

## Riset Data Risiko Kehamilan Menggunakan Decision Tree Dari Dataset Risk Maternal Health

**Widya Rahmatika Purwanti**

Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal,  
Indonesia

Email: [widyarahmatikap@gmail.com](mailto:widyarahmatikap@gmail.com)

### **Intisari**

Kehamilan adalah tahap penting dalam hidup seorang wanita yang membutuhkan perhatian khusus untuk memastikan keselamatan dan kesehatan ibu serta janin. Penelitian ini bertujuan untuk menganalisis potensi risiko kesehatan pada ibu hamil dengan menerapkan metode *Decision Tree* pada dataset Risiko Kesehatan *Maternal*. Keadaan kesehatan yang buruk pada ibu hamil dapat berakibat fatal, termasuk kematian ibu dan anak. Oleh karena itu, diperlukan pendekatan yang sistematis untuk memprediksi dan mengelola risiko tersebut. Metode *Decision Tree* dipilih karena kemampuannya dalam menghasilkan model yang jelas dan mudah dipahami untuk klasifikasi data kesehatan. Dalam penelitian ini, data kesehatan ibu hamil dianalisis dengan menggunakan algoritma *Decision Tree*, yang menghasilkan akurasi mencapai 100%. Temuan ini mengindikasikan bahwa model pohon keputusan yang diperoleh dapat dijadikan sebagai alat bantu dalam proses pengambilan keputusan terkait klasifikasi risiko kesehatan pada ibu hamil. Dengan demikian, penelitian ini memberikan kontribusi signifikan terhadap upaya pencegahan kondisi kesehatan yang tidak diinginkan selama kehamilan.

**Kata kunci:** Dataset, *Decision Tree*, Kesehatan Ibu Hamil, Risiko Kehamilan.

### **Abstract**

*Pregnancy is a crucial stage in a woman's life that requires special attention to ensure the safety and health of both mother and fetus. This study aims to analyze potential health risks in pregnant women by applying the Decision Tree method to the Maternal Health Risk dataset. Poor health in pregnant women can have fatal consequences, including maternal and child mortality. Therefore, a systematic approach is needed to predict and manage these risks. The Decision Tree method was chosen because of its ability to produce clear and easy-to-understand models for health data classification. In this study, maternal health data was analyzed using the Decision Tree algorithm, which achieved 100% accuracy. These findings indicate that the resulting decision tree model can be used as a tool in the decision-making process related to health risk classification in pregnant women. Thus, this study makes a significant contribution to efforts to prevent undesirable health conditions during pregnancy.*

**Keywords:** Dataset, Decision Tree, Maternal Health, Pregnancy Risk.

## PENDAHULUAN

Kehamilan merupakan tahap yang sangat penting dalam kehidupan seorang wanita, yang membawa tantangan dan risiko kesehatan yang signifikan. Berdasarkan data dari *World Health Organization* (WHO), setiap tahunnya sekitar 295.000 ibu meninggal akibat komplikasi yang berkaitan dengan kehamilan dan persalinan, dengan sebagian besar kejadian tersebut terjadi di negara-negara berkembang[9]. Risiko kesehatan selama kehamilan dapat dipengaruhi oleh berbagai faktor, termasuk kondisi medis yang sudah ada sebelumnya, usia ibu, tekanan darah, kadar gula darah, dan faktor yang lainnya[6]. Karena itu, pengelolaan risiko ini sangat krusial untuk menjamin kesehatan ibu dan bayi.

Dalam era *big data* dan kemajuan teknologi informasi, analisis data kesehatan menggunakan metode statistik dan algoritma pembelajaran mesin semakin umum digunakan untuk mengidentifikasi pola dan memprediksi risiko. Salah satu pendekatan yang efektif dalam klasifikasi data adalah *Decision Tree*. Metode ini menyediakan representasi *visual* yang mudah dipahami serta kemampuan untuk menangani data yang kompleks. *Decision Tree* berfungsi dengan membagi dataset menjadi *subset* berdasarkan atribut tertentu, yang mempermudah dalam mengidentifikasi faktor-faktor yang mempengaruhi risiko kesehatan[5].

Penelitian ini bertujuan untuk menerapkan metode *Decision Tree* pada dataset risiko kesehatan maternal untuk menganalisis dan memprediksi potensi risiko kesehatan pada ibu hamil. Dataset ini mencakup berbagai variabel yang relevan dengan kesehatan *maternal*, seperti usia ibu, tekanan darah sistolik dan diastolik, kadar gula darah, denyut jantung, suhu tubuh dan tingkat resiko. Dengan penerapan algoritma *Decision Tree*, diharapkan dapat diperoleh model yang tidak hanya tepat dalam memprediksi risiko, tetapi juga mudah dipahami oleh tenaga medis dan pihak-pihak terkait lainnya.

Melalui penelitian ini, diharapkan dapat memberikan pemahaman baru tentang faktor-faktor risiko kesehatan selama kehamilan serta memberikan kontribusi dalam upaya pencegahan komplikasi yang dapat membahayakan ibu dan bayi. Tujuannya agar mengurangi permasalahan atau dampak negatif yang timbul dari adanya suatu penyebab yang tidak diinginkan. Oleh karena itu, hasil penelitian ini diharapkan dapat menjadi landasan dalam pengambilan keputusan klinis dan penyusunan kebijakan kesehatan yang lebih efektif dalam konteks perawatan maternal.

## METODE PENELITIAN

Kerangka Kerja pada penelitian ini meliputi tahap proses dan langkah-langkah yang runtut untuk memproses dan menyelesaikan permasalahan yang ada.

### 1. Pengumpulan Data

Sekumpulan data atau yang biasa disebut dengan istilah dataset dapat dicari di berbagai *website* penyedia dataset yang banyak beredar di internet. Contoh *website* yang biasa digunakan untuk bahan mencari dataset diantaranya

seperti Kaggle, UCI Machine Learning Repository dan lain-lain. Mencari dataset bisa ditemukan dalam web yang gratis di internet bahkan bisa tersedia di web yang berbayar.

Dalam penelitian ini, data yang digunakan diperoleh dari dataset "*Maternal Health Risk*" yang tersedia di UCI Machine Learning Repository. Dataset ini terdiri dari 1015 record yang mencakup berbagai atribut kesehatan ibu hamil, seperti usia, tekanan darah sistolik dan diastolik, kadar gula darah, suhu tubuh, serta denyut jantung. Untuk memudahkan pemahaman dan pembahasan pada penelitian ini, nama-nama atribut asli dari dataset *Maternal Health Risk* akan diterjemahkan ke dalam Bahasa Indonesia sebagai berikut:

Tabel 1. Terjemahan Atribut Dataset Asli ke Bahasa Indonesia

Nama Atribut (Dataset Asli)	Nama Atribut (Diterjemahkan)	Keterangan / Satuan
<i>Age</i>	Usia	Tahun
<i>Systolicbp</i>	Tekanan Darah Sistolik	MmHg
<i>Diastolicbp</i>	Tekanan Darah Diastolik	MmHg
<i>Bs</i>	Kadar Gula Darah	Mmol/L
<i>Bodytemp</i>	Suhu Tubuh	Fahrenheit
<i>HearRate</i>	Detak Jantung	Bpm ( <i>Beats Per Minute</i> )
<i>RiskLevel</i>	Tingkat Risiko	Target Label ( <i>High, Mid, Low</i> )

Proses pengumpulan data dilakukan dengan memastikan bahwa semua atribut relevan dan dapat digunakan untuk analisis risiko kehamilan. Dataset yang digunakan untuk membuat jurnal ini dapat diunduh dari tautan berikut ini: <https://archive.ics.uci.edu/dataset/863/maternal+health+risk>

Tabel 2. Kumpulan Data Kesehatan Ibu Hamil

ID	Usia	Tekanan darah sistolik	Tekanan darah diastolik	Gula darah	Suhu tubuh	Denyut jantung	Tingkat risiko
1	25	130	80	15	98	86	Risiko tinggi
2	35	140	90	13	98	70	Risiko tinggi

ID	Usia	Tekanan darah sistolik	Tekanan darah diastolik	Gula darah	Suhu tubuh	Denyut jantung	Tingkat risiko
3	29	90	70	8	100	80	Risiko tinggi
4	30	140	85	7	98	70	Risiko tinggi
5	35	120	60	6,1	98	76	Risiko rendah
6	23	140	80	7,01	98	70	Risiko tinggi
7	23	130	70	7,01	98	78	Risiko sedang
8	35	85	60	11	102	86	Risiko tinggi
9	32	120	90	6,9	98	70	Risiko sedang
10	42	130	80	18	98	70	Risiko tinggi
...	...	...	...	...	...	...	...

Sumber : Dokumentasi Pribadi

Keterangan Kolom:

- a. ID: Identifikasi unik untuk setiap responden.
- b. Usia (Tahun): Usia ibu hamil saat kehamilan.
- c. Tekanan darah sistolik: nilai tekanan darah maksimal dengan satuan mmHg.
- d. Tekanan darah diastolik: nilai tekanan darah yang lebih rendah dengan satuan mmHg.
- e. Gula darah: kadar glukosa dalam darah.
- f. Suhu tubuh: suhu tubuh pada ibu hamil.
- g. Denyut jantung: detak jantung pada keadaan istirahat normal.
- h. Tingkat risiko: tingkat risiko yang diprediksi dengan mempertimbangkan atribut sebelumnya.

Tabel ini merupakan representasi dari dataset yang telah diproses untuk analisis lebih lanjut menggunakan metode *Decision Tree*. Data ini diharapkan dapat memberikan pemahaman tentang faktor-faktor risiko yang berdampak pada kesehatan ibu hamil.

## 2. *Decision Tree*

Metode analisis yang diterapkan dalam penelitian ini adalah *Decision Tree*, yang merupakan salah satu algoritma pembelajaran mesin yang efisien untuk klasifikasi. *Decision Tree* berfungsi dengan membagi dataset menjadi *subset* berdasarkan atribut-atribut yang paling relevan, sehingga menghasilkan model yang mudah dipahami dan diinterpretasikan.

Dalam konteks penelitian ini, *Decision Tree* digunakan untuk mengidentifikasi faktor-faktor risiko utama yang berkontribusi terhadap

kesehatan ibu hamil dan memprediksi kemungkinan komplikasi berdasarkan data yang dikumpulkan.

Perbandingan dengan Metode Lain: Dalam beberapa penelitian, *Decision Tree* dibandingkan dengan algoritma lain seperti *K-Nearest Neighbors* (KNN) dan *Naïve Bayes*. Hasilnya menunjukkan bahwa *Decision Tree* sering kali memberikan akurasi yang lebih tinggi[3].

### 3. *Penggunaan Aplikasi Rapidminer*

Analisis data dilakukan menggunakan aplikasi RapidMiner, sebuah platform analisis data yang menyediakan berbagai alat untuk pemrosesan dan analisis data. RapidMiner memungkinkan pengguna untuk membangun model *Decision Tree* secara visual, memudahkan proses pemilihan atribut, serta memberikan kemudahan dalam evaluasi model melalui metrik akurasi dan validasi silang[9].

## HASIL DAN PEMBAHASAN

### 1. Skema Pembagian Dataset

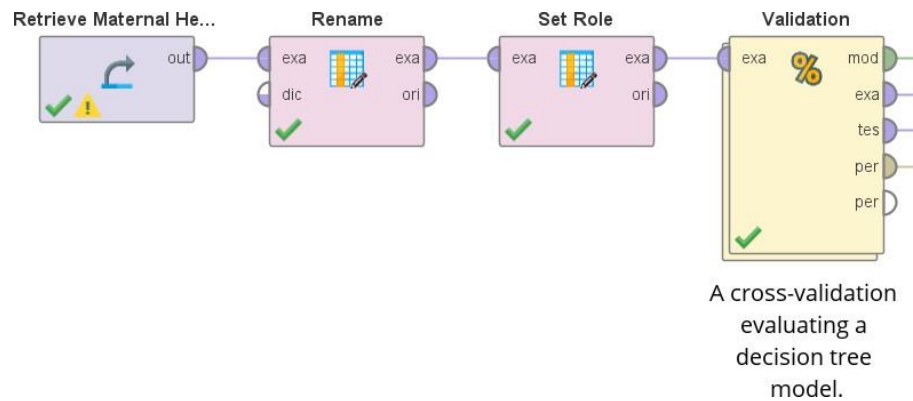
Sebelum melakukan tahap pengujian, dataset harus diolah agar mudah untuk diprediksi. Pengolahan data ini menghasilkan label dan pembagian data. Label berfungsi sebagai target yang ingin divalidasi dan akan diuji nantinya. Dataset yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Pembagian ini bertujuan untuk memastikan evaluasi yang baik terhadap model yang dibangun. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian berfungsi untuk menilai kinerja model setelah proses pelatihan[1]. Dalam penelitian ini, pembagian dataset menggunakan *10-fold cross validation* untuk menghindari *data leakage* dan menjamin reliabilitas, yang dimana dataset akan dibagi menjadi 10 bagian yang sama besar.

### 2. Pengujian Dataset

Pada bagian ini dataset siap untuk dilatih dan diuji menggunakan *cross validation*. *Cross validation* berisi dua kolom yang berguna untuk melatih dan menguji *dataset*. Operator *Decision Tree* akan berada di kolom pelatihan sedangkan operator *apply models* dan *performance* berada di kolom *testing*. Data yang masuk ke dalam operator *Decision Tree* yang merupakan sembilan lipatan data yang akan digunakan untuk melatih model pohon keputusan. Output dari pelatihan tersebut akan menghasilkan sebuah model statik atau prediktif yang merepresentasikan struktur dari pohon keputusan.

Pada tahap pengujian, model tersebut akan menyisakan satu lipatan terakhir yang akan diuji pada *apply model*, lipatan tersebut merupakan lipatan yang tidak dilatih pada operator *decision tree*. Ketika dataset diuji, operator *performance* akan menghitung berapa tebakan yang benar. Setelah sudah mencapai putaran terakhir, rapidMiner akan mencari rata-rata dari kesepuluh ujian tersebut.

Dengan demikian, penggunaan *Cross Validation* dalam *RapidMiner* sangat penting untuk mengevaluasi dan meningkatkan keandalan model prediksi dalam konteks penelitian risiko kesehatan ibu hamil. Proses ini memastikan bahwa hasil yang diperoleh adalah *representatif* dan dapat diandalkan untuk pengambilan keputusan klinis.



Sumber : Dokumen Pribadi  
Gambar 1. Process Dataset pada Rapidminer

### 3. Performance Vector

Di dalam *Performance vector* terdapat setidaknya empat kriteria pada hasil *performance* yaitu *accuracy*, *precision*, *recall*, dan *AUC*. Pengujian performa model dilakukan menggunakan metode *10-Fold Cross Validation*, metode tersebut dipilih untuk menghasilkan validitas yang terjamin tanpa data bias pemilihan data latih serta mengatasi isu *overfitting*.

accuracy: 69.72% +/- 4.52% (micro average: 69.72%)

	true high risk	true low risk	true mid risk	class precision
pred. high risk	231	7	36	84.31%
pred. low risk	11	347	171	65.60%
pred. mid risk	30	52	129	61.14%
class recall	84.93%	85.47%	38.39%	

Sumber : Dokumen Pribadi  
Gambar 2. Accuracy  
Berdasarkan Confusion Matrix diatas, dapat dianalisis sebagai berikut:



atribut tekanan darah sistolik, temuan tersebut telah mengkonfirmasi bahwa hipertensi adalah faktor risiko terbesar.

Salah satu keunggulan dari *Decision Tree* adalah kemampuannya untuk memberikan interpretasi yang jelas dan visualisasi yang mudah dipahami [7]. Struktur pohon yang dihasilkan mempermudah untuk memahami aturan keputusan tanpa memerlukan pemahaman teknis yang kompleks. Sebagai contoh, jika tekanan darah lebih dari 132,5 maka berisiko tinggi.

Meskipun akurasi global yang dihasilkan moderat, penerapan metode ini berhasil mengidentifikasi pola kasus berisiko tinggi dengan sensitivitas yang baik. Menurut Jauhari et.al (2025), Peneliti dapat mengoptimalkan pengolahan data kesehatan ibu hamil dan menghasilkan model prediktif yang lebih akurat[8].

#### 5. Pola Aturan Yang Dihasilkan

Berdasarkan struktur pohon keputusan yang terbentuk pada gambar ketiga, model menghasilkan aturan Keputusan yang dapat digunakan sebagai pendeteksi dini. Berikut adalah aturan utama yang dihasilkan oleh model:

##### a. Pola deteksi risiko tinggi (hipertensi)

Aturan ini menempatkan tekanan darah sebagai tolak ukur jika tekanan darah sistolik lebih dari 132,5 maka risikonya tinggi. Ibu hamil dengan tekanan darah sistolik diatas 132,5 mmHg diklasifikasikan sebagai ibu hamil dengan risiko tertinggi tanpa perlu melihat parameter lain. Hal ini selaras dengan pengetahuan medis yang menyatakan bahwa hipertensi adalah penyebab utama pada komplikasi maternal[2].

##### b. Pola deteksi risiko metabolik (gula darah)

Jika tekanan darah ibu hamil berada diambang batas normal, maka model akan memeriksa kadar gula darah sebagai penentu risikonya. Aturannya jika tekanan darah sistolik dibawah 132,5 mmHg dan gula darah lenih dari 9 mmol/L, maka ibu hamil tersebut diklasifikasikan sebagai risiko sedang atau bahkan tinggi. Hal itu dapat terjadi karena pengujian belum melibatkan parameter lain seperti suhu tubuh.

##### c. Pola kondisi aman (risiko rendah)

Kelas risiko rendah teridentifikasi pada pohon keputusan yang lebih dalam, hal tersebut mengharuskan ibu hamil lolos dari berbagai pemeriksaan parameter lain untuk dinyatakan sebagai ibu hamil dengan risiko rendah. Jika tekanan darah sistolik kurang dari 132,5 mmHG, gula darah kurang dari 7,9 mmol/L dan suhu tubuh kurang dari 99,5°F maka ibu hamil dinyatakan sebagai risiko rendah.

## KESIMPULAN

Berdasarkan hasil dari pengujian dan analisa yang telah dilakukan, algoritma *decision tree* pada dataset risiko kesehatan ibu hamil (*maternal health risk*) menghasilkan akurasi model sebesar 69,72% . meskipun akurasi tergolong moderat, model mengindikasikan efektivitas yang tinggi dalam mendeteksi kasus risiko tinggi (*high risk*) dengan persentase *recall* sebesar 84,93%. Hal ini menunjukkan bahwa metode tersebut dapat berguna sebagai alat bantu skrining awal ibu hamil. Berdasarkan struktur pohon keputusan, atribut tekanan darah sistolik teridentifikasi sebagai faktor utama dalam penentuan risiko. Menurut struktur pohon keputusan, ibu hamil dengan tekanan darah diastolik diatas 132,5 mmHg terindikasi ke dalam kategori risiko tinggi. Selain itu, kadar gula darah juga menjadi faktor yang dominan apabila tekanan darah ibu hamil normal.

Kelemahan utama pada penelitian ini yaitu pada kemampuan klasifikasi kategori ibu hamil dengan risiko sedang. Risiko tersebut sering terjadi *misclassification* ke kategori risiko rendah karena karakteristik yang hampir sama pada kedua kategori tersebut.

## SARAN

Untuk penelitian selanjutnya, disarankan untuk menerapkan metode lain misalnya *random forest* atau *gradient boosting*, ataupun melakukan penyeimbangan pada dataset (*resampling*) guna meningkatkan akurasi pada kategori risiko rendah maupun risiko sedang,

## DAFTAR PUSTAKA

- [1] Aditya, M., et al. (2024). Optimasi Klasifikasi Decision Tree Dengan Teknik Pruning untuk Mengurangi Overfitting pada Dataset Penyakit Jantung. *Jurnal Sistem Informasi (JSiI)*, 11(2), 87-96.
- [2] Astuti, E. R., Claudia, A J. G. (2024). Tinjauan Literatur: Penatalaksanaan Hipertensi Pada Ibu Hamil. *Journal Of Health Science And Research* , 6(2), 186-200.
- [3] Marwan, R. H., & Evicienna, E. (2023). Pkm Pada Aplikasi Desain Labeling Untuk Kemasan Makanan Olahan Lokal Jenis Ikan Peda Ukm Barakuda Di Pulau Payung, Kepulauan Seribu Dki Jakarta. *Jurnal Strategi Desain dan Inovasi Sosial*, 4(2), 127. DOI: [10.37312/jsdis.v4i2.6937](https://doi.org/10.37312/jsdis.v4i2.6937)
- [4] Amalia, H., et al. (2023). Prediksi Risiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree. *Jurnal Swabumi (Suara Wawasan Bernuansa Ilmiah)*, 11(1), 54-61.
- [5] Jauhari, M. T., Windarsyah, & Marleny, F. D. (2025). Penerapan Data Mining Algoritma Decision Tree Untuk Memprediksi Keterlambatan Pembayaran

- Piutang. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 13(3), 6828-6834. Universitas Lampung.
- [6] Kementerian Kesehatan RI. (2022). *Kesehatan Ibu dan Anak*. <https://ayosehat.kemkes.go.id/buku-kia-kesehatan-ibu-dan-anak>, diakses tanggal 12/4/2025.
- [7] Costa, V.G., Pedreira, C.E. Recent advances in decision trees: an updated survey. *Artif Intell Rev* 56, 4765–4800 (2023). <https://doi.org/10.1007/s10462-022-10275-5>.
- [8] Purnama, J. J., Hikmawati, N. K., & Rahayu, S. (2024). Analisis Algoritma Klasifikasi Untuk Mengidentifikasi Potensi Risiko Kesehatan Ibu Hamil. *Journal of Applied Computer Science and Technology (JACOST)*, 5(1), 120–127.
- [9] Wijaya, Y. A., Bahtiar, A., Kaslani, & R.N. (2021). Analisa Klasifikasi Menggunakan Algoritma Decision Tree pada Data Log Firewall. *Jurnal Sistem Informasi dan Manajemen*, 9(3), 256-264.
- [10] World Health Organization (WHO). (2022). *Maternal Health*. <https://www.who.int/health-topics/maternal-health>, diakses tanggal 12/4/2025.