



Pendekatan Interpretatif dalam Prediksi Persalinan Caesar Menggunakan Decision Tree pada Data Pelayanan Kesehatan Primer

Arif Susilo*, Asep Arwan Sulaeman

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}arif.susilo@pelitabangsa.ac.id, ²asep.arwan@pelitabangsa.ac.id

Email Penulis Korespondensi: arif.susilo@pelitabangsa.ac.id

Abstrak—Persalinan Caesar merupakan prosedur medis yang dilakukan pada kondisi tertentu untuk mengurangi risiko komplikasi pada ibu dan bayi. Namun, peningkatan angka persalinan Caesar yang tidak selalu didasarkan pada indikasi medis menimbulkan kebutuhan untuk memahami faktor-faktor yang memengaruhi metode persalinan secara lebih sistematis. Penelitian ini bertujuan untuk mengeksplorasi pola hubungan antara variabel klinis ibu hamil dengan metode persalinan menggunakan pendekatan data mining berbasis algoritma Decision Tree dan Random Forest. Dataset yang digunakan merupakan data sekunder dari tiga puskesmas, yaitu Puskesmas Mranti, Banyuurip, dan Bayan, dengan total 390 data ibu hamil. Penelitian dilakukan menggunakan kerangka Knowledge Discovery in Database (KDD) yang meliputi tahapan seleksi data, preprocessing, transformasi, pembagian dataset, penanganan ketidakseimbangan data menggunakan Synthetic Minority Over-sampling Technique (SMOTE), pemodelan, dan evaluasi model. Hasil penelitian menunjukkan bahwa model memiliki akurasi sebesar 88%, precision 58,82%, recall 66,67%, dan F1-score 62,50%. Meskipun nilai akurasi tergolong cukup tinggi, performa model dalam mendeteksi persalinan Caesar masih berada pada tingkat moderat. Temuan ini menunjukkan bahwa model lebih efektif dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas. Penelitian ini menekankan bahwa pendekatan data mining pada data pelayanan kesehatan primer memiliki potensi dalam mengidentifikasi pola awal, namun hasil yang diperoleh belum cukup untuk digunakan sebagai dasar pengambilan keputusan klinis secara langsung. Penelitian lanjutan dengan dataset yang lebih besar dan metode yang lebih adaptif diperlukan untuk meningkatkan performa model.

Kata Kunci: Persalinan Caesar; Data Mining; Decision Tree; Random Forest; SMOTE; Klasifikasi

Abstract—Caesarean delivery is a medical procedure performed under specific conditions to reduce risks for both mother and baby. However, the increasing rate of caesarean deliveries, which is not always based on medical indications, highlights the need to systematically understand the factors influencing delivery methods. This study aims to explore the relationship between clinical variables of pregnant women and delivery methods using a data mining approach based on Decision Tree and Random Forest algorithms. The dataset consists of secondary data collected from three primary healthcare centers (Puskesmas), namely Mranti, Banyuurip, and Bayan, with a total of 390 records. The study follows the Knowledge Discovery in Database (KDD) framework, including data selection, preprocessing, transformation, dataset splitting, handling class imbalance using Synthetic Minority Over-sampling Technique (SMOTE), modeling, and evaluation. The results show that the model achieved an accuracy of 88%, precision of 58.82%, recall of 66.67%, and an F1-score of 62.50%. Although the accuracy appears relatively high, the model's performance in identifying caesarean cases remains moderate. This indicates that the model is more effective in classifying the majority class than the minority class. This study highlights that data mining applied to primary healthcare data can provide valuable insights for early pattern identification. However, the obtained results are not sufficient for direct clinical decision-making. Future research with larger datasets and more adaptive methods is required to improve model performance.

Keywords: Caesarean Delivery; Data Mining; Decision Tree; Random Forest; SMOTE; Classification

1. PENDAHULUAN

Persalinan merupakan fase akhir dari proses kehamilan yang secara fisiologis dapat berlangsung secara normal melalui pervaginam [1]. Namun dalam praktik klinis, tidak semua proses persalinan dapat berjalan normal karena adanya kondisi medis tertentu seperti preeklamsia [2], ketuban pecah dini [3], kelainan letak janin [4], gangguan tekanan darah [5], anemia, maupun infeksi yang berpotensi membahayakan keselamatan ibu dan bayi. Dalam kondisi tersebut, tindakan Sectio Caesarea (SC) menjadi alternatif prosedur medis yang dilakukan untuk mengurangi risiko komplikasi [6]. Dalam beberapa dekade terakhir, angka persalinan Caesar menunjukkan tren peningkatan yang signifikan secara global. World Health Organization (WHO) merekomendasikan angka ideal tindakan Caesar berada pada kisaran 5–15% dari total persalinan [7]. Namun pada kenyataannya, angka tersebut telah melampaui batas yang direkomendasikan di berbagai negara, termasuk di Indonesia [8]. Peningkatan ini tidak hanya dipengaruhi oleh faktor medis, tetapi juga faktor non-medis seperti preferensi pasien, kemudahan penjadwalan, serta sistem pembiayaan kesehatan.

Di sisi lain, tindakan Caesar sebagai prosedur pembedahan memiliki risiko yang lebih tinggi dibandingkan persalinan normal, antara lain infeksi, perdarahan, komplikasi anestesi, serta masa pemulihan yang lebih lama [9]. Selain itu, peningkatan jumlah tindakan Caesar juga berdampak pada meningkatnya beban pembiayaan layanan kesehatan [10]. Oleh karena itu, diperlukan pendekatan yang mampu membantu memahami pola dan faktor yang memengaruhi metode persalinan secara lebih sistematis sebagai dasar untuk analisis awal [11]. Perkembangan teknologi informasi, khususnya data mining, memberikan peluang dalam mengolah data kesehatan menjadi informasi yang bermakna. Data pemeriksaan rutin ibu hamil yang tersimpan dalam rekam medis pada dasarnya merupakan data mentah yang tidak memiliki nilai strategis apabila tidak dianalisis [12]. Data seperti umur, tekanan darah, kadar hemoglobin, protein urine, dan parameter lainnya hanya berupa kumpulan angka yang berdiri sendiri. Melalui proses analitik, data tersebut dapat ditransformasikan menjadi informasi yang mampu menggambarkan pola hubungan antar variabel [13].



Salah satu metode klasifikasi yang banyak digunakan dalam data mining adalah Decision Tree, yang memiliki keunggulan dalam menghasilkan model yang bersifat interpretatif dan mudah dipahami [14] Namun, metode ini memiliki keterbatasan dalam hal performa dibandingkan metode ensemble seperti Random Forest yang mampu meningkatkan akurasi melalui kombinasi beberapa pohon keputusan [15]. Selain itu, permasalahan ketidakseimbangan data (class imbalance) sering terjadi dalam data kesehatan dan dapat memengaruhi performa model klasifikasi apabila tidak ditangani dengan baik [16]. Sebagian besar penelitian sebelumnya menggunakan dataset rumah sakit dengan variabel yang kompleks dan cenderung berfokus pada peningkatan akurasi model. Sementara itu, pemanfaatan data pemeriksaan rutin di fasilitas pelayanan kesehatan tingkat pertama seperti puskesmas masih relatif terbatas, padahal data tersebut tersedia secara sistematis dan mencerminkan kondisi nyata di lapangan [17]. Selain itu, penelitian yang menggabungkan pendekatan interpretatif dengan teknik peningkatan performa seperti metode ensemble dan penanganan ketidakseimbangan data masih belum banyak dilakukan dalam konteks data puskesmas [18], [19].

Penelitian ini menggunakan dataset yang diperoleh dari tiga puskesmas, yaitu Puskesmas Mranti, Puskesmas Banyuurip, dan Puskesmas Bayan, dengan total 390 data ibu hamil. Penggunaan data multi-puskesmas bertujuan untuk meningkatkan variasi data serta memberikan gambaran yang lebih representatif terhadap kondisi pelayanan kesehatan tingkat pertama. Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma Decision Tree dan Random Forest dalam memodelkan metode persalinan berdasarkan data pemeriksaan rutin ibu hamil, serta membandingkan kinerja kedua algoritma dengan mempertimbangkan penanganan ketidakseimbangan data menggunakan SMOTE [20], [21]. Fokus utama penelitian ini bukan pada pencapaian akurasi semata, melainkan pada pemahaman pola hubungan antar variabel klinis serta evaluasi performa model dalam konteks data pelayanan kesehatan primer [22]. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan pendekatan analitik berbasis data mining pada bidang kesehatan ibu dan anak, khususnya dalam pemanfaatan data pemeriksaan rutin di puskesmas sebagai sumber informasi untuk analisis awal. Selain itu, hasil penelitian ini dapat menjadi dasar bagi pengembangan model yang lebih robust dan memiliki performa yang lebih tinggi pada penelitian selanjutnya.

2. METODOLOGI PENELITIAN

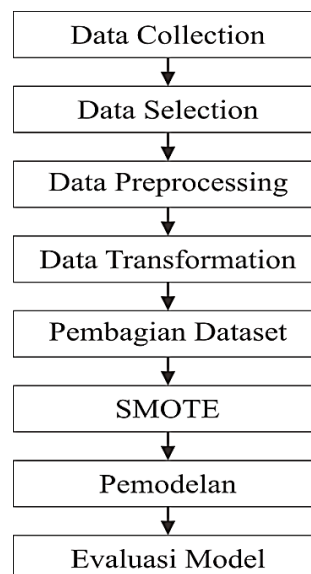
3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk membangun model klasifikasi metode persalinan berdasarkan data pemeriksaan ibu hamil.

3.2 Sumber dan Dataset

Dataset yang digunakan merupakan data sekunder yang diperoleh dari tiga puskesmas, yaitu Puskesmas Mranti, Banyuurip, dan Bayan, dengan total 390 data pasien.

3.3 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian berbasis KDD yang meliputi pengumpulan data, preprocessing, transformasi, pembagian dataset, penerapan SMOTE, pemodelan, dan evaluasi model.

2.3.1 Data Collection

Data dikumpulkan dari tiga puskesmas untuk meningkatkan variasi dan generalisasi model.

**Tabel 1.** Sumber dan Karakteristik Data

Komponen	Keterangan
Jenis Data	Data Sekunder
Sumber Data	Rekam medis Puskesmas Mranti, Banyuurip, dan Bayan
Jumlah Data	390 pasien
Periode Data	Data pemeriksaan trimester ibu hamil
Jumlah Atribut	8 atribut + 1 label
Jenis Label	Normal / Caesar

Tabel 1 menunjukkan sumber serta karakteristik dataset yang digunakan dalam penelitian ini. Data yang digunakan merupakan data sekunder yang diperoleh dari rekam medis di tiga puskesmas, yaitu Puskesmas Mranti, Banyuurip, dan Bayan. Dataset terdiri dari 390 data pasien ibu hamil yang berasal dari hasil pemeriksaan rutin pada masa kehamilan trimester. Setiap data memiliki 8 atribut yang merepresentasikan kondisi klinis ibu hamil, serta 1 label yang menunjukkan jenis metode persalinan, yaitu normal atau Caesar. Karakteristik data ini mencerminkan kondisi nyata di fasilitas pelayanan kesehatan tingkat pertama, sehingga diharapkan dapat memberikan gambaran yang lebih representatif dalam proses pemodelan klasifikasi metode persalinan.

2.3.2 Data Selection

Tahap ini memilih atribut yang relevan, seperti umur, tekanan darah, hemoglobin, protein urine, dan label persalinan.

Tabel 2. Dataset Hasil Seleksi Data

Umur	Sistol	Diastol	HB	HbSAg	Protein Urine	HIV	Biaya	Persalinan
28	120	80	12	Negatif	Negatif	Negatif	BPJS	Normal
34	140	90	10	Negatif	Positif	Negatif	BPJS	Caesar
25	110	70	13	Negatif	Negatif	Negatif	Umum	Normal
30	130	85	11	Negatif	Positif	Negatif	BPJS	Caesar
22	115	75	12	Negatif	Negatif	Negatif	Umum	Normal
36	150	95	9	Negatif	Positif	Negatif	BPJS	Caesar
27	118	78	12	Negatif	Negatif	Negatif	BPJS	Normal
32	135	88	11	Negatif	Positif	Negatif	Umum	Caesar
29	125	82	12	Negatif	Negatif	Negatif	BPJS	Normal
38	145	92	10	Negatif	Positif	Negatif	BPJS	Caesar

Tabel 2 menunjukkan hasil proses seleksi data yang digunakan dalam penelitian. Dataset terdiri dari 390 data ibu hamil dengan atribut pemeriksaan klinis yang relevan terhadap metode persalinan. Atribut yang digunakan meliputi umur, tekanan darah, kadar hemoglobin, status HbSAg, protein urine, HIV, serta jenis pembiayaan. Label yang digunakan adalah metode persalinan yang terdiri dari persalinan normal dan Caesar..

2.3.3 Data Preprocessing

Tahap preprocessing dilakukan untuk membersihkan data dari duplikasi, menangani nilai kosong, serta memastikan konsistensi data sebelum diproses lebih lanjut.

Tabel 3. Tahapan Data Preprocessing

No	Tahapan	Deskripsi
1	Cleaning	Menghapus data duplikat dan inkonsistensi
2	Missing Value	Menangani data kosong
3	Validasi	Memastikan nilai dalam rentang klinis
4	Normalisasi	Menyeragamkan format kategori
5	Filtering	Menghapus atribut tidak relevan

Tabel 3 menjelaskan tahapan-tahapan yang dilakukan dalam proses data preprocessing, mulai dari pembersihan data (*data cleaning*), penanganan nilai yang hilang, validasi nilai atribut, normalisasi format, hingga penyaringan atribut yang tidak relevan. Setiap tahapan memiliki tujuan untuk memastikan bahwa data yang digunakan dalam penelitian berada dalam kondisi yang bersih, konsisten, dan siap untuk dianalisis pada tahap berikutnya.

Tabel 4. Sebelum vs Sesudah Preprocessing

Aspek	Sebelum	Sesudah
Missing Value	Ada kemungkinan	Tidak ada
Duplikasi	Ada kemungkinan	Dihapus
Konsistensi	Tidak seragam	Seragam
Format	Campuran	Siap modeling



Tabel 4 menunjukkan perbandingan kondisi dataset sebelum dan sesudah dilakukan proses preprocessing. Sebelum preprocessing, dataset masih memiliki potensi permasalahan seperti inkonsistensi format, kemungkinan adanya data duplikat, serta nilai atribut yang tidak sesuai dengan rentang klinis. Setelah preprocessing, dataset menjadi lebih bersih, terstruktur, dan konsisten, sehingga dapat meningkatkan kualitas proses pemodelan serta mengurangi potensi kesalahan dalam analisis.

2.3.4 Data Transformation

Data numerik dikategorikan berdasarkan standar klinis, dan data kategorikal dilakukan proses encoding agar dapat digunakan dalam pemodelan.

Tabel 5. Aturan Transformasi Data

Variabel	Nilai	Kategori
Umur	<20	Muda
	20–35	Ideal
	>35	Risiko
HB	<11	Anemia
	11–13	Normal
Sistol	>13	Tinggi
	<140	Normal
	≥140	Tinggi

Tabel 5 menunjukkan aturan transformasi yang digunakan untuk mengubah data numerik ke dalam bentuk kategorikal berdasarkan batas nilai tertentu. Setiap batas nilai ditentukan dengan mempertimbangkan interpretasi klinis yang umum digunakan dalam pemeriksaan ibu hamil, seperti kategori usia risiko kehamilan, kondisi anemia berdasarkan kadar hemoglobin, serta klasifikasi tekanan darah. Transformasi ini bertujuan untuk mempermudah proses klasifikasi sekaligus meningkatkan interpretabilitas model.

Tabel 6. Encoding Data

Variabel	Kategori	Kode
Persalinan	Normal	0
	Caesar	1

Tabel 6 menjelaskan proses pengkodean (*encoding*) terhadap variabel kategorikal menjadi bentuk numerik. Proses ini dilakukan agar data dapat diproses oleh algoritma klasifikasi seperti Decision Tree dan Random Forest. Setiap kategori diberikan kode numerik yang konsisten sehingga tidak mengubah makna data, tetapi mempermudah proses komputasi dalam tahap pemodelan.

2.3.5 Pembagian Dataset

Dataset dibagi menjadi data training dan data testing dengan perbandingan 80:20 untuk melatih dan menguji model.

Tabel 7. Pembagian Dataset

Dataset	Normal	Caesar	Total
Testing	58	20	78
Total	292	98	390

Tabel 7 menunjukkan distribusi data antara data training dan data testing setelah dilakukan pembagian dataset. Data training mencakup 80% dari total dataset yang digunakan untuk membangun model, sedangkan data testing mencakup 20% yang digunakan untuk proses evaluasi. Pembagian ini memastikan bahwa model memiliki cukup data untuk mempelajari pola, sekaligus menyediakan data yang memadai untuk menguji performa model secara objektif.

2.3.6 SMOTE

Metode SMOTE digunakan untuk mengatasi ketidakseimbangan data dengan menambah data sintesis pada kelas minoritas pada data training.

Tabel 8. Setelah SMOTE

Dataset	Normal	Caesar	Total
Training (SMOTE)	234	234	468

Tabel 8 menunjukkan distribusi data pada data training setelah penerapan metode Synthetic Minority Over-sampling Technique (SMOTE). Sebelum dilakukan SMOTE, distribusi kelas pada data training tidak seimbang, di mana jumlah data persalinan normal lebih dominan dibandingkan persalinan Caesar. Kondisi ini berpotensi menyebabkan model cenderung bias terhadap kelas mayoritas dan kurang mampu mengenali pola pada kelas minoritas.



Jumlah data pada tahap pembagian dataset tetap sebesar 390 data. Namun, setelah penerapan metode SMOTE pada data training, jumlah data meningkat menjadi 468 karena adanya penambahan data sintetis pada kelas minoritas.

2.3.7 Pemodelan

Pada tahap pemodelan, data training yang telah melalui preprocessing, transformasi, dan penyeimbangan menggunakan SMOTE digunakan untuk membangun model klasifikasi metode persalinan. Penelitian ini menggunakan dua algoritma, yaitu Decision Tree (C4.5) dan Random Forest.

Decision Tree digunakan karena mampu menghasilkan model yang bersifat interpretatif melalui struktur pohon keputusan yang mudah dipahami. Pemilihan atribut terbaik pada setiap node dilakukan berdasarkan nilai entropy dan information gain, sehingga dapat menghasilkan aturan keputusan yang jelas.

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

Nilai entropy digunakan untuk mengukur seberapa homogen suatu dataset. Jika semua data dalam satu kelas sama, maka nilai entropy = 0 (tidak ada ketidakpastian).

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

Atribut dengan nilai information gain tertinggi akan dipilih sebagai node dalam pohon keputusan, karena dianggap paling mampu membedakan kelas data.

Random Forest digunakan sebagai metode pembandingan berbasis *ensemble learning* yang mampu meningkatkan performa model dengan menggabungkan beberapa pohon keputusan. Dengan menggunakan kedua algoritma tersebut, penelitian ini dapat membandingkan model yang bersifat interpretatif dengan model yang memiliki performa lebih tinggi.

2.3.8 Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kinerja algoritma dalam mengklasifikasikan metode persalinan. Evaluasi dilakukan menggunakan confusion matrix, yang membandingkan antara hasil prediksi model dengan data aktual pada data testing.

Confusion matrix terdiri dari empat komponen utama, yaitu:

1. True Positive (TP): data Caesar yang diprediksi benar sebagai Caesar
2. True Negative (TN): data Normal yang diprediksi benar sebagai Normal
3. False Positive (FP): data Normal yang diprediksi sebagai Caesar
4. False Negative (FN): data Caesar yang diprediksi sebagai Normal

Berdasarkan hasil pengujian model sebelumnya, diperoleh nilai:

1. TP = 10
2. TN = 78
3. FP = 7
4. FN = 5

Dari nilai tersebut, dihitung beberapa metrik evaluasi sebagai berikut:

1. Accuracy

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam memprediksi seluruh data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. Precision

Precision digunakan untuk mengukur tingkat ketepatan prediksi pada kelas positif (Caesar).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. Recall

Recall digunakan untuk mengukur kemampuan model dalam mendeteksi data positif secara benar.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

4. F1-Score

F1-score merupakan rata-rata harmonis antara precision dan recall.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

Berdasarkan hasil evaluasi, model menunjukkan tingkat akurasi yang cukup baik dalam mengklasifikasikan data secara keseluruhan. Namun demikian, nilai precision dan recall yang masih relatif moderat menunjukkan bahwa model belum optimal dalam mendeteksi kelas persalinan Caesar secara konsisten. Hal ini mengindikasikan bahwa model masih memiliki keterbatasan dalam mengenali pola pada kelas minoritas.



3. HASIL DAN PEMBAHASAN

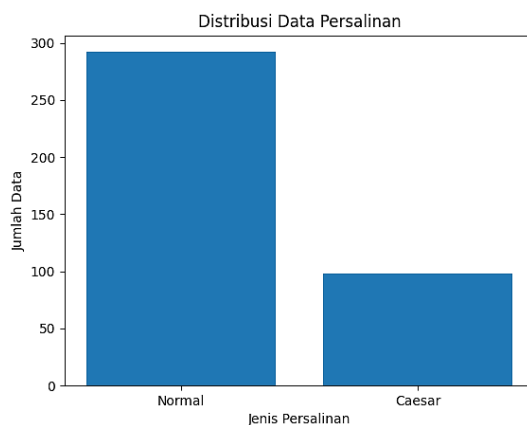
3.1 Distribusi Data Persalinan

Distribusi data digunakan untuk mengetahui komposisi jumlah data antara persalinan normal dan persalinan Caesar dalam dataset penelitian. Berdasarkan hasil pengolahan data, diperoleh total 390 data dengan distribusi sebagai berikut:

1. Normal: 292 data

2. Caesar: 98 data

Distribusi ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*), di mana persalinan normal lebih dominan dibandingkan Caesar. Kondisi ini berpotensi menyebabkan model bias terhadap kelas mayoritas.



Gambar 2. Distribusi Data Persalinan

Gambar 2 menunjukkan distribusi data persalinan dalam dataset penelitian yang terdiri dari dua kelas, yaitu persalinan normal dan persalinan Caesar. Terlihat bahwa jumlah data persalinan normal (292 data) lebih dominan dibandingkan persalinan Caesar (98 data). Perbedaan jumlah yang cukup signifikan ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) dalam dataset. Kondisi tersebut berpotensi menyebabkan model klasifikasi cenderung lebih akurat dalam memprediksi kelas mayoritas (normal) dibandingkan kelas minoritas (Caesar).

3.2 Hasil Prediksi Model

Hasil pengujian model terhadap data testing menunjukkan distribusi prediksi sebagai berikut:

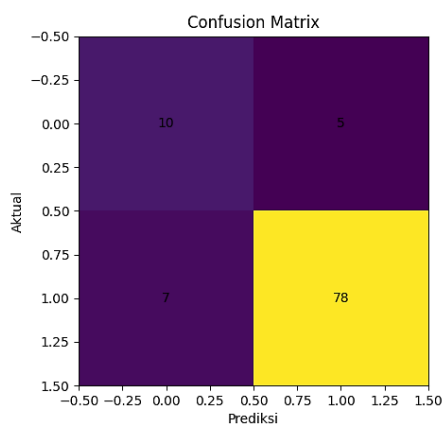
Tabel 9. Hasil Prediksi Model

Jenis Persalinan	Jumlah	Persentase
Normal	85	85%
Caesar	15	15%
Total	100	100%

Tabel 9 menunjukkan bahwa model lebih banyak memprediksi persalinan normal dibandingkan Caesar. Hal ini masih dipengaruhi oleh distribusi data yang didominasi oleh kelas normal.

3.3 Analisis Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi performa model secara lebih rinci dengan membandingkan hasil prediksi dan data aktual.



Gambar 3. Confusion Matrix



Gambar 3 menunjukkan confusion matrix yang menggambarkan perbandingan antara hasil prediksi model dan data aktual pada data testing. Confusion matrix terdiri dari empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Berdasarkan gambar, diperoleh nilai $TP = 10$, $TN = 78$, $FP = 7$, dan $FN = 5$. Nilai TN yang tinggi menunjukkan bahwa model mampu mengklasifikasikan persalinan normal dengan baik. Namun, masih terdapat kesalahan dalam mendeteksi persalinan Caesar, yang ditunjukkan oleh nilai FN dan FP . Kesalahan FN menunjukkan bahwa terdapat data persalinan Caesar yang diprediksi sebagai normal, sedangkan FP menunjukkan bahwa data normal diprediksi sebagai Caesar. Hal ini mengindikasikan bahwa model masih memiliki keterbatasan dalam mengenali pola pada kelas minoritas. Secara keseluruhan, confusion matrix memberikan gambaran bahwa performa model cukup baik pada kelas mayoritas, namun masih perlu ditingkatkan dalam mendeteksi kelas persalinan Caesar secara lebih akurat.

Tabel 10. Confusion Matrix

	Prediksi Caesar	Prediksi Normal
Aktual Caesar	10 (TP)	5 (FN)
Aktual Normal	7 (FP)	78 (TN)

Tabel 10 menunjukkan bahwa model cukup baik dalam mengklasifikasikan data normal (nilai TN tinggi), namun masih terdapat kesalahan dalam mendeteksi kasus Caesar.

3.4 Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score.

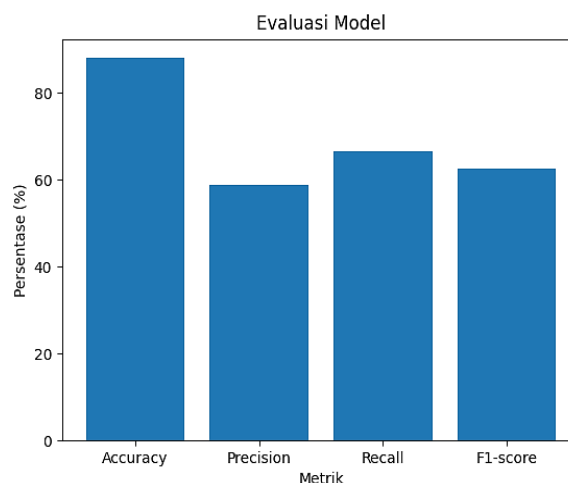
Tabel 11. Hasil Evaluasi Model

Metrik	Nilai
Accuracy	88%
Precision	58.82%
Recall	66.67%
F1-Score	62.50%

Tabel 11 menunjukkan hasil evaluasi performa model klasifikasi menggunakan metrik accuracy, precision, recall, dan F1-score. Nilai accuracy sebesar 88% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun, nilai precision sebesar 58,82% dan recall sebesar 66,67% menunjukkan bahwa kemampuan model dalam mendeteksi persalinan Caesar masih belum optimal. Hal ini mengindikasikan bahwa model masih menghasilkan kesalahan prediksi, terutama pada kelas minoritas. Nilai F1-score sebesar 62,50% mencerminkan keseimbangan antara precision dan recall yang masih berada pada tingkat moderat. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa meskipun model memiliki performa yang cukup baik secara umum, masih diperlukan peningkatan untuk meningkatkan akurasi deteksi pada persalinan Caesar. Perhitungan Accuracy berdasarkan rumus persamaan 3 sebagai berikut.

$$Accuracy = \frac{10 + 78}{10 + 78 + 7 + 5} = 88\%$$

Nilai accuracy menunjukkan bahwa model mampu mengklasifikasikan data dengan cukup baik secara keseluruhan. Namun, nilai precision dan recall yang masih moderat menunjukkan bahwa model belum optimal dalam mendeteksi persalinan Caesar.



Gambar 4. Evaluasi Model



Gambar 4 menunjukkan hasil evaluasi performa model klasifikasi berdasarkan empat metrik utama, yaitu accuracy, precision, recall, dan F1-score. Nilai accuracy mencapai sekitar 88%, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun demikian, nilai precision sebesar 58,82% dan recall sebesar 66,67% menunjukkan bahwa kemampuan model dalam mengidentifikasi kasus persalinan Caesar masih belum optimal. Hal ini mengindikasikan bahwa model masih menghasilkan kesalahan prediksi, baik berupa *false positive* maupun *false negative*, terutama pada kelas minoritas. Nilai F1-score sebesar 62,50% menggambarkan keseimbangan antara precision dan recall yang masih berada pada tingkat moderat. Secara keseluruhan, hasil ini menunjukkan bahwa meskipun model memiliki akurasi yang cukup baik, masih diperlukan peningkatan performa, khususnya dalam mendeteksi persalinan Caesar secara lebih akurat.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa model klasifikasi yang dibangun mampu mengidentifikasi pola hubungan antara variabel klinis ibu hamil dengan metode persalinan, meskipun dengan tingkat performa yang masih terbatas pada kelas minoritas. Nilai akurasi yang relatif tinggi (88%) mengindikasikan bahwa model secara umum mampu mengklasifikasikan data dengan baik. Namun, analisis lebih lanjut melalui precision, recall, dan F1-score menunjukkan bahwa kemampuan model dalam mendeteksi persalinan Caesar masih belum optimal. Hal ini mengindikasikan bahwa akurasi yang tinggi sebagian besar dipengaruhi oleh dominasi kelas persalinan normal dalam dataset.

Ketidakeimbangan data (class imbalance) menjadi faktor utama yang memengaruhi performa model. Meskipun metode SMOTE telah diterapkan untuk menyeimbangkan data training, hasil evaluasi menunjukkan bahwa model masih mengalami kesulitan dalam mengenali pola pada kelas minoritas. Hal ini menunjukkan bahwa penyeimbangan data secara kuantitatif belum sepenuhnya mampu merepresentasikan kompleksitas karakteristik klinis ibu hamil yang beragam, terutama pada kasus yang mengarah pada tindakan persalinan Caesar. Dari sisi interpretasi model, Decision Tree memberikan keunggulan dalam menghasilkan aturan keputusan yang transparan dan mudah dipahami. Atribut seperti tekanan darah, kadar hemoglobin, dan protein urine muncul sebagai faktor dominan dalam proses klasifikasi, yang secara klinis relevan dengan kondisi kehamilan berisiko. Hal ini memperkuat bahwa pendekatan interpretatif memiliki nilai penting dalam konteks pelayanan kesehatan primer, di mana tenaga medis membutuhkan model yang tidak hanya akurat, tetapi juga dapat dijelaskan secara logis.

Di sisi lain, penggunaan Random Forest memberikan peningkatan stabilitas model dengan mengurangi variansi melalui pendekatan ensemble. Meskipun demikian, peningkatan performa yang dihasilkan belum signifikan dalam meningkatkan kemampuan deteksi kelas minoritas. Hal ini menunjukkan bahwa peningkatan kompleksitas model tidak selalu berbanding lurus dengan peningkatan performa, terutama ketika kualitas dan representasi data masih terbatas. Jika dibandingkan dengan penelitian sebelumnya yang menggunakan dataset rumah sakit dan metode yang lebih kompleks, seperti Neural Network atau Support Vector Machine, penelitian ini menunjukkan pendekatan yang lebih sederhana namun memiliki keunggulan dalam interpretabilitas dan relevansi data. Penggunaan data dari fasilitas pelayanan kesehatan tingkat pertama (puskesmas) menjadi kontribusi penting, karena mencerminkan kondisi nyata di lapangan yang seringkali belum banyak dieksplorasi dalam penelitian data mining medis.

Namun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah dataset yang relatif terbatas masih menjadi kendala dalam membangun model yang lebih robust dan generalis. Kedua, meskipun data telah diambil dari tiga puskesmas, variasi karakteristik populasi masih belum cukup luas untuk merepresentasikan kondisi yang lebih beragam. Ketiga, proses transformasi data ke dalam bentuk kategorikal, meskipun berbasis interpretasi klinis, berpotensi mengurangi informasi yang sebenarnya dapat memberikan kontribusi lebih detail dalam proses klasifikasi. Secara keseluruhan, penelitian ini menunjukkan bahwa data pemeriksaan rutin ibu hamil yang sebelumnya hanya bersifat administratif dapat diolah menjadi informasi yang bermakna melalui pendekatan data mining. Namun, hasil yang diperoleh belum cukup kuat untuk digunakan sebagai sistem pendukung keputusan klinis secara langsung. Oleh karena itu, penelitian lanjutan dengan jumlah data yang lebih besar, fitur yang lebih kaya, serta eksplorasi metode yang lebih adaptif sangat diperlukan untuk meningkatkan performa model di masa mendatang..

4. KESIMPULAN

Pendekatan data mining menggunakan algoritma Decision Tree dan Random Forest pada data pemeriksaan ibu hamil di fasilitas pelayanan kesehatan tingkat pertama mampu mengidentifikasi pola hubungan antara variabel klinis dengan metode persalinan. Atribut seperti tekanan darah, kadar hemoglobin, dan protein urine terbukti menjadi faktor dominan dalam proses klasifikasi, yang secara klinis relevan dengan kondisi kehamilan berisiko. Berdasarkan hasil evaluasi model, diperoleh nilai accuracy sebesar 88%, precision 58,82%, recall 66,67%, dan F1-score 62,50%. Hasil ini menunjukkan bahwa meskipun model memiliki tingkat ketepatan yang cukup baik secara umum, kemampuan dalam mendeteksi persalinan Caesar masih berada pada tingkat moderat. Kondisi ini mengindikasikan bahwa model cenderung lebih optimal dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas, yang dipengaruhi oleh ketidakseimbangan distribusi data. Penerapan metode SMOTE pada data training memberikan kontribusi dalam mengurangi ketidakseimbangan kelas, namun belum sepenuhnya mampu meningkatkan performa model secara signifikan dalam mendeteksi kasus persalinan Caesar. Selain itu, penggunaan Random Forest sebagai metode ensemble memberikan peningkatan stabilitas model dibandingkan Decision Tree, meskipun belum menghasilkan peningkatan



performa yang substansial pada kelas minoritas. Pemanfaatan data pemeriksaan rutin di puskesmas memiliki potensi sebagai sumber informasi untuk analisis pola awal dalam bidang kesehatan ibu dan anak. Namun demikian, performa model yang diperoleh dalam penelitian ini belum memadai untuk digunakan sebagai sistem pendukung keputusan klinis secara langsung, mengingat masih adanya tingkat kesalahan prediksi yang signifikan pada kasus persalinan Caesar. Keterbatasan penelitian ini terletak pada jumlah dataset yang masih terbatas serta variasi data yang belum sepenuhnya merepresentasikan kondisi populasi yang lebih luas. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, menambah variasi sumber data, serta mengeksplorasi metode yang lebih adaptif untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas secara lebih akurat..

REFERENCES

- [1] Juliana Widyastuti Wahyuningsih and Intan Sari, "Perbandingan Tingkat Nyeri Pada Persalinan Normal Pervaginam Pada Ibu Inpartu yang Diberikan Terapi Hypnobirthing dan Tanpa Terapi Hypnobirthing," *Jurnal Kebidanan : Jurnal Ilmu Kesehatan Budi Mulia*, vol. 12, no. 2, pp. 124–131, Dec. 2022, doi: 10.35325/kebidanan.v12i2.317.
- [2] A. C. Abela and Andriana Kumala Dewi, "Karakteristik Ibu Bersalin dengan Preeklampsia di RSUD Kabupaten Natuna," *JIMKI: Jurnal Ilmiah Mahasiswa Kedokteran Indonesia*, vol. 12, no. 1, pp. 106–114, Aug. 2025, doi: 10.53366/jimki.v12i1.836.
- [3] A. Arum, I. N. Budiarta, F. O. Shariff, and A. Jhonet, "Hubungan Antara Usia Ibu Dengan Kejadian Ketuban Pecah Dini Di Rumah Sakit Artha Bunda Kabupaten Lampung Tengah," *Jurnal Ilmu Kedokteran dan Kesehatan*, vol. 11, no. 9, pp. 1726–1731, Oct. 2024, doi: 10.33024/jikk.v11i9.15596.
- [4] Vyona Wifiarni Putri, Desi Pramita Sari, and Resi Novia, "Hubungan Kelainan Letak dan Paritas dengan Kejadian Ketuban Pecah Dini di Rumah Sakit Harapan Bunda Kota Batam Tahun 2024," *Jurnal Kesehatan Amanah*, vol. 9, no. 2, pp. 233–242, Sep. 2025, doi: 10.57214/jka.v9i2.959.
- [5] N. Ghuman, J. Rheiner, B. E. Tendler, and W. B. White, "Hypertension in the Postpartum Woman: Clinical Update for the Hypertension Specialist," *The Journal of Clinical Hypertension*, vol. 11, no. 12, pp. 726–733, Dec. 2009, doi: 10.1111/j.1751-7176.2009.00186.x.
- [6] S. Fatimah and L. F. Oktaviana, "Faktor yang berhubungan dengan persalinan sectio caesaria," *THE JOURNAL OF Mother and Child Health Concerns*, vol. 4, no. 2, pp. 64–72, Dec. 2024, doi: 10.56922/mhc.v4i2.506.
- [7] C. M. Angolile, B. L. Max, J. Mushemba, and H. L. Mashauri, "Global increased cesarean section rates and public health implications: A call to action," *Health Sci. Rep.*, vol. 6, no. 5, May 2023, doi: 10.1002/hsr2.1274.
- [8] C. Roos Ety, E. Damani, and G. J. Nababan, "Faktor-Faktor yang Berhubungan Terhadap Persalinan dengan Kejadian Sectio Caesarea di Rumah Sakit Umum Dr. Hadrianus Sinaga Pangururan," *JURNAL HEALTH REPRODUCTIVE*, vol. 8, no. 2, pp. 13–23, Dec. 2023, doi: 10.51544/jrh.v8i2.4615.
- [9] Verenia, Jasmawati, and R. Firdaus, "The Effect of Early Mobilization Implementation on the Healing of Sectio Caesarea Operation Wounds at Harapan Insan Sendawar Hospital," *Formosa Journal of Science and Technology*, vol. 1, no. 8, pp. 1165–1184, Dec. 2022, doi: 10.55927/fjst.v1i8.2115.
- [10] F. A. Pribadi and E. Setijaninrum, "Analisis Prinsip Pembiayaan Kesehatan dalam Mendukung Cakupan Kesehatan Semesta di Indonesia," *Jejaring Administrasi Publik*, vol. 15, no. 2, pp. 60–78, Dec. 2023, doi: 10.20473/jap.v15i2.49582.
- [11] G. Abdurrahman and J. T. Wijaya, "Analisis Klasifikasi Kelahiran Caesar Menggunakan Algoritma Naive Bayes," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 4, no. 2, p. 46, Aug. 2019, doi: 10.32528/justindo.v4i2.2616.
- [12] R. M. Widhyasih, M. K. Nengsih, and Z. Zuraida, "Gambaran Hasil Pemeriksaan Anti-HIV dan HBsAg Metode Immunochromatography Tes Pada Ibu Hamil Di RSAB Harapan Kita Jakarta," *Anakes : Jurnal Ilmiah Analis Kesehatan*, vol. 9, no. 2, pp. 138–146, Sep. 2023, doi: 10.37012/anakes.v9i2.1856.
- [13] I. Sulihati, A. Syukur, and A. Marjuni, "Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2672.
- [14] I. C. Setia and T. Arifin, "Penentuan Penanganan Persalinan Caesar dengan Neural Network dan Particle Swarm Optimization," *SISTEMASI*, vol. 10, no. 2, p. 346, May 2021, doi: 10.32520/stmsi.v10i2.1235.
- [15] J. Gomes and M. Romão, "Evaluating Maturity Models in Healthcare Information Systems: A Comprehensive Review," *Healthcare*, vol. 13, no. 15, p. 1847, Jul. 2025, doi: 10.3390/healthcare13151847.
- [16] D. M. U. Atmaja, A. R. Hakim, A. Basri, and A. Ariyanto, "Klasifikasi Metode Persalinan pada Ibu Hamil Menggunakan Algoritma Random Forest Berbasis Mobile," *JRST (Jurnal Riset Sains dan Teknologi)*, vol. 7, no. 2, p. 161, Sep. 2023, doi: 10.30595/jrst.v7i2.16705.
- [17] H. Amalia, R. Rahmadanti, A. Syaiin, S. Salsabila4, Y. Yunita, and S. Sriyadi, "Prediksi Resiko Kesehatan Ibu Hamil Dengan Menggunakan Metode Decision Tree," *Swabumi*, vol. 11, no. 1, pp. 48–53, Mar. 2023, doi: 10.31294/swabumi.v11i1.15270.
- [18] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data," *Techno.Com*, vol. 23, no. 1, pp. 1–10, Feb. 2024, doi: 10.62411/tc.v23i1.9779.
- [19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [20] D. Nurmalasari, H. R. Yuliantoro, and D. H. Qudsi, "Improving Panic Disorder Classification Using SMOTE and Random Forest," *Journal of Applied Informatics and Computing*, vol. 8, no. 2, pp. 272–279, Oct. 2024, doi: 10.30871/jaic.v8i2.8315.
- [21] A. B. Alfianah, M. R. Pratama, M. Yunus, and V. Vestine, "Optimasi Keputusan Metode Persalinan dengan Algoritma C4.5," *J-REMI : Jurnal Rekam Medik dan Informasi Kesehatan*, vol. 6, no. 1, pp. 77–87, Dec. 2024, doi: 10.25047/j-remi.v6i1.4777.