

Perbandingan XGBoost dan Random Forest Menggunakan Seleksi Fitur ANOVA-MI Dalam Klasifikasi Kesehatan Janin Cardiotocography

Abednego Destyo Amanda, Angga Bayu Santoso*

Fakultas Teknik dan Ilmu Komputer, Prodi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹abednego_destyo_amanda@teknokrat.ac.id, ²anggabayu@teknokrat.ac.id

Email Penulis Korespondensi: anggabayu@teknokrat.ac.id

Submitted: 17/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

Abstrak—Studi ini membandingkan kinerja algoritma Random Forest dan XGBoost dalam mengklasifikasikan masalah kesehatan janin menggunakan data Kardiotokografi (CTG). Ketidakseimbangan jumlah data antar kelas, keberadaan fitur yang kurang relevan dan tantangan dalam mengidentifikasi kelas *Suspect*, yang memiliki karakteristik di antara kelas Normal dan *Pathological*, merupakan masalah utama pada dataset CTG. Kondisi ini penting karena tahap awal risiko gangguan kesehatan janin menentukan penanganan medis lebih lanjut yang direpresentasikan oleh kelas *Suspect*. Penelitian ini menggunakan teknik seleksi fitur ANOVA dan Mutual Information, serta metode oversampling ADASYN untuk menyeimbangkan data guna mengatasi masalah-masalah tersebut. Selain itu, Random Search digunakan untuk mengoptimasi parameter model guna meningkatkan kerjanya. Berbeda dari penelitian sebelumnya yang umumnya berfokus pada peningkatan akurasi, penelitian ini juga menekankan kemampuan model untuk mendeteksi kelas minoritas, khususnya kelas *Suspect*. Berdasarkan hasil penelitian, dalam hampir setiap skenario pengujian, XGBoost secara konsisten mengungguli kinerja Random Forest. Model XGBoost memperoleh akurasi optimal dari kombinasi ANOVA, ADASYN, dan hyperparameter tuning dengan akurasi sebesar 95,51%. Sementara itu, penerapan Mutual Information dengan ADASYN dan tuning cukup efektif dalam mengidentifikasi kelas *Suspect* dengan nilai *recall* yang lebih tinggi sebesar 81%. Namun, karena atribut kelas *Suspect* berada diantara atribut kelas Normal dan *Pathological*, model masih menghadapi tantangan untuk membedakannya secara optimal. Secara keseluruhan, penelitian ini menunjukkan bahwa kombinasi seleksi fitur yang tepat, penanganan ketidakseimbangan data, dan optimasi parameter dalam satu pipeline dapat meningkatkan performa model secara lebih seimbang. Penelitian ini diharapkan dapat mendukung pengambilan keputusan medis yang lebih objektif, khususnya dalam mendeteksi kondisi janin berisiko sejak tahap awal.

Kata Kunci: Kesehatan Janin; Cardiotocography; Random Forest; XGBoost; Seleksi Fitur; ADASYN

Abstract—This study compares the performance of Random Forest and XGBoost algorithms in classifying fetal health problems using Cardiotocography (CTG) data. The imbalance in the amount of data between classes, the presence of less relevant features, and the challenge in identifying the Suspect class, which has characteristics between the Normal and Pathological classes, are the main problems in the CTG dataset. This condition is important because the early stage of fetal health risk determines further medical treatment represented by the Suspect class. This study uses ANOVA and Mutual Information feature selection techniques, as well as the ADASYN oversampling method to balance the data to overcome these problems. In addition, Random Search is used to optimize model parameters to improve its performance. Unlike previous studies that generally focus on improving accuracy, this study also emphasizes the model's ability to detect minority classes, especially the Suspect class. Based on the results of the study, in almost every test scenario, XGBoost consistently outperforms Random Forest. The XGBoost model obtained optimal accuracy from the combination of ANOVA, ADASYN, and hyperparameter tuning with an accuracy of 95.51%. Meanwhile, the application of Mutual Information with ADASYN and tuning was quite effective in identifying the Suspect class with a higher recall value of 81%. However, because the Suspect class attribute lies between the Normal and Pathological class attributes, the model still faces challenges in optimally distinguishing them. Overall, this study shows that a combination of appropriate feature selection, handling data imbalance, and parameter optimization in a single pipeline can improve model performance more balanced. This research is expected to support more objective medical decision-making, especially in detecting fetal risk conditions from an early stage.

Keywords: Fetal Health; Cardiotocography; Random Forest; XGBoost; Feature Selection; ADASYN

1. PENDAHULUAN

Selama kehamilan, detak jantung janin adalah salah satu indeks penting kualitas hidup janin. Pemantauan detak jantung janin dapat mengungkapkan detail penting tentang kesehatan janin dan membantu mengidentifikasi kondisi seperti hipoksia atau stres janin. Organisasi Kesehatan Dunia (WHO) melaporkan bahwa kematian janin masih menjadi masalah utama bagi kesehatan ibu dan anak di seluruh dunia. Indonesia memiliki angka kematian bayi yang relatif tinggi, dibandingkan dengan banyak negara di kawasan Asia Tenggara [1]. Oleh karena itu, metode Cardiotocography digunakan sebagai teknik pemantauan untuk mengidentifikasi risiko penyakit pada janin selama kehamilan.

Penelitian ini menggunakan dataset *Fetal Health Classification* dari Kaggle. Salah satu permasalahan pada dataset ini adalah ketidakseimbangan data antar kelas, dengan data kelas Normal (N) jauh melebihi data kelas *Suspect* (S) dan kelas *Pathological* (P). Ketidakseimbangan ini dapat memengaruhi kinerja model klasifikasi, karena model cenderung menghasilkan prediksi yang bias terhadap kelas mayoritas dibandingkan kelas minoritas. Selain itu, tantangan utama terletak pada kesulitan dalam mengklasifikasikan kelas *Suspect*, karena karakteristiknya berada di antara kedua kelas lainnya, sehingga menimbulkan ambiguitas klinis dan menyebabkan terjadinya *overlap* fitur antar kelas [2].

Permasalahan tersebut menunjukkan bahwa diperlukan pendekatan komputasi yang mampu mengatasi ketidakseimbangan data dan meningkatkan akurasi klasifikasi, khususnya pada kelas *Suspect*. Metode klasifikasi

berbasis *machine learning* memungkinkan pengembangan model yang dapat memprediksi status kesehatan janin dengan lebih objektif berdasarkan evaluasi data CTG, karena mampu mendeteksi pola tersembunyi yang sulit ditemukan melalui analisis manual.

Oleh karena itu, diperlukan algoritma yang mampu menangani karakteristik khusus dataset CTG, seperti ketidakseimbangan kelas, kompleksitas hubungan antar fitur, serta *overlap* fitur pada kelas *Suspect*. Dalam penelitian klasifikasi kesehatan janin berbasis data CTG, algoritma Random Forest dan XGBoost dipilih karena kesesuaiannya dengan fitur-fitur tersebut. Random Forest efektif dalam menangani data dengan banyak fitur numerik serta mampu mengurangi *varians* dan *overfitting* melalui teknik *bagging*, sehingga lebih stabil dalam menghadapi distribusi data yang tidak seimbang dan pola non-linear pada data CTG [3].

Sementara itu, XGBoost sebagai algoritma berbasis *boosting* mampu meningkatkan performa model secara iteratif melalui optimisasi fungsi *loss* dan pemberian bobot pada kesalahan prediksi sebelumnya, sehingga lebih sensitif dalam membedakan kelas yang memiliki karakteristik saling berdekatan, seperti kelas *Suspect* dengan kelas Normal dan *Pathological*. Namun, performa XGBoost pada kelas minoritas dapat menurun pada dataset yang sangat tidak seimbang, sehingga diperlukan pendekatan tambahan untuk meningkatkan kemampuannya [4]. Dengan karakteristik tersebut, kedua algoritma ini dipilih secara spesifik untuk meningkatkan performa klasifikasi pada dataset CTG, khususnya dalam mengatasi kesulitan deteksi pada kelas *Suspect*.

Sejumlah penelitian serupa telah dilakukan sebelumnya untuk meningkatkan akurasi dalam mengklasifikasikan kondisi kesehatan janin dengan memanfaatkan *machine learning*. Salah satu penelitian, Campos et al. (2024) menjelaskan bahwa sinyal *Fetal Heart Rate* (FHR) rentan terkontaminasi oleh noise dan artefak sinyal yang dapat mengaburkan pola detak jantung janin. Kondisi ini menekankan pentingnya tahap preprocessing data [5]. Studi lain oleh Nazli et al. (2025) mengembangkan model LightGBM dengan fokus utama pada peningkatan nilai akurasi sebesar 91,34% sebagai indikator performa model, tetapi evaluasi dalam penelitian tersebut masih terbatas pada metrik akurasi tanpa mempertimbangkan metrik lain seperti *recall* dan *F1-score*[6].

Sementara itu, Ilham et al. (2025) menggunakan dataset berbeda yaitu *CTU-UHB Intrapartum Cardiotocography Database* dan menerapkan model seperti Random Forest. Meskipun kinerja Random Forest pada penelitian ini mencapai akurasi 93%, prediksi model cenderung bias ke kelas mayoritas karena tidak menyertakan strategi penanganan ketidakseimbangan kelas [7]. Kuzu dan Santur (2023) menggunakan dataset CTG dengan menggunakan pendekatan pembelajaran *ensemble* dan analisis korelasi antar fitur, namun tidak menggunakan prosedur seleksi fitur untuk mengurangi informasi yang berlebihan dalam dataset [8].

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa meskipun performa model klasifikasi pada data CTG telah menunjukkan hasil yang baik, masih terdapat beberapa keterbatasan yang belum teratasi secara menyeluruh. Secara umum, penelitian sebelumnya cenderung berfokus pada peningkatan akurasi tanpa memperhatikan performa pada kelas minoritas, khususnya kelas *Suspect*, yang memiliki karakteristik kompleks dan saling tumpang tindih dengan kelas lainnya. Selain itu, penanganan ketidakseimbangan data, seleksi fitur, dan optimasi model seringkali dilakukan secara terpisah, sehingga belum memberikan solusi yang komprehensif terhadap permasalahan klasifikasi pada data CTG. Untuk mengatasi keterbatasan tersebut, penelitian ini mengintegrasikan teknik penanganan ketidakseimbangan data dan seleksi fitur dalam satu pendekatan yang terstruktur. Salah satu metode yang digunakan untuk menangani ketidakseimbangan kelas adalah *Adaptive Synthetic Sampling* (ADASYN), yaitu teknik *oversampling* berbasis sintesis data yang menambahkan sampel buatan ke kelas minoritas secara adaptif. Teknik ini menghasilkan data sintesis dengan mempertimbangkan tingkat kesulitan setiap sampel minoritas, dimana sampel yang lebih sulit dipelajari akan mendapatkan proporsi data tambahan yang lebih besar [9].

Untuk menentukan fitur mana yang paling relevan dengan variabel target, seleksi fitur merupakan langkah penting untuk mengurangi redundansi informasi. Adapun peneliti saat ini menekankan penggunaan ANOVA dengan Mutual Information sebagai seleksi fitur. Penerapan ANOVA pada penelitian klasifikasi anemia menyatakan, akurasi hasil evaluasi model SVM dapat meningkatkan hingga 94,61%, sementara model tanpa seleksi fitur memiliki akurasi di bawah 90% [10].

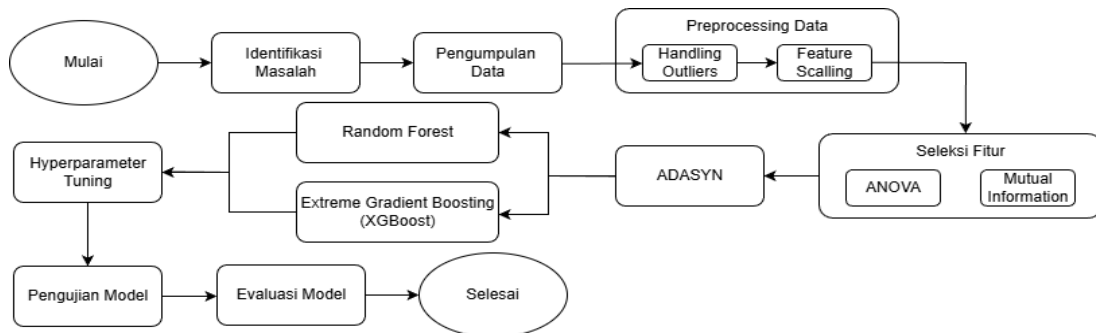
Pada penelitian lain, Mutual Information digunakan untuk meningkatkan performa klasifikasi Random Forest, dimana Mutual Information lebih baik dengan peningkatan akurasi sekitar 2% - 5% dibandingkan metode sebelumnya. Kedua metode ini digunakan untuk mengidentifikasi fitur yang paling relevan terhadap variabel target. Dengan demikian, kualitas input model dapat ditingkatkan sehingga membantu dalam membedakan karakteristik antar kelas, khususnya pada kelas *Suspect*. Penggunaan seleksi fitur juga berperan dalam meningkatkan efisiensi komputasi serta kinerja model klasifikasi [11].

Berdasarkan pemaparan diatas, penelitian ini bertujuan untuk mengevaluasi kinerja algoritma Random Forest dan XGBoost dalam klasifikasi kesehatan janin berbasis data *Cardiotocography* dengan pendekatan terintegrasi yang menggabungkan seleksi fitur dan penanganan ketidakseimbangan data. Secara khusus, studi ini menganalisis pengaruh penggunaan ANOVA dan Mutual Information sebagai metode seleksi fitur yang dikombinasikan dengan teknik *oversampling* ADASYN terhadap performa model, yang diukur menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Pendekatan ini diharapkan mampu meningkatkan kemampuan model dalam mendeteksi kelas *Suspect*, yang merupakan tantangan utama dalam klasifikasi data CTG, serta memberikan dasar analisis yang lebih objektif dan andal dalam mendukung pengambilan keputusan medis oleh para praktisi kesehatan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini, kinerja algoritma Random Forest dan XGBoost dalam mengkategorikan masalah kesehatan janin berdasarkan data *Cardiotocography* dibandingkan menggunakan pendekatan kuantitatif berdasarkan *machine learning*. Untuk menemukan fitur yang paling relevan dan meningkatkan kinerja model klasifikasi, kedua metode tersebut diintegrasikan menggunakan pendekatan seleksi fitur *Analysis of Variance* (ANOVA) dan *Mutual Information*. Penelitian ini menggunakan dataset *Fetal Health Classification* yang diperoleh dari Kaggle yang dibagi menjadi tiga kategori: Normal, *Suspect*, dan *Pathological*. Kemampuan prediksi kedua algoritma dan tingkat akurasi optimal dinilai sebelum pengambilan kesimpulan. Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Studi ini berdasarkan kerangka metodologis yang merujuk pada penelitian sebelumnya, gambar 1 menggambarkan alur langkah-langkah penelitian yang dilakukan dalam beberapa tahapan yaitu identifikasi masalah, pengumpulan data, pra-pemrosesan, seleksi fitur, pelatihan model, pengujian, dan evaluasi kinerja menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Tujuan dari metode ini adalah untuk menciptakan model klasifikasi dengan tingkat akurasi yang tinggi, stabil, dan kemampuan untuk menangkap karakteristik dari data yang diamati.

2.2 Pengumpulan Data

Dataset *Fetal Health Classification* yang digunakan dalam penelitian ini bersumber dari platform Kaggle, terdiri dari 2.126 data catatan hasil pemeriksaan prenatal yang berasal dari wanita hamil trimester ketiga. Dataset ini berisi 22 atribut: 21 fitur numerik dan 1 atribut label target yaitu *fetal_health* yang dirancang untuk membantu mengurangi angka kematian ibu dan bayi baru lahir. Fitur dataset berasal dari sinyal CTG yang diekstrak dan diklasifikasikan oleh tiga dokter kandungan ke dalam tiga kategori kesehatan janin: 1 = Normal, 2 = Suspect, dan 3 = Pathological. Deskripsi atribut dalam dataset kesehatan janin disajikan pada Tabel 1.

Tabel 1. Atribut Dataset Cardiotocography

Simbol Variabel	Deskripsi Variabel
BV	Baseline Value (FHR beats per minute)
AC	Accelerations (number of accelerations per second)
FM	Fetal Movement (number of fetal movement)
UC	Uterine Contractions (number of uterine contractions per second)
LD	Light Decelerations (number of light decelerations per second)
SD	Severe Decelerations (number of severe decelerations per second)
PD	Prolonged Decelerations (number of prolonged decelerations per second)
ASTV	Abnormal Short-Term Variability
MSTV	Mean Value of Short-Term Variability
ALTV	Percentage of Time with Abnormal Long-Term Variability
MLTV	Mean Value of Long-Term Variability
HW	Histogram Width (Width of FHR histogram)
HMax	Histogram Max (Maximum of FHR histogram)
Hmin	Histogram Min (Minimum of FHR histogram)
NP	Number of Histogram Peaks
NZ	Number of Histogram Zeroes
Hmo	Histogram Mode
Hme	Histogram Mean
Hmed	Histogram Median
HV	Histogram Variance
HT	Histogram Tendency
NSP	Fetal Health (Fetal state class code, N=Normal, S=Suspected, P=Pathological)

Berdasarkan Tabel 1, setiap variabel dalam dataset merupakan indikator penting yang digunakan untuk mendeskripsikan keadaan fisiologis janin. *Baseline value*, *short-term variability*, *long-term variability* dan jumlah *accelerations* digunakan untuk menilai kesehatan janin berdasarkan *Fetal Heart Rate*, sesuai dengan pedoman dari *National Institute of Child Health and Human Development*. Tetapi, tanda masalah penurunan denyut jantung janin dikaitkan dengan fitur seperti *light decelerations*, *severe decelerations* dan *prolongued decelerations*. Selain itu, distribusi pola detak jantung janin dijelaskan secara statistik menggunakan fitur berdasarkan histogram.

2.3 Preprocessing Data

Langkah penting dalam pembelajaran mesin adalah pra-pemrosesan data, yang mengubah data mentah menjadi format yang dapat digunakan oleh algoritma. Untuk meningkatkan kualitas dan konsistensi data, langkah ini melibatkan *data cleaning*, transformasi, dan normalisasi. Praproses juga membantu mengatasi masalah yang dapat memengaruhi kinerja model, seperti *noise*, nilai yang hilang, dan ketidakseimbangan data. Menurut studi terbaru, peningkatan akurasi, presisi, *recall*, dan *F1-score* model pembelajaran mesin secara langsung dipengaruhi oleh kualitas *preprocessing*.

2.3.1 Data Cleaning

Sebelum digunakan dalam proses pemodelan, tahap *preprocessing data* dilakukan untuk menjamin kualitas data. Tahap awal dalam proses ini yaitu, dilakukan pemeriksaan data duplikat dan *missing value* dalam dataset yang berjumlah 2.126 data. Tujuan dari prosedur ini adalah untuk menemukan kemungkinan inkonsistensi data yang dapat memengaruhi hasil analisis dan kinerja model. Untuk menjamin bahwa setiap entri dalam dataset berbeda dan menghindari bias dalam proses pembelajaran model, data yang ditemukan sebagai duplikat kemudian dieliminasi.

2.3.2 Splitting Data

Untuk menjaga proporsi distribusi kelas, data kemudian dibagi menjadi fitur (x) dan variabel target (y) sebelum dibagi menjadi 80% *data training* dan 20% *data testing* menggunakan teknik *stratified sampling* yang bertujuan menjaga distribusi kelas tetap seimbang di kedua subset data, sehingga model dapat mempelajari atribut setiap kelas secara proporsional dan tidak bias terhadap kelas Normal yang memiliki jumlah data paling dominan. Selain itu, tahap *data splitting* dilakukan untuk mencegah *data leakage*, keadaan di mana proses pelatihan model menggunakan informasi dari data uji secara tidak langsung.

2.3.3 Normalisasi Data

Untuk memastikan bahwa setiap fitur memiliki skala yang seragam sebelum digunakan dalam model *machine learning*, normalisasi data dilakukan sebagai bagian dari *preprocessing*. Prosedur *preprocessing* seperti normalisasi terbukti penting untuk meningkatkan kinerja model dan kualitas analisis data berdasarkan penelitian sebelumnya pada dataset CTG [12]. Penelitian ini menggunakan metode *StandardScaler* untuk normalisasi data, yang memodifikasi distribusi data sehingga memiliki nilai rata-rata 0 dan standar deviasi 1, sebagaimana dituliskan dalam rumus:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

2.3.4 Seleksi Fitur

Menemukan fitur yang tepat untuk digunakan dalam prediksi merupakan salah satu aspek krusial dari *machine learning*. Tujuan dari tahap seleksi fitur adalah menemukan atribut yang menawarkan informasi paling berkaitan dengan kesejahteraan janin. Berdasarkan karakteristik dataset yang memiliki fitur numerik dan variabel target kategorikal multi kelas, diterapkan ANOVA berdasarkan uji statistik untuk menguji setiap fitur antar kelas melalui nilai *F-statistic*, yang dirumuskan sebagai berikut:

$$F = \frac{MSB}{MSW} \quad (2)$$

F mewakili *F-statistic*, *Mean Square Between* (MSB) adalah variasi antar kelompok, dan *Mean Square Within* (MSW) adalah variasi dalam kelompok [13]. Penelitian ini menggunakan pendekatan *ranking* dengan mengurutkan nilai *F-statistic* dan memilih 10 fitur terbaik (*top-F*) dalam melatih model.

Selain ANOVA, teknik seleksi fitur *Mutual Information* berbasis teori informasi digunakan untuk mengukur tingkat korelasi setiap atribut yang paling informatif dengan variabel target, baik secara linier dan non-linear. Penggunaan ANOVA dan MI dilakukan secara komparatif untuk mengevaluasi pengaruh karakteristik hubungan fitur terhadap performa model. ANOVA efektif dalam mengidentifikasi perbedaan rata-rata signifikan antar kelas (hubungan linier), sedangkan MI mampu menangkap ketergantungan non-linear yang umum terdapat pada dataset CTG yang memiliki pola kompleks. Dengan demikian, kedua metode ini tidak bersifat redundan, melainkan digunakan untuk membandingkan efektivitas pendekatan linier dan non-linear dalam meningkatkan performa klasifikasi, khususnya pada kelas Suspect yang memiliki karakteristik berada di antara dua kelas lainnya. Selanjutnya, fitur diurutkan berdasarkan MI *score* dan dipilih 10 fitur terbaik untuk digunakan dalam pelatihan model klasifikasi. MI mengukur jumlah informasi yang dibagi antara fitur X_i dan label y , yang ditunjukkan oleh rumus berikut:

$$MI(X_i; y) = \sum x_i, y P(x_i, y). \log \frac{P(x_i, y)}{P(x_i)P(y)} \quad (3)$$

2.3.5 ADASYN (Adaptive Synthetic Sampling)

Pada dataset medis seperti CTG, ketidakseimbangan kelas (class imbalance) sering terjadi, dimana jumlah data kelas Normal jauh lebih besar dibandingkan kelas Suspect dan Pathological. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas. Untuk mengatasinya, digunakan teknik ADASYN (Adaptive Synthetic Sampling) untuk menyeimbangkan distribusi data pelatihan. ADASYN dikonfigurasi dengan *parameter* $n_neighbors = 5$, *sampling_strategy* = 'auto', dan *random_state* = 42. Parameter $n_neighbors$ dipilih untuk menjaga keseimbangan antara representasi lokal data dan stabilitas distribusi sintetis pada data CTG yang memiliki overlap antar kelas. Sementara itu, *sampling_strategy* digunakan untuk menyesuaikan jumlah sampel kelas minoritas, dan *random_state* untuk memastikan konsistensi hasil. ADASYN memberikan perhatian lebih pada data yang berada di dekat batas kelas atau yang sering salah diklasifikasikan, sehingga membantu meningkatkan kemampuan model dalam mempelajari pola yang lebih kompleks. Penerapan ADASYN dalam penelitian ini bertujuan agar model dapat mempelajari karakteristik setiap kelas secara lebih merata, sehingga meningkatkan performa klasifikasi terutama pada kelas minoritas.

2.4 Klasifikasi Random Forest

Salah satu algoritma *supervised machine learning* yang banyak digunakan adalah Random Forest, yang dapat diterapkan pada permasalahan regresi dan klasifikasi. Model ini membangun sejumlah pohon keputusan untuk menghasilkan target kelas berdasarkan hasil voting tertinggi dari setiap pohon. Random Forest menggunakan teknik *bagging* serta pemilihan fitur acak untuk membangun kumpulan pohon yang tidak berkorelasi, sehingga menghasilkan prediksi yang lebih akurat dibandingkan satu pohon keputusan tunggal. Setelah proses pembentukan model selesai, setiap pohon akan menerima data uji dan memprediksi kelas secara independen, yang kemudian digabungkan untuk menghasilkan keputusan akhir. Dalam penelitian ini, model Random Forest dilatih menggunakan data hasil ADASYN dengan transformasi label dari {1, 2, 3} menjadi {0, 1, 2} ke format multi-kelas. Prediksi akhir Random Forest ditentukan dengan mengambil voting mayoritas beberapa pohon keputusan yang dirumuskan sebagai berikut:

$$y = \text{mode}(h_1(x), h_2(x), \dots, h_n(x)) \quad (4)$$

$h_i(x)$ adalah hasil prediksi pohon ke- i berdasarkan data masukan x . Berdasarkan penelitian di bidang medis yang dilakukan oleh Widyastuty dan Azis (2024), algoritma Random Forest dalam klasifikasi *sleep disorders* mampu mencapai nilai *recall* sebesar 100%. Nilai *recall* yang sempurna ini menunjukkan bahwa model mampu mendeteksi seluruh kasus positif tanpa kesalahan [14]. Manfaat utama Random Forest adalah kemampuannya untuk menghasilkan model yang andal dan akurat dengan menggabungkan beberapa pohon keputusan. Metode ini dapat mendeteksi pola rumit dari banyak faktor yang memengaruhi kesehatan janin, termasuk detak jantung janin, akselerasi, dan deselerasi, dalam konteks dataset kardiologiografi (CTG).

2.4 Klasifikasi XGBoost

Extreme Gradient Boosting atau XGBoost adalah algoritma *ensemble learning* yang efektif dan dikenal karena kinerjanya yang sangat baik dalam berbagai masalah pembelajaran mesin, termasuk pemrosesan data terstruktur atau tabular. XGBoost merupakan algoritma berbasis *gradient boosting* yang mengoptimalkan fungsi *loss* terdiferensiasi untuk membangun *ensemble* dari sejumlah *weak learner* (umumnya berupa pohon keputusan) secara bertahap dan berurutan. Salah satu kelebihan XGBoost terletak pada implementasinya yang dirancang efisien dengan menggabungkan teknik regularisasi untuk mencegah *overfitting* dan strategi paralelisasi untuk mempercepat proses *training*. Dengan meningkatkan *ensemble* secara iteratif menggunakan optimasi *gradient descent*, XGBoost dapat mencapai performa optimal dalam hal akurasi prediksi, skalabilitas, dan efisiensi komputasi. Secara matematis, proses optimasi pada XGBoost dinyatakan dalam bentuk fungsi *loss* yang terdiri dari komponen kesalahan prediksi dan regularisasi sebagai berikut:

$$\text{Loss}(y, \hat{y}) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \mathcal{U}(f_k) \quad (5)$$

Dimana $l(y_i, \hat{y}_i)$ merupakan fungsi *loss* yang mengukur selisih antara nilai aktual dan hasil prediksi, sedangkan, $\mathcal{U}(f_k)$ merupakan komponen regularisasi yang berfungsi untuk mengontrol kompleksitas model sehingga dapat mencegah terjadinya *overfitting*. Berbagai penelitian telah dilakukan mengenai penerapan algoritma XGBoost dalam bidang kesehatan, dan algoritma ini menunjukkan kinerja yang baik dalam berbagai tugas klasifikasi medis. Model XGBoost bekerja lebih baik daripada teknik klasifikasi konvensional dalam memprediksi prognosis pasien serta mampu mencapai akurasi tinggi dalam klasifikasi data medis seperti deteksi aritmia berbasis ECG [15].

2.5 Hyperparameter

Pada penelitian ini optimasi parameter dilakukan pada kedua algoritma dengan menggunakan metode *Random Search*. Metode ini dipilih karena mampu mencari kombinasi parameter terbaik secara lebih efisien dibandingkan dengan pencarian manual atau *grid search*. Fungsi objektif yang digunakan dalam proses optimasi adalah rata-rata akurasi

hasil *cross-validation 5-fold* pada data latih yang telah melalui proses ADASYN. Dengan demikian, parameter yang diperoleh diharapkan tidak hanya sesuai terhadap data latih, tetapi memiliki kemampuan generalisasi yang lebih baik. Hasil optimasi parameter masing-masing algoritma ditampilkan pada Tabel 2.

Tabel 2. Hasil Hyperparameter Tuning

Model	Parameter
Random Forest	n_estimators=500, min_samples_split=5, min_samples_leaf=1, max_features=sqrt, max_depth=15, bootstrap=False
XGBoost	subsample=0.8, n_estimators=300, min_child_weight=1, max_depth=6, learning_rate=0.2, gamma=0.1, colsample_bytree=1.0

Berdasarkan Tabel 2, model Random Forest menggunakan 500 estimator dengan kedalaman maksimum 15, sedangkan XGBoost menggunakan 300 estimator dengan kedalaman maksimum 6. Perbedaan konfigurasi ini menunjukkan bahwa Random Forest cenderung menggunakan lebih banyak jumlah pohon untuk meningkatkan stabilitas prediksi. XGBoost menggunakan kedalaman pohon yang lebih terbatas untuk mencegah *overfitting* melalui proses *boosting* yang iteratif.

2.6 Evaluasi Model Klasifikasi

Bagian terakhir dari penelitian ini adalah mengevaluasi seberapa baik model klasifikasi bekerja menggunakan metode Random Forest dan XGBoost. Evaluasi dilakukan untuk menilai akurasi model dalam mengklasifikasikan kondisi janin berdasarkan data kardiotokografi menggunakan *confusion matrix*. Selain itu, metrik *accuracy*, *precision*, *recall*, dan *F1-score* digunakan untuk mengukur kinerja model dalam memprediksi kondisi janin secara akurat.

a. Accuracy

Akurasi didefinisikan sebagai proporsi jumlah data yang diklasifikasikan dengan benar oleh model terhadap total seluruh data yang diuji. Metrik ini rentan terhadap *imbalance class*, yang dapat mengakibatkan kesalahan interpretasi. Karena itu, selain akurasi diperlukan pengukuran metrik lain seperti *recall* dan *F1-score* yang dapat memperlihatkan performa model secara keseluruhan [16]. Rumus akurasi dihitung sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

b. Precision

Presisi sebagai rasio jumlah data yang diprediksi benar untuk suatu kelas terhadap semua jumlah data yang diprediksi sebagai kelas tersebut. Hal ini sangat penting untuk mengevaluasi kemampuan suatu model mengurangi prediksi positif palsu (*False Positive*) yang dalam konteks medis dapat mengakibatkan kesalahan diagnosis, terutama pada kondisi janin yang sebenarnya normal [17]. Rumus presisi dihitung sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

c. Recall

Recall menggambarkan sejauh mana model dapat mengidentifikasi semua kasus positif sebenarnya. Nilai recall yang tinggi mengindikasikan kasus positif yang dapat ditemukan oleh model. Hal ini sangat penting karena model perlu mengidentifikasi sebanyak mungkin janin berisiko *Pathological* dalam kasus seperti diagnosis janin. Rumus *recall* dihitung sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

d. F1-Score

Pada dataset dengan distribusi kelas yang tidak merata, *F1-score* digunakan untuk mengukur seberapa baik keseimbangan antara *precision* dan *recall*. Saat mengevaluasi performa model pada kelas minoritas, skor ini lebih representatif daripada akurasi. Rumus *F1-score* dihitung sebagai berikut:

$$F1 = \frac{2x(Precision \times Recall)}{Precision+Recall} \quad (9)$$

3. HASIL DAN PEMBAHASAN

3.1 Pemeriksaan Duplikasi dan Missing Value Data

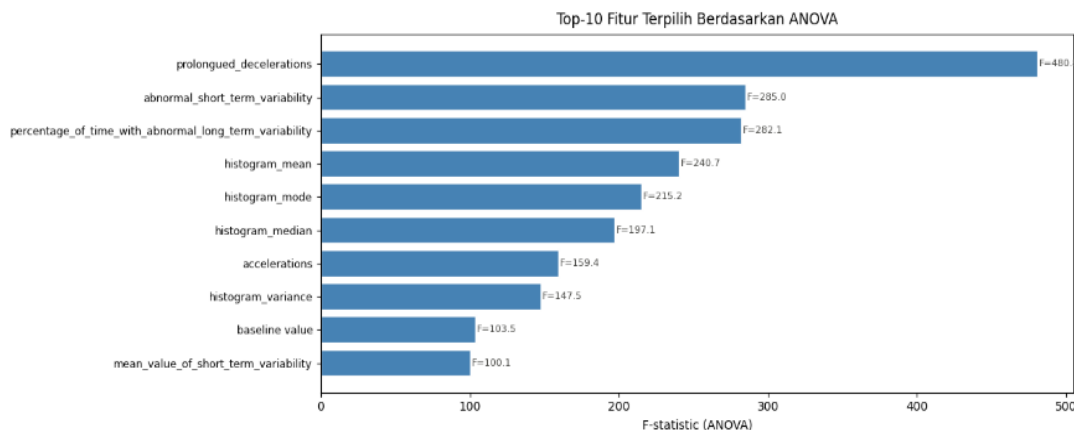
Langkah awal *cleaning data* mencakup pencarian data duplikat yang dapat memengaruhi hasil analisis. Hasil eksplorasi menemukan 13 baris data sebagai duplikat, kemudian data tersebut dihapus agar tidak mempengaruhi hasil analisis, karena dapat terlalu memengaruhi pola tertentu dalam model pelatihan. Sementara itu, tidak ditemukan *missing value* dalam dataset, sehingga tidak diperlukan penanganan khusus. Selanjutnya, dilakukan tahap pembagian data (*data splitting*) menjadi data pelatihan dan pengujian dengan proporsi 80% dan 20%, sehingga diperoleh 1.690 data latih dan 423 data uji dengan 21 fitur sebagai variabel input. Tahap ini bertujuan menghasilkan evaluasi model

yang lebih objektif serta mengurangi risiko model terlalu beradaptasi dengan data pelatihan sehingga tetap mampu melakukan generalisasi pada data baru

3.2 Seleksi Fitur

3.2.1 Seleksi Fitur dengan ANOVA

Seleksi fitur menggunakan *Analysis of Variance* (ANOVA) dilakukan untuk mengidentifikasi atribut yang paling relevan terhadap variabel target. Fitur kemudian diurutkan berdasarkan nilai F-statistik, dan 10 fitur teratas ditampilkan pada Gambar 2.

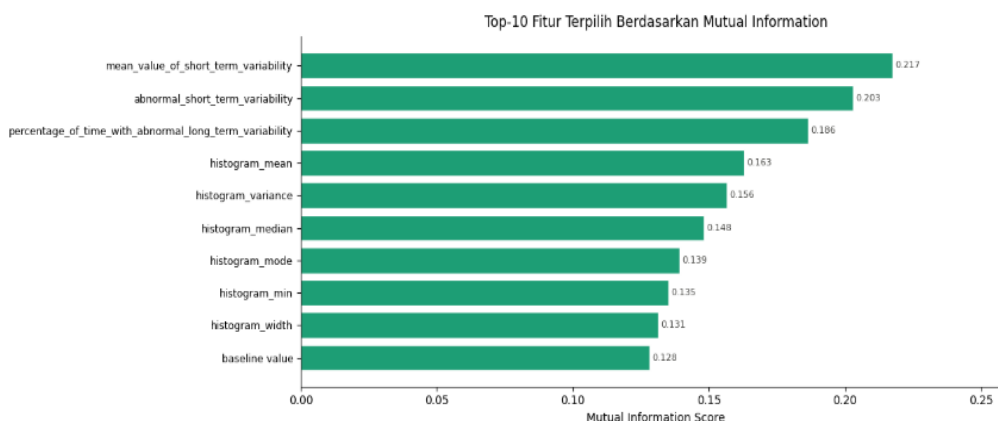


Gambar 2. Uji ANOVA

Berdasarkan Gambar 2, nilai F-statistik yang tinggi menunjukkan bahwa fitur yang dipilih memiliki perbedaan signifikan antar kelas, sehingga relevan dalam klasifikasi. Sebagian besar fitur yang dipilih terkait dengan perubahan variabilitas dan distribusi sinyal detak jantung janin memiliki peran penting dalam merepresentasikan kondisi fisiologis janin. Temuan ini mengindikasikan bahwa. Fitur-fitur tersebut berkaitan dengan respons janin terhadap kondisi stres atau hipoksia, sehingga menjadi indikator utama dalam membedakan tingkat kesehatan janin. Namun, karena ANOVA hanya menangkap hubungan linier antar fitur dan kelas, pemisahan kelas yang memiliki karakteristik berdekatan, seperti kelas *Suspect*, belum sepenuhnya optimal, sehingga diperlukan pendekatan lain yang mampu menangkap hubungan non-linear.

3.2.2 Seleksi Fitur dengan Mutual Information

Pada tahap selanjutnya, seleksi fitur dilakukan menggunakan Mutual Information untuk mengukur ketergantungan antara fitur dan variabel target. Berbeda dengan ANOVA, metode ini mampu menangkap hubungan non-linear, dengan fitur diurutkan berdasarkan skor Mutual Information seperti ditunjukkan pada Gambar 3.

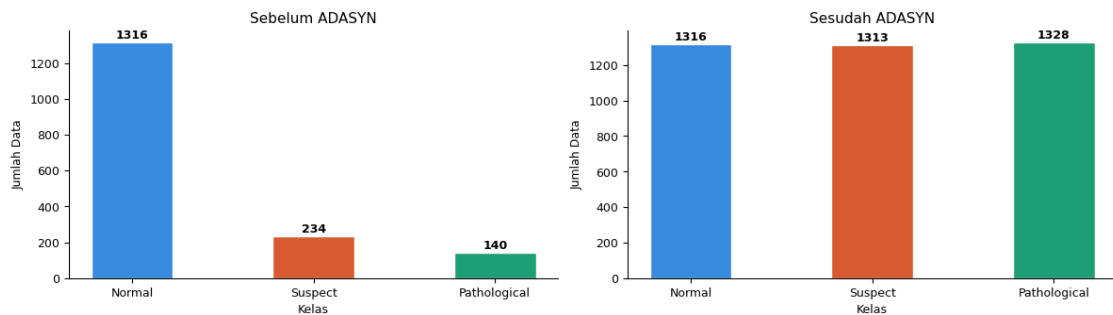


Gambar 3. Uji Mutual Information

Berdasarkan Gambar 3, sebagian besar fitur yang terpilih memiliki kesamaan dengan ANOVA, meskipun urutan kepentingannya berbeda, yang menunjukkan bahwa fitur tersebut berpengaruh baik secara linier maupun non-linear terhadap variabel target. Temuan ini menunjukkan bahwa hubungan antara fitur dan kondisi kesehatan janin tidak sepenuhnya bersifat linier, melainkan melibatkan interaksi kompleks antar variabel. Kemampuan Mutual Information dalam menangkap hubungan non-linear membuat model lebih sensitif terhadap pola, terutama pada kelas *Suspect* yang memiliki karakteristik *borderline*. Dengan demikian, penggunaan MI dapat meningkatkan pemisahan kelas yang sulit dibedakan dibandingkan pendekatan linier seperti ANOVA.

3.3 Penyeimbangan Data dengan ADASYN

Untuk melihat pengaruh penerapan ADASYN terhadap distribusi kelas, dilakukan visualisasi perbandingan jumlah data sebelum dan sesudah proses penyeimbangan. Perbandingan distribusi kelas tersebut disajikan pada Gambar 4.



Gambar 4. Distribusi kelas dengan ADASYN

Berdasarkan hasil visualisasi pada Gambar 4, sebelum penerapan ADASYN, terdapat ketidakseimbangan data dengan dominasi kelas mayoritas. Setelah diterapkan ADASYN, distribusi data menjadi lebih seimbang meskipun tidak sepenuhnya sama, dengan distribusi yang mendekati kelas mayoritas. Perubahan ini meningkatkan kemampuan model dalam mempelajari batas antar kelas, terutama pada kelas minoritas seperti *Suspect* dan *Pathological*. Dengan bertambahnya representasi data pada kelas minoritas, model menjadi lebih sensitif dalam mengenali pola pada kelas tersebut, sehingga berpotensi meningkatkan nilai *recall*. Namun, penambahan data sintetis juga dapat meningkatkan *overlap* antar kelas, yang berpotensi menurunkan akurasi keseluruhan. Dengan demikian, ADASYN mempengaruhi *trade-off* antara akurasi keseluruhan dan kemampuan deteksi kelas minoritas.

3.4 Evaluasi dan Perbandingan Semua Model dengan Seleksi Fitur

Tahap pemodelan dilakukan untuk membandingkan kinerja Random Forest dan XGBoost dalam klasifikasi kesehatan janin berbasis data CTG. Untuk meningkatkan kualitas input model, diterapkan teknik seleksi fitur ANOVA dan Mutual Information, serta ADASYN dan optimasi parameter dengan Random Search sehingga diperoleh empat kombinasi model. Hasil perbandingan akurasi empat model pengujian disajikan pada Tabel 3.

Tabel 3. Nilai Akurasi Model Klasifikasi

Model	Akurasi Model Klasifikasi		
	Tanpa ADASYN dan Random Search	Dengan ADASYN tanpa Random search	Dengan ADASYN dan Random Search
RF-ANOVA	94.80%	93.85%	93.85%
XGBoost-ANOVA	96.22%	95.51%	95.51%
RF-Mutual Information	92.91%	92.67%	91.96%
XGBoost-Mutual Information	92.91%	94.09%	94.33%

Seperti ditunjukkan pada Tabel 3, evaluasi dilakukan dalam tiga skenario: tanpa ADASYN dan tuning, dengan ADASYN tanpa tuning, serta kombinasi ADASYN dan tuning. Hasil menunjukkan bahwa XGBoost dengan seleksi fitur ANOVA mencapai akurasi tertinggi sebesar 96,22%, dengan *recall* kelas *Suspect* 81% dan *Pathological* 91%, yang mengindikasikan bahwa data awal cukup representatif. Penerapan ADASYN terbukti meningkatkan kemampuan model dalam mengklasifikasikan kelas minoritas, meskipun tidak selalu meningkatkan akurasi keseluruhan. Selain itu, hyperparameter tuning memberikan kontribusi positif terhadap performa model. Hal ini terlihat pada XGBoost dengan seleksi fitur Mutual Information yang mengalami peningkatan akurasi hingga 94,33% setelah penerapan ADASYN dan Random Search, yang mengindikasikan bahwa tuning membantu model dalam menyesuaikan kompleksitasnya terhadap distribusi data yang telah diubah, sehingga mampu meningkatkan kemampuan generalisasi terutama pada pola yang lebih kompleks. Secara keseluruhan, XGBoost menunjukkan performa lebih unggul dibandingkan Random Forest pada hampir seluruh skenario, karena keunggulan mekanisme *boosting* yang memungkinkan model untuk secara iteratif memperbaiki kesalahan prediksi, sehingga pendekatan *boosting* lebih efektif dibandingkan *bagging* dalam menangkap pola kompleks dan membedakan kelas dengan karakteristik yang saling berdekatan, seperti *Suspect*. Dengan mempertimbangkan keseimbangan akurasi, peningkatan *recall* *Suspect* 81% dan *Pathological* 97%, kombinasi XGBoost dengan Mutual Information, ADASYN, dan tuning menjadi pendekatan paling optimal untuk mendeteksi kelas *Suspect*.

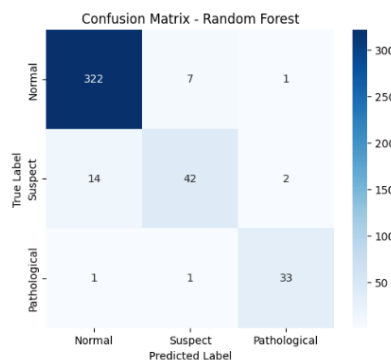
3.5 Evaluasi Confusion Matrix

Performa model dalam penelitian ini dinilai menggunakan klasifikasi yang dihasilkan dari *confusion matrix* yaitu *accuracy*, *precision*, *recall*, dan *F1-Score* serta melihat nilai *TP*, *TN*, *FP*, dan *FN*. Untuk mengetahui bagaimana proses

optimasi mempengaruhi performa algoritma, pengujian dilakukan pada beberapa model dengan konfigurasi dan nilai *hyperparameter* yang berbeda.

3.5.1 Evaluasi Random Forest dengan seleksi fitur ANOVA

Untuk menganalisis distribusi hasil prediksi model Random Forest dengan seleksi fitur ANOVA, digunakan *confusion matrix* yang disajikan pada Gambar 5. Secara umum, model menunjukkan kemampuan yang baik dalam mengklasifikasikan kelas Normal dan *Pathological*.



Gambar 5. Confusion Matrix Random Forest dengan ANOVA

Model Random Forest dengan ADASYN dan tuning berhasil mengidentifikasi kelas Normal dengan 322 prediksi akurat dan relatif sedikit kesalahan berdasarkan *confusion matrix* pada Gambar 5. Dengan 33 prediksi benar dan sangat sedikit kesalahan, model ini juga berkinerja cukup baik untuk kelas *Pathological*, yang menunjukkan bahwa kedua kelas tersebut memiliki pola yang lebih jelas dan mudah dikenali oleh model. Namun, pada kelas *Suspect*, model hanya menghasilkan 42 prediksi benar dan masih sering salah diklasifikasikan sebagai kelas Normal. Hal ini menunjukkan adanya *overlap* fitur antara kelas *Suspect* dan Normal, sehingga model kesulitan dalam membedakan kedua kelas tersebut. Dampak dari kondisi ini terlihat pada Tabel 3, di mana nilai *F1-Score* untuk kelas *Suspect* relatif lebih rendah yaitu sebesar 0,78 dibandingkan kelas lainnya.

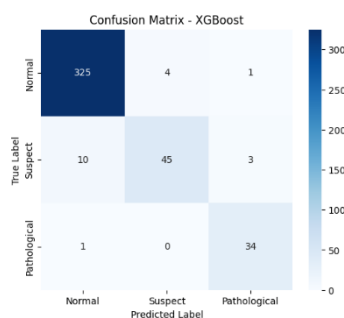
Tabel 3. Hasil Pemodelan Random Forest dengan ANOVA

	Precision	Recall	F1-Score	Support
0	0.96	0.98	0.97	330
1	0.84	0.72	0.78	58
2	0.92	0.94	0.93	35
accuracy			0.94	423
macro avg	0.90	0.88	0.89	423
weighted avg	0.94	0.94	0.94	423

Secara keseluruhan, model memiliki akurasi sebesar 0,94 berdasarkan Tabel 3. Meskipun nilai *recall* untuk kelas Normal dan *Pathological* masing-masing mencapai 0,98 dan 0,94, nilai recall untuk kelas *Suspect* hanya sebesar 0,72. Hal ini menegaskan bahwa kelas *Suspect* merupakan kelas yang paling sulit diklasifikasikan, karena memiliki karakteristik yang berada di antara dua kelas lainnya sehingga menimbulkan ambiguitas dalam proses prediksi.

3.5.2 Evaluasi XGBoost dengan seleksi fitur ANOVA

Berdasarkan *confusion matrix* pada Gambar 6, model XGBoost yang dikombinasi dengan ADASYN dan tuning menunjukkan performa klasifikasi yang sangat baik, dengan 325 data dalam kelas Normal diprediksi secara akurat dengan kesalahan yang sangat minim. Dengan 34 prediksi tepat dan hampir tanpa kesalahan, model ini menunjukkan kemampuannya yang akurat untuk mengidentifikasi keadaan janin berisiko dalam kelas *Pathological*.



Gambar 6. Confusion Matrix XGBoost dengan ANOVA

Pada Gambar 6, model berhasil mengidentifikasi 45 data dalam kelas *Suspect*, namun terdapat beberapa kesalahan, terutama pada titik data yang diproyeksikan sebagai Normal. Dibandingkan dengan model Random Forest, jumlah prediksi benar pada kelas *Suspect* mengalami peningkatan dari 0,72 menjadi 0,78, yang menunjukkan bahwa XGBoost lebih efektif mengenali pola kompleks pada kelas minoritas. Dalam konteks medis, kesalahan ini menjadi penting karena dapat berdampak pada keterlambatan deteksi kondisi janin yang berisiko.

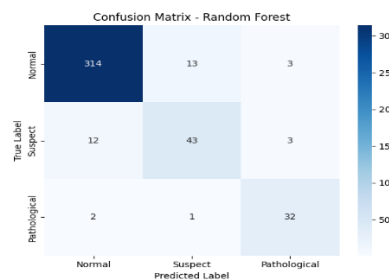
Tabel 4. Hasil Pemodelan XGBoost dengan ANOVA

	Precision	Recall	F1-Score	Support
0	0.97	0.98	0.98	330
1	0.92	0.78	0.84	58
2	0.89	0.97	0.93	35
accuracy			0.96	423
macro avg	0.93	0.91	0.92	423
weighted avg	0.95	0.96	0.95	423

Untuk memperkuat hasil tersebut, metrik evaluasi model disajikan pada Tabel 4. Berdasarkan tabel tersebut, model mencapai akurasi sebesar 0,96 dengan nilai *weighted average* sebesar 0,95, yang menunjukkan performa keseluruhan yang sangat baik. Selain itu, nilai *recall* untuk kelas *Pathological* sebesar 0,97 menunjukkan bahwa model sangat efektif dalam mengidentifikasi kondisi janin berisiko.

3.5.3 Evaluasi Random Forest dengan seleksi fitur Mutual Information

Random Forest dengan seleksi fitur Mutual Information yang dikombinasikan dengan ADASYN dan tuning pada Gambar 7 menunjukkan performa yang cukup baik. Sebanyak 314 data diklasifikasikan benar sebagai Normal, sementara sedikit kesalahan lainnya dianggap *Suspect* dan *Pathological*. Namun demikian, model menunjukkan kemampuan yang signifikan untuk mengenali kondisi janin berisiko dengan mengklasifikasikan 32 data benar sebagai *Pathological* dengan kesalahan yang relatif sedikit.



Gambar 7. Confusion Matrix Random Forest dengan Mutual Information

Berdasarkan Gambar 7, 43 data kelas *Suspect* berhasil diklasifikasikan dengan benar, namun masih terdapat beberapa kesalahan data yang diprediksi sebagai kelas Normal. Namun, jika dibandingkan dengan seleksi fitur ANOVA, *recall* Random Forest hanya menghasilkan 0.72. Sedangkan, Mutual information memberikan peningkatan kemampuan model mengenali kelas *Suspect* yang ditunjukkan oleh kenaikan recall sebesar 0.74 pada Tabel 5.

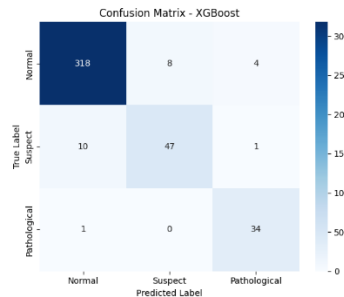
Tabel 5. Hasil Pemodelan Random Forest dengan Mutual Information

	Precision	Recall	F1-Score	Support
0	0.96	0.95	0.95	330
1	0.75	0.74	0.75	58
2	0.84	0.91	0.88	35
accuracy			0.92	423
macro avg	0.85	0.87	0.86	423
weighted avg	0.92	0.92	0.92	423

Berdasarkan Tabel 5, model mencapai akurasi sebesar 0,92 dengan nilai *weighted average* sebesar 0,92, yang menunjukkan performa model yang cukup stabil secara keseluruhan. Di sisi lain, penggunaan Mutual Information memberikan peningkatan pada kemampuan model dalam mengenali kelas *Suspect*. Hal ini dikarenakan Mutual Information dapat menangkap pola non-linear dibandingkan ANOVA, meskipun performanya masih belum optimal jika dibandingkan dengan kelas lainnya.

3.5.4 Evaluasi XGBoost dengan seleksi fitur Mutual Information

Menurut hasil pengujian model yang ditunjukkan pada Gambar 8, model XGBoost dengan seleksi fitur Mutual Information dan metode ADASYN dengan tuning menunjukkan jumlah prediksi akurat yang tinggi pada kelas Normal yaitu sebanyak 318 data berhasil diprediksi benar.



Gambar 8. Confusion Matrix XGBoost dengan Mutual Information

Untuk kelas *Suspect* pada Gambar 8, model mampu mengklasifikasikan 47 data dengan benar. Jika dibandingkan dengan XGBoost menggunakan seleksi fitur ANOVA, nilai recall pada kelas *Suspect* mengalami peningkatan dari 0,78 menjadi 0,81, yang menunjukkan bahwa Mutual Information memberikan kontribusi dalam meningkatkan kemampuan model dalam mengenali kelas minoritas.

Tabel 6. Hasil Pemodelan XGBoost dengan Mutual Information

	Precision	Recall	F1-Score	Support
0	0.97	0.96	0.97	330
1	0.85	0.81	0.83	58
2	0.87	0.97	0.92	35
accuracy			0.94	423
macro avg	0.90	0.92	0.91	423
weighted avg	0.94	0.94	0.94	423

Sementara itu berdasarkan hasil metrik pada Tabel 6, kelas *Pathological* memiliki nilai *recall* yang tetap tinggi dan stabil sebesar 0,97 pada kedua metode, yang mengindikasikan bahwa model telah mampu mengenali pola pada kelas ini dengan sangat baik.

3.6 Pembahasan

Penelitian ini menunjukkan bahwa Random Forest dan XGBoost mampu mengklasifikasikan kesehatan janin dengan baik pada kelas *Normal* dan *Pathological*, namun masih mengalami kesulitan pada kelas *Suspect* yang memiliki karakteristik data berada diantara kedua kelas tersebut. Secara umum, XGBoost menunjukkan performa lebih unggul dibandingkan Random Forest pada hampir setiap skenario pengujian, terutama dalam mendeteksi kelas minoritas, didukung oleh mekanisme boosting yang mampu memperbaiki kesalahan secara iteratif. Penerapan ADASYN meningkatkan kemampuan model dalam mengidentifikasi kelas minoritas, meskipun tidak selalu meningkatkan akurasi keseluruhan. Dalam konteks medis, peningkatan *recall* pada kelas *Suspect* dan *Pathological* menjadi penting untuk mendukung deteksi dini kondisi janin yang berpotensi mengalami komplikasi. Selain itu, seleksi fitur menggunakan ANOVA dan Mutual Information terbukti efektif, dengan Mutual Information lebih unggul dalam menangkap hubungan non-linear. Hal ini menunjukkan bahwa data CTG memiliki hubungan non-linear yang kompleks, sehingga pendekatan non-linear lebih efektif dalam meningkatkan performa model. Selain itu, kombinasi ADASYN dan seleksi fitur membantu model dalam mempelajari batas antar kelas minoritas yang sulit dibedakan. Berbeda dengan penelitian sebelumnya, studi ini mengintegrasikan seleksi fitur, penyeimbangan data, serta optimasi parameter dalam satu pendekatan, sehingga memberikan evaluasi yang lebih komprehensif terhadap kinerja model. Perbandingan dengan penelitian sebelumnya disajikan pada Tabel 7.

Tabel 7. Perbandingan dengan Beberapa Penelitian Sebelumnya

Penelitian Sebelumnya	Metode	Ruang Lingkup Penelitian	Akurasi
Nazli et al.[6]	LightGBM + SMOTE	Fetal Health	90.73%
Latif et al.[18]	RF + ADASYN	(CTG)	93.97%
Handayani dan Lailiah[19]	XGBoost + SMOTE		91.30%
Olayemi et al.[20]	Categorical Boosting + RFE		79.31%
Sulihati et al.[1]	Decision Tree + FFS		91.06%
Proposed Method	XGBoost + MI + ADASYN + Random Search		94.33%

Berdasarkan Tabel 7, metode yang diusulkan dalam penelitian ini, yaitu kombinasi XGBoost dengan seleksi fitur Mutual Information, ADASYN, dan Random Search, menghasilkan akurasi sebesar 94,33%, yang lebih tinggi dibandingkan metode LightGBM dengan SMOTE pada penelitian [6] sebesar 90,73%. Keunggulan tersebut dipengaruhi oleh kesesuaian metode dengan karakteristik dataset CTG yang tidak seimbang dan memiliki pola fitur kompleks, terutama pada kelas *Suspect*. Dalam kondisi ini, XGBoost mampu meningkatkan fokus pada data yang



sulit diklasifikasikan melalui mekanisme *boosting*, sehingga lebih efektif dalam mengenali kelas minoritas. Selain itu, ADASYN menghasilkan data sintetis secara adaptif pada area yang sulit dipelajari, sehingga membantu model dalam memahami batas antar kelas yang saling berdekatan, berbeda dengan SMOTE yang menghasilkan distribusi sintetis yang lebih merata tanpa mempertimbangkan kompleksitas data. Di sisi lain, Mutual Information mampu menangkap hubungan non-linear antara fitur dan kelas target, sehingga lebih efektif dalam merepresentasikan karakteristik kompleks sinyal detak jantung janin. Hal ini menjadi penting dalam membedakan kelas Suspect yang memiliki pola borderline. Kombinasi ketiga pendekatan ini menghasilkan performa klasifikasi yang lebih optimal dibandingkan metode sebelumnya.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi machine learning dalam analisis kardiotokografi (CTG) berpotensi meningkatkan objektivitas dan konsistensi dalam penilaian kesehatan janin, khususnya dalam mengidentifikasi kondisi berisiko sejak tahap awal. Berbeda dari pendekatan yang hanya berfokus pada akurasi global, penelitian ini menekankan pentingnya kemampuan model dalam mendeteksi kelas *Suspect*, yang secara klinis merepresentasikan kondisi transisi dan tahap awal gangguan kesehatan janin. Hasil penelitian mengindikasikan bahwa kombinasi XGBoost dengan seleksi fitur Mutual Information dan ADASYN merupakan pendekatan yang paling relevan untuk konteks klinis, karena mampu meningkatkan sensitivitas terhadap kelas *Suspect*. Peningkatan ini menunjukkan bahwa pendekatan berbasis hubungan non-linear lebih efektif dalam menangkap pola kompleks pada kondisi *borderline* yang seringkali sulit diidentifikasi melalui analisis konvensional. Dalam praktik medis, temuan ini memiliki implikasi penting. Klasifikasi *Suspect* tidak seharusnya dipandang sebagai ketidakpastian model, melainkan sebagai indikator kewaspadaan dini (*early warning*). Oleh karena itu, penggunaan sistem ini dapat mendukung proses triase awal dengan menandai pasien yang memerlukan pemantauan lebih intensif. Selain itu, hasil klasifikasi tersebut dapat mendorong dilakukannya evaluasi lanjutan, seperti pemeriksaan CTG berulang, USG, atau observasi klinis tambahan. Lebih lanjut, pemanfaatan model ini juga berpotensi mengurangi risiko keterlambatan intervensi dengan mengidentifikasi kondisi yang berpotensi berkembang menjadi patologis sejak tahap awal. Dengan demikian, model yang diusulkan tidak hanya berfungsi sebagai alat klasifikasi, tetapi sebagai sistem pendukung keputusan klinis yang proaktif, terutama dalam mendeteksi risiko pada tahap awal. Secara keseluruhan, penelitian ini menegaskan bahwa pendekatan yang mengutamakan sensitivitas terhadap kelas *Suspect* lebih sesuai untuk implementasi klinis dibandingkan pendekatan yang hanya mengejar akurasi tinggi. Untuk penelitian selanjutnya, pengembangan model dapat difokuskan pada integrasi data klinis tambahan dan eksplorasi metode yang lebih adaptif untuk semakin meningkatkan kemampuan deteksi kondisi borderline dalam praktik nyata.

REFERENCES

- [1] I. Sulihati, A. Syukur, and A. Marjuni, "Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2672.
- [2] M. Chen and Z. Yin, "Classification of Cardiotocography Based on the Apriori Algorithm and Multi-Model Ensemble Classifier," *Front. Cell Dev. Biol.*, vol. 10, May 2022, doi: 10.3389/fcell.2022.888859.
- [3] H. A. Zeini, D. Al-Jeznawi, H. Imran, L. F. A. Bernardo, Z. Al-Khafaji, and K. A. Ostrowski, "Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil," *Sustainability (Switzerland)*, vol. 15, no. 2, Jan. 2023, doi: 10.3390/su15021408.
- [4] G. Velarde *et al.*, "Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment," *Intelligent Systems with Applications*, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200354.
- [5] I. Campos, H. Gonçalves, J. Bernardes, and L. Castro, "Fetal Heart Rate Preprocessing Techniques: A Scoping Review," 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/bioengineering11040368.
- [6] 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, May 2025, doi: 10.3390/diagnostics15101250.
- [7] A. Ilham, T. A. P. Nagara, M. Kamaruddin, L. Khikmah, and T. Mantoro, "Fetal Health Risk Classification Using Important Feature Selection and Cart Model on Cardiotocography Data," *Informatika (Slovenia)*, vol. 49, no. 1, pp. 193–206, 2025, doi: 10.31449/inf.v49i1.5658.
- [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, Aug. 2023, doi: 10.3390/diagnostics13152471.
- [9] K. Wadhwa, R. Kumari, and A. Gosain, "Enhancing Model Performance in Hybrid Class Imbalance Techniques," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 288–297. doi: 10.1016/j.procs.2025.04.266.
- [10] T. N. Annisa, J. Jasmir, and N. Nurhadi, "Comparison of ANOVA and Chi-Square Feature Selection Methods to Improve Machine Learning Performance in Anemia Classification," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 1925–1940, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.5017.
- [11] S. Liu and M. Motani, "Improving Mutual Information based Feature Selection by Boosting Unique Relevance," Dec. 2022, [Online]. Available: <http://arxiv.org/abs/2212.06143>
- [12] F. Francis, S. Luz, H. Wu, S. J. Stock, and R. Townsend, "Machine learning on cardiotocography data to classify fetal outcomes: A scoping review," 2024, *Elsevier Ltd*. doi: 10.1016/j.compbio.2024.108220.
- [13] U. Sirisha *et al.*, "Dynamic Multi-Layer Perceptron for Fetal Health Classification Using Cardiotocography Data," *Computers, Materials and Continua*, vol. 80, no. 2, pp. 2301–2330, 2024, doi: 10.32604/cmc.2024.053132.



- [14] W. Widyastuty and M. A. Azis, "Classification and Evaluation of Sleep Disorders Using Random Forest Algorithm in Health and Lifestyle Dataset," *Compiler*, vol. 13, no. 1, p. 11, May 2024, doi: 10.28989/compiler.v13i1.2184.
- [15] D. Asmawati, L. Arif Sanjani, C. Dimas Renggana, C. Faticah, and T. Mustaqim, "Arrhythmia Classification with ECG Signal using Extreme Gradient Boosting (XGBoost) Algorithm," *Journal of Technology and Informatics (JoTI)*, vol. 6, no. 1, pp. 36–42, Oct. 2024, doi: 10.37802/joti.v6i1.792.
- [16] J. P. van Zyl and A. P. Engelbrecht, "Analysis of classification metric behaviour under class imbalance," *Egyptian Informatics Journal*, vol. 31, Sep. 2025, doi: 10.1016/j.eij.2025.100711.
- [17] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, vol. 27, no. 4S, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [18] A. Latif and Siti Khotimatul Wildah, "Optimization of Random Forest Model with SMOTE for Fetal Health Classification Based on Cardiotocography," *JURNAL TEKNOLOGI DAN OPEN SOURCE*, vol. 8, no. 1, pp. 390–397, Jun. 2025, doi: 10.36378/jtos.v8i1.4360.
- [19] K. Handayani *et al.*, "Comparison of XGboost, Extra Trees, and LightGBM with SMOTE for Fetal Health Classification," *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 3, 2024, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [20] O. C. Olayemi and O. O. Olasehinde, "Machine Learning Prediction of Fetal Health Status from Cardiotocography Examination in Developing Healthcare Contexts," *Journal of Computer Science Research*, vol. 6, no. 1, pp. 43–53, Mar. 2024, doi: 10.30564/jcsr.v6i1.6242.