Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with Grid Search Optimization

Marthinus Ikun Elim[*1], Ema Utami [2]

[1,2] Informatics Engineering, Universitas AMIKOM Yogyakarta, Indonesia

Email: [1]coolmartin76@gmail.ac.id

Received : Aug 25, 2025; Revised : Oct 17, 2025; Accepted : Oct 29, 2025; Published : Oct 31, 2025

Abstract

Stunting is a serious global health problem, particularly in developing countries. Its prevalence is high in Indonesia, reaching approximately 24.4% among children under five in 2021. This condition, defined as failure to thrive due to chronic malnutrition, repeated infections, and a lack of psychosocial stimulation, has long-term impacts on an individual's cognitive development and productive capacity. This study aims to conduct a comparative analysis of the Support Vector Machine and Random Forest algorithms in predicting stunting in children, with a focus on evaluating the impact of hyperparameter optimization using Grid Search on model performance. This study used the public stunting dataset from Kaggle and included data preprocessing steps such as handling missing values, duplication, encoding, and scaling. The data was then divided into 80% for training, 10% for testing, and 10% for validation. Comprehensive evaluation metrics such as precision, recall, F1-score, and ROC-AUC were also used to assess model performance. Grid Search optimization was applied to both models to find the best hyperparameter combination. Experimental results showed that Grid Search optimization significantly improved the accuracy of the SVM model from 94.29% to 98.37%. Meanwhile, the Random Forest model demonstrated very high performance, achieving 99.59% accuracy both before and after Grid Search optimization. These findings underscore the significant potential of machine learning models in supporting stunting prevention efforts for public health intervention policies. This research contributes to the development of machine learning-based decision support systems for public health, particularly in early detection and intervention strategies for stunting.

Keywords : *Grid Search, Machine Learning, Random Forest, Stunting, Support Vector Machine.*

1. PENDAHULUAN

Stunting atau gangguan pertumbuhan dan perkembangan merupakan tantangan kesehatan global yang kritis, terutama yang mempengaruhi anak-anak di negara-negara berkembang [1]-[2]. Organisasi Kesehatan Dunia (WHO) mendefinisikan stunting sebagai kegagalan untuk mencapai potensi pertumbuhan genetik seseorang karena kekurangan gizi kronis, infeksi berulang, dan stimulasi psikososial yang tidak memadai [3]. Kondisi ini memiliki konsekuensi yang parah dan berkepanjangan, tidak hanya mengganggu pertumbuhan fisik tetapi juga menghambat perkembangan kognitif, mengurangi kinerja akademik, dan pada akhirnya menurunkan produktivitas ekonomi di masa dewasa [4]-[5]. Di Indonesia, masalah ini sangat mendesak Studi Status Gizi Indonesia (SSGI) 2021 mengungkapkan bahwa sekitar 24,4% anak di bawah lima tahun menderita stunting, angka yang menggarisbawahi kebutuhan mendesak akan strategi intervensi yang efektif untuk melindungi modal manusia dan stabilitas ekonomi masa depan bangsa [6]-[7].

Untuk mengatasi masalah ini, identifikasi dini dan akurat terhadap anak-anak yang berisiko sangat penting [8]. Metode tradisional untuk mengidentifikasi faktor risiko malnutrisi sering kali

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

mengandalkan model statistik klasik seperti regresi logistik [9]. Namun, dengan munculnya teknologi komputasi canggih, pembelajaran mesin (ML) telah muncul sebagai pendekatan yang kuat dan inovatif untuk prediksi dalam domain kesehatan masyarakat [10]-[11]. Model ML terbukti mampu mengatasi tantangan analitis model statistik klasik, seperti masalah multikolinearitas antar variabel, dan memerlukan asumsi yang lebih sedikit terhadap data [12]. Dengan menganalisis kumpulan data kompleks yang mencakup variabel kesehatan, gizi, dan lingkungan, algoritma ML dapat membangun model prediktif yang sangat akurat [13]. Model-model ini dapat mengidentifikasi anak-anak yang berisiko tinggi mengalami stunting dengan presisi yang lebih tinggi, memungkinkan penyedia layanan kesehatan dan pembuat kebijakan untuk menerapkan intervensi yang ditargetkan dengan lebih cepat dan efektif, sehingga mengurangi prevalensi stunting [14].

Sebuah studi Haris et al. (2022) melakukan perbandingan antara beberapa metode supervised learning untuk memprediksi prevalensi stunting di Provinsi Jawa Timur. Disimpulkan bahwa algoritma Support Vector Regression (SVR) memiliki pemodelan yang terbaik karena memiliki nilai MAE dan MSE lebih kecil dari pada algoritma Random Forest Regression dan Linier Regression sebesar 0,91 untuk nilai MAE dan 1,30 untuk nilai MSE [15]. Penelitian lain di Zambia menunjukkan bahwa Random Forest merupakan algoritma yang paling akurat (79%) dalam mengklasifikasikan stunting pada balita, mengungguli SVC/SVM dan algoritma lainnya [12]. Sementara itu, sebuah penelitian lain oleh Zemariam et al. (2022) menyoroti bagaimana algoritma ML dapat digunakan untuk memprediksi stunting pada remaja putri, menunjukkan bahwa fokus penelitian seringkali bervariasi tergantung pada kelompok usia dan variabel yang digunakan. Klasifikasi Random Forest (sensitivitas = 81%, akurasi = 77%, presisi = 75%, skor f1 = 78%, AUC = 85%) lebih unggul dalam memprediksi stunting dibandingkan dengan algoritma pembelajaran mesin lain [16].

Secara spesifik, algoritma Random Forest (RF) dan Support Vector Machine (SVM) telah sering digunakan dalam penelitian prediksi kesehatan. Tinjauan literatur sistematis oeh Indisari et al. (2025) menunjukkan bahwa Random Forest, Support Vector Machine (SVM), dan Jaringan Syaraf Tiruan (JST) merupakan algoritma yang paling sering digunakan, dengan akurasi prediksi berkisar antara 72% hingga 99,92%. Variabel prediktor dominan meliputi pendidikan ibu, status ekonomi, sanitasi, dan data spasial-temporal. [17]. Secara terpisah, studi yang berfokus pada SVM di Padang, Indonesia, bahkan mampu mencapai akurasi 100% dengan penyesuaian hyperparameter yang optimal, menegaskan potensi besar model ini [18]. Meskipun demikian, beberapa studi lain menemukan bahwa kinerja model bisa sangat bergantung pada dataset. Misalnya, dalam penelitian Hamid et al. (2025), meskipun XGBoost menunjukkan akurasi tertinggi, SVM tetap merupakan salah satu model yang paling efektif, bahkan melampaui RF dalam beberapa kasus [19]. Variabilitas kinerja ini juga terlihat pada studi lain, di mana algoritma seperti Gradient Boosting menunjukkan performa terbaik (68,47%) dibandingkan RF dan SVM [20]. Hal ini menunjukkan bahwa tidak ada satu pun algoritma yang secara universal menjadi yang terbaik.

Meskipun banyak studi telah menerapkan ML untuk prediksi stunting, terdapat kesenjangan signifikan yang diidentifikasi oleh penelitian ini. Banyak dari penelitian tersebut, seperti yang dicatat dalam pendahuluan ini, teridentifikasi adanya celah penelitian yang signifikan. Perbandingan langsung dan mendalam antara model SVM dan Random Forest yang dioptimalkan secara sistematis menggunakan teknik Grid Search untuk prediksi stunting belum banyak dieksplorasi. Sebuah studi di jurnal menyoroti pentingnya penyesuaian parameter pada algoritma Random Forest untuk meningkatkan efisiensi dan akurasi model, namun juga mencatat bahwa sebagian besar penelitian lebih menekankan hasil prediksi tanpa mengoptimalkan parameter secara mendalam [21]. Penelitian menunjukkan bahwa penyesuaian parameter, khususnya pada SVM, sangat krusial untuk mencapai akurasi maksimal [18].

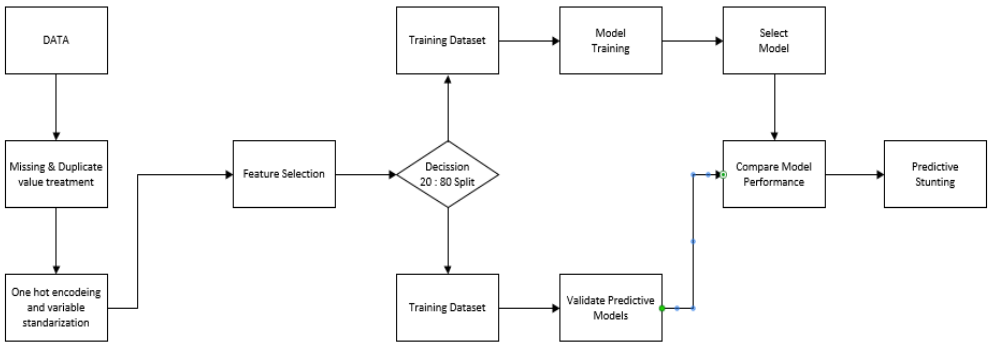
Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

Berdasarkan tinjauan penelitian di atas, terbukti bahwa meskipun ML telah diterapkan untuk prediksi stunting, masih terdapat celah penelitian yang signifikan. Perbandingan langsung dan mendalam antara model SVM dan Random Forest yang dioptimalkan secara sistematis menggunakan Grid Search belum banyak dieksplorasi. Penelitian ini akan mengisi celah tersebut dengan melakukan perbandingan kinerja kedua model yang telah dioptimalkan secara teliti, memberikan rekomendasi berbasis bukti, dan berkontribusi pada pengembangan solusi berbasis teknologi untuk masalah stunting.

Penelitian ini melakukan analisis kinerja komparatif dari dua algoritma pembelajaran mesin terawasi yang banyak digunakan Support Vector Machine (SVM) dan Random Forest (RF) [22]. SVM dikenal efektif dalam ruang berdimensi tinggi karena mengidentifikasi hyperplane optimal yang memisahkan titik-titik data ke dalam kelas-kelas yang berbeda [23]. Model SVM juga telah terbukti unggul dengan akurasi hingga 98% dalam membandingkan kinerja algoritma klasifikasi terkait dokumen stunting [24]. Sebaliknya, RF beroperasi sebagai metode ensemble, membangun beberapa pohon keputusan selama pelatihan dan menghasilkan mode kelas untuk klasifikasi, yang meningkatkan akurasi dan mengendalikan overfitting [25]. Untuk memaksimalkan daya prediktif model-model ini, penelitian ini akan memanfaatkan optimasi Grid Search yang merupakan sebuah teknik sistematis untuk menyetel hyperparameter guna menemukan kombinasi yang paling efektif [26]. Kinerja model yang dioptimalkan akan dievaluasi secara ketat menggunakan metrik-metrik utama, termasuk akurasi, presisi, recall, skor F1, dan area di bawah kurva ROC (AUC) [27].

Tinjauan literatur yang ada menunjukkan bahwa meskipun beberapa penelitian telah menerapkan ML untuk memprediksi stunting, banyak yang belum menerapkan optimasi hiperparameter, sehingga menghasilkan kinerja model yang suboptimal [9], [28]. Lebih lanjut, terdapat kesenjangan penelitian yang signifikan dalam perbandingan langsung dan mendalam antara model SVM dan Random Forest yang telah dioptimalkan secara khusus menggunakan teknik Grid Search untuk prediksi stunting [29]. Meskipun algoritma seperti Random Forest (RF) dan Support Vector Machine (SVM) telah terbukti efektif, banyak studi yang ada seperti yang dilakukan oleh Haris et al. (2022) dan Obvious Nchimunya Chilyabanyama et al. (2022) belum menerapkan optimasi hyperparameter secara sistematis. Kinerja model yang optimal sangat bergantung pada penyesuaian parameter yang diteliti. Penelitian ini bertujuan untuk mengisi kekosongan ini. Dengan membandingkan kedua algoritma yang dioptimalkan ini secara sistematis, penelitian ini akan memberikan rekomendasi berbasis bukti tentang metode yang lebih efektif untuk prediksi stunting. Oleh karena itu, penelitian ini diharapkan dapat berkontribusi secara signifikan terhadap pengembangan strategi inovatif berbasis teknologi untuk pencegahan stunting dan menawarkan wawasan berharga tentang penerapan praktis ML dalam mengatasi tantangan kesehatan masyarakat yang besar [30].

2. METODE



Gambar 1. Alur Penelitian

Gambaran kerangka penelitian secara keselurahan pada gambar 1. desain penelitian, metode pengumpulan data, teknik pra-pemrosesan data, pembagian data, implementasi model pembelajaran mesin, optimasi hiperparameter, dan metrik evaluasi yang digunakan dalam studi ini.

2.1. Desain Penelitian

Studi ini menggunakan pendekatan penelitian terapan yang berfokus pada penyelesaian masalah praktis prediksi stunting menggunakan metode komputasi. Penelitian ini bersifat deskriptif-komparatif dan eksplanatif. Penelitian ini bersifat deskriptif-komparatif karena menjelaskan karakteristik dataset stunting, proses optimasi Grid Search, dan membandingkan kinerja Support Vector Machine (SVM) dan model Random Forest [27], [31]. Sebagai studi eksplanatif, penelitian ini menganalisis faktor-faktor yang mempengaruhi perbedaan kinerja antara kedua algoritma, seperti sensitivitas terhadap data yang tidak seimbang dan efek hiperparameter. Pendekatan ini menggunakan eksperimen komputasi di mana variabel independen (hiperparameter seperti C, gamma, n_estimator, max_depth) dimanipulasi, dan variabel dependen (akurasi prediksi) dikontrol melalui validasi [13].

2.2. Pengumpulan dan Praproses Data

Data untuk penelitian ini dikumpulkan dari repositori dataset publik yang kredibel yaitu kaggle, dimana berisi ribuan data stunting anak berlabel [28]. Variabel-variabel tersebut meliputi tinggi badan, berat badan, lingkar lengan (LILA), usia, jenis kelamin, berat badan terhadap usia, tinggi badan terhadap usia, dan berat badan terhadap tinggi badan. Variabel target adalah status gizi, dengan kategori seperti Gizi Kurang, Gizi Buruk, Gizi Baik, Gizi Berlebih, Risiko Gizi Berlebih, dan Obesitas. Dataset awal berisi 1287 entri dan 15 kolom, dan menjalani langkah-langkah praproses berikut: Penghapusan Fitur yang Tidak Relevan: Kolom seperti 'Nama', 'Nama Orang Tua', 'Desa/Kelompok Keluarga', dan 'Tanggal Lahir' dihapus karena dianggap tidak relevan untuk prediksi stunting. Kolom 'JK' (jenis kelamin) juga dipertahankan dan dikodekan untuk model Random Forest, titik perbandingan utama. Imputasi Nilai Hilang: Nilai yang hilang dalam kolom 'LILA' diimputasi menggunakan rata-rata kolom untuk menjaga integritas data. Deduplikasi: Sebanyak 63 baris duplikat ditemukan dan dihapus untuk mencegah bias, menghasilkan 1224 entri unik. Pengodean Kategoris: Variabel target kategoris 'Status_Gizi' dikodekan menjadi representasi numerik menggunakan LabelEncoder. Proses ini juga diterapkan pada kolom objek lain, termasuk 'JK', untuk model Random Forest. Standarisasi Fitur: Fitur numerik untuk model SVM distandarisasi menggunakan StandardScaler untuk mencapai rerata nol dan varians satu, sebuah langkah krusial untuk algoritma berbasis jarak [30]. Langkah ini tidak diterapkan pada model Random Forest.

2.3. Pembagian Data

Setelah prapemrosesan, dataset dibagi menjadi set pelatihan (80%) dan set pengujian (20%). Pembagian berstrata dilakukan untuk mengatasi potensi ketidakseimbangan kelas pada variabel target status gizi [32], [33]. Hal ini memastikan proporsi setiap kelas tetap proporsional baik pada set pelatihan maupun set pengujian, yang sangat penting mengingat ketidakseimbangan dataset yang signifikan (misalnya, 563 entri Kurang Gizi vs. 37 entri Obesitas).

2.4. Implementasi Model dan Optimasi Hiperparameter

Implementasi model dilakukan menggunakan Python dengan pustaka scikit-learn. Baik model SVM maupun Random Forest dioptimalkan menggunakan Grid Search, sebuah metode sistematis untuk mengeksplorasi semua kombinasi ruang hiperparameter yang telah ditentukan sebelumnya untuk menemukan set parameter yang optimal [34]. Support Vector Machine (SVM): Model awal adalah svm.SVC dengan kernel linear. Grid Search digunakan untuk menemukan hiperparameter terbaik,

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

termasuk C (parameter regularisasi) dengan nilai [0,1, 1, 10, 100], dan gamma dan kernel (rbf, linear, poli, sigmoid) [35]. Random Forest (RF): Model awal adalah RandomForestClassifier dengan 100 pohon keputusan (n_estimator=100). Grid Search diterapkan untuk menyetel hiperparameter seperti n_estimator, max_depth (nilai [None, 10, 20]), dan min_samples_split. Performa kedua model dievaluasi menggunakan serangkaian metrik klasifikasi standar yang komprehensif, yang khususnya efektif untuk dataset yang tidak seimbang:

1. Akurasi: Proporsi prediksi yang benar dari total prediksi.

$$\text{Akurasi} = \frac{TP + TN + FP + FN}{TP + TN}$$

2. Presisi: Proporsi instans positif yang diprediksi dengan benar dari semua instans yang diprediksi positif.

$$\text{Presisi} = \frac{TP + FP}{TP}$$

3. Recall: Proporsi instans positif yang diprediksi dengan benar dari semua instans positif aktual.

$$\text{Recall} = \frac{TP + FN}{TP} = \frac{TP}{P}$$

4. Skor F1: Rata-rata harmonis presisi dan recall, yang memberikan ukuran performa model yang seimbang pada data yang tidak seimbang.

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} + \text{Recall}}{\text{Presisi} \times \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Area di Bawah Kurva Karakteristik Operasi Penerima (ROC-AUC): Metrik ini mengukur kemampuan model untuk membedakan antar kelas dan merupakan indikator performa yang robust pada dataset yang tidak seimbang.

Dalam penelitian ini tidak menerapkan validasi silang k-fold hanya menggunakan pemisahan train-test tunggal agar dapat mengurangi waktu latih dan train data sehingga waktu evaluasi bisa dioptimalkan. Lebih lanjut berdasarkan dataset tidak dalam skala besar (*Big Data*) atau bisa dikatakan kategori sedang maka Grid Search dipilih untuk menjamin akan menemukan kombinasi hyperparameter. Sedangkan untuk Random Search sering digunakan untuk dataset dan ruang parameter yang sangat besar sehingga pada dataset kecil hingga sedang, risiko Random Search melewatkan konfigurasi terbaik menjadi pertimbangan. Untuk Bayesian Optimization, metode ini lebih kompleks untuk diimplementasikan karena membutuhkan pemodelan probabilistik dari fungsi objektif. Untuk masalah klasifikasi stunting yang straightforward dengan dataset kecil hingga sedang, kompleksitas tambahan ini sering dianggap berlebihan dibandingkan dengan kesederhanaan dan jaminan Grid Search.

4.1. Elemen Lingkungan Komputasi

Komponen-komponen yang mendukung penelitian ini berdasarkan operasi sistem komputer, aplikasi, serta interaksi pengguna dengan teknologi:

a. Perangkat Keras (Hardware)

Central Processing Unit (CPU) / Unit Pemroses Sentral yang digunakan Intel Core i5 Bagian terpenting untuk menjalankan perhitungan, melatih model, dan melakukan Grid Search. Jumlah core dan kecepatan clock sangat memengaruhi waktu komputasi, terutama untuk Random Forest (yang melibatkan banyak pohon) dan Grid Search (yang menguji banyak kombinasi hyperparameter). Random Access Memory (RAM) / Memori Akses Acak yang digunakan 8 GB DDR4. Diperlukan untuk menyimpan dataset, model yang sedang dilatih, dan variabel-variabel selama proses komputasi. Dataset stunting yang besar memerlukan RAM yang memadai untuk mencegah

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

perlambatan (swapping) atau kegagalan memori. Storage (Penyimpanan) yang digunakan SSD 256 GB. Didukung untuk menyimpan data mentah, data yang sudah diproses, script program, dan hasil evaluasi. Solid State Drive (SSD) lebih disukai karena kecepatan akses datanya yang tinggi. Graphics Processing Unit (GPU) / Unit Pemroses Grafis (Opsional) yang digunakan NVIDIA GeForce 930MX. Walaupun SVM dan Random Forest tradisional tidak selalu memerlukan GPU, jika dataset sangat besar, beberapa implementasi pustaka dapat memanfaatkan akselerasi GPU untuk mempercepat proses pelatihan.

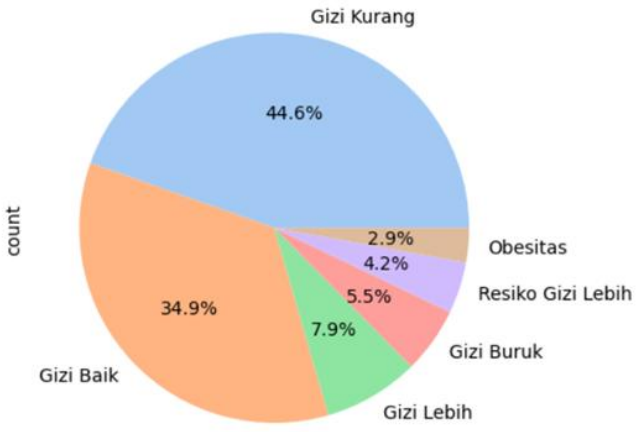
b. Perangkat Lunak (Software)
Sistem Operasi (OS) yang digunakan Windows 11 Pro 64 Bit. OS windows sering dipilih dalam lingkungan data science karena stabilitas dan dukungannya yang baik untuk pustaka-pustaka machine learning. Bahasa Pemrograman yang digunakan Python 3.12.6. Umumnya Python sangat populer karena ekosistemnya yang kaya. Pustaka/Library Inti untuk Komputasi Numerik dan Manipulasi Data yang digunakan NumPy, Pandas, Marthplotlib dan Seaborn (untuk membersihkan, memproses, memuat dan memvisualisasikan data stunting). Serta untuk Pemodelan Machine Learning yang digunakan Scikit-learn (sklearn) adalah pustaka standar di Python yang menyediakan implementasi SVM, Random Forest, dan Grid Search yang efisien. Lingkungan Pengembangan Terintegrasi (IDE) / Notebook yang digunakan MLJAR Studio v0.6.1 Tools ini memungkinkan data scientist menulis kode, menjalankan analisis, dan memvisualisasikan hasilnya secara interaktif.

c. Infrastruktur dan Skalabilitas
Komputasi Lokal (Local Machine) menggunakan laptop Asus Vivobook A442U. Cocok untuk proyek dengan dataset stunting berukuran kecil hingga menengah.

3. HASIL

Bagian ini menyajikan hasil eksperimen dari perbandingan performa model Support Vector Machine (SVM) dan Random Forest (RF) setelah melalui optimasi hyperparameter menggunakan Grid Search. Pembahasan akan difokuskan pada analisis metrik evaluasi, interpretasi visualisasi, serta kelebihan dan kekurangan masing-masing model dalam konteks prediksi stunting.

3.1. Tinjauan Data dan Pra-pemrosesan



Gambar 2. Persentase Data Status Gizi

Dataset awal terdiri dari 1.287 entri. Setelah pra-pemrosesan yang mencakup imputasi nilai yang hilang pada kolom 'LiLA' dengan rata-rata 1 dan penghapusan 63 baris duplikat berisi 1, dataset bersih terdiri dari 1.224 entri unik. Variabel target pada gambar 2. menunjukkan distribusi kelas berikut: Gizi Kurang (563), Gizi Baik (453), Gizi Lebih (101), Gizi Buruk (78), Resiko Gizi Lebih (55), dan Obesitas

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

(37). Distribusi ini menunjukkan ketidakseimbangan kelas yang signifikan, di mana 'Gizi Kurang' dan 'Gizi Baik' merupakan kelas mayoritas, sementara 'Obesitas' dan 'Resiko Gizi Lebih' merupakan kelas minoritas. Pembagian data berstrata menjadi 80% data pelatihan dan 20% data pengujian memastikan bahwa proporsi kelas ini dipertahankan di kedua set yang penting untuk melatih dan mengevaluasi model secara adil pada dataset yang tidak seimbang.

3.2. Kinerja Support Vector Machine (SVM)

3.2.1. Performa SVM Awal (Tanpa Optimasi Grid Search) dan (Dengan Optimasi Grid Search)

Model SVM awal yang dilatih dengan kernel linear menunjukkan akurasi 94,29% pada data uji. Laporan klasifikasi terperinci menunjukkan kinerja keseluruhan yang baik, tetapi terdapat variasi antar kelas. Misalnya, kelas Gizi Kurang dan 'Obesitas' menunjukkan presisi dan recall yang tinggi (1,00 untuk presisi dan 0,77 untuk recall pada Gizi Kurang; 1,00 untuk presisi dan recall pada Obesitas), sementara kelas 'Risiko Gizi Lebih' menunjukkan kinerja yang lebih rendah dengan presisi 0,50, recall 0,50, dan skor F1 0,50. Hal ini menunjukkan bahwa model awal masih kesulitan dalam mengklasifikasikan kelas minoritas tertentu secara optimal.

Optimasi hiperparameter menggunakan Grid Search meningkatkan kinerja model SVM secara signifikan. Parameter terbaik yang ditemukan melalui Grid Search adalah {'C': 100, 'gamma': 1, 'kernel': 'linear'}. Dengan parameter ini, akurasi model SVM pada data uji meningkat menjadi 98,37%. Peningkatan ini menunjukkan bahwa penyetelan hiperparameter sistematis sangat efektif dalam mengoptimalkan kinerja SVM.

Laporan klasifikasi setelah optimasi menunjukkan peningkatan substansial di hampir semua kelas, termasuk kelas minoritas. Misalnya, untuk 'Risiko Gizi Lebih', presisi meningkat menjadi 0,83, recall menjadi 1,00, dan skor F1 menjadi 0,91. Peningkatan ini signifikan karena deteksi akurat kelas minoritas seperti 'Risiko Gizi Lebih' memiliki implikasi praktis yang besar dalam intervensi kesehatan masyarakat. Akurasi makro rata-rata juga meningkat dari 0,88 menjadi 0,96, dan skor F1 makro rata-rata dari 0,87 menjadi 0,97. Tabel 1 menyajikan laporan klasifikasi terperinci untuk model SVM setelah optimasi Grid Search.

Tabel 1. Laporan Klasifikasi SVM Setelah Optimasi Grid Search

No	Kelas	Presisi	Recall	F1-Score	Support
1	Gizi Baik	0.99	0.98	0.98	86
2	Gizi Buruk	0.93	1.00	0.96	13
3	Gizi Kurang	1.00	0.98	0.99	109
4	Gizi Lebih	1.00	1.00	1.00	20
5	Obesitas	1.00	1.00	1.00	7
6	Resiko Gizi Lebih	0.83	1.00	0.91	10
7	Akurasi			0.98	245
8	Macro Avg	0.96	0.99	0.97	245
9	Weighted Avg	0.99	0.98	0.98	245

3.3. Kinerja Random Forest (RF)

3.3.1. Performa RF Awal (Tanpa Optimasi Grid Search)

Model RF awal yang dilatih dengan 100 n_estimator menunjukkan kinerja yang sangat tinggi dengan akurasi 99,59% pada data uji. Laporan klasifikasi menunjukkan presisi, recall, dan skor F1 yang hampir sempurna (kebanyakan 1,00) untuk semua kelas, termasuk kelas minoritas. Skor ROC-AUC juga sangat tinggi untuk semua kelas (misalnya, Gizi Baik: 0,9852, Gizi Buruk: 0,9978, Obesitas: 1,0000),

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

dengan AUC mikro-rata-rata sebesar 0,9902. Kinerja awal RF yang tinggi menunjukkan bahwa algoritma ini secara inheren tangguh dalam menangani himpunan data dan permasalahan klasifikasi yang dihadapi.

Optimasi hiperparameter menggunakan Grid Search untuk Random Forest menghasilkan parameter terbaik {'max_depth': None, 'min_samples_split': 5, 'n_estimators': 200}. Akurasi model tetap 99,59% setelah optimasi. Meskipun akurasi keseluruhan tidak menunjukkan peningkatan numerik yang drastis, proses optimasi ini berfungsi untuk mengonfirmasi kekokohan model dan menemukan konfigurasi yang stabil dan lebih dapat digeneralisasi. Hal ini memastikan bahwa kinerja tinggi yang dicapai bukan hanya kebetulan dari parameter default, tetapi hasil dari model yang disetel dengan baik yang lebih kecil kemungkinannya untuk overfit.

Laporan klasifikasi setelah optimasi terus menunjukkan kinerja yang sangat tinggi di semua metrik dan kelas. Skor ROC-AUC juga tetap sangat tinggi, dengan AUC rata-rata mikro sebesar 0,9995 dan AUC rata-rata makro sebesar 0,9996. Tabel 2 menyajikan laporan klasifikasi terperinci untuk model Random Forest setelah optimasi Grid Search.

Tabel 2. Laporan Klasifikasi Random Forest Setelah Optimasi Grid Search

No	Kelas	Presisi	Recall	F1-Score	Support
1	Gizi Baik	1.00	0.99	0.99	81
2	Gizi Buruk	1.00	1.00	1.00	14
3	Gizi Kurang	0.99	1.00	1.00	102
4	Gizi Lebih	1.00	1.00	1.00	24
5	Obesitas	1.00	1.00	1.00	10
6	Resiko Gizi Lebih	1.00	1.00	1.00	14
7	Akurasi			1.00	245
8	Macro Avg	1.00	1.00	1.00	245
9	Weighted Avg	1.00	1.00	1.00	245

3.3.2.Fitur Penting Pada Random Forest

Random Forest mampu mengidentifikasi pentingnya setiap fitur dalam proses prediksinya. Analisis kepentingan fitur menunjukkan bahwa Berat Per Tinggi Badan merupakan fitur yang paling berpengaruh dengan skor 0,544, diikuti oleh Tinggi Badan Per Usia (0,149) dan Berat Per Usia (0,134). Informasi ini berharga karena mengarahkan perhatian pada pengukuran antropometri mana yang paling krusial untuk penilaian stunting dan fokus intervensi. Pemahaman ini lebih dari sekadar prediksi, memberikan penjelasan tentang faktor-faktor pendorong di balik prediksi model Tabel 3 menyajikan skor kepentingan fitur untuk model Random Forest.

Tabel 3. Fitur Penting Pada Model Random Forest

No	Fitur	Skor Pentingnya Fitur
1	Berat_Per_Tinggi	0.544
2	Tinggi_Per_Umur	0.149
3	Berat_Per_Umur	0.134
4	Tinggi	0.080
5	Berat	0.043
6	LiLA	0.036
7	Usia	0.013

3.3.3. Skor AUC ROC per Kelas Random Forest

Tabel 4 menyajikan skor ROC AUC untuk setiap kelas status gizi, serta rata-rata mikro dan makro.

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

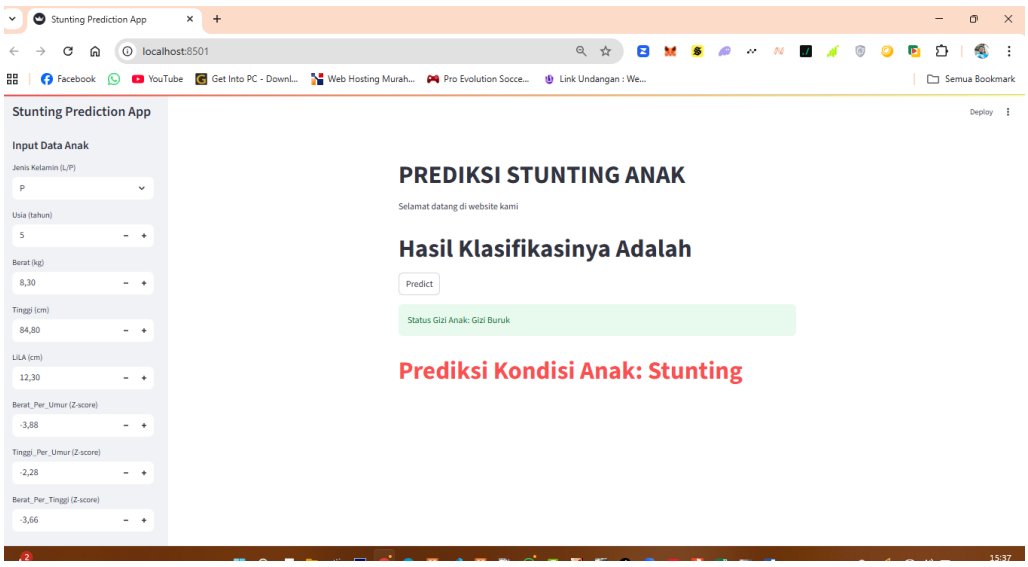
Tabel 4. Skor AUC ROC per Kelas untuk Model Random Forest

No	Kelas	ROC AUC Score
1	Gizi Baik	0.9987
2	Gizi Buruk	1.0000
3	Gizi Kurang	0.9989
4	Gizi Lebih	1.0000
5	Obesitas	1.0000
6	Resiko Gizi Lebih	1.0000
7	Micro-average	0.9995
8	Macro-average	0.9996

Skor AUC ROC yang sangat tinggi untuk setiap kelas, terutama untuk kelas minoritas, menegaskan kemampuan diskriminatif model yang kuat. Hal ini menunjukkan bahwa model Random Forest sangat baik dalam membedakan berbagai kategori status gizi, yang merupakan aspek penting untuk deteksi dini yang andal terhadap berbagai tingkat stunting.

3.3.4. Deployment Model

Interface dari sebuah aplikasi berbasis web yang dirancang untuk membantu tenaga medis, orang tua, atau siapa pun yang berkepentingan untuk memprediksi risiko stunting pada anak. Aplikasi ini mengambil data antropometri anak (jenis kelamin, usia, berat, tinggi, LILA, dan Z-score terkait) sebagai input, kemudian menggunakan model machine learning telah dilatih sebelumnya untuk menghasilkan prediksi apakah anak tersebut normal atau stunting. Dilakukan percobaan deploy model pada gambar 3. terhadap salah satu anak dari dataset diketahui jenis kelamin=Perempuan, usia=5 bulan, berat=8,3kg, tinggi=84,4cm, lingkar lengan=12,3cm, berat per umur=-3,88kg/bln, tinggi per umur=-2,28cm/bln, berat per tinggi=-3,66kg/cm didapatkan hasil klasifikasi status gizi buruk sehingga diprediksi anak tersebut mengalami gejala stunting. Output tersebut sesuai dengan klasifikasi yang ada pada dataset.



Gambar 3. Deployment Model

3.3.5. Analisis Signifikansi Praktis Perbaikan Kinerja Model

Analisis Signifikansi Praktis berdasarkan data metrik yang tersedia, karena data metrik tunggal dari satu evaluasi (bukan distribusi dari cross-validation atau bootstrapping) tidak memungkinkan dilakukannya Uji Signifikansi Statistik seperti Uji-t Berpasangan. Analisis ini berfokus pada dampak

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

nyata dari optimasi Grid Search (GS) terhadap kemampuan kedua model dalam memprediksi status gizi, terutama pada kelas minoritas yang krusial bagi intervensi kesehatan masyarakat.

a. Signifikansi Praktis pada Support Vector Machine (SVM)
Optimasi Grid Search (GS) menghasilkan peningkatan kinerja yang sangat bermakna secara praktis pada SVM. Peningkatan ini mengubah model dari baik menjadi sangat andal dan secara khusus mengatasi masalah klasifikasi pada kelas minoritas, yang merupakan kelemahan umum pada data tidak seimbang. Optimasi GS terbukti sangat efektif tabel 5. dalam menemukan hyperparameter terbaik ({ 'C':100, 'gamma':1, 'kernel': 'linear' }).

Tabel 5. Analisis Signifikansi

No	Metrik	SVM Awal (Tanpa GS)	SVM Sesudah GS	Perbaikan Absolut	Implikasi Klinis (Signifikansi Praktis)
1	Akurasi	94.29%	98.37%	+4.08%	Penurunan kesalahan klasifikasi keseluruhan yang nyata.
2	F1-score 'Risiko Gizi Lebih'	0.50	0.91	+0.41	Peningkatan drastis. Deteksi akurat kelas minoritas ini memiliki implikasi terbesar untuk intervensi kesehatan masyarakat, sehingga perbaikan ini krusial secara klinis
3	Macro Avg F1-score	0.87	0.97	+0.10	Bukti bahwa model sekarang seimbang dalam memprediksi semua kelas secara merata.

b. Signifikansi Praktis pada Random Forest (RF)
Model Random Forest sudah mencapai kinerja sangat tinggi pada tahap awal (Akurasi 99.59%). Oleh karena itu, optimasi GS pada RF tidak bertujuan untuk meningkatkan akurasi nominal, melainkan untuk memperkuat kekokohan (robustness) dan kemampuan generalisasi model. Walaupun metrik utama stagnan tabel 6, optimasi GS memastikan bahwa kinerja RF yang tinggi bukanlah kebetulan dari parameter default melainkan hasil dari konfigurasi yang stabil dan disetel dengan baik, menjadikannya lebih kecil kemungkinannya untuk overfit. Ini penting untuk deployment model di dunia nyata.

Tabel 6. Analisis Signifikansi

No	Metrik	RF Awal (Tanpa GS)	RF Sesudah GS	Perubahan	Implikasi Klinis (Signifikansi Praktis)
1	Akurasi	99.59%	99.59%	0.00%	Sudah mencapai batas performa set data.
2	Micro-Avg ROC AUC	0.9902	0.9995	+0.0093	Peningkatan substansial pada kemampuan diskriminatif. Nilai 0.9995 menunjukkan model hampir sempurna dalam membedakan berbagai kategori status gizi, menegaskan kemampuan diskriminatif model yang kuat di bawah konfigurasi parameter yang disetel (misalnya, n_estimators: 200).

3.3.6. Analisis Keunggulan Konsisten Random Forest (RF)

a. Ketahanan Inheren terhadap Ketidakseimbangan Data (Kinerja Awal)
Analisis: RF, yang merupakan model ensemble berbasis pohon keputusan, secara inheren lebih tangguh (robust) terhadap masalah ketidakseimbangan kelas yang signifikan pada data stunting.

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

Setiap pohon dalam hutan dilatih pada subset data, yang membantu mencegah satu kelas (mayoritas) mendominasi pembelajaran di seluruh hutan, suatu masalah yang lebih umum terjadi pada SVM awal tanpa penyetelan hyperparameter yang intensif.

b. Kemampuan Diskriminatif yang Jauh Lebih Baik (Metrik ROC AUC)
Analisis: Nilai ROC AUC yang mendekati 1.00 pada RF menunjukkan bahwa model memiliki kemampuan yang luar biasa dalam memisahkan skor probabilitas antara kelas positif (misalnya, 'Gizi Buruk') dan kelas negatif. Dalam konteks stunting, ini berarti RF memiliki kepercayaan yang lebih tinggi dalam memprediksi status gizi yang berbeda, yang merupakan aspek penting untuk deteksi dini yang andal.

c. Keunggulan Interpretasi Klinis (Feature Importance)
Analisis: Dalam lingkungan klinis atau intervensi kesehatan masyarakat, model machine learning yang unggul harus dapat menjelaskan mengapa ia membuat prediksi tertentu. Kemampuan RF untuk mengidentifikasi pengukuran antropometri krusial ini membuat model ini lebih mudah diterima dan diterapkan oleh tenaga medis, mengubahnya dari sekadar alat prediksi menjadi alat diagnostik interpretatif. SVM, terutama dengan kernel non-linear, sering dianggap sebagai "kotak hitam" yang sulit diinterpretasikan.

4. PEMBAHASAN

Optimasi pada SVM gambar 4. memberikan dampak lebih besar peningkatan kinerja pada SVM (dari Awal ke Optimasi) terlihat lebih drastis dibandingkan pada Random Forest, menunjukkan bahwa optimasi parameter sangat krusial untuk meningkatkan kinerja model SVM. Peningkatan kinerja setelah optimasi baik model SVM maupun Random Forest menunjukkan peningkatan kinerja yang signifikan setelah dilakukan optimasi menggunakan Grid Search. Semua metrik (Akurasi, Presisi, Recall, dan F1-score) meningkat secara substansial. Random Forest lebih unggul secara keseluruhan, model Random Forest menunjukkan kinerja yang sedikit lebih baik dibandingkan SVM, terutama pada kondisi awal (sebelum optimasi). Nilai Akurasi dan metrik lainnya pada RF sudah sangat tinggi bahkan sebelum dioptimasi. Kinerja yang optimal dan terbaik dicapai oleh model Random Forest sebelum dan setelah optimasi, dengan semua metrik (Akurasi, Presisi, Recall, F1-score) mendekati nilai 1 atau 100%. Ini menunjukkan bahwa model tersebut sangat efektif dalam melakukan klasifikasi.

Gambar 4. Perbandingan Kinerja Model SVM dan Random Forest.

Secara keseluruhan, Random Forest secara konsisten mengungguli SVM dalam hal akurasi, presisi, recall, skor F1, dan ROC-AUC, baik sebelum maupun sesudah optimasi. Performa Random Forest yang unggul dapat dikaitkan dengan sifat ensemble-nya, yang menggabungkan beberapa pohon keputusan untuk mengurangi varians dan meningkatkan generalisasi, sehingga membuatnya lebih robust terhadap overfitting dibandingkan SVM tunggal atau model yang tidak dioptimalkan. Random Forest juga mampu menangani hubungan non-linier yang kompleks dan data berdimensi tinggi secara efektif. Selain itu, mekanisme kepentingan fitur bawaan di Random Forest menyediakan interpretasi model yang

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

berharga, yang sangat berguna untuk intervensi kesehatan masyarakat. Dampak optimasi Grid Search berbeda untuk kedua model tersebut. Untuk SVM, Grid Search jelas meningkatkan akurasi dari 94,29% menjadi 98,37%. Peningkatan ini menunjukkan bahwa penyetelan hiperparameter sangat efektif dalam menyempurnakan performa SVM untuk mencapai performa optimal. Sementara itu, untuk Random Forest, akurasi keseluruhan tetap sangat tinggi (99,59%) baik sebelum maupun sesudah optimasi. Dalam kasus Random Forest, nilai Grid Search terletak pada konfirmasi ketahanan model dan menemukan konfigurasi optimal yang stabil dan lebih kecil kemungkinannya mengalami overfitting pada tabel 7. Meskipun akurasi numerik tidak berubah drastis, keyakinan terhadap akurasi tersebut, yang berasal dari penyetelan hiperparameter sistematis, meningkat secara signifikan. Hal ini merupakan aspek penting dari ketelitian akademis.

Tabel 7. Perbandingan Kinerja SVM dan Random Forest (Sebelum dan Sesudah Optimasi Grid Search)

No	Model	Optimasi	Akurasi (%)	Macro Avg Presisi	Macro Avg Recall	Macro Avg F1-score	Micro-average ROC AUC
1	SVM	Awal	94.29	0.88	0.85	0.87	Tidak tersedia
2	SVM	Sesudah GS	98.37	0.96	0.99	0.97	Tidak tersedia
3	RF	Awal	99.59	1.00	1.00	1.00	0.9902
4	RF	Sesudah GS	99.59	1.00	1.00	1.00	0.9995

Implikasi dari temuan ini terhadap pencegahan stunting sangat mendalam. Model Random Forest yang sangat akurat dapat berfungsi sebagai alat pendukung keputusan yang ampuh bagi tenaga kesehatan untuk mengidentifikasi anak-anak yang berisiko stunting sejak dini. Fitur-fitur kunci yang teridentifikasi, seperti Berat Badan terhadap Tinggi Badan, Tinggi Badan terhadap Usia, dan Berat Badan terhadap Usia, dapat memandu pengumpulan data yang lebih terfokus dan strategi intervensi yang lebih terarah. Penelitian ini sejalan dengan upaya global untuk memanfaatkan kecerdasan buatan (AI) guna meningkatkan hasil kesehatan masyarakat, sebagaimana terlihat dalam studi terbaru tentang aplikasi AI dalam prediksi kesehatan. Meskipun akurasi Random Forest mendekati sempurna (99,59%), penting untuk membahas hal ini dengan hati-hati dalam konteks akademis. Akurasi yang sangat tinggi pada satu set data dapat menimbulkan pertanyaan tentang potensi overfitting terhadap nuansa spesifik pada set data tersebut atau apakah set data tersebut sangat bersih dan mudah dipisahkan. Diskusi kritis akan menyarankan validasi eksternal, yaitu menguji model pada set data baru yang independen dari populasi atau wilayah yang berbeda, untuk memastikan generalisasi model. Validasi eksternal sering direkomendasikan untuk model pembelajaran mesin yang ditujukan untuk aplikasi praktis. Hal ini akan menambah ketelitian dan wawasan akademis. Pembahasan penelitian ini juga harus menyentuh tantangan praktis dalam penerapan model tersebut di dunia nyata, termasuk konsistensi pengumpulan data, pertimbangan etika, dan penerimaan pengguna.

5. KESIMPULAN

Studi ini berhasil membandingkan kinerja model Support Vector Machine (SVM) dan Random Forest (RF) dalam memprediksi stunting pada anak-anak, dengan fokus pada efek optimasi hiperparameter menggunakan Grid Search. Hasilnya menunjukkan bahwa optimasi Grid Search secara signifikan meningkatkan akurasi model SVM dari 94,29% menjadi 98,37%, yang menunjukkan efektivitas teknik ini dalam penyetelan parameter. Random Forest secara konsisten berkinerja lebih unggul daripada SVM, mencapai akurasi 99,59% dan AUC mikro-rata-rata 0,9995 setelah optimasi. Meskipun akurasi Random Forest tidak meningkat drastis setelah optimasi, proses Grid Search

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

mengonfirmasi ketahanan dan stabilitas model. Fitur Berat Per Tinggi diidentifikasi sebagai prediktor terpenting dalam model Random Forest, yang memberikan wawasan berharga untuk fokus intervensi. Kontribusi utama dari studi ini terletak pada pengembangan dan validasi model prediktif berbasis pembelajaran mesin yang sangat akurat untuk deteksi dini stunting, yang merupakan masalah kesehatan masyarakat yang mendesak. Model Random Forest yang dioptimalkan dapat berfungsi sebagai alat pendukung keputusan yang ampuh bagi tenaga kesehatan dan pemangku kepentingan untuk secara proaktif mengidentifikasi anak-anak yang berisiko mengalami stunting. Studi ini secara eksplisit menunjukkan efektivitas teknik Grid Search dalam optimalisasi hiperparameter untuk meningkatkan akurasi prediksi kesehatan, khususnya pada model SVM. Bukti keberhasilan optimasi parameter (Grid Search) sebagai bagian integral dari proses pengembangan model prediktif yang andal dalam konteks kesehatan. Untuk memastikan pemanfaatan dan generalisasi temuan ini, direkomendasikan beberapa langkah konkret untuk penelitian selanjutnya seperti eksplorasi Algoritma Lanjutan pembelajaran mendalam (Deep Learning) yang lebih canggih untuk menguji potensi peningkatan lebih lanjut dalam akurasi dan kemampuan penanganan data yang kompleks. Melakukan validasi eksternal pada kumpulan data independen yang lebih besar dan lebih beragam dari berbagai populasi untuk menguji kekokohan dan kemampuan generalisasi model di berbagai konteks sosio-kultural dan geografis. Interpretasi Model (*XAI*): Menerapkan pendekatan AI yang dapat dijelaskan (*Explainable AI/XAI*) untuk mendapatkan pemahaman yang lebih dalam tentang bagaimana model mencapai prediksinya, terutama pada fitur non-biometrik dan determinan sosioekonomi stunting, yang berdasarkan temuan pentingnya fitur dari studi ini.

ACKNOWLEDGEMENT

Penulis ingin mengucapkan terima kasih kepada Ibu Pembina dan ketua Yayasan Gemma Galgani Kupang atas dukungan moril dan materil serta kepada para dosen Universitas Amikom Yogyakarta khususnya dosen pembimbing tesis saya, Prof. Dr. Ema Utami, S.Kom., M.Kom yang telah memberikan dukungan bimbingan pengetahuan, wawasan, masukan serta arahan yang diberikan untuk menyelesaikan penelitian ini.

REFERENCES

[1] M. de Onis and F. Branca, “Childhood stunting: A global perspective,” May 01, 2016, *Blackwell Publishing Ltd*. doi: 10.1111/mcn.12231.

[2] F. Abdulla, M. M. A. El-Raouf, A. Rahman, R. Aldallal, M. S. Mohamed, and M. M. Hossain, “Prevalence and determinants of wasting among under-5 Egyptian children: Application of quantile regression,” *Food Sci Nutr*, vol. 11, no. 2, pp. 1073–1083, Feb. 2023, doi: 10.1002/fsn3.3144.

[3] WHO, “World Health Organization.” [Online]. Available: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-stunting-prevalence?utm_source=chatgpt.com

[4] S. Grantham-Mcgregor, Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp, “Child development in developing countries 1 Developmental potential in the fi rst 5 years for children in developing countries,” 2007. Accessed: Oct. 17, 2025. [Online]. Available: https://core.ac.uk/download/pdf/13103225.pdf

[5] C Scheffler, M. Hermanussen, B. Bogin, D. S. Liana, F. Taolin, and P. M. V. P. Cempaka, “Stunting is not a synonym of malnutrition,” May 2019, doi: https://doi.org/10.1038/s41430-019-0439-4.

[6] SSGI, *Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasiona, Provinsi dan Kabupaten/Kota Tahun 2021*. 2021. Accessed: Oct. 17, 2025. [Online]. Available: https://dinkes.acehprov.go.id/l-content/uploads/Hasil_SSGI_Tahun_2021_Tingkat_Kabupaten_Kota.pdf

[7] C. Mondon, P. Y. Tan, C. L. Chan, T. N. Tran, and Y. Y. Gong, “Prevalence, determinants, intervention strategies and current gaps in addressing childhood malnutrition in Vietnam: a

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

systematic review," *BMC Public Health*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12889-024-18419-8.

[8] A. Hamed, A. Hegab, and E. Roshdy, "Prevalence and factors associated with stunting among school children in Egypt," *Eastern Mediterranean Health Journal*, vol. 26, no. 7, pp. 787–793, 2020, doi: 10.26719/emhj.20.047.

[9] S. M. J. Rahman *et al.*, "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," *PLoS One*, vol. 16, no. 6 June 2021, Jun. 2021, doi: 10.1371/journal.pone.0253172.

[10] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: a practical introduction," *BMC Med Res Methodol*, vol. 19, no. 1, Mar. 2019, doi: 10.1186/s12874-019-0681-4.

[11] E.-S. M. El-Alfy and S. A. Mohammed, "A review of machine learning for big data analytics: bibliometric approach," *Technol Anal Strateg Manag*, vol. 32, no. 8, pp. 984–1005, Aug. 2020, doi: 10.1080/09537325.2020.1732912.

[12] O. N. Chilyabanyama *et al.*, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, Jul. 2022, doi: 10.3390/children9071082.

[13] H. Shen, H. Zhao, and Y. Jiang, "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," *Children*, vol. 10, no. 10, Oct. 2023, doi: 10.3390/children10101638.

[14] V. Desai, J. Cottrell, and L. Sowerby, "No longer a blank cheque: a narrative scoping review of physician awareness of cost," *Public Health*, vol. 223, pp. 15–23, Oct. 2023, doi: 10.1016/j.puhe.2023.07.009.

[15] M. S. Haris, A. N. Khudori, and W. T. Kusuma, "Perbandingan Metode Supervised Machine Learning untuk Prediksi Prevalensi Stunting di Provisi Jawa Timur," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 7, p. 1571, Dec. 2022, doi: 10.25126/jtiik.2022976744.

[16] A. B. Zemariam *et al.*, "Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms," *PLoS One*, vol. 20, no. 1 January, Jan. 2025, doi: 10.1371/journal.pone.0316452.

[17] E. Indrisari, H. Febiansyah, and B. Adiwinoto, "A Systematic Literature Review on the Application of Machine Learning for Predicting Stunting Prevalence in Indonesia (2020–2024)," 2025. doi: 10.32736/sisfokom.v14i3.2366.

[18] I. Rahmi, M. Susanti, H. Yozza, and F. Wulandari, "CLASSIFICATION OF STUNTING IN CHILDREN UNDER FIVE YEARS IN PADANG CITY USING SUPPORT VECTOR MACHINE," *Barekeng*, vol. 16, no. 3, pp. 771–778, Sep. 2022, doi: 10.30598/barekengvol16iss3pp771-778.

[19] M. As and E. R. Subhiyakto, "Performance Comparison of Random Forest , SVM , and XGBoost Algorithms with SMOTE for Stunting Prediction," vol. 9, no. 4, pp. 1163–1169, 2025, doi: https://doi.org/10.30871/jaic.v9i4.9701.

[20] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data Optimizing Stunting Data Classification Through Ensemble Learning on Multiclass Labels with Imbalance Data," 2024. doi: https://doi.org/10.62411/tc.v23i1.9779.

[21] 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.

[22] J. Sadhasivam, A. Rathore, I. Bose, S. Bhattacharjee, and S. Jayavel, "A Survey of Machine Learning Algorithms," *International Journal of Engineering Trends and Technology*, vol. 68, 2020, [Online]. Available: http://www.ijettjournal.org

[23] C. Cortes, V. Vapnik, and L. Saitta, "Support-Vector Networks Editor," Kluwer Academic Publishers, 1995. Accessed: Oct. 15, 2025. [Online]. Available: https://link.springer.com/article/10.1007/BF00994018

Vol. 6, No. 5, October 2025, Page. 5305-5319
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5285

[24] R. Kusumaningrum, T. A. Indihatmoko, S. R. Juwita, A. F. Hanifah, K. Khadijah, and B. Surarso, "Benchmarking of multi-class algorithms for classifying documents related to stunting," *Applied Sciences (Switzerland)*, vol. 10, no. 23, pp. 1–13, Dec. 2020, doi: 10.3390/app10238621.

[25] L. Breiman, "Random Forests," 2001. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[26] J. Bergstra, J. B. Ca, and Y. B. Ca, "Random Search for Hyper-Parameter Optimization Yoshua Bengio," 2012. [Online]. Available: http://scikit-learn.sourceforge.net.

[27] S. Abd El-Ghany and A. A. Abd El-Aziz, "A Robust Tuned Random Forest Classifier Using Randomized Grid Search to Predict Coronary Artery Diseases," *Computers, Materials and Continua*, vol. 75, no. 2, pp. 4633–4648, 2023, doi: 10.32604/cmc.2023.035779.

[28] S. Sutarmi, W. Warijan, T. Indrayana, D. P. P. B, and I. Gunawan, "Machine Learning Model For Stunting Prediction," *Jurnal Health Sains*, vol. 4, no. 9, pp. 10–23, 2023, doi: 10.46799/jhs.v4i9.1073.

[29] N. Fathirachman Mahing, A. Lazuardi Gunawan, A. Foresta Azhar Zen, F. Abdurrachman Bachtiar, and S. Agung Wicaksono, "Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 7, pp. 1527–1536, 2023, doi: 10.25126/jtiik.1078010.

[30] S. Lestari and F. Ilmu Komputer, "Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms," *Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, 2024, doi: 10.33395/v8i2.13448.

[31] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, "Machine Learning Method to Predict the Toddlers' Nutritional Status," *Jurnal Infotel*, vol. 16, no. 1, pp. 32–43, 2024, doi: 10.20895/infotel.v15i4.988.

[32] S. Belarouci, S. Bouchikhi, and M. A. Chikh, "Comparative study of balancing methods: case of imbalanced medical data," *Int J Biomed Eng Technol*, vol. 21, no. 3, p. 247, 2016, doi: 10.1504/IJBET.2016.078288.

[33] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 224–233, Dec. 2024, doi: 10.30591/jpit.v9i3.6640.

[34] F. Arden and C. Safitri, "Hyperparameter Tuning Algorithm Comparison with Machine Learning Algorithms," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Dec. 2022, pp. 183–188. doi: 10.1109/ICITISEE57756.2022.10057630.

[35] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput Surv*, vol. 31, no. 3, pp. 264–323, 1999, doi: 10.1145/331499.331504.