

TESIS
PERBANDINGAN KINERJA PREDIKSI STUNTING ANAK
MENGGUNAKAN SUPPORT VECTOR MACHINE VS
RANDOM FOREST DENGAN OPTIMASI GRID SEARCH



disusun oleh

MARTHINUS IKUN ELIM

24.55.1594

Konsentrasi : Digital Transformation Intelligence (DTI)

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

TESIS
PERBANDINGAN KINERJA PREDIKSI STUNTING ANAK
MENGGUNAKAN SUPPORT VECTOR MACHINE VS
RANDOM FOREST DENGAN OPTIMASI GRID SEARCH

PERFORMANCE COMPARISON OF CHILD STUNTING
PREDICTION SUPPORT VECTOR MACHINE VS RANDOM
FOREST WITH GRID SEARCH OPTIMIZATION

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana

Program Studi *PJJ Informatika*



disusun oleh

MARTHINUS IKUN ELIM

24.55.1594

Konsentrasi : Digital Transformation Intelligence (DTI)

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2026

HALAMAN PERSETUJUAN

**PERBANDINGAN KINERJA PREDIKSI STUNTING ANAK
MENGUNAKAN SUPPORT VECTOR MACHINE VS RANDOM
FOREST DENGAN OPTIMASI GRID SEARCH**

**PERFORMANCE COMPARISON OF CHILD STUNTING PREDICTION
SUPPORT VECTOR MACHINE VS RANDOM FOREST WITH GRID
SEARCH OPTIMIZATION**

yang disusun dan diajukan oleh

Marthinus Ikun Elm

24.55.1594

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 04 Maret 2026

Dosen Pembimbing,



Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

HALAMAN PENGESAHAN

**PERBANDINGAN KINERJA PREDIKSI STUNTING ANAK
MENGUNAKAN SUPPORT VECTOR MACHINE VS RANDOM
FOREST DENGAN OPTIMASI GRID SEARCH**

**PERFORMANCE COMPARISON OF CHILD STUNTING PREDICTION
SUPPORT VECTOR MACHINE VS RANDOM FOREST WITH GRID
SEARCH OPTIMIZATION**

yang disusun dan diajukan oleh

Marthinus Ikuu Elim

24.55.1594

Telah dipertahankan di depan Dewan Penguji
pada tanggal 04 Maret 2026

Susunan Dewan Penguji

Nama Penguji

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302258

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302198

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 04 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Marthinus Ikun Elim
NIM : 24.55.1594

Menyatakan bahwa Tesis dengan judul berikut:

Perbandingan Kinerja Prediksi Stunting Anak Menggunakan Support Vector Machine Vs Random Forest Dengan Optimasi Grid Search

Dosen Pembimbing : Prof. Dr. Ema Utami, S.Si., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 04 Maret 2026

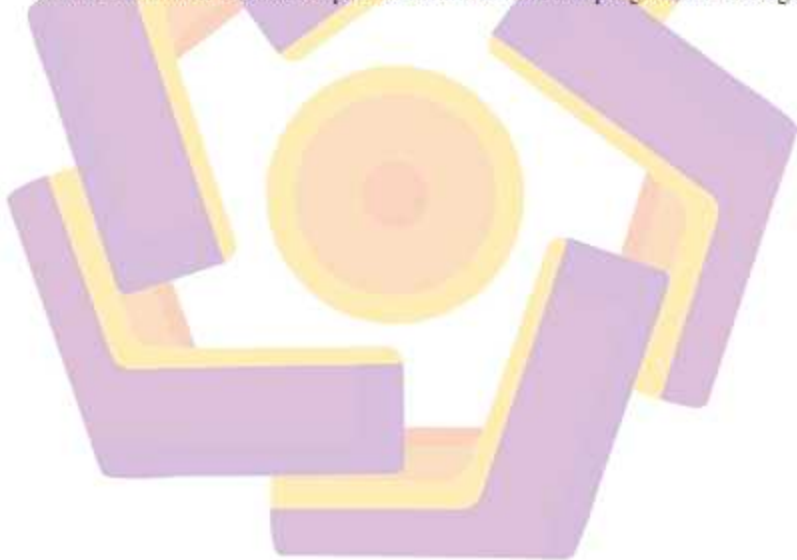
Yang Menyatakan,



Marthinus Ikun Elim

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, karya ini kupersembahkan untuk Pembina Yayasan Gemma Galgani Kupang yang telah memberikan beasiswa untuk melanjutkan studi ini. Kedua orang tuaku tercinta Bapak Yosep Elim dan Ibu Elisabeth Fetok, Istriku tercinta Yeni Yuniati Saku dan Kedua anakkku tercinta Cecilia Huanran Elim dan Gracia Jili Elim yang selalu memberikan kasih sayang, Doa dan dukungan tanpa lelah. Para dosen dan pembimbing yang telah membimbing dan memberi ilmu dengan penuh kesabaran. Sahabat-sahabat terbaik yang selalu ada dalam suka dan duka selama perjuangan kuliah. Almamater tercinta tempat aku menimba ilmu dan pengalaman berharga.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga Tesis yang berjudul "Perbandingan Kinerja Prediksi Stunting Anak Menggunakan Support Vector Machine VS Random Forest Dengan Optimasi Grid Search" ini dapat diselesaikan dengan baik. Penyusunan Tesis ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Magister Komputer pada Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa keberhasilan penyelesaian Tesis ini tidak lepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Ibu Prof. Dr. Ema Utami, S.Si., M.Kom. selaku Dosen Pembimbing I, yang telah meluangkan waktu dan memberikan arahan serta masukan berharga selama proses penyusunan hingga selesainya Tesis ini.
2. Bapak I Made Artha Agastya, S.T., M.Eng., PhD dan Bapak Robert Marco, S.T., M.T., Ph.D. selaku Tim Dosen Penguji yang telah memberikan kritik dan saran konstruktif demi kesempurnaan Tesis ini.
3. Bapak Dhani Ariatmanto, S.Kom., M.Kom., Ph.D. selaku Ketua Program Studi S2 PJJ Informatika, atas dukungan administratif dan kebijakan yang memudahkan penulis selama masa studi.
4. Ibu Pembina serta Jajaran pengurus Yayasan Gemma Galgani atas bantuan pendanaan yang diberikan sehingga studi ini bisa terselesaikan dengan baik.
5. Kedua orang tua penulis, Bapak Yosep Elim dan Ibu Elisabeth Fetok, Istri penulis Yeni Yuniati Saku dan Kedua anak penulis, Cecilia Huanran Elim dan Gracia Jili Elim dan keluarga besar serta teman seperjuangan yang senantiasa memberikan Doa, motivasi, dan kasih sayang yang tak terhingga.

Penulis berharap semoga Tesis ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya Digital Transformation Intelligence.

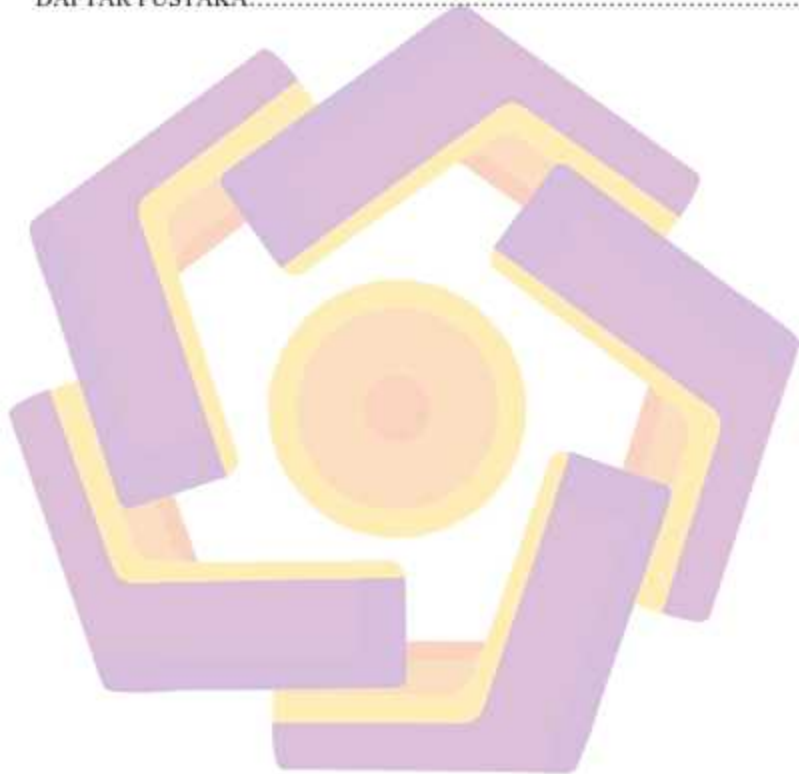
Yogyakarta, 04 Maret 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN TESIS.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMBANG DAN SINGKATAN.....	xi
DAFTAR ISTILAH.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	7
BAB 2 TINJAUAN PUSTAKA.....	10
2.1 Tinjauan Pustaka.....	10
2.2 Keaslian Penelitian.....	14
2.3 Landasan Teori.....	18
BAB 3 METODE PENELITIAN.....	31
3.1 Jenis, Sifat, dan Pendekatan Penelitian.....	31
3.2 Metode Pengumpulan Data.....	31
3.3 Metode Analisis Data.....	32
3.4 Alur Penelitian.....	33
BAB 4 HASIL PENELITIAN DAN PEMBAHASAN.....	40
4.1 Tinjauan Data dan Pra-pemrosesan.....	40
4.2 Kinerja Support Vector Machine (SVM).....	46
4.3 Kinerja Random Forest (RF).....	47
4.4 Fitur Penting Pada Random Forest.....	49
4.5 Skor AUC ROC per Kelas Random Forest.....	49

4.6 Deployment Model.....	50
4.7 Analisis Signifikansi Praktis Perbaikan Kinerja Model.....	51
4.8 Analisis Keunggulan Konsisten Random Forest (RF).....	53
4.9 Pembahasan.....	54
BAB 5 PENUTUP.....	58
5.1 Kesimpulan.....	58
5.2 Saran.....	59
DAFTAR PUSTAKA.....	60

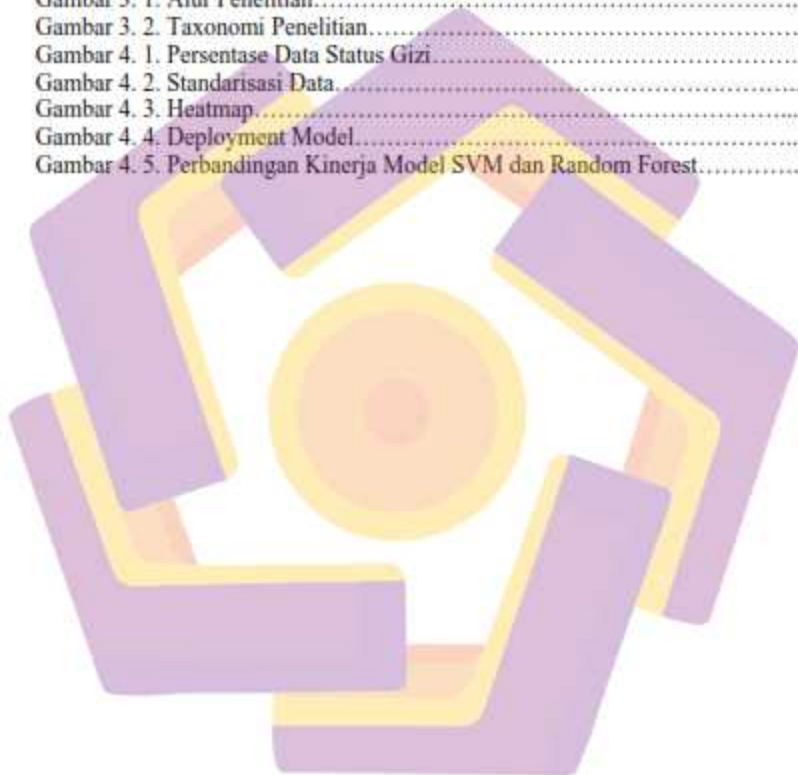


DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian Tuliskan Judul Tesis di Baris Ini.....	14
Tabel 3.1. Dataset.....	34
Tabel 3.2. Parameter Terbaik SVM with Grid Search.....	36
Tabel 3.3. Parameter Terbaik RF with Grid Search.....	37
Tabel 4.1. Missing Values.....	42
Tabel 4.2. Duplicate Values.....	42
Tabel 4.3. Tabel Fitur Reduksi Data.....	43
Tabel 4.4. One Hot Encoding.....	43
Tabel 4.5. Pemisahan Fitur dan Target.....	45
Tabel 4.6. Validasi Data.....	45
Tabel 4.7. Laporan Klasifikasi SVM Setelah Optimasi Grid Search.....	46
Tabel 4.8. Laporan Klasifikasi Random Forest Setelah Optimasi Grid Search.....	48
Tabel 4.9. Fitur Penting Pada Model Random Forest.....	48
Tabel 4.10. Laporan Skor ROC AUC tanpa Grid Search.....	49
Tabel 4.11. Laporan Skor ROC AUC dengan Grid Search.....	49
Tabel 4.12. Analisis Signifikansi SVM.....	51
Tabel 4.13. Analisis Signifikansi RF.....	52
Tabel 4.14. Perbandingan Kinerja SVM dan Random Forest.....	56

DAFTAR GAMBAR

Gambar 2. 1. Arsitektur Machine Learning.....	18
Gambar 2. 2. SVM Model.....	19
Gambar 2. 3. Flowchart SVM.....	20
Gambar 2. 4. Random Forest Model.....	25
Gambar 2. 5. Mekanisme kerja grid serach.....	28
Gambar 3. 1. Alur Penelitian.....	33
Gambar 3. 2. Taxonomi Penelitian.....	39
Gambar 4. 1. Persentase Data Status Gizi.....	40
Gambar 4. 2. Standarisasi Data.....	43
Gambar 4. 3. Heatmap.....	44
Gambar 4. 4. Deployment Model.....	50
Gambar 4. 5. Perbandingan Kinerja Model SVM dan Random Forest.....	55



DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence (Kecerdasan Buatan)
AUC	Area Under the Curve
BB/TB	Berat Badan menurut Tinggi Badan
BB/U	Berat Badan menurut Umur
GS	Grid Search
HPK	1.000 Hari Pertama Kehidupan
JST	Jaringan Syaraf Tiruan
LiLA	Lingkar Lengan Atas
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
RF	Random Forest
ROC	Receiver Operating Characteristic
SD	Standar Deviasi
SSGI	Studi Status Gizi Indonesia
SVM	Support Vector Machine
SVR	Support Vector Regression
TB/U	Tinggi Badan menurut Umur
WHO	World Health Organization

DAFTAR ISTILAH

Stunting	Masalah kurang gizi kronis
Wasting	Kondisi di mana berat badan anak sangat kurang
Prevalensi	kondisi kesehatan dalam populasi pada waktu tertentu
Underweight	Kondisi berat badan di bawah rata-rata standar pada usia tertentu
Malnutrisi	kondisi gizi yang tidak seimbang
Antropometri	Metode pengukuran dimensi fisik tubuh manusia
Machine Learning (ML)	Cabang kecerdasan buatan (AI)
Supervised Learning	Jenis pembelajaran mesin
Algoritma SVM	Metode klasifikasi mencari <i>hyperplane</i> terbaik
Random Forest, XGBoost, Gradient Boosting	Algoritma berbasis ensemble (gabungan)
Hyperparameter	Pengaturan eksternal model
Grid Search	Metode sistematis untuk menemukan kombinasi nilai hyperparameter terbaik
Presisi, Recall, F1-Score:	Ukuran kinerja jika data tidak seimbang
MAE/MSE	Ukuran kesalahan (<i>error</i>) untuk data regresi
ROC-AUC	Kurva yang mengukur kemampuan model
Imbalanced Data	Kondisi di mana jumlah data dalam satu kelas berbeda dengan kelas lain
Variabel Independen	Variabel target yang ingin diprediksi
Variabel Dependen	Variabel prediktor yang mempengaruhi hasil klasifikasi
Data Sekunder	Data yang dikumpulkan pihak lain sebelumnya
Imputasi	Mengisi data yang kosong/hilang
Encoding	Mengubah data kategori (huruf) menjadi angka
Scaling	Menyamakan skala angka antar variabel
Splitting	Membagi data menjadi data latih dan data uji

INTISARI

Stunting merupakan masalah kesehatan global yang serius, terutama di negara berkembang. Prevalensinya tinggi di Indonesia, mencapai sekitar 24,4% di antara anak-anak di bawah lima tahun pada tahun 2021. Kondisi ini, yang didefinisikan sebagai gagal tumbuh karena kekurangan gizi kronis, infeksi berulang, dan kurangnya stimulasi psikososial, memiliki dampak jangka panjang pada perkembangan kognitif dan kapasitas produktif individu. Studi ini bertujuan untuk melakukan analisis komparatif algoritma Support Vector Machine dan Random Forest dalam memprediksi stunting pada anak, dengan fokus pada evaluasi dampak optimasi hyperparameter menggunakan Grid Search pada kinerja model. Studi ini menggunakan dataset stunting publik dari Kaggle dan mencakup langkah-langkah pra-pemrosesan data seperti penanganan nilai yang hilang, duplikasi, pengkodean, dan penskalaan. Data kemudian dibagi menjadi 80% untuk pelatihan, 10% untuk pengujian, dan 10% untuk validasi. Metrik evaluasi komprehensif seperti presisi, recall, F1-score, dan ROC-AUC juga digunakan untuk menilai kinerja model. Optimasi Grid Search diterapkan pada kedua model untuk menemukan kombinasi hyperparameter terbaik. Hasil eksperimen menunjukkan bahwa optimasi Grid Search secara signifikan meningkatkan akurasi model SVM dari 94,29% menjadi 98,37%. Sementara itu, model Random Forest menunjukkan kinerja yang sangat tinggi, mencapai akurasi 99,59% baik sebelum maupun setelah optimasi Grid Search. Temuan ini menggarisbawahi potensi signifikan model pembelajaran mesin dalam mendukung upaya pencegahan stunting untuk kebijakan intervensi kesehatan masyarakat. Penelitian ini berkontribusi pada pengembangan sistem pendukung keputusan berbasis pembelajaran mesin untuk kesehatan masyarakat, khususnya dalam strategi deteksi dini dan intervensi untuk stunting.

Kata kunci: Grid Search, Machine Learning, Random Forest, Stunting, Support Vector Machine.

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.

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

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Stunting atau gangguan pertumbuhan dan perkembangan merupakan tantangan kesehatan global yang kritis, terutama yang mempengaruhi anak-anak di negara-negara berkembang (de Onis & Branca, 2016)-(Mondon et al., 2024). 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 (WHO, 2025). Stunting adalah gangguan pertumbuhan dan perkembangan yang dialami anak akibat kekurangan gizi kronis dan infeksi berulang, terutama dalam 1.000 Hari Pertama Kehidupan (HPK). Berdasarkan kriteria Klinis nak dinyatakan stunting jika nilai Z-score tinggi badan menurut umur (TB/U) kurang dari -2 standar deviasi (SD) dari kurva pertumbuhan standar WHO. Dampak Tidak hanya soal fisik yang pendek, stunting juga berkaitan dengan perkembangan otak yang tidak maksimal, yang dapat memengaruhi kemampuan kognitif di masa depan. Gizi buruk adalah bentuk terparah dari proses kekurangan gizi akut. Dalam literatur medis (seperti buku bagan Manajemen Terpadu Balita Sakit), gizi buruk sering dibagi menjadi dua kategori klinis: Marasmus (tampak sangat kurus) dan Kwashiorkor (disertai bengkak atau edema). Berdasarkan kriteria Klinis Indikator berat badan menurut tinggi badan (BB/TB) kurang dari -3 standar deviasi (SD). Lingkar Lengan Atas (LiLA) < 11,5 cm pada anak usia 6–59 bulan.

Adanya pitting edema bilateral (bengkak pada kedua punggung kaki/tangan) yang bersifat nutrisi. 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 (Grantham-Mcgregor et al., 2007)-(C Scheffler et al., 2019). 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 (SSGI, 2021)-(Mondon et al., 2024).

Tantangan utama dalam klasifikasi stunting, selain akurasi, adalah risiko terjadinya overfitting yang dipicu oleh kesalahan penyetelan parameter model. Tanpa optimasi yang tepat, model cenderung menjadi terlalu kompleks dan hanya "menghafal" pola spesifik pada data latih, sehingga kehilangan kemampuan generalisasi saat menghadapi data kesehatan baru di lapangan. Hal ini sangat krusial karena model yang mengalami overfitting mungkin menunjukkan akurasi tinggi pada saat pengujian internal, namun gagal memberikan prediksi yang valid bagi balita di dunia nyata. Kegagalan ini dapat menyebabkan kesalahan fatal dalam deteksi dini. Oleh karena itu, penggunaan Grid Search dalam penelitian ini bertujuan untuk mengontrol kompleksitas model dan menemukan parameter penalti serta kedalaman pohon yang paling optimal bagi SVM dan Random Forest, guna

memastikan model tetap stabil dan tidak terjebak pada kebisingan data (noise) (Müller & Guido, n.d.).

Untuk mengatasi masalah ini, identifikasi dini dan akurat terhadap anak-anak yang berisiko sangat penting (Hamed et al., 2020). Metode tradisional untuk mengidentifikasi faktor risiko malnutrisi sering kali mengandalkan model statistik klasik seperti regresi logistik (Rahman et al., 2021). Namun, dengan munculnya teknologi komputasi canggih, pembelajaran mesin (ML) telah muncul sebagai pendekatan yang kuat dan inovatif untuk prediksi dalam domain kesehatan masyarakat (Sidey-Gibbons & Sidey-Gibbons, 2019)-(El-Alfy & Mohammed, 2020). Model ML terbukti mampu mengatasi tantangan analitis model statistik klasik, seperti masalah multikolinearitas antar variabel, dan memerlukan asumsi yang lebih sedikit terhadap data (Chilyabanyama et al., 2022). Dengan menganalisis kumpulan data kompleks yang mencakup variabel kesehatan, gizi, dan lingkungan, algoritma ML dapat membangun model prediktif yang sangat akurat (Shen et al., 2023). 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 (Desai et al., 2023).

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 (Haris et al., 2022). 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 (Chilyabanyama et al., 2022). 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 fl = 78%, AUC = 85%) lebih unggul dalam memprediksi stunting dibandingkan dengan algoritma pembelajaran mesin lain (Zemariam et al., 2025).

Secara spesifik, algoritma Random Forest (RF) dan Support Vector Machine (SVM) telah sering digunakan dalam penelitian prediksi kesehatan. Tinjauan literatur sistematis oleh 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 (Indrisari et al., 2025). 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 (Rahmi et al., 2022). 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 (As & Subhiyanto, 2025). Variabilitas kinerja ini juga terlihat pada studi lain, di mana algoritma seperti Gradient Boosting menunjukkan performa terbaik (68,47%) dibandingkan RF dan SVM (Prasetyo & Nugroho, 2024). 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 (Mubarak et al., 2025). Penelitian menunjukkan bahwa penyesuaian parameter, khususnya pada SVM, sangat krusial untuk mencapai akurasi maksimal (Rahmi et al., 2022).

1.2. Rumusan Masalah

Rumusan masalah dari penelitian ini mempunyai fokus masalah, yaitu:

1. Bagaimana kinerja komparatif algoritma Support Vector Machine (SVM) dan Random Forest (RF) dalam memprediksi stunting pada anak sebelum dan sesudah menerapkan optimasi hyperparameter menggunakan teknik Grid Search?
2. Seberapa signifikan peningkatan akurasi dan metrik evaluasi lainnya (Presisi, Recall, F1-Score, dan ROC-AUC) pada model SVM setelah dioptimalkan dengan Grid Search?
3. Model manakah (SVM yang dioptimalkan atau Random Forest yang dioptimalkan) yang menunjukkan kinerja prediktif yang lebih optimal dan tangguh dalam menangani masalah klasifikasi stunting, khususnya pada kelas minoritas, pada dataset yang tidak seimbang?

1.3. Batasan Masalah

Bagian ini membatasi ruang lingkup agar penelitian tetap fokus dan tidak meluas ke luar konteks yang data/metodenya tersedia.

Batasan Permasalahan:

1. Algoritma: Penelitian ini difokuskan pada perbandingan kinerja dua algoritma supervised learning spesifik, yaitu Support Vector Machine (SVM) dan Random Forest (RF).
2. Sumber Data: Data yang digunakan adalah dataset publik sekunder bertema stunting anak yang diperoleh dari repositori Kaggle, bukan data primer yang diambil langsung dari lapangan.

3. Variabel Data: Variabel yang digunakan terbatas pada data antropometri dan demografi yang tersedia di dataset, seperti tinggi badan, berat badan, lingkar lengan (LILA), usia, jenis kelamin, serta rasio berat/tinggi badan.

Asumsi Dasar:

1. Kecukupan Data: Dataset dikategorikan sebagai skala sedang (bukan Big Data), sehingga penggunaan metode Grid Search diasumsikan cukup memadai dan efisien dibandingkan metode yang lebih kompleks seperti Bayesian Optimization.
2. Metode Validasi: Penelitian ini berasumsi bahwa pemisahan data latih dan uji (train-test split) tunggal (80:20) sudah cukup untuk evaluasi, dan tidak menerapkan k-fold cross-validation demi efisiensi waktu komputasi.

Batasan Solusi:

1. Teknik Optimasi: Optimasi hyperparameter hanya dilakukan menggunakan teknik Grid Search untuk menemukan kombinasi parameter terbaik.
2. Evaluasi Kinerja: Evaluasi kinerja model dibatasi pada metrik Akurasi, Presisi, Recall, F1-Score, dan ROC-AUC.
3. Lingkungan Pengembangan: Implementasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka scikit-learn.

1.4. Tujuan Penelitian

Bagian ini memuat penjelasan secara spesifik:

1. Solusi yang Ditawarkan:

1. Menerapkan optimasi hyperparameter menggunakan Grid Search pada algoritma SVM dan Random Forest untuk mengatasi masalah kinerja model yang suboptimal atau default.
2. Membangun model prediksi stunting yang mampu menangani ketidakseimbangan kelas (imbalanced data) pada dataset stunting.

Hal yang Ingin Dicapai:

1. Mengukur dan membandingkan secara langsung kinerja algoritma SVM dan Random Forest sebelum dan sesudah dilakukan optimasi Grid Search.
2. Mendapatkan konfigurasi hyperparameter terbaik yang dapat meningkatkan akurasi deteksi stunting, khususnya pada algoritma SVM yang terbukti meningkat signifikan setelah optimasi.
3. Mengidentifikasi fitur (variabel) yang paling berpengaruh (penting) dalam penentuan status stunting menggunakan fitur importance dari Random Forest.

1.5. Manfaat Penelitian

Bagian ini memuat penjelasan tentang:

Kontribusi Ilmiah:

1. Mengisi celah penelitian (research gap) mengenai perbandingan mendalam antara SVM dan Random Forest yang dioptimalkan secara sistematis dengan Grid Search, yang sebelumnya belum banyak dieksplorasi.
2. Memberikan bukti empiris bahwa Grid Search signifikan meningkatkan akurasi SVM (dari 94,29% menjadi 98,37%) dan memvalidasi kestabilan model Random Forest.

Kegunaan Sistem:

1. Sistem yang dihasilkan mampu memprediksi status gizi anak dengan akurasi yang sangat tinggi (mencapai 99,59% pada Random Forest), sehingga dapat diandalkan sebagai alat deteksi dini.
2. Sistem mampu memberikan informasi mengenai faktor antropometri (seperti berat per tinggi badan) yang paling krusial untuk dipantau.

Manfaat bagi Pengguna (Masyarakat/Tenaga Medis):

1. Membantu tenaga medis dan pembuat kebijakan dalam mengidentifikasi anak-anak yang berisiko tinggi mengalami stunting secara lebih presisi dan cepat.
2. Menyediakan purwarupa (prototype) aplikasi berbasis web yang memudahkan pengguna (orang tua atau kader kesehatan) untuk memasukkan data anak dan mendapatkan hasil prediksi status gizi secara instan.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian mengenai prediksi stunting menggunakan pendekatan Machine Learning telah banyak dilakukan dalam kurun waktu beberapa tahun terakhir. Tinjauan ini menguraikan enam penelitian relevan yang dipublikasikan antara tahun 2022 hingga 2025 untuk memetakan perkembangan metode dan mengidentifikasi celah penelitian yang ada.

Penelitian pertama dilakukan oleh Haris et al. (2022) yang membandingkan beberapa metode supervised learning untuk memprediksi prevalensi stunting di Jawa Timur. Kelebihan penelitian ini adalah kemampuannya menyoroti performa Support Vector Regression (SVR) yang terbukti memiliki nilai error (MAE dan MSE) lebih kecil dibandingkan Random Forest Regression dan Regresi Linier. Namun, fokus penelitian ini lebih pada pendekatan regresi untuk prevalensi wilayah, bukan klasifikasi status individu anak secara mendetail.

Dalam konteks internasional, Chilyabanyama et al. (2022) melakukan penelitian di Zambia yang berfokus pada klasifikasi stunting pada balita. Penelitian ini menemukan bahwa algoritma Random Forest memiliki akurasi tertinggi sebesar 79%, mengungguli algoritma SVC/SVM. Meskipun memberikan wawasan komparatif yang baik, akurasi yang dihasilkan masih di bawah 80%, yang menunjukkan masih adanya ruang untuk peningkatan performa model melalui optimasi lebih lanjut.

Studi selanjutnya oleh Zemariam et al. (2022) meneliti prediksi stunting pada kelompok demografi yang berbeda, yaitu remaja putri. Hasil penelitian menunjukkan bahwa Random Forest kembali lebih unggul dibandingkan algoritma lain dengan sensitivitas 81% dan Area Under Curve (AUC) sebesar 85%. Kekuatan penelitian ini terletak pada metrik evaluasi yang lengkap, namun keterbatasannya adalah fokus subjek yang spesifik pada remaja putri, sehingga generalisasi untuk balita perlu dikaji ulang.

Rahmi et al. (2022) melakukan penelitian spesifik menggunakan algoritma Support Vector Machine (SVM) di Padang, Indonesia. Penelitian ini menunjukkan hasil yang sangat impresif di mana model mampu mencapai akurasi 100% dengan penyesuaian hyperparameter yang optimal. Meskipun hasilnya sempurna, penelitian ini hanya berfokus pada satu algoritma (SVM) tanpa perbandingan langsung yang mendalam dengan metode ensemble seperti Random Forest pada dataset yang sama.

Sebaliknya, penelitian oleh Prasetyo dan Nugroho (2024) menemukan hasil yang kontradiktif. Dalam studi mereka, algoritma Gradient Boosting menunjukkan performa terbaik (68,47%), justru mengungguli Random Forest dan SVM yang memiliki performa lebih rendah. Kelemahan dari studi ini adalah rendahnya akurasi keseluruhan yang dicapai oleh model-model yang diuji, yang mengindikasikan bahwa tanpa optimasi yang tepat, algoritma populer pun bisa memberikan hasil suboptimal.

Terakhir, tinjauan literatur sistematis oleh Rahman et al. (2025) merangkum bahwa algoritma Random Forest (RF) memiliki kinerja terbaik dalam mengklasifikasikan anak-anak yang mengalami stunting, wasting, dan underweight dengan akurasi tertinggi masing-masing sebesar 88.3%, 87.7%, dan 85.7%. Studi hanya berfokus pada identifikasi faktor risiko malnutrisi menggunakan model klasik seperti regresi logistik dan tidak melakukan penyetelan hyper-parameter algoritma ML yang menghasilkan akurasi yang kurang memuaskan.

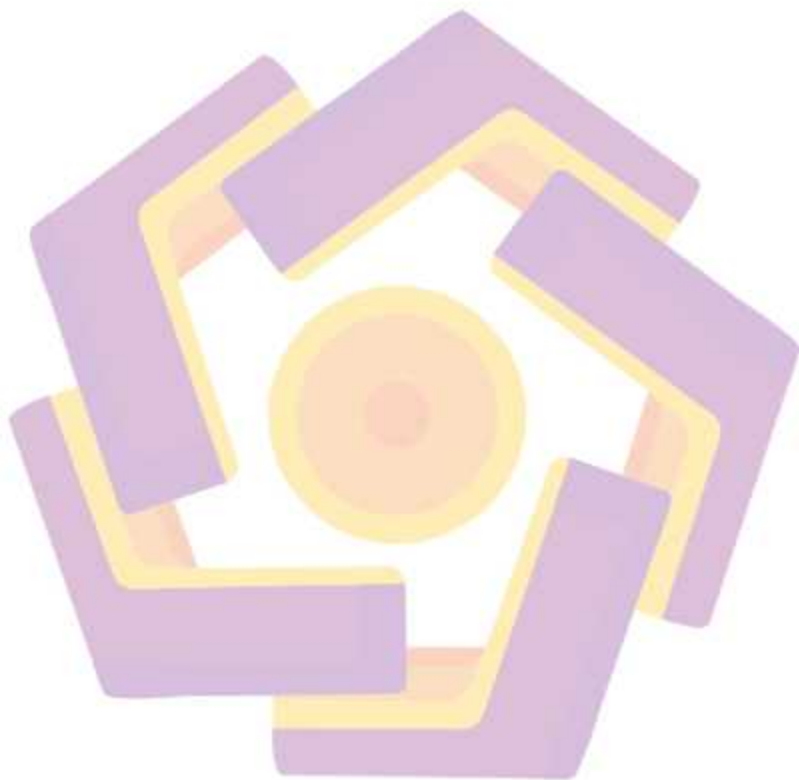
1. Posisi Penelitian

Berdasarkan telaah terhadap penelitian-penelitian di atas, teridentifikasi sebuah kesenjangan (gap) penelitian yang signifikan. Mayoritas penelitian sebelumnya, seperti yang dilakukan oleh Haris et al. (2022) dan Chilyabanyama et al. (2022), cenderung membandingkan algoritma dengan parameter default atau tanpa menerapkan teknik optimasi hyperparameter yang sistematis. Hal ini sering kali menghasilkan kinerja model yang suboptimal atau tidak menggambarkan potensi maksimal dari algoritma tersebut.

Selain itu, perbandingan langsung dan mendalam ("apple-to-apple") antara model SVM dan Random Forest yang keduanya sama-sama dioptimalkan menggunakan teknik Grid Search belum banyak dieksplorasi secara mendalam dalam konteks prediksi stunting.

Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan melakukan analisis komparatif antara SVM dan Random Forest yang telah melalui proses optimasi Grid Search. Kebaruan (novelty) dari penelitian ini terletak pada penerapan Grid Search untuk menemukan kombinasi hyperparameter terbaik guna

memastikan bahwa perbandingan kinerja dilakukan pada kondisi optimal masing-masing model, sehingga dapat memberikan rekomendasi berbasis bukti yang lebih akurat dan objektif untuk sistem pendukung keputusan kesehatan masyarakat.



2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Perbandingan Kinerja Prediksi Stunting Anak Menggunakan Support Vector Machine VS Random Forest Dengan Optimasi Grid Search

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Perbandingan Metode Supervised Machine Learning Untuk Prediksi Prevalensi Stunting Di Provinsi Jawa Timur	Haris et al., Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 2022	Membandingkan performa SVR, Random Forest Regression, dan Regresi Linier dalam memprediksi prevalensi stunting.	Hasil penelitian menunjukkan bahwa metode SVR memiliki nilai kesalahan terendah, dengan Mean Absolute Error (MAE) sebesar 0,91 dan Mean Squared Error (MSE) sebesar 1,30, mengungguli dua metode lainnya. Metode Support Vector Regression (SVR) terbukti paling unggul dengan nilai error (MAE dan MSE) terkecil.	Penelitian ini berfokus pada prediksi angka prevalensi (regresi) wilayah, bukan klasifikasi status individu anak.	Penelitian sebelumnya berfokus pada pendekatan regresi; penelitian ini berfokus pada klasifikasi status individu dengan penambahan tahap optimasi Grid Search.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	Classification Of Stunting In Children Under Five Years in Padang City Using Support Vector Machine	Rahmi et al., Berekeng Jurnal Ilmu Matematika dan Terapan, 2022	Untuk melakukan klasifikasi terhadap status stunting pada balita di Kota Padang menggunakan metode Support Vector Machine (SVM) serta menentukan parameter model terbaik (nilai cost dan gamma).	Penelitian ini berhasil membentuk model klasifikasi dengan parameter optimal yaitu nilai $cost = 10$ dan $gamma = 5$. Model tersebut memberikan hasil estimasi klasifikasi dengan tingkat akurasi yang sangat tinggi (mencapai 100% pada data yang diuji).	Penelitian ini memiliki keterbatasan dalam hal jumlah data atau variasi dataset yang digunakan. Saran untuk penelitian selanjutnya adalah menggunakan dataset yang lebih besar dan lebih seimbang (balanced) serta mempertimbangkan faktor-faktor lain yang lebih kompleks yang memengaruhi stunting.	Berbeda dengan penelitian sebelumnya yang mungkin hanya menggunakan algoritma standar atau metode statistik konvensional, penelitian ini secara spesifik menerapkan SVM dengan pencarian parameter (tuning) cost dan gamma untuk mendapatkan hyperplane pembatas yang paling optimal dalam memisahkan kelas stunting dan tidak stunting.
3	Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine	Zemariam et al., PLOS One, 2025	Memprediksi risiko stunting pada remaja putri menggunakan berbagai algoritma klasifikasi.	Random Forest menunjukkan performa terbaik dengan sensitivitas 81% dan AUC 85%.	Subjek penelitian terbatas pada remaja putri, sehingga karakteristik data berbeda dengan balita yang memiliki faktor pertumbuhan lebih dinamis.	Penelitian Zemariam et al. menggunakan model dasar pada subjek remaja; penelitian ini berfokus pada balita dan menguji efektivitas Grid Search dalam meningkatkan stabilitas model.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	learning algorithms					
4	Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia	Chilyabanyama et al., MDPI, 2022	Mengklasifikasikan status stunting pada balita di Zambia menggunakan algoritma Machine Learning.	Algoritma Random Forest memberikan akurasi tertinggi sebesar 79%, mengungguli model Support Vector Classification (SVC).	Akurasi yang dicapai masih di bawah 80%, sehingga diperlukan teknik optimasi parameter untuk meningkatkan performa model.	Penelitian Chilyabanyama et al. menggunakan parameter standar; sedangkan penelitian ini melakukan inovasi dengan optimasi Grid Search untuk mencari parameter terbaik pada SVM dan Random Forest.
5	Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data	Prasetyo dan Nugroho, Techno.com, 2024	Penelitian ini bertujuan untuk mengevaluasi kinerja berbagai metode machine learning dan teknik ensemble learning dalam mengklasifikasikan data stunting yang memiliki label multiclass (banyak kategori) dan masalah ketidakseimbangan	Berdasarkan hasil pengujian, algoritma Gradient Boosting menunjukkan performa terbaik dibandingkan metode lainnya dengan nilai akurasi sebesar 68,47%. Penelitian menyimpulkan bahwa teknik ensemble dapat membantu	Kelemahan utama yang disebutkan adalah rendahnya performa pada kategori label tertentu karena distribusi data yang sangat tidak seimbang (beberapa kategori memiliki jumlah sampel yang sangat sedikit). Peneliti menyarankan penggunaan teknik	Dibandingkan penelitian sebelumnya yang mungkin hanya menggunakan algoritma tunggal atau data yang seimbang, penelitian ini berfokus pada kompleksitas label multiclass dan imbalance data. Inovasinya terletak pada eksperimen penggunaan berbagai teknik Ensemble Learning (seperti Bagging dan Boosting) untuk

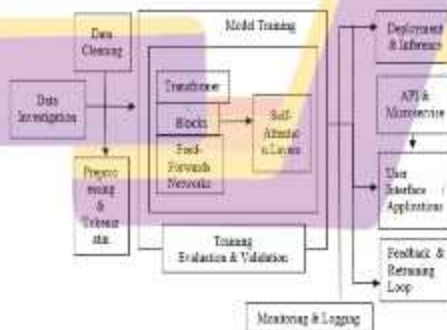
No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			data (imbalance data).	menangani kompleksitas data stunting, meskipun tantangan imbalance data tetap memengaruhi hasil secara signifikan.	penanganan imbalance data yang lebih maju seperti SMOTE atau penyesuaian bobot kelas pada penelitian berikutnya.	melihat mana yang paling tangguh menghadapi data dunia nyata yang tidak ideal.
6	Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach	Rahman et al, PLOS One, 2021	Untuk menggunakan algoritma machine learning (ML) guna mendeteksi faktor-faktor risiko malnutrisi (stunting, wasting, dan underweight) serta prediksinya pada anak-anak Bangladesh di bawah usia lima tahun.	Faktor-faktor risiko utama stunting, wasting, dan underweight Algoritma Random Forest (RF) memiliki kinerja terbaik dalam mengklasifikasikan anak-anak yang mengalami stunting, wasting, dan underweight dengan akurasi tertinggi masing-masing sebesar 88.3%, 87.7%, dan 85.7%. Identifikasi	Analisis interpretasi model ML (selain akurasi) seperti pentingnya fitur dari algoritma terbaik (RF) dapat memberikan wawasan yang lebih mendalam bagi pembuat kebijakan.	Penelitian rahman menyebutkan bahwa studi sebelumnya hanya berfokus pada identifikasi faktor risiko malnutrisi menggunakan model klasik seperti regresi logistik dan tidak melakukan penyetelan hyper-parameter algoritma ML yang menghasilkan akurasi yang kurang memuaskan. Penelitian ini berusaha mengatasi kelemahan ini dengan menggunakan kombinasi regresi logistik untuk

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				faktor risiko dan model prediksi ini diharapkan dapat membantu para pengambil kebijakan dalam upaya mengurangi malnutrisi pada anak-anak di bawah usia lima tahun di Bangladesh.		identifikasi faktor risiko dan algoritma ML dengan penyetelan yang lebih baik untuk prediksi.

2.3. Landasan Teori

1. Machine Learning

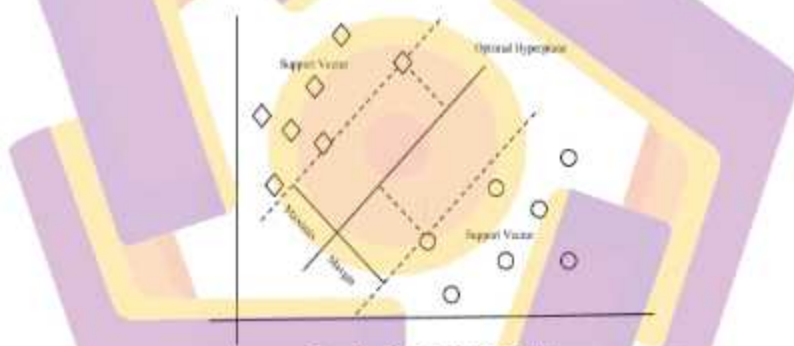
Pembelajaran mesin adalah subbidang kecerdasan buatan, yang secara luas didefinisikan sebagai kemampuan mesin untuk meniru perilaku manusia yang cerdas. Sistem kecerdasan buatan digunakan untuk melakukan tugas-tugas kompleks dengan cara yang mirip dengan cara manusia memecahkan masalah. Pembelajaran mesin adalah salah satu cara untuk menggunakan AI. Pembelajaran mesin didefinisikan pada tahun 1950-an oleh pelopor AI Arthur Samuel sebagai "bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit". "Fungsi sistem pembelajaran mesin dapat bersifat deskriptif, artinya sistem menggunakan data untuk menjelaskan apa yang terjadi; prediktif, artinya sistem menggunakan data untuk memprediksi apa yang akan terjadi; atau preskriptif, artinya sistem akan menggunakan data untuk membuat saran tentang tindakan apa yang harus diambil," tulis para peneliti (Brown, 2021).



Gambar 2. 1. Arsitektur Machine Learning

2. Support Vector Machine (SVM)

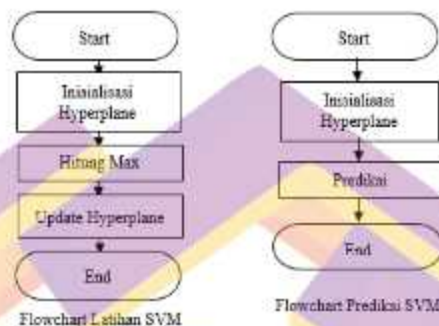
Support Vector Machine adalah metode pembelajaran terawasi yang digunakan untuk melakukan klasifikasi dan regresi. Konsep dasar dari SVM adalah mencari hiperplane (bidang) optimal yang memisahkan data dengan jelas ke dalam kategori-kategori yang berbeda. Dalam kasus klasifikasi citra, setiap citra dianggap sebagai suatu vektor fitur yang merepresentasikan piksel-piksel atau atribut tertentu dalam citra tersebut. SVM kemudian mencari hiperplane yang terbaik untuk memisahkan fitur-fitur dari kategori-kategori citra yang berbeda (Pamungkas, 2023).



Gambar 2. 2. SVM Model

Prinsip dasar pada SVM adalah linear classifier. Kemudian dikembangkan lebih lanjut agar dapat menyelesaikan permasalahan non-linear dengan memanfaatkan konsep kernel trick pada ruang yang berdimensi tinggi. SVM bekerja dengan memposisikan setiap data dalam sebuah peta berdimensi tinggi. Dengan memetakan setiap data dalam peta tersebut, diharapkan algoritma dapat memisahkan data-data tersebut menjadi dua kelas atau lebih dengan hyperplane. Hyperplane adalah sebuah garis yang memiliki satu dimensi lebih rendah dari

dimensi peta yang digunakan untuk memetakan data. Untuk membuat hyperplane dapat digunakan beberapa fungsi yang biasa disebut dengan fungsi kernel. Garis besar dari algoritma SVM dapat dipahami seperti pada gambar 2.3. berikut.



Gambar 2. 3. Flowchart SVM

Hyperparameter yang digunakan pada model SVM didapat melalui proses GridSearch CV untuk mencari hyperparameter terbaik. Hyperparameter yang dieksplorasi yaitu kernel: linear, poly, rbf, sigmoid dan C: 1, 2, 3 (Fathirachman Mahing et al., 2023).

SVM memiliki prinsip bisa melakukan klasifikasi kedalam dua kelompok dengan menentukan hyperplane yang tepat. Rumus perhitungan SVM adalah sebagai berikut (Givari et al., 2022) :

1. Titik data : $x_i = \{ x_1, x_2, \dots, x_n \} \in R^n$
2. Kelas data : $y_i \in \{-1, +1\}$
3. Pasangan data dan kelas : $\{(x_i, y_i)\}^{N_{i=1}}$
4. Maksimalkan fungsi:

$$L_d = \sum_{i=1}^N \alpha_i \cdot$$

$$\sum_{i=1}^N a_i y_i x_i \quad \sum_{i=1}^N a_i a_j (x_i, x_j)$$

syarat : $0 \leq a_i \leq C$ dan $\sum_{i=1}^N a_i y_i = 0$

5. Menghitung nilai w dan b :
6. $w = \sum_{i=1}^N a_i y_i x_i$ $b = -1/2 w \cdot x^* + w \cdot x^*$
7. Fungsi Keputusan klasifikasi sign ($f(x)$): $f(x) = w \cdot x + b$ atau $f(x) = \sum_{i=1}^{m_i-1}$

$$a_i y_i K(x, x_i) + b$$

Keterangan :

N : Jumlah data

N : Dimensi data atau banyaknya fitur

L_d : Dualitas Lagrange Multiplier

a_i : Nilai bobot setiap titik data

C : Nilai konstanta

m : Jumlah support vector/titik data yang memiliki $a_i > 0$

$K(x, x_i)$: fungsi kernel

Keuntungan Support Vector Machine :

1. Berfungsi dengan baik pada data yang kompleks: SVM sangat cocok untuk kumpulan data yang pemisahan antarkategorinya tidak jelas. SVM dapat menangani data linear dan non-linear secara efektif.
2. Efektif dalam ruang berdimensi tinggi: SVM berkinerja baik bahkan ketika ada lebih banyak fitur (dimensi) daripada sampel, membuatnya berguna untuk tugas seperti klasifikasi teks atau pengenalan gambar.

3. Menghindari overfitting: SVM berfokus pada pencarian batas keputusan (margin) terbaik antara kelas, yang membantu mengurangi risiko overfitting, terutama pada data berdimensi tinggi.
4. Serbaguna dengan kernel: Dengan menggunakan fungsi kernel yang berbeda (seperti fungsi basis linear, polinomial, atau radial), SVM dapat beradaptasi dengan berbagai jenis data dan memecahkan masalah yang kompleks.
5. Kuat terhadap outlier: SVM kurang terpengaruh oleh outlier karena berfokus pada vektor pendukung (titik data yang paling dekat dengan margin), yang membantu dalam menciptakan model yang lebih umum.

Kekurangan Support Vector Machine:

1. Lambat dengan kumpulan data besar: SVM dapat membutuhkan banyak komputasi dan lambat untuk dilatih, terutama ketika kumpulan datanya sangat besar.
2. Sulit disetel: Memilih kernel dan parameter yang tepat (seperti C dan γ) bisa jadi rumit dan sering kali memerlukan banyak percobaan dan kesalahan.
3. Tidak cocok untuk data yang berisik: Jika kumpulan data memiliki terlalu banyak kelas yang tumpang tindih atau kebisingan, SVM mungkin kesulitan bekerja dengan baik karena mencoba menemukan pemisahan yang sempurna.
4. Sulit ditafsirkan: Tidak seperti beberapa algoritma lainnya, model SVM tidak mudah ditafsirkan atau dijelaskan, terutama saat menggunakan kernel non-linier.

5. Membutuhkan banyak memori: SVM memerlukan penyimpanan vektor pendukung, yang dapat menghabiskan banyak memori, sehingga kurang efisien untuk kumpulan data yang sangat besar.

3. Random Forest

Random forest merupakan algoritma dalam machine learning yang menggunakan banyak pohon keputusan (Decision Tree) di mana Kumpulan pohon tersebut nantinya akan dilakukan analisa. Random forest dilakukan dengan membuat sampel data latih dan membuat decision tree dari sampel kemudian proses tersebut diulang hingga jumlah pohon yang diinginkan dibangun (Nisa Sofia Amriza et al., 2021). Setiap tree di Random forest akan mengeluarkan prediksi kelas. Pada model, kandidat prediksi adalah prediksi kelas dengan jumlah suara tertinggi. Semakin banyak pohon, semakin akurat dan lebih sedikit masalah overfitting.

Algoritma Random Forest

Random Forest dijalankan dalam dua tahap. Untuk membuat Random Forest, fase pertama adalah menggabungkan sebanyak N decision tree. Fase berikutnya adalah membuat prediksi untuk setiap tree yang dibuat pada fase awal. Kelebihan random forest:

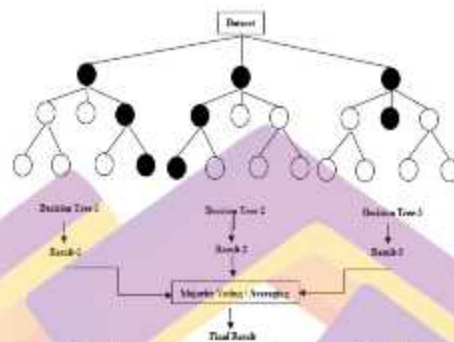
1. Mengurangi risiko overfitting: Pohon keputusan berisiko mengalami overfitting karena cenderung terlalu pas dengan semua sampel dalam data pelatihan. Namun, saat ada sejumlah besar pohon keputusan dalam hutan acak, pengklasifikasi tidak melakukan overfitting pada model karena rata-rata pohon yang tidak berkorelasi menurunkan varians keseluruhan dan kesalahan prediksi.

2. Memberikan fleksibilitas: Karena random forest dapat menangani tugas regresi dan klasifikasi dengan tingkat akurasi yang tinggi, metode ini populer di kalangan ilmuwan data. Feature bagging juga menjadikan pengklasifikasi random forest sebagai alat yang efektif untuk memperkirakan nilai yang hilang karena tetap akurat saat sebagian data hilang.
3. Mudah menentukan pentingnya fitur: Hutan acak memudahkan untuk mengevaluasi pentingnya variabel, atau kontribusi, terhadap model. Ada beberapa cara untuk mengevaluasi pentingnya fitur. Kepentingan Gini dan penurunan rata-rata ketidakmurnian (MDI) biasanya digunakan untuk mengukur seberapa besar akurasi model menurun ketika variabel tertentu dikecualikan. Namun, kepentingan permutasi, juga dikenal sebagai akurasi penurunan rata-rata (MDA), adalah ukuran penting lainnya. MDA mengidentifikasi penurunan rata-rata dalam akurasi dengan mengubah nilai fitur secara acak dalam sampel oob.

Kekurangan random forest :

1. Proses yang memakan waktu: Karena algoritma hutan acak dapat menangani kumpulan data besar, mereka dapat memberikan prediksi yang lebih akurat, tetapi dapat lambat dalam memproses data karena menghitung data untuk setiap pohon keputusan individual.
2. Memerlukan lebih banyak sumber daya: Karena hutan acak memproses kumpulan data yang lebih besar, mereka memerlukan lebih banyak sumber daya untuk menyimpan data tersebut.

3. Lebih kompleks: Prediksi pohon keputusan tunggal lebih mudah ditafsirkan dibandingkan dengan hutan pohon keputusan.



Gambar 2. 4. Random Forest Model

1. Algoritma mengambil sampel acak dari dataset yang tersedia.
2. Membangun pohon keputusan untuk setiap sampel yang dipilih. Setelah itu, hasil prediksi dari setiap pohon keputusan akan dihitung.
3. Untuk setiap hasil prediksi, dilakukan pemilihan. Modus, atau nilai yang paling sering muncul, digunakan untuk masalah klasifikasi, dan mean, atau nilai rata-rata, digunakan untuk masalah regresi.
4. Hasil prediksi yang paling banyak dipilih, atau suara terbanyak, akan dipilih oleh algoritma sebagai prediksi terakhir.

Root node merupakan simpul yang terletak paling atas, atau biasa disebut sebagai akar dari pohon keputusan. Internal node adalah simpul percabangan, dimana node ini mempunyai output minimal dua dan hanya ada satu input. Sedangkan leaf node atau terminal node merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai output. Pohon keputusan dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan

nilai information gain. Untuk menghitung nilai entropy digunakan rumus seperti pada persamaan 1, sedangkan nilai information gain menggunakan persamaan 2 (Nugroho & Emiliyawati, 2017)

$$\text{Persamaan 1: } Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Dimana Y adalah himpunan kasus dan $p(c|Y)$ merupakan proporsi nilai Y terhadap kelas c.

$$\text{Persamaan 2: } Information\ Gain(Y,a) = Entropy(Y) - \sum_{v \in Values(a)} |Y_v| / |Y_a| Entropy(Y_v) \quad (2)$$

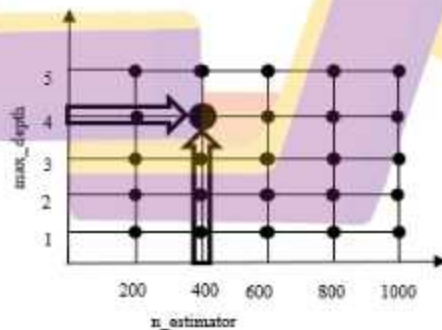
Dimana Values(a) merupakan semua nilai yang mungkin dalam himpunan kasus a. Y_v adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a. Y_a adalah semua nilai yang sesuai dengan a.

4. Optimasi Hyperparameter Grid Search

Hyperparameter adalah parameter yang nilainya ditentukan sebelum proses pelatihan model machine learning. Contoh hyperparameter termasuk parameter regularisasi (C) dan kernel pada Support Vector Machine (SVM), atau jumlah pohon (n_estimators) dan kedalaman maksimum (max_depth) pada Random Forest. Optimasi hyperparameter adalah proses mencari kombinasi hyperparameter terbaik untuk meningkatkan performa model.

Grid Search adalah salah satu teknik optimasi hyperparameter yang paling umum digunakan. Teknik ini bekerja dengan mencoba semua kombinasi hyperparameter yang telah ditentukan dalam rentang nilai tertentu. Prosesnya melibatkan:

1. Mendefinisikan Rentang Hyperparameter: Menentukan nilai atau rentang nilai yang akan diuji untuk setiap hyperparameter.
1. Contoh untuk SVM: $C = [0.1, 1, 10]$, $\text{kernel} = [\text{'linear'}, \text{'rbf'}]$, $\text{gamma} = [0.01, 0.1, 1]$.
2. Contoh untuk Random Forest: $\text{n_estimators} = [50, 100, 200]$, $\text{max_depth} = [5, 10, \text{None}]$, $\text{criterion} = [\text{'gini'}, \text{'entropy'}]$.
3. Membuat Grid Kombinasi: Grid Search akan menghasilkan semua kombinasi yang mungkin dari hyperparameter yang telah ditentukan.
4. Melatih dan Mengevaluasi Model: Untuk setiap kombinasi hyperparameter, model dilatih dan dievaluasi menggunakan metrik tertentu (misalnya, akurasi, presisi, recall, atau F1-score).
5. Memilih Kombinasi Terbaik: Kombinasi hyperparameter yang menghasilkan performa terbaik (berdasarkan metrik evaluasi) dipilih sebagai konfigurasi optimal.



Gambar 2. 5. Mekanisme kerja grid serach

Algoritma 1: Algoritma Grid Search pada Random Forest

1. Inisialisasi model Random Forest
2. Latih model awal Random Forest dengan data
3. Definisikan hyperparameter dalam sample space:

```
'n_estimators': [100, 250, 500, 750, 1000],  
'criterion': ['gini', 'entropy'],  
'max_features': [None, 'auto', 'sqrt', 'log2'],  
'max_depth': [1, 3, 5]
```

4. Model GS ← buat model (model Random Forest, sample space)
5. Latih model GS dengan data
6. Prediksi ← model GS (data uji)
7. Hitung matriks evaluasi:

Akurasi ← model GS (label uji, prediksi)

Presisi ← model GS (label uji, prediksi)

Recall ← model GS (label uji, prediksi)

8. Hasil akhir: akurasi, presisi, recall

Manfaat utama pencarian grid terletak pada sifatnya yang lengkap. Pencarian grid menjamin identifikasi kombinasi terbaik dalam ruang pencarian yang ditentukan. Metode ini khususnya bermanfaat untuk model dengan jumlah hiperparameter yang terbatas. Akan tetapi, biaya komputasinya yang tinggi menjadi kelemahan yang signifikan ketika berhadapan dengan ruang hiperparameter yang besar atau model yang kompleks.

Keuntungan grid search

1. Kelimpahan: Pencarian Grid bersifat menyeluruh dalam pencarian hiperparameter terbaik, karena menguji semua kemungkinan kombinasi. Ini memastikan bahwa dalam kumpulan nilai parameter yang ditentukan, kombinasi terbaik akan ditemukan.
2. Kesederhanaan: Metode ini mudah dipahami dan diterapkan. Prosesnya mudah, tanpa perlu penyesuaian yang rumit atau pengetahuan tingkat lanjut tentang perilaku model.
3. Reprodusibilitas: Karena Grid Search menguji kombinasi parameter yang tetap, hasilnya dapat direproduksi. Ini berarti peneliti atau pengembang lain dapat membuat ulang eksperimen yang sama dan mendapatkan hasil yang sama.

Kekurangan grid search

1. Biaya Komputasi: Kelemahan utama Grid Search adalah biaya komputasinya yang tinggi. Menguji semua kemungkinan kombinasi parameter dapat memakan banyak waktu, terutama untuk kumpulan data besar dan model yang kompleks.
2. Pencarian Grid: Pencarian Grid mungkin tidak menemukan kecocokan "terbaik" jika tidak disertakan dalam grid yang ditentukan. Kualitas hasil bergantung pada seberapa baik grid didefinisikan.
3. Penurunan Efisiensi dalam Ruang Berdimensi Tinggi: Seiring dengan meningkatnya jumlah hiperparameter, Pencarian Grid menjadi kurang efisien karena "kutukan dimensionalitas". Ini berarti bahwa jumlah kombinasi tumbuh secara eksponensial, membuat proses tersebut tidak praktis untuk banyak hiperparameter.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan penelitian terapan (applied research) yang berfokus pada penyelesaian masalah praktis, yaitu prediksi stunting menggunakan metode komputasi. Sifat penelitian ini adalah deskriptif-komparatif dan eksplanatif:

1. **Deskriptif-Komparatif:** Penelitian ini mendeskripsikan karakteristik dataset stunting dan membandingkan kinerja dua algoritma, yaitu Support Vector Machine (SVM) dan Random Forest.
2. **Eksplanatif:** Penelitian menganalisis faktor-faktor yang memengaruhi perbedaan kinerja antara kedua algoritma tersebut, seperti efek hiperparameter dan sensitivitas terhadap data yang tidak seimbang.

Pendekatan yang digunakan adalah eksperimen komputasi, di mana variabel independen (hiperparameter seperti C , γ , $n_estimator$) dimanipulasi untuk melihat pengaruhnya terhadap variabel dependen (akurasi prediksi).

3.2. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini menggunakan data sekunder yang diperoleh dari repositori publik.

1. **Sumber Data:** Dataset diambil dari Kaggle, sebuah repositori dataset publik yang kredibel <https://www.kaggle.com/datasets/mutashimrobbani/stuntingg>.

2. Karakteristik Data: Dataset awal berisi 1.287 baris entri data stunting anak berlabel dengan 15 kolom atribut.
3. Variabel Data:
 1. Fitur: Meliputi jenis kelamin, tinggi badan, berat badan, lingkaran lengan (LILA), usia, jenis kelamin, berat badan terhadap usia, tinggi badan terhadap usia, dan berat badan terhadap tinggi badan.
 2. Target: Status gizi dengan kategori seperti Gizi Kurang, Gizi Buruk, Gizi Baik, Gizi Berlebih, Risiko Gizi Berlebih, dan Obesitas.

3.3. Metode Analisis Data

Analisis data dilakukan menggunakan pendekatan Machine Learning dengan bantuan bahasa pemrograman Python (v3.12.6) dan pustaka Scikit-learn. Proses analisis meliputi:

1. Pra-pemrosesan Data (Data Preprocessing):
 1. Cleaning: Menghapus fitur tidak relevan (seperti Nama, Tanggal Lahir, Nama Ortu, Desa/Kel, RT, RW) dan menghapus 63 data duplikat.
 2. Imputasi: Mengisi nilai yang hilang pada kolom 'LILA' menggunakan nilai rata-rata (mean).
 3. Encoding: Mengubah variabel kategorikal (Status Gizi dan Jenis Kelamin) menjadi numerik menggunakan LabelEncoder.
 4. Scaling: Melakukan standarisasi fitur numerik menggunakan StandardScaler khusus untuk model SVM agar memiliki rerata nol dan varians satu.

1. Pengumpulan Data (Data Collection)

Proses dimulai dengan memasukkan data mentah yang diperoleh dari repositori publik Kaggle: <https://www.kaggle.com/datasets/mutashimrobbani/stunting> seperti ditunjukkan pada table 3.1.

Tabel 3. 1. Dataset

Na ma	J K	Us ia	Tgl Lahir	Ortu	Desa / Kel	R T	R W	Ber at	Tin ggi	LIL A	ZS BB /U	ZS TB /U	ZS BB/ TB	Stat us
B 1	P	3	04/09/ 2020	aji d	palebo n	1	2	8,3	79, 5	0	1,6 2	0,1 6	2,1 5	GK
B 2	P	4	18/11/ 2019	sarto no	palebo n	4	1	9,7	85, 6	0	1,8 3	0,7 8	2,0 2	GK
B 3	L	4	20/06/ 2019	tom my	palebo n	1	2	11	92, 5	14 ,4	1,8 1	0,2 1	2,6 2	GK
B 4	L	2	08/10/ 2021	suba kti g	palebo n	7	1	5,7	57		0,9 4	2,1 5	1,2 3	RKL
B 5	P	4	18/07/ 2019	aldy p.	palebo n	1	2	8,6	78, 4	13 ,5	3,2 6	3,6 7	1,4 7	GB A
...
B 128 7	L	1	23/08/ 2022	lis h	plamo ngan sari	2	1 2	7,9	68	0	1,2 1	2,0 1	-0,1	GB A

2. Pra-pemrosesan Data (Data Preprocessing)

Data mentah kemudian masuk ke tahap pra-pemrosesan untuk menjamin kualitas data:

1. Penanganan Missing & Duplicate Value: Dilakukan pemeriksaan dan penanganan terhadap nilai yang hilang (missing values) dengan menggunakan teknik pengisian nilai rata-rata (mean) menggunakan fungsi

fillna(data['...'].mean()) serta penghapusan data ganda (duplicate) menggunakan fungsi *drop_duplicates* untuk mencegah bias pada model.

2. Encoding dan Standarisasi: Tahap selanjutnya adalah One Hot Encoding dan Variable Standardization. Variabel kategorikal (seperti Status Gizi) diubah menjadi numerik menggunakan fungsi *LabelEncoder*. Khusus untuk model SVM, fitur numerik distandarisasi dengan fungsi *StandardScaler* agar memiliki skala yang seragam.

3. Reduksi Fitur (Reduction fitur)

Setelah data bersih, dilakukan Reduction fitur dengan menggunakan fungsi *data.drop* untuk memilah atribut yang paling relevan. Atribut yang tidak berkontribusi signifikan terhadap prediksi, seperti Nama, Tanggal Lahir, Nama Ortu, Desa/Kel, RT, RW dihapus pada tahap ini.

4. Pembagian Data (Data Splitting) Data yang telah diseleksi kemudian melalui proses Decision 80:20 Split. Dataset dibagi secara berstrata menjadi dua bagian: 80% sebagai Training Dataset untuk melatih model dan 20% sebagai data uji untuk validasi menggunakan fungsi *train_test_split*.

5. Pemodelan dan Validasi (Modeling & Validation)

Tahap ini berjalan secara paralel menggunakan data latih.

1. Pelatihan Model: Training Dataset digunakan dalam proses Model Training. Pada tahap ini, input data 80% ke dalam algoritma Support Vector Machine (SVM) dan Random Forest dilatih.

2. Seleksi Model: Dilakukan Select Model SVM dengan menerapkan optimasi Grid Search pada table 3.2 untuk mencari kombinasi hiperparameter terbaik (seperti nilai C, gamma, kernel) guna meningkatkan akurasi.

Tabel 3. 2. Parameter Terbaik SVM with Grid Search

Parameters	Value	Tipe Data	Skala / Rentang Nilai
C	1.0	Float	Umumnya 1–100
kernel	'linear'	String	'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'
degree	3	Integer	>0 (Umumnya 2 - 5)
gamma	'scale'	String/Float	'scale', 'auto', atau Float > 0
coef0	0.0	Float	Bebas (Umumnya 0.0)
shrinking	TRUE	Boolean	True atau False
probability	FALSE	Boolean	True atau False
tol	0.001	Float	> 0 (Default 1×10^{-3})
cache_size	200	Float	0.0 hingga ∞
class_weight	None	Dict / String	None atau 'balanced'
verbose	FALSE	Boolean	True atau False
max_iter	-1	Integer	-1 (Tanpa batas) atau Integer > 0
decision_function_shape	'ovr'	String	0.0 hingga ∞ (Non-negatif)
break_ties	FALSE	Boolean	True atau False
random_state	None	Integer / None	0 hingga $2^{32} - 1$ atau None

Dilakukan Select Model RF dengan menerapkan optimasi Grid Search pada table 3.3 untuk mencari kombinasi hiperparameter terbaik (seperti nilai `n_estimators`, `max_depth`, dan `min_samples_split`) guna meningkatkan akurasi.

Tabel 3. 3. Parameter Terbaik RF with Grid Search

Parameters	Value	Type Data	Skala / Rentang Nilai
n_estimators	100	Integer	Umumnya 1-100
criterion	'gini'	String	'gini', 'entropy', atau 'log_loss'
max_depth	None	Integer / None	1 - ∞ (tanpa batas) atau None
min_samples_split	2	Int / Float	Integer ≥ 2 atau Float [0, 1] (persentase)
min_samples_leaf	1	Int / Float	Integer 1 atau Float [0, 0.5]
min_weight_fraction_leaf	0.0	Float	0.0 hingga 0.5
max_features	'sqrt'	String/Int/Float	'sqrt', 'log2', None, Integer, atau Float [0, 1]
max_leaf_nodes	None	Integer / None	2 hingga ∞ atau None
min_impurity_decrease	0.0	Float	0.0 hingga ∞
bootstrap	TRUE	Boolean	True atau False
oob_score	FALSE	Boolean	True atau False
n_jobs	None	Integer-	-1 atau Integer ≥ 1 atau None
random_state	0	Integer	0 hingga $2^{(32)} - 1$ (untuk konsistensi)
verbose	0	Integer	≥ 0
warm_start	FALSE	Boolean	True atau False
class_weight	None	Dict / String	None, 'balanced', atau 'balanced_subsample'
ccp_alpha	0.0	Float	0.0 hingga ∞ (Non-negatif)
max_samples	None	Int / Float	None, Integer, atau Float [0, 1]
monotonic_cst	None	Array-like	None atau array berisi [-1, 0, 1]

- Validasi: Model yang sudah dilatih diuji menggunakan data 20% tadi dengan fungsi Y_{pred} untuk melihat kemampuannya memprediksi mengukur keandalannya terhadap data baru.

4. Evaluasi dan Komparasi (Evaluation & Comparison)

Hasil validasi dari kedua model dibawa ke tahap Compare Model Performance.

Kinerja SVM dan Random Forest dibandingkan berdasarkan metrik evaluasi seperti Akurasi, Presisi, Recall, F1-Score, dan ROC-AUC untuk menentukan model mana yang lebih unggul.

5. **Prediksi Akhir (Prediction)** Tahap terakhir adalah Predictive Stunting, dimana model terbaik digunakan untuk mengklasifikasikan data anak baru ke dalam kategori status gizi (Normal atau Stunting) sebagai output akhir sistem.
6. **Taxonomi** Menjelaskan struktur metodologi untuk sebuah penelitian yang membandingkan performa algoritma Machine Learning dalam memprediksi stunting pada anak pada gambar 3.2.

7. **Judul dan Topik Utama**

Research Title penelitian berfokus pada perbandingan performa antara dua algoritma, yaitu Support Vector Machine (SVM) dan Random Forest, yang dioptimalkan menggunakan Grid Search (dalam teks tertulis GA, namun labelnya menyebut Grid Search) untuk prediksi stunting pada anak. Research Topic penelitian ini berada di bawah payung besar Machine Learning.

8. **Klasifikasi Bidang (Sub Issue - Application & Type)**

Taksonomi ini membedakan berbagai aplikasi Machine Learning (seperti pengenalan suara, klasifikasi gambar, dan NLP), namun penelitian ini secara spesifik masuk ke dalam kategori Prediction (Prediksi) yang menjadi fokus utama penelitian. Algorithms di antara pilihan seperti KNN, Decision Tree, dan Gradient Boosting, penelitian ini secara khusus menyoroti SVM dan Random Forest.

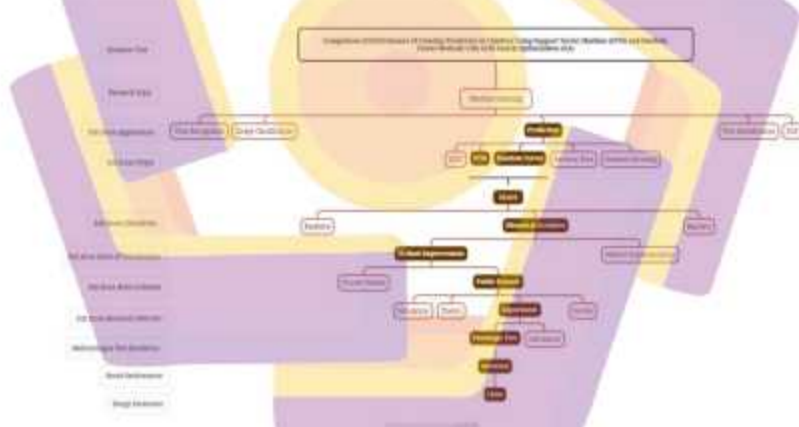
9. **Karakteristik Data dan Kontribusi (Objectives & Methodology)**

Model & Historical Archive penelitian ini menggunakan data historis (arsip), bukan data real-time atau big data. Method Improvement penelitian ini bertujuan untuk pengembangan metode, yang berarti tidak sekadar

menerapkan algoritma standar, tetapi melakukan optimasi (seperti Grid Search yang disebutkan di judul). Public Dataset Data yang digunakan berasal dari dataset publik, bukan data privat.

10. Alur Metodologi

Eksperimen menunjukkan langkah-langkah teknis pengujian dengan melakukan eksperimen terkontrol (bukan sekadar teori atau survei). Prototype Test menguji model yang telah dikembangkan. Accuracy parameter merupakan keberhasilan utama yang diukur adalah tingkat akurasi model. Yang terakhir class menjadi output akhir adalah klasifikasi (mungkin kategori stunting atau tidak stunting).



Gambar 3. 2. Taxonomi Penelitian

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

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.

10.1. Tinjauan Data dan Pra-pemrosesan

1. Tinjauan

Penelitian tersebut menggunakan dataset publik stunting dari website Kaggle link dataset <https://www.kaggle.com/datasets/mutashimrobbani/stunting> dengan jumlah 1288 entri data. Pengumpulan dataset stunting (variable independen: jenis kelamin, tinggi badan, berat badan, lingkar lengan (LILA), usia, jenis kelamin, berat badan terhadap usia, tinggi badan terhadap usia, dan berat badan terhadap tinggi badan) dan (variable dependen: status_gizi).

2. Pra-pemrosesan



Gambar 4. 1. Persentase Data Status Gizi

Dataset awal berjumlah dari 1.287 entri data. Variabel target pada gambar 4.1 menunjukkan distribusi kelas berikut: Gizi Kurang (563), Gizi Baik (453), Gizi Lebih (101), Gizi Buruk (78), Resiko Gizi Lebih (55), dan Obesitas (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.

Tabel 4. 1. Missing Values

No	JK	Status_gizi
1	JK	0
2	Usia	0
3	Berat	0
4	Tinggi	0
5	LiLA	199
6	Berat_Per_Umur	0
7	Tinggi_Per_Umur	0
8	Berat_Per_Tinggi	0
9	Status_Gizi	0

Tabel 4.1. proses awal mengecek apakah ada data yang kosong atau hilang di setiap kolom. Hasilnya menunjukkan terdapat 199 data yang kosong (missing values) pada kolom LiLA (Lingkar Lengan Atas). Data kosong tersebut diisi dengan nilai rata-rata (mean) pada kolom LiLA.

Tabel 4. 2. Duplicate Values

No	Data Before	Duplicate Value	Data After
1	1287	63	1243

Tahap ini mengecek dan menghapus baris data yang memiliki nilai persis sama (duplikat). Awalnya terdeteksi pada tabel 4.2. ada 63 data duplikat, kemudian dihapus dan data yang tinggal berjumlah 1224 baris data

Tabel 4. 3. Tabel Fitur Reduksi Data

J K	Usi a	Bera t	Ting gi	Lil A	Berat_Per_U mur	Tinggi_Per_U mur	Berat_Per_Tin ggi	Status_Gi zi
P	3	8.3	79.5	0.0	-1.62	-0.16	-2.15	Gizi Kurang
P	4	9.7	85.6	0.0	-1.83	-0.78	-2.02	Gizi Kurang
L	4	11.0	92.5	14. 4	-1.81	-0.21	-2.62	Gizi Kurang
L	2	5.7	57.0	Na N	-0.94	-2.15	1.23	Resiko Gizi Lebih
P	4	8.6	78.4	13. 5	-3.26	-3.67	-1.47	Gizi Baik

Proses berikut penghapusan (reduksi) beberapa kolom yang tidak relevan menggunakan perintah `data.drop`, kolom seperti 'Nama', 'Nama Ortu', 'Desa/Kel', 'Tgl Lahir', 'RT', dan 'RW' dibuang dari dataset karena tidak memiliki nilai prediktif untuk status gizi, sehingga menyisakan data inti untuk dianalisis tabel 4.3.

Tabel 4. 4. One Hot Encoding

No	Nama	Usia_0	Usia_1	Usia_2	Usia_3	Usia_4	Usia_5	Usia_6
1	Balita 1	0	0	0	1	0	0	0
2	Balita 2	0	0	0	0	1	0	0
3	Balita 3	0	0	0	0	1	0	0
4	Balita 4	0	0	1	0	0	0	0
5	Balita 5	0	0	0	0	1	0	0

Dalam penelitian ini salah satu faktor (fitur) yang sangat krusial. Namun, algoritma komputer seperti SVM dan Random Forest bekerja menggunakan perhitungan matematika. Jika kita membiarkan kolom usia berisi teks atau kategori angka biasa (seperti 0, 1, 2, ...), algoritma mungkin salah mengartikan bahwa usia

7 "lebih bernilai" daripada usia 0 secara matematis. One-Hot Encoding menyelesaikan masalah ini dengan mengubah kategori usia menjadi kolom biner (TRUE/FALSE atau 1/0) sehingga setiap kelompok usia (misalnya Usia_0 hingga Usia_7) dianggap setara secara bobot di awal proses tabel 4.4. Memungkinkan algoritma untuk melihat apakah pola stunting lebih sering terjadi pada kategori usia spesifik (misal: usia 2 tahun) tanpa terganggu oleh urutan angka.

```
[23]: scaler = StandardScaler()
[24]: scaler.fit(X)
[24]: StandardScaler
      Parameters
```

Gambar 4. 2. Standarisasi Data

Pada gambar 4.2, akan mengubah distribusi data agar memiliki rata-rata 0 dan varians 1, dan menghitung nilai rata-rata dan standar deviasi dari setiap kolom (fitur) yang ada di dalam dataset. Nilai-nilai ini akan disimpan di dalam objek scaler untuk digunakan pada tahap transformasi berikutnya.



Gambar 4. 3. Heatmap

Hubungan Sangat Kuat (Usia, Berat, Tinggi) terhadap Berat & Tinggi (0.88) Ini adalah korelasi tertinggi terlihat pada gambar 4.3. Semakin tinggi seorang anak, biasanya berat badannya juga bertambah. Usia & Tinggi (0.83) dan Usia & Berat (0.79) hubungan ini sangat kuat dan positif, menunjukkan pertumbuhan yang normal seiring bertambahnya usia. Hubungan Menengah terhadap Berat_Per_Tinggi & Berat_Per_Umur (0.63) Ada hubungan moderat di sini, yang berarti anak yang berat badannya di atas rata-rata untuk tingginya cenderung juga berat untuk usianya. Hubungan Lemah atau Negatif terhadap LiLA (Lingkar Lengan Atas): Variabel ini memiliki korelasi yang sangat rendah (bahkan negatif) dengan hampir semua variabel lainnya (seperti -0.11 dengan Usia).

Tabel 4. 5. Pemisahan Fitur dan Target

No	Data	Fitur (X)	Target (Y)
1	1224	7	1

Variabel X yang berisi semua data kecuali kolom 'Status_Gizi'. Dalam Machine Learning, X disebut sebagai Fitur (variabel independen/prediktor). Kolom 'Status_Gizi' dan menyimpannya di variabel Y. Y disebut sebagai Target (variabel dependen/label) yang ingin diprediksi oleh model tabel 4.5.

Tabel 4. 6. Validasi Data

No	Best Parameter	Value
1	max_depth	None
2	min_Sample_Split	5
3	n_estimator	50

Nilai yang ingin kita uji untuk model Random Forest, yaitu n_estimators atau jumlah pohon (50,100,200). Max_depth atau kedalaman maksimum setiap pohon (None, 10,20). Min_samples_split atau jumlah minimum sampel yang diperlukan

untuk membagi simpul internal (2,5,10). GridSearchCV akan mencoba setiap kemungkinan kombinasi dari param_grid yang kamu buat. Jika ada 3 nilai untuk setiap parameter (3x3x3), maka ada 27 kombinasi berbeda. Karena setiap kombinasi divalidasi 5 kali (cv=5), total ada 135 kali pelatihan yang dilakukan secara otomatis. Sehingga hasil validasi menemukan parameter terbaik (max_depth: None, min_samples_split: 5, n_estimators: 50) tabel 4.6.

4.2. Kinerja Support Vector Machine (SVM)

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 4.7. menyajikan laporan klasifikasi terperinci untuk model SVM setelah optimasi Grid Search.

Tabel 4. 7. 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
10	Akurasi SVM			0.9836734693877551	98.37%

4.3. Kinerja Random Forest (RF)

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 f1-score 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), 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 98,78% 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 0,9987 dan AUC rata-rata makro sebesar 0,9988. Tabel 4.8 menyajikan laporan klasifikasi terperinci untuk model Random Forest setelah optimasi Grid Search.

Tabel 4.8. Laporan Klasifikasi Random Forest Setelah Optimasi Grid Search

No	Kelas	Presisi	Recall	F1-Score	Support
1	Gizi Baik	1.00	0.96	0.98	76
2	Gizi Buruk	1.00	1.00	1.00	13
3	Gizi Kurang	0.97	1.00	0.99	115
4	Gizi Lebih	1.00	1.00	1.00	22
5	Obesitas	1.00	1.00	1.00	7
6	Resiko Gizi Lebih	1.00	1.00	1.00	12
7	Akurasi			0.99	245
8	Macro Avg	1.00	0.99	0.99	245
9	Weighted Avg	0.99	0.99	0.99	245
10	Akurasi Random Forest			0.9877551020408162	98.78%

4.4. 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 4.9 menyajikan skor kepentingan fitur untuk model

Tabel 4. 9. 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

4.5. Skor AUC ROC per Kelas Random Forest

Tabel 4. 10. Laporan Skor ROC AUC tanpa GS

No	Kelas	ROC AUC Score
1	Gizi Baik	0.9852
2	Gizi Buruk	0.9978
3	Gizi Kurang	0.9908
4	Gizi Lebih	1.0000
5	Obesitas	1.0000
6	Resiko Gizi Lebih	0.9957
7	Micro-average	0.9902

Tabel 4. 11. Laporan Skor ROC AUC dengan GS

No	Kelas	ROC AUC Score
1	Gizi Baik	0.9968
2	Gizi Buruk	1.0000
3	Gizi Kurang	0.9963
4	Gizi Lebih	1.0000
5	Obesitas	1.0000
6	Resiko Gizi Lebih	1.0000
7	Micro-average	0.9988
8	Macro-average	0.9987

Table 4.10. menunjukkan skor ROC AUC terhadap model RF tanpa Grid Search, sedangkan skor ROC AUC terhadap model RF tanpa Grid Search tabel 4.11 terlihat mengalami peningkatan signifikan. Tabel 4.4. menyajikan skor ROC AUC untuk setiap kelas status gizi, serta rata-rata mikro dan makro. 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.

4.6. 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 4.4. terhadap salah satu anak dari dataset diketahui jenis kelamin=Perempuan, usia = 5 bulan, berat = 8,3kg, tinggi = 84,4cm, lingkaran lengan = 12,3cm, berat per umur =-3,88 kg/bln, tinggi per umur =-2,28 cm/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 4. 4. Deployment Model

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

1. 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 4. 12. Analisis Signifikansi SVM

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.

2. Signifikansi Praktis pada Random Forest (RF)

Model Random Forest sudah mencapai kinerja sangat tinggi pada tahap awal (Akurasi 98.78%). 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 4.13., 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 4. 13. Analisis Signifikansi RF

No	Metrik	RF Awal (Tanpa GS)	RF Sesudah GS	Perubahan	Implikasi Klinis (Signifikansi Praktis)
1	Akurasi	98.78%	98.78%	0.00%	Sudah mencapai batas performa set data.
2	Micro-Avg ROC AUC	0.9902	0.9987	+0.0085	Peningkatan substansial pada kemampuan diskriminatif. Nilai 0.9987 menunjukkan model hampir sempurna dalam membedakan berbagai kategori status gizi, menegaskan kemampuan diskriminatif model yang kuat di bawah konfigurasi parameter yang disetel (misalnya, <code>n_estimators: 200</code>).

Substansial pada kemampuan diskriminatif. Nilai 0.9987 pada gambar 4.10. menunjukkan model hampir sempurna dalam membedakan berbagai kategori status gizi, menegaskan kemampuan diskriminatif model yang kuat di bawah konfigurasi parameter yang disetel.

2.8. Analisis Keunggulan Konsisten Random Forest (RF)

1. Ketahanan Inheren terhadap Ketidakseimbangan Data (Kinerja Awal)

RF yang merupakan model ensemble berbasis pohon keputusan, secara inheren lebih tangguh (*robust*) terhadap masalah ketidakseimbangan kelas yang signifikan pada data stunting. 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.

2. Kemampuan Diskriminatif yang Jauh Lebih Baik (Metrik ROC AUC)

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.

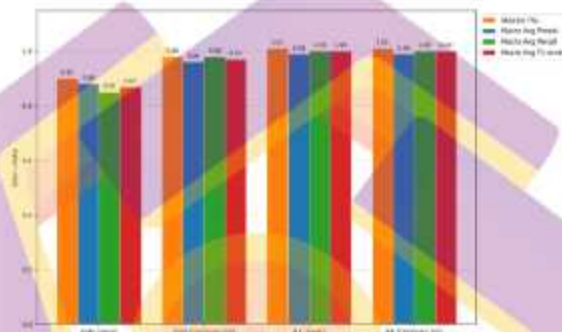
3. Keunggulan Interpretasi Klinis (Feature Importance)

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.

3.9. Pembahasan

Optimasi pada SVM gambar 4.14 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. 5. 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 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 (98,78%) 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. 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 4. 14. 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	98.78	1.00	0.99	0.99	0.9902
4	RF	Sesudah GS	98.78	1.00	0.99	0.99	0.9987

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.

BAB V

PENUTUP

5.1. 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 98,78% dan AUC mikro-rata-rata 0,9987 setelah optimasi. Meskipun akurasi Random Forest tidak meningkat drastis setelah optimasi, proses Grid Search 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.

5.2. Saran

1. Bagi penelitian selanjutnya, disarankan untuk mengembangkan penggunaan metode Support Vector Machine (SVM) untuk mengklasifikasikan lebih dari dua kelas (multi-class) agar kategori status gizi dapat dipetakan secara lebih detail.
2. Peneliti berikutnya dapat melakukan eksplorasi lebih lanjut dengan mencoba kombinasi nilai parameter (C dan γ) serta jenis kernel lain untuk menguji konsistensi akurasi pada dataset yang berbeda atau lebih luas.
3. Mengingat kompleksitas penentuan hyperplane pada banyak variabel prediktor, disarankan penggunaan teknik optimasi parameter yang lebih otomatis untuk meningkatkan efisiensi proses klasifikasi.

DAFTAR PUSTAKA

- As, M., & Subhiyakto, E. R. (2025). *Performance Comparison of Random Forest, SVM, and XGBoost Algorithms with SMOTE for Stunting Prediction*. 9(4), 1163–1169.
<https://doi.org/https://doi.org/10.30871/jaic.v9i4.9701>
- Brown, S. (2021). *Machine learning, explained*. MIT Management Sloan School.
<https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>
- Chilyabanyama, O. N., Chilengi, R., Simuyandi, M., Chisenga, C. C., Chirwa, M., Hamusonde, K., Saroj, R. K., Iqbal, N. T., Ngaruye, J., & Bosomprah, S. (2022). Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia. *Children*, 9(7).
<https://doi.org/10.3390/children9071082>
- C Scheffler, M. Hermanussen, B. Bogin, D. S. Liana, F. Taolin, & P. M. V. P. Cempaka. (2019). *Stunting is not a synonym of malnutrition*.
<https://doi.org/https://doi.org/10.1038/s41430-019-0439-4>
- de Onis, M., & Branca, F. (2016). Childhood stunting: A global perspective. In *Maternal and Child Nutrition* (Vol. 12, pp. 12–26). Blackwell Publishing Ltd.
<https://doi.org/10.1111/mcn.12231>
- Desai, V., Cottrell, J., & Sowerby, L. (2023). No longer a blank cheque: a narrative scoping review of physician awareness of cost. *Public Health*, 223, 15–23.
<https://doi.org/10.1016/j.puhe.2023.07.009>

- El-Alfy, E.-S. M., & Mohammed, S. A. (2020). A review of machine learning for big data analytics: bibliometric approach. *Technology Analysis & Strategic Management*, 32(8), 984–1005. <https://doi.org/10.1080/09537325.2020.1732912>
- Fathirachman Mahing, N., Lazuardi Gunawan, A., Foresta Azhar Zen, A., Abdurrachman Bachtiar, F., & Agung Wicaksono, S. (2023). Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(7), 1527–1536. <https://doi.org/10.25126/jtiik.1078010>
- Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *Nuansa Informatika*, 16(1), 141–149. <https://doi.org/10.25134/nuansa.v16i1.5406>
- Grantham-Mcgregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Child development in developing countries I Developmental potential in the first 5 years for children in developing countries. In *Lancet* (Vol. 369). <https://core.ac.uk/download/pdf/13103225.pdf>
- Hamed, A., Hegab, A., & Roshdy, E. (2020). Prevalence and factors associated with stunting among school children in Egypt. *Eastern Mediterranean Health Journal*, 26(7), 787–793. <https://doi.org/10.26719/emhj.20.047>
- Haris, M. S., Khudori, A. N., & Kusuma, W. T. (2022). Perbandingan Metode Supervised Machine Learning untuk Prediksi Prevalensi Stunting di Provinsi Jawa Timur. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(7), 1571. <https://doi.org/10.25126/jtiik.2022976744>

- Indrisari, E., Febiansyah, H., & Adiwinoto, B. (2025). A Systematic Literature Review on the Application of Machine Learning for Predicting Stunting Prevalence in Indonesia (2020–2024). In *Jurnal Sisfokom (Sistem Informasi dan Komputer)* (Vol. 14, Number 3, pp. 277–283). <https://doi.org/10.32736/sisfokom.v14i3.2366>
- Mondon, C., Tan, P. Y., Chan, C. L., Tran, T. N., & Gong, Y. Y. (2024). Prevalence, determinants, intervention strategies and current gaps in addressing childhood malnutrition in Vietnam: a systematic review. *BMC Public Health*, 24(1). <https://doi.org/10.1186/s12889-024-18419-8>
- Mubarok, A. H., Pujiono, P., Setiawan, D., Wicaksono, D. F., & Rimawati, E. (2025). Parameter Testing on Random Forest Algorithm for Stunting Prediction. *Sinkron*, 9(1), 107–116. <https://doi.org/10.33395/sinkron.v9i1.14264>
- Müller, A. C., & Guido, S. (n.d.). *Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python*.
- Nisa Sofia Amriza, R., Supriyadi, D., Jl Panjaitan No, P. DI, Purwokerto Selatan, K., Banyumas, K., & Tengah, J. (2021). Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media. *Jurnal JUPITER*, 13(2), 130–139.
- Nugroho, Y. S., & Emiliyawati, N. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. *Jurnal Teknik Elektro*, 9(1), 24–29.
- Prasetyo, E., & Nugroho, K. (2024). 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. In *Februari* (Vol. 23, Number 1).
<https://doi.org/https://doi.org/10.62411/tc.v23i1.9779>
- Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLoS ONE*, *16*(6 June 2021), 1–11.
<https://doi.org/10.1371/journal.pone.0253172>
- Rahmi, I., Susanti, M., Yozza, H., & Wulandari, F. (2022). CLASSIFICATION OF STUNTING IN CHILDREN UNDER FIVE YEARS IN PADANG CITY USING SUPPORT VECTOR MACHINE. *Barekeng*, *16*(3), 771–778.
<https://doi.org/10.30598/barekengvol16iss3pp771-778>
- Shen, H., Zhao, H., & Jiang, Y. (2023). Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children*, *10*(10).
<https://doi.org/10.3390/children10101638>
- Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, *19*(1).
<https://doi.org/10.1186/s12874-019-0681-4>
- SSGI. (2021). *Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi dan Kabupaten/Kota Tahun 2021*. https://dinkes.acehprov.go.id/1-content/uploads/Hasil_SSGI_Tahun_2021_Tingkat_Kabupaten_Kota.pdf
- WHO. (2025). *World Health Organization*.
https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-stunting-prevalence?utm_source=chatgpt.com

Zemariam, A. B., Abate, B. B., Alamaw, A. W., Lake, E. S., Yilak, G., Ayele, M., Tilahun, B. D., & Ngusie, H. S. (2025). Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms. *PLoS ONE*, 20(1 January), 1–27. <https://doi.org/10.1371/journal.pone.0316452>

