Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 24388-24396
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Prediksi Stunting Menggunakan Random Forest dan SMOTE Berdasarkan Data Posyandu

Selvi Nuriana[1], Fitri Yunita[2]
[12] Sistem Informasi, Teknik dan Ilmu Komputer, Universitas Islam Indragiri
[1]selvi190818@gmail.com , [2]fitriyun@gmail.com

Abstrak

Stunting merupakan salah satu permasalahan kesehatan yang masih menjadi perhatian di Indonesia karena dapat memengaruhi pertumbuhan fisik, perkembangan kognitif, dan kualitas sumber daya manusia di masa mendatang. Penelitian ini bertujuan membangun model prediksi status stunting pada balita menggunakan algoritma Random Forest serta menganalisis pengaruh penerapan Synthetic Minority Oversampling Technique *(SMOTE) terhadap performa model. Dataset yang digunakan merupakan data primer yang berasal dari 24 Posyandu di wilayah kerja Puskesmas Gajah Mada dengan jumlah data akhir sebanyak 1.330 balita yang terdiri atas 1.287 balita tidak stunting dan 43 balita stunting. Tahap penelitian meliputi* preprocessing *data, pembentukan variabel target,* train-test split*, penerapan SMOTE, pembangunan model Random Forest, dan evaluasi model menggunakan* accuracy, precision, recall, *dan* F1-score*. Hasil penelitian menunjukkan bahwa model Random Forest tanpa SMOTE menghasilkan* accuracy *sebesar 96,99%,* precision *sebesar 100%,* recall *sebesar 11,11%, dan* F1-score *sebesar 20%. Setelah penerapan SMOTE, performa model meningkat dengan* accuracy *sebesar 97,37%,* precision *sebesar 75%,* recall *sebesar 33,33%, dan* F1-score *sebesar 46,15%. Hasil tersebut menunjukkan bahwa SMOTE mampu meningkatkan kemampuan model dalam mendeteksi kasus stunting. Penelitian ini diharapkan dapat membantu kader Posyandu, tenaga kesehatan, Puskesmas, pemerintah daerah, dan pengelola Program Makan Bergizi Gratis dalam melakukan identifikasi dini balita yang berisiko mengalami stunting.*

Kata Kunci: *Stunting, Random Forest, SMOTE, Machine Learning, Prediksi Stunting, Posyandu.*

1. Latar Belakang

Stunting merupakan kondisi gagal tumbuh pada anak akibat kekurangan gizi kronis yang berlangsung dalam jangka waktu yang panjang, terutama pada periode 1.000 Hari Pertama Kehidupan. Kondisi ini ditandai dengan panjang atau tinggi badan anak yang berada di bawah standar usianya. Stunting tidak hanya memengaruhi pertumbuhan fisik, tetapi juga berdampak pada perkembangan kognitif, kemampuan belajar, produktivitas, serta meningkatkan risiko terjadinya penyakit kronis pada masa dewasa. Dampak jangka Panjang tersebut dapat memengaruhi kualitas sumber daya manusia di masa mendatang [1], [2].

Stunting masih menjadi salah satu permasalahan kesehatan yang mendapat perhatian khusus di Indonesia. Pemerintah terus melakukan berbagai upaya percepatan penurunan stunting melalui intervensi gizi dan peningkatan pelayanan kesehatan masyarakat. Berdasarkan Survei Kesehatan Indonesia (SKI) tahun 2023, prevalensi stunting nasional mengalami penurunan menjadi 19,8%, namun upaya pencegahan dan penanganan stunting tetap perlu dilakukan secara berkelanjutan. Kondisi tersebut menunjukkan bahwa identifikasi dini terhadap balita yang berisiko mengalami stunting masih menjadi kebutuhan yang penting [3], [4].

Perkembangan teknologi informasi telah mendorong pemanfaatan machine learning dalam bidang kesehatan untuk membantu proses klasifikasi dan prediksi berbagai kondisi kesehatan, termasuk status gizi pada balita. Teknologi ini memungkinkan sistem mempelajari pola dari data historis untuk menghasilkan prediksi terhadap data baru secara otomatis. Berbagai penelitian menunjukkan bahwa pendekatan machine learning mampu mendukung deteksi dini stunting dan membantu tenaga kesehatan dalam pengambilan keputusan yang lebih cepat dan akurat [5], [6]. Salah satu algoritma yang banyak digunakan adalah Random Forest karena memiliki kemampuan klasifikasi yang stabil, mampu menangani data dengan jumlah variabel yang cukup banyak, serta menghasilkan performa yang baik pada penelitian kesehatan dan prediksi stunting [7], [8]. Namun, penelitian prediksi stunting umumnya menghadapi permasalahan imbalanced data karena jumlah balita yang mengalami stunting jauh lebih sedikit dibandingkan balita tidak stunting. Kondisi tersebut menyebabkan model cenderung lebih mengenali kelas mayoritas sehingga diperlukan teknik penanganan khusus untuk meningkatkan kemampuan deteksi terhadap kelas minoritas[9], [10]. Salah satu teknik yang banyak digunakan untuk mengatasi permasalahan tersebut adalah

Synthetic Minority Oversampling Technique (SMOTE), yaitu metode yang membangkitkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang dan performa model dapat meningkat [9], [11].

Beberapa penelitian sebelumnya telah menerapkan Random Forest untuk prediksi stunting. Widyawati dkk. menunjukan bahwa Random Forest memiliki performa yang kompetitif dibandingkan Logistic Regression dan Support Vector Machine pada data antropometri balita [12]. Hasil yang serupa juga dilaporkan oleh Juwariyem dkk. yang menunjukkan bahwa algoritma Random Forest mampu menghasilkan performa klasifikasi yang baik pada prediksi status stunting balita [13]. Selain itu, Azriani dkk. mengembangkan model prediksi stunting pada populasi bayi di Indonesia dan menunjukkan bahwa pendekatan machine learning berpotensi digunakan sebagai alat identifikasi dini risiko stunting [14]. Akan tetapi, sebagian besar penelitian sebelumnya masih menggunakan dataset sekunder atau data berskala nasional sehingga belum banyak memanfaatkan data primer Posyandu secara langsung pada tingkat wilayah kerja Puskesmas. Selain itu, penggunaan variabel Program Makan Bergizi Gratis (MBG) sebagai salah satu fitur prediksi juga masih jarang diteliti. Program Makan Bergizi Gratis (MBG) merupakan salah satu program pemerintah yang bertujuan meningkatkan kualitas gizi masyarakat, khususnya pada kelompok rentan. Namun, pemanfaatan variabel MBG sebagai fitur dalam model prediksi stunting masih sangat terbatas sehingga berpotensi menjadi nilai tambah dalam penelitian ini.

Berdasarkan kondisi tersebut, penelitian ini menggunakan data primer yang berasal dari 24 Posyandu di wilayah kerja Puskesmas Gajah Mada yang dikumpulkan melalui pencatatan rutin kader Posyandu dan dilaporkan kepada Puskesmas setiap bulan. Selain menggunakan data primer Posyandu, penelitian ini juga mengintegrasikan variabel Program Makan Bergizi Gratis (MBG) serta membandingkan performa algoritma Random Forest sebelum dan sesudah penerapan SMOTE untuk meningkatkan kemampuan deteksi terhadap kasus stunting. Penelitian ini bertujuan membangun model prediksi status stunting pada balita menggunakan algoritma Random Forest, menganalisis pengaruh penerapan SMOTE terhadap performa model, serta mengidentifikasi tingkat kontribusi masing-masing variabel dalam proses prediksi. Hasil penelitian diharapkan dapat membantu kader Posyandu, tenaga kesehatan, Puskesmas, pemerintah daerah, dan pengelola Program Makan Bergizi Gratis dalam melakukan identifikasi dini balita yang berisiko mengalami stunting.

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis machine learning menggunakan algoritma Random Forest. Untuk mengatasi permasalahan ketidakseimbangan kelas (imbalanced data), diterapkan Synthetic Minority Oversampling Technique (SMOTE) pada data latih. Tahapan penelitian meliputi pengumpulan data, preprocessing, pembentukan variabel target, pembagian data, penerapan SMOTE, pembangunan model, dan evaluasi performa model. Seluruh proses pengolahan data dilakukan menggunakan bahasa pemrograman Python melalui Google Colaboratory.

2.1. Sumber dan Karakteristik Data

Dataset yang digunakan merupakan data primer balita yang berasal dari 24 Posyandu di wilayah kerja Puskesmas Gajah Mada. Data dikumpulkan oleh kader Posyandu melalui kegiatan pengukuran rutin yang dilaksanakan satu kali setiap bulan dan selanjutnya dilaporkan kepada Puskesmas Gajah Mada. Jumlah data awal yang diperoleh sebanyak 1.338 data balita. Setelah dilakukan proses data cleaning dan penghapusan data yang tidak memenuhi kriteria penelitian, jumlah data yang digunakan menjadi 1.330 data balita.

Tabel 1. Distribusi Data Balita

Status Balita	Jumlah
Tidak Stunting	1.287
Stunting	43
Total	1.330

Berdasarkan Tabel 1, jumlah balita tidak stunting jauh lebih besar dibandingkan jumlah balita stunting. Kondisi tersebut menunjukkan bahwa dataset memiliki distribusi kelas yang tidak seimbang (imbalanced data). Kondisi ini berpotensi menyebabkan model lebih cenderung memprediksi kelas mayoritas sehingga diperlukan teknik penanganan ketidakseimbangan data sebelum proses pelatihan model dilakukan [9], [11].

Penelitian ini menggunakan tujuh variabel fitur dan satu variabel target yang digunakan dalam proses pembangunan model prediksi stunting.

Tabel 2. Variabel Penelitian

Variabel	**Peran**
Jenis Kelamin	Fitur
Berat Badan Lahir	Fitur
Tinggi Badan Lahir	Fitur
Usia Balita	Fitur
Berat Badan	Fitur

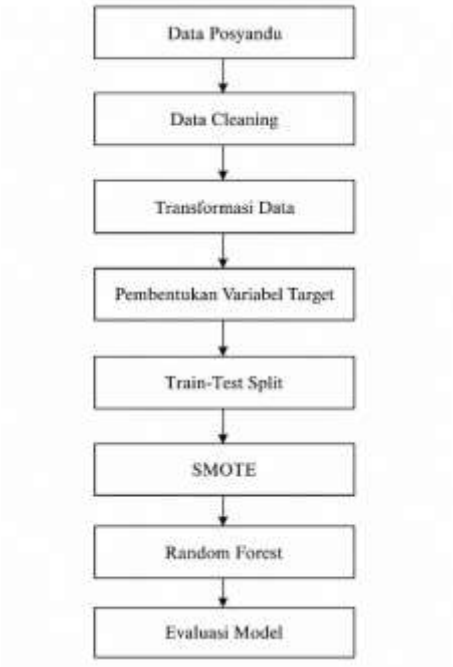
DOI: https://doi.org/10.31004/riggs.v5i2.12624

Tinggi Badan	Fitur
MBG	Fitur
Status Stunting	Target

Berdasarkan Tabel 2, penelitian ini menggunakan tujuh variabel fitur, yaitu jenis kelamin, berat badan lahir, tinggi badan lahir, usia balita, berat badan, tinggi badan, dan MBG sebagai variabel masukan (*input*) dalam proses pembangunan model. Sementara itu, status stunting digunakan sebagai variabel target (*output*) yang akan diprediksi oleh model klasifikasi. Pemilihan variabel tersebut didasarkan pada keterkaitannya dengan kondisi pertumbuhan dan perkembangan balita yang berpotensi memengaruhi status stunting.

2.2. Tahapan Penelitian

Tahapan penelitian dilakukan secara berurutan mulai dari pengumpulan data hingga evaluasi model yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian dilakukan secara bertahap mulai dari pengumpulan data Posyandu, preprocessing data yang meliputi data cleaning dan transformasi data, pembentukan variabel target, pembagian data latih dan data uji, penerapan SMOTE pada data latih, pembangunan model Random Forest, hingga evaluasi performa model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score.

2.3. Prepocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pembentukan model. Pada tahap data cleaning, ditemukan satu data kosong pada variabel Tinggi Badan Lahir. Nilai kosong tersebut ditangani menggunakan metode imputasi median sehingga tidak diperlukan penghapusan data. Variabel kategorikal kemudian ditransformasikan ke dalam bentuk numerik. Variabel jenis kelamin dikodekan menjadi laki-laki = 1 dan perempuan = 0, sedangkan variabel MBG dikodekan menjadi tidak = 0, ya = 1, dan tidak diketahui = 2. Selain itu, usia balita yang semula berbentuk teks dikonversi ke dalam satuan bulan. Sebagai contoh, usia 4 tahun 6 bulan dikonversi menjadi 54 bulan. Hasil preprocessing ini menghasilkan dataset yang lebih bersih dan siap digunakan pada tahap pembangunan model klasifikasi.

2.4. Pembangunan Model

Pembangunan model diawali dengan membagi dataset menjadi data latih (training data) dan data uji (testing data) menggunakan metode train-test split dengan proporsi 80% dan 20%. Teknik stratified sampling diterapkan untuk mempertahankan proporsi distribusi kelas pada kedua kelompok data sehingga representasi data tetap terjaga selama proses pelatihan dan pengujian model [15]. Selanjutnya, dilakukan penerapan Synthetic Minority Oversampling Technique (SMOTE) pada data latih untuk mengatasi ketidakseimbangan distribusi kelas. Teknik ini menghasilkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang dan kemampuan model dalam mendeteksi kasus stunting dapat ditingkatkan. Setelah proses penyeimbangan data selesai, algoritma Random Forest digunakan untuk membangun model klasifikasi status stunting. Algoritma ini dipilih karena memiliki kemampuan klasifikasi yang baik, mampu menangani data dengan banyak variabel, serta

DOI: https://doi.org/10.31004/riggs.v5i2.12624

relatif stabil dalam mengurangi risiko overfitting [8]. Parameter Random Forest yang digunakan pada penelitian ini ditunjukkan pada Tabel 3.

Tabel 3. Parameter Random Forest

Parameter	Nilai	Keterangan
n_estimators	100	Jumlah pohon keputusan yang dibangun
random_state	42	Nilai acak untuk menjaga konsistensi hasil

Berdasarkan tabel 3, nilai n_estimators ditetapkan sebanyak 100 untuk membangun 100 pohon keputusan (*decision tree*) pada algoritma Random Forest. Jumlah tersebut dipilih karena mampu memberikan keseimbangan antara performa prediksi dan efisiensi komputasi. Sementara itu, parameter random_state ditetapkan sebesar 42 untuk memastikan hasil eksperimen bersifat konsisten dan dapat direproduksi apabila proses pelatihan model dijalankan kembali pada kondisi yang sama.

Random Forest bekerja dengan membangun sejumlah decision tree menggunakan teknik bootstrap sampling. Setiap pohon keputusan menghasilkan prediksi yang kemudian digabungkan menggunakan mekanisme majority voting untuk menentukan hasil klasifikasi akhir. Pendekatan ini telah banyak digunakan pada penelitian prediksi stunting karena mampu menghasilkan performa klasifikasi yang baik dan konsisten [13].

2.5. Evaluasi Model

Evaluasi model dilakukan menggunakan Confusion Matrix, Accuracy, Precision, Recall, dan F1-Score. Confusion Matrix digunakan untuk mengetahui jumlah prediksi benar dan salah yang dihasilkan model.

Tabel 4. Struktur Confusion Matrix

Aktual/Prediksi	Tidak Stunting	Stunting
Tidak Stunting	TN	FP
Stunting	FN	TP

Berdasarkan nilai pada Confusion Matrix, performa model dihitung menggunakan empat metrik evaluasi.

1. Accuracy merupakan kemampuan model dalam melakukan prediksi secara keseluruhan.

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

2. Precision merupakan kemampuan model dalam menghasilkan prediksi positif yang benar.

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

3. Recall merupakan kemampuan model dalam mendeteksi seluruh kasus stunting.

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

4. F1-Score merupakan rata-rata harmonis antara precision dan recall.

$$F1 - Score = \frac{2\ x\ (Precision\ x\ Recall)}{(Precision\ +\ Recal)}$$

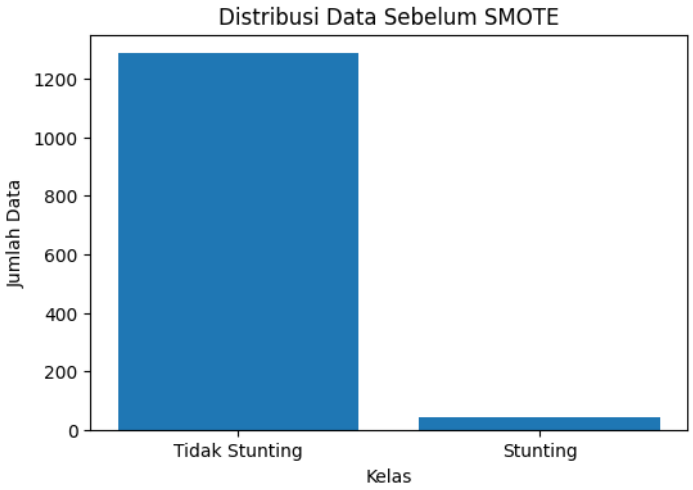
Keterangan;
TP (True Positive): balita stunting yang diprediksi benar sebagai stunting.
TN (True Negative): balita tidak stunting yang diprediksi benar sebagai tidak stunting.
FP (False Positive): balita tidak stunting yang salah diprediksi sebagai stunting.
FN (False Negative): balita stunting yang salah diprediksi sebagai tidak stunting.

3. Hasil dan Diskusi

3.1. Hasil Prepocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pembentukan model machine learning. Dataset awal yang diperoleh berjumlah 1.338 data balita yang berasal dari 24 Posyandu di wilayah kerja Puskesmas Gajah Mada. Pada tahap data cleaning, ditemukan satu data kosong pada variabel Tinggi Badan Lahir (TB Lahir). Data tersebut ditangani menggunakan metode imputasi median sehingga tidak diperlukan penghapusan data. Selain itu, data dengan kategori "-" dan Outlier pada variabel Tinggi Badan Menurut Umur (TB/U) dihapus karena tidak dapat digunakan sebagai kelas target dalam proses klasifikasi. Setelah proses pembersihan data selesai, jumlah data yang digunakan menjadi 1.330 data balita yang terdiri atas 1.287 balita tidak stunting dan 43 balita stunting.

DOI: https://doi.org/10.31004/riggs.v5i2.12624



Gambar 2. Distribusi Data Sebelum SMOTE

Berdasarkan Gambar 2, jumlah balita tidak stunting jauh lebih besar dibandingkan jumlah balita stunting. Perbedaan jumlah data yang cukup besar berpotensi menyebabkan model lebih dominan mempelajari karakteristik kelas mayoritas dan mengabaikan karakteristik kelas minoritas. Oleh karena itu, pada penelitian ini digunakan teknik Synthetic Minority Oversampling Technique (SMOTE) sebelum proses pelatihan model dilakukan. Pendekatan oversampling banyak digunakan pada penelitian klasifikasi data kesehatan karena mampu meningkatkan representasi kelas minoritas sehingga model dapat mempelajari pola data secara lebih optimal [16].

3.2. Hasil Penerapan Random Forest Tanpa SMOTE

Model Random Forest pertama dibangun menggunakan data asli tanpa penerapan teknik SMOTE. Evaluasi model dilakukan menggunakan Confusion Matrix dan empat metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score.

Tabel 5. Confusion Matrix Random Forest Tanpa SMOTE

Aktual/Prediksi	Tidak Stunting	Stunting
Tidak Stunting	257	0
Stunting	8	1

Berdasarkan Tabel 5, model Random Forest tanpa penerapan SMOTE berhasil mengklasifikasikan 257 balita tidak stunting dengan benar (True Negative) dan tidak menghasilkan kesalahan prediksi pada kelas tersebut (False Positive = 0). Namun, dari 9 balita yang sebenarnya mengalami stunting, model hanya mampu mengidentifikasi 1 balita sebagai stunting (True Positive), sedangkan 8 balita lainnya salah diklasifikasikan sebagai tidak stunting (False Negative). Hasil ini menunjukkan bahwa meskipun model memiliki tingkat akurasi yang tinggi, kemampuannya dalam mendeteksi kelas stunting masih sangat rendah. Kondisi tersebut disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah data balita tidak stunting jauh lebih banyak dibandingkan data balita stunting. Akibatnya, model cenderung mempelajari pola dari kelas mayoritas sehingga kemampuan dalam mengenali kelas minoritas menjadi kurang optimal. Temuan ini juga sejalan dengan nilai recall yang rendah pada hasil evaluasi model, sehingga diperlukan teknik penyeimbangan data, seperti Synthetic Minority Oversampling Technique (SMOTE), untuk meningkatkan kemampuan deteksi terhadap kasus stunting.

Tabel 6. Hasil Evaluasi Random Forest Tanpa SMOTE

Metrik	Nilai
Accuracy	96,99%
Precision	100%
Recall	11,11%
F1-Score	20%

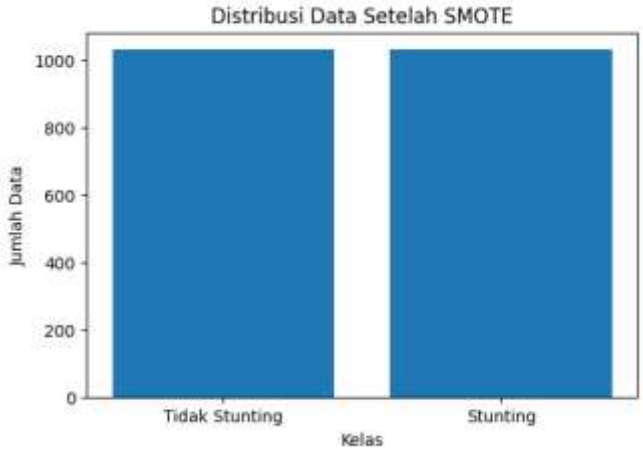
Berdasarkan Tabel 6, model Random Forest tanpa penerapan SMOTE menghasilkan accuracy sebesar 96,99%, precision sebesar 100%, recall sebesar 11,11%, dan F1-score sebesar 20%. Nilai accuracy yang tinggi menunjukkan bahwa model mampu melakukan klasifikasi dengan baik secara keseluruhan. Selain itu, nilai precision sebesar 100% mengindikasikan bahwa seluruh balita yang diprediksi sebagai stunting memang benar merupakan kasus stunting sehingga model tidak menghasilkan false positive. Namun, nilai recall yang hanya mencapai 11,11% menunjukkan bahwa sebagian besar balita yang sebenarnya mengalami stunting belum berhasil terdeteksi oleh model. Kondisi tersebut berdampak pada rendahnya F1-score, yaitu sebesar 20%, yang

DOI: https://doi.org/10.31004/riggs.v5i2.12624

menunjukkan bahwa keseimbangan antara nilai precision dan recall masih belum optimal. Hasil ini mengindikasikan bahwa ketidakseimbangan distribusi data menyebabkan model lebih banyak mempelajari karakteristik kelas mayoritas dibandingkan kelas minoritas sehingga diperlukan teknik penyeimbangan data, seperti Synthetic Minority Oversampling Technique (SMOTE), untuk meningkatkan kemampuan deteksi terhadap kasus stunting. Berdasarkan hasil tersebut, dapat disimpulkan bahwa akurasi yang tinggi belum tentu mencerminkan kemampuan model dalam mendeteksi kelas minoritas apabila distribusi data tidak seimbang.

3.3. Hasil Penerapan Random Forest dengan SMOTE

Pada penelitian ini, SMOTE diterapkan pada data training untuk menyeimbangkan distribusi jumlah data antara kelas stunting dan tidak stunting.



Gambar 3. Distribusi Data Setelah SMOTE

Berdasarkan Gambar 3, jumlah data pada kedua kelas menjadi seimbang, yaitu masing-masing sebanyak 1.030 data. Kondisi ini membantu model mempelajari karakteristik balita stunting secara lebih optimal. Hasil tersebut sejalan dengan penelitian Miranda dkk. yang menyatakan bahwa distribusi data yang lebih seimbang mampu meningkatkan proses pembelajaran model machine learning pada kasus prediksi status gizi balita [16].

Tabel 7. Confusion Matrix Random Forest dengan SMOTE

Aktual/Prediksi	Tidak Stunting	Stunting
Tidak Stunting	256	1
Stunting	6	3

Berdasarkan Tabel 7, model Random Forest setelah penerapan Synthetic Minority Oversampling Technique (SMOTE) berhasil mengklasifikasikan 256 balita tidak stunting dengan benar (True Negative) dan 3 balita stunting dengan benar (True Positive). Selain itu, terdapat 1 balita tidak stunting yang salah diprediksi sebagai stunting (False Positive) serta 6 balita stunting yang masih salah diklasifikasikan sebagai tidak stunting (False Negative). Dibandingkan dengan hasil sebelum penerapan SMOTE, jumlah balita stunting yang berhasil dideteksi meningkat dari 1 kasus menjadi 3 kasus, sedangkan jumlah False Negative menurun dari 8 menjadi 6 kasus. Hasil tersebut menunjukkan bahwa penerapan SMOTE membantu model mempelajari karakteristik kelas minoritas dengan lebih baik sehingga kemampuan dalam mengenali balita yang mengalami stunting mengalami peningkatan. Peningkatan tersebut menunjukkan bahwa proses oversampling berhasil memperkaya pola pada kelas minoritas sehingga model memperoleh informasi yang lebih representatif selama proses pelatihan. Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi pada kelas stunting yang menunjukkan bahwa proses identifikasi kasus stunting masih memiliki tantangan dan memerlukan pengembangan lebih lanjut.

Tabel 8. Hasil Evaluasi Random Forest dengan SMOTE

Metrik	Nilai
Accuracy	97,37%
Precision	75%
Recall	33,33%
F1-Score	46,15%

Berdasarkan Tabel 8, model Random Forest setelah penerapan Synthetic Minority Oversampling Technique (SMOTE) menghasilkan accuracy sebesar 97,37%, precision sebesar 75%, recall sebesar 33,33%, dan F1-score sebesar 46,15%. Dibandingkan dengan model tanpa SMOTE, nilai accuracy mengalami peningkatan dari 96,99% menjadi 97,37%, sedangkan recall meningkat dari 11,11% menjadi 33,33%. Peningkatan nilai recall menunjukkan bahwa model menjadi lebih mampu mendeteksi balita yang benar-benar mengalami stunting. Meskipun nilai

DOI: https://doi.org/10.31004/riggs.v5i2.12624

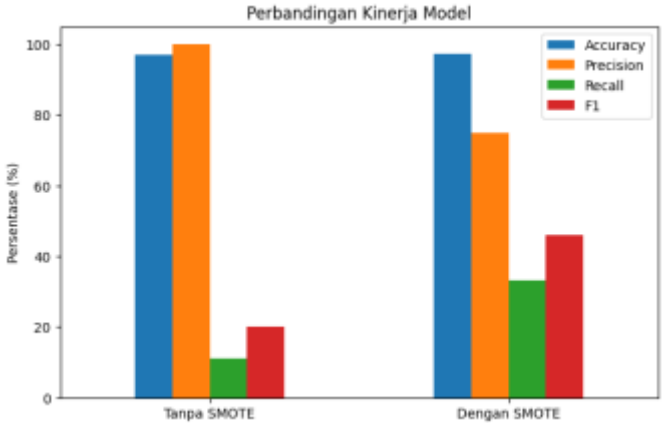
precision mengalami penurunan dari 100% menjadi 75%, kondisi tersebut masih dapat diterima karena model berhasil mengenali lebih banyak kasus stunting dibandingkan sebelumnya. Selain itu, nilai F1-score meningkat dari 20% menjadi 46,15%, yang menunjukkan bahwa keseimbangan antara precision dan recall menjadi lebih baik setelah penerapan SMOTE. Hasil ini menunjukkan bahwa teknik SMOTE efektif dalam mengatasi permasalahan ketidakseimbangan data sehingga kemampuan model dalam mengenali kelas minoritas mengalami peningkatan. Temuan tersebut sejalan dengan penelitian sebelumnya yang menyatakan bahwa penerapan SMOTE mampu meningkatkan sensitivitas model dalam mengenali kelas minoritas pada data yang tidak seimbang sehingga performa klasifikasi menjadi lebih optimal [13]. Hasil ini juga konsisten dengan penelitian Sugihartono dkk., yang menunjukkan bahwa penerapan teknik oversampling pada model ensemble learning mampu meningkatkan nilai recall dan F1-score dibandingkan penggunaan data asli yang tidak seimbang [17].

3.4. Perbandingan Kinerja Model

Perbandingan performa model sebelum dan sesudah penerapan SMOTE ditunjukkan pada Tabel 9.

Tabel 9. Perbandingan Kinerja Model

Metrik	Tanpa SMOTE	Dengan SMOTE
Accuracy	96,99%	97,37%
Precision	100%	75%
Recall	11,11%	33,33%
F1-Score	20%	46,15%



Gambar 4. Perbandingan Kinerja Model

Berdasarkan Tabel 9 dan Gambar 4, penerapan SMOTE meningkatkan nilai recall dan F1-score. Meskipun nilai precision mengalami penurunan, model menjadi lebih efektif dalam mengidentifikasi balita yang benar-benar berisiko mengalami stunting. Dalam konteks kesehatan, peningkatan kemampuan deteksi kasus stunting lebih diutamakan karena kegagalan mendeteksi balita yang membutuhkan perhatian dapat berdampak pada tumbuh kembang anak. Secara keseluruhan, kombinasi Random Forest dan SMOTE memberikan performa yang lebih baik dibandingkan penggunaan Random Forest tanpa penyeimbangan data. Hal ini menunjukkan bahwa penanganan ketidakseimbangan data merupakan langkah penting dalam pengembangan sistem prediksi stunting berbasis machine learning. Temuan tersebut didukung oleh penelitian El Furqany dkk. yang melaporkan bahwa penerapan metode penyeimbangan data pada model ensemble menghasilkan kemampuan klasifikasi yang lebih stabil, terutama dalam mendeteksi kelas minoritas [18]. Hasil tersebut menunjukkan bahwa penanganan ketidakseimbangan data tidak hanya meningkatkan kemampuan deteksi kasus stunting, tetapi juga menghasilkan model klasifikasi yang lebih seimbang dan lebih layak diterapkan sebagai sistem pendukung pengambilan keputusan di bidang kesehatan.

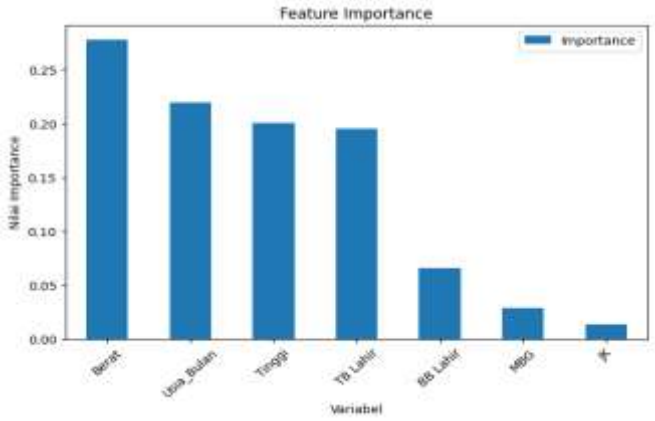
3.5. Analisis Feature Importance

Analisis feature importance dilakukan untuk mengetahui tingkat kontribusi masing-masing variabel terhadap proses prediksi status stunting.

Tabel 8. Hasil Feature Importance

Variabel	Importance
Berat	0,277
Usia_Bulan	0,220
Tinggi	0,201
TB Lahir	0,195
BB Lahir	0,065
MBG	0,028
JK	0,013

DOI: https://doi.org/10.31004/riggs.v5i2.12624



Gambar 5. Visualisasi Feature Importance

Berdasarkan Tabel 10 dan Gambar 5, berat badan merupakan variabel yang paling berpengaruh dalam proses prediksi status stunting, diikuti oleh usia balita, tinggi badan, dan tinggi badan lahir. Sementara itu, jenis kelamin memiliki kontribusi paling kecil.

Variabel Program Makan Bergizi Gratis (MBG) juga memberikan kontribusi terhadap model, meskipun nilainya masih relatif rendah dibandingkan variabel antropometri lainnya. Hal tersebut menunjukkan bahwa kondisi fisik dan pertumbuhan balita masih menjadi faktor dominan dalam proses identifikasi stunting, sejalan dengan penelitian sebelumnya yang menyatakan bahwa indikator antropometri merupakan prediktor utama dalam deteksi stunting[11], [17]. Temuan ini menunjukkan bahwa pendekatan berbasis machine learning tidak hanya mampu menghasilkan prediksi status stunting, tetapi juga memberikan informasi mengenai tingkat kontribusi masing-masing variabel. Informasi tersebut dapat dimanfaatkan sebagai dasar dalam mendukung pengambilan keputusan serta penyusunan strategi intervensi yang lebih tepat sasaran [19]. Selain meningkatkan akurasi prediksi, identifikasi variabel-variabel yang paling berpengaruh melalui pendekatan machine learning dapat membantu tenaga kesehatan dalam menentukan prioritas intervensi gizi bagi balita yang berisiko mengalami stunting [20].

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma Random Forest untuk memprediksi status stunting pada balita menggunakan 1.330 data primer yang berasal dari 24 Posyandu di wilayah kerja Puskesmas Gajah Mada. Hasil penelitian menunjukkan bahwa penerapan Synthetic Minority Oversampling Technique (SMOTE) mampu meningkatkan kinerja model, terutama pada kemampuan mendeteksi kasus stunting. Model Random Forest tanpa SMOTE menghasilkan accuracy 96,99%, precision 100%, recall 11,11%, dan F1-score 20%, sedangkan setelah penerapan SMOTE performa model meningkat menjadi accuracy 97,37%, precision 75%, recall 33,33%, dan F1-score 46,15%. Peningkatan nilai recall dan F1-score menunjukkan bahwa penanganan data tidak seimbang berperan penting dalam meningkatkan kemampuan model mengenali kelas minoritas. Hasil analisis feature importance juga menunjukkan bahwa berat badan, usia balita, tinggi badan, dan tinggi badan lahir merupakan variabel yang paling berpengaruh dalam prediksi stunting. Dengan demikian, model yang diusulkan berpotensi menjadi alat bantu pendukung pengambilan keputusan bagi kader Posyandu dan tenaga kesehatan dalam melakukan identifikasi dini balita yang berisiko mengalami stunting sehingga upaya pencegahan dan penanganan dapat dilakukan secara lebih cepat dan tepat. Pada penelitian selanjutnya, model dapat dikembangkan dengan menambahkan jumlah data, menggunakan algoritma machine learning lain, atau melakukan optimasi parameter Random Forest sehingga performa prediksi dapat ditingkatkan.

Referensi

[1] World Health Organization, “Child Growth.” [Online]. Available: https://www.who.int/health-topics/child-growth
[2] United Nations Children’s Fund, “Child Nutrition.” [Online]. Available: https://data.unicef.org/nutrition
[3] Kementerian Kesehatan Republik Indonesia, “Survei Kesehatan Indonesia (SKI) 2023,” Jakarta, 2024. [Online]. Available: https://www.badankebijakan.kemkes.go.id/hasil-ski-2023
[4] Kementerian Kesehatan Republik Indonesia, “Hasil Survei Status Gizi Indonesia (SSGI) 2022,” Jakarta, 2023. [Online]. Available: https://ayosehat.kemkes.go.id/publikasi-hasil-survei-status-gizi-indonesia-ssgi-2022
[5] W. R. Mgomezulu, P. Thangata, B. Mkandawire, and N. Amoah, “Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting,” *Hum. Nutr. Metab.*, vol. 42, no. August, p. 200340, 2025, doi: 10.1016/j.hnm.2025.200340.
[6] R. Kumar *et al.*, “A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions,” *Front. Artif. Intell.*, vol. 8, no. May, pp. 1–31, 2025, doi: 10.3389/frai.2025.1583459.
[7] A. H. Mubarok, P. Pujiono, D. Setiawan, D. F. Wicaksono, and E. Rimawati, “Parameter Testing on Random Forest Algorithm for Stunting Prediction,” *Sinkron*, vol. 9, no. 1, pp. 107–116, 2025, doi: 10.33395/sinkron.v9i1.14264.
[8] M. I. Elim and E. Utami, “Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with

DOI: https://doi.org/10.31004/riggs.v5i2.12624

Grid Search Optimization," *J. Tek. Inform.*, vol. 6, no. 5, pp. 5305–5319, 2025, doi: 10.52436/1.jutif.2025.6.5.5285.

[9] A. A. Dhani, T. A. Y. Siswa, and W. J. Pranoto, "Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting," *Teknika*, vol. 13, no. 2, pp. 264–272, 2024, doi: 10.34148/teknika.v13i2.875.

[10] R. Belferik, F. M. Sinaga, F. Ferawaty, M. A. S. Manullang, and T. Sinaga, "Addressing Class Imbalance in Stunting Classification Using SMOTE Enhanced Random Forest," *Sinkron*, vol. 9, no. 4, pp. 2108–2116, 2025, doi: 10.33395/sinkron.v9i4.15349.

[11] W. S. Lestari, Y. M. Saragih, and Caroline, "Comparison of Deep Neural Networks and Random Forest Algorithms for Multiclass Stunting Prediction in Toddlers," *Teknika*, vol. 13, no. 3, pp. 412–417, 2024, doi: 10.34148/teknika.v13i3.1063.

[12] S. B. D. Widyawati, P. Purwadi, and I. R. Yunita, "Comparative Analysis of Random Forest, Logistic Regression and SVM for Stunting Prediction Using Anthropometric Data," *J. Inf. Syst. Informatics*, vol. 7, no. 4, pp. 4271–4293, 2025, doi: 10.63158/journalisi.v7i4.1387.

[13] J. Juwariyem, S. Sriyanto, S. Lestari, and C. Chairani, "Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms," *Sinkron*, vol. 8, no. 2, pp. 947–955, 2024, doi: 10.33395/sinkron.v8i2.13448.

[14] D. Azriani, D. Agustian, Y. Zuhairini, I. N. Yulita, and M. Dhamayanti, "Prediction models for stunting at 2-years-old from Indonesian newborn population," *BMC Pediatr.*, vol. 25, no. 1, 2025, doi: 10.1186/s12887-025-06096-4.

[15] S. N. Syafa Iswahyudi and R. Eka Putra, "Sistem Deteksi Stunting pada Balita Berbasis Web Menggunakan Metode Random Forest," *J. Informatics Comput. Sci.*, vol. 6, no. 03, pp. 755–764, 2025, doi: 10.26740/jinacs.v6n03.p755-764.

[16] E. Miranda, M. Aryuni, A. Y. Zakiyyah, Y. E. Kurniawati, A. V. D. Sano, and M. Kumbangsila, "An early prediction model for toddler nutrition based on machine learning from imbalanced data," *Procedia Comput. Sci.*, vol. 245, pp. 263–271, 2024, doi: 10.1016/j.procs.2024.10.251.

[17] T. Sugihartono, D. Soetarno, and R. Sulaiman, "Ensemble Learning for Pediatric Stunting Detection : A Comparative Study of XGBoost , Random Forest , and LightGBM with Oversampling Techniques," *J. Inf. Syst. Informatics*, vol. 8, no. 2, pp. 1672–1692, 2026.

[18] N. El Furqany, M. Subianto, and A. Rusyana, "Hybrid Ensemble Learning with SMOTEENN and Soft Voting for Stunting Risk Prediction: A SHAP-Based Explainable Approach," *J. Appl. Data Sci.*, vol. 6, no. 4, pp. 2989–3004, 2025, doi: 10.47738/jads.v6i4.829.

[19] B. Rao, M. G. Hasan, B. Putturaya, A. Kamath, M. Aatif, and Y. M. Elmosaad, "Multidimensional Correlates of Childhood Stunting in India: A Spatial Machine Learning and Explainable AI Approach," *Stats*, vol. 9, no. 2, pp. 1–16, 2026, doi: 10.3390/stats9020034.

[20] M. M. Islam, N. M. Shoukot Jahan Kibria, S. Kumar, D. C. Roy, and M. R. Karim, "Prediction of undernutrition and identification of its influencing predictors among under-five children in Bangladesh using explainable machine learning algorithms," *PLoS One*, vol. 19, no. 12, pp. 1–22, 2024, doi: 10.1371/journal.pone.0315393.

DOI: https://doi.org/10.31004/riggs.v5i2.12624