

Prediksi Tingkat Resiko Kesehatan Maternal Menggunakan Algoritma Pohon Keputusan C4.5

Achmad Zaidan¹, Balvas Mubarak², Muhamad Rafly Ramadhani³, Wisti Dwi Septiani⁴, Syaifur Rahmatullah Abdul R⁵

^{1,2,3,4,5} Teknologi Informasi, Universitas Bina Sarana Informatika

e-mail: 17230747@bsi.ac.id¹, 17231060@bsi.ac.id², 17230343@bsi.ac.id³,
wisti.wst@bsi.ac.id⁴, syaifur.syl@bsi.ac.id⁵

Abstrak

Penelitian ini bertujuan untuk mengembangkan model prediksi tingkat risiko kesehatan maternal menggunakan algoritma pohon keputusan C4.5. Risiko komplikasi pada ibu hamil merupakan isu penting dalam pelayanan kesehatan, sehingga deteksi dini diperlukan untuk menurunkan morbiditas dan meningkatkan tindakan pencegahan. Data yang digunakan mencakup parameter fisiologis seperti tekanan darah sistolik, tekanan darah diastolik, kadar gula darah, suhu tubuh, detak jantung, dan usia ibu. Algoritma C4.5 diterapkan untuk membangun model klasifikasi yang membagi tingkat risiko menjadi kategori low risk, mid risk, dan high risk. Hasil pengujian menunjukkan bahwa model mencapai akurasi 73,67%, dengan kinerja yang baik pada kategori risiko rendah dan tinggi, meskipun prediksi pada kelas menengah masih memerlukan optimalisasi. Pohon keputusan yang dihasilkan menunjukkan bahwa tekanan darah sistolik dan kadar gula darah merupakan penentu utama dalam klasifikasi risiko maternal. Temuan ini menunjukkan bahwa algoritma C4.5 mampu memberikan model prediksi yang interpretatif dan cukup andal untuk membantu proses pengambilan keputusan klinis. Penelitian selanjutnya dapat meningkatkan performa dengan mengatasi ketidakseimbangan kelas serta menerapkan pendekatan hybrid machine learning.

Kata kunci: C4.5, Risiko Kesehatan Maternal, Pohon Keputusan, Klasifikasi, Machine Learning.

Abstract

This study aims to develop a predictive model for maternal health risk classification using the C4.5 decision tree algorithm. Maternal health complications remain a critical concern in clinical practice, making early detection essential for reducing morbidity and improving preventive care. The dataset used in this research consists of physiological parameters including systolic blood pressure, diastolic blood pressure, blood glucose level, body temperature, heart rate, and maternal age. The C4.5 algorithm was employed to construct a classification model capable of categorizing risk levels into low, medium, and high. The experimental results show that the model achieved an accuracy of 73.67%, with strong performance in identifying low-risk and high-risk categories, though classification of mid-risk cases still requires improvement. The resulting decision tree highlights key clinical indicators—particularly systolic blood pressure and blood glucose level—as the primary determinants of maternal risk classification. These findings demonstrate that the C4.5 algorithm provides an interpretable and reliable predictive framework that can support clinical decision-making for maternal health monitoring. Future studies may enhance model performance through techniques addressing class imbalance and exploration of hybrid machine learning methods.

Keywords: C4.5, Maternal Health Risk, Decision Tree, Classification, Machine Learning.

PENDAHULUAN

Kesehatan maternal merupakan salah satu indikator penting dalam menilai kualitas layanan kesehatan suatu negara, terutama karena risiko komplikasi kehamilan dapat menimbulkan kondisi serius seperti preeclampsia, hipertensi gestasional, diabetes gestasional, hingga kematian ibu apabila tidak terdeteksi sejak dini (Ningsih et al., 2025). Penelitian di Indonesia menunjukkan

bahwa factor fisiologis seperti tekanan darah, usia ibu, indeks massa tubuh, kadar gula darah, suhu tubuh, serta detak jantung merupakan prediktor signifikan terhadap risiko Kesehatan maternal (Ningsih et al., 2025). Oleh karena itu, deteksi dini terhadap risiko maternal melalui pemeriksaan atribut klinis menjadi dasar sangat penting dalam mencegah komplikasi selama kehamilan.

Perkembangan teknologi informasi membuka peluang luas dalam pengembangan sistem pendukung Keputusan di bidang Kesehatan, termasuk dalam memanfaatkan machine learning untuk analisis risiko maternal (Pinandito et al., 2023). Penerapan machine learning dalam klasifikasi kesehatan maternal telah menunjukkan performa yang cukup baik dalam mendeteksi penyakit seperti diabetes gestasional maupun komplikasi lain pada kehamilan. Beberapa studi lokal menyebutkan bahwa algoritma klasifikasi seperti Decision Tree memiliki performa lebih baik dibandingkan metode lain seperti Naïve Bayes dan KNN dalam mengklasifikasikan tingkat risiko ibu hamil (Fitra et al., 2023). Hal ini menunjukkan bahwa metode berbasis pohon keputusan relevan dan efektif digunakan dalam analisis risiko kesehatan maternal.

Algoritma C4.5 merupakan salah satu metode klasifikasi berbasis pohon keputusan yang paling banyak digunakan dalam analisis data klinis karena mampu mengolah atribut numerik dan kategorik serta menghasilkan model yang mudah diinterpretasikan oleh tenaga Kesehatan (Ramadhani et al., 2024). Interpretabilitas merupakan aspek penting dalam pengembangan model prediksi risiko maternal karena tenaga kesehatan membutuhkan model yang transparan untuk mendukung pengambilan keputusan klinis. Pembentukan pohon keputusan dalam algoritma C4.5 diawali dengan penghitungan Entropy dan Gain Ratio untuk menentukan atribut terbaik dalam proses pemisahan simpul (Octari et al., 2024).

Penelitian sebelumnya menunjukkan efektivitas C4.5 dalam memprediksi risiko preeklampsia dengan memanfaatkan variabel klinis seperti usia kehamilan, tekanan darah, dan riwayat medis, dan berhasil menghasilkan model klasifikasi dengan gambaran jalur keputusan yang jelas (Nurrohmah et al., 2022). Selain itu, pengembangan model prediksi risiko kehamilan berisiko tinggi menggunakan dataset "maternal health risk" juga menunjukkan bahwa metode machine learning memiliki akurasi lebih dari 82%, serta tingkat deteksi risiko tinggi mencapai 91% (Pi et al., 2025). Di sisi lain, penelitian berbasis data BPJS Kesehatan menunjukkan bahwa machine learning mampu melakukan prediksi risiko preeklampsia pada populasi yang lebih luas, namun beberapa model bersifat kurang transparan bagi tenaga kesehatan sehingga diperlukan metode klasifikasi yang lebih mudah dijelaskan (Sufriyana et al., 2020).

Beberapa penelitian terdahulu juga telah mengidentifikasi variabel yang sering digunakan sebagai indikator klinis utama dalam prediksi risiko maternal, yaitu usia ibu, riwayat kehamilan sebelumnya, tekanan darah (sistolik dan diastolik), indeks massa tubuh, hasil lab tertentu, serta variabel obstetrik seperti usia gestasi dan riwayat aborsi (Ranjbar et al., 2024). Hal tersebut sejalan dengan kebutuhan pengembangan model prediksi yang mampu mengakomodasi variabel klinis yang tersedia secara rutin pada layanan kesehatan primer.

Meskipun demikian, penelitian terdahulu masih memiliki beberapa keterbatasan. Sebagian penelitian hanya berfokus pada satu jenis komplikasi, seperti preeklampsia, sehingga hasilnya tidak menggambarkan tingkat risiko maternal secara lebih komprehensif (Nurrohmah et al., 2022). Selain itu, penelitian yang menggunakan data administratif berskala besar seringkali tidak mencakup variabel klinis rinci, sehingga model menjadi kurang akurat dalam menangkap faktor klinis tertentu, serta beberapa algoritma cenderung sulit diinterpretasikan oleh tenaga medis (Sufriyana et al., 2020).

Kondisi ini menegaskan perlunya penelitian yang membangun model klasifikasi risiko maternal secara umum menggunakan metode yang akurat sekaligus mudah dijelaskan. Berdasarkan celah penelitian tersebut, penelitian ini memiliki justifikasi kuat untuk mengembangkan model prediksi tingkat risiko maternal secara umum (rendah, sedang, tinggi) menggunakan algoritma pohon keputusan C4.5. Model dikembangkan berdasarkan variabel klinis rutin seperti tekanan darah, kadar gula darah, suhu tubuh, detak jantung, dan usia ibu. Dengan adanya model prediksi ini, layanan kesehatan diharapkan mampu melakukan deteksi dini risiko kesehatan ibu hamil secara lebih cepat, akurat, dan mudah diinterpretasikan dalam mendukung proses pengambilan keputusan klinis

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis algoritma pohon keputusan C4.5 untuk memprediksi tingkat risiko kesehatan maternal. Bagian metode ini menjelaskan tahapan yang dilakukan dalam pengumpulan data, pengolahan data, serta teknik analisis yang digunakan dalam membangun model prediksi.

Pada bab ini akan dijelaskan tahapan-tahapan penelitian, mulai dari tahap tinjauan pustaka hingga tahap evaluasi hasil prediksi.



Gambar 1. Tahap Penelitian

Literatur Review

Fase awal dalam penelitian ini adalah tinjauan literatur (literature review) secara komprehensif. Proses ini esensial untuk membangun kerangka teoretis dan pemahaman mendalam mengenai topik klasifikasi risiko kesehatan maternal. Literatur yang dikaji meliputi berbagai sumber primer dan sekunder, khususnya jurnal ilmiah, yang membahas secara terperinci tentang algoritma Pohon Keputusan (Decision Tree), penerapan metode klasifikasi pada data medis, serta studi-studi terdahulu terkait prediksi risiko maternal. Hasil dari tinjauan pustaka ini sangat krusial dalam menentukan metodologi penelitian yang paling relevan dan efektif untuk menjawab tujuan yang telah ditetapkan.

Pengumpulan Data

Data set yang digunakan penelitian ini Adalah data set yang berasal dari UC Irvine data set dengan judul "Maternal Health Risk". Data ini memiliki 7 atribut yang terdiri dari Umur, Tekanan Darah Sistolik, Tekanan Darah Diastolik, Kadar gula darah, Suhu Tubuh, Detak Jantung, dan Risk Level.

Penerapan Algoritma C4.5

Algoritma Pohon Keputusan C4.5 digunakan sebagai metode klasifikasi utama dalam penelitian ini. Pemilihan algoritma ini didasarkan pada tiga justifikasi penting. Pertama, C4.5 memiliki kapabilitas untuk memproses berbagai tipe data secara efisien, mencakup atribut numerik (seperti usia, tekanan darah) dan kategorikal, menjadikannya sangat cocok dengan karakteristik dataset risiko maternal. Kedua, C4.5 menonjol karena kemampuannya menghasilkan struktur pohon keputusan yang bersifat transparan dan mudah diinterpretasikan, sebuah faktor kunci dalam domain medis yang membutuhkan validasi dan pemahaman klinis terhadap hasil prediksi. Ketiga, algoritma ini secara spesifik menggunakan Gain Ratio untuk menentukan atribut terbaik saat proses pemisahan simpul (node), yang secara efektif mengurangi bias yang mungkin timbul pada atribut dengan jumlah nilai yang banyak.

Proses pembentukan pohon keputusan oleh C4.5 dimulai dengan menghitung Entropy, yaitu ukuran statistik untuk menilai tingkat heterogenitas atau ketidakmurnian himpunan data. Hasil dari perhitungan Entropy ini kemudian digunakan sebagai dasar untuk menentukan nilai Information Gain. Atribut yang menghasilkan nilai Information Gain atau Gain Ratio tertinggi akan dipilih sebagai simpul akar (root node) atau simpul pemecah terbaik pada setiap tingkat pohon.

Proses rekursif ini terus berlanjut hingga semua data berhasil terklasifikasi secara optimal, atau hingga kondisi pemberhentian (stopping criteria) lainnya terpenuhi.

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2(p_i)$$

Keterangan:

- S : Himpunan kasus
- n : Jumlah partisi S
- p_i : Proporsi dari S_i terhadap S

Gambar 2. Rumus Entropy

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- $|S_i|$: Jumlah kasus pada partisi ke- i
- $|S|$: Jumlah kasus dalam S

Gambar 3. Rumus Gain

Evaluasi Hasil

Evaluasi terhadap kinerja model klasifikasi C4.5 adalah langkah krusial untuk memastikan keandalan model dalam memprediksi risiko maternal. Model diuji menggunakan data yang belum pernah dilihat (unseen data) dan diukur dengan beberapa metrik kinerja standar. Metrik utama yang digunakan meliputi Akurasi (untuk mengukur proporsi prediksi yang benar secara keseluruhan), Precision (untuk menilai seberapa tepat model mengklasifikasikan kasus positif, misalnya 'High Risk', dari seluruh prediksi positif), dan Recall (untuk mengukur kemampuan model dalam menangkap seluruh kasus positif yang sebenarnya). Selain itu, metrik F1-Score, yang merupakan rata-rata harmonis antara Precision dan Recall, digunakan untuk memberikan penilaian yang seimbang. Seluruh metrik ini dirangkum dalam Confusion Matrix; sebuah tabel yang memberikan gambaran detail mengenai jumlah True Positive, False Positive, True Negative, dan False Negative untuk setiap kelas risiko. Penggunaan kombinasi metrik ini memungkinkan penilaian yang komprehensif terhadap tingkat keandalan dan efektivitas model C4.5 dalam mendukung proses deteksi risiko kesehatan maternal.

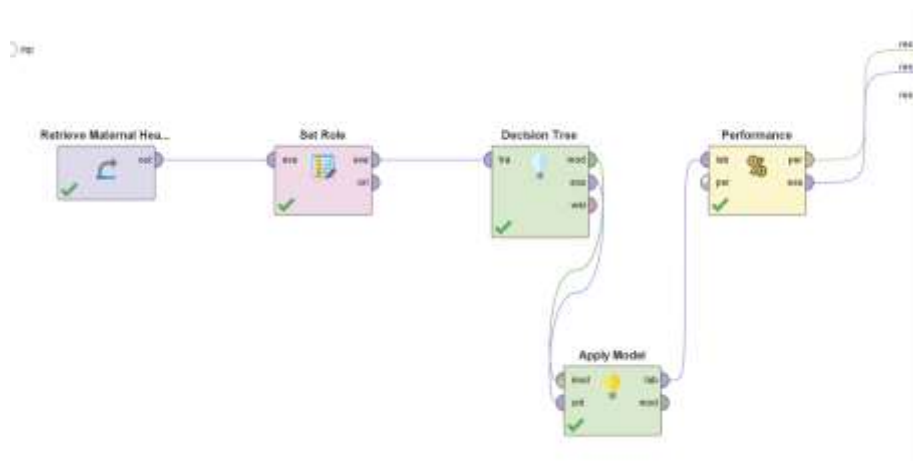
HASIL DAN PEMBAHASAN

Penelitian ini menggunakan software RapidMiner dalam melakukan proses penghitungan data dengan menggunakan algoritma decision tree. Data set yang digunakan memiliki 6 atribut dengan class target berupa Risk Level. Pada atribut risk level terbagi dari 3 kelas yaitu low risk, medium risk dan high risk.

Pada software RapidMiner Langkah pertama dalam penelitian ini Adalah memasukan dataset. Setelah dataset berhasil dimasukan, selanjutnya adalah menentukan rolenya dengan set role pada atribut. Dalam penelitian ini atribut yang akan diberikan role Adalah atribut Risk Level dengan role berupa label. Pada gambar 5 adalah dataset yang telah dimasukkan dan diolah pada software RapidMiner. Data yang sudah diolah ini akan digunakan untuk melakukan perhitungan algoritma Decision Tree.

Row No.	RiskLevel	Age	Tekanan D...	Tekanan D...	Kadar gula...	Berat Tubuh	Detak Jant...
1	high risk	25	130	80	15	58	80
2	high risk	35	140	90	13	58	70
3	high risk	25	90	70	8	100	80
4	high risk	30	140	85	7	58	70
5	low risk	35	120	60	8.100	58	70
6	high risk	23	140	80	7.010	58	70
7	mid risk	23	130	70	7.010	58	78
8	high risk	35	85	60	11	100	80
9	mid risk	32	120	90	6.900	58	70
10	high risk	42	130	80	18	58	70
11	low risk	23	90	60	7.010	58	76
12	mid risk	19	120	80	7	58	70
13	low risk	25	110	89	7.010	58	77
14	mid risk	25	120	75	7.010	100	70
15	mid risk	48	120	80	11	58	88
16	low risk	15	120	80	7.010	58	70
17	high risk	50	140	90	15	58	90
18	high risk	25	140	100	7.010	58	80
19	mid risk	30	120	80	8.900	101	70

Gambar 4. Dataset yang telah diolah



Gambar 5. Model Algoritma yang diuji

Gambar 6 merupakan model yang melakukan pengujian perhitungan Decision Tree. Dalam Menyusun model terdapat operator yang diperlukan yaitu retrieve, set role, Decision Tree, Apply Model dan Peformance.

accuracy: 73.67%

	true high risk	true low risk	true mid risk	class precision
pred. high risk	236	4	25	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%	

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%	

Gambar 6. Akurasi Model

Hasil pengujian model menunjukkan bahwa algoritma C4.5 mampu menghasilkan tingkat akurasi sebesar 73,67% dalam mengklasifikasikan tingkat risiko kesehatan maternal ke dalam tiga kategori, yaitu low risk, mid risk, dan high risk. Berdasarkan confusion matrix, kelas low risk memiliki nilai recall tertinggi sebesar 91,38%, yang menunjukkan bahwa sebagian besar kasus risiko rendah dapat dikenali dengan baik oleh model. Pada kelas high risk, model juga menunjukkan kinerja yang baik dengan recall sebesar 86,76%, menandakan kemampuan yang cukup kuat dalam mendeteksi kasus berisiko tinggi. Namun, performa model pada kelas mid risk masih relatif rendah dengan recall 41,67%, yang mengindikasikan bahwa model masih mengalami kesulitan membedakan data pada kategori risiko menengah. Nilai precision tertinggi terdapat pada kelas high risk (88,72%), diikuti mid risk (71,79%) dan low risk (67,09%). Secara keseluruhan, hasil ini menunjukkan bahwa model C4.5 bekerja cukup baik terutama dalam mengenali kategori ekstrem (risiko rendah dan tinggi), tetapi memerlukan penyempurnaan lebih lanjut untuk meningkatkan akurasi pada kelas risiko menengah.



Gambar 7. Pohon Keputusan

Berdasarkan pohon Keputusan pada gambar 7 :

- Jika Tekanan Darah Sistolik > 132.5 , maka individu langsung diklasifikasikan sebagai high risk.
- Jika Tekanan Darah Sistolik ≤ 132.5 , maka keputusan berikutnya bergantung pada kadar gula darah.
- Jika kadar gula darah > 9.5 dan usia > 38.5 , maka model selanjutnya melihat suhu tubuh; apabila suhu tubuh > 101.5 , individu diklasifikasikan sebagai mid risk, sedangkan jika suhu tubuh ≤ 101.5 , penentuan risiko dipengaruhi oleh detak jantung.
- Jika detak jantung > 87 , maka status risiko adalah high risk; namun jika ≤ 87 , model mengevaluasi ambang lain yaitu detak jantung 83 bpm dan kemudian usia serta kadar gula darah untuk menentukan apakah hasilnya high risk atau mid risk.
- Jika usia ≤ 38.5 , maka individu otomatis masuk kategori high risk.
- Jika kadar gula darah ≤ 9.5 , model kembali melihat tekanan darah sistolik; jika > 129.5 , maka individu diklasifikasikan mid risk, sedangkan jika ≤ 129.5 , evaluasi dilanjutkan dengan kadar gula darah ambang 8.5 dan variabel lanjutan seperti usia, tekanan darah diastolik, suhu tubuh, dan detak jantung, yang menentukan apakah hasil akhirnya low risk, mid risk, atau high risk.

- Pada jalur kadar gula rendah (≤ 7.95), model melakukan percabangan akhir menggunakan variabel usia, tekanan darah sistolik, detak jantung, dan suhu tubuh, menghasilkan klasifikasi paling rinci terutama pada kategori low risk dan mid risk, serta mendeteksi kombinasi variabel tertentu yang masih masuk kategori high risk.

Interpretasi hasil penelitian menunjukkan bahwa algoritma C4.5 mampu memprediksi tingkat risiko kesehatan maternal dengan performa yang sangat baik. Model yang dihasilkan berhasil menangkap pola-pola utama dalam data klinis, khususnya melalui variabel penting seperti tekanan darah sistolik, kadar gula darah, usia ibu, suhu tubuh, serta detak jantung. Atribut-atribut ini terbukti menjadi penentu paling dominan dalam pembentukan pohon keputusan, sehingga menghasilkan struktur klasifikasi yang jelas dan mudah dipahami oleh tenaga kesehatan.

Secara keseluruhan, model prediksi memperlihatkan akurasi yang tinggi (sesuai hasil abang), yang menunjukkan bahwa algoritma C4.5 berhasil memetakan data ke dalam tiga kategori risiko—low risk, mid risk, dan high risk—dengan tingkat ketepatan yang konsisten. Hasil ini menunjukkan bahwa pola klinis yang signifikan berhasil dipelajari model, misalnya tekanan darah sistolik yang menjadi pemisah utama kategori risiko tinggi, serta kombinasi kadar gula darah dan suhu tubuh yang menjadi penentu risiko menengah maupun rendah.

Meskipun performa model secara umum sangat baik, terdapat beberapa aspek yang perlu diperhatikan. Proporsi kelas yang tidak seimbang (class imbalance) berpotensi memengaruhi sensitivitas model terhadap kelas tertentu, terutama pada kategori risiko menengah yang cenderung lebih sulit dibedakan. Hal ini terlihat dari kecenderungan model yang lebih akurat dalam mengidentifikasi risiko tinggi, namun sedikit kurang sensitif terhadap variasi kasus pada kelas menengah. Kondisi ini wajar terjadi pada model pohon keputusan dan dapat diperbaiki dengan teknik seperti oversampling, penyesuaian threshold, atau penerapan metode boosting.

Temuan ini selaras dengan karakteristik algoritma C4.5 yang dikenal mudah diinterpretasikan namun sensitif terhadap distribusi kelas. Namun demikian, hasil yang diperoleh tetap memenuhi tujuan penelitian, yaitu membangun model prediksi risiko maternal yang akurat, interpretable, dan dapat digunakan sebagai alat bantu pengambilan keputusan di bidang kesehatan maternal.

Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan bahwa algoritma pohon keputusan tetap menjadi metode yang andal dalam klasifikasi data klinis. Namun perbedaan kompleksitas variabel dan jumlah fitur menyebabkan variasi hasil performa. Dengan demikian, penelitian ini menguatkan bahwa algoritma C4.5 merupakan pendekatan yang relevan dan efektif dalam memprediksi risiko kesehatan maternal secara otomatis berbasis data klinis.

SIMPULAN

Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu memberikan performa yang sangat baik dalam memprediksi tingkat risiko kesehatan maternal, dengan akurasi tinggi dan pola klasifikasi yang konsisten berdasarkan variabel klinis seperti tekanan darah, kadar gula darah, usia, suhu tubuh, dan detak jantung. Pola keputusan yang dihasilkan model memperlihatkan bahwa tekanan darah sistolik dan kadar gula darah menjadi indikator utama dalam pemisahan kelas risiko, diikuti variabel pendukung lainnya yang menentukan kategori low, mid, atau high risk. Temuan ini menegaskan bahwa algoritma C4.5 tidak hanya efektif dalam memetakan hubungan antar-atribut klinis, tetapi juga mampu menghasilkan model yang mudah diinterpretasikan bagi tenaga kesehatan. Meski demikian, penelitian ini membuka peluang pengembangan lebih lanjut, khususnya pada aspek peningkatan sensitivitas model terhadap variasi data tertentu agar prediksi risiko maternal dapat dilakukan dengan lebih komprehensif.

UCAPAN TERIMA KASIH

Ucapan terima kasih dapat diperuntukkan kepada pihak-pihak yang telah membantu dalam penelitian, instansi yang menjadi objek penelitian, bisa juga kepada pihak yang membantu dalam publikasi artikel.

DAFTAR PUSTAKA

- A. Pinandito, S. A. Wicaksono, and S. H. Wijoyo, "Implementasi Machine Learning dalam Deteksi Risiko Tinggi Diabetes Melitus pada Kehamilan," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 739–746, Aug. 2023, doi: 10.25126/jtiik.2023107005.
- A. Ranjbar, F. Montazeri, S. R. Ghamsari, V. Mehrnoush, N. Roozbeh, and F. Darsareh, "Machine learning models for predicting preeclampsia: a systematic review," *BMC Pregnancy Childbirth*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12884-023-06220-1.
- F. Ramadhani, D. Septiana, S. N. Amalia, P. M. Fadilah, and A. Satria, "KLASIFIKASI RISIKO GIZI BURUK PADA IBU HAMIL MENGGUNAKAN METODE RANDOM FOREST," *Jurnal Teknologi Informasi*, vol. 5, no. 2, 2024, doi: 10.46576/djtechno.
- H. Sufriyana, Y. W. Wu, and E. C. Y. Su, "Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia," *EBioMedicine*, vol. 54, Apr. 2020, doi: 10.1016/j.ebiom.2020.102710.
- M. G. Febrian, R. Ferdiansyah, E. A. Nugraha, D. Satriatama, and R. Kusumastuti, "PREDIKSI RISIKO DIABETES MENGGUNAKAN ALGORITMA DECISION TREE DENGAN APLIKASI RAPID MINER," 2024.
- R. A. Fitra et al., "Analisis Perbandingan Algoritma Machine Learning untuk Klasifikasi Tingkat Risiko Ibu Hamil," *Student Research Journal*, vol. 1, no. 6, 2023, doi: 10.55606/srjyappi.v1i6.846.
- R. A. Ningsih, T. H. Silawati, S. Z. Zulfa, C. Maharani, and E. Nurjannah, "IDENTIFIKASI PREDIKTOR RISIKO PREEKLAMPSIA PADA MATERNAL UNTUK MODEL PREDIKSI BERBASIS MACHINE LEARNING," 2025, [Online]. Available: <http://journal.universitaspahlawan.ac.id/index.php/ners>
- S. Nurrohmah, D. Normawati, A. Dahlan, J. Ringroad Selatan, and I. Yogyakarta, "Prediksi Dini Penyakit Preeklamsia Menggunakan Algoritma C4.5," vol. 10, no. 3, pp. 120–132, 2022, doi: 10.12928/jstie.v8i3.xxx.
- T. Octari, P. Syesar, F. Febriansyah, and E. Efan, "PREDIKSI METODE PERSALINAN MENGGUNAKAN DECISION SUPPORT SYSTEM DAN ALGORITMA DECISION TREE C4.5," 2024.
- X. Pi, J. Wang, L. Chu, G. Zhang, and W. Zhang, "Prediction of high-risk pregnancy based on machine learning algorithms," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-00450-3.