

Optimization of Stunting Risk Prediction Using a Hybrid Genetic-Machine Learning Model

Agustina Heryati^{1*}, Terttiaavini², Dona Marcelina³, Harsi Romli⁴

^{1,2} Sistem Informasi, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, Palembang, 30129, Indonesia

³ Magister Ilmu Komputer, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, Palembang, 30129, Indonesia

⁴ Akuntansi, Fakkultas Ekonomi, Universitas Indo Global Mandiri, Palembang, 30129, Indonesia

Informasi Artikel

Diterima : 20 Mei 2025
Revisi : 15 Juni 2025
Publikasi : 20 Juni 2025

Kata Kunci:

Model Hybrid
Stunting
Machine Learning
Genetic Algorithm
XGBoost
Prediksi Risiko

ABSTRAK

Stunting merupakan masalah gizi kronis yang masih menjadi isu prioritas nasional di Indonesia. Berdasarkan Survei Status Gizi Indonesia (SSGI) 2022, prevalensi stunting nasional mencapai 21,6%, dengan target penurunan menjadi 14% pada tahun 2024. Prediksi risiko stunting yang akurat masih menjadi tantangan, terutama di wilayah seperti Kota Palembang yang memiliki keragaman sosial-ekonomi dan karakteristik antropometrik yang kompleks. Penelitian ini mengembangkan model hybrid machine learning untuk prediksi risiko stunting, yang mengintegrasikan algoritma klasifikasi dengan metode optimasi Genetic Algorithm (GA) dalam seleksi fitur. Pendekatan hybrid ini bertujuan meningkatkan akurasi dan efisiensi model prediktif berbasis data numerik dan sosial-ekonomi. Sebanyak 6.000 sampel digunakan, dan setelah melalui proses preprocessing (trimming, winsorization, normalisasi, dan SMOTE), diperoleh 5.366 data bersih. Empat algoritma klasifikasi diuji, yaitu Decision Tree, K-Nearest Neighbor, Random Forest, dan XGBoost. Hasil terbaik diperoleh dari model XGBoost dengan akurasi 84,08%, recall 93%, dan F1-score 0,91 untuk kelas mayoritas. Dengan integrasi Genetic Algorithm, akurasi optimal mencapai 95,34% pada generasi ke-3 dalam proses seleksi fitur. Penelitian ini memberikan kontribusi berupa kerangka kerja prediktif berbasis hybrid machine learning yang dapat diadopsi oleh instansi kesehatan daerah untuk deteksi dini risiko stunting secara lebih tepat sasaran.

ABSTRACT

Stunting is a chronic nutritional problem that remains a national priority issue in Indonesia. According to the 2022 Indonesian Nutrition Status Survey (SSGI), the national stunting prevalence reached 21.6%, with a target reduction to 14% by 2024. Accurate prediction of stunting risk remains a challenge, particularly in regions like Palembang City, which exhibit diverse socio-economic conditions and complex anthropometric characteristics. This study develops a *hybrid machine learning* model for stunting risk prediction by integrating classification algorithms with a Genetic Algorithm (GA) for feature selection. The hybrid approach aims to enhance predictive accuracy and efficiency based on numerical and socio-economic data. A total of 6,000 samples were used, and after preprocessing (trimming, winsorization, normalization, and SMOTE), 5,366 clean data samples were obtained. Four classification algorithms were tested: Decision Tree, K-Nearest Neighbor, Random Forest, and XGBoost. The best performance was achieved by the XGBoost model, with an accuracy of 84.08%, recall of 93%, and F1-score of 0.91 for the majority class. By integrating the Genetic Algorithm, optimal accuracy reached 95.34% in the third generation of feature selection. This study contributes a *hybrid machine learning*-based predictive framework that can be adopted by local health institutions for more targeted early detection of stunting risk.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



* Penulis Koresponden

Email: agustina.heryati@uigm.ac.id

Cara sitasi IEEE::

A. Heryati, Terttiaavini, D. Marcelina, dan H. Romli, "Optimization of Stunting Risk Prediction Using a Hybrid Genetic-Machine Learning Model," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, p. 807-815, Bulan tahun. doi:10.30811/jaise.v5i2.6874

1. PENDAHULUAN

Stunting merupakan salah satu masalah kesehatan masyarakat yang hingga kini menjadi tantangan besar di Indonesia. Berdasarkan laporan Survei Status Gizi Indonesia (SSGI) tahun 2022 yang dirilis oleh Kementerian Kesehatan Republik Indonesia, prevalensi stunting nasional berada pada angka 21,6% dan ditargetkan turun hingga 14% pada tahun 2024 [1], [2]. Angka tersebut menunjukkan bahwa meskipun telah terjadi penurunan, stunting masih menjadi isu strategis nasional, khususnya karena dampaknya yang multidimensional terhadap pertumbuhan fisik, kognitif, dan produktivitas jangka panjang anak. Di Provinsi Sumatera Selatan, stunting masih menunjukkan prevalensi yang mengkhawatirkan, khususnya di kota-kota besar seperti Palembang. Hal ini diperparah oleh keterbatasan sistem pemantauan dan prediksi yang akurat untuk mengidentifikasi individu atau wilayah dengan risiko tinggi [3]. Kondisi ini menuntut pengembangan sistem prediksi yang tidak hanya andal secara statistik, tetapi juga mampu menangani kompleksitas dan heterogenitas data sosial-ekonomi yang mempengaruhi status gizi anak [4], [5].

Objek dalam penelitian ini adalah data stunting dari Kota Palembang yang mencakup variabel demografis dan antropometrik seperti *age* (usia), *birth weight* (berat badan lahir), *birth length* (panjang badan lahir), *body weight* (berat badan saat pengukuran), dan *body length* (panjang badan saat pengukuran). saat pengukuran [6],[7]. Data tersebut digunakan untuk mengembangkan model prediksi yang dapat mengklasifikasikan risiko stunting berdasarkan fitur-fitur numerik tersebut. Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana membangun model prediksi risiko stunting yang akurat dan efisien, dengan memperhitungkan tantangan seperti ketidakseimbangan data, outlier, dan relevansi fitur [8]. Tujuan dari penelitian ini adalah untuk mengoptimalkan prediksi risiko stunting melalui pengembangan model *hybrid machine learning* yang mengintegrasikan algoritma genetika (Genetic Algorithm/GA) untuk seleksi fitur dan model klasifikasi seperti Random Forest, XGBoost, Decision Tree, dan K-Nearest Neighbors (KNN) [9]. Tujuan khusus mencakup (1) membersihkan dan menyeimbangkan dataset, (2) menyeleksi fitur optimal menggunakan GA, dan (3) mengevaluasi kinerja model klasifikasi terhadap akurasi, precision, recall, dan F1-score [10]. Aplikasi yang digunakan untuk membangun model adalah google Colab menggunakan bahasa pemrograman Python. Dengan pendekatan ini, diharapkan sistem prediksi yang dihasilkan mampu memberikan dukungan keputusan yang lebih tepat bagi para pemangku kebijakan dalam upaya intervensi stunting [11].

Metode yang digunakan dalam penelitian ini terdiri dari lima tahap utama. Pertama, data preprocessing dilakukan melalui pembersihan data, penanganan outlier menggunakan teknik Winsorization dan Trimming, serta normalisasi data dengan Min-Max Scaling [12] [13]. Kedua, dilakukan pembagian dataset menjadi data latih dan uji dengan rasio 80:20. Ketiga, diterapkan teknik SMOTE (*Synthetic Minority Oversampling Technique*) untuk menangani ketidakseimbangan kelas. Keempat, seleksi fitur dilakukan menggunakan algoritma genetika, yang dikenal mampu mengeksplorasi ruang solusi nonlinier secara efisien untuk mencari subset fitur yang paling relevan. Terakhir, pemodelan dilakukan dengan empat algoritma populer: Decision Tree, KNN, Random Forest, dan XGBoost [14]. Pemilihan metode genetic algorithm dalam seleksi fitur didasarkan pada keunggulannya dalam mengidentifikasi fitur-fitur penting tanpa bergantung pada asumsi linieritas, serta kemampuannya untuk menghindari overfitting dan meningkatkan generalisasi model. Luo dan Xu (2021) menunjukkan bahwa Genetic Algorithm sangat efektif dalam menyaring fitur-fitur yang relevan untuk pemodelan prediktif, bahkan pada data berdimensi tinggi dan kompleksitas tinggi.

Sementara itu, algoritma XGBoost yang dikembangkan oleh Chen dan Guestrin (2016) dipilih karena merupakan metode boosting yang sangat efisien, akurat, dan telah terbukti unggul dalam berbagai kompetisi data mining dan penerapan nyata. XGBoost dirancang untuk menangani missing values, ketidakseimbangan data, serta memberikan kontrol yang fleksibel atas overfitting melalui parameter regularisasi, sehingga cocok untuk prediksi status gizi seperti stunting [15]. Kontribusi penelitian ini bersifat teoritis dan praktis. Dari sisi teoritis, penelitian ini memberikan pendekatan baru dalam penggabungan genetic algorithm dengan model machine learning untuk masalah prediksi stunting yang kompleks. Dari sisi praktis, penelitian ini menyediakan kerangka kerja (framework) untuk membangun sistem prediksi risiko stunting berbasis data yang dapat diterapkan oleh instansi kesehatan daerah, terutama di wilayah dengan keterbatasan sumber daya. Selain itu,

hasil penelitian ini dapat menjadi referensi bagi pengambil kebijakan dalam merancang intervensi berbasis bukti (evidence-based) guna menurunkan prevalensi stunting secara lebih tepat sasaran.

Berdasarkan beberapa penelitian sebelumnya, pendekatan prediktif terhadap masalah kesehatan masyarakat seperti stunting semakin banyak dikembangkan dengan memanfaatkan teknologi machine learning dan pemrosesan data berbasis fitur yang kompleks. WHO (2021) menegaskan bahwa prevalensi malnutrisi anak, termasuk stunting, masih menjadi tantangan global yang memerlukan solusi berbasis bukti dan inovasi teknologi dalam pemantauan dan intervensi [16]. Di tingkat nasional, data dari Survei Status Gizi Indonesia (SSGI) 2022 menunjukkan bahwa stunting masih menjadi masalah utama di banyak wilayah, termasuk Provinsi Sumatera Selatan, sehingga mendukung urgensi pengembangan sistem prediksi risiko secara tepat [17]. Sejalan dengan itu, Jaber et al. (2020) menunjukkan bahwa penerapan algoritma machine learning dalam kesehatan masyarakat mampu meningkatkan efektivitas prediksi dan deteksi dini berbagai kondisi kesehatan, termasuk dalam menangani isu-isu gizi [18], [19]. Lebih lanjut, Luo dan Xu (2021) mengungkapkan bahwa algoritma genetika (Genetic Algorithm) efektif dalam menyeleksi fitur-fitur relevan dari data berukuran besar dan berdimensi tinggi, sehingga menghasilkan model prediktif yang lebih efisien dan akurat [20], [21]. Dalam mendukung performa klasifikasi, algoritma XGBoost yang dikembangkan oleh Chen dan Guestrin (2016) telah terbukti sebagai pendekatan boosting yang sangat efisien, stabil, dan mampu menangani permasalahan data tidak seimbang yang umum terjadi dalam kasus kesehatan masyarakat. Oleh karena itu, integrasi antara metode seleksi fitur berbasis Genetic Algorithm dan pemodelan klasifikasi dengan XGBoost menjadi kombinasi yang menjanjikan dalam membangun sistem prediksi risiko stunting yang cerdas, akurat, dan dapat diterapkan pada berbagai konteks data sosial-ekonomi dan antropometrik. Dengan latar belakang yang kuat secara konseptual dan empiris ini, penelitian diharapkan dapat memberikan kontribusi signifikan terhadap literatur ilmiah dan implementasi nyata dalam bidang kesehatan masyarakat, khususnya terkait upaya pencegahan stunting secara sistematis dan berbasis teknologi.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode machine learning berbasis supervised classification yang dikombinasikan dengan algoritma genetika untuk proses seleksi fitur, guna mengoptimalkan akurasi prediksi risiko stunting. Adapun tahapan penelitian ini dijelaskan secara sistematis sebagai berikut.

2.1. Persiapan dan Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari data stunting Kota Palembang tahun 2024. Data tersebut telah melalui proses pembersihan dan validasi, sehingga tidak ditemukan adanya nilai yang hilang (*missing value*). Hal ini menunjukkan bahwa kualitas data cukup baik untuk digunakan dalam analisis lebih lanjut. Data tersebut disajikan dalam bentuk statistik deskriptif yang ditampilkan pada Tabel 1.

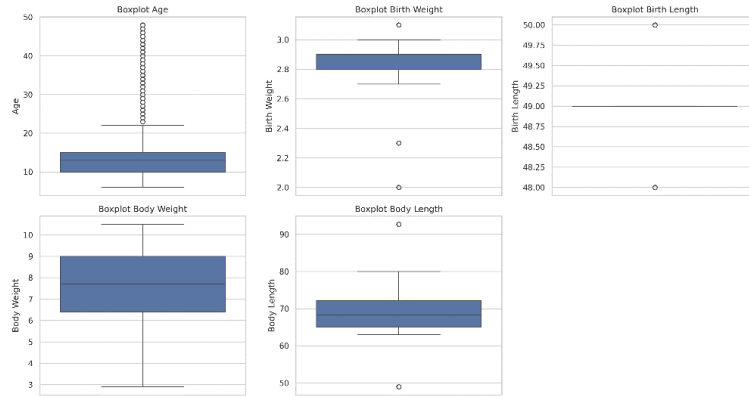
Tabel 1. Statistik Deskriptif data Stunting kota Palembang tahun 2024

	Age	Birth Weight	Birth Length	Body Weight	Body Length
count	6000	6000	6000	6000	6000
mean	14.255	2.7576	49.1174	7.63199	68.8549
std	7.923285	0.291092	0.439814	1.720461	8.90664
min	6	2	48	2.9	49
25%	10	2.8	49	6.4	65
50%	13	2.8	49	7.7	68.3
75%	15	2.9	49	9	72.2
max	48	3.1	50	10.5	92.7

Tabel 1 ini menyajikan statistik deskriptif dari lima variabel yang terkait dengan data stunting, yaitu *age*, *birth weight*, *birth length*, *body weight*, dan *body length*. Data terdiri dari 6.000 sampel untuk setiap variabel, yang mencerminkan populasi anak-anak yang diamati di Kota Palembang pada tahun 2024. Rata-rata usia anak dalam dataset ini adalah 14,26 bulan, dengan rentang usia dari 6 hingga 48 bulan, dan standar deviasi sebesar 7,92 bulan, yang menunjukkan adanya variasi usia yang cukup luas dalam sampel. Berat badan lahir berkisar antara 2,0 hingga 3,1 kg dengan rata-rata 2,76 kg, yang masih dalam batas normal, namun nilai minimum menunjukkan adanya kasus bayi lahir dengan berat rendah (< 2,5 kg) yang merupakan faktor risiko penting. Selain itu, nilai kuartil (25%, 50%, dan 75%) menunjukkan distribusi data yang cenderung normal untuk sebagian besar variabel, terutama *Birth Length* yang konsisten pada nilai 49 cm. Hal ini mengindikasikan bahwa variabel tersebut memiliki sebaran data yang relatif seragam dan dapat digunakan sebagai indikator yang stabil dalam analisis prediktif. Variasi yang signifikan pada variabel *age* dan *body length* perlu mendapat perhatian khusus karena dapat memengaruhi sensitivitas model prediksi, terutama dalam mengidentifikasi kelompok anak dengan risiko tinggi stunting. Analisis ini menjadi dasar penting dalam proses eksplorasi dan validasi model machine learning selanjutnya.

2.2 Eksplorasi dan Pemahaman Data

Hasil eksplorasi awal menunjukkan adanya nilai pencilan (outlier) pada beberapa variabel penting. Berdasarkan analisis tersebut, ditemukan sebanyak 754 outlier pada variabel *Age*, 2.332 outlier pada variabel *Birth Weight*, 2.072 outlier pada *Birth Length*, dan 1.287 outlier pada *Body Length*. Keberadaan outlier ini perlu ditangani secara tepat karena dapat memengaruhi akurasi dan validitas hasil analisis, khususnya dalam proses pengembangan model prediksi stunting [22]. Visualisasi keberadaan outlier pada masing-masing fitur ditampilkan dalam Gambar 1 melalui diagram boxplot yang memperlihatkan sebaran data dan titik-titik ekstrem yang berada di luar ambang batas normal.



Gambar 1. Visualisasi Outlier pada data Stunting

2.3. Penanganan Outlier dan Pembersihan Data

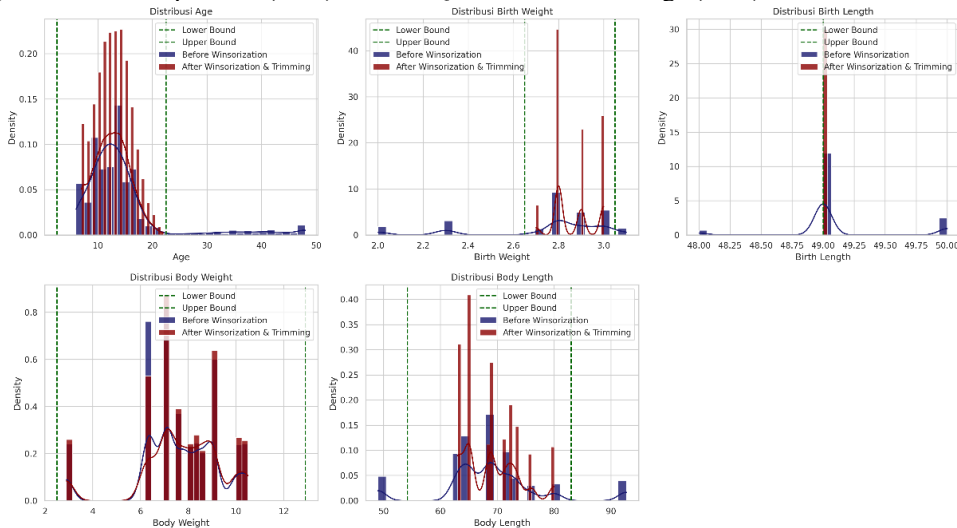
Pada tahap ini, dilakukan penanganan terhadap data outlier sebagai bagian dari proses pembersihan data untuk meningkatkan kualitas data dan akurasi model prediktif. Pembersihan outlier menggunakan teknik Winsorization dan Trimming, dengan rumus identifikasi batas outlier berdasarkan metode *Interquartile Range* (IQR) sebagai berikut:

$$IQR = Q3 - Q1 \tag{1}$$

$$lower_bound = Q1 - 1.5 * IQR \tag{2}$$

$$upper_bound = Q3 + 1.5 * IQR \tag{3}$$

dimana Q1 adalah kuartil pertama (25%) dan Q3 adalah kuartil ketiga (75%) dari distribusi data.



Gambar 2. Perbandingan kurva densitas sebelum dan sesudah penanganan outlier dengan teknik Winsorization dan Trimming

Pada proses Winsorization, nilai-nilai ekstrem yang berada di bawah persentil ke-5 dan di atas persentil ke-95 disesuaikan menjadi nilai pada persentil ke-5 dan ke-95. Ini bertujuan untuk mengurangi pengaruh nilai ekstrem tanpa menghapus data. Selanjutnya dilakukan Trimming, yaitu menghapus baris data

yang memiliki nilai di luar batas bawah dan atas yang telah dihitung berdasarkan rumus IQR di atas. Gabungan kedua metode ini bertujuan untuk membersihkan data dari outlier sekaligus menjaga integritas dan keseimbangan distribusi data yang akan digunakan dalam analisis lebih lanjut. Gambar 2 menunjukkan perbandingan kurva densitas (*density plot*) antara data sebelum dan sesudah dilakukan penanganan outlier menggunakan teknik Winsorization dan Trimming [23].

2.4. Transformasi dan Normalisasi Data

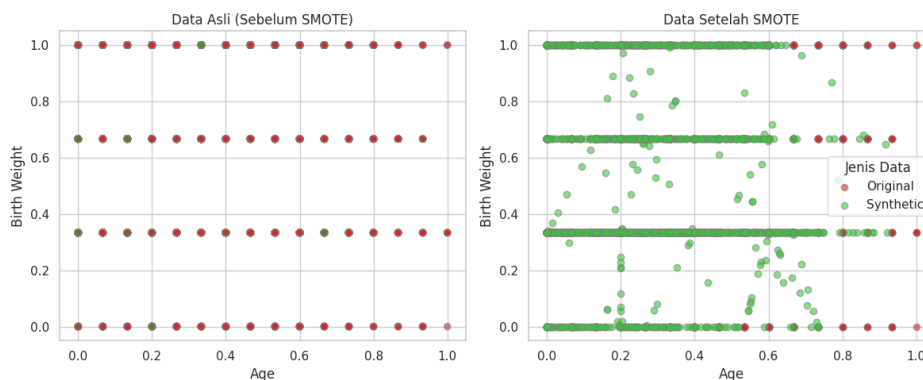
Transformasi dan normalisasi data dilakukan untuk memastikan bahwa fitur numerik memiliki skala yang seragam dan memenuhi persyaratan pemodelan. Proses ini hanya diterapkan pada fitur numerik yang relevan, sedangkan variabel kategorikal seperti jenis kelamin tidak dimasukkan dalam analisis. Normalisasi menggunakan metode Min-Max Scaling diterapkan untuk mengubah nilai setiap fitur numerik ke dalam rentang antara 0 hingga 1. Metode ini penting guna menghindari bias yang disebabkan oleh perbedaan skala antar fitur dalam dataset [24]. Rumus Min-Max Scaler secara umum adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Dimana X adalah nilai asli dari fitur; X_{min} adalah nilai minimum dari fitur tersebut dalam dataset; X_{max} adalah nilai maksimum dari fitur tersebut dalam dataset; X' adalah nilai fitur yang sudah dinormalisasi ke rentang 0 sampai 1. Hasil dari proses trimming dan normalisasi kemudian disimpan dalam file baru dengan nama *data_preprocessed.xlsx*. Penyimpanan data yang telah diproses ini bertujuan untuk menyediakan dataset akhir yang siap digunakan dalam tahap pemodelan selanjutnya. Dengan demikian, data yang digunakan pada tahap berikutnya memiliki konsistensi skala yang memadai, sehingga mendukung performa dan keandalan model prediksi.

2.5. Pembagian Data dan Penyeimbangan Dataset

Dataset yang telah melalui proses winsorization dan trimming berjumlah 5.366 data. Data tersebut dipisahkan menjadi fitur (x) yang terdiri dari variabel *age*, *birth weight*, *birth length*, *body weight*, dan *body length*, serta target (y) yaitu variabel Stunting. Pembagian dataset dilakukan dengan rasio 80:20, dimana 4.292 data digunakan sebagai data latih dan 1.074 data sebagai data uji. Pembagian ini memastikan model dapat dilatih secara optimal dan diuji pada data yang belum pernah dilihat untuk evaluasi yang valid. Untuk mengatasi ketidakseimbangan kelas target pada data latih, diterapkan metode SMOTE (*Synthetic Minority Oversampling Technique*). SMOTE menghasilkan sampel sintetis pada kelas minoritas sehingga distribusi kelas menjadi seimbang dan model mampu mempelajari pola dari kedua kelas secara efektif. Gambar 3 merupakan visualisasi scatter plot yang menunjukkan distribusi data sebelum dan setelah penerapan teknik SMOTE untuk menangani ketidakseimbangan kelas pada dataset stunting.



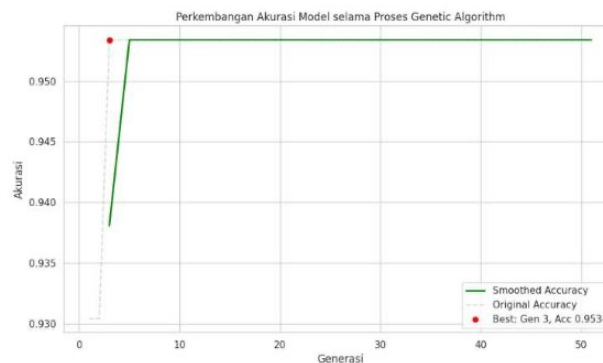
Gambar 3. Visualisasi scatter plot untuk distribusi data sebelum dan setelah penerapan teknik SMOTE

Data asli sebelum SMOTE pada grafik kiri menunjukkan distribusi titik merah yang tidak merata, terutama pada kelas minoritas seperti berat lahir rendah atau stunting, sehingga terjadi ketidakseimbangan kelas yang dapat menyebabkan model prediksi menjadi bias dan kurang akurat. Setelah SMOTE, grafik kanan memperlihatkan tambahan titik hijau berupa data sintetis yang tersebar merata di area kurang padat, sehingga distribusi kelas menjadi seimbang. Penyeimbangan ini memungkinkan model machine learning belajar lebih efektif dari kedua kelas, meningkatkan akurasi prediksi risiko stunting yang dipengaruhi oleh variabel *Age* dan *Birth Weight*.

2.6. Pemodelan dan Seleksi Fitur dengan Algoritma Genetic Algorithm (GA)

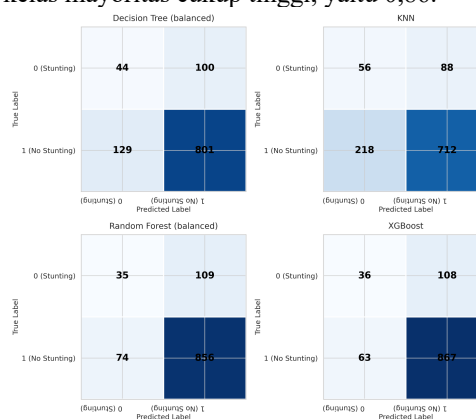
Dataset yang telah melalui proses winsorization dan trimming berjumlah 5.366 data, dengan lima fitur terpilih, yaitu *age*, *birth weight*, *birth length*, *body weight*, dan *body length*. Data tersebut kemudian dibagi menjadi data latih sebanyak 4.292 sampel dan data uji sebanyak 1.074 sampel dengan rasio 80:20. Pembagian ini bertujuan agar model dapat dilatih secara optimal sekaligus diuji pada data yang belum pernah dilihat, sehingga evaluasi performa menjadi lebih valid. Untuk mengatasi ketidakseimbangan kelas pada data latih, diterapkan metode SMOTE yang menghasilkan sampel sintetis pada kelas minoritas. Dengan demikian, distribusi kelas menjadi seimbang, memungkinkan model mempelajari pola dari kedua kelas secara efektif.

Pada tahap pemodelan dan seleksi fitur menggunakan Algoritma Genetic Algorithm, dua model yang dievaluasi secara mendalam adalah Random Forest dan XGBoost. Kedua model ini dilatih secara iteratif sehingga memungkinkan pemantauan performa selama proses pelatihan. Oleh karena itu, grafik akurasi (GA) selama training disajikan untuk memvisualisasikan perkembangan kemampuan model dalam mengenali pola data secara bertahap, sehingga membantu mengidentifikasi apakah model mengalami overfitting atau underfitting. Gambar 4 merupakan menyajikan grafik akurasi (GA) yang menunjukkan performa Random Forest dan XGBoost selama proses pelatihan secara bertahap.



Gambar 4. Grafik akurasi (GA) untuk Performa Random Forest dan XGBoost

Sementara itu, model Decision Tree (dengan `class_weight='balanced'`) dan K-Nearest Neighbors (KNN) dievaluasi menggunakan metrik akhir dan visualisasi confusion matrix, karena model tersebut tidak menggunakan proses pelatihan bertahap yang memungkinkan pembuatan grafik akurasi. Model Decision Tree mencapai akurasi sebesar 78,86% pada data uji. Dari confusion matrix, model berhasil mengklasifikasikan 803 sampel stunting dengan benar (true positive), namun terdapat 127 kasus stunting yang salah diklasifikasikan sebagai non-stunting (false negative). Pada kelas non-stunting, 44 sampel terdeteksi benar (true negative) dan 100 sampel salah diklasifikasikan (false positive). Precision dan recall pada kelas minoritas masing-masing sebesar 0,26 dan 0,31, yang menunjukkan tantangan dalam mendeteksi kelas tersebut, meskipun recall pada kelas mayoritas cukup tinggi, yaitu 0,86.



Gambar 5. Visualisasi Confusion Matrix untuk seluruh Model.

Model KNN menghasilkan akurasi 71,51%. Confusion matrix menunjukkan 712 prediksi stunting yang benar dan 218 yang salah (false negative), serta 56 prediksi non-stunting yang benar dan 88 yang salah (false positive). Meskipun recall pada kelas minoritas lebih baik (0,39), precision tetap rendah (0,20), mengindikasikan kesulitan model dalam mengidentifikasi kelas minoritas secara presisi. Hasil confusion matrix div visualisasikan dalam bentuk gambar untuk memberikan gambaran yang lebih jelas mengenai performa klasifikasi masing-masing model, khususnya dalam membedakan antara kelas stunting dan non-

stunting berdasarkan prediksi yang benar (true positive, true negative) maupun kesalahan prediksi (false positive, false negative) [25]. Gambar 5 menampilkan visualisasi confusion matrix untuk seluruh model.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Permodelan dengan Genetic Algorithm

Seleksi fitur merupakan tahap krusial dalam pengembangan model klasifikasi karena secara langsung memengaruhi performa dan kompleksitas model. Dalam studi ini, proses seleksi fitur dilakukan menggunakan Genetic Algorithm (GA), yaitu metode berbasis evolusi biologis yang bertujuan untuk mengidentifikasi subset fitur paling optimal secara iteratif.

Hasil evolusi yang diperoleh dari 50 generasi menunjukkan bahwa peningkatan akurasi paling signifikan terjadi pada generasi awal. Nilai akurasi tertinggi tercapai pada generasi ke-3 sebesar 95,34%, yang kemudian mengalami stagnasi hingga generasi ke-50. Fenomena ini menunjukkan bahwa populasi dalam proses evolusi telah mencapai titik konvergensi, di mana individu-individu terbaik tidak mengalami perubahan yang berarti.

Kemampuan GA dalam mencapai konvergensi pada generasi awal mengindikasikan efisiensi metode ini dalam mengidentifikasi kombinasi fitur optimal tanpa memerlukan iterasi berlebih. Dengan demikian, dapat disimpulkan bahwa Genetic Algorithm tidak hanya efektif dalam menyaring fitur-fitur yang relevan, tetapi juga efisien dari sisi komputasi. Efektivitas ini mendukung pemanfaatan GA sebagai strategi seleksi fitur dalam pengembangan model klasifikasi berbasis machine learning.

3.2. Hasil Evaluasi Model Klasifikasi

Untuk mengevaluasi pengaruh seleksi fitur terhadap kinerja model klasifikasi, dilakukan pengujian terhadap empat algoritma populer, yaitu: Decision Tree, K-Nearest Neighbors (KNN), Random Forest, dan Extreme Gradient Boosting (XGBoost). Proses pelatihan dilakukan pada data yang telah mengalami seleksi fitur dengan Genetic Algorithm. Evaluasi performa dilakukan dengan menggunakan metrik akurasi, precision, recall, dan F1-score, yang dihitung untuk masing-masing kelas (kelas minoritas/0 dan mayoritas/1). Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model pada Data Setelah Seleksi Fitur

Model	Accuracy	Precision (Class 0)	Recall (Class 0)	F1-Score (Class 0)	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)
Decision Tree	0.7868	0.25	0.31	0.28	0.89	0.86	0.87
KNN	0.7151	0.20	0.39	0.27	0.89	0.77	0.82
Random Forest	0.8296	0.32	0.24	0.28	0.89	0.92	0.90
XGBoost	0.8408	0.36	0.25	0.30	0.89	0.93	0.91

Model XGBoost menunjukkan kinerja terbaik dengan akurasi sebesar 84,08%, serta nilai F1-score tertinggi untuk kelas mayoritas (kelas 1) sebesar 0.91. Meskipun nilai recall dan F1-score pada kelas minoritas (kelas 0) masih tergolong rendah, model ini tetap unggul dalam hal stabilitas dan keseimbangan performa. Nilai precision pada kelas mayoritas konsisten sebesar 0.89 di semua model, menunjukkan bahwa keempat algoritma memiliki kemampuan serupa dalam mendeteksi instance positif. Namun, perbedaan signifikan terletak pada penanganan instance kelas minoritas, yang memengaruhi nilai recall dan F1-score pada kelas tersebut.

3.3. Pembahasan

Hasil eksperimen menunjukkan bahwa pendekatan seleksi fitur berbasis Genetic Algorithm mampu meningkatkan kinerja model klasifikasi secara signifikan. Konvergensi yang tercapai pada generasi awal mencerminkan efisiensi algoritma dalam menemukan kombinasi fitur paling relevan terhadap target klasifikasi, serta mengurangi kompleksitas model. Peningkatan akurasi model setelah dilakukan seleksi fitur menunjukkan bahwa fitur-fitur irrelevant atau redundan telah berhasil dieliminasi, sehingga proses pembelajaran model menjadi lebih fokus. Hal ini mendukung temuan bahwa proses seleksi fitur memiliki dampak positif terhadap generalisasi dan efektivitas prediksi.

Model XGBoost menunjukkan keunggulan dibandingkan model lain, yang sejalan dengan sifat algoritma tersebut sebagai metode ensemble boosting yang adaptif dalam memperbaiki kesalahan iteratif. Kemampuan XGBoost dalam mengelola informasi penting dari fitur yang telah terseleksi menjadikan algoritma ini sangat cocok untuk kasus klasifikasi kompleks. Namun demikian, performa model terhadap kelas minoritas masih menunjukkan kelemahan, terutama pada metrik recall dan F1-score. Ketidakseimbangan distribusi kelas (class imbalance) berdampak pada sensitivitas model dalam mengenali instance dari kelas tersebut. Oleh karena itu, penerapan teknik penanganan imbalance seperti SMOTE (*Synthetic Minority*

Oversampling Technique) atau penyesuaian bobot kelas (class weighting) perlu dipertimbangkan dalam penelitian lanjutan.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi risiko stunting dengan memanfaatkan pendekatan hybrid machine learning yang mengintegrasikan berbagai algoritma seperti Random Forest, Gradient Boosting, dan XGBoost. Model ini dibangun berdasarkan data sosial-ekonomi yang komprehensif dari Provinsi Sumatera Selatan, termasuk variabel-variabel seperti pendapatan rumah tangga, pendidikan orang tua, akses terhadap layanan kesehatan, serta infrastruktur wilayah. Proses feature engineering yang cermat dan seleksi variabel yang tepat terbukti meningkatkan akurasi model secara signifikan. Hasil evaluasi menunjukkan bahwa model hybrid memiliki performa prediksi yang lebih baik dibandingkan pendekatan tunggal, dengan tingkat akurasi yang tinggi dan kemampuan generalisasi yang baik. Temuan ini menunjukkan bahwa penggabungan data multidimensi dan algoritma cerdas mampu menghasilkan sistem pendukung keputusan yang efektif untuk mengidentifikasi wilayah atau kelompok masyarakat dengan risiko tinggi stunting. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam upaya mitigasi stunting di tingkat regional. Model yang dihasilkan dapat digunakan oleh pemerintah daerah dan pemangku kepentingan untuk menyusun kebijakan yang lebih tepat sasaran, berbasis data, dan bersifat preventif.

REFERENSI

- [1] Z. Wardani, D. Sukandar, Y. F. Baliwati, and H. Riyadi, "Sebuah alternatif: indeks stunting sebagai evaluasi kebijakan intervensi balita stunting di Indonesia," *Gizi Indonesia*, 2021, [Online]. Available: https://www.persagi.org/ejournal/index.php/Gizi_Indon/article/view/535
- [2] K. M. Amin, "Bahas Percepatan Penurunan Angka Prevalensi Stunting 14% pada 2024, Para Pemangku Kepentingan Pusat dan Daerah Berkumpul di Jakarta - TP2S," Wapresri.go.id. Accessed: May 20, 2025. [Online]. Available: <https://stunting.go.id/bahas-percepatan-penurunan-angka-prevalensi-stunting-14-pada-2024-para-pemangku-kepentingan-pusat-dan-daerah-berkumpul-di-jakarta/>
- [3] J. Fauziah, K. D. Trisnawati, Rini, Khansa P. S., dan Putri, Suci U, "Stunting: Penyebab, Gejala, dan Pencegahan," *Jurnal Parenting dan Anak*, 2024, [Online]. Available: <https://edu.pubmedia.id/index.php/jpa/article/view/220>
- [4] R. Hardinata, L. Oktaviana, F. F. Husain, Putri, S., and Kartiasih, F., "Analisis faktor-faktor yang memengaruhi stunting di Indonesia tahun 2021," *Seminar Nasional Official Statistic*, 2023, [Online]. Available: <https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/1867>
- [5] S. Helmyati, D. R. Atmaka, S. U. Wisnusanti, and M. Wigati, *Stunting: Permasalahan dan Penanganannya*. books.google.com, 2020. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=PK3qDwAAQBAJ&oi=fnd&pg=PA2&dq=stunting&ots=454H_i-WWG&sig=50p_Czrv1G1EPCxdN-5a_MD8dw0
- [6] E. Pujiana, I. Purnama Sari, V. Melia Mardika, and M. Putri, "Analisis Algoritma Back Propagation Dalam Prediksi Angka Kemiskinan Di Indonesia," *Pendekar : Jurnal Pendidikan Berkarakter*, vol. 3, no. 1, pp. 11–17, 2020, doi: 10.31764.
- [7] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 413–422, Feb. 2024, doi: 10.57152/malcom.v4i2.1210.
- [8] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 257–265, Jan. 2024, doi: 10.57152/malcom.V4i1.1078.
- [9] K. M. Hamdia, X. Zhuang, and T. Rabczuk, "An efficient optimization approach for designing machine learning models based on genetic algorithm," *Neural Comput Appl*, vol. 33, no. 6, pp. 1923–1933, Mar. 2021, doi: 10.1007/S00521-020-05035.
- [10] Y. Cheng *et al.*, "Optimized Credit Score Prediction via an Ensemble Model and SMOTEENN Integration," *2024 IEEE 7th International Conference on Information Systems and Computer Aided Education, ICISCAE 2024*, pp. 355–361, 2024, doi: 10.1109/ICISCAE62304.2024.10761352.
- [11] T. Terttiaavini, A. Heryati, A. A. Suryanto, E. Hertati, and M. Purba, "Data Mining Dan Business Intelligence," 1st ed., B. Santoso, Ed., Bengkulu, 2024. [Online]. Available: www.qianzysains.com
- [12] K. Cheng and D. S. Young, "An Approach for Specifying Trimming and Winsorization Cutoffs," *J Agric Biol Environ Stat*, vol. 28, no. 2, pp. 299–323, Jun. 2023, doi: 10.1007/S13253-023-00527-4.
- [13] V. Safak, "Min-Mid-Max Scaling, Limits of Agreement, and Agreement Score," *ArXiv*, Jun. 2020, Accessed: May 20, 2025. [Online]. Available: <https://arxiv.org/pdf/2006.12904>
- [14] T. E. Tallo and A. Musdholifah, "The Implementation of Genetic Algorithm in Smote (Synthetic Minority Oversampling Technique) for Handling Imbalanced Dataset Problem," *Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018*, Nov. 2018, doi: 10.1109/ICSTC.2018.8528591.
- [15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.
- [16] World Health Organization, "Levels and trends in child malnutrition: UNICEF/WHO/The World Bank Group joint child malnutrition estimates: key findings of the 2021 edition." Accessed: May 20, 2025. [Online]. Available: <https://www.who.int/publications/i/item/9789240025257>
- [17] Kemenkes, "Survei Status Gizi Indonesia (SSGI) 2022." Accessed: May 20, 2025. [Online]. Available: <https://layanandata.kemkes.go.id/katalog-data/ssgi/ketersediaan-data/ssgi-2022>
- [18] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidayatallah, and M. El-Amir, "A Review of Deep Learning Algorithms and Their Applications in Healthcare," *Algorithms 2022, Vol. 15, Page 71*, vol. 15, no. 2, p. 71, Feb. 2022, doi: 10.3390/A15020071.

-
- [19] T. Terttiaavini, "Pengembangan Aplikasi Bunda Care untuk Pemantau Tumbuh Kembang Anak Sebagai Inovasi Antisipatif Penanggulangan Stunting dengan Pendekatan Agile Development," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 547–555, Feb. 2024, doi: 10.57152/malcom.V4I2.1288.
- [20] J. Lee, Y. Choi, and Y. Yoon, "Genetic Algorithm-based Feature Selection for Machine Learning System Diagnosing Sarcopenia," *GECCO 2023 Companion - Proceedings of the 2023 Genetic and Evolutionary Computation Conference Companion*, pp. 71–72, Jul. 2023, doi: 10.1145/3583133.3596943.
- [21] M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel community detection based genetic algorithm for feature selection," *J Big Data*, vol. 8, no. 1, pp. 1–27, Dec. 2021, doi: 10.1186/S40537-020-00398-3.
- [22] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier Detection," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, Jun. 2020, doi: 10.1145/3381028.
- [23] J. Rashid and K. Waheed, "Missing Values and Outliers in Research Data," *Pakistan Postgraduate Medical Journal*, vol. 31, no. 04, pp. 167–167, Jun. 2020, doi: 10.51642/PPMJ.V31I04.404.
- [24] R. Addanki, A. McGregor, A. Meliou, and Z. Moumoulidou, "Improved Approximation and Scalability for Fair Max-Min Diversification," Jan. 2022, Accessed: May 20, 2025. [Online]. Available: <https://arxiv.org/pdf/2201.06678>
- [25] R. Susmaga, "Confusion Matrix Visualization," *Intelligent Information Processing and Web Mining*, pp. 107–116, 2004, doi: 10.1007/978-3-540-39985-8_12.