

---

# Analisis Perbandingan Algoritma KNN, SVM, dan Decision Tree dalam Memprediksi Kesehatan Janin

Mohamed<sup>1</sup>, Muhamad Zidan<sup>2</sup>, Muhamad Nopid Andriansyah<sup>3</sup>

<sup>1</sup>Department Sains and Teknologi, UIN Sunan Gunung Djati Bandung

<sup>2</sup>Department Sains and Teknologi, UIN Sunan Gunung Djati Bandung

<sup>3</sup>Department Sains and Teknologi, UIN Sunan Gunung Djati Bandung

---

## Article Info

### Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

---

### Keywords:

Decision Tree

Support Vector Machine

K-Nearest Neighbors

Kesehatan Janin

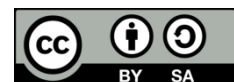
Kardiotokografi

---

## ABSTRACT

Angka kematian bayi yang baru lahir tetap menjadi isu yang sangat serius, khususnya di negara-negara yang sedang berkembang. Salah satu faktor utama yang berkontribusi adalah keterlambatan dalam mendeteksi keadaan kesehatan janin yang disebabkan oleh subjektivitas dalam analisis data kardiotokografi (CTG). Untuk menangani masalah ini, penelitian ini mengajukan metode otomatis yang berlandaskan machine learning untuk mengklasifikasikan status kesehatan janin ke dalam tiga kategori: normal, mencurigakan, dan patologis. Tiga algoritma pembelajaran terawasi yang diterapkan dalam penelitian ini meliputi K-Nearest Neighbors (KNN), Support Vector Machine (SVM), dan Decision Tree (DT). Penelitian ini mengikuti proses CRISP-DM mulai dari pemahaman data sampai evaluasi model. Dataset yang digunakan diambil dari CTU-UHB Intrapartum Cardiotocography Database, yang mencakup 2.126 sampel. Teknik augmentasi data SMOTE digunakan untuk mengatasi masalah ketidakseimbangan kelas. Evaluasi dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil dari eksperimen menunjukkan bahwa Decision Tree memberikan hasil terbaik dengan akurasi, presisi, recall, dan F1-score sebesar 91%, serta jumlah prediksi benar tertinggi (True Positive) yang mencapai 378 data. Temuan ini menunjukkan bahwa Decision Tree memiliki kemampuan lebih baik dalam menggeneralisasi data baru dan berpotensi untuk diterapkan dalam sistem pendukung keputusan klinis (CDSS).

*This is an open access article under the [CC BY-SA](#) license.*



---

## Corresponding Author:

Asep Ujang

Jurusan Teknik Informatika, UIN Sunan Gunung Djati Bandung

Email: asepujang@gmail.com

---

## 1. PENDAHULUAN

Kesehatan janin adalah indikator yang sangat penting dalam menilai kualitas pelayanan kesehatan ibu dan merupakan bagian integral dari usaha global untuk mengurangi angka kematian bayi baru lahir. Menurut laporan dari World Health Organization (WHO), pada tahun 2019 tercatat sekitar 2,4 juta kematian bayi baru lahir (dalam 28 hari pertama kehidupan), yang berkontribusi sebesar 47% dari total kematian anak di bawah usia lima tahun. Sekitar 6.700 kematian bayi baru lahir terjadi setiap harinya, dan sebagian besar disebabkan oleh komplikasi prematuritas, asfiksia saat lahir, atau infeksi [1]. Ini menegaskan pentingnya deteksi dini untuk menyelamatkan banyak nyawa bayi.

Situasi ini sangat relevan di negara berkembang, yang masih menghadapi berbagai tantangan dalam akses ke layanan kesehatan berkualitas dan kurangnya sistem pendukung keputusan klinis yang efektif [2]. Salah satu metode yang umum digunakan untuk memantau janin adalah Cardiotocography (CTG), metode ini menganalisis sinyal detak jantung janin dan kontraksi rahim untuk mengklasifikasikan janin ke dalam kategori normal, mencurigakan, atau patologis [3]. Namun, interpretasi CTG sangat tergantung pada penilaian klinis dan bisa berisiko subjektif serta kesalahan manusia, terutama dalam keadaan kritis [4], [5].

Dengan kemajuan teknologi, pendekatan kecerdasan buatan (AI) dan pembelajaran mesin (ML) muncul sebagai solusi yang menjanjikan untuk menganalisis data fisiologis CTG secara otomatis dan objektif. Berbagai studi menunjukkan bahwa algoritma ML seperti K-Nearest Neighbors (KNN), Support Vector Machine (SVM), dan Decision Tree (DT) dapat meningkatkan akurasi klasifikasi kondisi janin dibandingkan dengan metode tradisional [6], [8].

Setiap algoritma memiliki kelebihan dan kekurangannya. Meskipun KNN mudah diterapkan, algoritma ini sangat tergantung pada nilai  $k$  dan rentan terhadap outlier [9]. SVM efektif dalam menangani data dengan dimensi tinggi dengan memaksimalkan margin pemisah, namun memerlukan penyesuaian kernel yang sesuai [10]. Di sisi lain, DT unggul dalam interpretasi melalui struktur pohon if-then namun rentan mengalami *overfitting* jika tidak dilakukan pemangkasan yang tepat [11].

Penelitian sebelumnya, menunjukkan bahwa SVM dapat mencapai tingkat akurasi yang tinggi meskipun memerlukan penyetelan yang intensif [12]. Lalu membandingkan KNN, SVM, dan DT dan menemukan bahwa setiap algoritma memiliki keunggulan masing-masing pada berbagai metrik evaluasi [13]. Selain itu juga menunjukkan bahwa tahap pra-pemrosesan dan pemilihan fitur sangat penting untuk memastikan kestabilan model [14]. Pada pengembangan model *ensemble* yang menggabungkan berbagai algoritma untuk meningkatkan akurasi prediksi meskipun dengan mengorbankan kemudahan dalam interpretasi [15].

Tantangan utama dari penelitian ini adalah menemukan algoritma yang terbaik dengan mempertimbangkan keseimbangan antara akurasi, presisi, *recall*, *F1-score*, kemampuan interpretatif, dan efisiensi komputasi. Banyak penelitian sebelumnya menggunakan dataset dan teknik pra-pemrosesan yang berbeda, sehingga menyulitkan perbandingan yang langsung [14].

Dengan demikian, tujuan dari penelitian ini adalah untuk memberikan analisis komparatif yang komprehensif terhadap algoritma KNN, SVM, dan Decision Tree dalam memprediksi kondisi kesehatan janin berdasarkan data CTG. Evaluasi akan dilakukan secara konsisten menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Diharapkan dari penelitian ini dapat diperoleh rekomendasi mengenai algoritma terbaik untuk diterapkan dalam *Clinical Decision Support System* (CDSS) guna mendeteksi risiko kesehatan janin dengan cepat, objektif, dan praktis, terutama di fasilitas kesehatan dengan sumber daya yang terbatas.

### A. Penelitian Terdahulu

Untuk mendapatkan gambaran terkini dalam penelitian terkait, berikut disajikan tabel *state of the art* yang merangkum beberapa penelitian terdahulu. Tabel ini mencakup metode yang digunakan, inti pembahasan dari masing-masing penelitian, serta kelemahan yang masih terdapat pada pendekatan tersebut. Dengan demikian, dapat diidentifikasi celah penelitian (*research gap*) yang dapat menjadi landasan bagi penelitian selanjutnya.

Table 1. State of the art (SOTA)

Penulis	Metode	Inti Pembahasan	Kelemahan
Siswanto et al. [16]	K-Nearest Neighbors (KNN)	Mengklasifikasikan status janin dengan akurasi tinggi menggunakan data CTG dari UCI Repository. Kinerja model sangat ditentukan oleh pemilihan nilai parameter k dan proses preprocessing data.	Sangat dipengaruhi oleh nilai parameter k dan dapat terganggu oleh noise pada data.
Sharma dan Saini [11]	Support Vector Machine (SVM)	Efektif untuk klasifikasi tiga kelas kondisi janin. Pemilihan kernel, seperti radial basis function (RBF), secara signifikan mempengaruhi akurasi, terutama pada data berdimensi tinggi dan tidak linier.	Membutuhkan pemilihan kernel dan parameter yang tepat agar dapat bekerja secara optimal.
Mohapatra et al. [17]	Decision Tree (DT)	Menghasilkan model klasifikasi yang mudah dipahami oleh praktisi kesehatan karena strukturnya yang logis dan transparan.	Rentan terhadap overfitting, terutama jika tidak dilakukan teknik pruning yang tepat.
Singh dan Arora [12]	Perbandingan KNN, SVM, dan DT	Tidak ada algoritma yang secara konsisten unggul dalam semua metrik evaluasi. Pemilihan algoritma sebaiknya mempertimbangkan keseimbangan antara akurasi, interpretabilitas, dan efisiensi komputasi.	Studi ini menyoroti bahwa setiap algoritma memiliki keunggulan dan keterbatasan masing-masing, sehingga tidak ada satu solusi yang paling optimal untuk semua kasus.
Setiawan et al.[15]	Ensemble Learning	Mengintegrasikan beberapa algoritma (ensemble) dapat menghasilkan performa prediksi yang lebih stabil dan robust dibandingkan penggunaan algoritma tunggal.	Model ensemble sering kali mengorbankan interpretabilitas, yang merupakan aspek penting dalam lingkungan klinis.

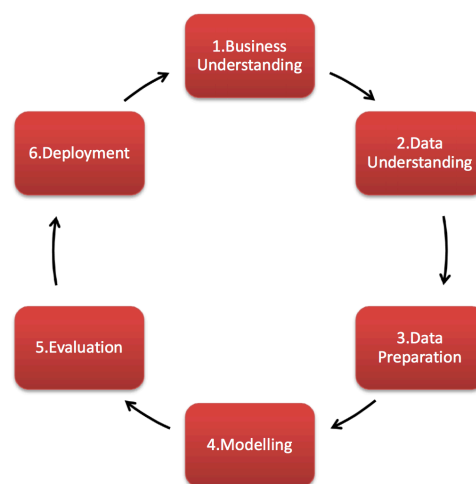
### A. Research Gap

Meskipun berbagai metode machine learning seperti K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), hingga pendekatan Ensemble Learning telah digunakan untuk klasifikasi status janin berdasarkan data CTG, masih terdapat beberapa celah penelitian yang dapat dieksplorasi lebih lanjut:

- a. **Keseimbangan antara Akurasi dan Interpretabilitas**  
Pendekatan seperti ensemble learning mampu meningkatkan akurasi dan stabilitas model, namun sering kali mengorbankan interpretabilitas yang sangat penting dalam konteks medis. Belum banyak studi yang berhasil menjembatani kedua aspek ini secara bersamaan.
- b. **Keterbatasan Adaptivitas Model terhadap Noise dan Ketidakseimbangan Data**  
Metode seperti KNN sangat rentan terhadap *noise* dan pemilihan parameter, sedangkan SVM dan DT membutuhkan proses tuning dan pruning yang kompleks. Gap ini menunjukkan perlunya metode yang lebih adaptif terhadap kualitas data dunia nyata yang sering tidak ideal.
- c. **Kurangnya Optimalisasi Multi-Objective**  
Penelitian sebelumnya cenderung fokus pada satu metrik (misalnya akurasi), sementara dalam aplikasi nyata dibutuhkan keseimbangan antara *sensitivity*, *specificity*, *precision*, dan *recall*. Belum ada pendekatan komprehensif yang mengoptimalkan berbagai metrik evaluasi secara bersamaan.

## 2. METODE

Penelitian ini mengadopsi desain eksperimental yang terstruktur menggunakan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM). Metodologi ini menyediakan pendekatan sistematis yang terbukti untuk memandu proyek penambangan data, memastikan bahwa setiap fase, mulai dari pemahaman masalah hingga penerapan solusi, ditangani secara komprehensif. Kerangka kerja CRISP-DM terdiri dari enam fase berulang: Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Penerapan [18].



Gambar 1: Kerangka kerja CRISP-DM

### 1. Business Understanding

Fase awal ini berfokus pada pemahaman tujuan proyek dari perspektif klinis dan menerjemahkannya ke dalam definisi masalah penambangan data. Tujuan utama dari penelitian ini adalah untuk mengembangkan dan mengevaluasi sebuah sistem pendukung keputusan klinis berbasis *machine learning* yang dapat membantu tenaga medis dalam mendiagnosis kesehatan janin secara lebih objektif dan konsisten. Latar belakang masalah ini berakar pada tantangan yang melekat dalam praktik klinis saat ini. Interpretasi data Kardiotokografi (CTG), yang merupakan metode standar untuk pemantauan janin, seringkali bersifat subjektif dan sangat bergantung pada pengalaman serta intuisi klinisi[19]. Variabilitas dalam interpretasi antar-klinisi dapat menyebabkan diagnosis yang tidak konsisten, yang berpotensi menunda intervensi yang diperlukan atau, sebaliknya, menyebabkan intervensi yang tidak perlu.

Kondisi ini diperparah oleh tingginya angka kematian neonatal global, terutama di negara-negara berkembang, di mana akses terhadap tenaga medis ahli mungkin terbatas. Laporan dari organisasi kesehatan dunia menunjukkan bahwa sebagian besar kematian bayi baru lahir disebabkan oleh keterlambatan atau

ketidaktepatan dalam mendeteksi gangguan kesehatan selama kehamilan. Oleh karena itu, masalah yang hendak dipecahkan adalah mengurangi subjektivitas dan potensi *human error* dalam analisis CTG melalui pengembangan sistem klasifikasi otomatis yang akurat, andal, dan efisien.

## 2. Data Understanding

Fase ini melibatkan pengumpulan data awal dan kegiatan untuk membiasakan diri dengan data, mengidentifikasi masalah kualitas data, menemukan wawasan pertama ke dalam data, atau mendeteksi subset yang menarik untuk membentuk hipotesis untuk informasi tersembunyi.

Penelitian ini menggunakan dataset publik yang berasal dari platform Kaggle. Dataset yang digunakan berjudul "Fetal Health Classification" dan diterbitkan oleh pengguna AndrewMVD. Dataset ini dipilih karena kelengkapan fitur dan relevansinya dengan tujuan penelitian untuk klasifikasi kesehatan janin.

Dataset ini memiliki total 2126 baris dan 22 fitur, yang seluruhnya berjenis data kuantitatif. Fitur-fitur tersebut terbagi menjadi 21 fitur independen dan 1 fitur dependen atau label. Fitur-fitur independen ini mencakup berbagai parameter statistik yang relevan secara klinis dari denyut jantung janin (*Fetal Heart Rate* - FHR) dan kontraksi rahim (*Uterine Contractions* - UC) yang diekstraksi dari sinyal CTG.

## 3. Data Preparation

Pada fase ini, data mentah yang telah dipahami dari fase sebelumnya diolah dan diubah menjadi format yang bersih, konsisten, dan optimal untuk proses pemodelan. Kualitas dan ketelitian dalam persiapan data secara langsung mempengaruhi kinerja, keandalan, dan keadilan dari model-model *machine learning* yang akan dibangun.

- A. **Pembersihan dan Pemisahan Variabel:** Langkah pertama adalah membersihkan data dengan menghapus baris duplikat untuk menjaga kualitas model. Sebanyak 13 data duplikat berhasil dihapus. Setelah itu, dataset dipisahkan menjadi variabel independen (X) dan variabel dependen (y). Variabel X berisi semua fitur prediktor, sedangkan variabel y berisi label target (*fetal\_health*).
- B. **Pembagian Dataset:** Untuk mengevaluasi kinerja generalisasi model secara objektif, dataset dibagi menjadi dua bagian yang independen: set pelatihan (*training set*) dan set pengujian (*testing set*). Pembagian dilakukan dengan perbandingan 80:20, di mana 80% data (1690 sampel) digunakan untuk data latih dan 20% sisanya (423 sampel) untuk data tes. Proses ini menggunakan fungsi *train\_test\_split* dari Scikit-learn dengan argumen *stratify=y* diaktifkan. Stratifikasi ini sangat penting untuk menjaga agar distribusi kelas 'Normal', 'Suspect', dan 'Pathological' tetap proporsional di kedua set, sehingga evaluasi model menjadi lebih andal dan tidak bias.
- C. **Augmentasi Data:** Data latih yang dihasilkan menunjukkan ketidakseimbangan kelas yang signifikan. Untuk mengatasi masalah ini, dilakukan augmentasi data pada data latih menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) [21]. SMOTE bekerja dengan menciptakan sampel sintesis baru untuk kelas-kelas minoritas ('Suspect' dan 'Pathological'), sehingga menyeimbangkan distribusi kelas dalam data latih. Proses ini membantu model agar tidak bias terhadap kelas mayoritas dan dapat mengenali pola dari kelas minoritas dengan lebih baik. Penting untuk dicatat bahwa augmentasi ini hanya diterapkan pada data latih untuk mencegah kebocoran data. Setelah proses augmentasi, jumlah sampel data latih bertambah menjadi 3948 sampel.
- D. **Normalisasi:** Langkah terakhir dalam persiapan data adalah normalisasi atau penskalaan fitur. Proses ini bertujuan untuk menyeragamkan rentang nilai dari fitur-fitur prediktor, yang sangat penting untuk algoritma yang sensitif terhadap jarak seperti KNN dan SVM [22], [23]. Penskalaan dilakukan menggunakan *StandardScaler* dari Scikit-learn. Objek *StandardScaler* di-fit hanya pada data latih (*X\_train*) untuk mempelajari parameter skala (rata-rata dan simpangan baku), kemudian digunakan untuk men-transform baik data latih maupun data tes (*X\_test*). Pendekatan ini memastikan bahwa informasi dari data tes tidak "bocor" ke dalam proses pelatihan dan mempermudah model dalam memprediksi data dengan menormalisasi distribusi nilai di setiap kolom.

## 4. Modeling

Fase ini merupakan inti dari eksperimen, di mana ketiga algoritma klasifikasi yang dipilih dibangun, dilatih secara sistematis, dan dioptimalkan untuk mencapai performa puncaknya. Proses ini dirancang untuk memastikan bahwa perbandingan akhir dilakukan antara versi terbaik dari setiap model

#### A. Pemilihan dan Implementasi Algoritma

Empat model klasifikasi *supervised learning* akan dilatih dan dievaluasi: K-Nearest Neighbors (KNN), Decision Tree, dan Support Vector Machine (SVM).

##### 1. K-Nearest Neighbors (KNN)

- Deskripsi: KNN adalah algoritma *instance-based learning* yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari 'k' tetangga terdekatnya. Jarak antar data dihitung menggunakan metrik seperti Jarak Euclidean.
- Rumus (Jarak Euclidean):

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2}$$

- Keterangan:
  - $d(x, x_i)$ : Jarak Euclidean antara titik uji dan data latih ke-  $x_i$
  - $x$ : Vektor fitur dari data uji
  - $x_i$ : Vektor fitur dari data latih ke-  $i$
  - $x_j$ : Nilai fitur ke-  $j$
  - $x_{ij}$ : Nilai fitur ke-  $j$  pada data latih ke-  $i$
  - $n$ : Jumlah fitur (dimensi data)
- Parameter: `n_neighbors=5`, `weights='uniform'`, `metric='minkowski'` dengan `p=2` (Euclidean).
- Kelebihan: Sederhana, mudah dipahami, dan tidak memerlukan fase pelatihan eksplisit (*lazy learning*).
- Kekurangan: Lambat pada dataset besar, sensitif terhadap *outlier* dan skala fitur (membutuhkan normalisasi).

##### 2. Decision Tree

- Deskripsi: Decision Tree adalah algoritma pembelajaran yang menggunakan struktur pohon untuk membuat keputusan berdasarkan fitur input. Setiap simpul (node) mewakili fitur, setiap cabang mewakili keputusan (rule), dan setiap daun (leaf) mewakili hasil akhir prediksi. Algoritma ini membagi data berdasarkan fitur yang memberikan informasi paling besar (misalnya, menggunakan *Gini Impurity* atau *Entropy* untuk klasifikasi).
- Rumus (Pemilihan Split terbaik):

$$Gain = Gini(parent) - \sum_j \frac{|D_j|}{|D|} Gini(D_j)$$

Misal dengan Gini Impurity:

$$Gini(D) = 1 - \sum_{i=1}^n p_i^2$$

- Parameter: `criterion='gini'` (atau `'entropy'` untuk pengukuran lain), `max_depth=3`, `min_samples_split=2`, `random_state=42`.
- Kelebihan: Mudah dipahami dan divisualisasikan, tidak memerlukan normalisasi data, dapat menangani data numerik dan kategorikal, dan cepat dalam proses pelatihan untuk dataset kecil hingga sedang.
- Kekurangan: Rentan terhadap overfitting jika tidak dilakukan pruning atau pembatasan kedalaman, tidak stabil terhadap perubahan kecil pada data (dapat menghasilkan pohon yang sangat berbeda), dan cenderung bias terhadap fitur dengan banyak kategori.

### 3. Support Vector Machine (SVM)

- Deskripsi: SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas dengan margin terbesar. Dengan menggunakan *kernel trick* (seperti RBF), SVM dapat menangani data non-linear secara efektif.
- Rumus (SVM Linear):

$$\frac{\min}{w, b} \frac{1}{2} ||w||^2 \text{ dengan syarat: } y_i(w \cdot x_i + b) \geq 1$$

- Keterangan:
  - $w$  : Vector Bobot
  - $b$  : Bias
  - $x_i$  : Data ke-  $i$
  - $y_i$  : Label data ke-  $i$
  - Margin =  $\frac{2}{||w||}$
- Parameter: C=1.0, kernel='rbf', gamma='scale'.
- Kelebihan: Akurat untuk data berdimensi tinggi dan efektif jika ada margin pemisahan yang jelas antar kelas.
- Kekurangan: Lambat untuk dataset besar, boros memori, dan sensitif terhadap pemilihan parameter kernel.

### 5. Evaluation

Pada fase ini, kinerja dari keempat model yang telah dilatih dievaluasi secara objektif menggunakan set pengujian. Untuk mempermudah dan menstandarisasi proses evaluasi, sebuah fungsi khusus dibuat untuk menghitung dan menampilkan metrik kinerja utama serta visualisasi matriks kebingungan.

Kinerja model diukur menggunakan empat metrik standar yang memberikan pandangan komprehensif, terutama pada dataset klasifikasi multi-kelas.

- Akurasi (Accuracy):** Proporsi total prediksi yang benar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Presisi (Precision):** Dari semua yang diprediksi positif, berapa persen yang benar-benar positif.

$$Precision = \frac{TP}{TP + FP}$$

- Recall (Sensitivity):** Dari semua yang sebenarnya positif, berapa persen yang berhasil diidentifikasi.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score:** Rata-rata harmonik dari Presisi dan Recall, menyeimbangkan kedua metrik.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Di mana TP (True Positive) adalah prediksi benar positif, TN (True Negative) adalah prediksi benar negatif, FP (False Positive) adalah prediksi salah positif, dan FN (False Negative) adalah prediksi salah negatif.

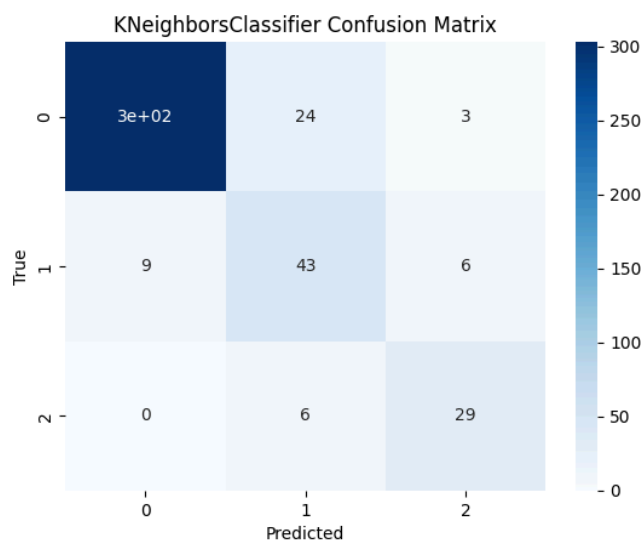
### 3. HASIL PENELITIAN

Dari hasil eksperimen yang telah dilakukan, berikut merupakan hasil evaluasi dari masing-masing model menggunakan metrik Accuracy, Precision, Recall, F1 Score :

Tabel 2 Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1 Score
KNN	89%	90%	89%	89%
SVM	89%	92%	89%	90%
Decision Tree	91%	91%	91%	91%

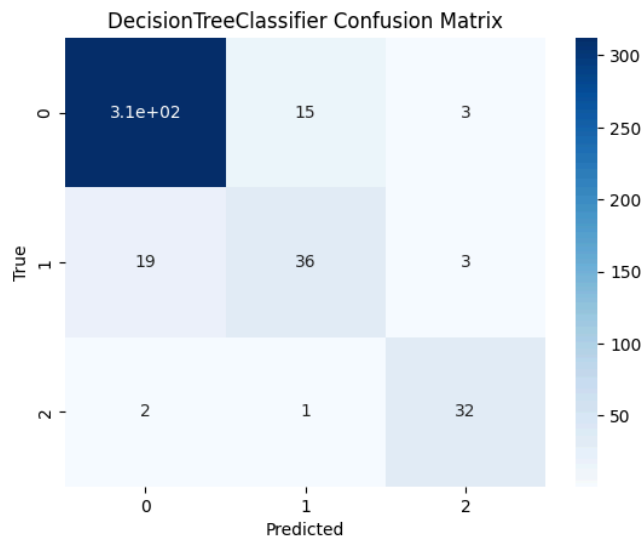
Dari tabel 1 , terlihat bahwa performa dari ketiga model berdasarkan evaluasi metrik sudah optimal, dibuktikan dengan nilai metrik yang berada di atas 85%, menandakan bahwa model tersebut konvergen. Model KNN memiliki nilai yang konsisten di semua metrik, sementara itu model SVM memiliki nilai presisi sebesar 92% dan nilai F1 sebesar 90%, hal tersebut disebabkan oleh , kemudian model Decision Tree memiliki nilai metrik yang serupa pada seluruh metrik yaitu sebesar 91 % dengan nilai metrik accuracy, recall, dan F1 yang lebih besar dibandingkan model KNN dan SVM yang mengimplikasikan bahwa model dapat memprediksi pola dengan lebih akurat dan tepat pada data baru.



Gambar 2 Confusion Matrix KNN

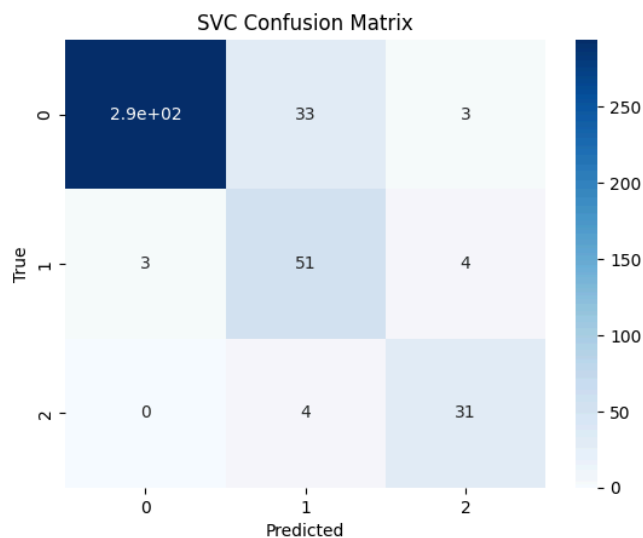
Terlihat dari diagram Confusion Matrix bahwa terdapat total 372 data yang telah terprediksi dengan benar yang menandakan bahwa sebanyak 372 data merupakan True , Positive (TP). label 0 memiliki TP sebanyak 300 data, FP sebanyak 27 data, FN sebanyak 9 dan TN sebanyak 72. label 1 memiliki TP sebanyak 43 data, FP sebanyak 15 data, FN sebanyak 30 data dan TN sebanyak 329 data. label 2 memiliki TP sebanyak 29 data, FP sebanyak 6 data, FN sebanyak 9 data dan TN sebanyak 343 data.





*Gambar 3 Confusion Matrix Decision Tree*

Terlihat dari diagram Confusion Matrix bahwa terdapat total 378 data yang telah terprediksi dengan benar yang menandakan bahwa sebanyak 372 data merupakan True , Positive (TP). label 0 memiliki TP sebanyak 310 data, FP sebanyak 9 data, FN sebanyak 21 dan TN sebanyak 68. label 1 memiliki TP sebanyak 36 data, FP sebanyak 22 data, FN sebanyak 16 data dan TN sebanyak 342 data. label 2 memiliki TP sebanyak 32 data, FP sebanyak 3 data, FN sebanyak 6 data dan TN sebanyak 346 data.



*Gambar 4 Confusion Matrix SVM*

Terlihat dari diagram Confusion Matrix bahwa terdapat total 372 data yang telah terprediksi dengan benar yang menandakan bahwa sebanyak 372 data merupakan True , Positive (TP). label 0 memiliki TP sebanyak 290 data, FP sebanyak 36 data, FN sebanyak 3 dan TN sebanyak 82. label 1 memiliki TP sebanyak 51 data, FP sebanyak 7 data, FN sebanyak 37 data dan TN sebanyak 321 data. label 2 memiliki TP sebanyak 31 data, FP sebanyak 4 data, FN sebanyak 7 data dan TN sebanyak 341 data.

Dari hasil diatas, terlihat bahwa model Decision Tree merupakan model yang optimal dalam memprediksi Kesehatan Janin, hal ini didukung dengan nilai metrik yang konsisten di besaran 91% dan

---

memiliki total data yang diprediksi dengan benar (TP) yang lebih besar yaitu 378 berdasarkan perbandingan besaran TP dari setiap confusion matrix pada masing-masing model.

Walaupun model lain seperti KNN dan SVM memiliki konvergensi yang bagus bahkan memiliki nilai precision yang lebih tinggi yaitu 92% oleh SVM, namun kedua model tersebut kurang optimal dalam memprediksi data di luar data latih dengan akurat jika dibandingkan dengan model Decision Tree.

Terlihat bahwa mayoritas dari hasil prediksi ketiga model cenderung sangat condong dengan label 0 dengan jarak yang jauh antar setiap label, hal ini mengindikasikan bahwa dataset yang digunakan untuk melatih model tidak seimbang terutama pada data dengan label 1 dan 2 sehingga dapat mempengaruhi hasil prediksi dari model. metode augmentasi data yang digunakan seperti (SMOTE) membantu model lebih cepat dalam mencapai konvergensi namun tidak mengatasi ketidakseimbangan dalam prediksi pada model. penambahan data latih dapat menjadi solusi untuk ketidakseimbangan data.

Model dijalankan menggunakan parameter standard dari fungsi yang disediakan oleh scikit-learn. Untuk bisa meningkatkan performa dari model maka hyperparameter tuning menggunakan tools seperti GridSearchCV dan RandomizedSearchCV untuk menemukan parameter terbaik berdasarkan acuan pada metrik yang ingin di optimalkan.

#### 4. CONCLUSION

Berdasarkan hasil evaluasi performa model klasifikasi terhadap data kesehatan janin, model Decision Tree menunjukkan kinerja yang paling optimal dibandingkan KNN dan SVM. Hal ini ditunjukkan oleh nilai metrik evaluasi yang konsisten pada angka 91% serta jumlah prediksi benar (True Positive) tertinggi, yaitu sebanyak 378 data.

Meskipun model SVM memiliki nilai precision tertinggi sebesar 92%, dan model KNN menunjukkan performa konvergen yang baik, kedua model tersebut masih kurang optimal dalam menggeneralisasi data di luar data latih dibandingkan dengan Decision Tree.

Distribusi prediksi dari ketiga model cenderung didominasi oleh kelas label 0, dengan jumlah prediksi yang jauh lebih tinggi dibanding label 1 dan 2. Hal ini menunjukkan adanya ketidakseimbangan kelas dalam dataset yang dapat memengaruhi performa model. Meskipun metode augmentasi seperti SMOTE telah diterapkan dan mampu mempercepat konvergensi model, pendekatan tersebut belum sepenuhnya mengatasi ketidakseimbangan dalam hasil prediksi.

Untuk meningkatkan performa model secara keseluruhan, penambahan jumlah data latih pada kelas minoritas serta penerapan hyperparameter tuning menggunakan alat seperti GridSearchCV dan RandomizedSearchCV direkomendasikan agar model dapat mencapai konfigurasi parameter terbaik sesuai dengan metrik evaluasi yang ditargetkan.

#### DAFTAR PUSTAKA

- [1] WHO, "Newborns: improving survival and well-being," WHO Fact Sheet, Sep. 2020.
  - [2] E. Lawn et al., "4 million neonatal deaths: When? Where? Why?," *Lancet*, vol. 365, no. 9462, pp. 891–900, 2005.
  - [3] A. Ribeiro et al., "Cardiotocography: Evaluation using machine learning techniques," *BMC Med. Inform. Decis. Mak.*, vol. 12 no. 1, 2012.
  - [4] D. M. Sholkamy, "The challenges of CTG interpretation," *Obstet. Gynecol. Int. J.*, vol. 9, no. 6, pp. 363–367, 2018.
  - [5] R. O. Morris et al., "A comparison of artificial intelligence techniques for fetal monitoring," *J. Biomed. Inform.*, vol. 85, pp. 76–86, 2018.
  - [6] L. Jiang et al., "Artificial Intelligence in Healthcare: Past, Present and Future," *Semin. Cancer Biol.*, vol. 86, pp. 1–13, 2022.
-

- 
- [7] F. S. Abin, T. Das, and P. Bhattacharyya, "Comparative Study of Machine Learning Models for Fetal Health Classification," *Procedia Comput. Sci.*, vol. 173, pp. 138–147, 2020.
  - [8] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
  - [9] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
  - [10] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
  - [11] R. Sharma and M. Saini, "Fetal Health Classification Using SVM," *Int. J. Comput. Appl.*, vol. 182, no. 44, pp. 5–9, 2019.
  - [12] J. Singh and A. Arora, "Comparison of ML Algorithms for CTG Data Classification," *Procedia Comput. Sci.*, vol. 167, pp. 1374–1383, 2020.
  - [13] M. Abin et al., "Performance Analysis of Machine Learning Models for Fetal Health Prediction," *Int. J. Appl. Eng. Res.*, vol. 15, no. 7, pp. 641–649, 2020.
  - [14] N. V. Chaurasia and S. Pal, "Performance Analysis of Machine Learning Techniques in Predicting Fetal Health," *Int. J. Comput. Appl.*, vol. 975, no. 8887, 2020.
  - [15] A. Setiawan, D. Cahyani, and A. Wijayanto, "Design of Decision Support System for Fetal Health Classification," *J. Phys. Conf. Ser.*, vol. 1566, 012070, 2020.
  - [16] D. Siswanto, R. S. Nugroho, and H. W. Prabowo, "Fetal Distress Classification Using K-Nearest Neighbor and Data Preprocessing," *Procedia Computer Science*, vol. 135, pp. 276–283, 2018. <https://doi.org/10.1016/j.procs.2018.08.178>
  - [17] D. Mohapatra, S. K. Swain, and S. Bhoi, "Machine Learning-based Fetal Health Classification using Decision Tree and Random Forest," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 5, no. 3, pp. 173–179, 2019.
  - [18] *CRISP-DM methodology*, Smart Vision Europe, [Online]. Available: <https://www.sv-europe.com/crisp-dm-methodology/>. [Accessed: Jul. 5, 2025].
  - [19] V. Subha, Dr.D.Murugan, J. Rani, and Dr.K.Rajalakshmi, "Comparative Analysis of Classification Techniques," *ResearchGate*, no. December, pp. 1–8, 2013.
  - [20] Anita, Asrul Abdullah, and Syarifah Putri Agustini Alkadri, "Classification of Fetal Health Using the K-Nearest Neighbor Method and the Relieff Feature Selection Method," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 2, pp. 986–989, 2025, doi: 10.59934/jaiea.v4i2.794.
  - [21] 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.
  - [22] M. M. Islam, M. Rokunojjaman, A. Amin, M. N. Akhtar, and I. H. Sarker, "Diagnosis and Classification of Fetal Health Based on CTG Data Using Machine Learning Techniques," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 491 LNICST, no. June, pp. 3–16, 2023, doi: 10.1007/978-3-031-34622-4\_1.
  - [23] M. M. R. Khan, R. B. Arif, A. B. Siddique, and M. R. Oishe, "Study and observation of the variation of accuracies of KNN, SVM, LMNN, ENN algorithms on eleven different datasets from UCI machine learning repository," *4th Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEICT 2018*, pp. 124–129, 2018, doi: 10.1109/CEEICT.2018.8628041.
-