

**IMPLEMENTASI PENDEKATAN MACHINE LEARNING
MENGUNAKAN HYPERPARAMETER TUNING UNTUK
MENDETEKSI PENYAKIT STUNTING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AKMAL RAFI FADHILLAH

21.11.4414

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**IMPLEMENTASI PENDEKATAN MACHINE LEARNING
MENGUNAKAN HYPERPARAMETER TUNING UNTUK
MENDETEKSI PENYAKIT STUNTING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AKMAL RAFI FADHILLAH

21.11.4414

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI PENDEKATAN MACHINE LEARNING
MENGUNAKAN HYPERPARAMETER TUNING UNTUK
MENDETEKSI PENYAKIT STUNTING**

yang disusun dan diajukan oleh

Akmal Rafi Fadhillah

21.11.4414

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Maret 2025

Dosen Pembimbing,



Robert Marco S.T., M.T., Ph.D.
NIK. 190302228

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI PENDEKATAN MACHINE LEARNING
MENGUNAKAN HYPERPARAMETER TUNING UNTUK
MENDETEKSI PENYAKIT STUNTING

yang disusun dan diajukan oleh

Akmal Rafi Fadhillah

21.11.4414

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Maret 2025

Susunan Dewan Penguji

Nama Penguji

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228

Tanda Tangan

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Maret 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Akmal Rafi Fadhillah

NIM : 21.11.4414

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI PENDEKATAN MACHINE LEARNING MENGUNAKAN HYPERPARAMETER TUNING UNTUK MENDETEKSI PENYAKIT STUNTING

Dosen Pembimbing: **Robert Marco, S.T., M.T., Ph.D.**

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, 12 Maret 2025

Yang Menyatakan,



METERAI
TEMPEL
57AMX227456ZB4

Akmal Rafi Fadhillah

KATA PENGANTAR

Alhamdulillah Rabbil 'Alamin. Puji Syukur penulis panjatkan kehadiran Allah SWT atas segala Rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul "IMPLEMENTASI PENDEKATAN MACHINE LEARNING MENGGUNAKAN HYPERPARAMETER TUNING UNTUK MENDETEKSI PENYAKIT STUNTING" ini diajukan untuk memenuhi salah satu syarat dalam menyelesaikan studi S1 di Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Dalam proses penyusunan skripsi ini, penulis banyak menerima bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. **Bapak Prof. M. Suyanto, M.M.**, selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan kesempatan kepada penulis untuk menuntut ilmu di universitas ini.
2. **Bapak Hanif Al Fatta, M.Kom., Ph.D.**, selaku Dekan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, atas segala dukungan dan fasilitas yang diberikan selama masa studi.
3. **Bapak Robert Marco, ST., MT., Ph.D.**, selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, serta motivasi selama proses penyusunan skripsi ini.
4. **Kedua orang tua tercinta**, yang selalu memberikan dukungan moral, doa, dan kasih sayang sepanjang perjalanan pendidikan penulis.

Yogyakarta, 14 Februari 2025

Penulis

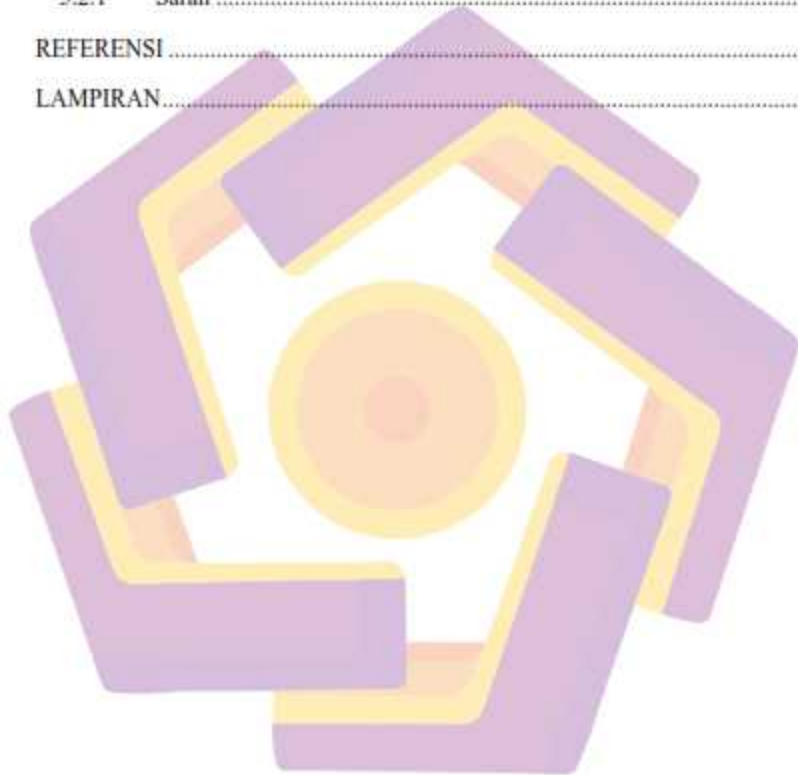
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH.....	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	17

2.2.1	<i>Stunting</i>	17
2.2.2	Data Mining	17
2.2.3	<i>Machine Learning</i>	17
2.2.4	<i>Hyperparameter Tuning</i>	18
2.2.5	<i>Grid search</i>	18
2.2.6	<i>Cross Validation</i>	18
2.2.7	<i>Label Encoding</i>	19
2.2.8	Normalisasi Data	19
2.2.9	Ketidakseimbangan Data	20
2.2.10	SMOTE	20
2.2.11	Klasifikasi	21
2.2.12	<i>K-Nearest Neighbor</i>	21
2.2.13	<i>Random Forest</i>	22
2.2.14	<i>Decision Tree</i>	23
2.2.15	Evaluasi Performa	24
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian	26
3.2	Alur Penelitian	26
3.2.1	Pengumpulan Data	27
3.2.2	<i>Preprocessing Data</i>	28
3.2.3	<i>Splitting Data</i>	28
3.2.4	Pelatihan Model	29
3.2.5	Parameter <i>Default</i> dan Parameter <i>Grid Search</i>	29
3.2.6	Evaluasi Performa	29
3.3	Alat dan Bahan	30

3.3.1	Data Penelitian	30
3.3.2	Tool dan Library	31
3.3.3	Hardware	32
3.3.4	Software	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Dataset	33
4.2	Preprocessing Data	33
4.2.1	Dropping Feature	34
4.2.2	Label Encoding	34
4.2.3	Normalisasi Data	35
4.2.4	Handling Imbalance Data	36
4.2.5	Feature Correlation Matrix	38
4.3	Splitting Data	39
4.4	Hyperparameter Tuning dan Pemodelan	39
4.4.1	K-Nearest Neighbor Sebelum Tuning	40
4.4.2	K-Nearest Neighbor Setelah Tuning	40
4.4.3	Random Forest Sebelum Tuning	41
4.4.4	Random Forest Setelah Tuning	41
4.4.5	Decision Tree Sebelum Tuning	42
4.4.6	Decision Tree Setelah Tuning	43
4.5	Evaluasi Performa	43
4.5.1	K-Nearest Neighbor Sebelum Tuning	44
4.5.2	K-Nearest Neighbor Setelah Tuning	45
4.5.3	Random Forest Sebelum Tuning	47
4.5.4	Random Forest Setelah Tuning	49

4.5.5	<i>Decision Tree Sebelum Tuning</i>	51
4.5.6	<i>Decision Tree Setelah Tuning</i>	52
BAB V PENUTUP		58
5.1	Kesimpulan	58
5.2.1	Saran	58
REFERENSI		59
LAMPIRAN.....		64



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	10
Tabel 3.1 Deskripsi Dataset	30
Tabel 4.1 Dataset	33
Tabel 4.2 Dataset Setelah Proses <i>Label Encoding</i>	34
Tabel 4.3 Komposisi <i>Splitting Data</i>	39
Tabel 4.4 Parameter <i>Grid Search K-Nearest Neighbor</i>	40
Tabel 4.5 Parameter <i>Grid Search Random Forest</i>	42
Tabel 4.6 Parameter <i>Grid Search Decision Tree</i>	43
Tabel 4.7 Hasil <i>K-Nearest Neighbor</i> dengan Parameter Default	45
Tabel 4.8 <i>Hyperparameter Optimal K-Nearest Neighbor</i>	45
Tabel 4.9 Hasil <i>K-Nearest Neighbor</i> dengan Parameter <i>Grid search</i>	47
Tabel 4.10 Hasil <i>Random Forest</i> dengan Parameter Default	48
Tabel 4.11 <i>Hyperparameter Optimal Random Forest</i>	49
Tabel 4.12 Hasil <i>Random Forest</i> dengan Parameter <i>Grid Search</i>	50
Tabel 4.13 Hasil <i>Decision Tree</i> dengan Parameter Default	52
Tabel 4.14 <i>Hyperparameter Optimal Decision Tree</i>	53
Tabel 4.15 Hasil <i>Decision Tree</i> dengan Parameter <i>Grid Search</i>	54
Tabel 4.16 Perbandingan Hasil Performa Parameter Default dan <i>Grid Search</i>	56

DAFTAR GAMBAR

Gambar 2.1 Distribusi Nilai Aktual	25
Gambar 3.1 Alur Penelitian	27
Gambar 4.1 Proses <i>Dropping Feature</i>	34
Gambar 4.2 Dataset Sebelum melalui Proses Normalisasi Data	35
Gambar 4.3 Dataset Setelah melalui Proses Normalisasi Data	36
Gambar 4.4 Distribusi Data Sebelum Penerapan SMOTE	36
Gambar 4.5 Distribusi Data Setelah Penerapan SMOTE	37
Gambar 4.6 <i>Heatmap</i> Korelasi Antar Fitur	38
Gambar 4.7 Confusion Matrix <i>K-Nearest Neighbor</i> Parameter Default	44
Gambar 4.8 Confusion Matrix <i>K-Nearest Neighbor</i> Parameter <i>Grid Search</i>	46
Gambar 4.9 Confusion Matrix <i>Random Forest</i> Parameter Default	48
Gambar 4.10 Confusion Matrix <i>Random Forest</i> Parameter <i>Grid Search</i>	50
Gambar 4.11 Confusion Matrix <i>Decision Tree</i> Parameter Default	51
Gambar 4.12 Confusion Matrix <i>Decision Tree</i> dengan Parameter <i>Grid Search</i>	54
Gambar 4.13 Perbandingan Akurasi Antar Algoritma	55

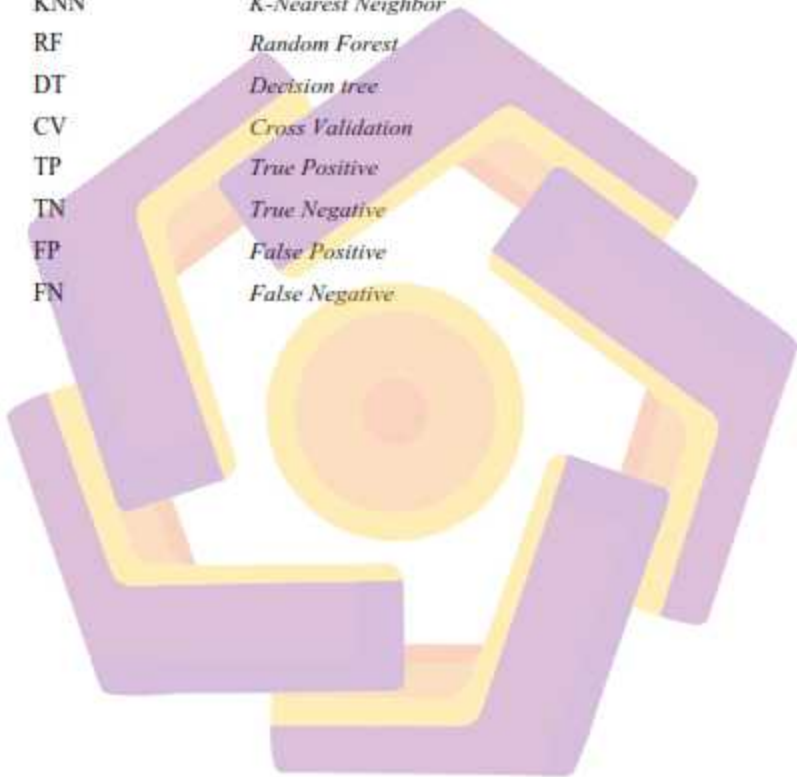


DAFTAR LAMPIRAN

Lampiran 1. Import <i>library</i>	64
Lampiran 2. Membaca dataset	64
Lampiran 3. Menghitung jumlah sampel	64
Lampiran 4. Melihat kategori <i>stunting</i> yang unik	64
Lampiran 5. Visualisasi distribusi fitur dengan <i>pairplot</i>	64
Lampiran 6. Mengkonversi label kategori menjadi numerik	64
Lampiran 7. Memisahkan fitur dan label	64
Lampiran 8. Normalisasi skala data	64
Lampiran 9. Membagi data menjadi data latih dan data uji	65
Lampiran 10. Menangani ketidakseimbangan data dengan SMOTE	65
Lampiran 11. Menampilkan statistik deskriptif setiap fitur	65
Lampiran 12. Proses penerapan KNN dengan parameter <i>default</i>	65
Lampiran 13. Proses penerapan KNN dengan <i>grid search</i>	65
Lampiran 14. Proses penerapan <i>Random Forest</i> dengan parameter <i>default</i>	66
Lampiran 15. Proses penerapan <i>Random Forest</i> dengan <i>grid search</i>	66
Lampiran 16. Proses penerapan <i>Decision Tree</i> dengan parameter <i>default</i>	66
Lampiran 17. Proses penerapan <i>Decision Tree</i> dengan <i>grid search</i>	67
Lampiran 18. Perbandingan akurasi antar algoritma	67

DAFTAR LAMBANG DAN SINGKATAN

SMOTE	<i>Synthetic Minority Oversampling Technique</i>
KDD	<i>Knowledge Discovery in Databases</i>
ML	<i>Machine learning</i>
KNN	<i>K-Nearest Neighbor</i>
RF	<i>Random Forest</i>
DT	<i>Decision tree</i>
CV	<i>Cross Validation</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>



DAFTAR ISTILAH



<i>Machine learning</i>	Kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa melalui pemrograman secara langsung
SMOTE	Metode resampling untuk menangani ketidakseimbangan data dengan cara menambahkan sampel pada kelas minoritas
<i>K-Nearest Neighbor</i>	Tetangga terdekat yang menentukan kelas berdasarkan mayoritas label dalam sejumlah K tetangga terdekatnya
<i>Euclidean Distance</i>	Metode pengukuran jarak yang digunakan dalam algoritma <i>K-Nearest Neighbor</i> untuk menghitung kedekatan antar data
<i>Random Forest</i>	Metode <i>ensemble learning</i> yang terdiri dari banyak pohon keputusan untuk meningkatkan akurasi prediksi
<i>Decision tree</i>	Struktur pohon yang digunakan dalam model prediksi, dimana setiap node mempresentasikan keputusan atau pembagian data
<i>Hyperparameter tuning</i>	Teknik optimasi parameter model <i>machine learning</i> untuk meningkatkan kinerja dan akurasi model
<i>Grid Search</i>	Teknik mencari kombinasi <i>hyperparameter</i> terbaik melalui pencarian sistematis
<i>Cross Validation</i>	Teknik evaluasi model yang membagi dataset untuk menguji dan melatih model

INTISARI

Stunting adalah bagian dari malnutrisi kronis yang menghambat pertumbuhan pada bayi dan balita, serta dapat menimbulkan gangguan hematologi, yang berdampak negatif pada perkembangan fisik dan kognitif anak. Menjadi sebuah tantangan bagi tenaga kesehatan untuk memperkenalkan diagnosis dini penderita penyakit *stunting* agar dapat dilakukan intervensi secara efektif. Dalam penelitian ini, digunakan dataset yang mencakup indikator kesehatan fisik anak, seperti usia, berat lahir, berat badan, panjang lahir, panjang badan dan pemberian ASI secara eksklusif. Penelitian ini mengeksplorasi penerapan algoritma *K-Nearest Neighbor*, *Random Forest* dan *Decision Tree* dalam klasifikasi *stunting* berdasarkan data antropometri dan faktor risiko lainnya. Ketiga algoritma tersebut dipilih karena kemampuannya dalam mengidentifikasi pola klasifikasi dari data numerik yang kompleks. Penelitian ini menerapkan berbagai teknik pemrosesan data, seperti *label encoding*, normalisasi data dan penggunaan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian. Penerapan dari beberapa model tersebut dikombinasikan menggunakan parameter *default* dan dioptimasi menggunakan teknik *grid search* untuk memperoleh hasil prediksi yang lebih akurat. Evaluasi dilakukan berdasarkan metrik *accuracy*, *precision*, *recall* dan *f1-score*. Hasil penelitian menunjukkan bahwa penerapan *grid search* dalam optimasi model mampu meningkatkan kinerja prediksi status *stunting* secara lebih optimal dibandingkan dengan penggunaan parameter *default*. Penelitian ini memberikan kontribusi dalam bidang perkembangan ilmu pengetahuan tentang kesehatan, terutama dalam upaya pencegahan dan penanganan *stunting* melalui pendekatan berbasis *machine learning*.

Kata kunci: *stunting*, balita, klasifikasi, prediksi, pembelajaran mesin

ABSTRACT

Stunting is part of chronic malnutrition which inhibits growth in infants and toddlers, and can cause hematological disorders which have a negative impact on children's physical and cognitive development. It is a challenge for health workers to carry out early diagnosis of stunting sufferers so that effective intervention can be carried out. In this study, a dataset was used that included indicators of children's physical health such as age, birth weight, birth weight, birth length, body length and exclusive breastfeeding. This research explores the application of the K-Nearest Neighbor, Random Forest and Decision tree algorithms in stunting classification based on anthropometric data and other risk factors. These three algorithms were chosen because of their ability to identify classification patterns from complex numerical data. This research applies various data processing techniques, such as label coding, data normalization and the use of Synthetic Minority Over-sampling Technique (SMOTE) to overcome data imbalances. Next, the dataset is divided into training data and test data with a proportion of 70% for training and 30% for testing. The application of several models is combined using default parameters and optimized using grid search techniques to obtain more accurate prediction results. Evaluation is carried out based on accuracy, precision, recall and f1-score metrics. The research results show that the application of grid search in model optimization is able to improve the performance of stunting status predictions more optimally compared to using default parameters. This research contributes to the development of health science, especially in efforts to prevent and treat stunting through a machine learning-based approach.

Keywords: stunting, toddlers, classification, prediction, machine learning