	-	-
8.	Posisi Bayi	Kondisi letak bayi normal atau sungsang	-	-
9.	Berat Bayi	Kondisi berat bayi melalui ultrasonografi	gram (gr)	2500 - 4000
10.	Plasenta Proveria	Kelainan posisi janin pada rahim bagian bawah, menutupi jalan lahir	-	-

Berdasarkan nilai referensi diatas maka terbentuk 23 kategori yang akan disesuaikan dengan data yang diproses. Kategori data persalinan dapat dilihat pada Tabel 3.2.

Tabel 3.2. Tabel Kategori Atribut

No	Atribut	Nilai Kontinyu	Kategori
1.	Tekanan Darah	TD < Normal	Rendah
2.	Tekanan Darah	Normal <= TD <= Normal	Normal
3.	Tekanan Darah	TD > 120/80	Tinggi
4.	Usia Ibu	Umur < 18	Dini

5.	Usia Ibu	18 <= Umur <= 35	Normal
6.	Usia Ibu	Umur > 35	Lanjut
7.	Gawat Janin	-	Ya
8.	Gawat Janin	-	Tidak
9.	Berat Bayi	2500 < BB	Kecil
10.	Berat Bayi	2500 <= BB <= 4000	Normal
11.	Berat Bayi	BB > 4000	Besar
12.	Posisi Bayi	-	Normal
13.	Posisi Bayi	-	Sungsang
14.	Banyak Kelahiran	1 - 3	Cukup
15.	Banyak Kelahiran	Kelahiran >= 4	Banyak
16.	Jarak Kelahiran	Jarak < 2	Dekat
17.	Jarak Kelahiran	Jarak >= 2	Normal
18.	Riwayat Caesar	-	Ya
19.	Riwayat Caesar	-	Tidak
20.	Riwayat Penyakit	-	Ya
21.	Riwayat Penyakit	-	Tidak
22.	Plasenta Proveria	-	Ya
23.	Plasenta Proveria	-	Tidak

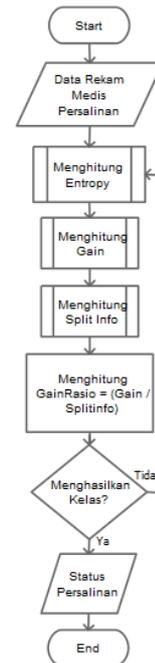
3.3 TAHAPAN DATA MINING

Analisis untuk mengolah data hingga siap digunakan harus melalui beberapa tahap perancangan *data mining* terlebih dahulu.

1. Pembersihan Data (*Cleanning*)
Melalui sejumlah data yang dikumpulkan terdapat beberapa data yang memiliki *field* kosong sehingga tidak akan digunakan. Selanjutnya data yang tidak relevan seperti jenis kelamin dan jumlah abortus tidak akan menjadi parameter atribut yang digunakan.
2. Integrasi Data (*Integration*)
Melalui kedua tempat ini akan digabungkan dari masing-masing basis datanya, menjadi data rekam medis yang dibutuhkan.
3. Seleksi Data (*Selection*)
Pada basis data tidak seluruhnya data digunakan, maka hanya data yang sesuai untuk analisis yang akan seleksi.
4. Transformasi Data (*Transformation*)
Transformasi data akan dikategorikan sesuai dengan kategori yang telah ditentukan pada Tabel 3.2.
5. Aplikasi Teknik *Data Mining*
Proses *mining* menggunakan metode *decision tree* dengan algoritma C4.5.
6. Evaluasi Pola dan Persentasi Pengetahuan
Pada tahap ini akan melihat apakah model yang telah terbentuk seperti pohon keputusan dan aturan (*rule*) sesuai atau tidak mendekati hasil yang diberikan oleh ahli kebidanan / kandungan.

3.4 PERANCANGAN ALGORITMA

Tahapan *flowchart* algoritma dari metode *Decision Tree* Algoritma C4.5 pada proses penentu status persalinan.



Gambar 3.1. Flowchart Algoritma C4.5

3.5 ANALISIS PROGRAM ALGORITMA

Sub aplikasi Algoritma C4.5 memiliki beberapa langkah sebagai berikut.

1. Menentukan data yang dipakai pada saat pelatihan. Transformasi data dengan kategori atribut berdasarkan dari nilai kategori yang telah ditentukan
2. Perhitungan *Entropy*
Berikut merupakan hasil realisasi untuk perhitungan nilai *entropy* total. Masing-masing total kasus diambil dari penerapan atribut data latih.

Total Kasus Persalinan _{Normal}	= 427
Total Kasus Persalinan _{Caesar}	= 118
Total Kasus	= 545

$$Entropy (semua) = 0,7538$$

$$= \left(\left(-\frac{427}{545} \right) * \log_2 \left(\frac{427}{545} \right) + \left(-\frac{118}{545} \right) * \log_2 \left(\frac{118}{545} \right) \right)$$

Perhitungan *entropy* atribut untuk perhitungan nilai kategori atribut tekanan darah normal. Masing-masing

total kasus diambil dari penerapan atribut data latih.

$$\text{Total Kasus Persalinan}_{\text{Normal|TD=Normal}} = 426$$

$$\text{Total Kasus Persalinan}_{\text{Caesar|TD=Normal}} = 103$$

$$\text{Total Kasus Kat_Atribut}_{\text{TD=Normal}} = 529$$

$$\text{Entropy}(td, normal) = 0,7112$$

$$= \left(\left(-\frac{426}{529} \right) * \log_2 \left(\frac{426}{529} \right) + \left(-\frac{103}{529} \right) * \log_2 \left(\frac{103}{529} \right) \right)$$

3. Perhitungan Info Gain

Perhitungan nilai gain atribut pada data rekam medis pasien yang diproses dijabarkan dengan contoh pada nilai atribut “tekanan darah” yang memiliki kategori ‘rendah’, ‘normal’, ‘tinggi’. Pada proses perhitungan gain, nilai entropy dari setiap kategori atribut harus diketahui.

$$\text{Total Tekanan_Darah}_{\text{Tinggi}} = 8$$

$$\text{Total Tekanan_Darah}_{\text{Rendah}} = 8$$

$$\text{Total Tekanan_Darah}_{\text{Normal}} = 529$$

$$\text{Hasil Entropy}_{\text{TD}_{\text{Rendah}}} = 0$$

$$\text{Hasil Entropy}_{\text{TD}_{\text{Normal}}} = 0,7112$$

$$\text{Hasil Entropy}_{\text{TD}_{\text{Tinggi}}} = 0,5436$$

$$\text{Gain}(semua, tekanandarah) = 0,0554$$

$$= \left(0,7538 - \left(\frac{8}{545} * 0 \right) + \left(\frac{529}{545} * 0,7112 \right) + \left(\frac{8}{545} * 0,5436 \right) \right)$$

4. Pehitungan SplitInfo

Nilai splitinfo atribut yang dihitung penerapannya adalah untuk tekanan darah.

$$\text{SplitInfo}(semua, tekanandarah) = 0,2205$$

$$\left(\left(-\frac{529}{545} \right) * \log_2 \left(\frac{529}{545} \right) + \left(-\frac{8}{545} \right) * \log_2 \left(\frac{8}{545} \right) + \left(-\frac{8}{545} \right) * \log_2 \left(\frac{8}{545} \right) \right)$$

5. Perhitungan Gain Ratio

Bentuk perhitungan gain ratio dari tekanan darah pada data rekam medis adalah sebagai berikut:

$$\text{Hasil Gain}(S, TD) = 0,0554$$

$$\text{Hasil SplitInfo}(S, TD) = 0,2205$$

$$\text{GainRatio}(semua, tekanan_darah) = \frac{0,0554}{0,2205} = 0,2514$$

6. Proses perhitungan entropy, gain, splitinfo, dan gainratio dilakukan pada semua atribut.

7. Pilih atribut sebagai akar dengan gain ratio tertinggi akan dipilih sebagai simpul. Atribut ‘riwayat penyakit’ memiliki gain ratio tertinggi akan dipilih sebagai root pada decision tree yang dibangun.

8. Buat cabang masing-masing nilai.

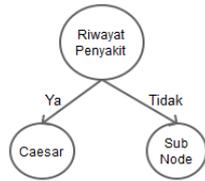
Tabel 3.3. Jumlah Kasus Riwayat Penyakit

Jenis		Kasus	Caesar	Normal
Riwayat Penyakit	Ya	26	26	0
	Tidak	519	92	427

Pada atribut (Riwayat Penyakit = Ya) memiliki jumlah kasus caesar 26 kasus dan normal 0 kasus, atribut (Riwayat Penyakit = Ya) memiliki keputusan kelas caesar. Selanjutnya pada atribut (Riwayat Penyakit = Tidak) memiliki jumlah kasus caesar 92 kasus dan normal 427 kasus, sehingga belum mendapatkan keputusan kelas dan masih perlu dilakukan perhitungan selanjutnya. Pada

atribut (Riwayat Penyakit = Tidak) akan dipilih suatu *subnode* menjadi cabang.

Hasil pembentukan *decision tree* dapat dilihat pada gambar 3.1.

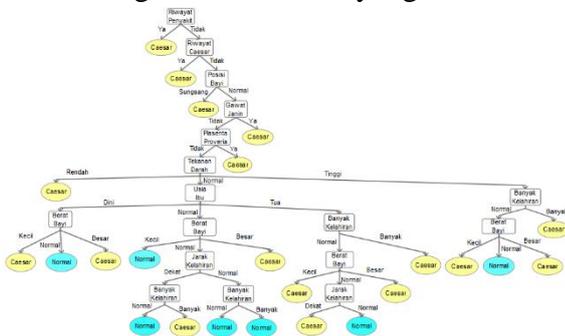


Gambar 3.2. Decision Tree Pemilihan Root

Setelah didapatkan suatu keputusan pada pohon maka akan membentuk suatu aturan dari pohon yang telah terbentuk.

Rule :
 IF r_penyakit = ya THEN keputusan = caesar

9. Bagi kasus dalam cabang. Pemisahan data yang dilakukan berdasarkan perhitungan sebelumnya. Sehingga seluruh data yang memiliki atribut (Riwayat Penyakit = Ya) tidak akan masuk pada perhitungan. Jadi hanya atribut dengan kategori (Riwayat Penyakit = Tidak) yang akan masuk pada perhitungan iterasi selanjutnya.
10. Ulangi untuk masing-masing cabang. Proses perulangan dilakukan dari perhitungan nilai *entropy*, *gain*, *splitinfo*, dan *gainratio* sampai semua kasus pada cabang memiliki kelas yang sama.



Gambar 3.3. Hasil Pohon

3.6 PERANCANGAN ANTARMUKA

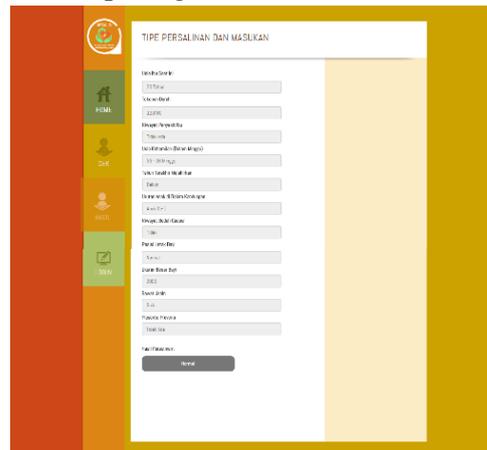
Rancangan antarmuka akan digunakan sebagai gambaran kepada perancang aplikasi dalam membentuk hasil saran persalinan.

Gambar 3.4. Rancangan Hasil Persalinan

4 IMPLEMENTASI

4.1 HASIL TAMPILAN SISTEM

Pada tahap ini merupakan implemetasi antarmuka yaitu halaman hasil saran persalinan pada gambar 4.1.



Gambar 4.1. Hasil APSALIN

4.2. HASIL PENGUKURAN KINERJA

Pada pengujian kebenaran aplikasi, dicoba dengan beberapa data rekam medis uji sebanyak 317 data. Perhitungan nilai akan diambil kasus caesar menjadi *actual class*. Kelas caesar menjadi kelas positif karena lebih dipentingkan oleh sistem harus dapat tepat mendeteksi pemberian kelas caesar.

Tabel 4.1. Confusion Matriks Pengukuran

		Asli		Total
		Caesar	Normal	
Prediksi	Caesar	TP 24	FP 1	25

	Norma 1	FN 3	TN 109	112
Total		27	110	137

Tabel 4.2. Hasil Akurasi C4.5

Pengukuran	Nilai
Akurasi (%)	97,0803
Precision	0,96
Recall	0,8889
F-Measure	0,9230
Sensitivity (TPR)	0,8889
Specificity (FPR)	0,0090
G-Mean	0,0894
AUC	0,9399

AUC (*Area Under the ROC Curve*) adalah ukuran kinerja yang populer dalam ketidakseimbangan kelas, nilai AUC yang tinggi menunjukkan perbedaan kinerja model yang lebih baik.

Pengukuran AUC menurut Gorunescu analisis hasil klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok. Hasil algoritma C4.5 termasuk klasifikasi sangat baik karena memiliki nilai AUC yang berada pada rentang 0.90-1.00.



Gambar 4.2. Kurva ROC C4.5

5 PENUTUPAN

5.1 KESIMPULAN

Berdasarkan hasil analisis dan pengujian yang telah dilakukan pada Tugas Akhir ini, maka dapat diambil beberapa kesimpulan, diantaranya:

1. Telah dibangun aplikasi *data mining* untuk prediksi penentuan status proses persalinan menggunakan algoritma C4.5. Melalui fungsi aplikasi yang telah berjalan sesuai kebutuhan yang didefinisikan pada analisis kebutuhan.

Melalui pengujian seluruh fungsi dapat dilihat bahwa pengujian seluruh fungsi telah diterima.

2. Pada aplikasi ini menggunakan 682 *record* dengan data *training* berjumlah 545 *record* dan data *testing* berjumlah 137 *record* data. Melalui data *testing* yang diujikan terdapat empat data *testing* dengan hasil berbeda dari keputusan asli.
3. Pada aplikasi penentuan status proses persalinan dengan Algoritma C4.5 memiliki tingkat *accuracy* sebesar 97,08%, nilai rata-rata *precision* 96%, *recall* 88,89%, *F-Measure* 92,30%, *G-mean* 0,0894 dan *AUC* 0,93995. Tingkat akurasi dapat berubah tergantung dengan data pelatihan dan data uji yang digunakan pada proses berjalan.

Hasil dari aplikasi *data mining* dalam menentukan status proses persalinan dengan menggunakan Algoritma C4.5 ini dapat dijadikan sebuah bahan pertimbangan untuk pengguna umum dalam menentukan proses persalinan yang akan dijalani.

5.2 SARAN

Saran yang dapat diberikan untuk penelitian selanjutnya adalah dapat dikembangkan terkait variabel yang digunakan, sehingga dapat menganalisis kesehatan pengguna umum secara lebih mendalam.

5.3 DAFTAR PUSTAKA

- [1] T. Kementerian Kesehatan, Info DATIN (Pusat Data dan Informasi Kementerian Kesehatan), Jakarta: Kementerian Kesehatan, 2014.
- [2] I. P. Damayanti and dkk, Buku Ajar Asuhan Kebidanan Komprehensif Pada Ibu Bersalinan dan Bayi Baru Lahir, Yogyakarta: Deepublish, 2014.

- [3] J. Han, M. Kamber and J. Pei, Data Mining Concept and Teqniques Third Edition, San Francisco: Morgan Kaufmann, 2012.
- [4] M. J. Zaki, Data Mining and Analysis (Fundamental Konsep and Algorithms, New York: Cambridge University, 2014.
- [5] S. P. Utari, "Implementasi Metode C4.5 Untuk Menentukan Guru," *Pelita Informatika Budi Darma*, pp. Volume : IX, Nomor: 3, 2015.
- [6] A. Saifudin and R. S. Wahono, "Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan," *Journal of Software Engineering, Vol. 1, No. 1, April 2015*, pp. 28-37, 2015.
- [7] M. Galar, A. Fernández, E. Barrenechea and F. Herrera, "EUSBoost: Enhancing ensembles for highly imbalanced datasets by evolutionary undersampling," *Pattern Recognition*, pp. 3460-3471, 2013.
- [8] F. Gorunescu, Data Mining: Concepts, Model, and Techniques Vol. 12, Berlin: Springer, 2011.

LAMPIRAN 1. HASIL PERHITUNGAN ITERASI PERTAMA

Jenis		Kasus	Caesar	Normal	Entropy	Gain	Split Info	Gain Ratio
Total		545	118	427	0,7538			-
Tekanan	Normal	529	103	426	0.7112			
Darah	Rendah	8	8	0	0	0,0554	0,2205	0,2514
	Tinggi	8	7	1	0.5436			
Berat Bayi	Besar	18	18	0	0			
	Kecil	29	11	18	0.9576	0,0839	0,5066	0,1657
	Normal	498	89	409	0.6773			
Gawat Janin	Tidak	536	109	427	0,7286			
	Ya	9	9	0	0	0,0372	0,1214	0,3064
Posisi Bayi	Normal	533	106	427	0,7197			
	Sungsang	12	12	0	0	0,0499	0,1526	0,3271
Riwayat Penyakit	Ya	26	26	0	0			
	Tidak	519	92	427	0,6741	0,1119	0,2766	0,4044
Usia	Muda	9	4	5	0,9911			
	Ideal	494	80	414	0,6389	0,1041	0,5112	0,2036
	Tua	42	34	8	0,7025			
Riwayat Proveria	Ya	3	3	0	0			
	Tidak	542	115	427	0,7456	0,0122	0,0492	0,2484
Jarak Kelahiran	Dekat	25	20	5	0.7219			
	Normal	520	98	422	0.6982	0,0544	0,2686	0,2026
Riwayat Caesar	Ya	24	24	0	0			
	Tidak	521	94	427	0,6810	0,1027	0,2605	0,3944
Urut Hamil	Normal	492	89	403	0.6820			
	Banyak	53	29	24	0.9936	0,0414	0,4602	0,0900