



Implementasi Data Mining dalam Menentukan Prediksi Status Resiko Persalinan pada Ibu Hamil menggunakan Algoritma C4.5

Dwidya Poernareksa ^{1*}, Nina Rahmadiliyani ¹

¹ Perakam Dan Informasi Kesehatan, Sekolah Tinggi Ilmu Kesehatan Husada Borneo, Indonesia

* Korespondensi: poernareksa2@gmail.com

Sitasi: Poernareksa, D.; dan Rahmadiliyani, N. (2025). Implementasi Data Mining dalam Menentukan Prediksi Status Resiko Persalinan pada Ibu Hamil menggunakan Algoritma C4.5. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 7(1), 123-132. <https://doi.org/10.35746/jtim.v7i1.619>

Diterima: 24-10-2024

Direvisi: 06-01-2025

Disetujui: 09-01-2025



Copyright: © 2025 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: High-risk of pregnancy refers to a situation where pregnancy will have a negative impact on the safety of the mother and baby. Since the beginning of pregnancy, high-risk pregnancy can be predicted by various factors such as the physical and psychological condition of the pregnant woman, nutritional intake, and congenital diseases. According to WHO, Indonesia ranks 5th in premature birth rates with 675,700 babies and this figure is 15.5% of the total birth rate in Indonesia. Estimates of high-risk pregnancies can be observed from patient medical record data, in this case, pregnancy data from pregnant women. Data that is processed into knowledge can be processed through the data mining process. The main objective of this study is to determine how data mining is implemented in determining the prediction of the birth process in pregnant women using the C4.5 algorithm. This research can provide knowledge about the combination of the Two Crows model and the C4.5 algorithm to predict the risk status of childbirth in pregnant women. The C4.5 algorithm is one of the most popular prediction techniques because it is easy for humans to interpret. The data analysis technique in this study uses the Two Crows model which is a development of the CRISP-DM model. The flow of the Two Crows model includes Understanding Business Problem, Building Data Mining Database, Data Explore, Prepare Data For Modeling, Building Model, and Evaluate Model. The data taken is examination data on pregnant women at the Health Center. Based on the results of the study, it was found that the highest root of the application of the C4.5 algorithm is in the height variable. The evaluation was carried out using a confusion matrix. From the evaluation results, it was found that the accuracy value reached 98.44%, the precision value reached 96%, and the recall value reached 100%.

Keywords: Data Mining, Pregnancy, Precision, Recall, Accurate, Prediction Algorithm, Decision Tree

Abstrak: Kehamilan berisiko terjadi ketika kehamilan akan membahayakan keselamatan ibu dan bayi. Sejak awal kehamilan, risiko kehamilan tinggi dapat diprediksi melalui faktor-faktor seperti kondisi fisik dan mental ibu hamil, pola makan, dan riwayat penyakit bawaan. Menurut WHO, Indonesia menempati peringkat ke-5 dalam angka kelahiran prematur dengan jumlah 675.700 bayi. Angka tersebut terbilang tinggi dan mewakili 15,5% dari keseluruhan angka kelahiran bayi di Indonesia. Perkiraan pada kehamilan berisiko dapat diamati dari data rekam medis pasien yang dalam hal ini adalah data kehamilan dari ibu hamil. Data yang diolah menjadi pengetahuan bisa diolah melalui proses data mining. Adapun tujuan utama pada penelitian ini adalah untuk mengetahui bagaimana implementasi data mining dalam menentukan prediksi proses kelahiran pada ibu hamil menggunakan algoritma C4.5. Penelitian ini dapat memberikan pengetahuan tentang kombinasi model Two Crows dan Algoritma C.45 untuk melakukan prediksi status resiko persalinan pada Ibu

Hamil. Algoritma C.45 adalah salah satu teknik prediksi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Teknik analisis data pada penelitian ini menggunakan model Two Crows yang merupakan pengembangan dari model CRISP-DM. Alur dari model Two Crows ini antara lain adalah Understanding Business Problem, Build Data Mining Database, Data Explore, Prepare Data For Modelling, Build Model, dan Evaluate Model. Data yang diambil adalah data pemeriksaan pada Ibu hamil di Puskesmas. Berdasarkan hasil penelitian, didapatkan bahwa akar tertinggi dari penerapan algoritma C4.5 terdapat pada variabel tinggi badan. Evaluasi dilakukan dengan confusion matrix. Dari hasil evaluasi, didapatkan bahwa nilai akurasi mencapai 98,44%, nilai presisi mencapai 96%, dan nilai recall mencapai 100%.

Kata kunci: Data Mining; Ibu Hamil; Presisi, Recall, Akurasi, Algoritma Prediksi, Pohon Keputusan

1. Pendahuluan

Seiring dengan berkembangnya teknologi pemanfaatannya dapat ditemukan dalam berbagai bidang, salah satunya pada bidang kesehatan. Perkembangan pada bidang ini ditemukan pula pada proses persalinan ibu hamil. Berdasarkan Survei Demografi Kesehatan Indonesia (SDKI) tahun 2012, angka kematian ibu di Indonesia termasuk tinggi, yaitu 395 per 100.000 kelahiran hidup [1]. Menurut World Health Organization (WHO) sejak tahun 2020, sekitar 800 perempuan meninggal disebabkan dengan adanya komplikasi selama kehamilan dan persalinan [2]. Angka kematian ibu ini adalah salah satu indikator untuk melihat kesehatan ibu. Kematian ibu selama masa kehamilan atau setelah persalinan dengan berbagai macam penyebab. Lebih dari setengah juta ibu berusia 15-49 tahun meninggal karena masalah terkait kehamilan dan persalinan. [3]. Banyak masalah yang menjadi penyebab adanya meningkatnya tingkat kematian. Pertama, kurangnya edukasi mengenai kesehatan obstetri dan ginekologi. Kedua, kurangnya layanan darurat obstetri dan neonatal dasar dan komprehensif yang tersedia bagi mereka yang membutuhkan secara tepat waktu. Ketiga, karena masih banyak yang mengatasi masalah kesehatan tersebut dengan pengobatan tradisional yang tidak jarang kontra produktif dengan keadaan tubuh, sehingga bila sudah mengalami gejala yang cukup serius baru datang ke dokter [4].

Saat ini, sistem perawatan kesehatan menggunakan sistem informasi untuk mengelola data perawatan kesehatan atau pasien. Sistem ini menghasilkan sejumlah besar data medis yang mungkin berisi catatan pasien elektronik. Data ini mungkin terdiri dari banyak informasi dan pengetahuan tak ternilai yang tersembunyi dalam basis data tersebut, yang dapat digunakan untuk mendukung pengambilan keputusan klinis yang efektif. Namun, penemuan pola dan pengetahuan tersembunyi dari basis data tersebut sering kali tidak dimanfaatkan [5]. Salah satu cara yang dapat dilakukan adalah dengan mengolah data agar dapat menjadi informasi berharga dengan memanfaatkan metode data mining [6]. Data mining adalah fase dalam mengurutkan informasi dengan menghubungkan setiap contoh dalam setiap indeks informasi besar dengan banyak informasi, Penambangan informasi menggunakan kerangka sampel yang dibuat dengan memeriksa informasi terdekat yang diperlukan untuk mengekstrak informasi dari sekumpulan data yang dicari untuk membantu menemukan informasi yang perlu ditemukan Pilihan teknik atau perhitungan penambangan informasi harus dilakukan berdasarkan alasan umum dan periodik yang penting dalam praktik [7]. Pengolahan data dilakukan dengan teknik data mining dengan menebak kemungkinan yang terjadi berdasarkan kejadian sebelumnya. Teknik data mining memungkinkan Anda membuat prediksi berdasarkan data masa lalu. Penambangan data menemukan pola dan tren yang berguna dalam kumpulan data yang besar. Teknologi data mining adalah alat untuk menambang data dalam database besar untuk mengidentifikasi informasi berguna dan pengetahuan yang relevan [8]. Data mining bisa dikatakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database [9].

Penelitian yang mengangkat studi kasus serupa sebelumnya telah dilakukan, seperti penelitian pemilihan metode melahirkan dengan memanfaatkan kinerja metode Naïve Bayes dibantu dengan algoritma optimasi Particle Swarm Object (PSO) yang memperoleh hasil akurasi sebesar 70,83% [10]. Selanjutnya ada pula penelitian [11] mengenai sistem prediksi kondisi kelahiran bayi menggunakan klasifikasi Naïve Bayes dengan memanfaatkan 110 data training dan 55 data testing sehingga memperoleh hasil akurasi 92,73%. Kemudian selanjutnya penelitian yang memanfaatkan algoritma C4.5 ini telah diterapkan dalam penelitian status proses persalinan ibu hamil, dalam penelitian ini menggunakan 682 record data rekam medis dan menghasilkan akurasi prediksi maksimum yaitu sebesar 97,08% [1].

Berdasarkan tiga penelitian sebelumnya dapat dilihat bahwa penggunaan metode C4.5 memperoleh akurasi terbaik dalam hal studi kasus prediksi persalinan ibu hamil. Akan tetapi model yang digunakan pada penelitian ini mengacu pada model KDD, penelitian ini menggunakan model Two Crows dalam runtutannya. Hal ini diharapkan dapat memberikan hasil yang lebih maksimal pada saat evaluasi algoritma data mining. Kelebihan dari metode C4.5 ini dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, serta efisien ketika menangani atribut yang berbentuk diskret ataupun numerik [12] Tidak hanya itu algoritma C4.5 mempermudah peneliti memprediksi sebuah nilai menjadi informasi dan algoritma ini pula merupakan pengembangan dari ID3 dengan membentuk pohon keputusan untuk proses prediksi perkiraan sebuah kejadian [13]

Berdasarkan penjelasan maupun penelitian yang sebelumnya telah dilakukan. Maka penelitian yang dilakukan adalah melakukan implementasi data mining dalam menentukan prediksi status resiko persalinan pada ibu hamil menggunakan algoritma C4.5. Adapun tujuan penelitian ini adalah untuk mengetahui seberapa akurat prediksi status resiko yang dihasilkan dengan memanfaatkan algoritma C4.5 berdasarkan tingkat akurasi, presisi, dan recall yang dihasilkan pada algoritma.

2. Bahan dan Metode

Data yang digunakan dalam penelitian ini adalah database resiko ibu hamil yang diperoleh dari salah satu Puskesmas di Kota Banjarbaru. Database merupakan data kehamilan dari beberapa pasien yang terdata di Puskesmas di kota Banjarbaru. Database diperoleh dalam bentuk Microsoft excel, dan akan di ubah dalam bentuk MySQL untuk mempermudah jalannya algoritma C4.5 dan proses evaluasi data. Database terdiri dari tahun 2021 sampai dengan 2024 yang berjumlah 492 data. Data tersebut akan dijadikan data latih dan data uji dari penelitian. Database didapatkan dari sistem informasi puskesmas yang terdapat pada Puskesmas di Kota Banjarbaru. Data yang ada berisi tentang ANC beserta data-data yang lain terkait status resiko Ibu Hamil. Jenis penelitian ini adalah penelitian kuantitatif dengan metode klasifikasi. Penelitian ini bermaksud menerapkan metode klasifikasi prediksi dalam data proses kelahiran pada ibu hamil menggunakan algoritma C4.5.

Adapun teknik pengumpulan dan analisis data yang digunakan pada penelitian ini menggunakan model Two Crows Corporation. Adapun tahapan Two Crows Corporation yaitu pengembangan dari model CRISP-DM [14]. Adapun tahapan dari model ini terdiri dari Understanding Business Problem, Build Data Mining Database, Data Explore, Prepare Data For Modelling, Build Model, dan Evaluate Model. Pada tahap pertama, ini dititik beratkan untuk memahami mengenai data dan alur bisnis dari penelitian. Penentuan pemahaman dari permasalahan bisnis, dan tujuan data mining harus jelas. Perbedaan yang ada pada model CRISP-DM dan Two Crows terdapat pada tahap kedua, Pada Model CRISP-DM, tahapan kedua dinamakan Data Understanding, Pada tahap ini dilakukan proses pengumpulan data yang dilanjutkan dengan analisis data dan penilaian kualitas data yang digunakan dalam penelitian [15]. Sedangkan pada Two Crows pada

tahap kedua disebut Build Data Mining Database. Tugas tahap ini adalah membuat database yang akan digunakan dalam proses data mining. Tugas yang perlu dilakukan pada database ini meliputi pengumpulan data, deskripsi data, pemilihan data, penilaian kualitas data dan pembersihan data, integrasi dan integrasi. Pada model Two Crows, tahapan ini digunakan untuk melakukan penyelidikan dan mengidentifikasi informasi penting mengenai data yang akan digunakan pada saat pencarian pengetahuan di proses data mining. Sedangkan tahapan ketiga pada model CRISP-DM adalah data preparation. Pada fase ini, kita memilih data dan mengelompokkan atribut atau bidang yang dipilih ke dalam data untuk digunakan dalam penelitian [16]. Pada Model Two Crows, tahap ketiga adalah data explore, Tujuan tahapan ini untuk melakukan penyelidikan dan mengidentifikasi informasi penting mengenai data yang akan digunakan pada saat pencarian pengetahuan di proses data mining [14]. Pada tahap keempat, model Two Crows menggunakan tahapan prepare data for modelling. Penggunaan seluruh data merupakan langkah yang ideal, namun kurang praktis dalam pengerjaan data mining. Untuk itu, pemilihan attribut yang dipakai sangat berpengaruh dalam proses data mining. Pada tahap kelima dinapkan modelling. Dalam tahap ini, berbagai macam teknik pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis yang sesuai. Adapun Teknik yang digunakan yaitu dengan teknik classification, menggunakan algoritma C4.5. Sedangkan terakhir merupakan tahap evaluasi model. Pada tahap ini, model yang sudah dibuat diuji dan dievaluasi keakuratan dan generalitasnya. Penelitian ini menggunakan proses evaluasi dengan confusion matrix. Confussion Matrix digunakan untuk menentukan nilai akurasi, presisi, dan recall dari sebuah data uji yang dihasilkan [14]

3. Hasil

Penelitian ini bertujuan untuk mengetahui pola klasifikasi dari pohon keputusan dalam menentukan resiko ibu hamil dan mengukur tingkat akurasi, presisi dan recall dari pohon keputusan yang dihasilkan pada algoritma C4.5. Algoritma ini didasarkan pada tingkat gain ratio yang diperoleh dari perhitungan yang nantinya akan berpengaruh pada akar pengambilan keputusan. Gain ratio yang tertinggi nantinya akan menjadi akar utama dari pohon keputusan. Tahap evaluasi didapatkan dengan cara membandingkan data testing yang tersedia, dengan pola pohon keputusan yang telah terbentuk dari perhitungan algoritma pada data training. Untuk pengukuran tingkat akurasi, presisi, dan recall dari data yang telah diuji menggunakan confusion matrix. Confussion matrix didasarkan pada perbandingan jumlah data yang dianggap benar dan salah oleh aplikasi, dengan data aktual yang ada, kemudian dihitung dalam bentuk matriks.

Pada penelitian ini terdapat 492 total data yang terdiri dari 428 data training (tahun 2022 – 2024) dan 64 data testing (tahun 2024). Data tersebut diolah kedalam model metode C4.5, pada pengimplementasian metode ini terdapat beberapa langkah yang dilakukan yaitu memilih atribut sebagai akar sehingga menghasilkan nilai entropi

Tabel 1. Hasil Perhitungan Entropi

No	Atribut	Nilai Atribut	Jumlah Kasus			Entropi
			Total	Tidak Aman	Aman	
1	Total	Total	428	181	247	0.9828
2	riwayat_penyakit	Ada	54	51	3	0.3095
3	riwayat_penyakit	tidak ada	374	130	244	0.9319
4	usia	diatas 35 tahun	31	30	1	0.2056
5	usia	dibawah 35 tahun	397	151	246	0.9583
6	tb	diatas 150 cm	331	92	239	0.8526
7	tb	dibawah 150 cm	97	89	8	0.4108
8	varitas	diatas 3 anak	43	42	1	0.1594

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi
9	varitas	dibawah 3 anak	385	139	246	0.9435
10	kelahiran_sebelumnya	caesar diatas 12 bulan	35	22	13	0.9518
11	kelahiran_sebelumnya	caesar dibawah 12 bulan	11	9	2	0.684
12	kelahiran_sebelumnya	keguguran	15	7	8	0.9968
13	kelahiran_sebelumnya	normal diatas 6 bulan	169	75	94	0.9909
14	kelahiran_sebelumnya	normal dibawah 6 bulan	8	7	1	0.5436
15	kelahiran_sebelumnya	tidak diketahui	190	61	129	0.9055

Tabel 2. Hasil Perhitungan Entropi Setiap Atribut Utama

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain
1	Total	Total	428	181	247	0.9828	
2	riwayat_penyakit	Ada	54	51	3	0.3095	0.1294
3	riwayat_penyakit	tidak ada	374	130	244	0.9319	0.1294
4	usia	diatas 35 tahun	31	30	1	0.2056	0.079
5	usia	dibawah 35 tahun	397	151	246	0.9583	0.079
6	tb	diatas 150 cm	331	92	239	0.8526	0.2303
7	tb	dibawah 150 cm	97	89	8	0.4108	0.2303
8	varitas	diatas 3 anak	43	42	1	0.1594	0.1181
9	varitas	dibawah 3 anak	385	139	246	0.9435	0.1181
10	kelahiran_sebelumnya	caesar diatas 12 bulan	35	22	13	0.9518	0.0491
11	kelahiran_sebelumnya	caesar dibawah 12 bulan	11	9	2	0.684	0.0491
12	kelahiran_sebelumnya	keguguran	15	7	8	0.9968	0.0491
13	kelahiran_sebelumnya	normal diatas 6 bulan	169	75	94	0.9909	0.0491
14	kelahiran_sebelumnya	normal dibawah 6 bulan	8	7	1	0.5436	0.0491
15	kelahiran_sebelumnya	tidak diketahui	190	61	129	0.9055	0.0491

Tabel 3. Hasil Perhitungan Gain Information dan Split Info

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
1	Total	Total	428	181	247	0.9828			0
2	riwayat_penyakit	Ada	54	51	3	0.3095	0.1294	0.5468	0.2366

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
3	riwayat_penyakit	tidak ada	374	130	244	0.9319	0.1294	0.5468	0.2366
4	usia	diatas 35 tahun	31	30	1	0.2056	0.079	0.3749	0.2107
5	usia	dibawah 35 tahun	397	151	246	0.9583	0.079	0.3749	0.2107
6	tb	diatas 150 cm	331	92	239	0.8526	0.2303	0.7721	0.2983
7	tb	dibawah 150 cm	97	89	8	0.4108	0.2303	0.7721	0.2983
8	varitas	diatas 3 anak	43	42	1	0.1594	0.1181	0.4705	0.251
9	varitas	dibawah 3 anak	385	139	246	0.9435	0.1181	0.4705	0.251
10	kelahiran_sebelumnya	caesar diatas 12 bulan	35	22	13	0.9518	0.0491	1.7573	0.0279
11	kelahiran_sebelumnya	caesar dibawah 12 bulan	11	9	2	0.684	0.0491	1.7573	0.0279
12	kelahiran_sebelumnya	keguguran	15	7	8	0.9968	0.0491	1.7573	0.0279
13	kelahiran_sebelumnya	normal diatas 6 bulan	169	75	94	0.9909	0.0491	1.7573	0.0279
14	kelahiran_sebelumnya	normal dibawah 6 bulan	8	7	1	0.5436	0.0491	1.7573	0.0279
15	kelahiran_sebelumnya	tidak diketahui	190	61	129	0.9055	0.0491	1.7573	0.0279

Karena atribut varitas adalah atribut dengan gain ratio tertinggi di bawah akar dari tb, maka atribut varitas dipilih menjadi atribut akar setelah atribut tb (node akar 1.1 dan 1.2). Kemudian dilakukan perhitungan untuk atribut varitas yang dapat dilihat pada Tabel 4 :

Tabel 4. Perhitungan Atribut Varitas

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
1	Total	Total	428	181	247	0.8526			0
2	riwayat_penyakit	Ada	54	51	3	0.3843	0.1913	0.5318	0.3597
3	riwayat_penyakit	tidak ada	374	130	244	0.6994	0.1913	0.5318	0.3597
4	usia	diatas 35 tahun	31	30	1	0.2352	0.1342	0.397	0.338
5	usia	dibawah 35 tahun	397	151	246	0.7596	0.1342	0.397	0.338
6	varitas	diatas 3 anak	43	42	1	0.1831	0.2015	0.4962	0.4061

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
7	varitas	dibawah 3 anak	385	139	246	0.7082	0.2015	0.4962	0.4061
8	kelahiran_sebelumnya	caesar diatas 12 bulan	35	22	13	0.9988	0.0972	1.749	0.0556
9	kelahiran_sebelumnya	caesar dibawah 12 bulan	11	9	2	0.7219	0.0972	1.749	0.0556
10	kelahiran_sebelumnya	keguguran	15	7	8	0.9183	0.0972	1.749	0.0556
11	kelahiran_sebelumnya	normal diatas 6 bulan	169	75	94	0.9255	0.0972	1.749	0.0556
12	kelahiran_sebelumnya	normal dibawah 6 bulan	8	7	1	0.7219	0.0972	1.749	0.0556
13	kelahiran_sebelumnya	tidak diketahui	190	61	129	0.5436	0.0972	1.749	0.0556

Karena atribut kelahiran_sebelumnya adalah atribut dengan gain ratio tertinggi di bawah akar dari varitas, maka atribut kelahiran_sebelumnya dipilih menjadi atribut akar setelah atribut varitas (node akar 1.1.1 dan 1.1.2). Selain itu, atribut kelahiran sebelumnya pada nilai normal dibawah 6 bulan dan tidak diketahui bernilai 0, maka pohon keputusan langsung mengidentifikasi tidak dan atribut tersebut sudah menjadi daun. Kemudian dilakukan perhitungan untuk atribut kelahiran sebelumnya. Berikut adalah tabel perhitungan untuk atribut kelahiran_sebelumnya, yang mana dapat dilihat pada Tabel 5.

Tabel 5. Perhitungan kelahiran_sebelumnya

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
1	Total	Total	428	181	247	0.8113			0
2	riwayat_penyakit	Ada	54	51	3	0	0.1226	0.8113	0.1511
3	riwayat_penyakit	tidak ada	374	130	244	0.9183	0.1226	0.8113	0.1511
4	usia	didas 35 tahun	31	30	1	0	0.1226	0.8113	0.1511
5	usia	dibawah 35 tahun	397	151	246	0.9183	0.1226	0.8113	0.1511

Karena atribut Riwayat penyakit adalah atribut dengan gain ratio tertinggi di bawah akar dari kelahiran_sebelumnya, maka atribut tim dipilih menjadi atribut akar setelah atribut Riwayat penyakit. Untuk perhitungan dari atribut Riwayat_penyakit dapat dilihat pada Tabel 6.

Tabel 6. Perhitungan riwayat penyakit

No	Atribut	Nilai Atribut	Jumlah Kasus Total	Jumlah Kasus Tidak Aman	Jumlah Kasus Aman	Entropi	Information Gain	Split Info	Gain Ratio
1	Total	Total	428	181	247	0.9183			0
2	usia	diatas 35 tahun	31	30	1	0	0	0	0
3	usia	dibawah 35 tahun	397	151	246	0.9183	0	0	0

Setelah dilakukan perhitungan untuk semua atribut prediksi, terbentuklah pohon keputusan. Setelah dilakukan pembentukan pohon keputusan, maka akan dibentuk rule prediksi untuk semua atribut yang sudah ada di dalam pohon keputusan sesuai dengan arah akar node, dan daun dari pohon keputusan. Setelah dilakukan pembentukan rule dari pohon keputusan C4.5, akan dilakukan proses pengujian data testing terhadap rule yang ada. Gambar hasil pengujian data testing dapat dilihat pada Gambar 1.

Usia	Tinggi Badan	Varitas	Riwayat penyakit	Persalinan	Kelahiran Sebelumnya	Keputusan Asli	Keputusan C4.5	ID Rule C4.5
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
diatas 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	45
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
diatas 35 tahun	dibawah 150 cm	dibawah 3 anak	ada	belum	tidak diketahui	tidak aman	tidak aman	41
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	50
dibawah 35 tahun	dibawah 150 cm	diatas 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	43
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	keguguran	aman	aman	36
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
dibawah 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	tidak aman	tidak aman	52
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	caesar	caesar diatas 12 bulan	aman	aman	34
diatas 35 tahun	diatas 150 cm	diatas 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	10
diatas 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	30
dibawah 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	50
dibawah 35 tahun	dibawah 150 cm	diatas 3 anak	ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	41
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
diatas 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	30
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
diatas 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	caesar	caesar diatas 12 bulan	tidak aman	tidak aman	45
dibawah 35 tahun	diatas 150 cm	diatas 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	10
dibawah 35 tahun	diatas 150 cm	diatas 3 anak	tidak ada	normal	normal diatas 6 bulan	tidak aman	tidak aman	10
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	normal diatas 6 bulan	aman	aman	37
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	normal	keguguran	aman	aman	36
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39
dibawah 35 tahun	dibawah 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	tidak aman	tidak aman	52
dibawah 35 tahun	diatas 150 cm	dibawah 3 anak	tidak ada	belum	tidak diketahui	aman	aman	39

Gambar 1. Pengujian Data Testing

Proses pengujian dilakukan sampai seluruh data testing berhasil diidentifikasi oleh sistem rule dari C4.5. Setelah semua data testing telah diuji, akan dilakukan perhitungan dari tingkat akurasi, presisi, dan recall dari data testing. Untuk tabel penilaian keseluruhan data testing bisa dilihat pada tabel 7.

Tabel 7. Penilaian Kinerja

	Klasifikasi	Prediksi		
		Aman	Tidak Aman	Jumlah
Aktual	Aman	39	1	40
	Tidak Aman	0	24	24
	Jumlah	39	25	64

Setelah dilakukan perhitungan confusion matrix, didapatkan hasil presisi sebesar 96%. Nilai recall 100% dan nilai akurasi sebesar 98,44%.

4. Pembahasan

Setelah model algoritma C4.5 dijalankan dan menghasilkan pohon keputusan yang berkaitan dengan resiko kehamilan yang dialami pasien, didapatkan akar-akar yang menjadi atribut pengambilan keputusan. Pengetahuan yang diperoleh dari informasi yang diolah adalah sebagai berikut:

1. Jumlah nilai akurasi yang dihasilkan pada penelitian ini sebesar 98,44%. Nilai ini disebabkan karena jumlah data yang diujikan secara rata-rata benar, tidak dititik beratkan pada keputusan yang bernilai benar dan salah pada sistem dan data aktual. Hal ini sejalan dengan penelitian sebelumnya [17] yang mana mendapatkan hasil akurasi terbaik ketika mengimplementasikan metode C4.5 pada studi kasus serupa. Pada penelitian serupa [4] menggunakan algoritma Naive Bayes, akurasi yang dihasilkan sebesar 92%. Pada penelitian dari [17], algoritma C4.5 terbukti memiliki keandalan yang lebih baik dibandingkan Naive Bayes dengan kasus serupa.
2. Jumlah nilai presisi yang dihasilkan pada penelitian ini sebesar 96%. Nilai ini disebabkan karena untuk perhitungan presisi, dititik beratkan pada nilai yang bernilai benar dan salah yang terklasifikasi pada sistem. Pada sistem yang ada, nilai yang diidentifikasi bernilai "aman" yang sesuai dengan sistem dan data aktual 1, dan nilai yang diidentifikasi bernilai "tidak aman" sebesar 24, sehingga didapatkan hasil 96%.
3. Jumlah nilai *recall* yang dihasilkan pada penelitian ini sebesar 100%. Nilai ini disebabkan karena untuk perhitungan nilai *recall*, dititik beratkan kepada perbandingan jumlah data yang benar oleh sistem dengan jumlah data aktual yang muncul. Hal lain yang menyebabkan nilai *recall* tinggi, karena kelas tujuan yang dicari tergolong *majority bounding*, yaitu kelas yang mempunyai nilai tujuan "aman" yang banyak daripada nilai tujuan yang "tidak aman". Keterbatasan pada penelitian ini terletak pada atribut yang digunakan, pada penelitian ini sebagian besar menggunakan data terkait demografis dan pemeriksaan fisik pasien, sedangkan pada penelitian sebelumnya seperti yang dilakukan oleh [3] dan [18], sebagian besar menggunakan data pemeriksaan fisik.

5. Kesimpulan

Pohon keputusan akar tertinggi yang merupakan attribute awal pohon keputusan diambil pada jumlah gain ratio terbanyak pada setiap iterasi dari perhitungan algoritma C4.5. Berdasarkan percobaan, gain ratio tertinggi dimiliki oleh atribut *tb* dengan jumlah gain ratio sebesar 0,2983. Nilai akurasi yang didapatkan pada penelitian ini sebesar 98,44%. Nilai ini disebabkan karena jumlah data yang diujikan secara rata-rata benar, tidak dititik beratkan pada keputusan yang bernilai benar dan salah pada sistem dan data aktual. Sedangkan nilai presisi yang didapatkan pada penelitian ini adalah 96%, dan nilai *recall* yang dihasilkan pada penelitian ini sebesar 100%. Sementara hal yang masih perlu diperbaiki dalam penelitian ini adalah atribut masih terbatas pada data dari hasil pemeriksaan pasien dan demografi pasien. Saran untuk penelitian selanjutnya bisa ditambah atribut lain agar pohon keputusan yang dihasilkan bisa lebih beragam. Algoritma yang dipakai bisa diperbaharui menjadi algoritma C5.0 dan bisa dibuat perbandingan evaluasi kinerja dari dua algoritma prediksi yang lama maupun yang terbaru.

Ucapan Terima Kasih: Ucapan terima kasih yang tak terhingga kepada pihak yang terlibat pada penelitian kali ini.

Referensi

- [1] M. P. S. D. Ningsih and B. Noranita, "Status Proses Persalinan Menggunakan Algoritma C4.5," *J Masy Inform*, vol. 9, no. 1, pp. 1–13, 2018, doi: 10.14710/jmasif.9.1.31478.
- [2] E. Ekayanti and M. Mutmainah, "ANALISIS KEHAMILAN RESIKO 4T PADA IBU HAMIL DI DESA PASIRGINTUNG KECAMATAN CIKULUR KABUPATEN LEBAK TAHUN 2023," *Bina Gener J Kesehat*, vol. 15, no. 2, pp. 58–70, Mar. 2024, doi: 10.35907/bgjk.v15i2.319.
- [3] M. Y. Pusadan, A. Ghifari, and Y. Anshori, "Implementasi Data Mining untuk Prediksi Status Proses Persalinan pada Ibu Hamil Menggunakan Algoritma Naive Bayes," *Technomedia J*, vol. 8, no. 1 Juni, pp. 137–153, 2023, doi: 10.33050/tmj.v8i1.1980.
- [4] V. Khoirunnisa and S. Lestari, "Implementasi Klasifikasi Kehamilan Beresiko Dengan Metode Naive Bayes Pada Puskesmas Kelurahan Malaka Jaya," *J Indones Manaj Inform dan Komun*, vol. 4, no. 3, pp. 1680–1693, 2023, doi: 10.35870/jimik.v4i3.396.
- [5] E. N. Ekwonwune, C. I. Ubochi, and A. E. Duroha, "Data Mining as a Technique for Healthcare Approach," *Int J Commun Netw Syst Sci*, vol. 15, no. 09, pp. 149–165, 2022, doi: 10.4236/ijcns.2022.159011.
- [6] H. Amalia, R. Rahmadanti, A. Syaiin, S. Salsabila4, Y. Yunita, and S. Sriyadi, "Prediksi Resiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree," *Swabumi*, vol. 11, no. 1, pp. 48–53, 2023, doi: 10.31294/swabumi.v11i1.15270.
- [7] B. I. Nugroho, N. P. Lestari, R. D. Kurniawan, and G. Gunawan, "Tinjauan Pustaka Sistematis: Data Mining Dalam Bidang Kesehatan," *J Ekon Teknol dan Bisnis*, vol. 1, no. 1, pp. 14–27, 2022, doi: 10.57185/jetbis.v1i1.2.
- [8] A. Saifudin, "Metode Data Mining Untuk Seleksi Calon Mahasiswa PADA PENERIMAAN MAHASISWA BARU DI UNIVERSITAS PAMULANG," vol. 10, no. 1, pp. 25–36, 2018, doi: <https://doi.org/10.24853/jurtek.10.1.25-36>.
- [9] Q. Hasanah, A. Andrianto, and M. A. Hidayat, "Sistem Informasi Posyandu Ibu Hamil dengan Penerapan Klasifikasi Resiko Kehamilan Menggunakan Metode Naive Bayes," *Berk Sainstek*, vol. 6, no. 1, p. 1, 2018, doi: 10.19184/bst.v6i1.7554.
- [10] H. Amalia, R. Z. Agungsyah, F. A. Lestari, A. F. Lestari, A. Puspita, and S. Sriyadi, "Peningkatan Kinerja Metode Naive Bayes Dengan Particle Swarm Object Untuk Dataset Pemilihan Metode Melahirkan," *Swabumi*, vol. 11, no. 2, pp. 156–161, 2023, doi: 10.31294/swabumi.v11i2.17138.
- [11] H. A. Damar Rani and S. Zuhri, "Sistem Prediksi Kondisi Kelahiran Bayi menggunakan Klasifikasi Naive Bayes," *Joined J (Journal Informatics Educ*, vol. 3, no. 2, p. 48, 2020, doi: 10.31331/joined.v3i2.1432.
- [12] P. Perner, *Advances in Data Mining. Applications and Theoretical Aspects: 9th Industrial Conference, ICDM 2009, Leipzig, Germany, July 20 - 22, 2009. Proceedings*. in *Lecture Notes in Computer Science*. Leipzig: Springer Berlin Heidelberg, 2009. [Online]. Available: <https://books.google.co.id/books?id=GRZYV4sDzH0C>
- [13] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradig - J Komput dan Inform*, vol. 23, no. 2, 2021, doi: <https://doi.org/10.31294/p.v23i2.11395>.
- [14] T. Crows, *Introduction to Data Mining and Knowledge Discovery*. USA: Two Crows Corporation, 1999. [Online]. Available: <papers2://publication/uuid/87DAF9E6-09CC-4831-8AB5-F0F97F467BFA>
- [15] B. Muhammad, "Implementasi Data Mining untuk Prediksi Standar Hidup Layak Berdasarkan Tingkat Kesehatan dan Pendidikan Masyarakat," *J SISKOM-KB (Sistem Komput dan Kecerdasan Buatan)*, vol. 2, no. 2, pp. 33–37, 2019.
- [16] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter," *Komputa J Ilm Komput dan Inform*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34010/komputa.v12i1.9418.
- [17] M. B. Hanif, H. A. D. Rani, A. Rifai, and Gallet Guntoro Setiaji, "Komparasi Metode Naive Bayes dan C4 . 5 Pada Klasifikasi Persalinan Prematur," *Joined J*, vol. 5, no. 1, pp. 54–65, 2022, doi: <https://doi.org/10.31331/joined.v5i1.2242>.
- [18] M. D. Setyaningsih, A. Wahyuni, and A. Y. Kuntoro, "Implementasi Data Mining C4.5 Untuk Klasifikasi Faktor Resiko Kesehatan Pada Ibu Hamil," *J Ilmu Komput dan Bisnis*, vol. 13, no. 2a, pp. 67–77, 2022, doi: 10.47927/jikb.v13i2a.376.