



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 823–832

Penerapan Algoritma C4.5 pada RapidMiner untuk Resiko Kesehatan Ibu Hamil

Muhammad Fajar Sidik, Hasbi Firmansyah, Wahyu Asriyani, Ali Sofyan

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

DOI : <https://doi.org/10.32672/mister.v3i1.4004>

How to Cite this Article

APA : Sidik, M. F., Firmansyah M.Kom, H., Asriyani M.Pd, W., & Sofyan M.Kom, A. (2025). Penerapan Algoritma C4.5 pada RapidMiner untuk Resiko Kesehatan Ibu Hamil. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 823 – 832. <https://doi.org/10.32672/mister.v3i1.4004>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister/index>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Penerapan Algoritma C4.5 pada RapidMiner untuk Resiko Kesehatan Ibu Hamil

Muhammad Fajar Sidik¹, Hasbi Firmansyah², Wahyu Asriyani³, Ali Sofyan⁴

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Pancasakti Tegal, Tegal, Indonesia^{1,2,3,4}

*Email:

mfajarsidik64@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com, alisofyan.bbs@mail.com

Diterima: 08-12-2025

| Disetujui: 18-12-2025

| Diterbitkan: 20-12-2025

Abstract

Identifying delivery risk in pregnant women is essential for early detection of potential pregnancy complications. However, manual risk assessment is often limited by subjective interpretation and the complexity of interrelated health parameters. This study aims to apply the C4.5 algorithm to classify delivery risk levels using a maternal health risk dataset containing blood pressure, blood sugar, body temperature, heart rate, and maternal age. The research method includes data preprocessing, decision tree construction, and model evaluation through a confusion matrix. The results show that SystolicBP and Blood Sugar (BS) are the most influential attributes in determining risk categories. The C4.5 model achieves an accuracy of 73.67%, with the best performance observed in the high-risk and low-risk classes. Although the mid-risk class demonstrates lower recall, the model still provides an objective preliminary overview of pregnancy risk levels. These findings indicate that the C4.5 algorithm can support health practitioners by offering a data-driven tool for evaluating delivery risk.

Keywords: data mining, C4.5 algorithm, pregnancy risk, classification, maternal health

Abstrak

Penentuan risiko persalinan pada ibu hamil merupakan aspek penting dalam upaya deteksi dini komplikasi kehamilan. Namun, proses identifikasi risiko sering terkendala oleh keterbatasan analisis manual dan kompleksitas faktor kesehatan yang saling berpengaruh. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam melakukan klasifikasi tingkat risiko persalinan berdasarkan dataset maternal health risk yang mencakup parameter seperti tekanan darah, kadar gula darah, suhu tubuh, denyut jantung, dan usia ibu. Metode penelitian meliputi preprocessing data, pembentukan model pohon keputusan, dan evaluasi performa menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa atribut SystolicBP dan Blood Sugar (BS) menjadi faktor paling dominan dalam proses klasifikasi. Model C4.5 menghasilkan akurasi sebesar 73,67%, dengan performa tertinggi pada kelas high risk dan low risk. Meskipun kelas mid risk menunjukkan nilai recall yang lebih rendah, model tetap mampu memberikan gambaran awal mengenai tingkat risiko kehamilan secara lebih objektif. Temuan ini menegaskan bahwa algoritma C4.5 dapat mendukung proses evaluasi risiko persalinan sebagai alat bantu analisis bagi tenaga kesehatan.

Kata kunci: data mining, C4.5, risiko kehamilan, klasifikasi, maternal health

PENDAHULUAN

Risiko kesehatan ibu hamil merupakan salah satu tantangan prioritas dalam pembangunan kesehatan di Indonesia, terutama karena tingginya angka komplikasi kehamilan yang dapat berujung pada kematian ibu maupun bayi. Deteksi dini terhadap faktor risiko sangat penting untuk meminimalkan potensi komplikasi melalui tindakan pencegahan dan intervensi medis yang tepat. Dalam beberapa tahun terakhir, pemanfaatan *data mining* telah meningkat signifikan di sektor kesehatan karena kemampuannya dalam mengolah data kompleks dan besar untuk menghasilkan wawasan prediktif yang akurat. Algoritma C4.5 menjadi salah satu metode klasifikasi yang paling banyak digunakan pada penelitian terapan di Indonesia, terutama karena kemampuannya menangani data kategorikal dan numerik, menghasilkan pohon keputusan yang interpretatif, serta mudah diterapkan pada perangkat lunak analisis seperti RapidMiner. Beberapa penelitian lokal, dalam mengklasifikasikan faktor risiko kesehatan ibu hamil (Poernareksa & Rahmadiliyani, 2025). dalam mengidentifikasi penyakit jantung menggunakan RapidMiner (No et al., 2023), menunjukkan bahwa C4.5 dapat memberikan hasil klasifikasi yang stabil dan mudah dipahami oleh tenaga kesehatan.

Penerapan C4.5 dalam bidang kesehatan ibu dan anak juga banyak dibahas dalam berbagai studi yang menunjukkan peran penting metode ini dalam prediksi risiko persalinan, komplikasi kehamilan, serta status kesehatan balita. Poernareksa dan Rahmadiliyani (Mujiyono & Syaiful, 2023) membuktikan bahwa algoritma C4.5 dapat mengidentifikasi status risiko persalinan secara efektif, sementara Umam et al. (Nuraisana et al., 2024) menerapkannya dalam deteksi dini stunting balita dengan akurasi yang kompetitif. Di sisi lain, penelitian kesehatan ibu seperti prediksi pre-eklampsia yang dilakukan oleh Nurrohman dan Normawati (Pada et al., 2022) memperlihatkan bahwa model pohon keputusan mampu menangkap hubungan antar variabel klinis yang signifikan. Selain itu, penelitian dari Politeknik Negeri Jember menunjukkan penggunaan C4.5 untuk menentukan metode persalinan yang paling aman berdasarkan riwayat medis ibu (Yamantri & Ahmad, 2024). Penerapan lebih luas pada ranah prediksi kelahiran prematur juga ditemukan pada studi Rosida et al. (Fatah & Diagnosis, 2025), yang menegaskan bahwa struktur pohon keputusan dapat membantu proses klinis dalam memahami determinan risiko tanpa memerlukan pendekatan statistik yang kompleks. Penelitian-penelitian lain, seperti yang dilakukan oleh peneliti di Universitas Wiraraja (Keputusan et al., 2024), juga mempertegas bahwa C4.5 merupakan salah satu algoritma yang unggul dalam klasifikasi data kesehatan berbasis atribut klinis dan demografis.

Melihat tren tersebut, potensi penerapan C4.5 untuk mengklasifikasikan risiko kesehatan ibu hamil menjadi semakin relevan, terutama di era digitalisasi layanan kesehatan. Model C4.5 tidak hanya memberikan akurasi yang baik, tetapi juga menyediakan *explainable output* dalam bentuk rule atau pohon keputusan yang dapat dimanfaatkan oleh tenaga medis untuk memahami pola faktor risiko secara lebih praktis. Dukungan perangkat seperti RapidMiner juga memungkinkan proses pelatihan model yang lebih cepat dan mudah tanpa memerlukan kemampuan pemrograman yang mendalam, sebagaimana ditunjukkan oleh berbagai penelitian terapan C4.5 pada bidang kesehatan dan layanan publik (Handayani et al., 2021), (Wahid, 2025). Oleh karena itu, penelitian ini berupaya menerapkan algoritma C4.5 menggunakan RapidMiner dalam mengklasifikasikan risiko kesehatan ibu berdasarkan atribut klinis, dengan tujuan menghasilkan model yang akurat, mudah ditafsirkan, dan dapat digunakan sebagai alat bantu pengambilan keputusan dalam pelayanan kesehatan ibu hamil. Dengan memanfaatkan temuan dari penelitian-penelitian sebelumnya, studi ini diharapkan dapat berkontribusi pada pengembangan sistem pendukung keputusan yang lebih responsif dan efisien dalam mendeteksi risiko kesehatan ibu di fasilitas pelayanan kesehatan.

METODE PENELITIAN

Metode penelitian pada studi ini dirancang untuk menghasilkan model klasifikasi risiko persalinan pada ibu hamil menggunakan algoritma C4.5. Tahapan penelitian dilakukan melalui beberapa fase utama, yaitu pengumpulan dataset, preprocessing data, pembentukan model menggunakan algoritma C4.5, dan evaluasi performa model dengan metrik klasifikasi. Alur lengkap metode penelitian digambarkan dalam bentuk tahapan berurutan agar proses analisis dapat dipahami secara komprehensif.

Analisa Permasalahan

Permasalahan utama dalam penelitian ini adalah bagaimana memprediksi tingkat risiko kesehatan ibu berdasarkan variabel fisiologis seperti tekanan darah, kadar gula, detak jantung, suhu tubuh, dan usia. Permasalahan ini muncul karena banyak kasus komplikasi kehamilan tidak terdeteksi lebih awal akibat kurangnya sistem klasifikasi otomatis berbasis data kesehatan ibu (Sarjana & Informatika, 2022). Penggunaan data mining menjadi solusi efektif untuk mengidentifikasi pola kesehatan dan melakukan klasifikasi risiko secara akurat (Octari et al., 2024).

Algoritma C4.5 dipilih karena keunggulannya dalam menangani atribut numerik, kemampuan menghasilkan model interpretable, serta performa klasifikasi yang stabil pada kasus kesehatan, sebagaimana dibuktikan pada penelitian prediksi risiko kehamilan dan deteksi penyakit berbasis decision tree (Arkamil & Jambak, 2023), (Octaviani et al., 2025). Sistem klasifikasi berbasis machine learning juga terbukti mampu meningkatkan deteksi dini risiko kehamilan dengan akurasi yang lebih tinggi dibandingkan analisis konvensional (Salim et al., 2024).

Penelitian ini dilakukan melalui beberapa tahapan, yang diilustrasikan pada diagram alur berikut:



Gambar 1 Alur Flow

1. Pengumpulan Dataset

Table 1 Dataset Maternal Health Risk

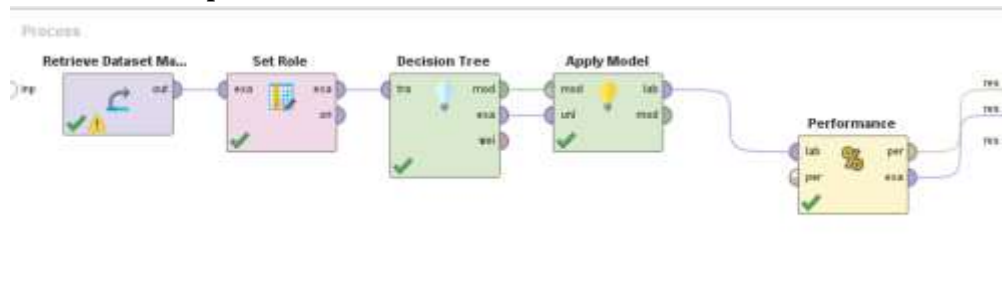
No	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
1	25	130	80	15	98	86	high risk
2	35	140	90	13	98	70	high risk
3	29	90	70	8	100	80	high risk
4	30	140	85	7	98	70	high risk
5	35	120	60	6.1	98	76	low risk
6	23	140	80	7.01	98	70	high risk
7	23	130	70	7.01	98	78	mid risk
8	35	85	60	11	102	86	high risk
9	32	120	90	6.9	98	70	mid risk
10	42	130	80	18	98	70	high risk
....			...		...		...
1015	32	120	65	6	101	76	mid risk

Pada tahap awal penelitian, data yang digunakan adalah dataset *Maternal Health Risk* yang berisi sejumlah atribut penting seperti **Age**, **Systolic Blood Pressure**, **Diastolic Blood Pressure**, **Heart Rate**, **Body Temperature**, serta **Blood Sugar (BS)**. Dataset tersebut diperoleh dari sumber terpercaya seperti **UCI Machine Learning Repository** atau dari hasil pengumpulan data di lapangan. Dataset ini dipilih karena memiliki karakteristik lengkap untuk memprediksi tingkat risiko kesehatan ibu hamil berdasarkan variabel fisiologis yang berkaitan langsung dengan kondisi kehamilan.

2. Preprocessing Data

Sebelum data digunakan untuk pemodelan, dilakukan proses *preprocessing* agar dataset siap diproses oleh algoritma. Tahapan ini meliputi kegiatan **cleaning**, yaitu memperbaiki dan menangani *missing value* atau data kosong yang dapat memengaruhi hasil model. Selain itu dilakukan pula **normalization** apabila rentang nilai antaratribut tidak seragam sehingga model dapat melakukan perhitungan secara optimal. Selanjutnya dilakukan **labeling**, yaitu menentukan variabel target atau atribut yang akan diprediksi, yakni **Risk Level** yang diklasifikasikan menjadi Low, Medium, dan High. Jika ditemukan data ekstrem, maka dilakukan **outlier detection** guna memastikan data berada dalam rentang yang wajar sehingga proses pelatihan model berjalan stabil.

3. Pengolahan Data di RapidMiner



Gambar 2 Tahapan Proses

Proses analisis dilakukan menggunakan perangkat lunak **RapidMiner**, sesuai dengan alur pada gambar proses yang terdiri dari beberapa operator utama. Tahap pertama adalah **Retrieve Dataset**, yaitu mengambil dataset yang telah dipersiapkan. Kemudian dilakukan **Set Role**, yaitu menentukan atribut mana yang berfungsi sebagai *label* atau variabel target. Setelah itu, dataset digunakan untuk melatih model dengan operator **Decision Tree**, yang dalam penelitian ini menggunakan algoritma **C4.5**. Model yang terbentuk kemudian diterapkan pada data menggunakan operator **Apply Model**, dan hasilnya dievaluasi melalui operator **Performance** yang menghasilkan metrik penting seperti akurasi dan pengukuran lainnya.

4. Pembentukan Model Decision Tree (C4.5)

Pada tahap ini RapidMiner membentuk struktur pohon keputusan berdasarkan prinsip-prinsip algoritma C4.5. Proses pemilihan atribut dilakukan menggunakan perhitungan entropy, information gain, dan gain ratio, untuk mencari atribut dengan kemampuan terbaik dalam membagi data. Atribut dengan nilai gain ratio tertinggi akan dipilih sebagai root node, diikuti dengan pembentukan node-node cabang hingga seluruh data dapat dipetakan pada kelas risiko yang tepat. Hasil proses ini berupa struktur pohon keputusan yang menunjukkan hubungan logis antar variabel terhadap tingkat risiko.

5. Penerapan Model dan Evaluasi

Model yang telah terbentuk kemudian diterapkan pada data uji menggunakan operator **Apply Model**. Setelah itu dilakukan evaluasi performa model menggunakan operator **Performance**, yang menghasilkan berbagai metrik pengukuran. Evaluasi mencakup **akurasi**, **confusion matrix**, **precision**, **recall**, serta **F1-score**, sehingga kualitas model dapat dianalisis secara menyeluruh. Pengukuran ini memastikan bahwa model mampu memprediksi tingkat risiko kesehatan ibu hamil dengan baik dan dapat berfungsi secara konsisten.

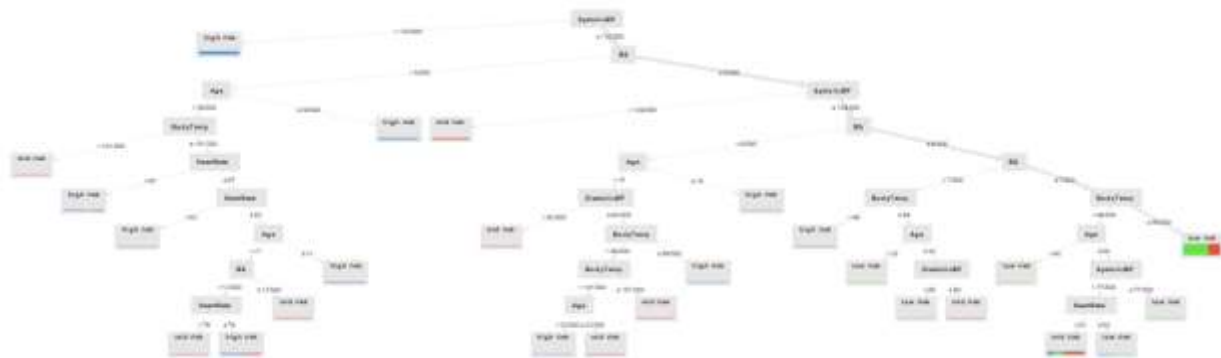
6. Interpretasi Hasil dan Analisis Prediksi

Tahapan terakhir adalah menginterpretasi hasil dari pohon keputusan yang terbentuk. Pohon keputusan menghasilkan sejumlah **aturan keputusan (rules)** yang digunakan untuk memprediksi tingkat risiko berdasarkan nilai-nilai atribut. Dari interpretasi tersebut dapat diidentifikasi atribut mana yang memiliki pengaruh paling signifikan terhadap risiko kesehatan ibu hamil. Selain itu, hasil prediksi yang dihasilkan model juga dianalisis untuk melihat seberapa baik sistem mendeteksi kategori risiko Low, Medium, atau High.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pemodelan algoritma C4.5 menggunakan RapidMiner beserta pembahasannya. Seluruh hasil percobaan ditampilkan dalam bentuk gambar dan tabel sesuai standar penulisan. Analisis dilakukan untuk melihat kinerja model dalam memprediksi risiko kesehatan ibu berdasarkan dataset maternal health risk.

Hasil Permodelan Algoritma C4.5



Gambar 3 Hasil Pohon Keputusan (Decision Tree)

Pemodelan algoritma C4.5 menghasilkan struktur pohon keputusan yang memberikan gambaran menyeluruh mengenai bagaimana atribut-atribut fisiologis yang ada pada dataset digunakan untuk menentukan tingkat risiko maternal. Pada hasil visualisasi yang ditampilkan melalui perangkat RapidMiner, atribut **Systolic Blood Pressure (SystolicBP)** dipilih sebagai node akar (*root node*). Terpilihnya atribut ini menunjukkan bahwa tekanan darah sistolik merupakan faktor dominan dalam proses pemilihan data serta menjadi indikator awal yang paling berpengaruh terhadap penentuan tingkat risiko. Hal ini sejalan dengan banyak penelitian medis sebelumnya yang menyatakan bahwa tekanan darah tinggi merupakan salah satu faktor paling signifikan dalam penilaian risiko kehamilan.

Setelah proses pemilihan pada node akar, pohon keputusan bercabang menuju atribut lain seperti **Blood Sugar (BS)**, **Age**, **Body Temperature**, **DiastolicBP**, dan **HeartRate**. Masing-masing atribut tersebut berperan dalam mempersempit kategori risiko hingga akhirnya menghasilkan keputusan akhir berupa *low risk*, *mid risk*, atau *high risk*. Setiap cabang yang terbentuk menggambarkan keterkaitan antara kondisi fisiologis ibu dengan tingkat risiko kehamilan. Misalnya, nilai SystolicBP di atas ambang tertentu dikombinasikan dengan nilai BS yang tinggi cenderung mengarah pada risiko tinggi (*high risk*), sedangkan usia muda dan tekanan darah normal lebih cenderung mengarah pada risiko rendah (*low risk*).

Selain itu, struktur pohon yang terbentuk juga menunjukkan bahwa beberapa atribut muncul berulang pada berbagai cabang, menandakan bahwa informasi pada atribut tersebut sangat berkontribusi dalam keseluruhan proses klasifikasi. Contohnya, atribut BS muncul di beberapa level pohon, yang mengindikasikan bahwa kadar gula darah memiliki peran penting dalam membedakan risiko *mid risk* dari *high risk*. Atribut Age juga muncul sebagai faktor penentu antara risiko rendah dan menengah, yang menunjukkan bahwa usia ibu tetap menjadi variabel penting meskipun kontribusinya tidak sekuat tekanan darah.

Secara keseluruhan, hasil pemodelan ini tidak hanya menghasilkan pohon keputusan yang komprehensif, tetapi juga memberikan wawasan mengenai hubungan antarvariabel fisiologis yang dapat digunakan sebagai referensi tambahan dalam proses penilaian klinis. Kejelasan struktur pohon juga membuat model lebih mudah dipahami, karena proses pengambilan keputusan dapat ditelusuri secara jelas dari akar hingga

Daun.

Hasil Evaluasi Model

PerformanceVector

```
PerformanceVector:
accuracy: 73.67%
ConfusionMatrix:
True:  high risk      low risk      mid risk
high risk:    236      4        26
low risk:     12      371      170
mid risk:     24      31       140
classification_error: 26.33%
ConfusionMatrix:
True:  high risk      low risk      mid risk
high risk:    236      4        26
low risk:     12      371      170
mid risk:     24      31       140
```

Gambar 4 Hasil Performance Vector

Berdasarkan hasil *Performance Vector* yang dihasilkan oleh RapidMiner, model klasifikasi menggunakan algoritma C4.5 memperoleh tingkat akurasi sebesar **73,67%**, yang menunjukkan bahwa sebagian besar data uji berhasil diprediksi sesuai dengan kelas risiko sebenarnya. Evaluasi performa ditampilkan melalui *confusion matrix* yang membandingkan prediksi model dengan kondisi aktual untuk tiga kategori risiko, yaitu **high risk**, **low risk**, dan **mid risk**. Pada kelas *high risk*, model berhasil mengidentifikasi 236 kasus dengan benar, meskipun masih terdapat kesalahan klasifikasi sebanyak 4 kasus yang diprediksi sebagai *low risk* dan 26 kasus yang diprediksi sebagai *mid risk*. Untuk kelas *low risk*, performa model menunjukkan hasil sangat baik dengan 371 prediksi benar, namun terdapat 12 kasus yang salah diprediksi sebagai *high risk* dan 170 kasus salah diklasifikasikan sebagai *mid risk*. Sementara itu, tingkat kesalahan terbesar terlihat pada kelas *mid risk*, di mana hanya 140 kasus berhasil diprediksi dengan benar dan sisanya—24 kasus diprediksi sebagai *high risk* serta 31 kasus diprediksi sebagai *low risk*.

Keseluruhan kesalahan klasifikasi sebesar **26,33%** menunjukkan bahwa atribut pada kelas *mid risk* cenderung lebih sulit dipisahkan secara jelas oleh model C4.5, sehingga sering terjadi overlapping dengan dua kelas lainnya. Meski demikian, kinerja yang ditampilkan pada kelas *high risk* dan *low risk* tetap menunjukkan konsistensi yang baik. Secara umum, hasil evaluasi ini menggambarkan bahwa algoritma C4.5 mampu memberikan performa klasifikasi yang cukup efektif untuk memetakan tingkat risiko kesehatan maternal, terutama pada kategori risiko tinggi dan rendah, meskipun diperlukan peningkatan lebih lanjut pada akurasi kategori risiko menengah.

Tabel 1 Hasil Evaluasi Model C4.5

Accuracy:73,67%

	True high risk	True low risk	True mid risk	Class precision
Pred. high risk	236	4	26	88.72%
Pred. low risk	12	371	170	67.09%
Pred. mid risk	24	31	140	71.79
Class recall	86.76%	91.38%	41.67%	

Evaluasi performa model dilakukan menggunakan metode **split validation** dengan pembagian 80% data sebagai data latih dan 20% data sebagai data uji. Melalui tahap evaluasi ini, diperoleh nilai akurasi sebesar **73,67%**, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan tepat. Nilai akurasi ini

dapat dikategorikan cukup baik untuk dataset medis yang bersifat kompleks dan memiliki batas antar Kelas yang relatif berdekatan.

Dari hasil confusion matrix, terlihat bahwa model memberikan performa yang berbeda-beda pada masing-masing kelas risiko. Pada kelas **high risk**, model menunjukkan performa sangat baik dengan nilai precision sebesar 88,72% dan recall sebesar 86,76%. Tingginya nilai pada kedua metrik ini menunjukkan bahwa model mampu mengenali sebagian besar kasus risiko tinggi secara konsisten, dengan kesalahan prediksi yang minimal. Hal ini sangat penting mengingat deteksi kasus risiko tinggi merupakan prioritas utama dalam sistem pendukung keputusan maternal.

Pada kelas **low risk**, recall mencapai 91,38%, yang berarti model mampu mengidentifikasi sebagian besar data risiko rendah secara benar. Namun, precision pada kelas ini tidak setinggi kelas high risk, yang menunjukkan bahwa beberapa data dari kelas lain kemungkinan terprediksi sebagai low risk. Kondisi ini masih bisa diterima dalam konteks kesehatan, namun perlu perhatian jika diterapkan pada skenario nyata yang membutuhkan ketelitian lebih tinggi pada setiap kelas.

Hasil berbeda ditemukan pada kelas **mid risk**, dengan nilai recall hanya sebesar 41,67%. Rendahnya nilai recall menunjukkan bahwa model mengalami kesulitan dalam mengenali data dengan risiko menengah dan sering salah mengelompokkannya ke kelas risiko rendah atau tinggi. Hal ini dapat disebabkan oleh distribusi data yang tidak seimbang atau pola fisiologis pada kelas risiko menengah yang tidak terlalu berbeda dari dua kelas lainnya. Kondisi ini umum terjadi pada dataset medis yang memiliki rentang nilai fisiologis yang saling tumpang tindih.

Dengan demikian, meskipun nilai akurasi keseluruhan cukup baik, hasil evaluasi per kelas menunjukkan bahwa model membutuhkan penyempurnaan lebih lanjut agar dapat memberikan performa yang seimbang pada semua kelas risiko, terutama pada kelas mid risk.

Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma C4.5 merupakan metode yang efektif untuk mengklasifikasikan tingkat risiko maternal berdasarkan variabel fisiologis seperti tekanan darah, kadar gula darah, usia ibu, hingga suhu tubuh dan denyut jantung. Keunggulan utama dari algoritma ini adalah kemampuannya menghasilkan model yang mudah diinterpretasikan melalui struktur pohon keputusan. Interpretabilitas ini sangat penting dalam bidang kesehatan karena praktisi medis membutuhkan pemahaman yang jelas mengenai faktor-faktor apa saja yang berperan dalam penentuan risiko.

Dari struktur pohon yang dihasilkan, terlihat bahwa beberapa atribut memiliki pengaruh lebih besar dibandingkan atribut lainnya. Hal ini terlihat dari kemunculan atribut SystolicBP dan BS pada beberapa cabang pohon. Pola ini menunjukkan bahwa kondisi tekanan darah dan kadar gula darah merupakan indikator penting dalam menentukan tingkat risiko ibu, terutama risiko tinggi. Selain itu, atribut usia dan suhu tubuh turut berkontribusi dalam membedakan risiko rendah dari risiko menengah. Secara keseluruhan, struktur pohon memberikan gambaran yang logis dan selaras dengan penelitian medis sebelumnya, sehingga hasil ini dapat dipercaya dan dijadikan rujukan.

Namun demikian, performa model yang kurang memuaskan pada kelas mid risk menunjukkan adanya tantangan yang perlu diperhatikan lebih dalam. Rendahnya nilai recall pada kelas ini mengindikasikan bahwa batas antara nilai-nilai fisiologis pada kelas mid risk dan kelas lain relatif kabur. Hal ini dapat diperburuk oleh ketidakseimbangan jumlah data pada masing-masing kelas. Oleh karena itu, penelitian lanjutan dapat fokus pada perbaikan distribusi kelas, penggunaan teknik oversampling seperti SMOTE, atau bahkan memanfaatkan algoritma lain yang lebih kuat seperti Random Forest atau Gradient Boosting.

Selain itu, interpretasi hasil evaluasi ini juga memberikan wawasan tambahan mengenai bagaimana model dapat digunakan dalam praktik. Model dengan performa tinggi pada kelas high risk dapat sangat membantu dalam mendeteksi kondisi berbahaya secara cepat. Meskipun demikian, upaya peningkatan diperlukan agar model tidak mengabaikan kasus risiko menengah yang juga memiliki potensi berkembang menjadi kondisi yang lebih serius.

Secara keseluruhan, model C4.5 yang digunakan dalam penelitian ini dapat berfungsi sebagai alat bantu prediksi awal yang cukup akurat, transparan, dan relevan. Model ini dapat dikembangkan lebih lanjut untuk

diintegrasikan ke dalam sistem pendukung keputusan (Decision Support System) di bidang kesehatan maternal guna membantu tenaga medis dalam proses identifikasi risiko secara lebih efektif dan berbasis data.

KESIMPULAN

Penelitian ini membuktikan bahwa algoritma **C4.5** dapat diterapkan secara efektif untuk melakukan klasifikasi tingkat risiko persalinan pada ibu hamil dengan menggunakan dataset maternal health risk yang mencakup parameter tekanan darah, kadar gula darah, suhu tubuh, denyut jantung, dan usia ibu. Hasil analisis menunjukkan bahwa atribut **SystolicBP** dan **Blood Sugar (BS)** merupakan faktor paling dominan yang memengaruhi klasifikasi risiko, sehingga penilaian risiko persalinan dapat lebih fokus pada parameter ini. Model C4.5 yang dibangun berhasil mencapai **akurasi sebesar 73,67%**, dengan performa terbaik pada kelas *high risk* dan *low risk*, sementara kelas *mid risk* memiliki nilai recall yang lebih rendah.

Temuan ini menegaskan bahwa penggunaan algoritma C4.5 tidak hanya mampu memberikan gambaran awal mengenai tingkat risiko kehamilan secara lebih objektif, tetapi juga berpotensi menjadi alat bantu analisis yang praktis bagi tenaga kesehatan dalam proses deteksi dini komplikasi persalinan. Dengan implementasi model ini, tenaga medis dapat melakukan pemantauan yang lebih terstruktur, meningkatkan efektivitas pengambilan keputusan, serta membantu prioritas tindakan medis bagi ibu hamil yang berisiko tinggi. Meskipun masih terdapat keterbatasan pada akurasi kelas *mid risk*, penelitian ini memberikan landasan untuk pengembangan model prediktif yang lebih kompleks dengan tambahan parameter kesehatan atau dataset yang lebih besar di masa depan.

DAFTAR PUSTAKA

- Arkamil, R. F., & Jambak, M. I. (2023). *Klasifikasi Tindakan Persalinan Pada Pasien Ibu Bersalin Menggunakan Metode Decision Tree C4 . 5*. 4(1), 513–523. <https://doi.org/10.30865/klik.v4i1.1168>
- Fatah, Z., & Diagnosis, P. (2025). *Penerapan Data Mining Untuk Prediksi Diagnosis Demam Berdarah Dengan Algoritma Decision Tree*. 3(May), 74–82. <https://doi.org/10.59435/jocstec.v3i2.437>
- Handayani, K., Latif, A., Firdaus, M. R., Hasan, F. N., Informasi, T., Bina, U., Informatika, S., Akuntansi, S. I., Bina, U., Informatika, S., Komputer, T., Bina, U., Informatika, S., Komputer, I., Bina, U., Informatika, S., & Pusat, J. (2021). *Komparasi Algoritma C4 . 5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah*. 10, 676–687. <https://doi.org/10.32520/stmsi.v10i3.1440>
- Keputusan, O., Persalinan, M., Alfianah, A. B., Pratama, M. R., Yunus, M., Vestine, V., Studi, P., Informasi, M., Kesehatan, J., & Jember, P. N. (2024). *Optimasi Keputusan Metode Persalinan dengan Algoritma C4.5*. 6(1), 77–87. <https://doi.org/10.25047/j-remi.v6i1.4777>
- Mujiyono, S., & Syaiful, N. (2023). *Prediction of Nutritional Status of Toddlers Using C4 . 5 Algorithm*. 12(1), 41–52. <https://doi.org/10.35335/mandiri.v12i1.229>
- No, V., Hal, A., Agus, I. M., Gunawan, O., Ayu, I. D., Saraswati, I., Gede, I. D., Agung, R., & Eka, I. P. (2023). *Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4 . 5 Dengan Rapidminer*. 5(2), 73–83. <https://doi.org/10.47233/jteksis.v5i2.775>
- Nuraisana, N., Halawa, S. W., & Harun, M. (2024). *Implementasi Algoritma C4 . 5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Sekip Lubuk Pakam*. 23, 130–137. <https://doi.org/10.53513/jis.v23i1.9596>
- Octari, T., Syesar, P., Febriansyah, F., & Efan, E. (2024). *PREDIKSI METODE PERSALINAN MENGGUNAKAN DECISION SUPPORT PREDICTION OF DELIVERY METHOD USING DECISION SUPPORT SYSTEM AND*. 01(02), 12–19. <https://doi.org/10.47927/jikb.v13i2a.376>

- Octaviani, D. A., Widiastuti, D., Amelia, R., Salam, A., & Nuswantoro, U. D. (2025). *Implementasi Data Mining Untuk Memprediksi Dalam Kehamilan Poltekkes Kemenkes Semarang , Indonesia Implementation Of Data Mining To Predict Pre-Eclampsia In Pregnancy* Alamat korespondensi : 18(1), 29–38. <https://doi.org/10.32763/ps5qpj75>
- Pada, S., Kesehatan, P. O. S., Pkd, D., & Website, B. (2022). *Implementasi algoritma c4.5 untuk mendeteksi stunting pada pos kesehatan desa (pkd) berbasis website (studi kasus : desa purbayasa kecamatan tonjong)*. <https://doi.org/10.58436/ijir.v4i2.1907>
- Poernareksa, D., & Rahmadiliyani, N. (2025). *JTIM : Jurnal Teknologi Informasi dan Multimedia Implementasi Data Mining dalam Menentukan Prediksi Status Resiko Persalinan pada Ibu Hamil menggunakan Algoritma*. 7(1), 123–132. <https://doi.org/10.35746/jtim.v7i1.619>
- Salim, S., Mashrafi, A., & Tafakori, L. (2024). Predicting maternal risk level using machine learning models. *BMC Pregnancy and Childbirth*. <https://doi.org/10.1186/s12884-024-07030-9>
- Sarjana, J., & Informatika, T. (2022). *Prediksi Dini Penyakit Preeklamsia Menggunakan Algoritma*. 10(3), 120–131. <https://doi.org/10.12928/jstie.v10i3.24187>
- Wahid, M. F. (2025). *Penerapan Data Mining Menggunakan Algoritma C4 . 5 untuk Klasifikasi Penyakit Paru-Paru Universitas Ibrahimy*. 3(5), 646–653. <https://doi.org/10.61722/jipm.v3i5.1513>
- Yamantri, A. B., & Ahmad, A. (2024). *Penerapan Algoritma C4 . 5 Untuk Prediksi Faktor Risiko Obesitas Pada Penduduk Dewasa*. 2(c), 118–125. <https://doi.org/10.70052/jka.v2i3.341>