



## Optimasi Keputusan Metode Persalinan dengan Algoritma C4.5

Adinda Bunga Alfianah\*, Mudafiq Riyan Pratama, Muhammad Yunus, Veronica Vestine

Program Studi Manajemen Informasi Kesehatan, Jurusan Kesehatan, Politeknik Negeri Jember  
*adinda.alfianah@gmail.com, mudafiq.riyan@polije.ac.id, m.yunus@polije.ac.id, veronikavestine@polije.ac.id*

---

### Keywords:

*C4.5 Algorithm,  
Data Mining,  
Classification,  
Childbirth*

### ABSTRACT

*In 2018, the Indonesian Ministry of Health conducted a survey revealing that the rate of deliveries via Caesarean Section (C-Section) had exceeded the WHO's maximum standard of 17.6%. At Aulia Hospital in Pekanbaru, the prevalence of C-Sections reached 76% per 1000 births, significantly higher than the international benchmark. This study aims to analyze the factors influencing delivery methods using the C4.5 algorithm. The research employs a quantitative analytic approach with Secondary Data Analysis (SDA). A dataset of 500 records with 11 variables was utilized, including maternal age, gestational age, hypertension, hemoglobin, glucose levels, delivery history, fetal position, Cephalopelvic Disproportion (CPD), premature rupture of membranes (PROM), oligohydramnios, and estimated fetal weight (EFW). The C4.5 algorithm demonstrated 92% accuracy in predicting delivery methods, with a precision of 75%, indicating its ability to correctly identify necessary C-Sections. Furthermore, it achieved a recall of 100%, reflecting its effectiveness in identifying all actual C-Section cases. The rule tree analysis highlighted delivery history as the primary determinant in selecting the delivery method. These findings are expected to support medical decision-making processes regarding delivery methods and improve the management of high C-Section rates.*

---

### Kata Kunci

*Algoritma C4.5,  
Data Mining,  
Klasifikasi,  
Persalinan*

### ABSTRAK

Pada tahun 2018, Kementerian Kesehatan Indonesia melakukan survei yang menunjukkan jumlah kelahiran dengan metode *Sectio Caesarea* telah melebihi batas maksimal standar yang ditetapkan oleh WHO, yaitu 17,6%. Di Aulia Hospital Pekanbaru prevalensi persalinan *Sectio Caesarea* mencapai 76% per 1000 kelahiran, yang menunjukkan angka yang jauh lebih tinggi dibandingkan dengan standar internasional. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi metode persalinan menggunakan algoritma C4.5. Jenis penelitian ini adalah penelitian kuantitatif analitik dengan metode Analisis Data Sekunder (ADS). Dataset yang digunakan sebanyak 500 data dengan 11 variabel. Variabel-variabel ini mencakup usia ibu, usia kehamilan, hipertensi, hemoglobin, glukosa, riwayat *partus*, letak bayi, *Cephalopelvic Disproportion* (CPD), ketuban pecah dini (KPD), *oligohidramnion*, dan taksiran berat janin (TBJ). Hasil pengujian bahwa *accuracy* algoritma C4.5 dalam memprediksi metode persalinan yakni 92%. Nilai *precision* yang mengukur seberapa tepat prediksi algoritma menentukan persalinan *Sectio Caesarea* yang seharusnya dilakukan adalah sebesar 75%. Sementara itu, nilai *recall* yang mengukur kemampuan algoritma dalam memprediksi semua kasus persalinan *Sectio Caesarea* yang sebenarnya adalah 100%. Hasil dari *rule tree* menunjukkan bahwa riwayat *partus* merupakan kunci utama dalam penentuan persalinan. Temuan ini diharapkan dapat membantu dalam pengambilan keputusan medis terkait

---

metode persalinan dan mendorong pengelolaan yang lebih baik terhadap tingginya angka *Sectio Caesarea*.

---

**Korespondensi Penulis:**

Adinda Bunga Alfianah,  
Politeknik Negeri Jember,  
Jl. Mastrip PO BOX 164 Jember, Jawa Timur  
Telepon: +6285738426743  
Email: adinda.alfianah@gmail.com

**Submitted : 26-03-2024; Accepted : 23-09-2024;  
Published : 31-12-2024**

*Copyright (c) 2024 The Author (s)*  
*This article is distributed under a Creative Commons  
Attribution-ShareAlike 4.0 International License (CC BY-SA  
4.0)*

---

## 1. PENDAHULUAN

Persalinan adalah proses pengeluaran janin dan plasenta yang telah cukup bulan [1]. Dua jenis persalinan yang umum adalah persalinan pervaginam yang merupakan persalinan normal dan *Sectio Caesarea* yang merupakan persalinan melalui operasi perut [2]. Penggunaan *Sectio Caesarea* telah meningkat secara signifikan dalam dua dekade terakhir, bahkan melebihi rekomendasi WHO yang mengisyaratkan bahwa persalinan ini seharusnya hanya sekitar 5-15% dari total persalinan [3]. Di Indonesia, prevalensi *Sectio Caesarea* bahkan melampaui batas maksimal WHO, seperti yang terlihat dalam survei Kementerian Kesehatan pada tahun 2018 di Aulia Hospital Pekanbaru, dengan prevalensi mencapai 76% per 1000 kelahiran.

Penelitian sebelumnya oleh Aprina dan Puri (2016) mengidentifikasi beberapa penyebab umum persalinan *Sectio Caesarea*, seperti ketuban pecah dini dan kelainan letak janin [4]. Namun, keputusan untuk melakukan persalinan ini memerlukan pertimbangan yang matang, karena memiliki risiko kematian dan komplikasi yang lebih tinggi daripada persalinan *pervaginam* [5]. Untuk mengatasi prevalensi yang tinggi, kemajuan teknologi seperti *Machine Learning* dapat digunakan untuk mengidentifikasi risiko kehamilan dan memprediksi jenis persalinan yang sesuai.

Salah satu metode dalam penerapan *machine learning* adalah *data mining* yang merupakan salah satu metode analisis data untuk menemukan pola atau aturan yang terkandung dalam sejumlah besar data. Dalam konteks kesehatan, teknik *data mining* menjadi penting karena dapat digunakan untuk mendeteksi penyakit secara dini dan diharapkan akan tercipta sistem yang lebih efektif dalam mendukung pengambilan keputusan medis yang tepat. Salah satu teknik *data mining* yang populer adalah klasifikasi, dimana setiap variabel dapat dipetakan ke dalam kelas yang sudah ditentukan sebelumnya dengan membuat aturan dari *data training* [6]. Pada *data training* tersebut harus disertakan dengan beberapa variabel yang menjadi faktor risiko dari persalinan *Sectio Caesarea*. Pada penelitian ini menggunakan beberapa variabel, yaitu usia ibu, usia kehamilan, hipertensi, hemoglobin, glukosa, riwayat partus, letak bayi, Cephalopelvic Disproportion (CPD), ketuban pecah dini (KPD), oligohidramnion, dan taksiran berat janin (TBJ).

Klasifikasi dalam *data mining* memiliki cara perhitungan dengan berbagai metode. Salah satu metode klasifikasi yang umum digunakan dalam dunia medis adalah algoritma C4.5. Anggara (2018) dalam Fatchan et al (2021) menyatakan bahwa Algoritma C4.5 menggunakan struktur pohon keputusan untuk mengklasifikasikan data berdasarkan aturan-aturan yang diperoleh dari *data training* [7]. Dengan memanfaatkan variabel-variabel risiko yang telah disebutkan sebelumnya, algoritma C4.5 dapat membantu tenaga medis dalam mengevaluasi potensi risiko pada setiap kasus kehamilan dan membantu dalam menentukan metode persalinan *pervaginam* atau *Sectio Caesarea*. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi metode persalinan menggunakan algoritma C4.5 Dengan demikian, penerapan algoritma C4.5 dapat berpotensi mengurangi prevalensi persalinan *Sectio Caesarea* yang tidak sesuai, serta meningkatkan keselamatan dan kesehatan ibu serta bayi selama proses persalinan.

## 2. METODE PENELITIAN

### 2.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif analitik yang menggunakan metode analisis data sekunder (ADS). ADS adalah sebuah metode yang mengandalkan data sekunder sebagai sumber data utama [8]. Tujuan dari penelitian ini adalah guna memperoleh informasi yang lebih mendalam tentang faktor-faktor yang mempengaruhi metode persalinan, serta untuk mengidentifikasi pola atau hubungan yang tersembunyi dalam data. Hasil penelitian ini dapat digunakan untuk menginformasikan pengambilan keputusan atau kebijakan terkait metode persalinan. Selain itu, penelitian ini juga berkontribusi terhadap pengembangan teori atau pengetahuan baru di bidang penelitian kesehatan [9].

## 2.2 Metode Pengumpulan Data

Penelitian dimulai dengan melakukan pengambilan data rekam medis pasien melahirkan tahun 2019 dan 2020 yang berasal dari penelitian Armi dan Andriyani (2021), yaitu dari <https://bit.ly/dataasetrm> dengan jumlah data sebanyak 500 data [10]. Pemilihan jumlah dataset berdasarkan hasil penelitian Widystuti dan Darmawan (2018) yang menyatakan bahwa penggunaan dataset minimal 100 sampai dengan 1000 [11]. Semakin banyak dataset yang digunakan, maka semakin tinggi nilai akurasi yang didapatkan dan setiap dataset memiliki kesesuaian algoritma yang berbeda [12].

Dalam penelitian ini, data yang digunakan telah diolah sedemikian rupa sehingga bebas dari informasi yang dapat mengungkapkan identitas pribadi individu yang terlibat. Pengolahan data dilakukan dengan teliti untuk memastikan bahwa seluruh data privasi telah dihapus atau dianonimkan, sehingga penelitian ini terbebas dari potensi masalah etik terkait pelanggaran privasi. Data penelitian ini diperoleh dari penelitian Armi dan Andriyani (2021) yang dengan tegas menghormati hak privasi partisipan dan tidak melibatkan penyalahgunaan informasi [10]. Dengan demikian, penelitian ini berkomitmen penuh terhadap prinsip-prinsip etika penelitian.

## 2.3 Analisis Data

Langkah berikutnya setelah pengumpulan data, dilakukan proses *preprocessing* yang dimulai dengan menangani data yang tidak lengkap atau inkonsisten, termasuk mengganti atau menghapus data yang kosong. Selanjutnya, dilakukan pemilihan atribut (*selection*) yang relevan. Dataset yang telah diproses kemudian dibagi menjadi dua bagian: data latih (*training*) dan data uji (*testing*). *Data training* digunakan untuk membangun model, sementara *data testing* digunakan untuk mengevaluasi kinerja model dalam memprediksi metode persalinan [13]. Kemudian data diolah menggunakan *software RapidMiner* untuk dilakukan *preprocessing*, membangun pohon keputusan, dan mengevaluasi model yang dibangun dengan algoritma C4.5. Termasuk mengukur *accuracy*, *precision*, dan *recall*. Pohon keputusan tersebut dapat digunakan sebagai bentuk visualisasi proses pengambilan keputusan dan dapat mempermudah identifikasi variabel-variabel yang mempengaruhi metode persalinan, sehingga memudahkan interpretasikan hasil dari pengambilan keputusan klinis. Pengujian dan evaluasi dilakukan menggunakan *Confusion Matrix* guna untuk mengetahui *accuracy*, *recall*, dan *precision*.

## 3. HASIL DAN ANALISIS

### 3.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data ibu melahirkan sebanyak sebanyak 500 data dengan 11 atribut yaitu usia ibu, usia kehamilan, hipertensi, hemoglobin, glukosa, riwayat *partus*, letak bayi, *Cephalopelvic Disproportion* (CPD), ketuban pecah dini (KPD), *oligohidramnion*, dan taksiran berat janin (TBJ) dan 1 kelas yaitu persalinan. Pemilihan jumlah dataset berdasarkan hasil penelitian Widystuti dan Darmawan (2018) bahwa penggunaan dataset minimal 100 sampai dengan 1000 [11]. Semakin banyak dataset yang digunakan, maka semakin tinggi nilai akurasi yang didapatkan dan setiap dataset memiliki kesesuaian algoritma yang berbeda [12]. Berikut adalah dataset penelitian yang dipaparkan pada tabel 1 di bawah ini.

Tabel 1. Dataset Penelitian

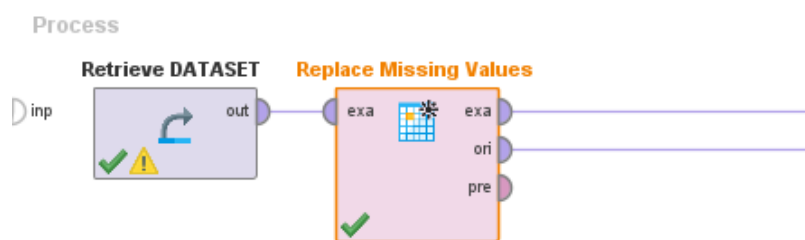
No	Nama Pasien	Usia Ibu	Usia Kehamilan	Hipertensi	Hemoglobin	Glukosa	...	Persalinan
1.	Pasien 1	Normal	Normal	Tidak	Normal	Normal	...	Normal
2.	Pasien 2	Normal	Normal	Tidak	Tidak Normal	Normal	...	Normal
3.	Pasien 3	Normal	Normal	Tidak	Normal	Normal	...	Normal
4.	Pasien 4	Normal	Normal	Tidak	Normal	Normal	...	SC
5.	Pasien 5	Normal	Normal	Tidak	Normal	Tidak Normal	...	SC
6.	Pasien 6	Normal	Normal	Tidak	Normal	Normal	...	Normal

No	Nama Pasien	Usia Ibu	Usia Kehamilan	Hipertensi	Hemoglobin	Glukosa	...	Persalinan
7.	Pasien 7	Normal	Normal	Tidak	Normal	Normal	...	SC
8.	Pasien 8	Normal	Normal	Tidak	Normal	Normal	...	Normal
9.	Pasien 9	Normal	Normal	Tidak	Normal	Normal	...	Normal
...	.....	.....	.....	.....	.....	.....	...	.....
...	.....	.....	.....	.....	.....	.....	...	.....
498.	Pasien 498	Beresiko	Beresiko	Tidak	Normal	Normal	...	Normal
499.	Pasien 499	Beresiko	Normal	Tidak	Normal	Normal	...	SC
500.	Pasien 500	Normal	Normal	Tidak	Tidak Normal	Normal	...	SC

### 3.2 Preprocessing

Dari dataset diatas langkah selanjutnya adalah *preprocessing*. *Preprocessing* adalah teknik pertama sebelum melakukan *data mining*. Tahapan *preprocessing* yaitu *replace*, *selection*, dan *split data*. Berikut adalah tahapan dalam *preprocessing*.

#### 3.2.1 Data Replace (Mengganti Data)

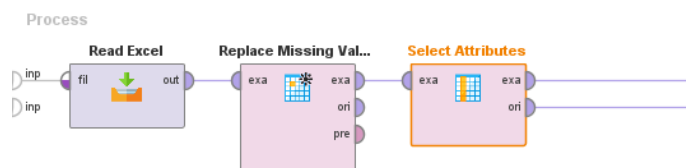


Gambar 1. Proses *Data Replace*

Pada tahap ini penting untuk menangani data valid yang tidak lengkap, kosong, hilang, atau menghilangkan ketidakkonsistenan. Pada proses ini dilakukan karena ditemukan data yang inkonsisten dalam dataset yaitu pada variabel hipertensi, riwayat partus, CPD, KPD, dan *oligohidramnion*. Proses *replace missing value* dilakukan dengan cara merata-ratakan data yang ada, kemudian diisikan pada data yang hilang menggunakan *RapidMiner*.

#### 3.2.2 Data Selection (Pemilihan Data)

Data yang dikumpulkan tidak semua akan digunakan karena hanya data yang akan dianalisis saja yang diperlukan. Proses pemilihan variabel untuk meminimalkan jumlah data yang digunakan dalam proses *data mining* dengan tetap mewakili data asli.



Gambar 2. Proses *Selection*

Variabel yang akan dihapus yaitu nama, dan hamil primi. Gambar di atas adalah proses pemilihan variabel yang digunakan dalam proses *data mining* menggunakan *Rapidminer*. Proses pemilihan atribut yang akan digunakan dalam penelitian didasarkan pada faktor risiko terjadinya persalinan. Atribut hamil

primi dihapus karena sudah terwakilkan oleh atribut riwayat *partus* dengan nilai tidak ada. Dan atribut nama tidak digunakan karena merupakan identitas dari pasien yang harus dirahasiakan.

### 3.2.3 Split Data (Pembagian Data)

*Split data* digunakan untuk membagi dataset menjadi dua, yaitu untuk *data training* dan *data testing* [14]. *Data training* digunakan sebagai dasar pengetahuan untuk membentuk model data yang menghasilkan pohon keputusan. Sementara itu, *data testing* dipakai untuk mengevaluasi model dan mengukur tingkat ketepatannya [15]. Rasio yang akan dibandingkan pada penelitian ini yaitu rasio perbandingan 95:5, 90:10, 85:15, 80:20, 75:25, 70:30, 65:35, 60:40, dan 50:50. Dalam penelitian ini, peneliti menggunakan *sampling* tipe *stratified sampling*. *Stratified sampling* merupakan proses membagi dataset secara *random subset* tetapi dengan memperhatikan distribusi kelas yang sama [16]. Setelah dilakukan *preprocessing data*, data sudah siap untuk proses klasifikasi. Data yang sudah dilakukan *preprocessing* disimpan dalam dataset dapat dilihat pada tautan <https://bit.ly/3RRtL5Y>.

Pada tahap penentuan *data training* dan *data testing*, tujuannya adalah untuk uji validitas terhadap data yang telah dipelajari selama tahap *training* dan *testing* dengan menyisipkan data baru. Hal ini bertujuan untuk menentukan nilai akurasi yang optimal dan mendeteksi nilai *error* yang muncul [14]. Pengujian tipe *sampling* pada penelitian ini meliputi 3 jenis, yaitu, *linier sampling*, *shuffled sampling*, dan *stratified sampling*. Hasil dari evaluasi proses percobaan penentuan *data split* yang paling sesuai dalam memprediksi cara bersalin dapat dilihat pada penjabaran berikut.

#### 1. Linier Sampling

*Linier sampling* adalah teknik pengambilan data secara *linier* tanpa merubah urutan dalam dataset. Artinya, *linier sampling* secara sederhana mempartisi *ExampleSet* tanpa melakukan perubahan pada urutan dataset. Berikut adalah hasil pengujian algoritma C4.5 dengan jenis *linier sampling* menggunakan tools *RapidMiner*.

Tabel 2. Akurasi Pengujian Pengaruh Jumlah *Data Training* dan *Data Testing* dengan *Linier Sampling*

No	Perbandingan	Jumlah Data	Accuracy (%)
1	95:5	328:172	86.05
2	90:10	450:50	84
3	85:15	425 : 75	86.67
4	80:20	400:100	88
5	75:25	375:125	87.20
6	70:30	350:150	86
7	65:35	325 /175	86.29
8	60:40	300:200	86.50
9	50:50	250:250	88.40

Berdasarkan Tabel di atas terlihat bahwa hasil akurasi yang diperoleh dengan *linier sampling* menggunakan metode C4.5 menghasilkan tingkat akurasi yang berbeda-beda pada tahap pengujian tergantung pada jumlah *data training* dan *data testing*. Hasil pengujian menunjukkan bahwa perbedaan jumlah *data training* dan perbandingan *data testing* berpengaruh terhadap nilai akurasi. Diketahui ketika menggunakan *linier sampling*, perbandingan *data training* dan *data testing* 50:50 dengan nilai akurasi mencapai 88,40%. Sedangkan akurasi terendah terlihat pada perbandingan *data training* dan *data testing* sebesar 70:30, dengan akurasi 84%.

Dari hal tersebut diketahui bahwa apabila *data training* dan *data testing* memiliki distribusi yang seimbang, model dapat belajar dengan baik dan memberikan prediksi yang akurat jika data diambil secara urut. Hal ini juga selaras dengan hasil penelitian Azis et al (2020) bahwa perbandingan 50:50 memberikan nilai performa paling baik yakni sebesar 82% [17]. Namun penelitian tersebut tidak menyampaikan teknik *sampling* yang digunakan. Sehingga peneliti berpendapat bahwa perbandingan tersebut tidak dapat digunakan sebagai patokan dalam menentukan perbandingan rasio *data training* dan *data testing* dikarenakan hal tersebut tergantung pada variasi datanya.

#### 2. Shuffled Sampling

*Shuffled sampling* atau pengambilan data acak, melibatkan pemilihan *instance* dari dataset secara acak. Pengumpulan data dilakukan tanpa mempertimbangkan atribut atau kelas tertentu, sehingga setiap

kasus mempunyai peluang yang sama untuk dipilih. Berikut adalah hasil pengujian algoritma C4.5 dengan jenis *shuffled sampling* menggunakan *tools RapidMiner*.

Tabel 3. Akurasi Pengujian Pengaruh Jumlah *Data Training* dan *Data Testing* dengan *Shuffled Sampling*

No	Perbandingan	Jumlah Data	Accuracy (%)
1	95:5	328:172	85.47
2	90:10	450:50	76
3	85:15	425 : 75	84
4	80:20	400:100	87
5	75:25	375:125	88.20
6	70:30	350:150	82
7	65:35	325 /175	85.71
8	60:40	300:200	83
9	50:50	250:250	81.60

Berdasarkan tabel di atas terlihat bahwa hasil akurasi yang diperoleh metode C4.5 dengan *shuffle sampling* menghasilkan tingkat akurasi yang berbeda-beda pada tahap pengujian tergantung pada jumlah *data training* dan *data testing*. Hasil pengujian menunjukkan bahwa perbedaan jumlah *data training* dan perbandingan *data testing* berpengaruh terhadap nilai akurasi. Diketahui bahwa ketika menggunakan *shuffled sampling*, perbandingan *data training* dan *data testing* 75:25 mencapai nilai akurasi sebesar 88,20%. Sedangkan nilai terendah terlihat pada rasio *data training* dan *data testing* sebesar 90:10 dengan rata-rata akurasi sebesar 76%. Dari hal tersebut diketahui bahwa pengambilan data secara acak dapat memberikan variasi yang berguna untuk melatih model. Hal ini juga selaras dengan hasil penelitian Junior (2021) dalam Utomo dan Prathivi (2024), bahwa klasifikasi dengan algoritma *decision tree* menghasilkan akurasi sebesar 85.93% dengan rasio *data training* dan *data testing* sebesar 75:25 [18].

### 3. Stratified Sampling

*Stratified sampling* merupakan jenis *sampling* yang melakukan pengambilan data secara acak, namun perbedaannya dengan *shuffled sampling* adalah *stratified sampling* memastikan distribusi atribut yang sama dengan seluruh dataset. Berikut adalah hasil pengujian algoritma C4.5 dengan tipe *sampling stratified sampling* menggunakan *tools RapidMiner*.

Tabel 4. Akurasi Pengujian Pengaruh Jumlah *Data Training* dan *Data Testing* dengan *Stratified Sampling*

No	Perbandingan	Jumlah Data	Accuracy (%)
1	95:5	328:172	85.47
2	90:10	450:50	92
3	85:15	425 : 75	86.67
4	80:20	400:100	85
5	75:25	375:125	84.80
6	70:30	350:150	82
7	65:35	325 /175	84
8	60:40	300:200	83
9	50:50	250:250	85.60

Berdasarkan Tabel di atas terlihat bahwa hasil akurasi yang dihasilkan dengan menggunakan metode C4.5 dengan tipe *Stratified Sampling* menghasilkan tingkat akurasi yang berbeda-beda pada tahap pengujian tergantung pada jumlah data, *data training*, dan *data testing*. Hasil pengujian menunjukkan bahwa perbedaan perbandingan jumlah *data training* dan *data testing* mempengaruhi nilai akurasi. Rasio *data training* dan *data testing* 90:10, tingkat akurasi mencapai nilai maksimum yaitu 92%. Sedangkan akurasi terendah terlihat pada rasio 70:30 dengan akurasi 82%. Dari hal tersebut diketahui jika ukuran sampel terdistribusi sama, model dapat menghasilkan prediksi yang baik pada *data testing*.

Hasil penelitian Yuliani (2020) menunjukkan bahwa semakin banyak *data training* yang digunakan, semakin baik hasilnya [19]. Hal ini ditunjukkan dengan hasil pengujian algoritma C4.5 dengan pengujian *Confusion Matrix*; akurasi 72,97% dicapai dengan 90% *data training* dan 10% *data testing*. Hasil pengujian

dengan dampak akurasi tertinggi dari jumlah *data training* dan *data testing*, serta jenis sampel, pada algoritma C4.5 ditunjukkan pada tabel 5 di bawah ini.

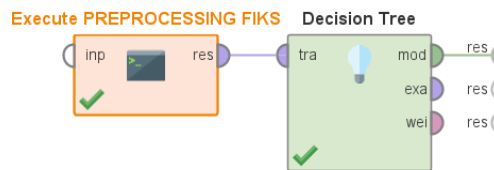
Tabel 5. Pengaruh Jumlah *Data Training* dan *Data Testing* serta Tipe *Sampling* Algoritma C4.5

Type Sampling	C4.5	
	Perbandingan	Akurasi (%)
<i>Linier Sampling</i>	50:50	88.40
<i>Shuffled Sampling</i>	75:25	88.20
<i>Stratified Sampling</i>	90:10	92

Berdasarkan tabel di atas, tingkat akurasi terbaik dalam memprediksi cara persalinan dengan metode C4.5 menggunakan perbandingan rasio *data training* dan *data testing* 90:10 dengan tipe *sampling stratified sampling*. Penggunaan *stratified sampling* harus dipastikan bahwa setiap kelompok atau sub-kelompok terdistribusi secara proporsional dalam sampel. Dengan kata lain, ketika kita menggunakan *stratified sampling*, dataset direpresentasikan dengan baik saat sampel diambil. Hal tersebut membantu mencegah adanya bias yang muncul jika suatu kelas atau atribut dominan salah satu dalam sampel.

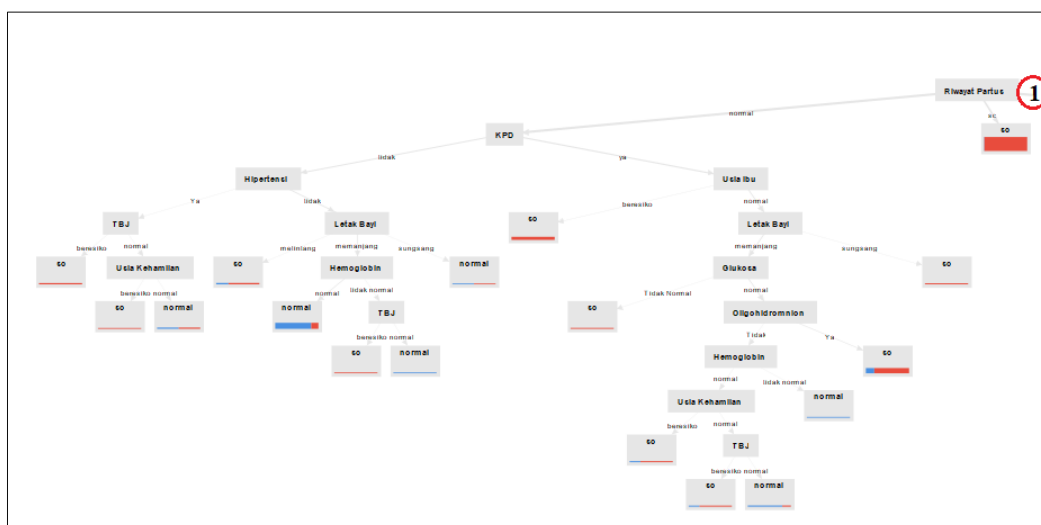
**3.3 Proses Klasifikasi**

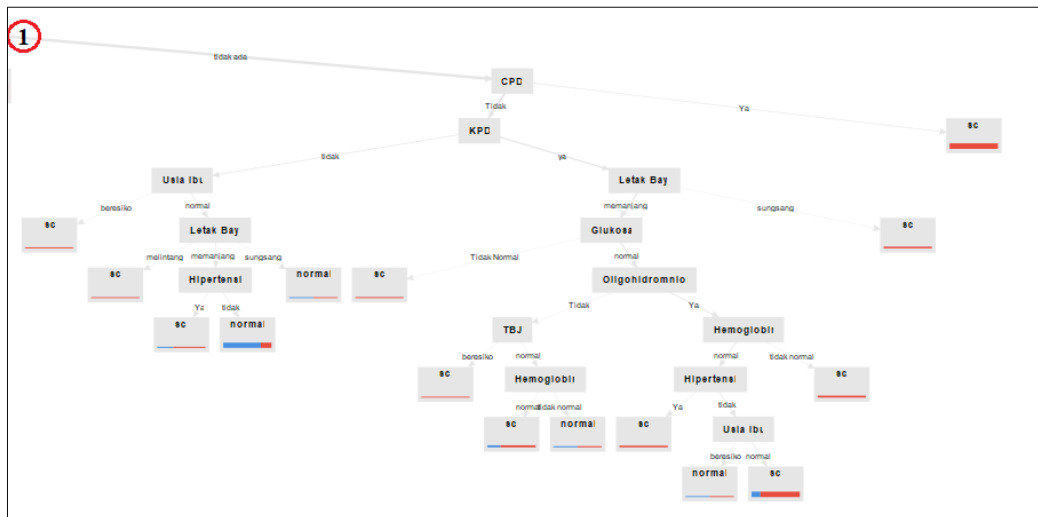
Algoritma klasifikasi yang dipilih dalam penelitian ini adalah C4.5. Pada tahap ini dilakukan pembentukan model pohon keputusan menggunakan nilai *entropy total*, *entropy* setiap nilai pada atribut, dan gain dari setiap atribut. Nilai gain tertinggi akan menjadi akar pertama serta selanjutnya sebagai cabang. Proses dilakukan untuk semua atribut sehingga pohon keputusan terbentuk. Pemodelan algoritma C4.5 dilakukan menggunakan *tools RapidMiner Studio* sesuai dengan gambar di bawah ini.



Gambar 3. Model Pohon Keputusan di *RapidMiner Studio*

Dari proses di atas, kemudian menghasilkan *rule* algoritma C4.5 dalam bentuk pohon keputusan seperti gambar dibawah ini.





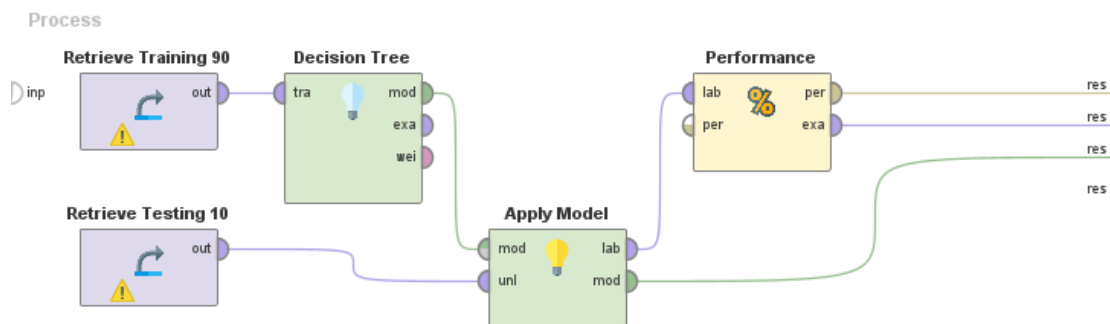
Gambar 4. Model Pohon Keputusan Algoritma C4.5

Gambar di atas merupakan hasil pemodelan dari pengolahan *data training* dan *data testing* yang membentuk pohon keputusan. Akar dari pohon keputusan tersebut adalah riwayat persalinan yang merupakan faktor utama penentu metode persalinan. Hal tersebut selaras dengan penelitian Rochjati (2023) bahwa riwayat persalinan digunakan sebagai salah satu faktor penentu metode persalinan [20].

1. Jika pasien memiliki riwayat *partus Sectio Caesarea*, maka algoritma akan mengidentifikasi persalinan berikutnya sebagai persalinan SC.
2. Jika pasien memiliki riwayat *partus* normal dan terjadi ketuban pecah dini dan usia ibu beresiko, maka algoritma akan mengidentifikasi persalinan berikutnya sebagai persalinan SC.
3. Jika pasien tidak memiliki riwayat *partus* dan pasien mengalami *Chepalopelvic disporpotion*, maka algoritma akan mengidentifikasi persalinan berikutnya sebagai persalinan SC.
4. Dan peran lainnya sesuai dengan gambar 4.

### 3.4 Pengujian dan Evaluasi

Pengujian dilakukan dengan menguji *data testing*. Setelah dilakukan *learning* pada *data training*, maka selanjutnya dilakukan adalah pengujian *data testing*. Pengujian metode dilakukan dengan menguji tingkat *accuracy*, *precision*, dan *recall* dari model yang telah dibuat pada tahap pemodelan. Setelah dilakukan *learning* pada *data training*, maka selanjutnya yang dilakukan adalah pengujian *data testing*. Peneliti menggunakan *stratified sampling* dengan rasio *data training* dan *data testing* adalah 90:10 atau 450:50 data, tingkat akurasi mencapai nilai maksimum yaitu 92%. Proses pengujian menggunakan aplikasi *RapidMiner* dengan menambahkan *data testing*, *apply model*, dan *performance*.



Gambar 5. Workflow Uji Kinerja Model Pohon Keputusan

Hasil pengujian dengan *RapidMiner* dapat dilihat pada tabel 5. Kolom persalinan faktual merupakan data diagnosa yang sebenarnya terjadi pada pasien, sedangkan persalinan prediktif merupakan hasil klasifikasi yang diprediksi oleh model C4.5.



Tabel 6. Hasil Pengujian

No	Nama Pasien	Usia Ibu	Usia Kehamilan	Hipertensi	Hemoglobin	Glukosa	.....	Persalinan Faktual	Persalinan Prediktif
1	Pasien 3	Normal	Normal	Tidak	Normal	Normal	.....	Normal	Normal
2	Pasien 8	Normal	Normal	Tidak	Normal	Normal	.....	Normal	Normal
3	Pasien 13	Normal	Normal	Tidak	Normal	Normal	.....	SC	SC
4	Pasien 15	Normal	Normal	Tidak	Tidak Normal	Normal	.....	SC	SC
5	Pasien 35	Normal	Normal	Tidak	Normal	Normal	.....	Normal	Normal
6	Pasien 52	Normal	Normal	Tidak	Normal	Normal	.....	SC	SC
7	Pasien 86	Normal	Normal	Tidak	Normal	Normal	.....	SC	SC
8	Pasien 94	Normal	Normal	Tidak	Normal	Normal	.....	Normal	Normal
9	Pasien 115	Beresik	Normal	Tidak	Normal	Normal	.....	Normal	Normal
10 ...	.....	.....	.....	.....	.....	.....	.....	.....	.....
11 ...	.....	.....	.....	.....	.....	.....	.....	.....	.....
50	Pasien 458	Normal	Normal	Tidak	Normal	Normal	.....	Normal	Normal

Salah satu metode yang digunakan untuk mengukur performa *data mining* dengan teknik klasifikasi adalah *Confusion Matrix*. Evaluasi dengan *Confusion Matrix* dilakukan dengan cara memprediksi tingkat kebenaran data. *Confusion matrix* dapat digunakan untuk mengukur tingkat *accuracy*, *recall*, dan *precision*. Hasil nilai *Confusion Matrix* ditunjukkan pada tabel berikut.

Tabel 7. *Confusion Matrix*

		Actual Class	
		+ (Positive)	- (Negative)
Prediction Class	+ (Positive)	12	4
	- (Negative)	0	34

Dari tabel di atas menunjukkan bahwa menurut prediksi algoritma C4.5, terdapat 12 kasus normal yang diprediksi normal (TP), 34 kasus SC yang diprediksi SC (TN), 4 kelas SC yang diprediksi normal (FP), dan dalam hal ini tidak ada data normal yang diprediksi sebagai SC (FN). Nilai positif yang dihasilkan dari perhitungan *Confusion Matrix* yang dilakukan pada proses klasifikasi dengan metode C4.5 adalah “persalinan normal”, sedangkan nilai negatif yang dihasilkan dari perhitungan *Confusion Matrix* yang dilakukan pada proses klasifikasi adalah “persalinan SC”.

Berdasarkan tabel tersebut terlihat bahwa tingkat akurasi pada penggunaan algoritma C4.5 adalah sebesar 92% dan dapat dihitung untuk mencari nilai *accuracy*, *recall*, dan *precision* pada persamaan sebagai berikut.

### 1. Perhitungan *accuracy* pada *Confusion Matrix*

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{(TP+FP+TN+FN)} \times 100\% = \frac{34+12}{(50)} \times 100\% \\
 &= \frac{46}{50} \times 100\% \\
 &= 92\%
 \end{aligned}$$

### 2. Perhitungan *precision* pada *Confusion Matrix*

$$\begin{aligned}
 Precision &= \frac{TP}{(TP+FP)} \times 100\% = \frac{12}{(12+4)} \times 100\% \\
 &= \frac{12}{16} \times 100\% \\
 &= 75\%
 \end{aligned}$$

### 3. Perhitungan *recall* pada *Confusion Matrix*

$$\begin{aligned}
 Recall &= \frac{TP}{(TP+FN)} \times 100\% = \frac{12}{(12+0)} \times 100\% \\
 &= \frac{12}{12} \times 100\% \\
 &= 100\%
 \end{aligned}$$

Dari hasil perhitungan *Confusion Matrix* yang dilakukan pada proses klasifikasi dengan metode C4.5, dihasilkan ringkasan nilai sebagai berikut:

1. Hasil perhitungan *Confusion Matrix* dengan metode C4.5 pada Tabel 7 menghasilkan nilai *accuracy* 92%, *precision* 75%, dan *recall* 100%. *Accuracy* merupakan hasil prediksi benar (baik positif maupun negatif) dengan keseluruhan data. Artinya persalinan yang diprediksi benar normal dan benar SC dari keseluruhan metode persalinan adalah 92%.
2. *Precision* merupakan hasil prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Artinya 75% algoritma C4.5 tepat dalam memprediksi persalinan SC yang seharusnya dilakukan SC. Terdapat 15% kasus persalinan normal yang salah diklasifikasikan sebagai SC, sehingga mempengaruhi *precision*.
3. Sedangkan *recall* merupakan hasil prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Artinya nilai *recall* 100% berarti bahwa kasus yang membutuhkan persalinan SC berhasil diidentifikasi dengan benar oleh model dan tidak ada satupun kasus SC yang terlewatkan oleh model. Dari segi kemampuan sistem dalam memulihkan informasi, tingkat keberhasilannya sangat tinggi dalam menemukan semua kasus SC yang benar-benar terjadi. Pada percobaan kali ini, kualitas klasifikasi dinilai cukup berhasil karena mencapai nilai presisi dan *recall* yang tinggi.

### 3.5 Analisis Kode Diagnosis

Pengkodean kasus persalinan harus sesuai aturan ICD-10. Menurut WHO, koding kasus persalinan terdiri dari kode kondisi ibu (O00-O75), metode persalinan (O80-O84), dan *Outcome of delivery Z37*-digunakan sebagai kode tambahan untuk mengetahui hasil persalinan. Persalinan normal tanpa komplikasi direpresentasikan dengan kode O80.9 "*Spontaneous delivery, NOS*". Dan dalam ICD-9-CM, kondisi tersebut direpresentasikan dengan kode 73.59 "*other manually assisted delivery*". Sedangkan persalinan SC direpresentasikan dengan kode O82.9 "*Delivery by caesarean section, unspecified*". Sementara itu, untuk ICD-9-CM, persalinan SC direpresentasikan dengan kode 74.99 "*other caesarean section of unspecified type*". Sebagai kode tambahan, Z37.9 "*Outcome of delivery, unspecified*" juga dapat digunakan.

## 4. KESIMPULAN

Dari penelitian yang telah dilakukan pembuatan algoritma menggunakan data ibu melahirkan untuk prediksi metode persalinan diperoleh kesimpulan bahwa *accuracy* merupakan hasil prediksi benar (baik positif maupun negatif) dengan keseluruhan data. Artinya persalinan yang diprediksi benar normal dan SC dari keseluruhan metode persalinan adalah 92%. *Precision* adalah hasil prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif menghasilkan nilai sebesar 75%. Artinya semakin tinggi nilai presisi, semakin sedikit kasus persalinan normal salah diklasifikasikan sebagai SC. *Recall* adalah hasil prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif dan menghasilkan nilai sebesar 100%. Artinya tidak ada satupun kasus SC yang salah klasifikasi.

Riwayat partus menjadi penentu awal dalam menentukan persalinan berikutnya. Hal tersebut dilakukan untuk meminimalisir seperti infeksi, perdarahan, atau komplikasi lainnya. Kode persalinan terdiri dari kode kondisi ibu, metode persalinan, dan *outcome of delivery*. Pada penelitian selanjutnya dapat menggunakan metode lain agar mendapatkan nilai akurasi yang lebih baik dan dapat mengatasi masalah waktu yang lama pada saat proses klasifikasi, dapat mengatasi masalah *noisy data* yang bisa mengakibatkan akurasi menurun, dan dapat menentukan nilai K yang lebih baik. Penelitian ini dapat dikembangkan menjadi sebuah sistem yang dapat digunakan untuk deteksi dini teknik persalinan menggunakan algoritma C4.5 karena memiliki nilai *precision* yang tinggi.

## REFERENSI

- [1] A. K. Pratiwi, "Faktor Risiko yang Mempengaruhi Metode Persalinan di Rumah Sakit Ibu dan Anak Sitti Khadijah 1 Makassar," *UMI Med. J.*, vol. 4, no. 2, pp. 63–69, Dec. 2019, doi: 10.33096/umj.v4i2.71.
- [2] H. Rosyati, "Indikasi Janin Terhadap Persalinan Seksio Sesarea di Rumah Sakit Islam Jakarta Cempaka Putih," *J. Kedokt. dan Kesehat.*, vol. 10, no. 2, pp. 99–104, 2014.
- [3] WHO, "Death rates from caesarean section far higher in developing countries, finds major global study," 2019. <https://www.qmul.ac.uk/media/news/2019/smd/death-rates-from-caesarean-section-far-higher-in-developing-countries-finds-major-global-study.html> (accessed Mar. 24, 2024).
- [4] A. Aprina and A. Puri, "Faktor-faktor yang Berhubungan dengan Persalinan Sectio Caesarea di RSUD dr. H. Abdul Moeloek Provinsi Lampung," *J. Kesehat.*, vol. 7, no. 1, pp. 90–96, Apr. 2016, doi:

- 10.26630/jk.v7i1.124.
- [5] L. Indawati, D. R. Dewi, A. E. Pramono, and Y. Maryati, *Manajemen Informasi Kesehatan V: Sistem Klaim dan Asuransi Pelayanan Kesehatan*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2018.
- [6] D. Ardiansyah and W. Walim, "Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner," *J. Inkofar*, vol. 1, no. 2, pp. 5–12, Jan. 2018, doi: 10.46846/jurnalinkofar.v1i2.29.
- [7] M. Fatchan, N. Tedi, Alfian, Kurniawan, and E. Widodo, "Perbandingan Dalam Memprediksi Penyakit Liver Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor," *J. Pelita Teknol.*, vol. 16, no. 1, pp. 15–21, 2021.
- [8] C. D. Sucipto, *Metodologi Penelitian Kesehatan*. Yogyakarta: Gosyen Publishing, 2020.
- [9] Sugiyono, *Metode Penelitian Kuantitatif Kualitatif, dan R&D*. Bandung: Alfabeta, 2019.
- [10] P. F. Armi and Y. Andriyani, "Klasifikasi Teknik Persalinan Berdasarkan Data Rekam Medis Menggunakan Algoritma Naive Bayes," Universitas Riau, 2021.
- [11] W. Widystuti and B. Darmawan, "Pengaruh Jumlah Data Set terhadap Akurasi Pengenalan dalam Deep Convolutional Network," in *Konferensi Nasional Sistem Informasi 2018*, STMIK Atma Luhur Pangkalpinang, 2018, pp. 634–638.
- [12] Anita, A. Wicaksono, and T. N. Padilah, "Pengaruh Jumlah Record Dataset Terhadap Algoritma Klasifikasi Berdasarkan Data Customer Churn," *J. Ilm. Inform.*, vol. 6, no. 1, pp. 1–10, Jun. 2021, doi: 10.35316/jimi.v6i1.1223.
- [13] A. A. Robbani, A. M. Siregar, and D. S. Kusumaningrum, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma C4.5," *Sci. Student J. Information, Technol. Sci.*, vol. 3, no. 1, pp. 76–82, 2022, [Online]. Available: <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/424/338>
- [14] G. A. Syafarina, "Penerapan Algoritma Neural Network Dalam Menentukan Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan," *Technol. J. Ilm.*, vol. 7, no. 2, pp. 80–88, Jun. 2016, doi: 10.31602/tji.v7i2.618.
- [15] A. Nuroctaviani, E. P. Satia, and D. Sonia, "Analisis Penggunaan Telemedicine pada Pendaftaran Rekam Medis Klinik Pratama Medika Antapani," *Cerdika J. Ilm. Indones.*, vol. 1, no. 8, pp. 910–916, Aug. 2021, doi: 10.36418/cerdika.v1i8.149.
- [16] P. K. Arieska and N. Herdiani, "Pemilihan Teknik Sampling Berdasarkan Perhitungan Efisiensi Relatif," *J. Stat. Univ. Muhammadiyah Semarang*, vol. 6, no. 2, pp. 166–171, 2018.
- [17] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, Aug. 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [18] M. Utomo and R. Prathivi, "Perbandingan Algoritma Support Vector Machine dan Decision Tree untuk Klasifikasi Performa Perusahaan," *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 105–114, 2024, doi: <https://doi.org/10.47065/bits.v6i1.5278>.
- [19] R. Yuliani, "Penerapan Algoritma C4.5 Berbasis AdaBoost untuk memprediksi Financial Distress Perusahaan," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2020.
- [20] P. Rochjati, *Skrining antenatal pada ibu hamil: pengenalan faktor risiko: deteksi dini ibu hamil risiko tinggi*. Surabaya: Airlangga University Press, 2023.