Populer: Jurnal Penelitian Mahasiswa
Volume. 5 Nomor. 3 September 2026
E-ISSN: 2963-5306, P-ISSN: 2962-116X, Hal 96-102
DOI: https://doi.org/10.58192/populer.v5i3.5468
Tersedia: https://journal.unimar-amni.ac.id/index.php/Populer

Implementasi Prediksi Stunting Menggunakan Algoritma Xgboost dan Smote Berdasarkan Data BKKBN

Sabila Wahyu Pratama[1], Ghufron[2]
[1-2]Universitas Islam Sultan Agung
Email: sabilarealme2283@gmail.com[1], ghufron@unissula.ac.id[2]

**Penulis Korespondensi: sabilarealme2283@gmail.com*

Abstract. *The management and intervention of stunting cases in toddlers still face major challenges due to the high class imbalance in risk assessment data, where the population of normal toddlers significantly dominates the at-risk class. This condition often causes machine learning classification models to suffer from prediction bias and reduced accuracy. This study aims to build a Machine Learning classification model based on the XGBoost algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE) to predict stunting risk and analyze its primary key determinants (Feature Importance) using Stunting Risk Family (KRS) data from BKKBN Central Java. The dataset comprises 7,605 family data records with evaluation performed on 2,771 pure test data points. The evaluation results demonstrate that the combination of XGBoost and SMOTE achieved an Accuracy of 88.20% and a Precision of 99.91%. Furthermore, Gain analysis confirms that Gender (score 0.3232) and Sanitation/Latrine Facilities (score 0.2182) are the two most dominant risk factors for stunting. These findings demonstrate the potential of leveraging artificial intelligence systems to support efficient decision-making for healthcare workers within a smart governance framework.*

Keywords: *Feature Importance; Machine Learning; SMOTE; Stunting; XGBoost.*

Abstrak. Pengelolaan dan penanganan kasus stunting pada balita masih menghadapi tantangan besar akibat tingginya ketimpangan distribusi data (*imbalanced data*) pada rekam medis keluarga berisiko, di mana populasi kelas balita normal jauh lebih mendominasi dibandingkan kelas berisiko stunting. Kondisi tersebut dapat menyebabkan model klasifikasi kecerdasan buatan mengalami kecenderungan bias dan menurunkan akurasi prediksi. Penelitian ini bertujuan membangun model klasifikasi *Machine Learning* berbasis algoritma XGBoost yang dikombinasikan dengan teknik penyeimbangan data *Synthetic Minority Over-sampling Technique* (SMOTE) untuk memprediksi risiko stunting serta menganalisis faktor penentu utamanya (*Feature Importance*) berdasarkan data Keluarga Berisiko Stunting (KRS) BKKBN Provinsi Jawa Tengah. Dataset yang digunakan mencakup 7.605 rekam data keluarga dengan pengujian pada 2.771 data uji murni. Hasil evaluasi menunjukkan kombinasi XGBoost dan SMOTE berhasil mencapai nilai Akurasi sebesar 88,20% dan Presisi sebesar 99,91%. Selain itu, analisis Gain mengonfirmasi bahwa Jenis Kelamin (skor 0,3232) dan Sanitasi/Fasilitas Tempat BAB (skor 0,2182) merupakan dua faktor pemicu paling dominan terhadap risiko stunting. Temuan ini membuktikan potensi pemanfaatan sistem kecerdasan buatan dalam mendukung efisiensi pengambilan keputusan bagi petugas kesehatan berbasis *smart governance*.

Kata kunci: *Feature Importance; Machine Learning; SMOTE; Stunting; XGBoost.*

1. LATAR BELAKANG

Stunting, atau kondisi gagal tumbuh pada balita akibat kekurangan gizi kronis, merupakan salah satu masalah kesehatan masyarakat paling serius di Indonesia. Dampak stunting tidak hanya terbatas pada pertumbuhan fisik yang terhambat, tetapi juga memengaruhi perkembangan kognitif dan meningkatkan risiko penyakit kronis di masa depan. Berdasarkan buku resmi *Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka* yang dirilis oleh Badan Kebijakan Pembangunan Kesehatan Kementerian Kesehatan RI, prevalensi stunting nasional berhasil diturunkan menjadi 19,8% (Kementerian Kesehatan RI, 2025). Di tingkat lokal, Kota Semarang telah berhasil menekan angka stunting hingga mencapai 10,3%. Meskipun sudah

Naskah Masuk: 18 Juli 2026; Revisi: 25 Agustus 2026; Diterima: 18 September 2026; Tersedia : 21 September 2026

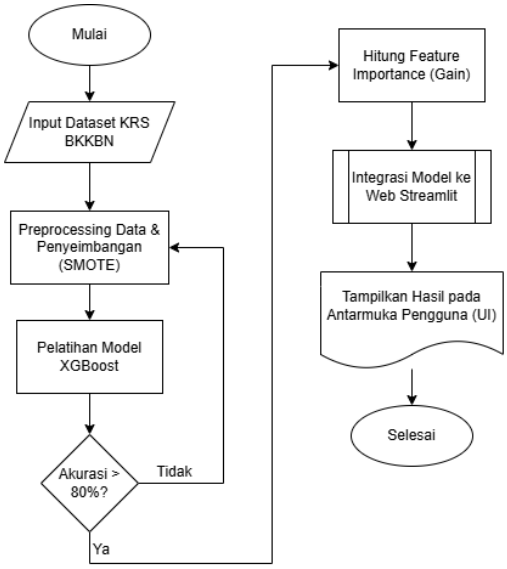
berada di bawah rata-rata nasional, pemetaan sisa kasus di lapangan tetap krusial. Terlebih lagi, studi analisis spasial terbaru oleh (Nugroho dkk., 2026) membuktikan bahwa determinan stunting di wilayah Jawa Tengah sangat heterogen dan bervariasi secara lokal antar kabupaten/kota, sehingga memerlukan pendekatan yang lebih spesifik.

Dalam implementasi percepatan penurunan stunting di lapangan, intervensi seringkali terkendala oleh efisiensi dan akurasi sasaran. Penanganan selama ini cenderung bersifat reaktif dan disamaratakan (misalnya hanya berfokus pada Pemberian Makanan Tambahan), padahal literatur membuktikan bahwa kondisi sanitasi dan higiene yang buruk juga menjadi faktor risiko hulu yang sangat signifikan memicu infeksi dan mengganggu penyerapan nutrisi (Putri & Lismayanti, 2025).

Untuk memecahkan kerumitan masalah yang bersifat multidimensional ini, pendekatan berbasis data (*data-driven*) menggunakan *Machine Learning* (ML) sangat diperlukan. Algoritma canggih seperti XGBoost terbukti sangat efektif untuk klasifikasi data medis yang kompleks dibandingkan metode statistik konvensional (Chen & Guestrin, 2016); Pangestu & Mujiyono, 2025). Penelitian ini tidak hanya menguji performa XGBoost yang dipadukan dengan teknik SMOTE untuk menyeimbangkan data (Sugihartono dkk., 2025), tetapi juga membongkar pola penyebab melalui analisis *Feature Importance*. Pendekatan ini diharapkan mampu memberikan panduan bagi pemerintah daerah untuk mengeksekusi program intervensi yang presisi dan tepat sasaran.

2. KAJIAN TEORITIS

Deteksi dini stunting merupakan masalah klasifikasi yang kerap berhadapan dengan ketimpangan kelas data (*imbalanced class*), di mana jumlah balita sehat jauh lebih banyak daripada yang stunting. Algoritma *Extreme Gradient Boosting* (XGBoost) dipilih karena arsitekturnya mengintegrasikan teknik regularisasi L1 dan L2 yang efektif untuk mencegah *overfitting*, serta smampu menangani nilai data yang hilang secara otomatis (Chen & Guestrin, 2016). Berbagai studi komparasi, seperti yang dilakukan oleh (Hadisuwarno dkk., 2025) (Puspita dkk., 2025), secara konsisten menempatkan XGBoost sebagai algoritma dengan performa prediksi stunting yang superior dibandingkan *Random Forest* maupun *Decision Tree*.



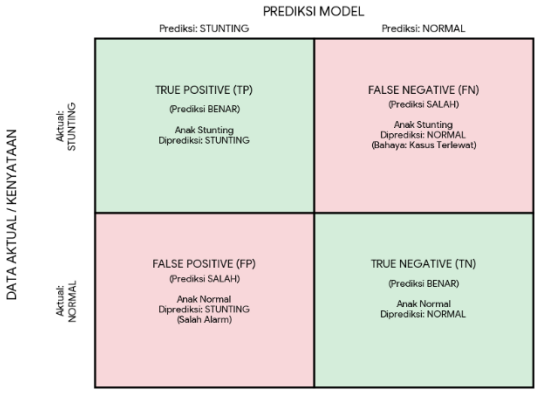
Gambar 1. Alur Pemrosesan Data XGBoost

Secara arsitektural, alur pemrosesan data XGBoost memproses vektor fitur masukan melalui serangkaian *decision tree* yang dibangun secara bertahap (*boosting*), di mana setiap pohon mengoreksi *residual error* dari pohon sebelumnya. Untuk menghindari bias akurasi mayoritas, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan guna membangkitkan data sintetis pada kelas minoritas (Sugihartono dkk., 2025). Lebih lanjut, model ini menerapkan konsep *Explainable AI* (XAI) melalui metode perhitungan *Gain*. Seperti yang ditekankan oleh (Puspanagara, 2025), transparansi model AI melalui ekstraksis *Feature Importance* adalah syarat mutlak agar keputusan algoritma (yang umumnya bersifat *black-box*) dapat diinterpretasikan dan dipercaya oleh praktisi di lapangan.

3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif eksperimental. Sumber data sekunder berupa rekapan Keluarga Berisiko Stunting (KRS) Tahun 2025 diperoleh dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Provinsi Jawa Tengah. Dataset mentah disaring khusus untuk balita berusia 0-60 bulan, menghasilkan total 7.605 baris data bersih setelah melalui tahap *preprocessing* dan *feature encoding*. Dataset ini memuat 11 variabel prediktor (fitur demografi, sosial-ekonomi, dan sanitasi) serta 1 variabel target (Risiko Stunting).

Ketimpangan data diatasi menggunakan teknik SMOTE (Sugihartono dkk., 2025), lalu dipartisi menggunakan rasio 80:20 (80% atau 11.081 baris sebagai Data Latih, dan 20% atau 2.771 baris sebagai Data Uji murni). Model XGBoost dilatih dan dievaluasi kinerjanya menggunakan metode *Confusion Matrix* (Kase & Prasetyo, 2025).



Gambar 2. Matriks Evaluasi Klasifikasi Deteksi Risiko Stunting

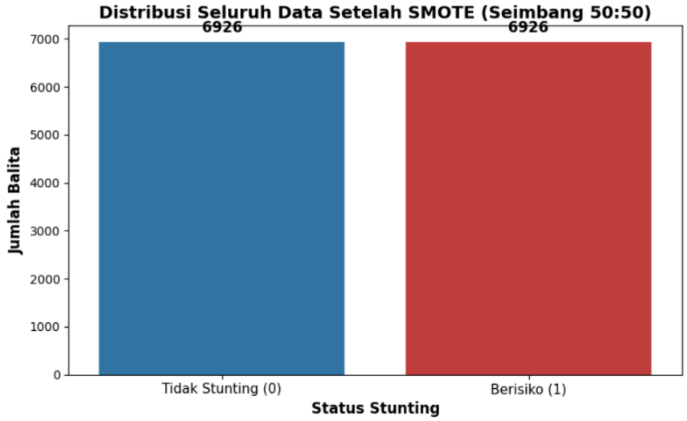
Tahapan akhir adalah ekstraksi *Feature Importance* untuk memeringkat faktor risiko paling dominan, yang kemudian disematkan ke dalam *deployment* aplikasi *web* interaktif menggunakan *framework* Streamlit (Pangestu & Mujiyono, 2025).

4. HASIL DAN PEMBAHASAN

Berdasarkan pengolahan data dan pemodelan yang dilakukan, diperoleh hasil sebagai berikut:

Eksplorasi dan Penyeimbangan Data (SMOTE)

Analisis awal menunjukkan ketimpangan kelas yang sangat ekstrem, di mana balita berstatus "Tidak Stunting" mendominasi dengan 6.926 baris (91,07%), sementara "Berisiko Stunting" hanya berjumlah 679 baris (8,93%). Jika dibiarkan, hal ini akan menyebabkan *accuracy paradox*.



Gambar 3. Distribusi Data setelah Penerapan SMOTE

Setelah penerapan algoritma SMOTE, distribusi data pada kelas latih berhasil diseimbangkan menjadi 50:50, sehingga model XGBoost memiliki pijakan yang adil untuk mempelajari karakteristik kerentanan stunting secara komprehensif.

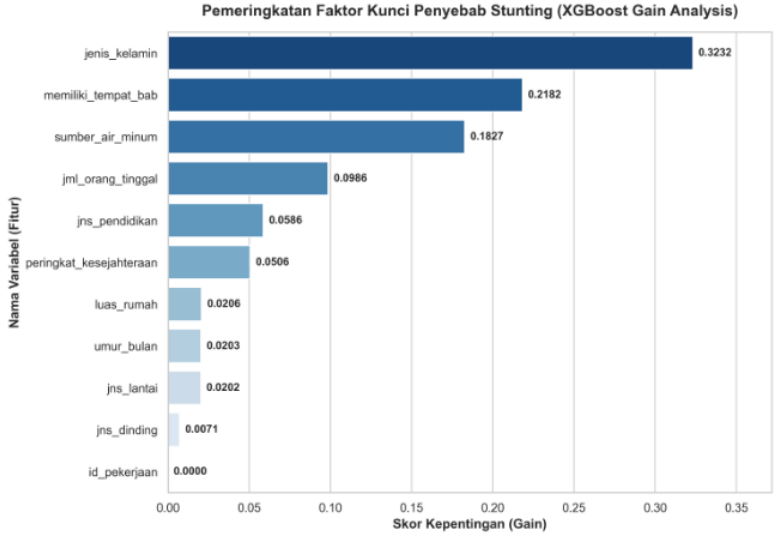
Evaluasi Kinerja Model XGBoost

Pengujian model dilakukan menggunakan 2.771 baris data uji murni yang diisolasi dari proses pelatihan. Hasil klasifikasi menunjukkan bahwa model XGBoost mampu memprediksi status gizi balita dengan tingkat Akurasi sebesar 88,20%.

Lebih dari itu, model mencatatkan nilai Presisi 99,91%, yang berarti peringatan risiko yang dihasilkan hampir tidak memiliki alarm palsu (*False Positive* hanya 1 kasus). Titik ekuilibrium model ditunjukkan oleh F1-Score sebesar 86,73% dan Recall 76,63%. Capaian ini sejaslan dengan penelitian (Hadisuwarno dkk., 2025) yang menegaskan keunggulan XGBoost dalam menangani data terstruktur stunting setelah dilakukan penyeimbangan.

Analisis Faktor Kunci Penyebab Stunting (*Feature Importance*)

Untuk membuka *black-box* algoritma, ekstraksi nilai *Gain* dilakukan untuk mengungkap determinan stunting di wilayah penelitian (Puspanagara, 2025).

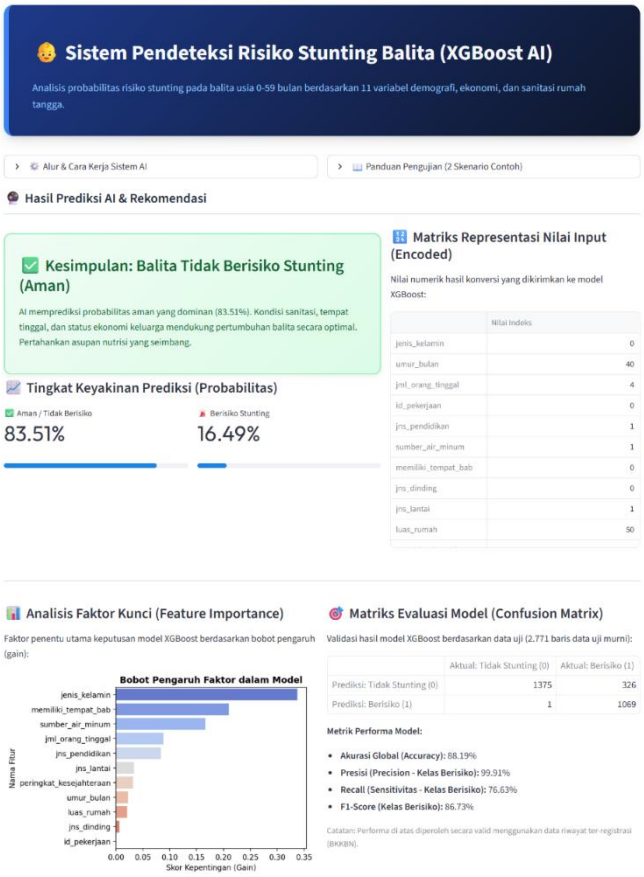


Gambar 4. Faktor-faktor Penyebab Stunting

Berdasarkan hasil visualisasi analisis *Feature Importance*, Jenis Kelamin (skor Gain 0,3232) dan ketersediaan Fasilitas Tempat BAB/Sanitasi (skor Gain 0,2182) adalah prediktor yang paling berkontribusi.Secara sosiomedis, temuan ini mengonfirmasi tinjauan literatur dari (Putri & Lismayanti, 2025) , di mana ketiadaan sanitasi yang layak secara langsung memicu paparan infeksi kronis berulang (seperti diare) yang merusak sistem penyerapan usus balita, sehingga memperparah kondisi malnutrisi terlepas dari jumlah asupan makannya. Variabel Tingkat Pendidikan dan Sumber Air Minum menyusul sebagai faktor struktural.

Implementasi Antarmuka Prediksi Berbasis Web

Mengacu pada pendekatan operasional yang dicontohkan oleh (Pangestu & Mujiyono, 2025), model XGBoost ini berhasil dikemas ke dalam purwarupa (*prototype*) aplikasi *web* interaktif menggunakan Streamlit.



Gambar 5. Antarmuka Website

Sistem dirancang secara transparan, tidak hanya mengeluarkan vonis diagnosis akhir, tetapi juga melampirkan persentase probabilitas (*predict_proba*) dan grafik *Feature Importance* secara langsung di layar pengguna untuk mendukung pengambilan keputusan intervensi yang presisi di tingkat puskesmas.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa algoritma XGBoost yang dikombinasikan dengan teknik SMOTE sangat efektif dalam memprediksi risiko stunting, menghasilkan akurasi 88,20% dan presisi 99,91%. Teknik *Feature Importance* juga berhasil membongkar pola sosiomedis, membuktikan bahwa selain asupan gizi, risiko stunting secara dominan dipicu oleh faktor Jenis Kelamin dan buruknya ketersediaan Fasilitas Tempat BAB (Sanitasi).

Untuk pengembangan selanjutnya, disarankan agar wilayah observasi BKKBN diperluas lintas provinsi. Pengintegrasian variabel lingkungan keluarga ini dengan rekapan medis antropometri e-PPGBM (seperti Berat Badan Lahir dan riwayat PMT), serta implementasi optimasi *hyperparameter* dinamis seperti *GridSearchCV*, sangat

direkomendasikan untuk mereduksi angka *False Negative* dan meningkatkan ketajaman diagnosis sistem secara komprehensif.

DAFTAR REFERENSI

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Hadisuwarno, M. A. H., Martono, K. T., & Adriono, E. (2025). Komparasi performa model machine learning algoritma XGBoost dan Random Forest pada studi kasus mendeteksi stunting. AITI: Jurnal Teknologi Informasi, 22(2), 266–278.

Kase, C. P. W., & Prasetyo, S. Y. J. (2025). Analisis Faktor Risiko Stunting pada Balita di Desa Kesetnana Menggunakan Metode Random Forest. Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 6(3), 1556–1566.

Kementerian Kesehatan RI. (2025). Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka. Badan Kebijakan Pembangunan Kesehatan Kementerian Kesehatan RI.

Nugroho, I. W., Febriansyah, C. R. E., Ramadhani, O. S., Murti, B. S., Sabil, S., & Zuhdi, S. (2026). Analisis Pola Global dan Lokal Determinan Prevalensi Stunting di Jawa Tengah Tahun 2024 Menggunakan Geographically Weighted Regression. Jurnal Kesehatan Masyarakat Indonesia, 21(1), 48–56.

Pangestu, A., & Mujiyono, S. (2025). Implementasi Algoritma XGBoost Untuk Prediksi Status Gizi Balita Berbasis Website. Jurnal Algoritma, 22(2), 176–187. https://doi.org/10.33364/algoritma/v.22-2.2390

Puspanagara, A. L. (2025). Penerapan Explainable AI untuk Prediksi Performa Akademik Mahasiswa Menggunakan Random Forest dan SHAP. Infoman's: Jurnal Ilmu-ilmu Informatika dan Manajemen, 19(1).

Puspita, D. M., Santoso, K. I., Triyono, A., Supriyadi, E., & Nugroho, A. S. (2025). Algoritma Random Forest, Decision Tree dan XGboost Untuk Klasifikasi Stunting Pada Balita. Transformatika, 23(1), 67–76.

Putri, A., & Lismayanti, L. (2025). Faktor-Faktor Risiko Yang Mempengaruhi Stunting Pada Balita: Literatur Riview. Jurnal Ilmiah Penelitian Mahasiswa, 3(6), 1098–1107.

Sugihartono, T., Wijaya, B., Marini, Alkayes, A. F., & Anugrah, H. A. (2025). Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN. Journal of Applied Data Sciences, 6(1), 667–682. https://doi.org/10.47738/jads.v6i1.494