



Klasifikasi Kondisi Janin Berdasarkan Data Kardiotogram Menggunakan Algoritma Naive Bayes

Isruel Syah Utama, Elin Haerani*, Fitri Wulandari, Siti Ramadhani

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11850114441@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin.suska.ac.id, ³fitri.wulandari@uin-suska.ac.id,

⁴siti.ramadhani@uin-suska.ac.id

Email Penulis Koresponden: elin.haerani@uin.suska.ac.id

Abstrak—Kesehatan janin selama kehamilan merupakan aspek krusial yang dapat dipantau melalui data kardiotogram (KTG), namun interpretasi manual terhadap data ini sering mengalami kendala akibat ketidakseimbangan distribusi kelas. Penelitian ini bertujuan untuk membangun model klasifikasi kondisi janin menggunakan algoritma Naive Bayes yang dikombinasikan dengan teknik Synthetic Minority Over-sampling Technique (SMOTE) guna mengatasi ketimpangan jumlah data antar kelas. Data KTG yang digunakan diperoleh dari situs Kaggle, terdiri atas 2.126 data dengan tiga kelas target: Normal, Suspect, dan Pathological. Proses pengolahan data dilakukan melalui tahapan Knowledge Discovery in Databases (KDD), mencakup seleksi data, pembersihan, normalisasi, pembagian data dalam empat rasio (70:30, 80:20, 85:15, dan 90:10), penerapan SMOTE, serta evaluasi model menggunakan metrik akurasi dan F1-Macro. Hasil penelitian menunjukkan bahwa rasio 80:20 menghasilkan akurasi tertinggi sebesar 79,81%, sedangkan rasio 90:10 memberikan nilai F1-Macro tertinggi sebesar 0,6788. Temuan ini menunjukkan bahwa meskipun akurasi relatif stabil, F1-Macro memberikan gambaran performa yang lebih seimbang antar kelas, khususnya pada kelas minoritas. Penerapan SMOTE terbukti efektif dalam memperbaiki distribusi kelas dan meningkatkan sensitivitas model. Penelitian ini menjadi dasar awal dalam pengembangan sistem klasifikasi kondisi janin yang lebih andal dan adaptif, serta membuka peluang eksplorasi lebih lanjut terhadap algoritma pembandingan dan parameter SMOTE.

Kata kunci: Kardiotogram; Naive Bayes; SMOTE; Klasifikasi Janin; Ketidakseimbangan Data

Abstract—Fetal health during pregnancy is a crucial aspect that can be monitored through cardiocography (CTG) data; however, manual interpretation of this data often encounters challenges due to class imbalance. This study aims to develop a fetal condition classification model using the Naive Bayes algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE) to address the disparity in class distribution. The CTG dataset, obtained from Kaggle, consists of 2,126 records categorized into three target classes: Normal, Suspect, and Pathological. Data processing followed the Knowledge Discovery in Databases (KDD) framework, including data selection, cleaning, normalization, splitting into four ratios (70:30, 80:20, 85:15, and 90:10), SMOTE application, and model evaluation using accuracy and F1-Macro metrics. The results showed that the 80:20 ratio yielded the highest accuracy at 79.81%, while the 90:10 ratio produced the highest F1-Macro score of 0.6788. These findings indicate that although accuracy remained relatively stable, the F1-Macro metric provided a better representation of performance across all classes, especially minority ones. The application of SMOTE proved effective in balancing class distribution and enhancing model sensitivity. This study serves as a foundational step in developing a more reliable and adaptive fetal condition classification system and highlights opportunities for further exploration of alternative algorithms and SMOTE parameter optimization.

Keywords: Cardiocography; Naive Bayes; SMOTE; Fetal Classification; Imbalanced Data

1. PENDAHULUAN

Pemantauan kondisi janin selama kehamilan merupakan aspek penting dalam perawatan kesehatan ibu dan anak [1]. Salah satu metode yang digunakan untuk memantau kondisi janin adalah melalui pengukuran Kardiotogram (KTG), yang mencatat variabilitas detak jantung janin, gerakan janin, kontraksi rahim, serta beberapa parameter lainnya yang dapat menggambarkan respons janin terhadap lingkungannya [2]. Data KTG memberikan gambaran yang lebih jelas tentang kondisi fisiologis janin, yang sangat berguna dalam mendeteksi masalah potensial, seperti adanya stres pada janin yang dapat membahayakan keselamatannya [3]. Oleh karena itu, pengolahan dan analisis data KTG menjadi sangat penting dalam dunia kedokteran untuk membuat keputusan yang cepat dan tepat dalam perawatan janin.

Kehamilan memerlukan pemantauan yang cermat terhadap kondisi janin untuk mendeteksi kemungkinan komplikasi yang dapat mempengaruhi kesejahteraannya. KTG digunakan untuk memantau kondisi janin, merekam detak jantung janin (FHR) serta kontraksi rahim (UC)[4]. Data yang dihasilkan oleh KTG memungkinkan tenaga medis untuk mengidentifikasi potensi risiko terhadap janin. Namun, interpretasi data KTG secara manual oleh tenaga medis dapat menimbulkan tantangan, terutama dalam kondisi ketidakseimbangan jumlah data antar kelas dalam dataset [5].

Berbagai penelitian telah dilakukan dalam klasifikasi kondisi janin menggunakan data KTG, seperti yang dilakukan oleh [6] yang menggunakan *Support Vector Machines* (SVM) dan penelitian oleh [7] yang menggunakan teknik Deep Learning menunjukkan hasil yang sangat baik dalam klasifikasi, namun memerlukan data dalam jumlah besar serta waktu pelatihan yang panjang, serta belum mengatasi masalah ketidakseimbangan kelas yang dapat menurunkan kualitas prediksi pada kelas minoritas. Di sisi lain, [8] dan [9] mengaplikasikan teknik SMOTE untuk mengatasi ketidakseimbangan kelas dalam dataset KTG, namun penelitian tersebut tidak membandingkan teknik tersebut dengan metode lain seperti Naive Bayes yang lebih efisien dalam hal waktu komputasi. Penelitian ini bertujuan untuk menggabungkan Naive Bayes dengan teknik SMOTE, serta mengevaluasi model secara menyeluruh menggunakan precision, recall, dan F1-score, yang selama ini jarang dilakukan secara komprehensif pada penelitian sebelumnya.

Ketidakeimbangan kelas dapat mempengaruhi akurasi model klasifikasi, karena model cenderung lebih memprediksi kelas mayoritas [10]. Teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dapat meningkatkan



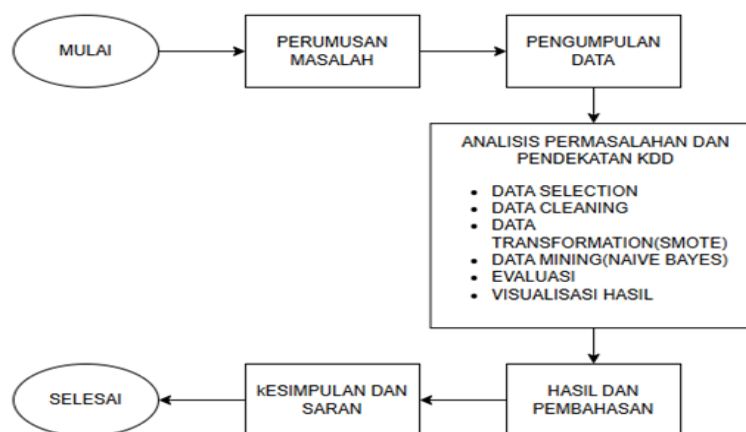
jumlah data pada kelas minoritas dan memperbaiki performa model [11]. Naive Bayes dipilih karena kesederhanaannya, efektivitasnya dalam menangani data yang memiliki dimensi tinggi, serta kemampuannya dalam memberikan probabilitas yang berguna untuk klasifikasi [12]. Untuk menjamin bahwa proses klasifikasi dilakukan secara terorganisir dan sistematis, penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) [13]. KDD adalah kerangka kerja berurutan yang komprehensif dan berperan penting dalam mengekstraksi pengetahuan meliputi pemeliharaan kualitas data, pemilihan strategi pemodelan yang tepat, serta pemahaman mendalam terhadap pola-pola klasifikasi yang diperoleh [14].

Penelitian ini berkontribusi pada pengembangan sistem klasifikasi otomatis yang dapat digunakan oleh tenaga medis dalam memantau kondisi janin dengan lebih akurat. Keberhasilan sistem ini dapat membantu mengurangi beban kerja tenaga medis dalam menganalisis data KTG dan meningkatkan deteksi dini terhadap kondisi janin yang berisiko tinggi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan terlihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan dari Penelitian yang di rencanakan digambarkan dalam Gambar 1.

a. Perumusan Masalah

Data KTG sering kali memiliki ketidakseimbangan kelas. Hal ini menyebabkan model klasifikasi cenderung memprediksi kelas mayoritas dan mengabaikan kondisi janin yang lebih jarang ditemukan [15]. Penelitian ini bertujuan untuk mengatasi masalah ketidakseimbangan kelas dengan menggunakan algoritma Naive Bayes dan SMOTE. Dengan pendekatan ini, diharapkan dapat meningkatkan akurasi prediksi kondisi janin, terutama untuk kelas suspect dan pathological, serta membantu dalam pengambilan keputusan medis yang lebih tepat.

b. Pengumpulan Data

Dataset KTG yang diklasifikasikan oleh D. Campos dkk ini, diperoleh sebagai data sekunder dari situs Kaggle, dalam format .csv. Dataset ini terdiri dari 2.126 catatan pemeriksaan janin yang mencakup 22 atribut, yaitu 21 fitur numerik dan 1 atribut target, yaitu fetal_health. Atribut target tersebut mengklasifikasikan kondisi janin ke dalam tiga kategori: 1 = Normal, 2 = Suspect, dan 3 = Pathological

c. Analisis Permasalahan dan Penerapan KDD

Bagaimana penerapan data mining untuk mengklasifikasikan kondisi janin berdasarkan data KTG efisien menggunakan Algoritma Naïve Bayes dan Teknik SMOTE untuk mengatasi distribusi data yang tidak seimbang [16]. untuk menjawab permasalahan secara sistematis, penelitian ini mengadopsi kerangka kerja KDD. Setiap tahap dalam KDD berfungsi sebagai panduan utama dalam analisis data dan pengambilan keputusan [17]. Berdasarkan pada tahapan KDD:

1. Data Selection

Berdasarkan tujuan analisis, seleksi data berperan dalam pemilihan data yang tepat dari kumpulan data besar dan menetapkan label yang dibutuhkan.

2. Data Cleaning

Data cleaning bertujuan untuk memperbaiki kualitas data yang tidak valid, hilang, atau tidak konsisten.

3. Data Transformation.

Transformasi data bertujuan untuk mempersiapkan data dalam format yang tepat untuk proses pemodelan seperti normalisasi dan pembagian data.

4. Data Mining



Data Mining menggunakan teknik analisis data untuk menemukan pola atau informasi tersembunyi dalam dataset yang besar.

5. Evaluation

Menilai efektivitas algoritma dalam melakukan klasifikasi kondisi janin dengan *classification report*.

6. Visualisasi Model

Memberikan gambaran yang lebih jelas mengenai bagaimana model bekerja dalam mengklasifikasikan kondisi janin melalui confusion matrix dan grafik.

2.2 Synthetic Minority Over-sampling Technique (SMOTE)

Masalah ketidakseimbangan kelas, dimana jumlah data pada kelas minoritas jauh lebih sedikit dibanding kelas mayoritas, dapat menyebabkan model klasifikasi menjadi bias. SMOTE merupakan teknik oversampling yang digunakan untuk menyeimbangkan distribusi kelas dengan menambah sampel sintetis pada kelas minoritas di dataset pelatihan [18]. Ketidakseimbangan kelas dalam dataset diatasi dengan menghasilkan data sintetis pada kelas minoritas, bertujuan menambah jumlah sampel tanpa mengurangi kualitas data, sehingga model dapat belajar lebih efektif dari data yang seimbang [19]. Proses ini dilakukan sebagai berikut:

Secara umum, proses SMOTE [20] terdiri dari langkah-langkah sebagai berikut:

a. Menentukan Tetangga Terdekat

Untuk setiap sampel pada kelas minoritas, SMOTE mencari beberapa tetangga terdekat dari sampel tersebut berdasarkan metrik jarak, seperti Euclidean Distance.

b. Pembuatan Sampel Sintetis.

Setelah tetangga terdekat ditemukan, SMOTE menghasilkan sampel sintetis dengan cara menginterpolasi antara titik data yang ada dan tetangga terdekatnya. Proses ini menciptakan titik baru di antara titik data yang ada, berdasarkan nilai rata-rata dari fitur-fitur mereka.

Formula untuk pembuatan sampel sintetis adalah:

$$New\ Sample = Sample_i + \lambda \times (Neighbor_j - Sample_i) \quad (1)$$

Dimana:

$Sample_i$ adalah sampel kelas minoritas yang ada.

$Neighbor_j$ adalah salah satu tetangga terdekat dari sampel tersebut.

λ adalah parameter acak yang digunakan untuk menghasilkan sampel sintetis di antara sampel i dan tetangga j .

2.3 Algoritma Naive Bayes

Algoritma Naive Bayes adalah metode klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes dengan asumsi bahwa fitur-fitur dalam data bersifat independen satu sama lain (naive independence assumption) [21]. Algoritma ini menghitung probabilitas posterior dari setiap kelas berdasarkan fitur-fitur input, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi [22].

Tahapan algoritma Naive Bayes:

a. Hitung probabilitas prior tiap kelas $P(C_k)$,

$$P(C_k) = \frac{\text{Jumlah data pada kelas } C_k}{\text{Total data latih}} \quad (2)$$

b. Hitung rata-rata μ_k dan varians σ_k^2 tiap fitur perkelas,

c. Menghitung Probabilitas Kondisional dengan Distribusi Gaussian:

$$P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}\right) \quad (3)$$

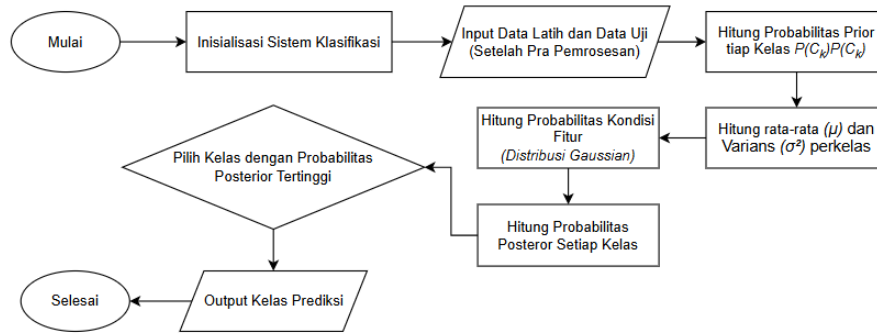
d. Menghitung Probabilitas Posterior, mengalikan probabilitas prior kelas dan semua probabilitas kondisional fitur

$$P(C_k|X) = \frac{P(C_k)}{P(X)} \prod_{i=1}^n P(x_i|C_k) \quad (4)$$

e. Menentukan Kelas dengan Probabilitas Tertinggi.

2.4 Flowchart

Diagram *flowchart* memberikan gambaran visual tentang tahapan-tahapan dalam suatu algoritma atau proses. Alat ini kerap digunakan oleh analis sistem untuk menjelaskan logika operasional serta mendeteksi potensi masalah teknis. Simbol yang terdapat dalam *flowchart* dihubungkan oleh garis alur untuk menunjukkan urutan eksekusi [23]. *Flowchart* Naive Bayes [15] dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Naive Bayes

Gambar 2 menunjukkan alur proses kerja algoritma Naive Bayes dalam melakukan klasifikasi. Proses dimulai dengan inisialisasi sistem klasifikasi, kemudian memasukkan data latih dan data uji yang telah melalui tahap pra-pemrosesan. Selanjutnya, algoritma menghitung probabilitas prior untuk setiap kelas, yaitu probabilitas dasar keberadaan kelas tersebut. Setelah itu, dihitung rata-rata dan varians dari setiap fitur untuk masing-masing kelas, yang digunakan dalam distribusi *Gaussian* untuk memperkirakan probabilitas kondisi fitur. Kemudian, probabilitas posterior untuk setiap kelas dihitung berdasarkan data uji. Pada tahap berikutnya, kelas dengan probabilitas posterior tertinggi dipilih sebagai kelas prediksi untuk data uji. Proses diakhiri dengan mengeluarkan hasil prediksi kelas tersebut.

3. HASIL DAN PEMBAHASAN

Pengembangan model klasifikasi kondisi janin dengan algoritma Naive Bayes menggunakan dataset dengan judul "Cardiotocography (CTG)" diambil dari situs kaggle. 21 atribut fitur numerik sebagai parameter fisiologis dan 1 atribut target klasifikasi janin (normal, suspect dan pathological).

Tahapan KDD:

a. Data Selection

Menyeleksi variabel $x = 21$ atribut fitur numerik dan variabel $y = 1$ target klasifikasi.

b. Data Cleaning

Tidak terdapat kekosongan nilai dan *outlier* pada data dianggap membawa informasi penting dari sisi medis, sehingga akan ditangani dengan normalisasi

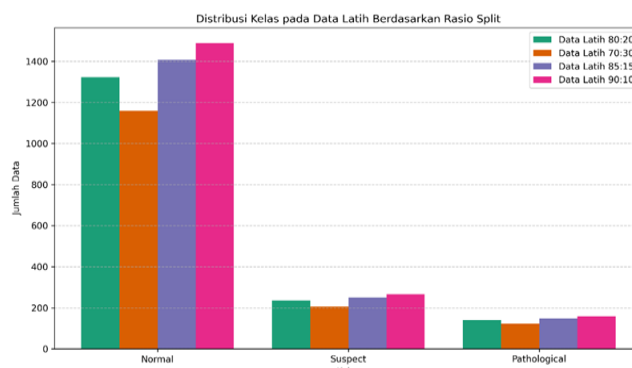
c. Data Transformation

Normalisasi dengan *min-max scaling* agar nilai berada pada rentang 0 sampai 1. Setelah itu pembagian data seperti pada Tabel 1 berikut ini:

Tabel 1. Distribusi Data Dengan 4 proporsi

Pengujian	Proporsi	Data Latih	Normal	Suspect	Pathological	Support
1	70:30	1488	1158	207	123	638
2	80:20	1700	1323	235	142	426
3	85:15	1807	1407	251	149	319
4	90:10	1913	1489	266	158	213

Tabel 1 menunjukkan distribusi data latih dan data uji yang digunakan dalam pengujian model berdasarkan empat proporsi pembagian data yang berbeda. Pembagian data ini mempengaruhi jumlah sampel yang digunakan untuk pelatihan (data latih) dan pengujian (data uji), berikut akan di tampilkan grafik sesuai Tabel 1.



Gambar 4. Grafik distribusi kelas berdasarkan proporsi pengujian

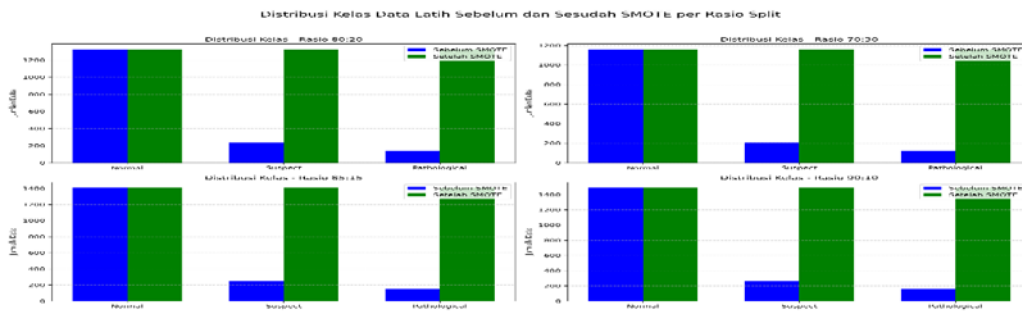


Gambar 4 menampilkan grafik distribusi kelas, berdasarkan grafik dapat disimpulkan bahwa distribusi kelas terdapat ketidakseimbangan kelas (*class imbalance*). SMOTE berperan dalam penyeimbangan data dan menghasilkan distribusi kelas sebagai berikut.

Tabel 2. Setelah Penerapan SMOTE

Proporsi	Normal	Suspect	Pathological
70:30	1158	1158	1158
80:20	1323	1323	1323
85:15	1406	1406	1406
90:10	1489	1489	1489

Dapat terlihat pada Tabel 2 distribusi kelas setelah di terapkan teknik SMOTE, data memiliki jumlah yang seimbang di setiap kelas. Grafik distribusi kelas berdasarkan Tabel 2 dapat dilihat pada Gambar 5



Gambar 5. Distribusi kelas setelah SMOTE

Gambar 5 menampilkan distribusi kelas setelah di terapkan teknik SMOTE. Sebelum penerapan SMOTE, distribusi kelas tidak seimbang, dengan kelas Normal memiliki jumlah yang jauh lebih banyak dibandingkan kelas Suspect dan Pathological. Setelah penerapan SMOTE, jumlah sampel pada kelas minoritas meningkat, membuat distribusi kelas lebih seimbang dan memberikan model kesempatan untuk belajar lebih baik pada kelas-kelas yang lebih sedikit jumlahnya

d. Data Mining

Alur *flowchart* Naive Bayes:

1. Inisialisasi Sistem Klasifikasi
Inisialisasi sudah di persiapan untuk proses klasifikasi
2. Input Data Latih dan Data Uji (Setelah Pra-Pemrosesan)
Data latih dan data uji berhasil di input
3. Perhitungan Probabilitas Prior Tiap Kelas
Setelah didapat probabilitas prior, hasilnya akan digunakan sebagai dasar klasifikasi
4. Perhitungan Rata-Rata (μ) dan Varians (σ^2) Tiap Fitur per Kelas
Hasil dari perhitungan (μ) dan (σ^2) akan di gunakan sebagai parameter gaussian
5. Perhitungan Probabilitas Kondisi Fitur (Likelihood)
Probabilitas dihitung dengan parameter (μ) dan (σ^2) dengan distribusi Gaussian.
6. Perhitungan Probabilitas Posterior Setiap Kelas
Hitung likelihood dan probabilitas gabungan untuk menghasilkan nilai posterior
7. Pemilihan Kelas dengan Probabilitas Posterior Tertinggi
Posterior tertinggi dipilih yang kemudian akan menjadi hasil prediksi
8. Output Kelas Prediksi
Hasil prediksi merupakan output dari klasifikasi

e. Evaluasi

Hasil prediksi ditampilkan pada Tabel 4.

Tabel 4. Hasil Prediksi dengan Proporsi 70:30, 80:20, 85:15, dan 90:10

Uji Coba	Proporsi	Actual \ Pred	Normal	Suspect	Pathological
Uji 1	70:30	Normal	406	70	13
		Suspect	6	74	7
		Pathological	2	22	29
Uji 2	80:20	Normal	274	49	9
		Suspect	5	48	6
		Pathological	1	16	10
Uji 3	85:15	Normal	204	39	5



Uji Coba	Proporsi	Actual \ Pred	Normal	Suspect	Pathological
Uji 4	90:10	Suspect	3	36	5
		Pathological	1	12	14
		Normal	132	29	5
		Suspect	1	25	3
		Pathological	1	6	11

Tabel 4 menampilkan hasil prediksi berupa confusion matrix dari proporsi berbeda. Semakin besar proporsi, prediksi benar akan meningkat, kelas Normal sangat terlihat peningkatan prediksi. Sedangkan, kesalahan klasifikasi antar kelas menurun, pada kelas Suspect dan Pathological terlihat jelas penurunan kesalahan prediksi.

1. Evaluasi Model Naive Bayes untuk Rasio 70:30

Classification Report:

Tabel 5. Evaluasi Performa Model rasio 70:30

Kelas	Precision	Recall	F1-Score	Support
Normal (1.0)	0.98	0.82	0.89	497
Suspect (2.0)	0.42	0.83	0.56	88
Pathological (3.0)	0.59	0.55	0.57	53
Accuracy			0.7962	638
F1 Macro			0.6725	
Macro Avg	0.66	0.73	0.67	638
Weighted Avg	0.87	0.80	0.82	638

Tabel 5 menunjukkan performa model Naive Bayes dengan pembagian data pelatihan dan pengujian sebesar 70:30. Model mencapai akurasi sebesar 79,62% pada data pengujian. Kelas Normal (1.0) memiliki precision dan recall yang tinggi, masing-masing 0,98 dan 0,82, sedangkan kelas Suspect dan Pathological menunjukkan precision yang lebih rendah (0,42 dan 0,59) dengan recall yang lebih baik (0,83 dan 0,55). Nilai F1 Macro sebesar 0,6725 menunjukkan keseimbangan antara precision dan recall secara keseluruhan.

2. Evaluasi Model Naive Bayes untuk Rasio 80:20

Classification Report:

Tabel 6. Evaluasi Performa Model rasio 80:20

Kelas	Precision	Recall	F1-Score	Support
Normal (1.0)	0.98	0.83	0.90	332
Suspect (2.0)	0.42	0.81	0.56	59
Pathological (3.0)	0.55	0.51	0.53	35
Accuracy			0.7981	426
F1 Macro			0.6610	
Macro Avg	0.65	0.72	0.66	426
Weighted Avg	0.87	0.80	0.82	426

Tabel 6 menyajikan hasil evaluasi model dengan rasio pembagian data 80:20. Akurasi yang diperoleh adalah 79,81%. Kelas Normal mempertahankan precision dan recall yang tinggi (0,98 dan 0,83), sementara kelas Suspect dan Pathological memiliki precision yang lebih rendah (0,42 dan 0,55) dengan recall moderat (0,81 dan 0,51). F1 Macro sebesar 0,6610 mengindikasikan keseimbangan performa model di seluruh kelas.

3. Evaluasi Model Naive Bayes untuk Rasio 85:15

Classification Report:

Tabel 7. Evaluasi Performa Model rasio 85:15

Kelas	Precision	Recall	F1-Score	Support
Normal (1.0)	0.98	0.82	0.89	248
Suspect (2.0)	0.41	0.82	0.55	44
Pathological (3.0)	0.58	0.52	0.55	27
Accuracy			0.7962	319
F1 Macro			0.6645	
Macro Avg	0.66	0.72	0.66	319
Weighted Avg	0.87	0.80	0.82	319

Pada tabel 7 ditampilkan evaluasi model dengan rasio pembagian data 85:15. Model menunjukkan akurasi 79,62%. Precision dan recall pada kelas Normal tetap tinggi (0,98 dan 0,82), namun kelas Suspect dan Pathological masih



memiliki precision yang rendah (0,41 dan 0,58) dengan recall yang moderat (0,82 dan 0,52). F1 Macro sebesar 0,6645 menggambarkan keseimbangan performa antar kelas.

4. Evaluasi Model Naive Bayes untuk Rasio 90:10

Classification Report:

Tabel 8. Evaluasi Performa Model rasio 90:10

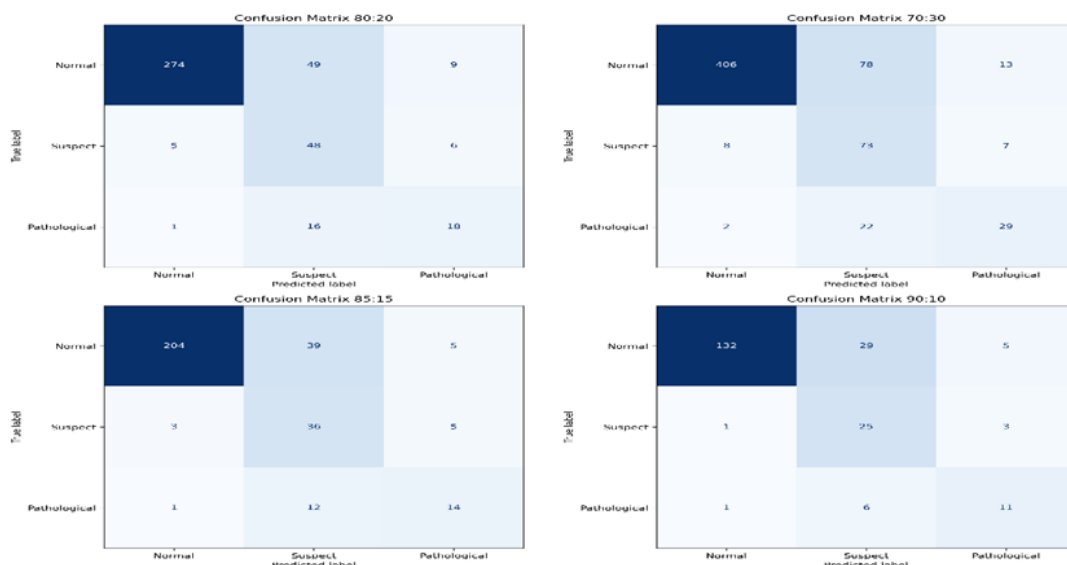
Kelas	Precision	Recall	F1-Score	Support
Normal (1.0)	0.99	0.80	0.88	166
Suspect (2.0)	0.42	0.86	0.56	29
Pathological (3.0)	0.58	0.61	0.59	18
Accuracy			0.7887	213
F1 Macro			0.6788	
Macro Avg	0.66	0.76	0.68	213
Weighted Avg	0.87	0.79	0.81	213

Tabel 8 memperlihatkan evaluasi model dengan rasio data pelatihan dan pengujian 90:10. Model mencapai akurasi sebesar 78,87%. Kelas Normal memiliki precision dan recall yang baik (0,99 dan 0,80), sedangkan kelas Suspect dan Pathological menunjukkan precision yang rendah (0,42 dan 0,58) dengan recall masing-masing 0,86 dan 0,61. F1 Macro sebesar 0,6788 menunjukkan performa yang cukup seimbang di seluruh kelas meskipun dengan rasio pengujian yang lebih kecil.

Hasil evaluasi model Naive Bayes dengan variasi rasio pembagian data latih dan uji menunjukkan bahwa rasio 80:20 menghasilkan akurasi tertinggi sebesar 79,81%. Hal ini mengindikasikan bahwa proporsi data pelatihan dan pengujian pada rasio tersebut memberikan keseimbangan yang baik, memungkinkan model untuk belajar secara efektif sekaligus menguji kemampuannya dengan data yang representatif. Di sisi lain, rasio 90:10 memperlihatkan nilai F1 Macro terbaik sebesar 0,6788, yang mencerminkan kemampuan model dalam menjaga keseimbangan antara precision dan recall di seluruh kelas, terutama pada kelas minoritas seperti Suspect dan Pathological. Kondisi ini menunjukkan bahwa meskipun data pengujian lebih sedikit, model tetap dapat mempertahankan performa yang seimbang antar kelas. Dengan demikian, rasio 80:20 dapat dipandang sebagai konfigurasi yang optimal untuk mencapai akurasi keseluruhan terbaik, sementara rasio 90:10 memberikan keunggulan dalam hal keseimbangan performa antar kelas. Penerapan teknik SMOTE juga berperan penting dalam mendukung proses klasifikasi dengan menyeimbangkan distribusi kelas pada dataset, sehingga model mampu mengenali pola pada kelas minoritas yang sebelumnya kurang terwakili, memperkuat kestabilan dan keandalan hasil klasifikasi kondisi janin.

f. Visualisasi Hasil

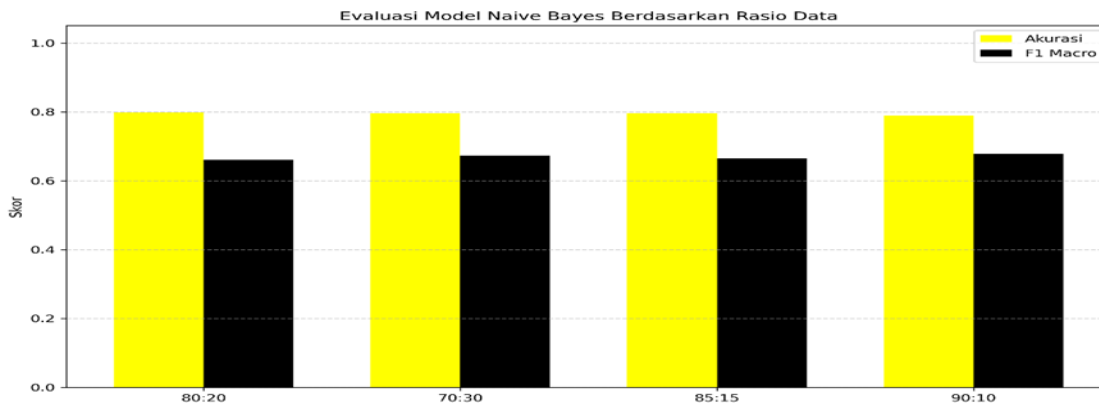
1. Confusion Matrix



Gambar 6. Confusion Matrix model Naive Bayes

Gambar 6 menunjukkan confusion matrix hasil klasifikasi menggunakan algoritma Naive Bayes pada empat rasio pembagian data latih dan data uji: 70:30, 80:20, 85:15, dan 90:10.

2. Gravitik Evaluasi



Gambar 7. Grafik Evaluasi Model

Gambar 7 memperlihatkan grafik nilai akurasi relatif yang stabil di sekitar 0.79–0.80 untuk semua rasio data. Namun, nilai F1 Macro cenderung sedikit lebih rendah, mencerminkan ketidakseimbangan performa antar kelas. Rasio data 70:30 menghasilkan skor F1 Macro tertinggi, menunjukkan bahwa model lebih seimbang dalam memprediksi semua kelas pada rasio tersebut. Grafik ini menegaskan bahwa meskipun akurasi tinggi, penting juga memperhatikan metrik lain seperti F1 Macro, terutama saat menangani dataset yang tidak seimbang.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi kondisi janin berdasarkan data KTG menggunakan algoritma Naive Bayes untuk menjaga efisiensi waktu komputasi, sehingga cocok untuk diterapkan dalam skenario dengan data besar yang dikombinasikan dengan teknik SMOTE untuk mengatasi ketidakseimbangan kelas pada dataset KTG, dengan cara menghasilkan data sintesis untuk kelas minoritas, sangat membantu meningkatkan sensitivitas dan recall pada kelas minoritas, serta memungkinkan model untuk belajar dengan lebih baik dari data minoritas. Penggunaan SMOTE meningkatkan keseimbangan performa model tanpa harus mengorbankan data mayoritas. Proses klasifikasi ini dilakukan secara terorganisir dan sistematis dengan menggunakan KDD yang menjamin setiap tahapan dilakukan dengan langkah yang tepat. Evaluasi pada empat rasio data latihan dan uji, yaitu 70:30, 80:20, 85:15, dan 90:10 menunjukkan bahwa rasio 80:20 menghasilkan akurasi tertinggi sebesar 79,81%, sementara rasio 90:10 memberikan nilai F1-Macro tertinggi sebesar 0,6788, yang mencerminkan performa lebih seimbang antar kelas, khususnya pada kelas minoritas. Penerapan SMOTE terbukti efektif dalam menyeimbangkan distribusi data dan meningkatkan sensitivitas model. Meskipun hasil klasifikasi cukup stabil dan akurat, model masih memiliki keterbatasan pada precision kelas minoritas. Oleh karena itu, penelitian selanjutnya disarankan untuk membandingkan algoritma lain seperti K-Nearest Neighbors (K-NN), Random Forest, atau Support Vector Machines (SVM), serta mengeksplorasi berbagai parameter SMOTE untuk lebih meningkatkan akurasi dan keseimbangan performa model dalam mengklasifikasikan kondisi janin secara lebih efektif.

REFERENCES

- [1] T. R. P. Lestari, "Achievement of Mother and Baby Health Status As One of the Successes of Mother and Child Health Programs," *Kajian*, vol. 25, no. 1, pp. 75–89, 2020, doi: 10.22212/kajian.v25i1.1889.
- [2] Sutrani Syarif, "Pemanfaatan Teknologi Tentang Menghitung Denyut Jantung Janin Di Desa Tanakaraeng Kabupaten Gowa," *J. Pelayanan dan Pengabd. Masy. Indones.*, vol. 2, no. 2, pp. 204–208, 2023, doi: 10.55606/jppmi.v1i2.617.
- [3] Y. A. Aji *et al.*, "Sistem Pengukuran Detak Jantung Janin Melalui Elektrokardiogram Abdominal dan Android," *Indones. J. Appl. Phys.*, vol. 12, no. 2, p. 279, 2022, doi: 10.13057/ijap.v12i2.65287.
- [4] E. Wati, S. A. Sari, and N. L. Fitri, "Penerapan Pendidikan Kesehatan tentang Tanda Bahaya Kehamilan untuk Meningkatkan Pengetahuan Ibu Hamil Primigravida Di Wilayah Kerja UPTD Puskesmas Purwosari Kec. Metro Utara," *J. Cendikia Muda*, vol. 3, no. 2, pp. 226–234, 2023.
- [5] S. Gholampour, "Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 2, pp. 827–841, 2024, doi: 10.3390/make6020039.
- [6] F. Francis, S. Luz, H. Wu, S. J. Stock, and R. Townsend, "Machine learning on cardiotocography data to classify fetal outcomes: A scoping review," *Comput. Biol. Med.*, vol. 172, no. February, p. 108220, 2024, doi: 10.1016/j.compbiomed.2024.108220.
- [7] Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang, and M. M. V. Chalapati, "Cardiotocography Data Analysis for Fetal Health Classification Using Machine Learning Models," *IEEE Access*, vol. 12, no. February, pp. 26005–26022, 2024, doi: 10.1109/ACCESS.2024.3364755.
- [8] I. Nazli, E. Korbeko, S. Dogru, E. Kugu, and O. K. Sahingoz, "Early Detection of Fetal Health Conditions Using Machine Learning for Classifying Imbalanced Cardiotocographic Data," *Diagnostics*, vol. 15, no. 10, pp. 1–26, 2025, doi: 10.3390/diagnostics15101250.
- [9] A. Kuzu and Y. Santur, "Early Diagnosis and Classification of Fetal Health Status from a Fetal Cardiotocography Dataset Using Ensemble Learning," *Diagnostics*, vol. 13, no. 15, pp. 1–15, 2023, doi: 10.3390/diagnostics13152471.
- [10] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-03430-5.



- [11] N. Sharfina and N. G. Ramadhan, "Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 33, 2023, doi: 10.31328/jointecs.v8i1.4456.
- [12] S. M. Habib, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, pp. 248–258, 2022, doi: 10.32672/jnkti.v5i2.4191.
- [13] N. Singhal and Himanshu, "A Review on Knowledge Discovery from Databases," *Lect. Notes Electr. Eng.*, vol. 860, no. January, pp. 457–464, 2022, doi: 10.1007/978-981-16-9488-2_43.
- [14] A. G. Shaaban, M. H. Khafagy, M. A. Elmasry, H. El-Beih, and M. H. Ibrahim, "Knowledge discovery in manufacturing datasets using data mining techniques to improve business performance," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 3, pp. 1736–1746, 2022, doi: 10.11591/ijeecs.v26.i3.pp1736-1746.
- [15] S. Bhatia and J. Malhotra, "Naïve bayes classifier for predicting the novel coronavirus," *Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mob. Networks, ICICV 2021*, no. Icicv, pp. 880–883, 2021, doi: 10.1109/ICICV50876.2021.9388410.
- [16] V. No, Z. Amri, M. Rodi, M. N. Wathani, A. Bagja, "Infotek : Jurnal Informatika dan Teknologi Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN," *J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 193–204, 2025, doi: 10.29408/jit.v8i1.27975 e-ISSN.
- [17] F. K. Nasser and S. F. Behadili, "A Review of Data Mining and Knowledge Discovery Approaches for Bioinformatics," *Iraqi J. Sci.*, vol. 63, no. 7, pp. 3169–3188, 2022, doi: 10.24996/ijcs.2022.63.7.37.
- [18] R. Kembang Hapsari and T. Surabaya, "Implementasi Algoritma SMOTE Sebagai Penyelesaian Imbalance Hight Dimensional Datasets," *Pros. Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, pp. 427–427, 2022, doi: 10.31284/p.snestik.2022.2868.
- [19] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.
- [20] A. Ilham *et al.*, "CFCM-SMOTE: A Robust Fetal Health Classification to Improve Precision Modeling in Multiclass Scenarios," *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, pp. 471–486, 2024, doi: 10.12785/ijcds/160137.
- [21] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naïve Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
- [22] M. V. Anand, B. Kiranbala, S. R. Srividhya, K. C., M. Younus, and M. H. Rahman, "Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer," *Mob. Inf. Syst.*, vol. 2022, p. 7, 2022, doi: 10.1155/2022/2436946.
- [23] D. Abdullah, K. Asmi, and I. G. A. K. Warmayana, *Perancangan dan Pembuatan Aplikasi File Server Berbasis Web Menggunakan Metode Interpolation Search*. SEFA Bumi Persada, 2020.