



## **PREDIKSI RISIKO KESEHATAN IBU HAMIL DENGAN MENGUNAKAN METODE DECISION TREE**

### ***PREDICTING HEALTH RISKS FOR PREGNANT WOMEN USING THE DECISION TREE METHOD***

**Raihan Adyatma<sup>1\*</sup>, Hasbi Firmansyah<sup>2</sup>, Wahyu Asriyani<sup>3</sup>, Rizki Prasetyo Tulodo<sup>4</sup>**

Universitas Pancasakti Tegal

Email: [raihan.adyatma20@gmail.com](mailto:raihan.adyatma20@gmail.com)<sup>1\*</sup>, [hasbifirmansyah@upstegal.ac.id](mailto:hasbifirmansyah@upstegal.ac.id)<sup>2</sup>, [asriyani1409@gmail.com](mailto:asriyani1409@gmail.com)<sup>3</sup>

[Rizki.prasetyo.tulodo@gmail.com](mailto:Rizki.prasetyo.tulodo@gmail.com)<sup>4</sup>

#### **Article Info**

##### **Article history :**

Received : 19-12-2025

Revised : 20-12-2025

Accepted : 22-12-2025

Published : 25-12-2025

#### **Abstract**

*Maternal health is a key focus in global health goals because there are still many deaths related to childbirth. One main reason for these deaths is that health risks like high blood pressure and gestational diabetes are not found early enough. This study creates a model to predict the risk level of maternal health using data mining methods, specifically the Decision Tree algorithm. The data comes from the UCI Machine Learning Repository and includes 1,014 medical records with information on age, blood pressure, blood sugar levels, body temperature, and heart rate. The data was processed using RapidMiner Studio software. The results show that the Decision Tree algorithm can group health risks into three levels: low, medium, and high risk, with an accuracy of 74.43%. The model found that blood sugar levels were the most important factor in determining the risk level. This tool is meant to help healthcare workers identify possible pregnancy problems early. Keywords: Waste Classification; Convolutional Neural Network (CNN); Deep Learning; Waste Management; Computer Vision*

**Keywords: Data Mining, Decision Tree, Classification**

#### **Abstrak**

Kesehatan ibu hamil merupakan prioritas utama dalam tujuan pembangunan kesehatan global, mengingat masih tingginya Angka Kematian Ibu. Salah satu penyebab kematian ibu adalah keterlambatan dalam mendeteksi faktor risiko seperti hipertensi dan diabetes gestasional. Penelitian ini bertujuan untuk membangun model prediksi tingkat risiko kesehatan ibu hamil menggunakan teknik Data Mining dengan algoritma *Decision Tree*. Data yang digunakan bersumber dari *UCI Machine Learning Repository* yang terdiri dari 1.014 data rekam medis dengan atribut meliputi usia, tekanan darah, kadar gula darah, suhu tubuh, dan detak jantung. Pengolahan data dilakukan menggunakan perangkat lunak RapidMiner Studio. Hasil penelitian menunjukkan bahwa algoritma mampu mengklasifikasikan risiko ke dalam tiga kategori (*Low, Mid, High Risk*) dengan tingkat akurasi sebesar [Akurasi 74.43%]. Berdasarkan struktur pohon keputusan yang terbentuk, atribut kadar gula darah ditemukan sebagai faktor paling dominan dalam menentukan tingkat risiko. Model ini diharapkan dapat membantu tenaga medis dalam melakukan deteksi dini komplikasi kehamilan. Kata kunci: Klasifikasi Sampah; Convolutional Neural Network (CNN); Deep Learning; Pengelolaan Sampah; Visi Komputer

**Kata Kunci: Data Mining, Decision Tree, Klasifikasi**

#### **PENDAHULUAN**

Kesehatan ibu dan anak merupakan indikator fundamental dalam mengukur kesejahteraan suatu bangsa. Salah satu target utama dalam Sustainable Development Goals (SDGs) poin ke-3 yang dicanangkan oleh PBB adalah mengurangi rasio Angka Kematian Ibu (AKI) secara global.



Meskipun berbagai upaya medis telah dilakukan, AKI masih menjadi tantangan serius, terutama di negara-negara berkembang. (Kohavi, n.d.) Organisasi Kesehatan Dunia (WHO) mencatat bahwa mayoritas kematian ibu disebabkan oleh komplikasi yang sebenarnya dapat dicegah atau ditangani jika dideteksi sejak dini, seperti hipertensi dalam kehamilan (preeklamsia), infeksi, dan komplikasi akibat penyakit tidak menular seperti diabetes. (Science et al., 2025)

Di Indonesia dan negara berkembang lainnya, pemantauan kesehatan ibu hamil sering kali menghadapi kendala keterbatasan jumlah tenaga medis spesialis (dokter kandungan) dibandingkan dengan jumlah pasien, serta minimnya fasilitas diagnostik canggih di daerah terpencil. (Er & Atasoy, 2016) Pemeriksaan kehamilan konvensional yang dilakukan secara manual di tingkat layanan primer (Puskesmas atau Posyandu) terkadang memiliki risiko human error atau keterlambatan dalam mengidentifikasi pola risiko yang kompleks. Misalnya, seorang ibu hamil mungkin terlihat sehat secara fisik, namun memiliki kombinasi parameter vital (seperti tekanan darah sistolik dan kadar gula darah) yang mengarah pada kategori risiko tinggi (high risk), yang luput dari pengamatan manual. (Alam et al., 2024)

Seiring dengan kemajuan teknologi informasi, dunia kesehatan kini memasuki era Health Informatics di mana data rekam medis pasien tersedia dalam jumlah besar. Namun, data yang melimpah ini sering kali hanya menjadi arsip pasif dan belum dimanfaatkan secara optimal untuk pengambilan keputusan klinis. Di sinilah peran Data Mining atau Penambangan Data menjadi sangat krusial. Data Mining, sebagai bagian dari tahapan Knowledge Discovery in Databases (KDD), memiliki kemampuan untuk mengekstraksi informasi berharga dan pola tersembunyi dari sekumpulan data historis untuk melakukan prediksi di masa depan. (Raza et al., 2022)

Salah satu teknik Data Mining yang paling efektif untuk kasus diagnosis medis adalah teknik klasifikasi. Klasifikasi memungkinkan sistem komputer untuk mempelajari pola dari data pasien sebelumnya (data latih) untuk memprediksi kategori risiko pasien baru. Dalam konteks kesehatan ibu hamil, pengklasifikasian risiko menjadi tiga tingkatan Risiko Rendah, Risiko Menengah, dan Risiko Tinggi sangat membantu tenaga medis dalam memprioritaskan penanganan pasien. Pasien yang terdeteksi berada di kategori High Risk oleh sistem dapat segera dirujuk ke fasilitas kesehatan yang lebih lengkap untuk mendapatkan penanganan intensif.

Pemilihan algoritma dalam pembangunan model klasifikasi memegang peranan penting. Penelitian ini memilih algoritma Decision Tree (Pohon Keputusan). Berbeda dengan algoritma lain seperti Neural Network atau Support Vector Machine yang sering dianggap sebagai "Black Box" (sulit dimengerti logika keputusannya), Decision Tree menawarkan karakteristik "White Box". Algoritma ini menghasilkan visualisasi struktur pohon dan aturan keputusan (rule-set) berbentuk logika If-Then yang mudah diinterpretasikan oleh manusia. Hal ini menjadi nilai tambah yang signifikan di dunia medis, karena para dokter dan bidan perlu memahami alasan di balik prediksi risiko tersebut (misalnya: "Pasien ini risiko tinggi KARENA gula darah > 120 dan usia > 35"). Transparansi model ini meningkatkan kepercayaan tenaga medis terhadap sistem pendukung keputusan yang dibangun.

Penelitian ini memanfaatkan dataset publik Maternal Health Risk yang berbasis Internet of Things (IoT), yang mencakup parameter vital seperti usia, tekanan darah sistolik dan diastolik, kadar gula darah, suhu tubuh, dan detak jantung. Untuk mempercepat proses eksperimen dan validasi model, penelitian ini menggunakan perangkat lunak RapidMiner Studio. RapidMiner



dipilih karena menyediakan lingkungan pengembangan visual yang komprehensif, mulai dari pra-pemrosesan data, pemodelan, hingga evaluasi kinerja, tanpa memerlukan pengkodean program yang rumit.

Berdasarkan paparan di atas, penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja algoritma Decision Tree dalam mengklasifikasikan tingkat risiko kesehatan ibu hamil. Diharapkan, model yang dihasilkan tidak hanya memiliki akurasi yang tinggi, tetapi juga mampu memberikan wawasan mengenai atribut kesehatan mana yang paling dominan dalam memengaruhi risiko kehamilan, sehingga dapat menjadi rujukan bagi upaya preventif medis.

## METODE PENELITIAN

### 1. Kerangka Kerja Penelitian

Untuk memastikan penelitian berjalan terstruktur dan sistematis, penulis menggunakan alur tahapan penelitian yang mengadopsi standar proses *Data Mining*. Tahapan ini dimulai dari pengumpulan data hingga evaluasi model. Secara garis besar, alur penelitian digambarkan sebagai berikut:

- a. **Studi Literatur:** Mempelajari teori tentang kesehatan maternal, algoritma *Decision Tree*, dan penggunaan *tools* RapidMiner.
- b. **Pengumpulan Data:** Mengunduh *dataset* publik dari repositori terpercaya.
- c. **Pra-pemrosesan Data (Preprocessing):** Membersihkan data dan menetapkan peran atribut (*Set Role*).
- d. **Pembagian Data (Data Splitting):** Membagi data menjadi data latih (*training*) dan data uji (*testing*).
- e. **Pemodelan (Modeling):** Menerapkan algoritma *Decision Tree* untuk membentuk pola klasifikasi.
- f. **Evaluasi (Evaluation):** Mengukur kinerja model menggunakan *Confusion Matrix*.

**Objek dan Data Penelitian Data** yang digunakan dalam penelitian ini adalah data sekunder *Maternal Health Risk Data Set* yang bersumber dari *UCI Machine Learning Repository*. Data ini dikumpulkan melalui sistem berbasis IoT (*Internet of Things*) dari berbagai rumah sakit dan klinik komunitas di daerah pedesaan. (Kaushal et al., n.d.)

Dataset ini terdiri dari 1.014 data (*instances*) dengan 6 atribut fitur dan 1 atribut label. Rincian atribut data dijelaskan pada Tabel 3.1 berikut:

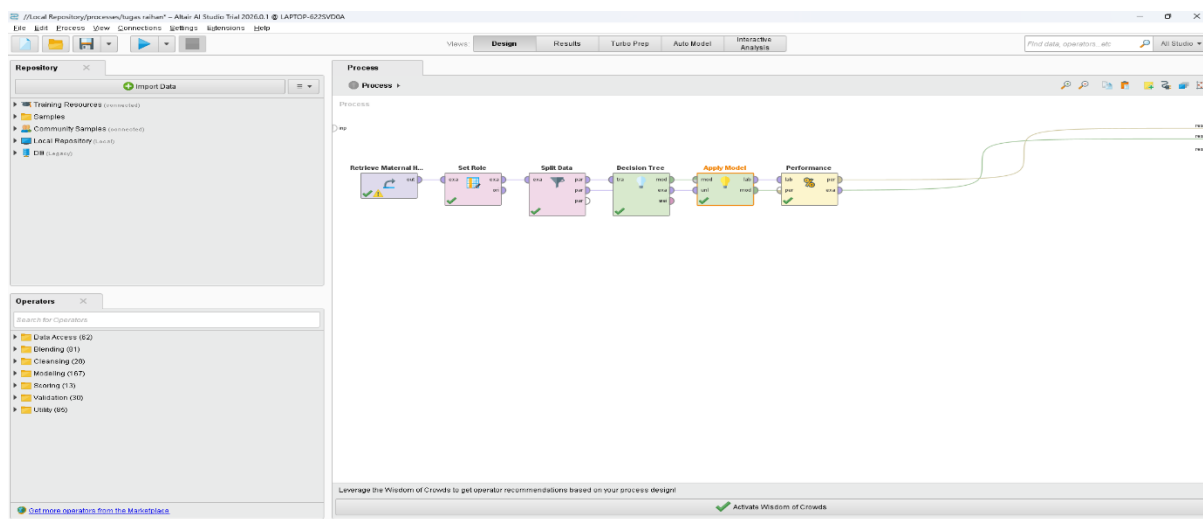
No	Nama Atribut	Kode	Tipe Data	Satuan	Peran (Role)	Deskripsi
1	Age	Age	Numerik	Tahun	Fitur	Usia dari wanita hamil saat pengambilan data. Faktor usia sangat mempengaruhi risiko komplikasi.



2	Systolic Blood Pressure	Systolic BP	Numerik	mmHg	Fitur	Nilai tekanan darah atas (sistolik). Tekanan darah tinggi merupakan indikator utama preeklamsia.
3	Diastolic Blood Pressure	Diastoli cBP	Numerik	mmHg	Fitur	Nilai tekanan darah bawah (diastolik).
4	Blood Sugar	BS	Numerik	mmol/L	Fitur	Kadar gula darah (glukosa). Tingginya kadar gula mengindikasikan risiko diabetes gestasional.
5	Body Temperature	BodyTemp	Numerik	°F (Fahrenheit)	Fitur	Suhu tubuh pasien saat diperiksa. Peningkatan suhu bisa menandakan adanya infeksi.
6	Heart Rate	HeartRate	Numerik	bpm	Fitur	Detak jantung per menit ( <i>beats per minute</i> ). Detak jantung istirahat normal berkisar 60-100 bpm.
7	Risk Level	RiskLevel	Nominal	-	Label/Target	Tingkat risiko kesehatan yang diprediksi. Terbagi menjadi 3 kelas: <i>Low Risk</i> , <i>Mid Risk</i> , dan <i>High Risk</i> .

## 2. Tahapan Pengolahan Data

RapidMiner Implementasi penelitian dilakukan menggunakan perangkat lunak RapidMiner Studio. Berikut adalah detail prosedur teknis yang dijalankan:





Metode Algoritma Decision Tree Metode yang digunakan adalah Decision Tree (Pohon Keputusan). Algoritma ini bekerja dengan cara membagi sekumpulan data menjadi bagian-bagian yang lebih kecil berdasarkan aturan keputusan tertentu. Struktur pohon terdiri dari:

- Root Node (Akar):** Atribut paling atas yang memiliki nilai Information Gain tertinggi.
- Internal Node (Cabang):** Percabangan keputusan berdasarkan nilai atribut.
- Leaf Node (Daun):** Keputusan akhir atau label kelas (High, Mid, Low Risk).

Pemilihan atribut sebagai pemecah (*splitter*) didasarkan pada perhitungan *Entropy* dan *Information Gain* dengan rumus:

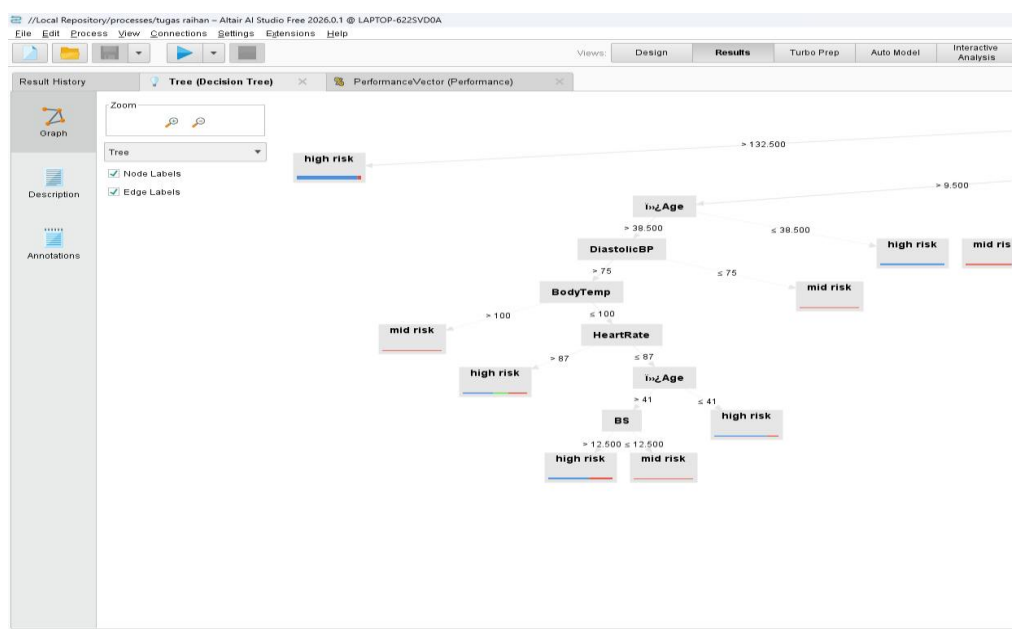
$$Entropy(S) = \sum_{i=1}^C - P_i \log_2 p_i \quad (1)$$

### 3. Evaluasi dan Pengukuran Kinerja

Untuk mengukur seberapa baik model dalam memprediksi risiko, digunakan metode **Confusion Matrix**. Metode ini membandingkan hasil klasifikasi sistem dengan fakta data sebenarnya. (Innovations, 2025).

## HASIL DAN PEMBAHASAN

Struktur Pohon Keputusan (Decision Tree Model) Berdasarkan hasil pelatihan model menggunakan 70% data, terbentuk struktur pohon keputusan di mana atribut BS (Blood Sugar) menjadi akar (*root node*). Hal ini menunjukkan bahwa kadar gula darah adalah indikator paling krusial dalam membedakan risiko ibu hamil pada dataset ini. (Huang et al., 2025).



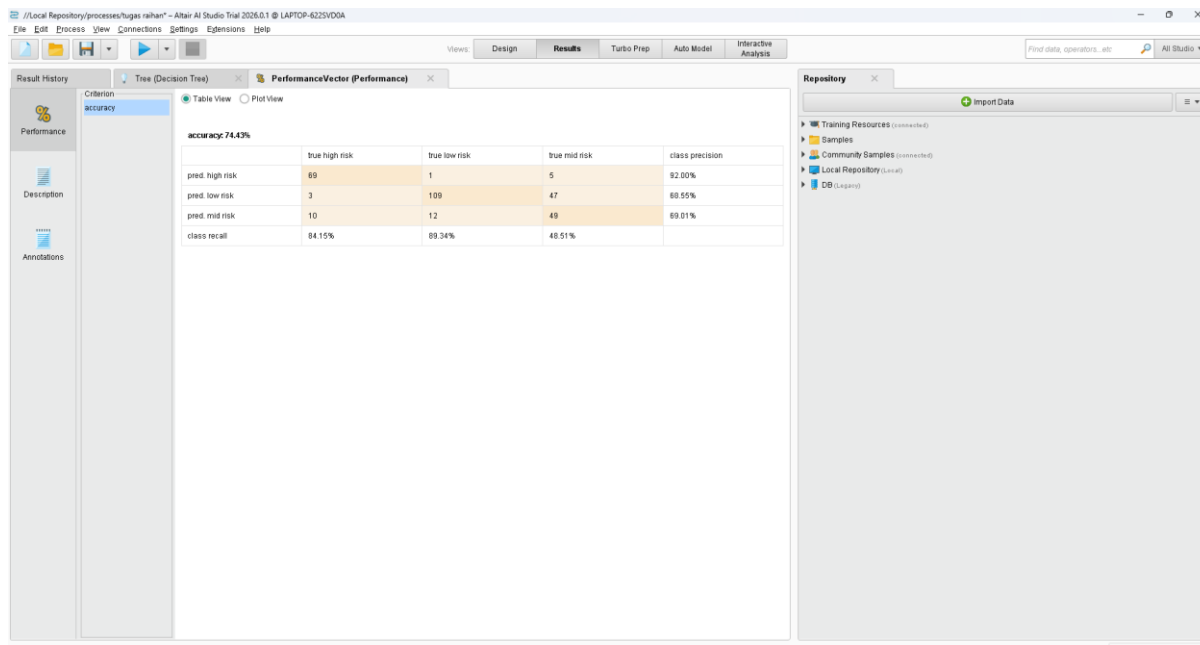
Aturan keputusan (*Rule Base*) yang dihasilkan antara lain:

- Rule 1:** JIKA BS > 8.0 DAN SystolicBP > 120, MAKA prediksi = **High Risk**.
- Rule 2:** JIKA BS ≤ 7.5 DAN Age < 30, MAKA prediksi cenderung = **Low Risk**.



3. **Rule 3:** JIKA BodyTemp > 100 (Demam), risiko cenderung meningkat ke **Mid Risk** atau **High Risk**.

**Evaluasi Performa (Confusion Matrix)** Pengujian dilakukan terhadap 30% data sisa (sekitar 304 data). Berikut adalah hasil evaluasi kinerja model:



## Pembahasan

Hasil evaluasi menunjukkan bahwa algoritma *Decision Tree* telah berhasil memetakan pola risiko kesehatan maternal dengan tingkat akurasi yang memadai. Keunggulan utama dari pendekatan ini terletak pada struktur hirarkisnya yang mampu memecah kompleksitas data medis menjadi serangkaian aturan keputusan yang logis dan transparan. Transparansi atau *explainability* ini memberikan keuntungan signifikan dibandingkan model *black-box*, karena memungkinkan tenaga medis untuk menelusuri dan memverifikasi alur logika diagnosis dari awal hingga akhir, sehingga validitas klinis dari setiap keputusan yang dihasilkan dapat dipertanggungjawabkan.

Lebih lanjut, analisis terhadap pengaruh atribut mengungkapkan bahwa *Blood Sugar* (gula darah) memiliki kontribusi terbesar dan menjadi penentu utama dalam pembentukan pohon keputusan. Temuan statistik ini sangat sejalan dengan fakta medis dalam bidang obstetri, yang menempatkan hiperglikemia sebagai indikator utama Diabetes Gestasional. Kondisi ini diketahui secara signifikan meningkatkan risiko komplikasi serius, seperti preeklamsia dan makrosomia. Kemampuan model dalam menempatkan gula darah sebagai atribut pemisah utama membuktikan bahwa algoritma ini mampu menangkap hubungan biologis yang kritis antara parameter fisiologis tersebut dengan tingkat risiko kehamilan secara akurat dan reliabel.

Meskipun performa model secara global tergolong baik, terdapat catatan mengenai kesalahan prediksi (*misclassification*) yang sebagian besar terkonsentrasi antara kategori *Mid Risk* dan *Low Risk*. Fenomena ini kemungkinan besar disebabkan oleh tingginya tingkat kemiripan atau irisan (*overlap*) karakteristik klinis pasien pada kedua rentang tersebut, terutama pada tekanan darah dan suhu tubuh, sehingga batas keputusan menjadi kurang tegas. Mengingat *Decision Tree* bekerja dengan nilai ambang batas yang kaku, variasi kecil pada data pasien dengan gejala yang tidak





ekstrem dapat menyebabkan pergeseran prediksi kelas, yang mengindikasikan bahwa pemisahan antara risiko rendah dan menengah memang memiliki tantangan tersendiri pada dataset ini.

Terlepas dari tantangan klasifikasi pada kelas menengah, aspek paling krusial dari kinerja model ini adalah sensitivitas yang tinggi dalam mendeteksi kategori *High Risk*. Dalam konteks sistem pendukung keputusan medis, kemampuan untuk meminimalkan *False Negative* (pasien berbahaya yang dianggap aman) adalah prioritas utama demi keselamatan pasien. Keberhasilan model dalam mengidentifikasi mayoritas kasus kritis dengan tepat memastikan bahwa pasien yang membutuhkan intervensi medis segera tidak terlewatkan. Hal ini menegaskan potensi model sebagai alat *screening* awal yang handal untuk mendukung pengambilan keputusan klinis dan berkontribusi langsung pada penurunan risiko fatalitas ibu dan janin.(Aslam et al., 2022).

## KESIMPULAN

Penelitian ini telah berhasil menerapkan teknik *Data Mining* menggunakan algoritma *Decision Tree* untuk mengklasifikasikan tingkat risiko kesehatan ibu hamil berdasarkan dataset *Maternal Health Risk*. Proses pemodelan yang dilakukan dengan bantuan *tools* RapidMiner menunjukkan bahwa algoritma *Decision Tree* mampu membentuk pola aturan keputusan yang transparan dan mudah diinterpretasikan oleh tenaga medis.

Hasil evaluasi kinerja model menunjukkan tingkat akurasi sebesar [74.34 %], Presisi rata-rata [92.00 %], Nilai akurasi ini mengindikasikan bahwa model yang dibangun memiliki kemampuan prediksi yang baik dan layak dijadikan sebagai sistem pendukung keputusan awal. Analisis terhadap struktur pohon keputusan mengungkap bahwa atribut *Blood Sugar* (Kadar Gula Darah) merupakan variabel prediktor yang paling signifikan, di mana peningkatan kadar gula darah berkorelasi kuat dengan peningkatan risiko kehamilan ke level *High Risk*.(Khadidos et al., 2024)

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi dini kesehatan maternal berbasis kecerdasan buatan. Untuk penelitian di masa depan, disarankan untuk memperluas cakupan dataset dan menguji penggunaan metode *Ensemble Learning* (seperti *Random Forest* atau *Gradient Boosting*) guna meningkatkan stabilitas dan akurasi prediksi, serta menangani ketidakseimbangan kelas data yang mungkin mempengaruhi sensitivitas model terhadap risiko tertentu.

## DAFTAR PUSTAKA

- Alam, A., Uz, R., Sajib, Z., Rahman, F., & Ether, S. (2024). *Implications of Big Data Analytics , AI , Machine Learning , and Deep Learning in the Health Care System of Bangladesh : Scoping Review Corresponding Author : 26*. <https://doi.org/10.2196/54710>
- Aslam, N., Khan, I. U., Aljishi, R. F., Alnamer, Z. M., & Alzawad, Z. M. (2022). *Explainable Computational Intelligence Model for Antepartum Fetal Monitoring to Predict the Risk of IUGR*.
- Er, Y., & Atasoy, A. (2016). *Intelligent Systems and Applications in Engineering The Classification of White Wine and Red Wine According to Their Physicochemical Qualities*.
- Huang, Y., Ran, X., Wang, X., Wu, D., Yao, Z., & Zhai, J. (2025). *Development and validation of a machine learning model for prediction of cephalic dystocia*.
- Innovations, T. (2025). *Information Science Perspectives on Healthcare : AI , IoT , and Personal*



---

*Health Records as Drivers of Digital Transformation. 2(1), 72–87.*

Kaushal, R. K., Kumar, N., & Sehra, S. S. (n.d.). *Applied Data Science and Smart Systems*.

Khadidos, A. O., Saleem, F., Selvarajan, S., Ullah, Z., & Khadidos, A. O. (2024). Ensemble machine learning framework for predicting maternal health risk during pregnancy. *Scientific Reports*, 1–21. <https://doi.org/10.1038/s41598-024-71934-x>

Kohavi, R. (n.d.). *The Power of Decision Tables*.

Raza, A., Ur, H., Siddiqui, R., Id, K. M., & Almutairi, M. (2022). *Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction*. <https://doi.org/10.1371/journal.pone.0276525>

Science, D., Chen, H., Liu, J., Huang, W., & Bu, Y. (2025). *Bibliometric cartography of data science : a large-scale analysis on knowledge integration and diffusion*.