



Implementasi Data Mining Untuk Memprediksi Pre-Eklampsia Dalam Kehamilan Menggunakan Algoritma C4.5

Dhita Aulia Octaviani^{1✉}, Dhias Widiastuti², Rizky Amelia³, Abu Salam⁴

¹Poltekkes Kemenkes Semarang, Indonesia

²Universitas Dian Nuswantoro, Indonesia

E-mail/HP: dhitaaulia@poltekkes-smg.ac.id / 081325061986

INFO ARTIKEL	ABSTRAK
Diterima: Des 2024 Disetujui: Feb 2025 Dipublikasi: Mei 2025 Keyword: pre-eklampsia, prediksi, algoritma C4.5, data mining	Pre-eklampsia (PE), menjadi penyebab utama kematian ibu yang kasusnya terus meningkat. Perkembangan teknologi dan informasi dapat dimanfaatkan dalam bidang kesehatan, diantaranya implementasi Data Mining algoritma C4.5 untuk membangun model klasifikasi dalam bentuk pohon keputusan, di mana setiap node internal mewakili atribut, setiap cabang mewakili nilai atribut, dan setiap daun mewakili kelas keputusan, sehingga dapat diimplementasikan untuk penentuan prediksi status pre-eklampsia melalui data rekam medis kehamilan dan persalinan. Penelitian ini mengidentifikasi faktor resiko PE menggunakan analisis data mining dengan algoritma C4.5, untuk menentukan faktor risiko yang paling berpengaruh terhadap kejadian PE dengan tahapan: seleksi data, pembersihan data, transformasi data, mencari pola data, dan evaluasi hasil. Algoritma C4.5 menunjukkan bahwa Riwayat PE mempunyai korelasi erat dengan kejadian PE dengan nilai 0,71, diikuti oleh Riwayat hipertensi ibu dan Riwayat hipertensi dalam keluarga. Hasil performa prediksi menggunakan C4.5 mencapai nilai akurasi 96,93%, sensitivitas 96,93%, presisi 96,84% dan F1-Score 96,88%. Dapat disimpulkan bahwa Ibu tanpa Riwayat PE, paritas >2 dan obesitas, maka ibu akan mengalami PE. Ibu dengan Riwayat hipertensi/PE dan terdapat Riwayat hipertensi/PE dalam keluarga, paritas >2, diabetes, dan jarak kelahiran < 2 tahun, maka ibu akan mengalami PE. C4.5 sudah memberikan hasil yang baik, namun perlu membandingkan kinerjanya dengan algoritma lain dalam data mining, seperti <i>Random Forest</i>, <i>Support Vector Machine (SVM)</i>, atau <i>K-Nearest Neighbors (KNN)</i> untuk mengetahui apakah ada algoritma lain yang lebih unggul dalam hal akurasi dan sensitivitas prediksi.

DOI: 10.32763/ps5qpj75

Implementation Of Data Mining To Predict Pre-Eclampsia In Pregnancy Using The C4.5 Algorithm

ABSTRACT

*Preeclampsia (PE) is a leading cause of maternal mortality, with cases continuing to rise. Advances in technology and information can be leveraged in the healthcare sector, including the implementation of **Data Mining** using the **C4.5 algorithm** to build a classification model in the form of a decision tree. In this model, each internal node represents an attribute, each branch represents attribute values, and each leaf represents a decision class. This approach can be applied to predict preeclampsia status through pregnancy and delivery medical records. This study identifies PE risk factors using data mining analysis with the C4.5 algorithm to determine the most influential risk factors contributing to PE occurrence. The process includes data selection, data cleaning, data transformation, pattern discovery, and result evaluation. The C4.5 algorithm shows that a **history of PE** has a strong correlation with PE occurrence, with a value of **0.71**, followed by a **history of maternal hypertension** and a **family history of hypertension**. The prediction performance using C4.5 achieved an **accuracy of 96.93%**, **sensitivity of 96.93%**, **precision of 96.84%**, and an **F1-score of 96.88%**. It can be concluded that mothers without a history of PE but with **parity >2** and **obesity** are likely to experience PE. Mothers with a **history of hypertension/PE**, a **family history of hypertension/PE**, **parity >2**, **diabetes**, and **birth spacing <2 years** are also likely to develop PE. While C4.5 has delivered good results, it is necessary to compare its performance with other data mining algorithms, such as **Random Forest**, **Support Vector Machine (SVM)**, or **K-Nearest Neighbors (KNN)**, to determine whether other algorithms outperform C4.5 in terms of prediction accuracy and sensitivity.*



Jurnal Kesehatan

Published by UPPM Poltekkes Kemenkes Ternate
p-ISSN 1907-6401 e-ISSN 2597-7520



Alamat korespondensi:

Poltekkes Kemenkes Semarang, Semarang – *Central Java* , Indonesia

Email: dhitaaulia@poltekkes-smg.ac.id

© 2025 Poltekkes Kemenkes Ternate

Pendahuluan

Angka Kematian Ibu (AKI) merupakan salah satu indikator untuk melihat keberhasilan upaya kesehatan ibu. AKI adalah rasio kematian ibu selama masa kehamilan, persalinan, dan nifas yang disebabkan oleh kehamilan, persalinan, dan nifas atau pengelolaannya tetapi bukan karena sebab-sebab lain seperti kecelakaan atau insidental di setiap 100.000 kelahiran hidup (Achadi, 2019). AKI di Indonesia berdasarkan Survey Angka Sensus tahun 2015 sebanyak 305 per 100.000 KH dan menjadikan Indonesia peringkat kedua sebagai negara dengan kematian ibu tertinggi di Asia Tenggara (Sri Utami et al., 2020).

Hipertensi dalam kehamilan, termasuk pre-eklampsia (PE) dan eklampsia, merupakan penyebab utama kematian ibu yang sebenarnya dapat dicegah dan merupakan salah satu komplikasi kehamilan yang kasusnya semakin meningkat (Lindayani, 2018). Pre-eklampsia berkontribusi pada tingginya angka kesakitan dan kematian, baik pada ibu maupun janin, termasuk dapat mengakibatkan gangguan pertumbuhan janin, resiko terjadinya dysplasia bronkopulmonalis, serebral palsy, persalinan premature, dan meningkatkan kejadian persalinan dengan Tindakan. Ibu hamil yang tidak tertangani dengan baik, dapat mengalami komplikasi berat seperti eklampsia, sindrom HELLP, gagal ginjal, bahkan kematian (Lindayani, 2018; Phipps et al., 2019; Salan, 2017).

Perkembangan teknologi dan informasi dapat dimanfaatkan dalam berbagai bidang, termasuk di dalamnya bidang kesehatan. Proses data mining merupakan salah satu proses penggalian data melalui perangkat lunak yang membantu pendukung keputusan penentuan status pre-eklampsia melalui data rekam medis kehamilan dan atau persalinan. Tahapan – tahapan dalam pengambilan dan pengolahan data meliputi: seleksi data, permbersihan data, transformasi data, mencari pola data, dan evaluasi hasil (M.Adams, 2015; Muzakir & Wulandari, 2016). Melalui proses data mining, diharapkan dapat ditemukan suatu potensi yang lebih dari sekedar data, namun juga dapat membantu menganalisis penentuan status pre-eklampsia pada seorang ibu hamil (Jhee et al., 2019a; Tahir et al., 2018).

Dalam upaya menurunkan angka kematian ibu (AKI), identifikasi faktor risiko yang dapat memicu komplikasi selama kehamilan, seperti preeklamsia, sangat penting. Salah satu pendekatan yang efektif adalah penerapan algoritma machine learning, khususnya algoritma C4.5, yang merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) dan dikembangkan oleh Ross Quinlan. Algoritma ini membangun model klasifikasi dalam bentuk pohon keputusan yang mudah dipahami dan diinterpretasikan (Nurrohmah et al., 2022)

Penelitian sebelumnya telah menunjukkan efektivitas algoritma C4.5 dalam memprediksi risiko kesehatan. Misalnya, penelitian oleh Nurrohmah dan Normawati (2022) menggunakan algoritma C4.5 untuk mendeteksi dini preeklamsia pada ibu hamil. Penelitian ini melibatkan 870 data pasien dengan atribut seperti pendidikan, pekerjaan, usia, usia kehamilan, tekanan darah, berat badan, jenis kehamilan, jumlah kelahiran, riwayat aborsi, riwayat persalinan, riwayat penyakit, dan proteinuria. Hasil pengujian menunjukkan bahwa sistem prediksi yang dibangun memiliki akurasi 81,38%, precision 78,37%, recall 79,69%, dan f1-score 78,73% (Nurrohmah et al., 2022)

Selain itu, penelitian oleh Ma'mur dan Maulana (2025) mengoptimalkan algoritma C4.5 menggunakan *Particle Swarm Optimization (PSO)* untuk meningkatkan akurasi prediksi risiko kesehatan kehamilan. Hasilnya menunjukkan bahwa model C4.5 yang dioptimalkan dengan PSO memiliki tingkat akurasi 71,65%, lebih tinggi dibandingkan model C4.5 standar dengan akurasi 67,49% (Ma'mur & Maulana, 2024)

Penggunaan algoritma C4.5 dalam penelitian ini penting karena kemampuannya menangani data yang tidak lengkap dan menghasilkan aturan klasifikasi yang jelas serta mudah dipahami. Selain itu, algoritma ini memiliki performa yang baik dalam hal akurasi dan efisiensi, sehingga cocok untuk analisis data kesehatan yang kompleks. Dengan demikian, penerapan algoritma C4.5 dalam memprediksi risiko preeklamsia diharapkan dapat memberikan kontribusi signifikan dalam upaya menurunkan AKI di Indonesia.



Metode

1. Desain Penelitian

Penelitian ini menggunakan desain case-control yang bertujuan untuk mengidentifikasi hubungan antara faktor risiko dengan kejadian pre-eklampsia pada ibu hamil. Penelitian ini dinyatakan layak etik sesuai Surat Keterangan Laik Etik yang diterbitkan oleh Komite Etik Penelitian Kesehatan Poltekkes Kemenkes Semarang No. 0665/EA/KEPK/2024.

2. Pengumpulan Data

Data diperoleh dari catatan medis ibu hamil dan bersalin yang mengalami pre-eklampsia/eklamsia dari dua rumah sakit di Kota Semarang. Pengumpulan data berlangsung dari Juni hingga Juli 2024 dengan jumlah total data sebanyak 1114 rekam medis. Terdapat sembilan faktor risiko awal (sebelum decoding), yaitu:

Tabel 1. Atribut/Faktor Resiko PE (sebelum dilakukan *decoding*)

No	Atribut/Faktor Resiko
1.	Usia
2.	Paritas
3.	Jarak Kelahiran
4.	Riwayat Hipertensi
5.	Riwayat PE
6.	Obesitas
7.	Riwayat DM
8.	Riwayat Hipertensi/PE dalam Keluarga
9.	Sosial ekonomi Rendah

Sumber: Data RM responden

3. Tahap Analisis Data

Analisis data dalam penelitian ini menggunakan algoritma C4.5 dengan tahapan berikut:

a. Pra Pemrosesan Data

Tahap ini dilakukan untuk mengubah data mentah (raw data) menjadi data siap olah yang dapat diterima algoritma *machine learning* (Azmi & Dahria, 2013; Jhee et al., 2019b; M.Adams, 2015). Data penelitian terdiri dari dua jenis, yaitu numerik dan kategorik. Data kategorik kemudian dikonversi menggunakan teknik *one hot encoding*, sehingga atribut berubah dari 9 atribut menjadi 16 atribut numerik sebagai berikut:

Tabel 2. Atribut/Faktor Resiko PE (setelah dilakukan *one hot encode*)

No.	Atribut/Faktor Resiko
1.	Usia_Reproduksi Sehat
2.	Usia_Reproduksi tidak Sehat
3.	Paritas
4.	Jarak Kelahiran
5.	Riwayat Hipertensi_Ya
6.	Riwayat Hipertensi_Tidak
7.	Riwayat PE_Ya
8.	Riwayat PE_Tidak
9.	Obesitas_Ya
10.	Obesitas_Tidak
11.	Riwayat DM_Ya
12.	Riwayat DM_Tidak
13.	Riwayat Hipertensi/PE dalam Keluarga_Ya
14.	Riwayat Hipertensi/PE dalam Keluarga_Tidak
15.	Sosial ekonomi < UMR
16.	Sosial ekonomi > UMR

Sumber: Data RM responden

b. Seleksi Fitur

Feature selection dilakukan menggunakan metode *Mutual Information* untuk mengukur relevansi tiap fitur terhadap kejadian pre-eklampsia (Manoochehri et al., 2021). Tujuan seleksi fitur antara lain adalah mereduksi dimensi data, meningkatkan akurasi dan efisiensi model,

mencegah overfitting, serta mempermudah interpretasi model (M.Adams, 2015; Mardi, 2017; Muzakir & Wulandari, 2016). Fitur seleksi memiliki beberapa fungsi, di antaranya:

- 1) Mengurangi dimensi data: membantu menghilangkan fitur-fitur yang tidak relevan atau kurang penting, sehingga model lebih sederhana dan lebih mudah diinterpretasikan.
- 2) Meningkatkan akurasi model: dengan menghilangkan fitur yang tidak relevan atau saling berkorelasi tinggi, model dapat lebih fokus pada fitur yang benar2 berkontribusi pada hasil prediksi, sehingga mengurangi resiko *overfitting* dan meningkatkan akurasi pada data baru (*generalization*).
- 3) Mengurangi waktu pemrosesan: jumlah data yang perlu diproses berkurang, maka waktu proses dan prediksi model juga menjadi lebih cepat.
- 4) Mengurangi *overfitting*: Jika data terlalu banyak dan tidak relevan, dapat menyebabkan model terlalu menyesuaikan diri pada data (*overfit*). Dengan fitur seleksi, model lebih fokus pada pola yang benar-benar signifikan.
- 5) Mempermudah interpretasi model: model dengan sedikit fitur lebih mudah dipahami dan diinterpretasikan
- 6) Meningkatkan efisiensi penyimpanan data

Pada penelitian ini, fitur seleksi mereduksi data-data yang sekiranya tidak diperlukan/tidak relevan dalam pemrosesan data (dengan nilai korelasi 0). Jika nilai korelasi mendekati angka 1, berarti faktor/atribut tersebut mempunyai hubungan/korelasi (Tahir et al., 2018). Metode analisis korelasi dalam penelitian ini dilakukan dengan menggunakan visualisasi heatmap, yang berfungsi untuk mengidentifikasi tingkat hubungan antar variabel prediktor terhadap variabel hasil (kejadian preeklampsia). Heatmap menyajikan nilai korelasi dalam bentuk visual berupa gradasi warna, di mana semakin gelap atau semakin intens warna yang muncul menunjukkan semakin kuat hubungan antar variabel tersebut. Sebaliknya, warna yang semakin terang atau redup menunjukkan hubungan yang lebih lemah atau rendah. Dengan heatmap, peneliti dapat secara cepat dan jelas menentukan atribut atau variabel mana yang memiliki hubungan paling signifikan dalam memprediksi kejadian preeklampsia, sehingga memudahkan dalam tahap seleksi fitur sebelum model klasifikasi dibangun. Hal ini layaknya penghitungan menggunakan *odd ratio* pada *case control*, yang menunjukkan seberapa besar suatu faktor berpengaruh terhadap suatu kejadian.

Tabel 3. Faktor resiko/atribut sebelum dilakukan reduksi

No.	Atribut/Faktor Resiko	Nilai Korelasi
1.	Riwayat PE_Tidak	0.122888
2.	Riwayat PE_Ya	0.109965
3.	Riwayata Hipertensi/PE dalam Keluarga_Ya	0.021182
4.	Riwayat Hipertensi_Tidak	0.017653
5.	Paritas	0.017246
6.	Riwayat Hipertensi_Ya	0.009369
7.	Riwayat Hipertensi/PE dalam Keluarga_Tidak	0.004114
8.	Obesitas_Ya	0.002168
9.	Riwayat DM_Ya	0.001669
10.	Riwayat DM_Tidak	0.001641
11.	Jarak Kelahiran	0.000280
12.	Usia_Reproduksi Sehat	0.000000
13.	Usia_Reproduksi tidak Sehat	0.000000
14.	Obesitas_Tidak	0.000000
15.	Sosial ekonomi < UMR	0.000000
16.	Sosial ekonomi > UMR	0.000000

Tabel 4. Faktor resiko/atribut sesudah dilakukan reduksi

No.	Atribut/Faktor Resiko	Nilai Korelasi
1.	Riwayat PE_Tidak	0.122888
2.	Riwayat PE_Ya	0.109965
3.	Riwayat Hipertensi/PE dalam Keluarga_Ya	0.021182
4.	Riwayat Hipertensi_Tidak	0.017653
5.	Paritas	0.017246

6.	Riwayat Hipertensi_Ya	0.009369
7.	Riwayat Hipertensi/PE dalam Keluarga_Tidak	0.004114
8.	Obesitas_Ya	0.002168
9.	Riwayat DM_Ya	0.001669
10.	Riwayat DM_Tidak	0.001641
11.	Jarak Kelahiran	0.000280

c. Membangun Pohon Keputusan dengan C4.5

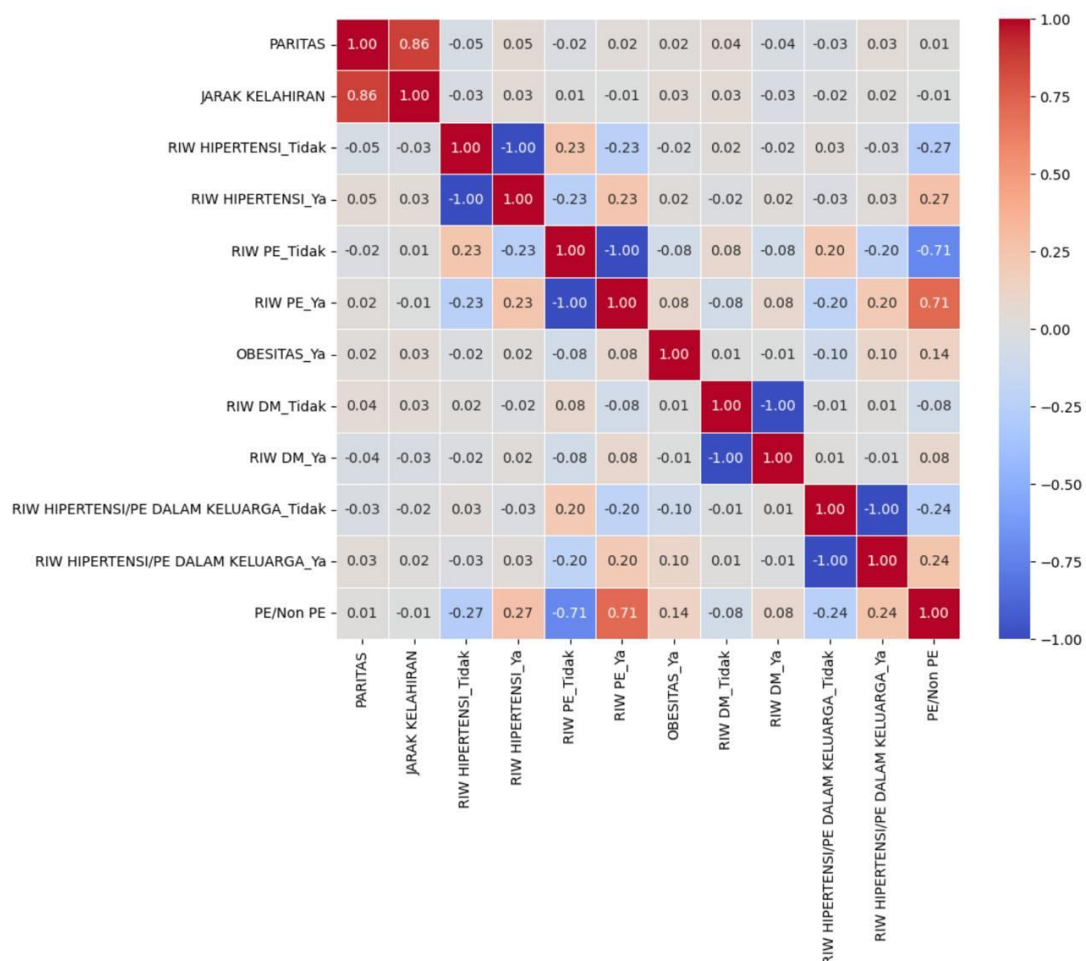
Algoritma C4.5 digunakan untuk membangun model prediktif berupa *decision tree* guna mengidentifikasi faktor risiko yang paling berpengaruh terhadap kejadian pre-eklampsia. *Decision tree* dihasilkan berdasarkan proses splitting atribut-atribut yang sudah terseleksi melalui *feature selection* sebelumnya.

d. Evaluasi Performa Model

Performa model yang dihasilkan dievaluasi menggunakan parameter-parameter akurasi, presisi, sensitivitas, dan f1-score untuk memastikan keakuratan prediksi faktor risiko pre-eklampsia.

Hasil dan Pembahasan

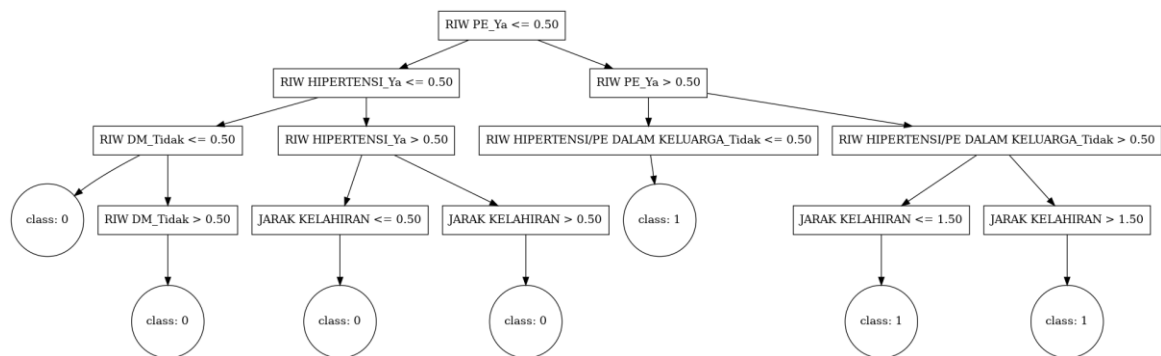
Setelah dilakukan tahap pre-processing dan feature selection, jumlah fitur atau atribut yang digunakan dalam analisis menggunakan algoritma C4.5 menjadi 11 atribut dari sebelumnya 16 atribut. Reduksi ini mempertahankan atribut-atribut yang relevan berdasarkan nilai korelasi Mutual Information, model pohon keputusan (*decision tree*) yang dihasilkan oleh algoritma C4.5 mengidentifikasi beberapa faktor risiko yang memiliki pengaruh signifikan terhadap kejadian pre-eklampsia, dengan faktor risiko Riwayat pre-eklampsia sebelumnya (Riwayat PE) dan Riwayat hipertensi dalam keluarga muncul sebagai faktor risiko paling dominan berdasarkan nilai korelasi yang tertinggi. Faktor-faktor tersebut menunjukkan potensi besar sebagai prediktor terjadinya pre-eklampsia.



Gambar 1. Visualisasi faktor predisposisi terjadinya PE

Semakin gelap warna yang ditampilkan, maka semakin erat hubungan antara faktor risiko dengan kejadian PE. Dari gambar 1 dapat dilihat bahwa riwayat PE mempunyai korelasi erat dengan kejadian PE dengan nilai 0,71. Diikuti oleh riwayat hipertensi dan riwayat hipertensi dalam keluarga.

Penelitian ini mengembangkan model klasifikasi menggunakan algoritma pohon keputusan C4.5 dengan metode seleksi fitur berbasis Mutual Information (MI). Model ini bertujuan untuk memprediksi risiko preeklampsia (PE) berdasarkan faktor risiko yang dimiliki ibu hamil.



Bagan 1. *Decision tree* faktor risiko pencetus PE

Berdasarkan *decision tree* yang terbentuk pada Bagan 1, terdapat dua cabang utama yang menentukan klasifikasi berdasarkan faktor risiko. Secara garis besar, pohon keputusan yang dihasilkan mengidentifikasi riwayat preeklampsia sebelumnya (Riwayat PE) sebagai faktor risiko paling dominan dalam menentukan klasifikasi:

1. Ibu tanpa Riwayat Preeklampsia ($PE \leq 0,50$)

Pada kelompok ini, faktor risiko yang menjadi penentu klasifikasi adalah riwayat hipertensi, diabetes mellitus (DM), serta jarak kelahiran:

- Ibu tanpa riwayat hipertensi dan tanpa riwayat diabetes mellitus diklasifikasikan sebagai tidak mengalami preeklampsia (class 0).
- Ibu dengan riwayat hipertensi tetapi dengan jarak kelahiran lebih dari 0,5 (lebih dari 2 tahun) tetap diklasifikasikan tidak mengalami preeklampsia (class 0).

Interpretasi Klinis:

Pada ibu yang tidak memiliki riwayat PE sebelumnya, kehadiran hipertensi dan diabetes menjadi faktor penting yang perlu diperhatikan. Namun, hasil model menunjukkan bahwa dalam kondisi tanpa riwayat PE sebelumnya, risiko preeklampsia tidak langsung tinggi, kecuali disertai faktor tambahan tertentu seperti paritas tinggi (>2) dan obesitas. Penelitian lain menunjukkan bahwa wanita dengan paritas tinggi (>2) mengalami risiko lebih tinggi mengalami preeklampsia karena penurunan fungsi sistem reproduksi akibat persalinan berulang, menyebabkan gangguan fungsi endotel dan perfusi plasenta (Ayu Lestari et al., 2024). Demikian pula dengan obesitas, yang secara signifikan meningkatkan risiko PE akibat beban metabolik dan kardiovaskular tambahan selama kehamilan (Persson et al., 2016).

2. Ibu dengan Riwayat Preeklampsia ($PE > 0,50$)

Pada kelompok ini, faktor penentu klasifikasi berikutnya adalah riwayat hipertensi atau PE dalam keluarga serta jarak kelahiran antar kehamilan:

- Ibu dengan riwayat PE dan memiliki riwayat hipertensi/PE dalam keluarga, dengan jarak kelahiran $\leq 1,5$ (kurang dari 2 tahun) diklasifikasikan mengalami preeklampsia (class 1).
- Ibu dengan riwayat PE dan memiliki riwayat hipertensi/PE dalam keluarga, dengan jarak kelahiran lebih dari 1,5 tahun juga diklasifikasikan mengalami preeklampsia (class 1).

Interpretasi Klinis:

Model mengindikasikan bahwa riwayat PE pada kehamilan sebelumnya merupakan faktor risiko utama. Risiko tersebut semakin meningkat jika terdapat riwayat hipertensi atau PE di keluarga, jarak kelahiran yang pendek (<2 tahun), dan faktor seperti paritas tinggi serta diabetes

mellitus. Hal ini didukung oleh penelitian yang menunjukkan bahwa riwayat hipertensi/PE pada keluarga meningkatkan risiko preeklampsia melalui predisposisi genetik yang memengaruhi regulasi tekanan darah dan fungsi endotel (Lee et al., 2022). Paritas tinggi dan diabetes mellitus juga meningkatkan risiko PE melalui gangguan fungsi vaskular, inflamasi kronis, serta stres oksidatif yang mengganggu plasentasi (Longhitano et al., 2022; Abraham & Romani, 2022). Jarak antar kehamilan yang pendek (<2 tahun) dapat mengurangi kemampuan tubuh ibu untuk memulihkan kondisi fisiologis sepenuhnya, meningkatkan risiko PE pada kehamilan berikutnya (Ni et al., 2023).

Evaluasi Performa Model

Performa model algoritma C4.5 yang dikembangkan dievaluasi dengan beberapa parameter, yakni akurasi, presisi, sensitivitas (recall), dan nilai F1-score pada data uji (test set):

1. **Akurasi:** 96,93%, menunjukkan bahwa model berhasil mengklasifikasikan dengan benar sebagian besar kasus risiko preeklampsia.
2. **Presisi:** 96,84%, menunjukkan bahwa prediksi positif preeklampsia sangat akurat, dengan angka false positive yang rendah.
3. **Sensitivitas (Recall):** 96,93%, mencerminkan kemampuan tinggi model dalam mendeteksi kasus preeklampsia secara benar (true positive).
4. **F1-Score:** 96,88%, menunjukkan keseimbangan yang optimal antara presisi dan sensitivitas, menjadikan model ini kuat dalam mengidentifikasi kasus preeklampsia secara efektif.

Performa yang tinggi ini menunjukkan bahwa algoritma C4.5 dengan seleksi fitur Mutual Information mampu memberikan hasil prediksi risiko preeklampsia yang andal, sehingga berpotensi digunakan sebagai alat deteksi dini risiko PE di Indonesia.

Kesimpulan Faktor Risiko Utama dari Model

Berdasarkan pohon keputusan (decision tree) yang terbentuk, beberapa faktor risiko yang paling dominan dalam klasifikasi risiko preeklampsia adalah:

1. **Riwayat preeklampsia pada kehamilan sebelumnya:** Merupakan faktor risiko dominan dengan dampak signifikan terhadap prediksi kejadian PE.
2. **Riwayat hipertensi/PE dalam keluarga:** Faktor ini memperkuat risiko preeklampsia, terutama pada ibu yang memiliki riwayat PE sebelumnya.
3. **Paritas tinggi (>2):** Berkontribusi signifikan terhadap peningkatan risiko preeklampsia, khususnya pada ibu tanpa riwayat PE sebelumnya.
4. **Obesitas:** Meningkatkan risiko preeklampsia pada ibu dengan paritas tinggi dan tanpa riwayat PE sebelumnya.
5. **Jarak kelahiran antar kehamilan (<2 tahun):** Berhubungan erat dengan peningkatan risiko PE pada ibu dengan riwayat PE sebelumnya.

Hasil ini menjadi dasar penting bagi tenaga kesehatan dalam upaya pencegahan dan deteksi dini preeklampsia, terutama bagi ibu hamil yang memiliki kombinasi faktor risiko tersebut.

Penutup

Berdasarkan hasil penelitian dan analisis data menggunakan algoritma C4.5 dengan metode seleksi fitur Mutual Information, dapat disimpulkan bahwa faktor risiko utama yang secara signifikan meningkatkan prediksi kejadian preeklampsia adalah riwayat preeklampsia sebelumnya, obesitas, paritas tinggi (>2), riwayat hipertensi atau preeklampsia dalam keluarga, diabetes mellitus, serta jarak antar kelahiran kurang dari dua tahun. Secara khusus, ibu tanpa riwayat preeklampsia sebelumnya tetap berisiko mengalami preeklampsia apabila memiliki kombinasi faktor paritas tinggi (>2) disertai obesitas. Sedangkan pada ibu dengan riwayat hipertensi atau preeklampsia sebelumnya, risiko preeklampsia meningkat secara signifikan bila disertai riwayat hipertensi atau preeklampsia dalam keluarga, paritas tinggi, diabetes mellitus, dan jarak antar kelahiran yang pendek (<2 tahun).

Model klasifikasi menggunakan decision tree algoritma C4.5 dalam penelitian ini berhasil mencapai tingkat akurasi sebesar 96,93%, dengan presisi 96,84%, sensitivitas 96,93%, dan F1-score sebesar 96,88%. Kinerja model yang tinggi tersebut menunjukkan bahwa model ini sangat efektif digunakan

untuk memprediksi risiko preeklampsia berdasarkan faktor risiko yang telah diidentifikasi. Oleh karena itu, hasil penelitian ini dapat dijadikan pedoman bagi tenaga kesehatan dalam melakukan deteksi dini serta pemantauan lebih intensif terhadap ibu hamil yang memiliki kombinasi faktor risiko tersebut, guna menurunkan risiko preeklampsia dan mendukung upaya pengurangan angka kematian ibu di Indonesia.

Daftar Pustaka

- Achadi, E. L. (2019). Kematian Maternal dan Neonatal di Indonesia. *Rakerkernas 2019*, 1–47.
- Ayu Lestari, D., Juniarty, E., Fitriyah, A., & Kebidanan Rangga Huasada, A. (2024). Hubungan Paritas Dan Obesitas Dengan Kejadian Preeklampsia Pada Ibu Hamil Di Rumah Sakit Umum Daerah Kota Prabumulih. *Jurnal Ilmu Kebidanan Dan Kesehatan*, 15(2), 88–94.
- Azmi, Z., & Dahria, M. (2013). Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan. *Saintikom*, 12, 157–164. http://demo.pohonkeputusan.com/files/Decision_Tree_Berbasis_Algoritma_Untuk_Pengambilan_Keputusan.pdf?i=1
- Jhee, J. H., Lee, S., Park, Y., Lee, S. E., Kim, Y. A., Kang, S. W., Kwon, J. Y., & Park, J. T. (2019a). Prediction model development of late-onset preeclampsia using machine learning-based methods. *PLoS ONE*, 14(8). <https://doi.org/10.1371/journal.pone.0221202>
- Jhee, J. H., Lee, S., Park, Y., Lee, S. E., Kim, Y. A., Kang, S. W., Kwon, J. Y., & Park, J. T. (2019b). Prediction model development of late-onset preeclampsia using machine learning-based methods. *PLoS ONE*, 14(8), 1–12. <https://doi.org/10.1371/journal.pone.0221202>
- Lee, K., Brayboy, L., & Tripathi, A. (2022). Pre-eclampsia: a Scoping Review of Risk Factors and Suggestions for Future Research Direction. In *Regenerative Engineering and Translational Medicine* (Vol. 8, Issue 3, pp. 394–406). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s40883-021-00243-w>
- Lindayani, I. K. (2018). Skrining pre eklampsia. *Jurnal Ilmiah Kebidanan*, 6(1), 1–6. <https://www.ejournal.poltekkes-denpasar.ac.id/index.php/JIK/article/download/1056/372>
- Longhitano, E., Siligato, R., Torreggiani, M., Attini, R., Masturzo, B., Casula, V., Matarazzo, I., Cabiddu, G., Santoro, D., Versino, E., & Piccoli, G. B. (2022). The Hypertensive Disorders of Pregnancy: A Focus on Definitions for Clinical Nephrologists. In *Journal of Clinical Medicine* (Vol. 11, Issue 12). MDPI. <https://doi.org/10.3390/jcm11123420>
- M.Adams, D. J. H. N. (2015). Data Mining Data mining. *Mining of Massive Datasets, January 2013*, 5–20. <https://doi.org/10.1002/9781118445112.stat06466.pub2>
- Ma'mur, K., & Maulana, A. E. (2024). Optimasi PSO untuk Meningkatkan Performa Algoritma C4.5 dalam Memprediksi Risiko Kesehatan Kehamilan. *Jurnal Informatika Universitas Pamulang*, 9(4), 178–186. <https://doi.org/10.32493/informatika.v9i4.46039>
- Manoochchhri, Z., Manoochchhri, S., Soltani, F., Tapak, L., & Sadeghifar, M. (2021). Predicting preeclampsia and related risk factors using data mining approaches: A cross-sectional study. *International Journal of Reproductive BioMedicine*, 19(11), 959–968. <https://doi.org/10.18502/ijrm.v19i11.9911>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>
- Ni, W., Gao, X., Su, X., Cai, J., Zhang, S., Zheng, L., Liu, J., Feng, Y., Chen, S., Ma, J., Cao, W., & Zeng, F. (2023). Birth spacing and risk of adverse pregnancy and birth outcomes: A systematic review and dose–response meta-analysis. In *Acta Obstetrica et Gynecologica Scandinavica* (Vol. 102, Issue 12, pp. 1618–1633). John Wiley and Sons Inc. <https://doi.org/10.1111/aogs.14648>

- Nurrohmah, S., Normawati, D., Dahlan, A., Ringroad Selatan, J., & Yogyakarta, I. (2022). *Prediksi Dini Penyakit Preeklamsia Menggunakan Algoritma C4.5*. 10(3), 120–132. <https://doi.org/10.12928/jstie.v8i3.xxx>
- Persson, M., Cnattingius, S., Wikström, A. K., & Johansson, S. (2016). Maternal overweight and obesity and risk of pre-eclampsia in women with type 1 diabetes or type 2 diabetes. *Diabetologia*, 59(10), 2099–2105. <https://doi.org/10.1007/s00125-016-4035-z>
- Phipps, E. A., Thadhani, R., Benzing, T., & Karumanchi, S. A. (2019). Pre-eclampsia: pathogenesis, novel diagnostics and therapies. *Nature Reviews Nephrology*, 15(5), 275–289. <https://doi.org/10.1038/s41581-019-0119-6>
- Salan, Y. D. C. (2017). Biomarker Terkini Dalam Usaha Memprediksi Preeklampsia. *Berkala Kedokteran*, 13(1), 119. <https://doi.org/10.20527/jbk.v13i1.3448>
- Sri Utami, B., Utami, T., Sekar Siwi, A., & Harapan Bangsa Purwokerto Jl Raden Patah No, U. (2020). Hubungan Riwayat Hipertensi Dan Status Gizi Dengan Kejadian Preeklamsia Pada Ibu Hamil : Literature Review Bakti Sri Utami, Tin Utami, Adiratna Sekar Siwi. *Jurnal Ilmu Keperawatan Maternitas*, 3(2). <https://doi.org/10.32584/jikm.v3i2.703>
- Tahir, M., Badriyah, T., & Syarif, I. (2018). Classification Algorithms of Maternal Risk Detection For Preeclampsia With Hypertension During Pregnancy Using Particle Swarm Optimization. *EMITTER International Journal of Engineering Technology*, 6(2).

