

**ANALISIS PERBANDINGAN METODE KNN DAN RANDOM
FOREST MENGGUNAKAN HYPERPARAMETER TUNING
UNTUK DETEKSI PENDERITA STUNTING**

SKRIPSI

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



disusun oleh

AYU FATIMAH

22.11.4770

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN METODE KNN DAN RANDOM
FOREST MENGGUNAKAN HYPERPARAMETER TUNING
UNTUK DETEKSI PENDERITA STUNTING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AYU FATIMAH

22.11.4770

Kepada

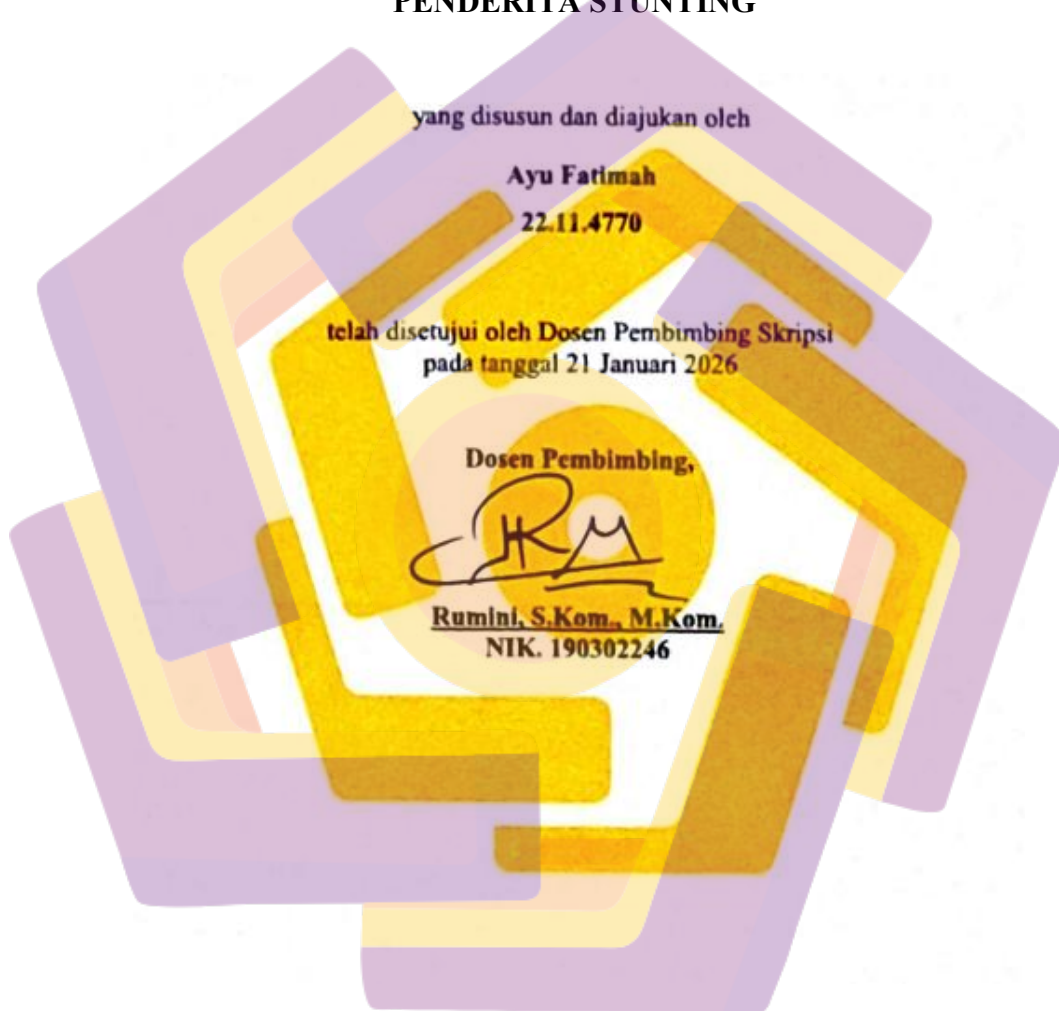
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN METODE KNN DAN RANDOM FOREST
MENGUNAKAN HYPERPARAMETER TUNING UNTUK DETEKSI
PENDERITA STUNTING**



HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN METODE KNN DAN RANDOM FOREST
MENGUNAKAN HYPERPARAMETER TUNING UNTUK DETEKSI
PENDERITA STUNTING**

yang disusun dan diajukan oleh
Ayu Fatimah
22.11.4770
Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
<u>Anna Balta, S.Kom., M.Kom.</u> NIK. 190302290	
<u>Dina Maulina, S.Kom., M.Kom.</u> NIK. 190302250	
<u>Rumini, S.Kom., M.Kom.</u> NIK. 190302246	

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 January 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ayu Fatimah
NIM : 22.11.4770

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS PERBANDINGAN METODE KNN DAN RANDOM FOREST
MENGUNAKAN HYPERPARAMETER TUNING UNTUK DETEKSI
PENDERITA STUNTING**

Dosen Pembimbing : Rumini, S.Kom., 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, 21 Januari 2026

Yang Menyatakan,



Ayu Fatimah

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas limpahan rahmat, hidayah, dan kemudahan yang diberikan, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Penelitian berjudul “ANALISIS PERBANDINGAN METODE KNN DAN RANDOM FOREST MENGGUNAKAN HYPERPARAMETER TUNING UNTUK DETEKSI PENDERITA STUNTING” ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Proses penyusunan skripsi ini tidak lepas dari bantuan dan dukungan banyak pihak. Oleh karena itu, penulis menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta, atas kesempatan dan lingkungan akademik yang kondusif bagi penulis dalam menempuh pendidikan.
2. Bapak Prof. Dr. Kusrini, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer, atas dukungan, fasilitas, dan pelayanan akademik yang mempermudah proses studi penulis.
3. Ibu Rumini, S.Kom., M.Kom., selaku pembimbing skripsi, atas arahan, masukan, dan bimbingan ilmiah yang sangat berarti selama proses penelitian berlangsung.
4. Kedua orang tua tercinta, atas doa, dukungan moral, dan pengorbanan yang menjadi sumber kekuatan bagi penulis.
5. Penulis juga berterima kasih kepada seseorang yang senantiasa hadir memberi perhatian dan dukungan selama proses ini berlangsung.

Yogyakarta, 21 Januari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	18
BAB III METODE PENELITIAN	25

3.1	Alur Penelitian	25
3.1.1	Pengumpulan Dataset.....	26
3.1.2	<i>Preprocessing</i> Data.....	26
3.1.3	<i>Splitting</i> Data	27
3.1.4	Pelatihan Model	28
3.1.5	Parameter <i>Default</i> dan Parameter <i>Grid Search</i>	40
3.1.6	Evaluasi Performa.....	41
3.2	Alat dan Bahan.....	41
3.2.1	Data Penelitian	42
3.2.2	<i>Tools</i> dan <i>Library</i>	43
3.2.3	<i>Hardware</i>	43
3.2.4	<i>Software</i>	44
BAB IV HASIL DAN PEMBAHASAN		45
4.1	Dataset.....	45
4.2	<i>Preprocessing</i> Data.....	46
4.2.1	<i>Dropping Feature</i>	46
4.2.2	<i>Label Encoding</i>	47
4.2.3	Normalisasi Data.....	48
4.2.4	<i>Feature Correlation Matrix</i>	49
4.3	<i>Split</i> Data.....	50
4.4	<i>Hyperparameter Tuning</i> dan Pemodelan	51
4.4.1	<i>K-Nearest Neighbor</i> Sebelum <i>Tuning</i>	51
4.4.2	<i>K-Nearest Neighbor</i> Setelah <i>Tuning</i>	52
4.4.3	<i>Random Forest</i> Sebelum <i>Tuning</i>	53
4.4.4	<i>Random Forest</i> Setelah <i>Tuning</i>	53

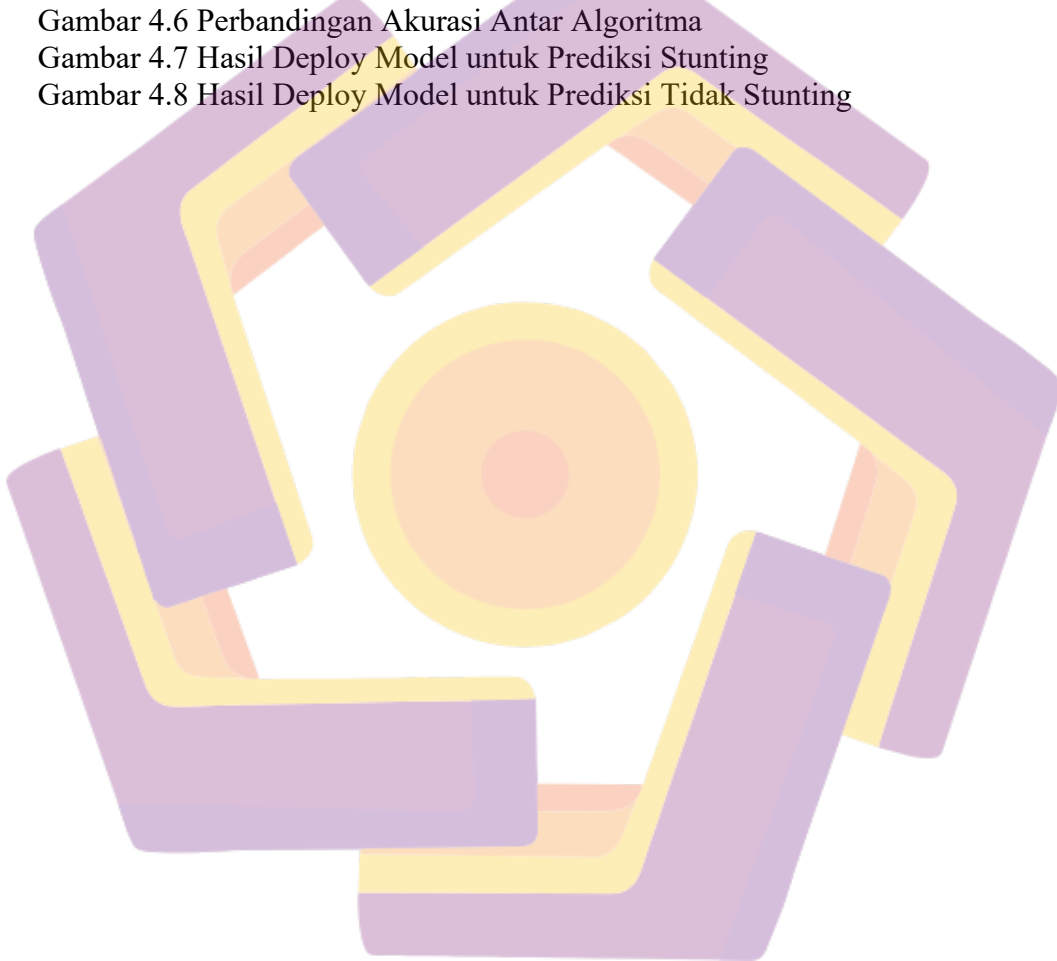
4.5	Evaluasi Performa	54
4.5.1	<i>K-Nearest Neighbor</i> Sebelum <i>Tuning</i>	55
4.5.2	<i>K-Nearest Neighbor</i> Setelah <i>Tuning</i>	57
4.5.3	<i>Random Forest</i> Sebelum <i>Tuning</i>	58
4.5.4	<i>Random Forest</i> Setelah <i>Tuning</i>	59
4.5.5	Hasil Deployment Prediksi Stunting.....	64
BAB V PENUTUP		67
5.1	Kesimpulan	67
5.2	Saran	68
REFERENSI		69
LAMPIRAN.....		73

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	12
Tabel 2.2. Distribusi Nilai Aktual	23
Tabel 3.1 Data Training	28
Tabel 3.2 Data Testing	29
Tabel 3.3 Hasil Jarak Euclidean	31
Tabel 3.4 Analisis Tetangga Terdekat dan Stunting	31
Tabel 3.5 Data Training	33
Tabel 3.6 Data Testing	33
Tabel 3.7 Pengambilan Kelompok Breastfeeding 1	34
Tabel 3.8 Pengambilan Kelompok Breastfeeding 0	34
Tabel 3.9 Uji threshold	35
Tabel 3.10 Perbandingan Informasi Gain	37
Tabel 3.11 Deskripsi Dataset	42
Tabel 4.1 Dataset	45
Tabel 4.2 Dataset Setelah Proses Label Encoding	47
Tabel 4.3 Data sebelum di Normalisasi	48
Tabel 4.4 Data sebelum di Normalisasi	48
Tabel 4.5 Komposisi Splitting Data	50
Tabel 4.6 Parameter Grid Search K-Nearest Neighbor	52
Tabel 4.7 Parameter Grid Search Random Forest	54
Tabel 4.8 Hasil K-Nearest Neighbor dengan Parameter Default	56
Tabel 4.9 Hasil K-Nearest Neighbor dengan Parameter Grid Search	57
Tabel 4.10 Hasil Random Forest dengan Parameter Default	59
Tabel 4.11 Hyperparameter Optimal Random Forest	60
Tabel 4.12 Hasil Random Forest dengan Parameter Grid Search	62
Tabel 4.13 Perbandingan Hasil Performa Parameter Default dan Grid Search	63

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	25
Gambar 3.2 Pohon Keputusan Random Forest	39
Gambar 4.1 Proses Dropping Feature	46
Gambar 4.2 Heatmap Korelasi Antar Fitur	49
Gambar 4.3 Confusion Matrix K-Nearest Neighbor Parameter Default	55
Gambar 4.4 Confusion Matrix Random Forest Parameter Default	58
Gambar 4.5 Confusion Matrix Random Forest Parameter Grid Search	61
Gambar 4.6 Perbandingan Akurasi Antar Algoritma	63
Gambar 4.7 Hasil Deploy Model untuk Prediksi Stunting	65
Gambar 4.8 Hasil Deploy Model untuk Prediksi Tidak Stunting	66

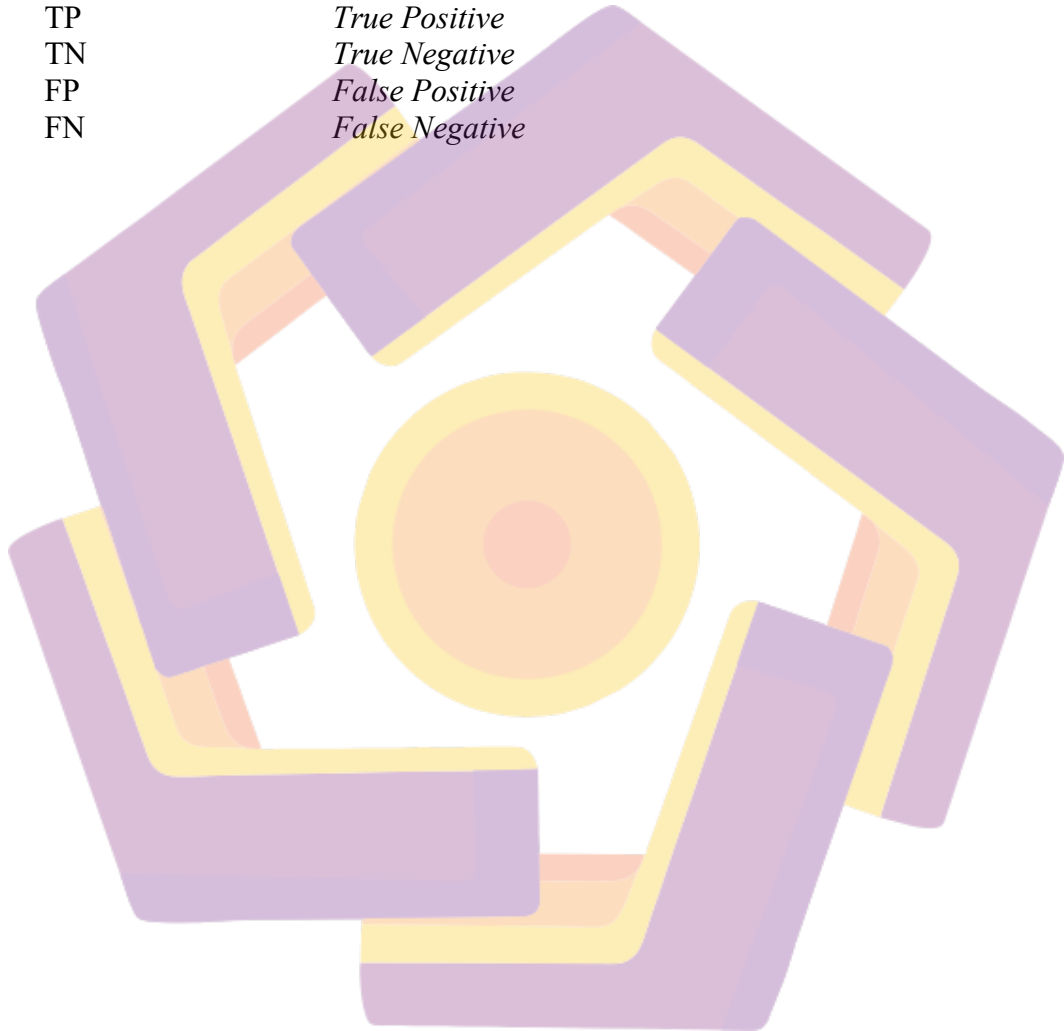


DAFTAR LAMPIRAN

Lampiran 1. Import pustaka untuk preprocessing, analisis, dan pemodelan	73
Lampiran 2. Pendefinisian nama file sumber dan keluaran dataset	73
Lampiran 3. Memuat dataset dari file CSV dan menyimpan ke format Excel	73
Lampiran 4. Membaca dataset Excel dan menampilkan 10 baris awal	73
Lampiran 5. Menghitung jumlah kategori pada kolom <i>Stunting</i>	73
Lampiran 6. Menampilkan nilai unik pada kolom <i>Stunting</i>	74
Lampiran 7. Mengonversi label kategorikal menjadi numerik	74
Lampiran 8. Menghitung korelasi fitur dan menampilkan heatmap	74
Lampiran 9. Memisahkan fitur (X) dan label (y)	74
Lampiran 10. Normalisasi data menggunakan MinMaxScaler	75
Lampiran 11. Membagi data menjadi data latih dan data uji	75
Lampiran 12. Menampilkan statistik deskriptif setiap fitur	75
Lampiran 13. Penerapan model KNN dengan parameter default	76
Lampiran 14. Penerapan model KNN dengan Grid Search	76
Lampiran 15. Penerapan model Random Forest dengan parameter default	77
Lampiran 16. Penerapan model Random Forest dengan Grid Search	77
Lampiran 17. Perbandingan akurasi antar algoritma	78

DAFTAR LAMBANG DAN SINGKATAN

KDD	<i>Knowledge Discovery in Databases</i>
ML	<i>Machine learning</i>
KNN	<i>K-Nearest Neighbor</i>
RF	<i>Random Forest</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>Stunting</i>	Kondisi gagal tumbuh akibat kekurangan gizi kronis.
<i>Machine Learning (ML)</i>	Algoritma yang belajar dari data untuk membuat prediksi.
<i>Data Mining</i>	Proses penemuan pola atau informasi berharga dari data besar.
<i>Random Forest (RF)</i>	Metode <i>ensemble</i> (gabungan banyak pohon keputusan) untuk klasifikasi.
<i>K-Nearest Neighbor (KNN)</i>	Algoritma klasifikasi berdasarkan jarak kedekatan ke sejumlah K tetangga terdekat.
<i>Grid Search</i>	Teknik optimasi untuk mencari kombinasi hyperparameter terbaik secara sistematis.
<i>Hyperparameter Tuning</i>	Proses penyesuaian parameter untuk meningkatkan kinerja model.
<i>Cross Validation (CV)</i>	Teknik evaluasi model dengan membagi data berulang kali untuk pengujian kinerja.
<i>Label Encoding</i>	Proses konversi data kategorikal menjadi nilai numerik.
Klasifikasi	Proses prediksi label kelas data baru berdasarkan model yang telah dilatih.
<i>True Positive (TP)</i>	Positif sejati: Data positif diprediksi benar sebagai positif.
<i>False Positive (FP)</i>	Positif palsu: Data negatif salah diprediksi sebagai positif.
<i>Majority Voting</i>	Hasil prediksi di mana kelas yang mendapat suara terbanyak dari beberapa model atau beberapa tetangga (pada KNN) akan menjadi prediksi akhir.

INTISARI

Stunting merupakan masalah kesehatan masyarakat yang serius dan memerlukan sistem deteksi dini yang efektif berbasis data. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma Machine Learning, yaitu *K-Nearest Neighbor* (KNN) dan *Random Forest* (RF), dalam mengklasifikasikan status stunting pada balita. Dataset yang digunakan melibatkan berbagai fitur prediktor seperti berat badan lahir dan panjang lahir. Untuk memaksimalkan kinerja model dan mengurangi potensi overfitting, kedua metode dioptimasi menggunakan teknik *Hyperparameter Tuning* dengan *Grid Search*. Berdasarkan evaluasi menggunakan *5-Fold Cross Validation*, algoritma *Random Forest* yang telah melalui proses tuning menunjukkan kinerja terbaik dengan nilai recall sebesar 98,13%, yang mencerminkan kemampuan model dalam mendeteksi hampir seluruh kasus balita yang mengalami stunting. Selain itu, *Random Forest* juga menghasilkan nilai precision sebesar 98,46% dan accuracy sebesar 98,46%, sementara algoritma *K-Nearest Neighbor* setelah tuning memperoleh nilai accuracy sebesar 97,92%. Recall dijadikan metrik utama karena dihitung secara khusus pada kelas stunting, sedangkan accuracy, precision, dan F1-score digunakan sebagai metrik pendukung untuk menggambarkan kinerja model secara keseluruhan pada kedua kelas. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode *ensemble* yang teroptimasi, khususnya *Random Forest* dengan *Grid Search*, merupakan pendekatan yang paling efektif dan andal dalam mendeteksi kasus *stunting* pada balita, sehingga direkomendasikan untuk digunakan sebagai sistem pendukung deteksi dini *stunting* dalam konteks kesehatan masyarakat.

Kata Kunci: *Stunting, Random Forest, K-Nearest Neighbor, Klasifikasi, Hyperparameter Tuning.*

ABSTRACT

Stunting is a serious public health problem that requires an effective data-driven early detection system. This study aims to compare the performance of two Machine Learning algorithms, namely K-Nearest Neighbor (KNN) and Random Forest (RF), in classifying stunting status in children. The dataset used involves various predictive features such as birth weight and birth length. To improve model performance and reduce the risk of overfitting, both algorithms were optimized using Hyperparameter Tuning with the Grid Search method. Based on the evaluation using 5-Fold Cross Validation, the Random Forest algorithm after tuning demonstrated the best performance, achieving a recall value of 98.13%, which reflects the model's ability to detect nearly all cases of children experiencing stunting. In addition, Random Forest achieved a precision of 98.46% and an accuracy of 98.46%, while the K-Nearest Neighbor algorithm after tuning achieved an accuracy of 97.92%. Recall was selected as the primary evaluation metric because it is calculated specifically for the stunting class, whereas accuracy, precision, and F1-score were used as supporting metrics to describe the overall model performance across both classes. Based on these results, it can be concluded that optimized ensemble methods, particularly Random Forest with Grid Search, represent the most effective and reliable approach for detecting stunting in children, and are therefore recommended for use as a decision-support system for early stunting detection in the public health context.

Keywords: *Stunting, Random Forest, K-Nearest Neighbor, Classification, Hyperparameter Tuning.*