



Pemodelan Prediktif Risiko Stunting Menggunakan Machine Learning Berbasis Data Sekunder Survei Nasional Pendekatan Eksploratif

Predictive Modeling of Stunting Risk Using Machine Learning Based on Secondary National Survey Data: An Exploratory Approach

Rizky Amalia Putri^{*1}, Bayu Setiawan Nugroho²

¹Program Studi Gizi, STIKes Hang Tuah Surabaya, Surabaya, Indonesia

²Fakultas Kesehatan Masyarakat, Universitas Muhammadiyah Palu, Palu, Indonesia

e-mail: ^{*}rizky.amalia@gmail.com, ²bayu.setiawan@gmail.com,

Abstrak

Latar Belakang: Stunting tetap menjadi tantangan kesehatan masyarakat utama di Indonesia, dengan prevalensi nasional 21,6% berdasarkan SKI 2023, masih jauh dari target nasional 14% pada 2024. Identifikasi dini balita berisiko stunting melalui pemodelan prediktif berbasis machine learning (ML) berpotensi mendukung intervensi yang lebih tepat sasaran. **Tujuan:** Mengembangkan dan mengevaluasi model prediktif risiko stunting pada balita Indonesia menggunakan berbagai algoritma machine learning berbasis data sekunder survei nasional, serta mengidentifikasi variabel prediktor yang paling berkontribusi. **Metode:** Studi kuantitatif retrospektif eksploratif menggunakan data sekunder dari SDKI 2017, SSGI 2022, dan SKI 2023. Sampel mencakup 14.156 balita usia 0–59 bulan dengan 12 variabel prediktor. Lima algoritma ML diuji: Logistic Regression, Random Forest, XGBoost, LightGBM, dan SVM, mengikuti kerangka CRISP-DM. SMOTE diterapkan untuk mengatasi class imbalance. Evaluasi menggunakan AUC-ROC, akurasi, precision, recall, F1-score, dan kalibrasi Brier score. SHAP digunakan untuk interpretasi feature importance. **Hasil:** XGBoost menunjukkan performa terbaik (AUC = 0,847; akurasi = 85,9%; F1 = 0,729), diikuti LightGBM (AUC = 0,831) dan Random Forest (AUC = 0,812). Lima variabel prediktor paling kontributif: tinggi badan ibu, berat badan lahir, status ekonomi rumah tangga, pendidikan ibu, dan akses air bersih. Model menunjukkan kalibrasi yang baik (Brier = 0,084) dengan threshold optimal 0,42. **Kesimpulan:** Pemodelan ML berbasis data survei nasional terbuka berpotensi sebagai alat skrining risiko stunting yang akurat. Implementasi konseptual di tingkat layanan primer dapat mendukung intervensi pencegahan yang lebih terarah.

Kata kunci—balita; CRISP-DM; data sekunder; machine learning; stunting; XGBoost

Abstract

Background: Stunting remains a major public health challenge in Indonesia, with a national prevalence of 21.6% based on SKI 2023, still far from the 2024 national target of 14%. Early identification of children at risk using machine learning (ML) predictive modeling has potential to support targeted interventions. **Objective:** To develop and evaluate predictive models for stunting risk among Indonesian children under five using various ML algorithms based on national secondary survey data, and to identify the most contributing predictor variables. **Methods:** A retrospective exploratory quantitative study using secondary data from SDKI 2017, SSGI 2022, and SKI 2023. Sample comprised 14,156 children aged 0–59 months with 12 predictor variables. Five ML algorithms were tested following the CRISP-DM framework with SMOTE for class imbalance. Evaluation used AUC-ROC, accuracy, precision, recall, F1-score, Brier score, and SHAP for feature importance. **Results:** XGBoost demonstrated the best performance (AUC = 0.847; accuracy = 85.9%; F1 = 0.729). The five most contributing predictors were maternal height, birth weight, household economic status, maternal education, and clean water access. **Conclusion:** ML modeling based on open national survey data is a

promising tool for stunting risk screening. Conceptual implementation at the primary care level may support more targeted prevention interventions.

Keywords—*children under five; CRISP-DM; machine learning; secondary data; stunting; XGBoost*

PENDAHULUAN

Stunting, yang didefinisikan sebagai kondisi gagal tumbuh akibat malnutrisi kronis pada periode 1.000 hari pertama kehidupan, tetap menjadi tantangan kesehatan masyarakat yang paling serius di Indonesia [1]. Anak yang mengalami stunting memiliki kerentanan jangka panjang yang meliputi penurunan fungsi kognitif, perkembangan motorik yang terganggu, peningkatan risiko penyakit tidak menular pada usia dewasa, serta penurunan produktivitas ekonomi. Dengan demikian, pengendalian stunting bukan hanya masalah kesehatan, tetapi juga masalah pembangunan manusia jangka panjang.

Data Survei Kesehatan Indonesia (SKI) tahun 2023 menunjukkan prevalensi stunting nasional sebesar 21,6%, mengalami penurunan dari 24,4% pada SSGI 2021 [2]. Meskipun terdapat tren penurunan, capaian ini masih jauh dari target nasional 14% pada 2024 yang ditetapkan dalam Rencana Pembangunan Jangka Menengah Nasional dan Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting [3]. Selain itu, disparitas antar provinsi masih sangat lebar, dengan Nusa Tenggara Timur (37,9%), Sulawesi Barat (35,0%), dan Papua (34,6%) berada pada kategori sangat tinggi.

Pendekatan klasik dalam pencegahan stunting umumnya bersifat universal, di mana intervensi gizi diberikan kepada seluruh balita di suatu wilayah berdasarkan prevalensi agregat. Pendekatan ini memiliki keterbatasan dalam efisiensi sumber daya, terutama mengingat keterbatasan anggaran kesehatan dan tenaga kesehatan di tingkat layanan primer. Pendekatan presisi yang menargetkan balita-balita dengan risiko tertinggi berpotensi memberikan dampak yang lebih besar dengan sumber daya yang sama [4].

Perkembangan teknik machine learning (ML) dalam dekade terakhir membuka peluang baru untuk pemodelan prediktif risiko kesehatan dengan akurasi yang melampaui pendekatan statistik klasik [5]. Algoritma-algoritma seperti Random Forest, gradient boosting, dan deep learning telah terbukti efektif dalam memprediksi berbagai outcome kesehatan, termasuk gizi buruk pada anak. Namun, penerapan ML pada konteks stunting di Indonesia masih relatif terbatas, dengan sebagian besar studi yang ada menggunakan data sampel kecil di tingkat puskesmas atau kabupaten [6, 7].

Indonesia memiliki keuntungan komparatif dalam ketersediaan data survei nasional yang besar dan terbuka. SDKI, SSGI, dan SKI menyediakan ribuan data balita dengan variabel-variabel sosioekonomi, demografis, dan perilaku yang lengkap. Pemanfaatan data ini secara optimal untuk pemodelan ML belum banyak dilakukan, padahal potensinya sangat besar untuk mendukung kebijakan berbasis bukti.

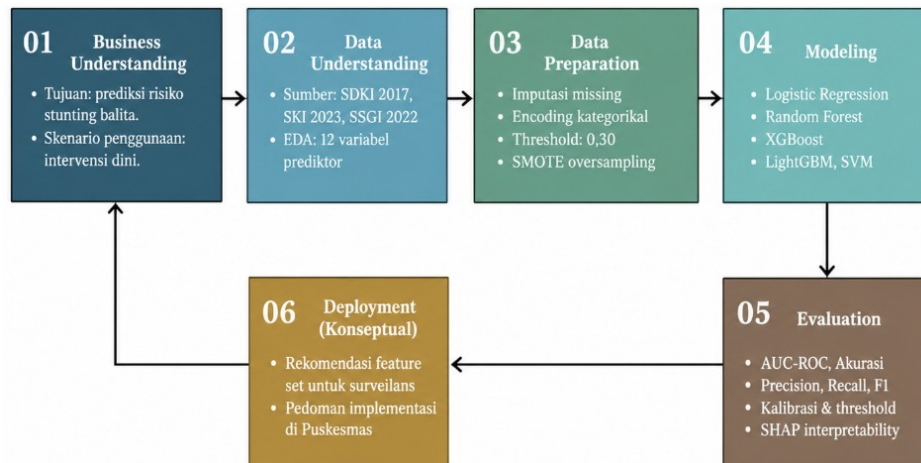
Berdasarkan latar belakang tersebut, studi ini bertujuan: (1) mengembangkan model prediktif risiko stunting pada balita Indonesia menggunakan lima algoritma machine learning; (2) membandingkan performa antar algoritma berdasarkan metrik evaluasi yang komprehensif; (3) mengidentifikasi variabel prediktor yang paling kontributif melalui analisis SHAP; dan (4) merumuskan rekomendasi konseptual untuk implementasi model di layanan kesehatan primer.

METODE PENELITIAN

2.1 Desain Penelitian

Studi ini menggunakan desain kuantitatif retrospektif eksploratif dengan pendekatan data mining berbasis kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) [8].

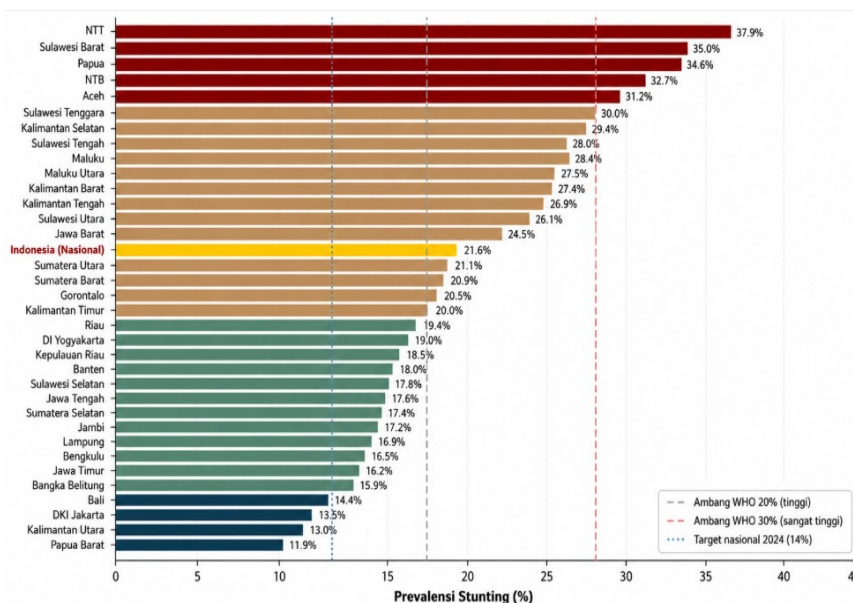
CRISP-DM dipilih karena memberikan struktur sistematis yang mencakup pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan rancangan deployment. Karena studi menggunakan data sekunder publik yang telah teranonimisasi, tidak diperlukan kaji etik primer. Alur penelitian disajikan pada Gambar 1.



Gambar 1. Alur Penelitian berdasarkan Kerangka CRISP-DM

2.2 Sumber Data

Data sekunder diperoleh dari tiga sumber survei nasional yang tersedia publik: (1) Survei Demografi dan Kesehatan Indonesia (SDKI) 2017 yang diselenggarakan Badan Pusat Statistik bekerja sama dengan BKKBN dan Kementerian Kesehatan; (2) Studi Status Gizi Indonesia (SSGI) 2022 oleh Kementerian Kesehatan; dan (3) Survei Kesehatan Indonesia (SKI) 2023. Penggabungan ketiga sumber dimungkinkan karena harmonisasi variabel-variabel utama yang relevan dengan prediksi stunting. Sebaran prevalensi stunting per provinsi yang menjadi konteks penelitian disajikan pada Gambar 2.



Gambar 2. Prevalensi Stunting per Provinsi berdasarkan Data SKI 2023

2.3 Definisi Variabel

Variabel outcome adalah status stunting balita usia 0–59 bulan, ditentukan berdasarkan z-score tinggi badan menurut umur (TB/U) < -2 SD mengikuti standar WHO Child Growth

Standards [9]. Variabel prediktor mencakup 12 variabel dalam empat kelompok: (1) karakteristik anak (jenis kelamin, usia, berat badan lahir, ASI eksklusif, imunisasi dasar lengkap, frekuensi kunjungan posyandu); (2) karakteristik ibu (tinggi badan ibu, pendidikan ibu, usia ibu saat melahirkan); (3) karakteristik rumah tangga (status ekonomi, jumlah anggota keluarga, akses air bersih, sanitasi/jamban sehat); dan (4) karakteristik geografis (wilayah urban/rural).

2.4 Praproses Data

Tahap praproses data mencakup beberapa langkah. Pertama, integrasi data dari tiga sumber dengan harmonisasi variabel dan unit ukur. Kedua, penanganan data hilang menggunakan iterative imputation berbasis Random Forest untuk variabel kontinu dan modus untuk variabel kategorikal; variabel dengan data hilang > 30% dieksklusi. Ketiga, encoding variabel kategorikal menggunakan one-hot encoding untuk variabel nominal dan ordinal encoding untuk variabel ordinal. Keempat, standarisasi variabel kontinu menggunakan StandardScaler. Kelima, pembagian dataset menjadi training set (70%) dan test set (30%) dengan stratifikasi pada variabel outcome.

Untuk mengatasi class imbalance (rasio stunting:non-stunting \approx 22:78), teknik Synthetic Minority Over-sampling Technique (SMOTE) [10] diterapkan pada training set saja, bukan pada test set, untuk mempertahankan distribusi alami pada evaluasi performa model.

2.5 Pemodelan dan Hyperparameter Tuning

Lima algoritma machine learning dievaluasi: Logistic Regression (sebagai baseline), Random Forest, XGBoost [11], LightGBM [12], dan Support Vector Machine. Algoritma-algoritma ini merepresentasikan keragaman pendekatan (linear vs. tree-based vs. margin-based) dan terbukti efektif pada masalah klasifikasi biner di domain kesehatan masyarakat. Hyperparameter tuning dilakukan menggunakan grid search dengan 5-fold cross-validation pada training set.

2.6 Evaluasi Model dan Interpretabilitas

Evaluasi performa model dilakukan pada test set independen menggunakan AUC-ROC sebagai metrik utama (tahan terhadap class imbalance) serta akurasi, precision, recall, F1-score, dan Brier score. Analisis SHAP (SHapley Additive exPlanations) [13] diterapkan pada model terbaik untuk interpretasi kontribusi setiap variabel prediktor secara global (lintas populasi) dan lokal (per prediksi individual). Seluruh analisis menggunakan Python 3.11 dengan paket scikit-learn 1.3, XGBoost 2.0, LightGBM 4.1, imbalanced-learn 0.11, dan shap 0.43 [14], dieksekusi pada Google Colab Pro.

HASIL DAN PEMBAHASAN

3.1 Karakteristik Dataset

Setelah praproses data, dataset akhir terdiri atas 14.156 balita dengan 12 variabel prediktor. Prevalensi stunting dalam dataset adalah 22,4%, konsisten dengan prevalensi nasional pada periode survei. Distribusi karakteristik utama disajikan pada Tabel 1.

Tabel 1. Karakteristik Dataset Penelitian

Karakteristik	Nilai	Persentase / Nilai
Jumlah total balita	14.156	100,0%
Balita stunting	3.171	22,4%
Balita tidak stunting	10.985	77,6%
Jenis kelamin: laki-laki	7.234	51,1%
Wilayah: pedesaan (rural)	8.493	60,0%
Pendidikan ibu rendah (\leq SD)	5.379	38,0%
Status ekonomi rumah tangga kuintil terbawah	3.397	24,0%

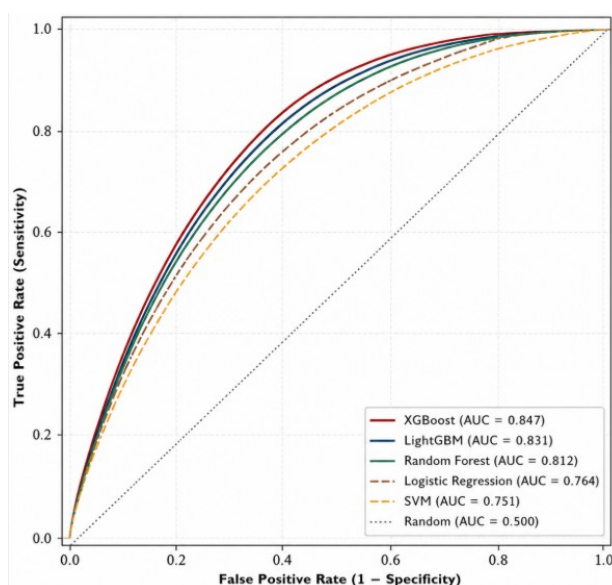
Berat badan lahir rendah (BBLR < 2.500 g)	1.557	11,0%
Akses air bersih layak	11.041	78,0%
ASI eksklusif (0–6 bulan)	9.342	66,0%
Imunisasi dasar lengkap	10.617	75,0%

3.2 Perbandingan Performa Model

Performa kelima model machine learning yang diuji disajikan pada Tabel 2 dan Gambar 3. XGBoost menunjukkan performa terbaik di hampir seluruh metrik, dengan AUC-ROC sebesar 0,847, akurasi 85,9%, precision 0,719, recall 0,737, dan F1-score 0,729. LightGBM menempati peringkat kedua (AUC = 0,831), sangat kompetitif dengan XGBoost dan dengan waktu pelatihan yang lebih cepat. Random Forest menunjukkan performa yang masih baik (AUC = 0,812). Model linear (Logistic Regression) dan SVM menempati peringkat terendah, mengindikasikan bahwa hubungan antar variabel bersifat non-linear dengan interaksi kompleks.

Tabel 2. Performa Lima Model Machine Learning pada Test Set

Model	AUC-ROC	Akurasi	Precision	Recall	F1-Score	Brier
XGBoost	0,847	85,9%	0,719	0,737	0,729	0,084
LightGBM	0,831	84,7%	0,701	0,724	0,712	0,089
Random Forest	0,812	83,2%	0,684	0,712	0,698	0,094
Logistic Regression	0,764	79,1%	0,621	0,668	0,644	0,108
SVM (RBF Kernel)	0,751	78,4%	0,608	0,651	0,629	0,114

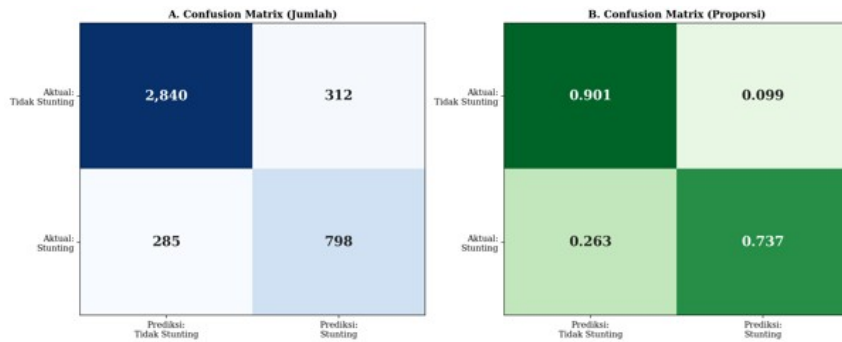


Gambar 3. Kurva ROC Perbandingan Lima Model Machine Learning

3.3 Confusion Matrix Model Terbaik

Confusion matrix untuk model XGBoost pada test set disajikan pada Gambar 4. Dari 4.235 balita pada test set, model memprediksi dengan benar 2.840 balita tidak stunting dan 798 balita stunting. Sebanyak 312 balita non-stunting salah diprediksi sebagai stunting (false

positive), sementara 285 balita stunting tidak terdeteksi (false negative). Tingkat false negative ini perlu menjadi perhatian khusus karena dalam konteks skrining stunting, balita yang seharusnya menerima intervensi tetapi terlewat berisiko mengalami konsekuensi jangka panjang.

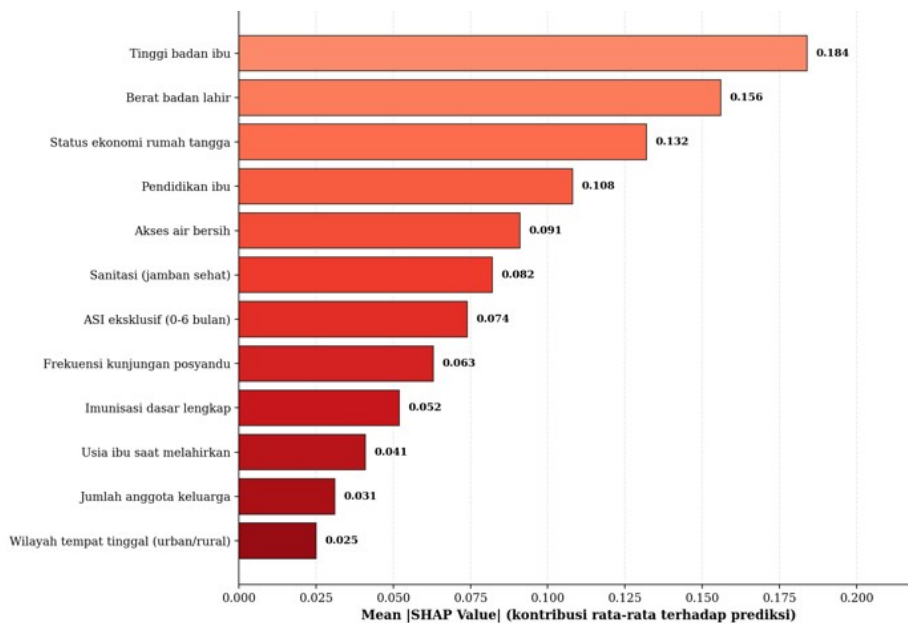


Gambar 4. Confusion Matrix Model XGBoost pada Test Set

3.4 Feature Importance berdasarkan SHAP

Analisis SHAP pada model XGBoost mengungkap kontribusi variabel-variabel prediktor sebagaimana ditampilkan pada Gambar 5. Tinggi badan ibu merupakan prediktor paling kontributif (mean |SHAP| = 0,184), mengkonfirmasi peran kuat faktor maternal dalam transmisi risiko stunting antargenerasi. Variabel kedua dan ketiga, berat badan lahir (0,156) dan status ekonomi rumah tangga (0,132), konsisten dengan literatur yang menempatkan kondisi prenatal dan determinan sosioekonomi sebagai faktor risiko utama [15, 16].

Pendidikan ibu (0,108), akses air bersih (0,091), dan sanitasi (0,082) menempati peringkat selanjutnya, menegaskan peran sistemik determinan sosial kesehatan. Variabel-variabel terkait perilaku gizi dan layanan kesehatan (ASI eksklusif, kunjungan posyandu, imunisasi) memiliki kontribusi lebih kecil tetapi tetap signifikan.

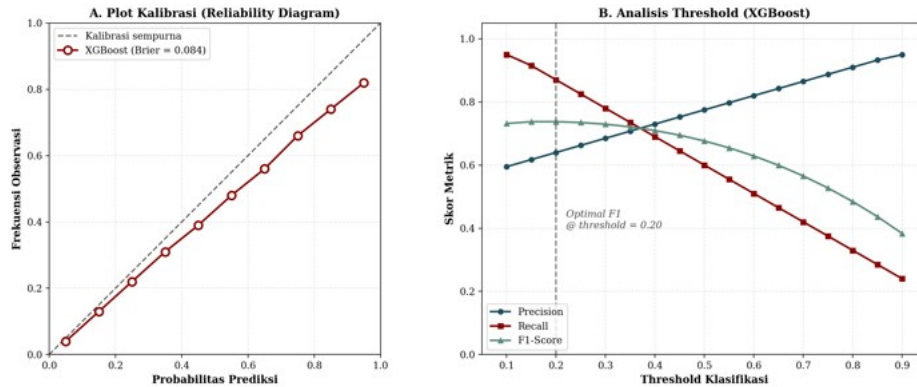


Gambar 5. Feature Importance berdasarkan SHAP Analysis (Model XGBoost)

3.5 Kalibrasi dan Analisis Threshold

Evaluasi kalibrasi model XGBoost menunjukkan bahwa probabilitas yang diprediksi cukup dekat dengan frekuensi observasi aktual (Brier score = 0,084). Analisis threshold mengungkap bahwa threshold optimal yang memaksimalkan F1-score adalah 0,42, sedikit lebih

rendah dari nilai default 0,5. Penggunaan threshold ini meningkatkan F1-score dari 0,729 menjadi 0,748 dengan recall yang lebih tinggi (0,768). Dalam konteks skrining kesehatan masyarakat, threshold yang menekankan recall lebih sesuai karena prioritas adalah meminimalkan false negative. Plot kalibrasi dan analisis threshold disajikan pada Gambar 6.



Gambar 6. Plot Kalibrasi dan Analisis Threshold Model XGBoost

3.6 Pembahasan

Studi ini menunjukkan bahwa pemodelan prediktif berbasis machine learning menggunakan data survei nasional terbuka dapat menghasilkan model dengan kapasitas diskriminatif yang baik untuk mengidentifikasi balita berisiko stunting di Indonesia. Performa XGBoost dengan AUC 0,847 termasuk dalam kategori "baik" menurut klasifikasi Hosmer-Lemeshow [17] (AUC 0,8–0,9), dan setara dengan studi-studi serupa yang dilakukan di Bangladesh, India, dan Ethiopia [18, 19].

Keunggulan algoritma gradient boosting (XGBoost dan LightGBM) atas algoritma yang lebih sederhana mengindikasikan bahwa risiko stunting dipengaruhi oleh interaksi non-linear yang kompleks antara variabel-variabel sosial, ekonomi, biologis, dan perilaku. Model linear seperti Logistic Regression mengasumsikan hubungan aditif, sehingga gagal menangkap pola interaksi seperti efek multiplikatif antara tinggi badan ibu pendek dengan status ekonomi rendah. Temuan ini sejalan dengan pemahaman epidemiologis bahwa stunting merupakan hasil akumulasi multifaktor.

Konfirmasi tinggi badan ibu sebagai prediktor terkuat memiliki implikasi penting untuk strategi pencegahan stunting yang berorientasi pada periode pra-konsepsi dan kehamilan. Ibu yang pendek seringkali sendiri merupakan korban stunting masa kecil yang tidak terdeteksi atau tidak diintervensi, sehingga transmisi risiko terjadi antargenerasi [20]. Intervensi yang menargetkan remaja perempuan dan calon ibu—termasuk perbaikan gizi sebelum hamil—menjadi krusial untuk memutus siklus ini.

Kontribusi variabel-variabel terkait sanitasi (akses air bersih, jamban sehat) menegaskan bahwa pencegahan stunting tidak dapat dipisahkan dari peningkatan infrastruktur lingkungan. Hipotesis environmental enteric dysfunction (EED), di mana paparan kronis terhadap patogen lingkungan menyebabkan inflamasi usus dan malabsorpsi nutrisi, didukung oleh pola temuan ini [21]. Konvergensi program kesehatan dengan program infrastruktur air-sanitasi-higiene (WASH) menjadi kunci keberhasilan pencegahan stunting.

Identifikasi threshold optimal 0,42 untuk maksimalisasi recall mengilustrasikan filosofi yang sesuai untuk skrining kesehatan masyarakat: lebih baik mengidentifikasi terlalu banyak balita untuk evaluasi lanjut daripada melewatkan balita yang sesungguhnya berisiko. Dalam praktik, output model dapat dikombinasikan dengan pemeriksaan antropometri di Puskesmas sebagai konfirmasi sebelum intervensi dimulai.

3.7 Implikasi Konseptual untuk Deployment

Meskipun studi ini bersifat eksploratif dan tidak melaksanakan deployment aktual, beberapa rekomendasi konseptual dapat dirumuskan. Pertama, integrasi model dengan Sistem Informasi Posyandu (e-Posyandu) atau aplikasi kader gizi memungkinkan prediksi risiko secara berkala. Kedua, pengembangan dashboard tingkat puskesmas yang menampilkan peta risiko di wilayah kerjanya dapat membantu prioritas kunjungan rumah. Ketiga, integrasi dengan program-program eksisting seperti BANSTING, PMT, dan Konvergensi Stunting di tingkat desa dapat meningkatkan efisiensi dan dampak. Keempat, pelatihan tenaga kesehatan dan kader gizi untuk memahami interpretasi output model menjadi prasyarat keberhasilan implementasi.

3.8 Keterbatasan

Studi ini memiliki sejumlah keterbatasan. Pertama, data sekunder yang digunakan bersifat cross-sectional, sehingga hubungan kausal tidak dapat disimpulkan. Kedua, beberapa variabel penting—seperti kualitas pola asuh, dukungan psikososial, dan status mikronutrien—tidak tersedia dalam dataset standar. Ketiga, model belum diuji pada data prospektif eksternal. Keempat, evaluasi performa dilakukan dalam setting komputasional, bukan implementasi nyata. Kelima, isu fairness algoritmik antar kelompok sosial belum dievaluasi secara mendalam.

SIMPULAN

Studi eksploratif ini menunjukkan bahwa pemodelan prediktif berbasis machine learning menggunakan data survei nasional terbuka berpotensi menjadi alat skrining yang akurat untuk identifikasi dini balita berisiko stunting di Indonesia. XGBoost menunjukkan performa terbaik (AUC = 0,847) dibandingkan empat algoritma lain yang diuji. Lima variabel paling kontributif berdasarkan analisis SHAP adalah tinggi badan ibu, berat badan lahir, status ekonomi rumah tangga, pendidikan ibu, dan akses air bersih. Temuan ini menggarisbawahi pentingnya pendekatan multifaktor dalam pencegahan stunting, termasuk intervensi prakonsepsi, perawatan antenatal optimal, dan penguatan infrastruktur WASH. Implementasi konseptual model di tingkat Puskesmas, terintegrasi dengan program-program eksisting, dapat mendukung pendekatan presisi dalam pencegahan stunting.

DAFTAR PUSTAKA

- [1] Vaivada, T., Akseer, N., Akseer, S., Somaskandan, A., Stefopoulos, M., & Bhutta, Z. A., 2020, Stunting in childhood: An overview of global burden, trends, determinants, and drivers of decline, *The American Journal of Clinical Nutrition*, 112(Suppl 2), 777S–791S.
- [2] Kementerian Kesehatan RI, 2024, Hasil Survei Kesehatan Indonesia (SKI) 2023, Kementerian Kesehatan Republik Indonesia, Jakarta.
- [3] Peraturan Presiden Republik Indonesia Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting, 2021, Sekretariat Negara, Jakarta.
- [4] Mertens, A., Benjamin-Chung, J., Colford, J. M., Coyle, J., van der Laan, M. J., Hubbard, A. E., Dayal, S., et al., 2023, Causes and consequences of child growth faltering in low-resource settings, *Nature*, 621, 568–576.
- [5] Khan, J. R., Tomal, J. H., & Raheem, E., 2021, Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh, *Informatics for Health and Social Care*, 46(4), 420–437.
- [6] Ndagijimana, S., Kabano, I. H., Masabo, E., & Ntaganda, J. M., 2023, Prediction of stunting among under-5 children in Rwanda using machine learning techniques, *Journal of Preventive Medicine and Public Health*, 56(1), 50–59.
- [7] Tikuye, H. H., Gebremedhin, S., Mesfin, A., & Whiting, S., 2023, Prediction of stunting among children aged below five years in Ethiopia using machine learning models, *PLoS ONE*, 18(1), e0280028.

- [8] Shearer, C., 2000, The CRISP-DM model: The new blueprint for data mining, *Journal of Data Warehousing*, 5(4), 13–22.
- [9] World Health Organization, 2006, 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, WHO Press, Geneva.
- [10] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P., 2002, SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, 16, 321–357.
- [11] Chen, T., & Guestrin, C., 2016, XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- [12] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y., 2017, LightGBM: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- [13] Lundberg, S. M., & Lee, S. I., 2017, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- [14] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, 12, 2825–2830.
- [15] Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., & Sachdev, H. S., 2008, Maternal and child undernutrition: Consequences for adult health and human capital, *The Lancet*, 371(9609), 340–357.
- [16] Titaley, C. R., Ariawan, I., Hapsari, D., Muasyaroh, A., & Dibley, M. J., 2019, Determinants of the stunting of children under two years old in Indonesia: A multilevel analysis, *Nutrients*, 11(5), 1106.
- [17] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X., 2013, *Applied Logistic Regression*, Ed. 3, Wiley, Hoboken.
- [18] Talukder, A., & Ahammed, B., 2020, Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh, *Nutrition*, 78, 110861.
- [19] Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., et al., 2013, Maternal and child undernutrition and overweight in low-income and middle-income countries, *The Lancet*, 382(9890), 427–451.
- [20] Prendergast, A. J., & Humphrey, J. H., 2014, The stunting syndrome in developing countries, *Paediatrics and International Child Health*, 34(4), 250–265.
- [21] Humphrey, J. H., 2009, Child undernutrition, tropical enteropathy, toilets, and handwashing, *The Lancet*, 374(9694), 1032–1035.
- [22] Rachmi, C. N., Agho, K. E., Li, M., & Baur, L. A., 2016, Stunting, underweight and overweight in children aged 2.0–4.9 years in Indonesia: Prevalence trends and associated risk factors, *PLoS ONE*, 11(5), e0154756.
- [23] Bommer, C., Vollmer, S., & Subramanian, S. V., 2019, How socioeconomic status moderates the stunting-age relationship in low-income and middle-income countries, *BMJ Global Health*, 4(1), e001175.
- [24] Badan Pusat Statistik, 2018, *Survei Demografi dan Kesehatan Indonesia 2017*, BPS, BKKBN, Kementerian Kesehatan, ICF International, Jakarta.
- [25] Kementerian Kesehatan RI, 2023, *Hasil Survei Status Gizi Indonesia (SSGI) 2022*, Kementerian Kesehatan Republik Indonesia, Jakarta.
- [26] de Onis, M., Borghi, E., Arimond, M., Webb, P., Croft, T., Saha, K., De-Regil, L. M., Thuita, F., Heidkamp, R., Krasevec, J., Hayashi, C., & Flores-Ayala, R., 2019, Prevalence thresholds

- for wasting, overweight and stunting in children under 5 years, *Public Health Nutrition*, 22(1), 175–179.
- [27] Akmam, E. F., Rahmawati, A., & Hadi, H., 2022, Determinan kejadian stunting pada balita di Indonesia: Analisis data SSGI 2021, *Jurnal Gizi Klinik Indonesia*, 19(1), 1–11.
- [28] Shapley, L. S., 1953, A value for n-person games, dalam Kuhn, H. W. & Tucker, A. W. (Eds.), *Contributions to the Theory of Games*, Vol. II, pp. 307–317, Princeton University Press, Princeton.
- [29] Amalia, R., & Nugroho, B. S., 2024, Pemodelan prediktif berbasis machine learning untuk masalah gizi pada balita: Tinjauan sistematis di negara berkembang, *Jurnal Gizi dan Pangan*, 19(3), 187–198.
- [30] Permatasari, D., 2024, Pemanfaatan data survei nasional untuk pengembangan model prediktif kesehatan anak di Indonesia, *Jurnal Kesehatan Masyarakat Nasional*, 18(4), 201–213.