



# Klasifikasi Kondisi Janin Menggunakan Algoritma K-Nearest Neighbors dan Teknik SMOTE Berdasarkan Data Kardiotogram

Dede Fadillah, Elin Haerani\*, Fitri Wulandari, Fadhilah Syafria

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

Email: <sup>1</sup>11850112495@students.uin-suska.ac.id, <sup>2,\*</sup>elin.haerani@uin.suska.ac.id, <sup>3</sup>fitri.wulandari@uin-suska.ac.id,

<sup>4</sup>fadhilah.syafria@uin-suska.ac.id

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

**Abstrak**-Kesehatan janin merupakan aspek penting dalam upaya menurunkan angka kematian bayi, di mana kardiotokogram (KTG) digunakan untuk memantau kondisi janin melalui rekaman denyut jantung dan kontraksi rahim. Namun, interpretasi data KTG secara manual masih menghadapi tantangan, terutama karena distribusi kelas yang tidak seimbang. Penelitian ini bertujuan untuk mengembangkan model klasifikasi kondisi janin menggunakan algoritma K-Nearest Neighbors (K-NN) yang dikombinasikan dengan teknik Synthetic Minority Over-sampling Technique (SMOTE). Dataset yang digunakan berasal dari Kaggle, terdiri atas 2.126 data pemeriksaan KTG dengan tiga kelas: Normal, Suspect, dan Pathological. Data diolah melalui tahapan Knowledge Discovery in Databases (KDD) yang mencakup seleksi data, pembersihan data, normalisasi, pembagian data, penyeimbangan dengan SMOTE, dan klasifikasi menggunakan K-NN. Model diuji pada empat rasio pembagian data latih dan uji (70:30, 80:20, 85:15, dan 90:10) menggunakan metrik akurasi dan F1-score macro. Hasil menunjukkan bahwa rasio 85:15 menghasilkan akurasi tertinggi sebesar 89,7%, sedangkan rasio 90:10 menghasilkan F1-score macro tertinggi sebesar 0,83. Temuan ini menunjukkan bahwa rasio 85:15 memberikan keseimbangan optimal antara kemampuan pembelajaran dan evaluasi model, sementara F1-score tertinggi pada rasio 90:10 mencerminkan sensitivitas model yang lebih tinggi terhadap kelas minoritas. Kombinasi K-NN dan SMOTE terbukti efektif dalam mengatasi ketidakseimbangan data mendukung kestabilan model dalam proses klasifikasi kondisi janin secara keseluruhan.

**Kata kunci:** Kardiotogram; K-Nearest Neighbors; SMOTE; Klasifikasi Janin; Ketidakseimbangan Data

**Abstract**-Fetal health is a crucial aspect in reducing infant mortality rates, where cardiotocography (CTG) is used to monitor fetal condition through recordings of fetal heart rate and uterine contractions. However, manual interpretation of CTG data still faces challenges, particularly due to imbalanced class distribution. This study aims to develop a classification model for fetal conditions using the K-Nearest Neighbors (K-NN) algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE). The dataset used, sourced from Kaggle, consists of 2,126 CTG examinations categorized into three classes: Normal, Suspect, and Pathological. The data processing follows the Knowledge Discovery in Databases (KDD) process, including data selection, cleaning, normalization, splitting, balancing with SMOTE, and classification using K-NN. The model was evaluated using four training-testing split ratios (70:30, 80:20, 85:15, and 90:10) with accuracy and macro F1-score as metrics. The results indicate that the 85:15 split ratio achieved the highest accuracy of 89.7%, while the 90:10 ratio yielded the highest macro F1-score of 0.83. These findings suggest that the 85:15 ratio offers an optimal balance between model training and evaluation, whereas the highest F1-score at 90:10 reflects greater model sensitivity to minority classes. The combination of K-NN and SMOTE proved effective in addressing data imbalance and supports model stability in the overall classification process of fetal conditions.

**Keywords:** Cardiotocography; K-Nearest Neighbors; SMOTE; Fetal Classification; Imbalanced Data

## 1. PENDAHULUAN

Kematian janin masih menjadi isu penting dalam kesehatan ibu dan anak. Berdasarkan data WHO, pada tahun 2015 jumlah kematian janin secara global diperkirakan mencapai 3,82 hingga 22,14 juta jiwa. Di Indonesia, angka kematian bayi mencapai 32 per 1.000 kelahiran hidup (25,2%), lebih tinggi dibanding negara tetangga seperti Singapura, Thailand, dan Malaysia [1]. Salah satu indikator utama kesehatan janin adalah denyut jantungnya. Detak jantung janin normal berkisar antara 120–160 denyut per menit, dan kelainan seperti takikardia atau bradikardia dapat mengindikasikan risiko [2]. Salah satu sasaran global yang tercantum dalam Tujuan Pembangunan Berkelanjutan (SDGs) adalah mengurangi tingkat kematian ibu dan janin [3]. Oleh karena itu, pemantauan denyut jantung janin sangat penting dan umumnya dilakukan dengan teknik kardiotografi.

Kardiotogram (KTG) adalah rekam medis detak jantung dan kontraksi rahim [4] yang berguna untuk mendeteksi gangguan suplai oksigen pada janin [5] yang dihasilkan dari teknik kardiotografi. Namun, analisis data KTG masih menghadapi suatu masalah, seperti beban kerja tenaga medis yang tinggi dan potensi kesalahan dalam interpretasi manual, terutama pada data yang kompleks dan tidak seimbang antar kelas. Tantangan utama dalam dataset KTG adalah ketidakseimbangan distribusi kelas, di mana kelas "Normal" jauh lebih dominan dibandingkan dengan kelas "Suspect" dan "Pathological".

Berbagai penelitian telah dilakukan terhadap data KTG seperti pada penelitian [6] dan [7] yang berfokus untuk meningkatkan akurasi melalui pemilihan nilai  $k$  optimal pada algoritma K-NN, namun belum banyak yang memperhatikan dampak ketidakseimbangan kelas secara sistematis dalam klasifikasi kondisi janin. Studi seperti [8] dan [9] menunjukkan pentingnya penanganan imbalance, namun belum menggabungkan pendekatan tersebut secara terstruktur dalam konteks pemrosesan data KTG. Penelitian ini mengkombinasikan metode *lazy learning* seperti K-Nearest Neighbors (K-NN) dengan teknik penyeimbangan seperti Synthetic Minority Over-sampling Technique (SMOTE) secara sistematis.

Algoritma K-NN telah terbukti efektif dalam klasifikasi multi-kelas di bidang medis [10], termasuk dalam prediksi kondisi janin, karena kemampuannya mengenali pola lokal dan tidak memerlukan asumsi distribusi data tertentu serta



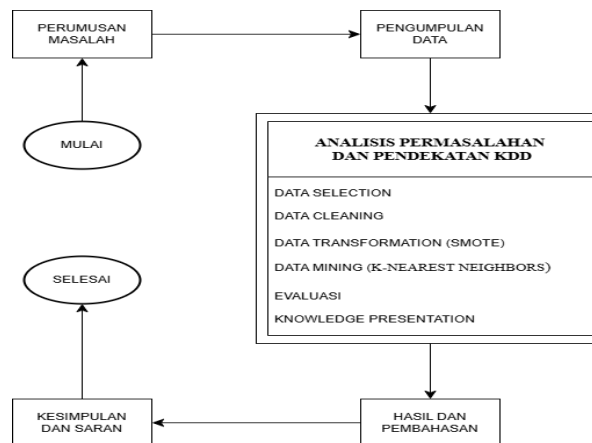
memiliki karakteristik yang sederhana, tidak memerlukan proses pelatihan model secara eksplisit (*lazy learning*), serta efektif dalam menangani data numerik dan klasifikasi multi-kelas seperti pada dataset Beberapa studi terdahulu menunjukkan bahwa K-NN mampu memberikan performa yang tinggi dalam klasifikasi data KTG [11]. Untuk mengatasi tantangan utama terhadap ketidakseimbangan distribusi kelas, penelitian ini memanfaatkan teknik SMOTE, yang bekerja dengan menambahkan data sintesis pada kelas minoritas tanpa melakukan duplikasi, sehingga model dapat belajar secara lebih seimbang [12]. Pendekatan data mining digunakan untuk membantu menganalisis data secara otomatis dan efisien [13] yang diterapkan dalam konteks proses *Knowledge Discovery in Databases* (KDD) [14]. Dengan menerapkan KDD, proses pembangunan model tidak hanya terfokus pada pencapaian akurasi akhir, tetapi juga memastikan kualitas data, strategi pemodelan, dan pemahaman menyeluruh terhadap pola klasifikasi yang ditemukan [15].

Penelitian ini mengusulkan kombinasi algoritma K-NN dengan teknik SMOTE untuk klasifikasi data KTG yang tidak seimbang agar dapat memberikan dasar pengembangan sistem klasifikasi otomatis yang potensial untuk diintegrasikan dalam sistem pendukung keputusan medis, sehingga dapat meringankan beban analisis manual, serta meningkatkan kemampuan deteksi dini terhadap janin yang berpotensi mengalami risiko tinggi, sehingga pengambilan keputusan medis dapat dilakukan lebih cepat dan tepat.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan terlihat pada Gambar 1.



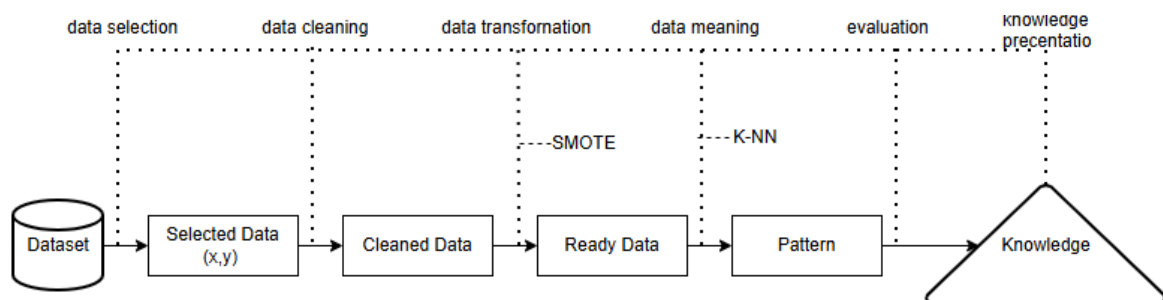
Gambar 1. Tahapan Penelitian

Langkah-langkah penting dalam tahapan penelitian untuk mencapai tujuan penelitian.

- Perumusan Masalah: Diawali dengan merumuskan permasalahan yang akan diteliti.
- Pengumpulan Data: Memperoleh data yang mendukung proses penelitian.
- Pendekatan KDD: Memastikan proses klasifikasi berlangsung secara sistematis dan terstruktur.
- Hasil dan Pembahasan: Menjelaskan hasil berdasarkan proses klasifikasi sesuai pendekatan KDD.
- Kesimpulan dan Saran: menyimpulkan hasil klasifikasi dan saran penelitian selanjutnya.

### 2.2 Pendekatan Knowledge Discovery in Databases (KDD)

Langkah-langkah dalam proses pendekatan KDD.



Gambar 2. Pendekatan KDD

Gambar 2 menunjukkan tahapan proses KDD (Knowledge Discovery in Databases) secara berurutan dari dataset mentah hingga menjadi pengetahuan[16]. Berikut penjelasan singkat tiap tahapnya:

- Data Selection

Memilih data yang relevan dari dataset besar, biasanya berupa fitur (x) dan label (y) sesuai tujuan analisis.



- b. Data Cleaning  
Membersihkan data dari kesalahan atau nilai kosong agar data konsisten dan dapat dianalisis.
- c. Data Transformation  
Mengubah data ke dalam format atau skala yang sesuai untuk pemodelan, seperti normalisasi (*min-max scaling*), pembagian data dan penyeimbangan data (SMOTE).
- d. Data Meaning (Data Mining)  
Proses inti untuk menemukan pola atau model dari data yang telah disiapkan [17], seperti klasifikasi dengan Algoritma K-NN
- e. Evaluation  
Mengevaluasi model atau pola yang ditemukan apakah valid dan bermanfaat berdasarkan metrik tertentu.
- f. Knowledge Presentation  
Menyajikan hasil akhir dalam bentuk yang mudah dipahami, seperti grafik dan laporan.

### 2.3 Teknik Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE adalah teknik oversampling yang bertujuan untuk mengatasi masalah ketidakseimbangan kelas (class imbalance) pada dataset pelatihan, khususnya dalam konteks klasifikasi. Masalah ini terjadi ketika jumlah sampel dari satu kelas (biasanya kelas minoritas) jauh lebih sedikit dibandingkan kelas lainnya, yang dapat menyebabkan model cenderung bias terhadap kelas mayoritas [18].

Berbeda dengan pendekatan oversampling konvensional yang hanya menduplikasi data minoritas, SMOTE menghasilkan sampel sintetis baru dengan cara melakukan interpolasi antara sampel minoritas yang ada dan tetangga terdekatnya dalam ruang fitur [19].

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

- a. Untuk setiap sampel minoritas  $x_i$ , algoritma akan memilih  $k$  tetangga terdekatnya dari kelas yang sama.
- b. Salah satu tetangga  $x_{zi}$  dipilih secara acak dari  $k$  tetangga tersebut.
- c. Sampel baru  $x_{new}$  kemudian dibentuk berdasarkan interpolasi linier antara  $x_i$  dan  $x_{zi}$  menggunakan persamaan berikut:

$$x_{baru} = x_i + \delta X (x_{zi} - x_i) \quad (1)$$

Interpolasi dalam proses SMOTE dilakukan dengan menggunakan sampel minoritas asli yang dilambangkan sebagai  $x_i$  dan salah satu tetangganya yang dipilih, yaitu  $x_{zi}$ . Bilangan acak  $\delta \in [0,1]$  digunakan untuk menentukan posisi interpolasi antara kedua titik tersebut. Hasil dari proses ini adalah titik baru  $x_{baru}$  yang terletak di antara  $x_i$  dan  $x_{zi}$ , sehingga tetap mempertahankan karakteristik lokal dari data minoritas tanpa membuat duplikasi langsung.

### 2.4 Algoritma K-Nearest Neighbors (K-NN)

K-NN adalah metode *lazy learning* yang tidak memerlukan proses pelatihan model secara eksplisit dan efektif dalam menangani data numerik serta klasifikasi multi-kelas seperti yang terdapat pada dataset KTG, melainkan menyimpan seluruh data pelatihan dan melakukan prediksi berdasarkan kedekatan data uji dengan data pelatihan yang ada. K-NN bekerja dengan mencari sejumlah  $K$  tetangga terdekat dari suatu data uji, lalu menentukan kelas berdasarkan mayoritas label dari tetangga-tetangga tersebut [20].

Adapun langkah-langkah algoritma K-NN adalah sebagai berikut:

- a. Menentukan jumlah  $K$  (jumlah tetangga terdekat yang digunakan).
- b. Menghitung jarak antara data uji dengan seluruh data pelatihan.
- c. Mengurutkan data pelatihan berdasarkan jarak terdekat ke data uji.
- d. Mengambil  $K$  tetangga terdekat.
- e. Menentukan kelas berdasarkan mayoritas label dari  $K$  tetangga tersebut.

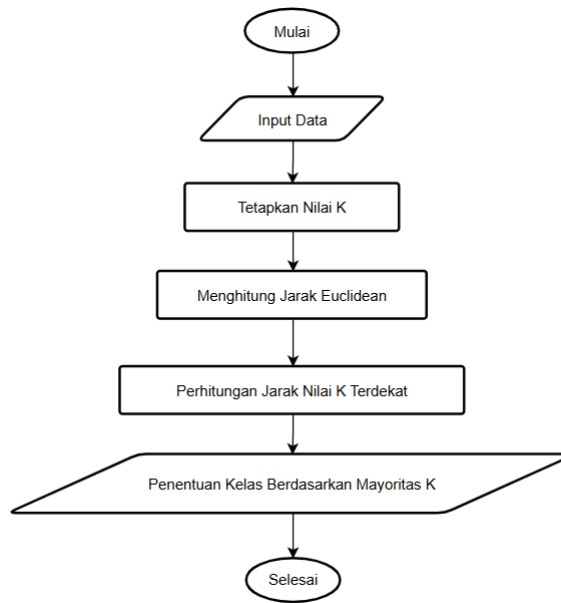
Salah satu metode yang umum digunakan dalam menghitung jarak antara data uji dan data pelatihan adalah Euclidean Distance. Metode ini mengukur jarak lurus dalam ruang multidimensi, dengan rumus sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Rumus Euclidean Distance digunakan untuk menghitung jarak antara dua titik dalam ruang multidimensi. Dalam konteks ini,  $d(x,y)$  merepresentasikan jarak antara data uji  $x$  dan data latih  $y$ . Setiap  $x_i$  dan  $y_i$  adalah nilai dari fitur ke- $i$  pada data uji dan data latih secara berturut-turut. Simbol  $\sum_{i=1}^n$ ,  $i=1, n$  menunjukkan proses penjumlahan dari dimensi pertama hingga ke- $n$ , di mana  $n$  adalah jumlah total fitur dalam dataset. Selisih antara nilai fitur data uji dan data latih pada setiap dimensi dihitung dengan  $(x_i - y_i)^2$ , kemudian semua hasilnya di hitung dengan akar kuadrat, sehingga menghasilkan nilai jarak Euclidean. Metode ini umum digunakan karena mampu mengukur kedekatan antara dua sampel secara akurat dalam ruang fitur yang memiliki banyak dimensi.

### 2.5 Flowchart K-NN

*Flowchart* menggambarkan urutan langkah dalam sebuah proses atau algoritma secara visual. Flowchart digunakan oleh analis untuk menjelaskan logika sistem dan mengantisipasi kendala teknis. Simbol-simbol dalam diagram menunjukkan proses, dan dihubungkan dengan garis alur. Gambar 3 menyajikan flowchart algoritma K-NN [21].



Gambar 3. Flowchart K-NN

Pada Gambar 3 Flowchart K-NN, menggambarkan proses kerja algoritma K-NN. Proses dimulai dengan memasukkan data yang akan diklasifikasikan, kemudian menetapkan nilai K sebagai jumlah tetangga terdekat yang akan diperhitungkan. Selanjutnya, algoritma menghitung jarak Euclidean antara data uji dan seluruh data pelatihan. Setelah itu, dipilih K data dengan jarak terdekat. Langkah terakhir adalah menentukan kelas dari data uji berdasarkan mayoritas kelas dari K tetangga terdekat tersebut. Proses ini berakhir setelah kelas ditentukan.

### 3. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle dengan nama "Cardiotocography (CTG)" (<https://www.kaggle.com>) dan tersedia dalam format .csv. Dataset ini terdiri dari 2.126 data dan mencakup 22 atribut, yaitu 21 fitur numerik yang menggambarkan parameter fisiologis janin, seperti denyut jantung janin dan kontraksi uterus, serta 1 atribut target yang berfungsi untuk mengklasifikasikan kondisi janin ke dalam tiga kategori: Normal, Suspect, dan Pathological. Data CTG dipersiapkan untuk digunakan pada pengembangan model klasifikasi kondisi janin berbasis algoritma K-NN.

Berikut ini adalah penjabaran hasil penelitian berdasarkan pendekatan KDD:

a. Data Selection

Label fetal\_health sebagai variabel target (y) dan 21 atribut sebagai fitur (X) dalam proses pemodelan.

b. Data Cleaning

1. Cek Missing Value: Diperoleh hasil bahwa tidak terdapat missing.
2. Deteksi Outlier (z-score): Diperoleh hasil data yang berada di luar rentang normal (*outlier*). Namun, mengingat data ini berasal dari bidang medis, outlier tersebut berpotensi membawa informasi klinis penting. Oleh karena itu, penanganan yang diambil bukan dengan menghapus data, melainkan dengan melakukan normalisasi.

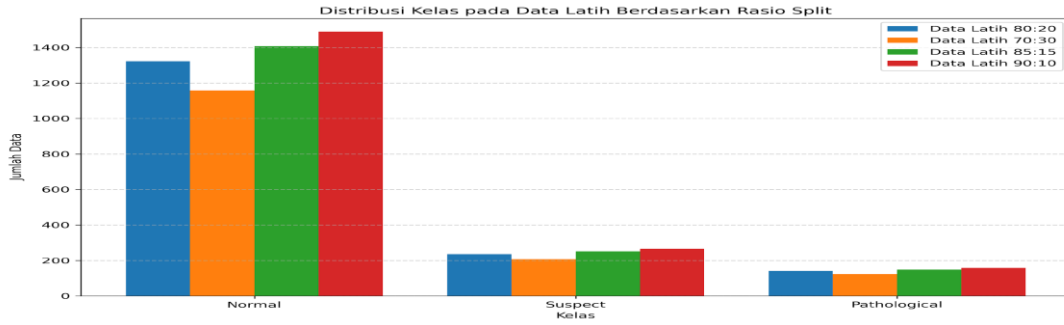
c. Data Transformation

1. Normalisasi (min-max scaling): Diperoleh hasil data dengan seluruh nilai fitur kini berada dalam rentang 0 hingga 1. Data yang telah dinormalisasi kemudian siap digunakan dalam proses selanjutnya.
2. Pembagian Data: Diperoleh hasil pembagian data sebagai berikut:

Tabel 1. Distribusi Data Setelah Dibagi Menjadi 4 Percobaan

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

Tabel 1 menunjukkan hasil dari pembagian data, dengan distribusi kelas sebagai berikut:



Gambar 4. Grafik distribusi kelas berdasarkan rasio percobaan

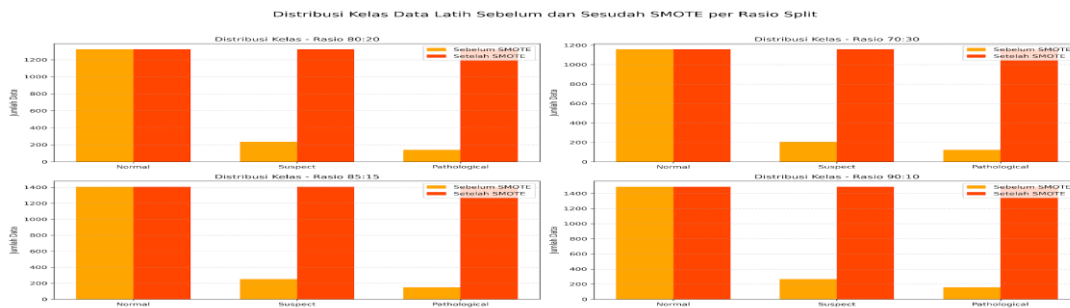
Gambar 4 menunjukkan adanya ketidakseimbangan terhadap distribusi kelas (*class imbalance*).

3. Penyeimbangan Data (SMOTE): Diperoleh distribusi data sebagai berikut:

Tabel 2. Distribusi Kelas Setelah Diseimbangkan oleh Teknik 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

Visualisasi distribusi kelas sebelum dan sesudah *SMOTE* dalam bentuk grafik berdasarkan Tabel 2 akan ditampilkan pada Gambar 5.



Gambar 5. Distribusi kelas setelah SMOTE

Distribusi kelas pada data latih menjadi lebih seimbang (*balance*) seperti yang terlihat pada Gambar 5.

d. Data Mining

Penerapan algoritma K-NN berdasarkan *flowchart*:

1. Mulai

Seluruh komponen teknis sudah disiapkan dan data sudah melewati tahap pra-pemrosesan

2. Input Data

Data latih dan data uji yang sudah melewati pra-pemrosesan di input

3. Tetapkan Nilai K terbaik

Nilai K dengan angka ganjil dengan rentang 5-15 diuji menggunakan GridSearchCV dengan validasi silang sebanyak lima lipatan (5-fold cross-validation) mendapatkan hasil:

Tabel 3. Rentang K=5-15 angka ganjil

K	Bobot	F1 Macro	Akurasi
5	distance	0.9619	0.9620
7	distance	0.9586	0.9587
9	distance	0.9541	0.9541
11	distance	0.9511	0.9511
13	distance	0.9488	0.9489
15	distance	0.9465	0.9466

Berdasarkan Tabel 3 dapat disimpulkan bahwa k terbaik adalah k = 5, distance memberikan F1 macro = 0.9619 dan Akurasi = 0.9620.

4. Menghitung Jarak Euclidean

Setiap data uji yang jaraknya sudah dihitung ke seluruh data latih menggunakan Euclidean distance.



5. Perhitungan Jarak Nilai K Terdekat  
Hasil perhitungan jarak diurutkan dari yang terkecil, lalu ambil K=5 atau 5 data latih dengan jarak terdekat
  6. Penentuan Kelas Berdasarkan Mayoritas K  
Lakukan voting mayoritas terhadap kelas dari kelima data terdekat. Hasil voting menunjukkan bahwa sebagian besar tetangga memiliki label kelas yang sama
  7. Selesai  
Output dari proses klasifikasi adalah label prediksi untuk setiap data uji
- e. Evaluasi  
Pada tahap evaluasi menampilkan hasil prediksi dengan confusion matrix sebagai berikut:

**Tabel 4.** Confusion Matrix Model K-NN pada skenario proporsi 70:30, 80:20, 85:15, dan 90:10

Uji Coba	Proporsi	Actual \ Pred	Normal	Suspect	Pathological
Uji 1	70:30	Normal	455	39	2
		Suspect	12	70	6
		Pathological	3	7	43
Uji 2	80:20	Normal	303	26	3
		Suspect	10	45	4
		Pathological	1	5	29
Uji 3	85:15	Normal	229	18	1
		Suspect	6	34	4
		Pathological	1	3	23
Uji 4	90:10	Normal	145	17	1
		Suspect	2	25	2
		Pathological	0	2	16

Tabel 4 menunjukkan confusion matrix dari empat percobaan dengan proporsi data pelatihan dan pengujian berbeda. Semakin besar proporsi data pelatihan, jumlah prediksi yang benar meningkat, terutama pada kelas Normal, sementara kesalahan klasifikasi antar kelas menurun, khususnya pada kelas Suspect dan Pathological.

**Tabel 5.** Evaluasi Performa Model

Rasio	Kelas	Precision	Recall	F1-Score	Support
70:30	Normal (1.0)	0.97	0.92	0.94	497
	Suspect (2.0)	0.60	0.80	0.69	88
	Pathological (3.0)	0.83	0.81	0.82	53
	Macro Avg	0.80	0.84	0.82	638
	Weighted Avg	0.91	0.89	0.90	638
	Accuracy			0.890	
80:20	Normal (1.0)	0.96	0.91	0.94	332
	Suspect (2.0)	0.59	0.76	0.67	59
	Pathological (3.0)	0.81	0.83	0.82	35
	Macro Avg	0.79	0.83	0.81	426
	Weighted Avg	0.90	0.88	0.89	426
	Accuracy			0.885	
85:15	Normal (1.0)	0.97	0.92	0.95	248
	Suspect (2.0)	0.62	0.77	0.69	44
	Pathological (3.0)	0.82	0.85	0.84	27
	Macro Avg	0.80	0.85	0.82	319
	Weighted Avg	0.91	0.90	0.90	319
	Accuracy			0.897	
90:10	Normal (1.0)	0.99	0.89	0.94	166
	Suspect (2.0)	0.57	0.86	0.68	29
	Pathological (3.0)	0.84	0.89	0.86	18
	Macro Avg	0.80	0.88	0.83	213
	Weighted Avg	0.92	0.89	0.90	213
	Accuracy			0.887	

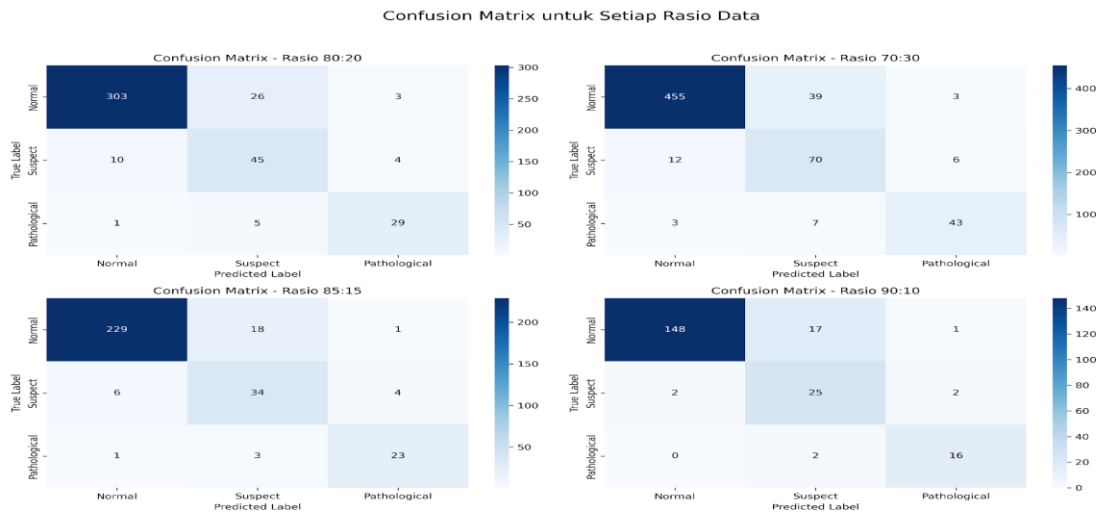
Berdasarkan Tabel 5, model K-NN menunjukkan performa klasifikasi yang cukup baik. Berdasarkan hasil evaluasi dari beberapa kali uji coba dengan rasio (70:30, 80:20, 85:15, 90:10), uji coba ke-3 dengan rasio 85:15 memberikan nilai akurasi tertinggi yaitu 89.7% dengan F1-macro sebesar 0.82 menunjukkan performa yang paling optimal. Pada uji coba ke-3 ini, model K-NN mampu memberikan keseimbangan terbaik antara kemampuan klasifikasi dan generalisasi, memberikan representasi fitur yang cukup pada data pelatihan sekaligus menyediakan data uji yang

memadai untuk mengukur kinerja model secara akurat. Oleh karena itu, uji coba ke-3 dapat dijadikan sebagai konfigurasi yang direkomendasikan dalam proses pengembangan model klasifikasi kondisi janin.

Sebagai bagian dari strategi pengembangan model, penerapan SMOTE dalam penelitian ini berperan penting dalam menangani ketidakseimbangan kelas pada dataset kondisi janin, memberikan peluang bagi model untuk mengenali pola pada kelas minoritas yang sebelumnya kurang terwakili, sehingga model dapat bekerja dengan basis data yang lebih lengkap dan beragam. Dengan demikian, SMOTE mendukung kestabilan model dalam proses klasifikasi kondisi janin.

## f. Knowledge Presentation

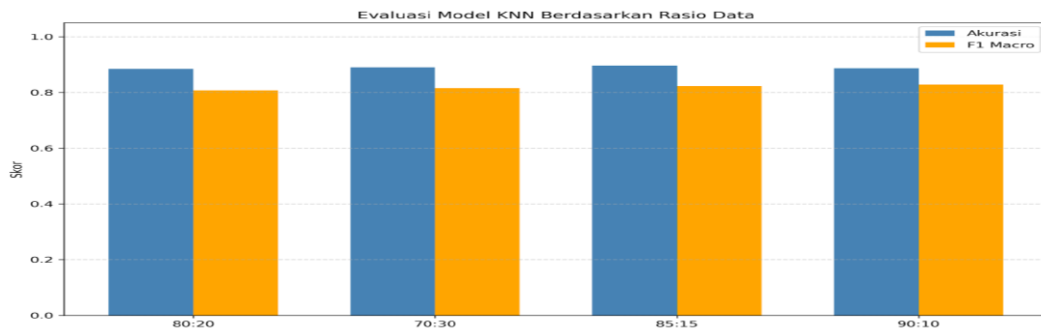
### 1. Confusion Matrix



**Gambar 6.** Confusion Matrix model K-NN dari keempat percobaan (70:30, 80:20, 85:15, dan 90:10)

Gambar 6 menunjukkan confusion matrix dari keempat percobaan klasifikasi menggunakan model K-NN dengan 4 kali uji coba. Setiap confusion matrix menggambarkan performa model dalam memprediksi tiga kelas kondisi janin: Normal, Suspect, dan Pathological.

### 2. Gravig Evaluasi



**Gambar 7.** Grafik Evaluasi Model

Gambar 7, grafik evaluasi model KNN pada berbagai rasio pembagian data menunjukkan performa yang stabil dengan akurasi tertinggi pada rasio 85:15, yang menawarkan keseimbangan optimal antara data latih dan uji. Rasio ini memungkinkan model mengenali pola ketiga kelas dengan lebih baik, terutama pada dataset Kardiotogram yang memiliki ketidakseimbangan kelas. Meskipun rasio 90:10 menghasilkan F1-score macro tertinggi, akurasinya sedikit lebih rendah karena data uji yang kecil. Rasio 70:30 memberikan evaluasi lebih kuat karena data uji lebih besar, sementara 80:20 menunjukkan performa terendah akibat data latih dan uji yang kurang ideal. Secara keseluruhan, rasio 85:15 adalah pilihan terbaik untuk performa dan generalisasi model KNN dalam mengklasifikasikan kondisi janin.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi kondisi janin berdasarkan data KTG dengan menggunakan algoritma K-NN yang efektif dalam menangani data numerik dan klasifikasi multi-kelas, yang dikombinasikan dengan teknik SMOTE untuk mengatasi ketidakseimbangan kelas serta berdampak baik terhadap sensitivitas dan recall kelas yang di seimbangkan tanpa berdampak buruk terhadap kelas mayoritas. Tahapan klasifikasi dilaksanakan dengan



pendekatan KDD agar lebih sistematis dan setiap langkah ditangani dengan tepat sehingga lebih terorganisir. Evaluasi model pada berbagai rasio 70:30, 80:20, 85:15, dan 90:10, pembagian data latih dan uji menunjukkan akurasi tertinggi sebesar 89,7% dan F1-macro 0,82 pada rasio 85:15, yang menandakan keseimbangan optimal antara proses pelatihan dan evaluasi. Penerapan SMOTE berperan penting dalam memberikan representasi data yang lebih seimbang, khususnya pada kelas minoritas seperti Suspect dan Pathological, sehingga mendukung kestabilan model dalam proses klasifikasi. Meskipun model ini mampu memberikan prediksi yang akurat dan konsisten, penelitian ini masih memiliki keterbatasan terkait perbandingan dengan algoritma lain serta eksplorasi parameter SMOTE yang lebih mendalam. Oleh karena itu, penelitian selanjutnya disarankan untuk membandingkan performa K-NN dengan algoritma lain seperti Random Forest, Support Vector Machines (SVM), atau XGBoost, serta mengembangkan integrasi dengan antarmuka visualisasi dan sistem pendukung keputusan klinis guna meningkatkan akurasi dan aplikasi praktis model dalam diagnosis kondisi janin.

## REFERENCES

- [1] I. Sulihati, A. Syukur, and A. Marjuni, "Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1658–1664, 2022, doi: 10.47065/bits.v4i3.2672.
- [2] M. Minarti and R. Risnawati, "Posisi Ibu Hamil Memengaruhi Akurasi Pengukuran Kesejahteraan Janin," *J. Bidan Cerdas*, vol. 2, no. 3, pp. 170–176, 2020, doi: 10.33860/jbc.v2i3.93.
- [3] Meti Patimah, "Pendidikan Kesehatan Ibu Hamil Tentang Ketidaknyamanan Pada Kehamilan Trimester I dan Penatalaksanaannya," *Din. J. Pengabd. Kpd. Masy.*, vol. 4, no. 3, pp. 570–578, 2020, doi: 10.31849/dinamisia.v4i3.3790.
- [4] I. F. Nurahmadan, A. Agusta, P. A. Winarno, B. H. Sazali, Y. Thurfah, and A. Rosaliah, "Perbandingan Algoritma Machine Learning Untuk Klasifikasi Denyut Jantung Janin," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 02, no. 1, pp. 733–740, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1411>
- [5] I. Harun *et al.*, "Mendorong Partisipasi Perempuan Dalam Perencanaan Ekonomi Pembangunan Kampung Takengon Barat Kec. Lut Tawar Encouraging," *J. Pengabd. Pada Masy. Indones.*, vol. 2, no. 3, pp. 186–194, 2023, doi: 10.55606/jppmi.v2i3.603.
- [6] M. Mawaddah, A. Homaidi, and L. F. Lidimillah, "Implementation of the K-Nearest Neighbors Method for Determining Fetal Health Status Implementasi Metode K-Nearest Neighbors Untuk Penentuan," *J. Tek. Inform.*, vol. 5, no. 4, pp. 329–336, 2024, doi: 10.52436/1.jutif.2024.5.4.2173.
- [7] A. K. Wardhani *et al.*, "Optimasi Nilai K Pada Algoritma K-Nearest," *J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 44–51, 2025, doi: 10.54840/jstech.v5i1.360.
- [8] 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.
- [9] A. Surya Firmansyah, A. Aziz, and M. Ahsan, "Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3341–3347, 2024, doi: 10.36040/jati.v7i6.7257.
- [10] R. T. Prasetyo, "Seleksi Fitur dan Optimasi Parameter K-NN Berbasis Algoritma Genetika Pada Dataset Medis," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 213–221, 2020, doi: 10.51977/jti.v2i2.319.
- [11] A. D. Monica and S. Sulastri, "Klasifikasi Hasil Cardiotocography (CTG) Ibu Hamil untuk Memprediksi Kesehatan Janin," *J. Teknol. Sist. Inf. dan Apl.*, vol. 6, no. 3, pp. 443–457, 2023, doi: 10.32493/jtsi.v6i3.31548.
- [12] S. E. Situmeang and N. P. Savina, "Analisis Perbandingan Metode Decision Tree, Random Forest, dan Support Vector Machine (SVM) dalam Memprediksi Kesehatan Janin," *Dep. Stat. Inst. Teknol. Sepuluh Nop.*, pp. 1–8, 2024, doi: 10.12962/j27213862.vvix.xxxx.
- [13] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes," *J. SAINTEKOM*, vol. 13, no. 1, pp. 42–54, 2023, doi: 10.33020/saintekom.v13i1.352.
- [14] 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/ijs.2022.63.7.37.
- [15] 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.
- [16] D. Papakyriakou and I. S. Barbounakis, "Data Mining Methods: A Review," *Int. J. Comput. Appl.*, vol. 183, no. 48, pp. 5–19, 2022, doi: 10.5120/ijca2022921884.
- [17] L. Cao and C. Zhang, "Domain driven data mining," *Data Min. Knowl. Discov. Technol.*, vol. 7, no. 4, pp. 196–223, 2008, doi: 10.4018/978-1-59904-960-1.ch009.
- [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] S. W. Binabar and Ivandari, "Optimasi Parameter K pada Algoritma KNN untuk Deteksi Penyakit Kanker Payudara," *IC-Tech*, vol. XII, no. 2, pp. 11–18, 2017.
- [21] 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.