Information System Journal (INFOS)
Vol. 9, No. 2, November 2026, pp. 69-78
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2026v9i02.2947

PEMODELAN SISTEM PENDUKUNG KEPUTUSAN KLINIS BERBASIS EXPLAINABLE AI (XAI) UNTUK MITIGASI STUNTING

Paris Hizriyan Saputra [1)], Heru Yuni Prasetyo [2)], Mochamad Dwijanto [3)], Walman Franky [4)], Anshar [5)], Kusrini [6)], I Made Artha Agastya [7)]

[1)2)3)4)5)] *Program Studi Magister PJJ Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta*
[6)7)] *Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta*

email: p.hizriyan@students.amikom.ac.id [1)], h.yuni@students.amikom.ac.id [2)], m.dwijanto@students.amikom.ac.id [3)], walmanfranky@students.amikom.ac.id [4)], anshar@students.amikom.ac.id [5)], kusrini@amikom.ac.id [6)], artha.agastya@amikom.ac.id [7)]

INFO ARTIKEL

Riwayat Artikel:

Diterima Juli, 2026
Revisi November, 2026
Terbit November, 2026

Penulis Korespondensi:

Paris Hizriyan Saputra
Program Studi Magister PJJ Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

Email:
p.hizriyan@students.amikom.ac.id

ABSTRAK

Model *machine learning* untuk deteksi stunting sering bersifat *black box* dan rawan kebocoran label karena status stunting ditetapkan dari *z-score* tinggi badan menurut umur (TB/U). Penelitian ini mengembangkan Sistem Pendukung Keputusan Klinis berbasis *Explainable AI* yang memisahkan penetapan status gizi deterministik dari skor risiko skrining. Data 40.071 balita Kabupaten Jeneponto dibersihkan (50,3% nilai berat badan bertipe tanggal dipulihkan) dan dibagi 80:20 sebelum *SMOTE* diterapkan dalam validasi silang. Label terbukti identik dengan aturan *z-score* TB/U < −2; rumus *WHO* tanpa *machine learning* mencapai akurasi 95,0%, sehingga akurasi 95,8–100% pada model yang memuat tinggi badan merupakan efek kebocoran. Model *XGBoost* tanpa tinggi badan mencapai *AUC-ROC* 0,829 serta sensitivitas 80,3% dan spesifisitas 67,6% pada ambang skrining. *TreeSHAP* mengatribusikan 51,8% kontribusi pada berat badan dan 39,3% pada usia. Aplikasi web menetapkan status stunting sesuai Permenkes No. 2/2020 dan memberikan rekomendasi dari *knowledge base* berrujukan pedoman; kesiapan klinis menunggu uji penerimaan pengguna.

Kata Kunci :
Stunting; *Explainable AI*; *TreeSHAP*; *Clinical Decision Support System*; Kebocoran Label

ABSTRACT

Machine learning models for stunting detection are often black boxes and prone to label leakage because stunting status is defined by the height-for-age z-score (HAZ). This study developed an Explainable AI-based Clinical Decision Support System that separates deterministic nutritional status assessment from a screening risk score. Records of 40,071 children from Jeneponto Regency were cleaned (recovering 50.3% of weights stored as dates) and split 80:20 before applying SMOTE within cross-validation. The label was identical to HAZ < −2; the WHO formula alone reached 95.0% accuracy, so the 95.8–100% accuracy of models including height reflects leakage. An XGBoost model without height achieved AUC-ROC 0.829, with 80.3% sensitivity and 67.6% specificity at the screening threshold. TreeSHAP attributed 51.8% of contributions to weight and 39.3% to age. The web application determines stunting status following Ministry of Health Regulation No. 2/2020 and provides guideline-referenced recommendations; clinical readiness awaits user acceptance testing.

***Keywords*:**
Stunting; Explainable AI; TreeSHAP; Clinical Decision Support System; Label Leakage

1. PENDAHULUAN

Stunting merupakan gangguan pertumbuhan *linear* kronis pada balita akibat interaksi kekurangan gizi jangka panjang, infeksi berulang, dan defisiensi stimulasi psikososial [1]. Kondisi ini ditetapkan ketika *z-score* tinggi badan menurut umur (TB/U) berada di bawah −2 SD dari median standar pertumbuhan *WHO* [2] dan berdampak pada perkembangan kognitif, produktivitas, serta risiko penyakit tidak menular di usia dewasa [3]. Survei Status Gizi Indonesia (SSGI) 2022 melaporkan prevalensi stunting 21,6%, masih di atas ambang 20% [4], dan Indonesia termasuk negara dengan beban stunting tertinggi di Asia Tenggara [5]. Kabupaten Jeneponto, Sulawesi Selatan, mencatat lebih dari 16.000 balita stunting pada periode 2021–2024 [6].

Deteksi dini melalui pengukuran antropometri di posyandu masih terkendala inkonsistensi pengukuran dan keterbatasan tenaga terlatih [7]. *Machine learning* (ML) menawarkan solusi potensial: Wicaksono et al. melaporkan akurasi *XGBoost* 97,57% pada data 40.071 balita Jeneponto [6], dan studi lain di Indonesia menunjukkan keunggulan algoritma *ensemble* berbasis pohon [8], [9]. Namun, model ML umumnya bersifat *black box* sehingga tidak layak dipakai untuk keputusan bertaruhan tinggi tanpa mekanisme interpretasi [10]. SHAP (*SHapley Additive exPlanations*) memberikan kerangka berbasis teori permainan untuk mengkuantifikasi kontribusi setiap fitur terhadap prediksi individual [11] dan telah dipakai untuk menjelaskan model stunting di Ethiopia dan Mesir, meskipun belum terintegrasi dalam alur kerja tenaga kesehatan [12], [13].

Kinerja yang sangat tinggi juga perlu diuji terhadap kebocoran label (*target leakage*), yaitu kondisi ketika fitur masukan memuat informasi pembentuk label sehingga model tampak akurat tanpa kemampuan prediktif yang sebenarnya [14]; kebocoran semacam ini merupakan salah satu penyebab utama hasil ML yang tidak dapat direproduksi [15]. Pada data antropometri, status stunting ditetapkan dari *z-score* TB/U, sedangkan *z-score* tersebut merupakan fungsi deterministik dari jenis kelamin, usia, dan tinggi badan [2], [16]. Model yang menerima tinggi badan dan usia dapat mencapai akurasi mendekati sempurna hanya dengan merekonstruksi rumus tersebut, dan penjelasan XAI-nya akan menempatkan kedua variabel itu sebagai faktor dominan tanpa memberi wawasan klinis baru.

Penelitian ini mengembangkan *Clinical Decision Support System* (CDSS) berbasis *Explainable AI* (XAI) yang menangani persoalan tersebut secara eksplisit, dengan kontribusi: (1) pembuktian kuantitatif kebocoran label pada dataset Jeneponto melalui tiga skenario fitur dengan protokol yang membagi data sebelum *oversampling*; (2) arsitektur yang menetapkan status gizi secara deterministik sesuai Permenkes No. 2 Tahun 2020, terpisah dari skor risiko skrining *XGBoost* tanpa fitur pembentuk label yang dijelaskan dengan *TreeSHAP*; (3) *knowledge base* rekomendasi yang berrujukan pedoman, berversi, dan mensyaratkan validasi ahli gizi; serta (4) instrumen uji penerimaan pengguna (UAT) yang terintegrasi dalam sistem.

2. METODOLOGI PENELITIAN

2.1 Dataset dan Praproses Data

Dataset berasal dari data pertumbuhan balita Kabupaten Jeneponto yang dipublikasikan Wicaksono et al. [6], berisi 40.071 rekaman balita usia 0–60 bulan tanpa identitas anak. Label adalah kategori *Height for Age* dengan 16.204 rekaman (40,4%) berlabel Stunted. Variabel dan perannya dirangkum pada Tabel 1.

Tabel 1. Variabel Dataset dan Perannya dalam Penelitian

Variabel	Deskripsi	Peran
Gender, *Age* (*Month*), *Weight*	Jenis kelamin, usia (bulan), berat badan (kg)	Input S1–S3; perhitungan *z-score*
Height	Panjang/tinggi badan (cm)	Input S1–S2; tidak dipakai model sistem (S3)
Height for Age	Kategori TB/U (*Stunted*/*Not Stunted*)	Label (target)
Z-Score H/A	*z-score* TB/U, dasar label	Input S2 (uji kebocoran eksplisit)
Weight for Age, *Z-Score W/A*	Kategori dan *z-score* BB/U	Validasi pemulihan berat badan
Weight for Height, *Z-Score W/H*	Kategori dan z-score BB/TB	Tidak dipakai

Sebelum pembagian data, praproses dibatasi pada langkah yang tidak bergantung pada distribusi data. Sebanyak 20.158 nilai *Weight* (50,3%) tersimpan sebagai tanggal *Excel* akibat salah tafsir format regional (*locale*), misalnya 8,2 kg terbaca sebagai 8 Februari. Karena seluruhnya berbulan 1–9, nilai tersebut dipulihkan tanpa ambiguitas sebagai hari + bulan/10; validasi terhadap kolom *Z-Score W/A* memberikan korelasi

Information System Journal (INFOS)
Vol. 9, No. 2, November 2026, pp. 69-78
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2026v9i02.2947

r = 0,984. *Z-score TB/U*, BB/U, dan BB/TB dihitung ulang dengan metode LMS standar *WHO* 2006 yang diadopsi Permenkes No. 2 Tahun 2020 [2], [16]:

$$z = \frac{\left(\frac{X}{M}\right)^{L} - 1}{L \cdot S}, \tag{1}$$

dengan X nilai pengukuran serta L, M, dan S parameter *Box-Cox*, median, dan koefisien variasi referensi *WHO* menurut jenis kelamin dan usia (atau panjang/tinggi badan untuk BB/TB). Sebanyak 55 rekaman dengan nilai biologis tidak plausibel (|z TB/U| > 6, z BB/U < −6 atau > 5, |z BB/TB| > 5) dikeluarkan sesuai *flag WHO* [17] sebagai pengganti *winsorization*, dan dua rekaman tanpa usia dibuang, sehingga data bersih berjumlah 40.014 rekaman (prevalensi Stunted 40,4%). Sebanyak 9.923 rekaman memiliki kombinasi jenis kelamin, usia, berat, dan tinggi yang identik dengan rekaman lain dan mungkin merupakan kunjungan berulang.

2.2 Pembagian Data, *SMOTE*, dan Skenario Fitur

Data dibagi menjadi data latih (80%; n = 32.011) dan data uji (20%; n = 8.003) secara terstratifikasi sebelum praproses apa pun yang bergantung pada data. Pembagian dilakukan berbasis grup sehingga rekaman dengan kombinasi jenis kelamin, usia, berat, dan tinggi yang identik selalu berada pada partisi yang sama (tumpang-tindih grup = 0). Ketidakseimbangan kelas ditangani dengan *SMOTE* [18] yang ditempatkan di dalam *pipeline* pelatihan (*SMOTE* → *XGBoost*), sehingga selama validasi silang 5-*fold SMOTE* hanya di-*fit* pada lipatan latih, sementara lipatan validasi dan data uji tetap berdistribusi asli; *oversampling* sebelum pembagian data diketahui menghasilkan estimasi yang terlalu optimistis [19]. Setiap skenario juga dilatih tanpa *SMOTE* sebagai pembanding.

Tiga skenario fitur dibandingkan. S1 memuat jenis kelamin, usia, berat, dan tinggi badan, konfigurasi yang juga dipakai pada analisis awal penelitian ini; karena label merupakan fungsi jenis kelamin, usia, dan tinggi melalui *z-score* TB/U, S1 mengandung kebocoran implisit. S2 menambahkan *Z-Score H/A* sehingga mengandung kebocoran eksplisit (label = 1[Z-Score H/A < −2]). S3 hanya memuat jenis kelamin, usia, dan berat badan, dan diposisikan sebagai skor risiko skrining, misalnya pada penimbangan rutin ketika panjang/tinggi badan (PB/TB) tidak diukur. Pembanding yang disertakan adalah aturan *z-score* TB/U *WHO* tanpa ML, pengklasifikasi kelas mayoritas, dan regresi logistik dengan *SMOTE* pada jenis kelamin, usia, berat, dan z-score BB/U.

2.3 *XGBoost*, Evaluasi, dan *TreeSHAP*

XGBoost membangun pohon keputusan secara sekuensial; setiap pohon baru mengoreksi residual pohon sebelumnya dengan meminimalkan fungsi objektif teregularisasi [20]:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \tag{2}$$

dengan l fungsi *logistic loss* antara label y_i dan prediksi $\hat{y}_i$, f_t pohon pada iterasi ke-t, dan Ω suku regularisasi:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{3}$$

dengan T jumlah daun, w_j bobot daun ke-j, serta γ dan λ parameter penalti. Hiperparameter dipilih dengan *GridSearchCV* 5-*fold* berbasis grup pada data latih (metrik *AUC-ROC*) atas *n_estimators* ∈ {200, 400}, *max_depth* ∈ {3, 5, 7}, *learning_rate* ∈ {0,05; 0,1}, dan *subsample* ∈ {0,8; 1,0}. Model terbaik dievaluasi satu kali pada data uji dengan akurasi, *balanced accuracy*, sensitivitas, spesifisitas, PPV, NPV, F1, *AUC-ROC*, dan skor Brier; interval kepercayaan (IK) 95% diperoleh dari 500 sampel *bootstrap*. Untuk S3 ditetapkan pula ambang skrining, yaitu ambang tertinggi dengan sensitivitas ≥ 80% pada prediksi *out-of-fold* data latih.

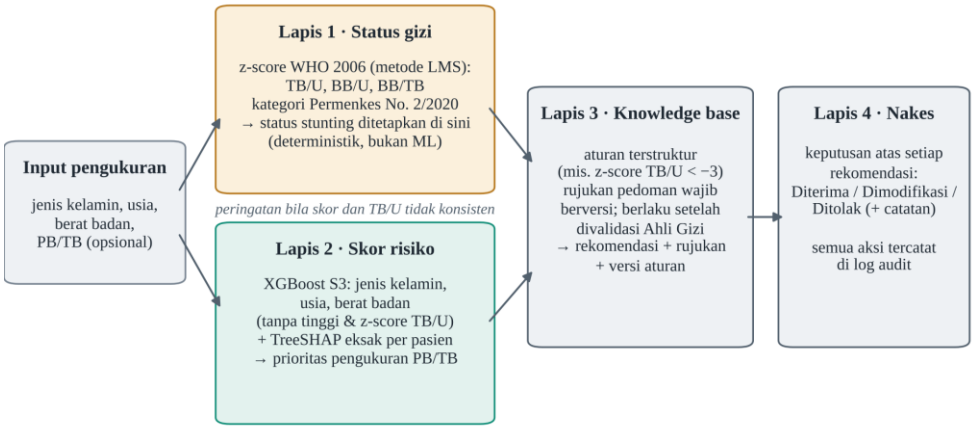
TreeSHAP menghitung nilai Shapley φ_i secara eksak untuk model berbasis pohon [11], [21]. Pada *XGBoost*, nilai *SHAP* berada dalam ruang *log-odds* dan bersifat aditif:

$$f(x) = \varphi_0 + \sum_{i=1}^{M} \varphi_i, \qquad p(x) = \frac{1}{1 + e^{-f(x)}}, \tag{4}$$

dengan φ_0 nilai dasar, M jumlah fitur, dan $p(x)$ skor risiko. Kepentingan global dihitung sebagai porsi rata-rata $|\varphi_i|$ pada data uji. *SHAP* menjelaskan perilaku model, bukan kausalitas; pada model yang bocor, atribusinya hanya mencerminkan rumus pembentuk label.

2.4 Arsitektur Sistem dan Rancangan UAT

CDSS terdiri atas empat lapis (Gambar 1) [22]. Lapis 1 menghitung z-*score* dengan Persamaan (1) dan menetapkan kategori Permenkes No. 2 Tahun 2020; status stunting (pendek: $-3 \leq z < -2$; sangat pendek: $z < -3$) ditetapkan di sini, bukan oleh model. Lapis 2 menghitung skor risiko S3 beserta *TreeSHAP* per pasien; bila PB/TB tidak diukur, sistem hanya menampilkan skor risiko, dan bila skor tidak konsisten dengan TB/U, sistem menampilkan peringatan. Lapis 3 adalah mesin aturan *knowledge base* (KB), dan Lapis 4 mencatat keputusan tenaga kesehatan (Diterima, Dimodifikasi, atau Ditolak) serta log audit. Sistem dibangun dengan *React* + *Vite* dan *FastAPI* (*Python*) dengan basis data *SQLite* sembilan tabel, termasuk *kb_revisi* untuk jejak versi aturan dan *uat_respon* untuk data UAT; angka kinerja dibaca langsung dari laporan pelatihan.



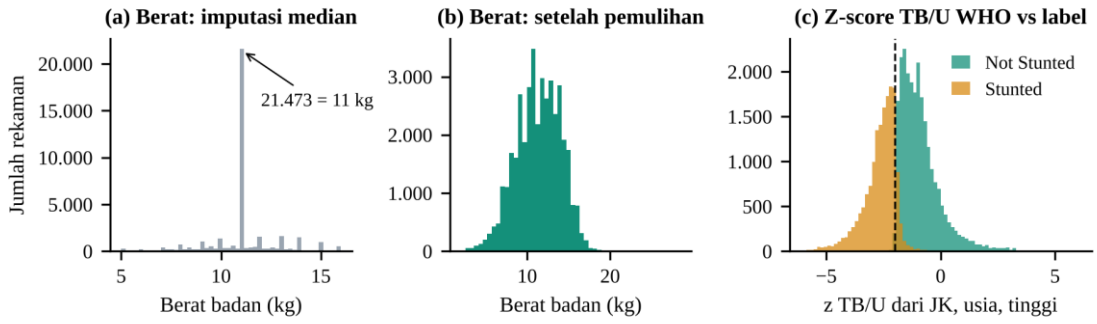
Gambar 1. Arsitektur CDSS: Status Stunting Ditetapkan secara Deterministik, Terpisah dari Skor Risiko Model

UAT dirancang dengan enam skenario tugas (menambah data balita, menjalankan analisis lengkap, menafsirkan TB/U dan *TreeSHAP*, mencatat keputusan atas rekomendasi, skrining tanpa PB/TB, dan menemukan prevalensi per desa), *System Usability Scale* (SUS) 10 butir [23], serta dua butir persepsi klinis (kegunaan penjelasan *TreeSHAP* dan kesesuaian rekomendasi). Kriteria penerimaan ditetapkan sebelum pengujian: minimal lima responden per peran (bidan/tenaga kesehatan, ahli gizi, dokter), rata-rata SUS ≥ 68 [24], keberhasilan setiap tugas $\geq 80\%$, dan skor kegunaan XAI ≥ 4 dari 5.

3. HASIL DAN PEMBAHASAN

3.1. Kualitas Data dan Bukti Kebocoran Label

Pada *pipeline* awal, nilai berat bertipe tanggal gagal dikonversi lalu diimputasi median (11 kg), sehingga 21.473 rekaman (53,6%) bernilai tepat 11 kg (Gambar 2a); setelah pemulihan, distribusi berat kembali kontinu dengan median 11,5 kg (Gambar 2b). Label Height for Age 100% identik dengan aturan Z-Score H/A < -2. Z-score TB/U yang dihitung ulang dari jenis kelamin, usia, dan tinggi berkorelasi $r = 0{,}983$ dengan Z-Score H/A, dan kedua kelas terpisah hampir sempurna pada ambang -2 SD (Gambar 2c). Pada data uji, aturan tanpa ML ini mencapai akurasi 95,0% (sensitivitas 87,9%, spesifisitas 99,8%) dan AUC-ROC 0,989 bila z-score dipakai sebagai skor. Informasi untuk memprediksi label dengan demikian hampir seluruhnya terkandung dalam jenis kelamin, usia, dan tinggi badan; tumpang-tindih kecil di sekitar ambang kemungkinan berasal dari perbedaan perhitungan z-score pada sumber data.



Gambar 2. Distribusi Berat Badan (a) dengan Imputasi Median dan (b) setelah Pemulihan; (c) Z-Score TB/U yang Dihitung Ulang menurut Label Dataset

Information System Journal (INFOS)
Vol. 9, No. 2, November 2026, pp. 69-78
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2026v9i02.2947

3.2. Kinerja Model

Tabel 2., menyajikan kinerja seluruh skenario dan pembanding pada data uji.

Tabel 2. Kinerja Skenario Fitur dan Pembanding pada Data Uji (n=8.003; Ambang 0,5)

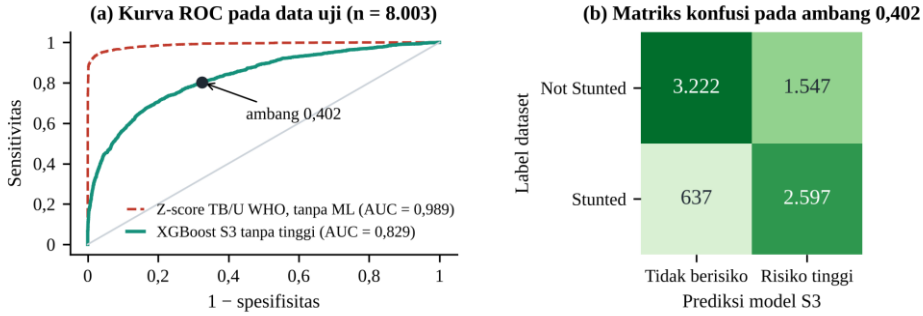
Skenario	Fitur	*SMOTE*	Akurasi [IK 95%]	Sens.	Spes.	*AUC-ROC* [IK 95%]	*AUC CV*
S1 (bocor implisit)	JK, Usia, BB, TB	Ya	95,81% [95,40–96,29]	95,0%	96,4%	0,993 [0,992–0,994]	0,992
S2 (bocor eksplisit)	S1 + *Z-Score H/A*	Ya	100%	100%	100%	1,000	1,000
S3 (bebas kebocoran)	**JK, Usia, BB**	**Ya**	**75,82% [75,02–76,73]**	**72,3%**	**78,2%**	**0,829 [0,821–0,838]**	**0,822**
S3 (bebas kebocoran)	JK, Usia, BB	Tidak	76,93% [76,15–77,90]	63,9%	85,7%	0,829 [0,821–0,839]	0,823
Regresi logistik	JK, Usia, BB, z BB/U	Ya	74,88%	74,1%	75,4%	0,821	–
Kelas mayoritas	–	–	59,59%	0,0%	100%	0,500	–
Aturan *WHO* (tanpa ML)	JK, Usia, TB	–	95,01%	87,9%	99,8%	0,989†	–

JK = jenis kelamin; BB = berat badan; TB = tinggi badan; *AUC CV* = rata-rata AUC validasi silang 5-*fold* pada data latih. †Z-score TB/U dipakai sebagai skor kontinu.

S1 mencapai akurasi 95,81% dan S2 100%, tetapi keduanya tidak dapat dibaca sebagai keberhasilan model: S1 hanya unggul 0,8 poin dari aturan *WHO* tanpa ML, sedangkan S2 menerima label dalam bentuk ambang sederhana. Akurasi tinggi pada dataset ini, termasuk 97,57% yang dilaporkan sebelumnya [6], perlu ditafsirkan dengan memperhatikan fitur yang digunakan. Lebar IK 95% akurasi S1 (0,9 poin) juga menunjukkan bahwa selisih sekitar 0,1 poin antara model berparameter baku dan hasil *tuning* tidak berarti.

S3 mencapai akurasi 75,82% dan *AUC-ROC* 0,829 (IK 95% 0,821–0,838), konsisten dengan *AUC* validasi silang (0,822 ± 0,007) sehingga tidak ada indikasi *overfitting*. S3 jauh di atas kelas mayoritas dan sedikit di atas regresi logistik (*AUC* 0,821), yang menunjukkan bahwa sebagian besar sinyal tanpa tinggi badan sudah tercakup dalam defisit berat menurut usia. *SMOTE* tidak mengubah *AUC*, tetapi menggeser titik operasi pada ambang 0,5 (sensitivitas 63,9% → 72,3%; spesifisitas 85,7% → 78,2%) dan membuat probabilitas lebih tinggi dari prevalensi sebenarnya (Brier 0,166 vs 0,161), sehingga skor diperlakukan sebagai penanda prioritas, bukan risiko absolut.

Pada ambang skrining 0,402, S3 mencapai sensitivitas 80,3%, spesifisitas 67,6%, PPV 62,7%, dan NPV 83,5% (Tabel 3 dan Gambar 3). Sebanyak 4.144 dari 8.003 balita (51,8%) ditandai berisiko tinggi, dan 637 dari 3.234 balita stunted (19,7%) tidak tertandai. Karena PPV dan NPV bergantung pada prevalensi, sedangkan prevalensi dataset (40,4%) jauh di atas prevalensi nasional, nilai ini tidak dapat langsung digeneralisasi. S3 berguna untuk memprioritaskan pengukuran PB/TB, bukan menggantikannya.



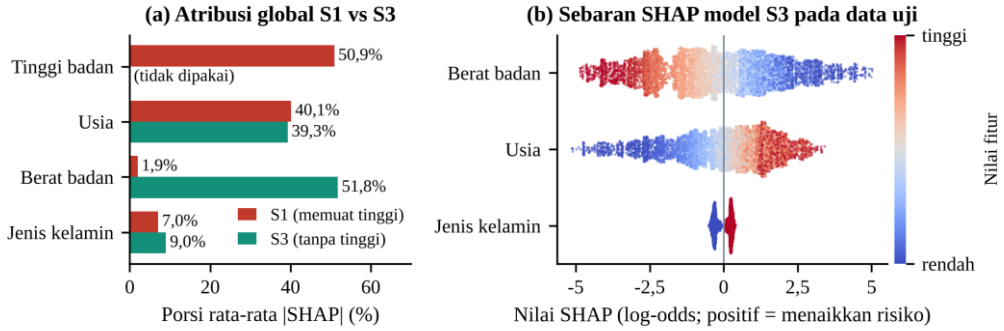
Gambar 3. (a) Kurva *ROC* Model S3 dan *Z-Score TB/U* WHO; (b) Matriks Konfusi S3 pada Ambang Skrining 0,402

Tabel 3. Kinerja Model S3 (dengan *SMOTE*) pada Dua Ambang Keputusan

Metrik	Ambang 0,5	Ambang skrining 0,402
Akurasi / *balanced accuracy*	75,82% / 75,25%	72,71% / 73,93%
Sensitivitas / spesifisitas	72,29% / 78,21%	80,30% / 67,56%
PPV / NPV	69,23% / 80,63%	62,67% / 83,49%
F1	0,707	0,704
TP / FP / FN / TN	2.338 / 1.039 / 896 / 3.730	2.597 / 1.547 / 637 / 3.222

3.3. Analisis Interpretabilitas *TreeSHAP*

Pada S1, *TreeSHAP* mengatribusikan 50,9% kontribusi pada tinggi badan dan 40,1% pada usia (Gambar 4a). Atribusi ini tidak dapat dibaca sebagai temuan klinis, karena label TB/U dibentuk dari kedua variabel tersebut; *TreeSHAP* hanya merekonstruksi rumus *z-score*. Pada S3, berat badan (51,8%) dan usia (39,3%) menjadi kontributor utama, diikuti jenis kelamin (9,0%). Nilai SHAP berat berkorelasi negatif dengan berat badan (r = −0,98) dan nilai *SHAP* usia berkorelasi positif dengan usia (r = 0,93) (Gambar 4b), sehingga model pada dasarnya menangkap defisit berat menurut usia. Hal ini konsisten dengan pedoman yang menjadikan pemantauan berat badan, termasuk berat badan tidak naik, sebagai dasar penapisan balita bermasalah gizi [25].



Gambar 4. (a) Porsi Rata-rata |*SHAP*| pada S1 dan S3;
(b) Sebaran Nilai *SHAP* S3 pada Data Uji (untuk Jenis Kelamin, Merah=Laki-laki)

Gambar 5., menampilkan analisis Kasus Contoh A (laki-laki, 40 bulan, 9,0 kg, 80,0 cm). Lapis deterministik menghasilkan *z-score* TB/U −4,80 SD (sangat pendek), BB/U −4,08 SD (berat badan sangat kurang), dan BB/TB −2,00 SD (tepat pada batas gizi kurang); status stunting berat ditetapkan dari TB/U, bukan dari model. Skor risiko S3 adalah 0,99, dengan kontribusi *TreeSHAP* berat badan +2,30, usia +1,91, dan jenis kelamin +0,18 sehingga $f(\mathbf{x}) = 4{,}40$ dan $p = 0{,}988$ sesuai Persamaan (4). Dua aturan KB terpicu, yaitu KB-02 (stunting berat) dan KB-11 (berat badan sangat kurang), masing-masing dengan rujukan pedoman, versi aturan, dan kolom keputusan tenaga kesehatan.



Gambar 5. Halaman Analisis Stunting untuk Kasus Contoh A: Status Gizi Deterministik dan Skor Risiko dengan Penjelasan *TreeSHAP*

3.4. Implementasi Sistem CDSS

Setiap aturan KB berupa kondisi terstruktur atas variabel terukur dan wajib mencantumkan dokumen rujukan (Tabel 4 dan Gambar 6). Ambang kategori mengikuti Permenkes No. 2 Tahun 2020 [16]; tindakan mengikuti Juknis PMT Berbahan Pangan Lokal edisi ketiga [25], yaitu PMT 14 hari untuk berat badan tidak naik, 28 hari untuk berat badan kurang, 56 hari untuk gizi kurang, rujukan stunting ke rumah sakit setelah satu siklus PMT, dan gizi buruk bukan sasaran PMT lokal; prioritas usia di bawah dua tahun mengikuti Perpres No. 72 Tahun 2021 [26]. Aturan yang bersumber dari model diberi label “Model CDSS” agar tidak tertukar dengan pedoman.

Information System Journal (INFOS)
Vol. 9, No. 2, November 2026, pp. 69-78
E-ISSN : 2655-142X , P-ISSN : 2655-190X, DOI:https://doi.org/10.24076/infosjournal.2026v9i02.2947

Tabel 4. Aturan *Knowledge Base* dan Rujukannya

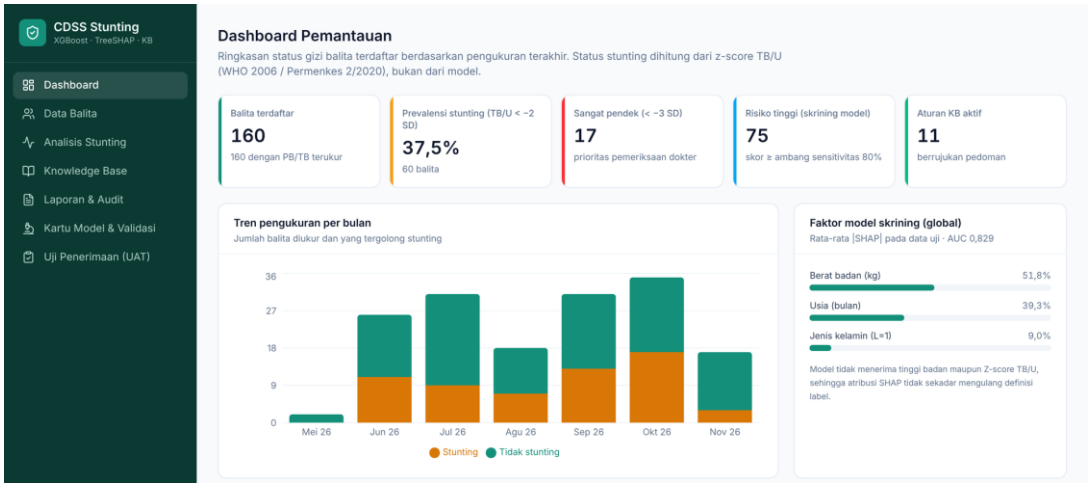
Kode	Kondisi	Tindakan utama (ringkas)	Sumber
KB-01	$-3 \leq z$ TB/U < -2	Konfirmasi PB/TB; periksa dokter; rujuk RS setelah satu siklus PMT	[16], [25]
KB-02	z TB/U < -3	Prioritas pemeriksaan dokter; rujuk RS setelah satu siklus PMT atau lebih cepat	[16], [25]
KB-03	$-3 \leq z$ BB/U < -2	Konfirmasi BB/TB; PMT lokal 28 hari; pantau BB mingguan	[16], [25]
KB-04	$-3 \leq z$ BB/TB < -2	PMT lokal 56 hari; periksa dokter; pantau BB mingguan	[16], [25]
KB-05	z BB/TB < -3	Bukan sasaran PMT lokal; tata laksana gizi buruk	[16], [25]
KB-06	BB tidak naik	PMT lokal 14 hari; rujuk RS bila tetap tidak naik	[25]
KB-07	Risiko tinggi, PB/TB belum diukur	Ukur PB/TB terstandar; skor bukan diagnosis	Model CDSS
KB-08	Risiko tinggi, z TB/U ≥ -2	Periksa ulang usia dan teknik pengukuran	Model CDSS
KB-09	Usia < 24 bulan, salah satu z < -2	Prioritas intervensi spesifik 1000 HPK	[26]
KB-10	Ketiga z ≥ -2	Pemantauan pertumbuhan rutin bulanan	[25]
KB-11	z BB/U < -3	Konfirmasi BB/TB; periksa dokter; pantau BB mingguan	[16], [25]

Setiap perubahan aturan menaikkan versi, menyimpan *snapshot*, alasan, dan pengubahnya pada tabel kb_revisi, serta mengembalikan status ke Draf; aturan baru berlaku setelah divalidasi peran Ahli Gizi, dan setiap rekomendasi mencatat versi aturannya. Dengan demikian, sistem tidak memperbarui dirinya secara otomatis; kesesuaian dengan pedoman terbaru bergantung pada proses validasi tersebut. Pada prototipe ini, sebelas aturan awal diaktifkan agar alur dapat diuji, tetapi seluruhnya ditandai "Belum divalidasi pakar".



Gambar 6. Halaman Knowledge Base dengan Alur Tata Kelola: Usulan Rujukan Wajib, Validasi Pakar, dan Jejak Audit

Modul lain mencakup *Dashboard* (Gambar 7; status gizi dari pengukuran terakhir, prevalensi per kelompok usia, dan atribusi global model), Laporan & Audit (distribusi kategori, prevalensi per desa, rekap keputusan tenaga kesehatan, dan log aktivitas), Kartu Model & Validasi (protokol, bukti kebocoran, dan kinerja seluruh skenario), serta Uji Penerimaan yang memuat instrumen pada Subbab 2.4. Hingga naskah ini disusun, UAT belum dilaksanakan sehingga belum ada data penerimaan pengguna yang dapat dilaporkan. Data pada Gambar 7 adalah 160 balita demo yang diambil acak dari data uji dengan nama samaran, bukan angka populasi; pada dataset lengkap, jumlah rekaman berlabel *Stunted* adalah 16.204 dari 40.071 (40,4%).



Gambar 7. Bagian Atas Halaman Dashboard: Indikator Utama, Tren Pengukuran, da Atribusi Global Model (Data Demo dengan Nama Samaran)

3.5. Perbandingan dan Keterbatasan

Tabel 5., membandingkan penelitian ini dengan studi terdahulu; angka kinerja tidak dapat dibandingkan langsung karena dataset, target, dan fitur berbeda. Dibandingkan studi pada dataset yang sama [6], penelitian ini tidak menghasilkan akurasi lebih tinggi, tetapi menunjukkan bahwa akurasi tinggi dapat dicapai dari fitur pembentuk label dan menyediakan model skrining yang bebas dari fitur tersebut dalam antarmuka klinis. Dibandingkan LIME [12], *TreeSHAP* memberikan atribusi eksak yang aditif terhadap keluaran model.

Tabel 4. Aturan *Knowledge Base* dan Rujukannya

Peneliti	Dataset	Metode	Kinerja	Keterbatasan	Posisi penelitian ini
[6]	Jeneponto, 40.071	*XGBoost*, *SHAP*	Akurasi 97,57%	Tanpa antarmuka klinis	Uji kebocoran; model tanpa tinggi
[8]	IFLS, 5.093	kNN, RF, LR	F1 84,5%	Data 1997–2014	Data terkini + XAI
[9]	97 rekam medis	NB, MLP, J48, SVM	Akurasi 100%	Sampel sangat kecil	8.003 data uji + IK 95%
[12]	EDHS, 11.121	Extra Tree + MLP, LIME	Akurasi 94%	LIME kurang konsisten	*TreeSHAP* eksak
[13]	DHS Mesir, 37.051	*XGBoost*, GB, RF	Akurasi 90%	Data 2014	Rekomendasi berrujukan
Penelitian ini	Jeneponto, 40.014	*XGBoost*, *TreeSHAP*, KB	AUC 0,829	Validasi internal; UAT belum	–

Nilai XAI pada kasus stunting dengan demikian terletak pada kemampuannya mengungkap kebocoran dan menjelaskan model yang memperkirakan sesuatu yang belum diukur, bukan pada konfirmasi bahwa tinggi badan dan usia penting. Keterbatasan penelitian ini meliputi validasi internal pada satu kabupaten, ketiadaan identitas anak sehingga pengukuran berbeda dari anak yang sama dapat berada di kedua partisi, probabilitas yang tidak dikalibrasi, ketiadaan variabel sosial-ekonomi, serta isi aturan KB yang belum divalidasi pakar.

4. KESIMPULAN

Penelitian ini menghasilkan prototipe CDSS berbasis web yang memisahkan penetapan status stunting secara deterministik (*z-score WHO*, Permenkes No. 2/2020) dari skor risiko skrining *XGBoost* tanpa fitur pembentuk label (*AUC* 0,829), dilengkapi penjelasan *TreeSHAP* per-pasien dan rekomendasi berbasis *knowledge base* yang berversi dan berrujukan pedoman. Label stunting pada dataset Jeneponto identik dengan aturan *Z-Score H/A* < −2 dan dapat direproduksi dengan akurasi 95,0% tanpa ML, sehingga akurasi 95,8–100% pada model yang memuat tinggi badan atau *z-score* TB/U merupakan efek kebocoran. Tanpa tinggi badan, berat badan dan usia menjadi kontributor utama (51,8% dan 39,3%), dan pada ambang skrining model mendeteksi 80,3% balita *stunted* dengan spesifisitas 67,6%. Kesiapan penggunaan klinis belum dapat disimpulkan dan akan dievaluasi melalui UAT bersama tenaga kesehatan Kabupaten Jeneponto menggunakan instrumen yang telah terintegrasi dalam sistem.

Penelitian selanjutnya disarankan untuk melaksanakan UAT dan validasi aturan KB oleh ahli gizi dan dokter, melakukan validasi eksternal dan temporal, mengkalibrasi probabilitas model, serta memanfaatkan data longitudinal dengan identitas anak dan variabel sosial-ekonomi.

DAFTAR PUSTAKA

[1] W. H. Organization, "Reducing stunting in children: equity considerations for achieving the Global Nutrition Targets 2025," 2018. Available: https://iris.who.int/bitstream/handle/10665/260202/9789241513647-eng.pdf

[2] W. H. Organization, WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. World Health Organization, 2006. Available: https://iris.who.int/bitstream/handle/10665/43413/924154693X_eng.pdf

[3] C. G. Victora, P. Christian, L. P. Vidaletti, G. Gatica-Domínguez, P. Menon, and R. E. Black, "Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda," Lancet, vol. 397, no. 10282, pp. 1388–1399, 2021. Available: https://www.thelancet.com/action/showPdf?pii=S0140-6736%2821%2900394-9

[4] R. I. Kemenkes, "Hasil survei status gizi Indonesia (SSGI) 2022," 2022, Kemenkes.

[5] W. H. O. UNICEF, "WHO, World Bank Levels and trends in child malnutrition," Jt. child malnutrition Estim. New York, NY United Nations Int. Child. Fund, 2023. Available: https://www.who.int/publications/i/item/9789240073791

[6] A. Wicaksono, D. Prasetyo, Y. Mar'atullatifah, D. U. Iswavigra, H. Mahmudah, and A. Hapsari, "Data Analysis and Explainable Machine Learning for Stunting Prediction," J. Artif. Intell. Leg. Technol., vol. 1, no. 1, pp. 35–44, 2025. Available: https://journal.sah.co.id/index.php/JAILT/article/view/4/4

[7] M. Amelia, "Determinan Stunting pada Anak di Indonesia: Sebuah Tinjauan Literatur Sistematis," Mejor. Med. J. Awatara, vol. 3, no. 4, pp. 199–205, 2025.

[8] 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, p. 718, 2025. Available: https://link.springer.com/article/10.1186/s12887-025-06096-4

[9] A. Fadlil and M. Ma'ruf, "Machine Learning-Based Early Prediction of Stunting Risk: A Comparative Study," Methods Sci. Technol. Stud., vol. 1, no. 1, pp. 24–31, 2025. Available: https://journal.futuristech.co.id/index.php/msts/article/view/351

[10] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell., vol. 1, no. 5, pp. 206–215, 2019. Available: https://www.nature.com/articles/s42256-019-0048-x

[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst., vol. 30, 2017. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[12] T. K. Wudu, A. A. Endalew, and A. A. Dires, "Explainable hybrid machine learning model for predicting stunting and identifying key risk factors among Ethiopian children under five," Sci. Rep., 2026. Available: https://www.nature.com/articles/s41598-026-46417-w

[13] A. Hendy et al., "Supervised machine learning for classification and prediction of stunting among under-five Egyptian children," BMC Pediatr., vol. 25, no. 1, p. 681, 2025. Available: https://link.springer.com/article/10.1186/s12887-025-06138-x

[14] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Trans. Knowl. Discov. Data, vol. 6, no. 4, pp. 1–21, 2012, doi: 10.1145/2382577.2382579.

[15] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.

[16] Kementerian Kesehatan Republik Indonesia, Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang Standar Antropometri Anak. Jakarta: Kementerian Kesehatan RI, 2020.

[17] World Health Organization and United Nations Children's Fund, Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. Geneva: World Health Organization, 2019.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[19] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araújo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches," IEEE Comput. Intell. Mag., vol. 13, no. 4, pp. 59–76, 2018, doi: 10.1109/MCI.2018.2866730.

[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794. Available: https://dl.acm.org/doi/abs/10.1145/2939672.2939785
[21] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.
[22] E. S. Berner and T. J. La Lande, "Overview of clinical decision support systems," in Clinical decision support systems: Theory and practice, Springer, 2016, pp. 1–17. Available: https://link.springer.com/chapter/10.1007/978-3-319-31913-1_1
[23] J. Brooke, "SUS: A 'quick and dirty' usability scale," in Usability Evaluation in Industry, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. London: Taylor & Francis, 1996, pp. 189–194.
[24] J. R. Lewis and J. Sauro, "Item benchmarks for the System Usability Scale," J. Usability Stud., vol. 13, no. 3, pp. 158–167, 2018.
[25] Direktorat Jenderal Kesehatan Primer dan Komunitas, Keputusan Direktur Jenderal Kesehatan Primer dan Komunitas Nomor HK.02.02/B/576/2025 tentang Petunjuk Teknis Pemberian Makanan Tambahan Berbahan Pangan Lokal bagi Ibu Hamil dan Balita Bermasalah Gizi. Jakarta: Kementerian Kesehatan RI, 2025. Available: https://kesprimkom.kemkes.go.id/assets/uploads/contents/others/2025Kepdirjenkesprimkom576.pdf
[26] Pemerintah Republik Indonesia, *Peraturan Presiden Republik Indonesia Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting.* Jakarta, 2021