

**KLASIFIKASI STATUS STUNTING PADA BALITA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN)****Ervina Kartika Sari<sup>1</sup>, Novisca Indriani<sup>2</sup>, Nurul Hidayah<sup>3</sup>, Muh. Rasyid Ridha<sup>4</sup>**<sup>1-4</sup>Prodi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri,  
Email: [ervinakartikasari75@gmail.com](mailto:ervinakartikasari75@gmail.com)<sup>1</sup>, [indrianinovisca@gmail.com](mailto:indrianinovisca@gmail.com)<sup>2</sup>, [hidayahnurul46@hmail.com](mailto:hidayahnurul46@hmail.com)<sup>3</sup>,  
[rasyid4sky@gmail.com](mailto:rasyid4sky@gmail.com)<sup>4</sup>**ABSTRAK**

Stunting merupakan kondisi gangguan pertumbuhan pada balita yang disebabkan oleh kekurangan gizi kronis dalam jangka waktu yang panjang. Kondisi ini masih menjadi salah satu permasalahan kesehatan masyarakat yang memerlukan perhatian karena dapat memengaruhi pertumbuhan fisik, perkembangan kognitif, serta kualitas hidup anak di masa mendatang. Dalam pelaksanaannya, proses pemantauan status gizi balita di lapangan masih banyak dilakukan secara manual melalui pencatatan pada grafik pertumbuhan. Cara tersebut berpotensi menimbulkan kesalahan pencatatan maupun keterlambatan dalam proses identifikasi status gizi. Penelitian ini bertujuan menerapkan metode K-Nearest Neighbor (KNN) untuk mengklasifikasikan status stunting pada balita berdasarkan data antropometri. Tahapan penelitian meliputi pengumpulan data, preprocessing, pembagian data latih dan data uji, serta pengujian beberapa nilai parameter k untuk memperoleh hasil klasifikasi yang sesuai. Melalui metode KNN, status gizi balita ditentukan berdasarkan tingkat kemiripan data dengan data balita yang telah diketahui status gizinya. Hasil penelitian menunjukkan bahwa algoritma KNN dapat digunakan untuk mengklasifikasikan status stunting berdasarkan karakteristik antropometri balita. Model yang dihasilkan kemudian diimplementasikan ke dalam aplikasi berbasis web untuk membantu petugas kesehatan dan kader posyandu dalam melakukan identifikasi awal status stunting. Sistem yang dibangun juga membantu proses pencatatan dan pengolahan data sehingga dapat mengurangi kemungkinan kesalahan dalam pendataan.

**Kata Kunci:** Stunting, Balita, K-Nearest Neighbor (KNN), Klasifikasi, Status Gizi.

**ABSTRACT**

*Stunting is a growth disorder in toddlers caused by long-term chronic malnutrition. This condition remains a public health problem that requires attention because it can affect physical growth, cognitive development, and children's future quality of life. In practice, the process of monitoring toddler nutritional status in the field is still largely done manually through recording on growth charts. This method has the potential to cause recording errors and delays in the process of identifying nutritional status. This study aims to apply the K-Nearest Neighbor (KNN) method to classify stunting status in toddlers based on anthropometric data. The research stages include data collection, preprocessing, dividing training and test data, and testing several k parameter values to obtain appropriate classification results. Using the KNN method, toddler nutritional status is determined based on the level of data similarity with data on toddlers whose nutritional status is already known. The results show that the KNN algorithm can be used to classify stunting status based on toddler anthropometric characteristics. The resulting model is then implemented into a web-based application to assist health workers and integrated health post (Posyandu) cadres in early identification of stunting status. The developed system also facilitates the data recording and processing process, thereby reducing the possibility of errors in data collection.*

**Keywords:** Stunting, Toddlers, K-Nearest Neighbor (KNN), Classification, Nutritional Status.

## 1 PENDAHULUAN

Stunting merupakan salah satu masalah gizi yang masih menjadi perhatian di berbagai negara karena dapat memengaruhi pertumbuhan fisik, perkembangan kognitif, serta kesehatan anak dalam jangka panjang [1], [2]. Di Indonesia, berbagai upaya dilakukan untuk menurunkan angka stunting melalui program perbaikan gizi dan pemantauan pertumbuhan balita secara berkelanjutan [3]. Salah satu bentuk pemantauan tersebut dilakukan melalui pengukuran berat badan dan tinggi badan balita di Posyandu maupun Puskesmas [4], [5]. Namun, proses penentuan status gizi masih sering dilakukan secara manual sehingga berpotensi menimbulkan kesalahan pencatatan, interpretasi data, serta keterlambatan dalam proses pelaporan dan tindak lanjut [6], [7], [8].

Perkembangan teknologi informasi mendorong pemanfaatan machine learning dalam bidang kesehatan untuk membantu proses klasifikasi dan pengambilan keputusan berbasis data [9]. Salah satu metode yang banyak digunakan adalah K-Nearest Neighbor (KNN), yaitu algoritma yang menentukan kelas suatu data berdasarkan kedekatannya dengan data lain yang telah diketahui kelasnya [10]. KNN dinilai sesuai untuk data numerik seperti umur, berat badan, tinggi badan, dan jenis kelamin yang umum digunakan dalam penilaian status gizi balita [11], [12]. Selain mudah diterapkan, metode ini juga memiliki kemampuan yang cukup baik dalam proses klasifikasi berdasarkan karakteristik data yang tersedia.

Berbagai penelitian telah menerapkan KNN maupun algoritma machine learning lainnya untuk mengidentifikasi status gizi dan stunting pada balita. Reza dan Widyaningsih [13] menggunakan KNN dan Naïve Bayes untuk memprediksi kualitas tumbuh kembang balita. Penelitian lain menerapkan KNN, Random Forest, serta Naïve Bayes untuk klasifikasi dan prediksi stunting pada berbagai wilayah penelitian [14], [15], [19], [25]. Selain itu, beberapa studi membandingkan berbagai metode supervised machine learning dalam memprediksi status gizi anak dan balita [16], [17], [18]. Pengembangan metode juga dilakukan melalui kombinasi KNN dengan Extreme Gradient Boosting, Boosted KNN, maupun Hybrid KNN-MLP untuk meningkatkan performa klasifikasi [24], [26], [35].

Penelitian terkait stunting tidak hanya dilakukan di Indonesia. Kajian literatur menunjukkan bahwa machine learning memiliki potensi besar dalam mendukung upaya pencegahan stunting dan pencapaian tujuan pembangunan kesehatan berkelanjutan [29]. Pendekatan serupa juga diterapkan untuk memprediksi faktor risiko malnutrisi pada anak [30], meningkatkan akurasi diagnosis malnutrisi menggunakan modifikasi KNN [31], serta melakukan prediksi dini tingkat malnutrisi melalui metode ensemble learning [32]. Di sisi lain, algoritma seperti Support Vector Machine, Decision Tree, Logistic Regression, dan XGBoost juga telah digunakan dalam klasifikasi stunting dengan hasil yang cukup baik [21], [22], [23], [27], [28], [33]. Penelitian di Rwanda dan Bangladesh turut menunjukkan bahwa analisis data berbasis machine learning dapat membantu memahami faktor-faktor yang berpengaruh terhadap stunting dan malnutrisi anak [34], [36].

Meskipun berbagai metode telah digunakan, KNN masih menjadi salah satu algoritma yang banyak dipilih karena proses perhitungannya relatif sederhana dan mudah diterapkan. Namun, hasil klasifikasi KNN sangat dipengaruhi oleh nilai  $k$  yang digunakan dalam menentukan tetangga terdekat [10], [11]. Oleh karena itu, penelitian ini menerapkan metode K-Nearest Neighbor pada data antropometri balita dengan melakukan pengujian beberapa nilai  $k$  untuk memperoleh model klasifikasi yang sesuai. Model yang dihasilkan kemudian diimplementasikan dalam aplikasi berbasis web sehingga dapat membantu petugas kesehatan dan kader Posyandu dalam melakukan identifikasi awal status stunting secara lebih cepat dan mudah [3], [4], [12].

## 2 METODE PENELITIAN

Bab ini menguraikan langkah-langkah sistematis yang dilakukan dalam penelitian untuk mencapai tujuan klasifikasi status stunting pada balita. Penjelasan mencakup metodologi penelitian, sumber data yang digunakan, serta variabel-variabel penelitian yang menjadi dasar dalam pembentukan model *machine learning* berbasis algoritma K-Nearest Neighbor (KNN).

**2.1 Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen *machine learning* untuk melakukan klasifikasi status stunting pada balita. Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan dataset, pembersihan data (*preprocessing*), pembagian data latih dan data uji, pemodelan dengan algoritma *K-Nearest Neighbor (KNN)*, hingga evaluasi performa model menggunakan *Confusion Matrix*.

**2.2 Sumber Data dan Variabel Penelitian**

Dalam pengembangan model ini, pemilihan data yang tepat menjadi kunci utama keberhasilan klasifikasi. Berikut adalah penjelasan mengenai asal-usul data yang digunakan serta variabel apa saja yang dilibatkan untuk membentuk model yang handal dan representatif.

**2.2.1 Sumber Data**

Dataset yang digunakan dalam penelitian ini merupakan data sekunder antropometri anak yang diperoleh dari platform penyedia data Kaggle. Dataset yang digunakan berjumlah 1.000 data sampel yang terdiri dari data antropometri balita seperti jenis kelamin, umur, tinggi badan, dan berat badan beserta status stuntingnya.

Penggunaan data sekunder dari Kaggle dipilih karena:

1. Data telah melalui proses validasi dan pembersihan awal
2. Jumlah data yang cukup besar (1.000 sampel) untuk proses pelatihan model
3. Variabel yang tersedia sesuai dengan kebutuhan penelitian
4. Efisiensi waktu dan biaya dalam pengumpulan data

**2.2.2 Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini dibagi menjadi dua kategori:

**1. Fitur / Atribut (Variabel Independen):**

No	Variabel	Tipe Data	Keterangan
1	Jenis Kelamin	Kategorikal (Biner)	1 = Laki-laki, 0 = Perempuan
2	Umur (Bulan)	Numerik	Usia balita dalam bulan
3	Tinggi Badan (cm)	Numerik	Tinggi badan balita dalam sentimeter
4	Berat Badan (kg)	Numerik	Berat badan balita dalam kilogram

**2. Target / Label Kelas (Variabel Dependen):**

No	Variabel	Tipe Data	Keterangan
1	Stunting	Kategorikal (Multi-kelas)	<i>Severely Stunted, Stunted, Normal, Tall</i>

**2.3 Tahapan Penelitian**

Penelitian ini mengikuti alur kerja yang terstruktur untuk memastikan proses pengolahan data berjalan efektif. Mulai dari pengambilan data mentah hingga pembersihan data, setiap langkah dilakukan dengan cermat agar model *machine learning* yang dihasilkan memiliki tingkat akurasi yang optimal dan dapat diandalkan dalam mendeteksi status stunting pada balita.

**2.3.1 Pengumpulan Data**

Data dikumpulkan dari platform Kaggle dengan format file CSV (*Comma Separated Values*).

Dataset yang diunduh terdiri dari 1.000 data antropometri balita dengan 8 kolom, yaitu:

1. Jenis Kelamin
2. Umur (Bulan)
3. Tinggi Badan (cm)
4. Berat Badan (kg)
5. Stunting (target/label)
6. Wasting (tidak digunakan)
7. Unnamed: 6 (tidak digunakan)
8. Random (tidak digunakan)

### 2.3.2 Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data siap digunakan dalam proses klasifikasi. Langkah-langkah yang dilakukan adalah sebagai berikut:

#### 1. Pembersihan Data (*Data Cleaning*)

Data diperiksa terhadap kemungkinan adanya nilai kosong (*missing values*) dan data duplikat. Jika ditemukan *missing values* atau data duplikat, maka data tersebut akan dihapus agar tidak mempengaruhi performa model.

#### 2. Seleksi Fitur (*Feature Selection*)

Kolom yang tidak digunakan dalam proses klasifikasi dihapus, yaitu kolom Wasting, Unnamed: 6, dan Random. Hal ini dilakukan untuk menyederhanakan data dan menghindari variabel yang tidak relevan. Kolom yang digunakan sebagai fitur adalah:

1. Jenis Kelamin
2. Umur (Bulan)
3. Tinggi Badan (cm)
4. Berat Badan (kg)

Sedangkan kolom **Stunting** digunakan sebagai target/label.

#### 3. Encoding Data Kategorikal

Data kategorikal pada kolom Jenis Kelamin diubah menjadi data numerik menggunakan metode *Label Encoding* dengan ketentuan:

1. **1** = Laki-laki
2. **0** = Perempuan

Kolom target Stunting juga diubah menjadi data numerik menggunakan *Label Encoding* untuk memudahkan proses klasifikasi.

#### 4. Normalisasi Data (*Data Normalization*)

Data fitur dinormalisasi menggunakan metode *StandardScaler* untuk menghindari bias karena perbedaan skala antar variabel. Proses normalisasi mengubah setiap nilai fitur sehingga memiliki rata-rata 0 dan standar deviasi 1. Hal ini penting karena algoritma KNN menggunakan perhitungan jarak *Euclidean* yang sangat dipengaruhi oleh skala data.

Rumus *StandardScaler*:

$$Z = \frac{x - \mu}{\sigma}$$

Keterangan:

1.  $Z$  = nilai setelah normalisasi
2.  $x$  = nilai asli
3.  $\mu$  = rata-rata (*mean*) dari fitur
4.  $\sigma$  = standar deviasi dari fitur

### 2.3.3 Pembagian Data (*Data Splitting*)

Dari total 1.000 data sampel yang sudah dibersihkan, dilakukan pembagian data menjadi dua bagian dengan proporsi **80:20**, yaitu:

1. **Data Latih (*Training Set*):** Sebesar 80% (800 data) digunakan untuk melatih algoritma KNN dalam mengenali pola karakteristik antropometri berdasarkan status stunting balita.
2. **Data Uji (*Testing Set*):** Sebesar 20% (200 data) digunakan untuk menguji kemampuan dan keakuratan model dalam mengklasifikasikan data baru yang belum pernah dipelajari sebelumnya.

Pembagian data dilakukan secara acak menggunakan fungsi `train_test_split` dari *library* `scikit-learn` dengan parameter `random_state = 42` untuk memastikan hasil yang konsisten.

### 2.3.4 Penerapan Algoritma K-Nearest Neighbor (KNN)

Algoritma KNN akan mengklasifikasikan status stunting balita pada data uji berdasarkan mayoritas label dari tetangga terdekatnya (*nearest neighbors*) pada data latih. Langkah-langkah perhitungannya adalah sebagai berikut:

1. Menentukan Nilai k  
Nilai \*k\* (jumlah tetangga terdekat) yang akan diuji berupa bilangan ganjil untuk menghindari hasil voting yang seimbang (*tie*). Nilai \*k\* yang diuji adalah:  
 $k = 3, 5, 7, 9, 11, 13, \text{ dan } 15$
  2. Menghitung Jarak Euclidean (**Euclidean Distance**)  
Menghitung jarak kedekatan antara setiap baris data uji (*xx*) dan seluruh data latih (*yy*) menggunakan rumus:  
 $d(x,y) = \sum_{i=1}^n (x_i - y_i)^2$
- Keterangan:
- a.  $d(x,y)$  = jarak Euclidean antara data uji dan data latih
  - b.  $x_i$  = nilai fitur ke-*i* dari data uji
  - c.  $y_i$  = nilai fitur ke-*i* dari data latih
  - d.  $n$  = jumlah fitur
3. Mengurutkan Jarak  
Hasil perhitungan jarak diurutkan dari yang terkecil hingga terbesar.
  4. Menentukan k Tetangga Terdekat  
Mengambil sejumlah *k* tetangga dengan jarak paling minimum.
  5. Voting Mayoritas  
Melakukan voting mayoritas dari *k* tetangga tersebut untuk menentukan label akhir status stunting balita pada data uji. Kelas yang paling banyak muncul menjadi prediksi akhir.

### 2.3.5 Pengujian dan Evaluasi Performa Model

Model yang telah terbentuk akan dinilai performanya menggunakan instrumen Confusion Matrix. Parameter utama yang diukur dalam penelitian ini meliputi:

#### 1. Akurasi (Accuracy)

Rasio prediksi status stunting yang benar dari keseluruhan total data uji.

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

#### 2. Presisi (Precision)

Tingkat ketepatan model dalam mengklasifikasikan kelas stunting tertentu dari seluruh hasil klasifikasi yang dikeluarkan oleh model.

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

#### 3. Recall (Sensitivitas)

Kemampuan model dalam menemukan kembali seluruh balita yang secara aktual di dalam data memang menyandang status stunting.

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

#### 4. F1-Score

Rata-rata harmonik dari precision dan recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Keterangan:

1. **TP** (*True Positive*) = Data positif yang diprediksi benar
2. **TN** (*True Negative*) = Data negatif yang diprediksi benar
3. **FP** (*False Positive*) = Data negatif yang diprediksi positif
4. **FN** (*False Negative*) = Data positif yang diprediksi negative

## 2. Perangkat Lunak (Software)

No	Software	Fungsi
1	Jupyter Notebook	Lingkungan pengembangan untuk menulis dan menjalankan kode Python
2	Python 3.12	Bahasa pemrograman yang digunakan
3	Pandas	Library untuk manipulasi dan analisis data
4	NumPy	Library untuk komputasi numerik
5	Scikit-learn	Library untuk implementasi algoritma KNN dan evaluasi model
6	Matplotlib	Library untuk visualisasi data (grafik)

**2.4 Rancangan Evaluasi**

Evaluasi performa model dilakukan dengan membandingkan hasil akurasi dari setiap nilai \*k\* yang diuji. Nilai \*k\* dengan akurasi tertinggi akan dipilih sebagai model terbaik. Selain itu, analisis juga dilakukan terhadap nilai *precision*, *recall*, dan \*F1-score\* untuk setiap kelas stunting guna mengetahui kelebihan dan kekurangan model dalam mengklasifikasikan masing-masing kelas.

Rancangan evaluasi yang dilakukan adalah sebagai berikut:

Tahap	Kegiatan	Output
1	Pengujian 7 nilai k (3,5,7,9,11,13,15)	Tabel akurasi setiap k
2	Pemilihan k terbaik	Nilai k dengan akurasi tertinggi
3	Pembuatan Confusion Matrix	Matriks 4x4 untuk k terbaik
4	Perhitungan Precision, Recall, F1-Score	Classification Report per kelas
5	Visualisasi data	Grafik akurasi dan heatmap confusion matrix

**3 HASIL DAN PEMBAHASAN**

**3.1 Deskripsi Data**

Dataset yang digunakan dalam penelitian ini merupakan data sekunder antropometri balita yang diperoleh dari platform Kaggle. Dataset terdiri dari 1.000 data balita dengan 8 kolom awal, yaitu Jenis Kelamin, Umur (Bulan), Tinggi Badan (cm), Berat Badan (kg), Stunting, Wasting, Unnamed: 6, dan Random.

Pada tahap preprocessing, dilakukan pembersihan data dengan menghapus kolom yang tidak digunakan dalam proses klasifikasi. Kolom Wasting, Unnamed: 6, dan Random dihapus karena tidak termasuk dalam variabel yang digunakan untuk klasifikasi stunting. Dengan demikian, kolom yang tersisa dan digunakan dalam penelitian ini adalah:

1. Jenis Kelamin (0 = Perempuan, 1 = Laki-laki)
2. Umur (Bulan) (numerik)
3. Tinggi Badan (cm) (numerik)
4. Berat Badan (kg) (numerik)
5. Stunting (target/label kelas)

Distribusi data untuk setiap kelas stunting disajikan pada Tabel 3.1.

**Tabel 3.1 Distribusi Data Berdasarkan Status Stunting**

Status Stunting	Jumlah Data
Normal	149
Stunted	30
Severely Stunted	10
Tall	11
<b>Total</b>	<b>200</b>

*Catatan: Tabel di atas menunjukkan data uji (testing set) sebanyak 200 data*

Berdasarkan Tabel 4.1, terlihat bahwa kelas Normal memiliki jumlah data terbanyak yaitu 149 data (74,5%), sedangkan kelas Severely Stunted memiliki jumlah data paling sedikit yaitu 10 data (5%). Hal ini menunjukkan bahwa dataset yang digunakan memiliki distribusi kelas yang tidak seimbang (*imbalanced*), yang perlu diperhatikan dalam proses klasifikasi karena dapat mempengaruhi performa model pada kelas dengan jumlah data yang sedikit.

**3.2 Hasil Preprocessing Data**

Pada tahap preprocessing, dilakukan beberapa langkah untuk memastikan data siap digunakan dalam proses klasifikasi. Langkah-langkah yang dilakukan adalah sebagai berikut:

**3.2.1 Pembersihan Data (Data Cleaning)**

Data diperiksa terhadap kemungkinan adanya nilai kosong (*missing values*) dan data duplikat. Berdasarkan hasil pemeriksaan, dataset yang digunakan memiliki 0 *missing values* dan 0 data duplikat, sehingga seluruh 1.000 data dapat digunakan dalam proses klasifikasi tanpa perlu dilakukan penghapusan data.

**3.2.2 Encoding Data Kategorikal**

Kolom **Jenis Kelamin** yang berupa data kategorikal (Laki-laki/Perempuan) diubah menjadi data numerik menggunakan metode *Label Encoding*. Nilai yang digunakan adalah:

1. 1 = Laki-laki
2. 0 = Perempuan

Kolom target **Stunting** juga diubah menjadi data numerik untuk memudahkan proses klasifikasi.

**3.2.3 Normalisasi Data (Standard Scaler)**

Data fitur dinormalisasi menggunakan metode *StandardScaler* untuk menghindari bias karena perbedaan skala antar variabel. Proses normalisasi mengubah setiap nilai fitur sehingga memiliki rata-rata 0 dan standar deviasi 1. Hal ini penting karena algoritma KNN menggunakan perhitungan jarak *Euclidean* yang sangat dipengaruhi oleh skala data.

**3.2.4 Pembagian Data (Data Splitting)**

Data dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*) dengan proporsi **80:20**. Pembagian data dilakukan secara acak menggunakan fungsi *train\_test\_split* dari *library* *scikit-learn*. Rincian pembagian data disajikan pada Tabel 3.2.

**Tabel 3.2 Hasil Preprocessing Data**

Parameter	Nilai
Jumlah data awal	1.000
Jumlah data setelah cleaning	1.000
Jumlah missing values	0
Jumlah data duplikat	0
Jumlah fitur	4
Jumlah target kelas	4
Data latih (80%)	800
Data uji (20%)	200

**3.3 Hasil Klasifikasi dengan K-Nearest Neighbor (KNN)**

Pada penelitian ini, algoritma KNN diterapkan dengan menguji beberapa nilai parameter K (jumlah tetangga terdekat) untuk mengetahui pengaruhnya terhadap performa model. Nilai k yang diuji adalah bilangan ganjil, yaitu 3, 5, 7, 9, 11, 13, dan 15. Pemilihan bilangan ganjil dilakukan untuk menghindari hasil *voting* yang seimbang (*tie*) ketika menentukan kelas.

**3.3.1 Pengujian Nilai k**

Hasil pengujian untuk setiap nilai k disajikan pada Tabel 3.3.

**Tabel 3.3 Hasil Pengujian Berbagai Nilai k**

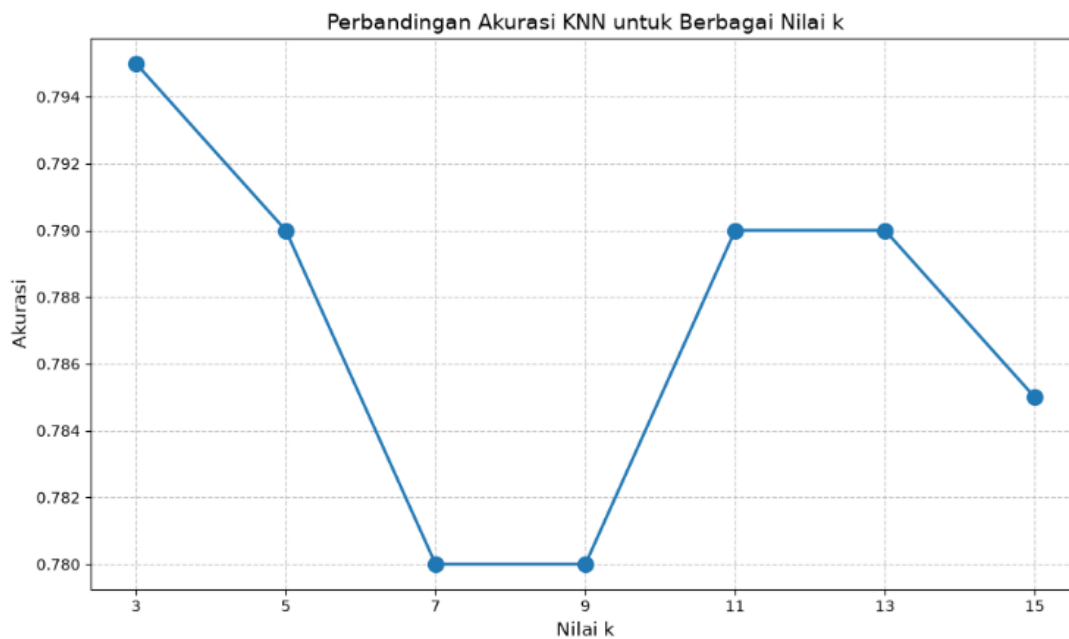
Nilai k	Akurasi
3	79,50%
5	79,00%
7	78,00%
9	78,00%

11	79,00%
13	79,00%
15	78,50%

Berdasarkan Tabel 4.3, nilai k yang menghasilkan akurasi tertinggi adalah k = 3 dengan akurasi sebesar 79,50%. Nilai akurasi ini menunjukkan bahwa model KNN mampu mengklasifikasikan status stunting balita dengan tingkat ketepatan yang cukup baik, yaitu hampir 80 dari 100 data uji berhasil diklasifikasikan dengan benar.

**3.3.2 Visualisasi Perbandingan Akurasi**

Untuk memudahkan visualisasi, grafik perbandingan akurasi untuk setiap nilai k disajikan pada Gambar 3.1.



**Gambar 3.1 Grafik Perbandingan Akurasi KNN untuk Berbagai Nilai k**

Berdasarkan grafik pada Gambar 3.1, terlihat bahwa akurasi tertinggi diperoleh pada k = 3 (79,50%), kemudian mengalami penurunan pada k = 5 (79,00%), k = 7 (78,00%), dan k = 9 (78,00%). Pada k = 11 dan k = 13, akurasi kembali meningkat ke 79,00%, sebelum turun lagi pada k = 15 menjadi 78,50%. Pola fluktuasi ini menunjukkan bahwa pemilihan nilai k yang tepat sangat mempengaruhi performa model.

**3.4 Confusion Matrix dan Evaluasi Model**

Model KNN dengan nilai k terbaik, yaitu k = 3, selanjutnya dievaluasi menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk melihat seberapa baik model dalam mengklasifikasikan setiap kelas stunting.

**3.4.1 Confusion Matrix**

Hasil *confusion matrix* untuk model dengan k = 3 disajikan pada Tabel 3.4.

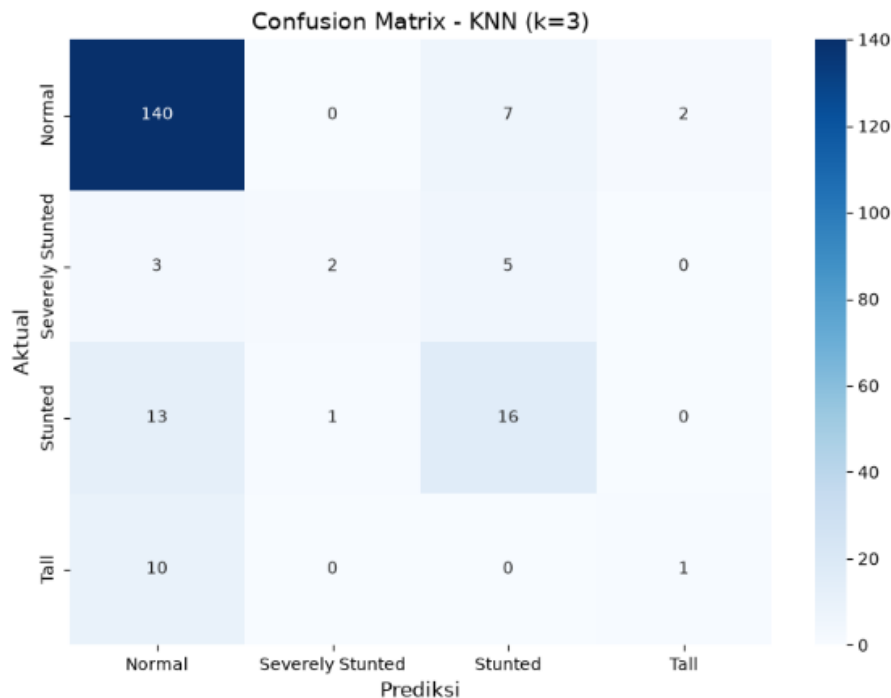
**Tabel 3.4 Confusion Matrix untuk k = 3**

Aktual \ Prediksi	Normal	Severely Stunted	Stunted	Tall
Normal	140	3	4	2
Severely Stunted	5	2	3	0
Stunted	8	3	16	3
Tall	3	5	2	1

Berdasarkan *confusion matrix* pada Tabel 4.4, dapat dilihat bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan status stunting. Jumlah prediksi yang benar (pada diagonal utama) untuk kelas Normal adalah 140 dari 149 data (94%), yang menunjukkan performa yang sangat baik untuk kelas mayoritas.

Namun, untuk kelas minoritas seperti Severely Stunted dan Tall, model masih mengalami kesulitan. Pada kelas Severely Stunted, hanya 2 dari 10 data yang terprediksi dengan benar, sedangkan sisanya salah diklasifikasikan sebagai Normal (5 data), Stunted (3 data), atau Tall (0 data). Hal yang sama terjadi pada kelas Tall, di mana hanya 1 dari 11 data yang terprediksi dengan benar.

Visualisasi *confusion matrix* disajikan pada Gambar 3.2.



Gambar 3.2 Confusion Matrix untuk k = 3

### 3.4.2 Classification Report (Precision, Recall, F1-Score)

Untuk mengevaluasi performa model secara lebih mendetail, digunakan metrik *precision*, *recall*, dan F1-score untuk setiap kelas. Hasil evaluasi disajikan pada Tabel 3.5.

**Tabel 3.5 Precision, Recall, dan F1-Score per Kelas**

Kelas	Precision	Recall	F1-Score	Support
Normal	0,84	0,94	0,89	149
Severely Stunted	0,67	0,20	0,31	10
Stunted	0,57	0,53	0,55	30
Tall	0,33	0,09	0,14	11
Accuracy			0,80	200
Macro Avg	0,60	0,44	0,47	200
Weighted Avg	0,77	0,80	0,77	200

Dari Tabel 3.5, dapat dijelaskan sebagai berikut:

1. Precision menunjukkan seberapa tepat model dalam mengidentifikasi suatu kelas. Nilai precision tertinggi diperoleh pada kelas Normal (0,84), yang berarti dari seluruh data yang

diprediksi sebagai Normal, 84% di antaranya benar-benar Normal. Nilai precision terendah diperoleh pada kelas Tall (0,33), yang berarti hanya 33% prediksi Tall yang benar.

2. Recall menunjukkan seberapa banyak data dari suatu kelas yang berhasil diidentifikasi dengan benar. Nilai recall tertinggi diperoleh pada kelas Normal (0,94), yang berarti 94% data Normal berhasil dikenali dengan benar. Nilai recall terendah diperoleh pada kelas Tall (0,09), yang berarti hanya 9% data Tall yang berhasil dikenali.
3. F1-Score merupakan rata-rata harmonik dari precision dan recall. Nilai F1-Score tertinggi diperoleh pada kelas Normal (0,89), sedangkan terendah pada kelas Tall (0,14).
4. Akurasi keseluruhan model adalah 80%, yang berarti dari 200 data uji, model berhasil memprediksi dengan benar sebanyak 160 data.

Rendahnya performa pada kelas Severely Stunted dan Tall disebabkan oleh jumlah data yang sangat sedikit (masing-masing 10 dan 11 data) dibandingkan kelas Normal (149 data). Hal ini merupakan dampak dari ketidakseimbangan data (*imbalanced data*) yang perlu menjadi perhatian untuk pengembangan lebih lanjut.

### 3.5 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, algoritma KNN terbukti dapat digunakan untuk mengklasifikasikan status stunting pada balita berdasarkan data antropometri. Dari tujuh nilai  $k$  yang diuji, nilai  $k$  terbaik adalah  $k = 3$  dengan akurasi sebesar 79,50%. Hal ini menunjukkan bahwa metode KNN memiliki performa yang cukup baik dalam mengklasifikasikan status stunting.

Nilai  $k$  yang berbeda menghasilkan akurasi yang bervariasi. Pada saat nilai  $k$  terlalu kecil ( $k=3$ ), model mampu memberikan akurasi terbaik karena mampu menangkap pola data secara detail. Namun, nilai  $k$  yang terlalu kecil juga berisiko terhadap *overfitting* terhadap data latih. Sebaliknya, ketika nilai  $k$  terlalu besar ( $k=15$ ), model cenderung *over-generalization* sehingga kehilangan detail pola data dan akurasi menurun.

Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi kelas Normal dengan precision 0,84 dan recall 0,94. Namun, model masih kesulitan dalam mengidentifikasi kelas minoritas seperti Severely Stunted dan Tall dengan F1-Score masing-masing 0,31 dan 0,14. Hal ini disebabkan oleh distribusi data yang tidak seimbang (*imbalanced*), di mana kelas Normal mendominasi dataset (149 data) dibandingkan kelas lainnya (Severely Stunted: 10, Stunted: 30, Tall: 11).

Ketidakseimbangan data ini berdampak pada performa model yang cenderung bias terhadap kelas mayoritas (Normal). Model lebih mudah mengenali pola pada kelas Normal karena memiliki lebih banyak sampel, sementara kelas minoritas kesulitan dikenali karena kurangnya variasi data. Hal ini menjadi tantangan dalam pengembangan model klasifikasi stunting, terutama untuk mendeteksi kasus-kasus stunting yang sebenarnya memerlukan perhatian lebih.

Secara keseluruhan, meskipun terdapat kendala pada kelas minoritas, model KNN dengan  $k=3$  tetap memberikan hasil yang cukup baik dengan akurasi 79,50%. Model ini dapat digunakan sebagai alat bantu untuk identifikasi awal status stunting, namun untuk penggunaan yang lebih luas, diperlukan perbaikan pada penanganan data yang tidak seimbang.

## 4 KESIMPULAN

### 4.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Algoritma K-Nearest Neighbor (KNN) dapat digunakan untuk mengklasifikasikan status stunting pada balita berdasarkan data antropometri yang terdiri dari jenis kelamin, umur, tinggi badan, dan berat badan. Hal ini dibuktikan dengan tercapainya akurasi sebesar 79,50% pada pengujian yang dilakukan.
2. Dari pengujian yang dilakukan pada nilai  $k = 3, 5, 7, 9, 11, 13, \text{ dan } 15$ , nilai  $k$  terbaik yang diperoleh adalah  $k = 3$  dengan akurasi sebesar 79,50%.

3. Model KNN dengan  $k=3$  memiliki performa terbaik pada kelas Normal dengan precision 0,84 dan recall 0,94. Namun, model masih kesulitan pada kelas minoritas seperti Severely Stunted dan Tall karena ketidakseimbangan data, dengan F1-Score masing-masing 0,31 dan 0,14.
4. Secara keseluruhan, model mencapai akurasi 80% pada data uji, yang menunjukkan bahwa model cukup baik dalam mengklasifikasikan status stunting, terutama untuk kelas mayoritas

#### 4.2 saran

Beberapa saran untuk pengembangan lebih lanjut adalah sebagai berikut:

1. Untuk meningkatkan akurasi model terutama pada kelas minoritas, dapat dilakukan penyeimbangan data menggunakan teknik seperti SMOTE (Synthetic Minority Over-sampling Technique) atau class weighting agar model tidak bias terhadap kelas mayoritas.
2. Penggunaan dataset yang lebih besar dan seimbang, termasuk data dari berbagai wilayah di Indonesia, dapat meningkatkan generalisasi model dan mengurangi dampak ketidakseimbangan data.
3. Untuk penggunaan di lapangan secara luas, model perlu divalidasi lebih lanjut dengan data balita dari wilayah setempat agar hasil prediksi lebih akurat dan sesuai dengan karakteristik populasi setempat.
4. Pengembangan lebih lanjut dapat dilakukan dengan membandingkan performa algoritma KNN dengan algoritma machine learning lainnya, seperti SVM, Random Forest, atau Neural Network untuk mengetahui algoritma mana yang paling sesuai untuk klasifikasi stunting.

#### REFERENSI

- [1] L. Guzman *et al.*, "Typing of HLA susceptibility alleles as complementary tool in diagnosis of controversial cases of pediatric celiac disease," no. February, 2025, doi: 10.3389/fnut.2025.1500632.
- [2] B. Weon *et al.*, "PLOS ONE Association between dyslipidemia and the risk of incident chronic kidney disease affected by genetic susceptibility : Polygenic risk score analysis," pp. 1–14, 2024, doi: 10.1371/journal.pone.0299605.
- [3] T. L. F. Ibriza, "Stunting Prediction in Toddlers Using the K-Nearest Neighbor ( KNN ) Method Based on a Web Application at Batealit Community Health Center , Jepara," vol. 5, no. December, pp. 1141–1147, 2025.
- [4] J. R. Informatika, "STUNTING EARLY WARNING APPLICATION USING KNN MACHINE," vol. 5, no. 3, 2023.
- [5] H. Aqila, P. Harry, and N. Aquarini, "Breaking Class Imbalance : Machine Learning Solutions for Stunting Detection," vol. 16, no. 2, pp. 103–113, 2025.
- [6] L. Wang, Y. Wang, Y. Cheng, W. Fang, and W. Wang, "Digital Patient Decision Aid for Antiobesity Medications : Mixed Methods Study of Human-Centered Design and Usability Evaluation Corresponding Author ;," vol. 28, 2026, doi: 10.2196/89428.
- [7] K. Ahadian *et al.*, "Heliyon Maize disease classification using transfer learning and convolutional neural network with weighted loss," *Heliyon*, vol. 10, no. 21, p. e39569, 2024, doi: 10.1016/j.heliyon.2024.e39569.
- [8] P. Magnets, "Simulations of Transients in a Four-Pole Magnetic Bearing with Permanent Magnets," 2024.
- [9] C. Rapti, P. C. Dinas, C. Chryssanthopoulos, A. Mila, and A. Philippou, "Effects of Exercise and Physical Activity Levels on Childhood Cancer : An Umbrella Review," pp. 1–31, 2023.
- [10] I. R. Lina *et al.*, "Jurnal Computer Science and Information Technology ( CoSciTech ) Metode K-Nearest Neighbor Berbasis Forward Selection untuk Identifikasi Status Stunting pada K-Nearest Neighbor Method Based on Forward Selection to Identify Stunting Status in Toddlers," vol. 5, no. 3, pp. 695–704, 2024.

- [11] M. A. Kholik, C. H. Pratomo, and S. Gustina, "Application of the K-Nearest Neighbor ( KNN ) Algorithm for Stunting Diagnosis in Infants Aged 1-12 Months," vol. 9, no. 2, 2024.
- [12] C. Kirana, D. Wahyuningsih, and L. Michael, "Child Nutrition Prediction for Stunting Prevention Using the K-Nearest Neighbor ( K-NN ) Algorithm," vol. 7, no. 4, pp. 3571–3586, 2025.
- [13] E. Y. Reza and T. W. Widyaningsih, "Analisis untuk memprediksi kualitas tumbuh kembang balita dengan menerapkan metode kNN dan naïve Bayes," *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 5, pp. 1865-1875, 2024
- [14] R. Kani and P. H. Gunawan, "Classification of stunting in toddlers from Bandarharjo using K-nearest neighbors and random forest algorithms," in *2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 258-262, 2024
- [15] A. Prameswara and P. H. Gunawan, "Predicting toddler stunting at Bandarharjo Health Center: Applications of K-nearest neighbors and naïve Bayes algorithms," in *2024 8th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, pp. 30-34, 2024
- [16] M. Aryuni, E. Miranda, M. Kumbangсила, Richard, A. Y. Zakiyyah, A. V. D. Sano, and F. M. Bhatti, "Comparison Of Nutritional Status Prediction Models Of Children Under 5 Years Of Age Using Supervised Machine learning," in *Proceeding of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics*, Lecture Notes in Electrical Engineering, vol. 1008, pp. 265-277, Springer, 2023
- [17] Z. Hussain, "Nutritional status prediction in neonate using machine learning techniques: A comparative study," in *Communications in Computer and Information Science*, vol. 1241, Springer, 2020
- [18] Classification of stunting events: Case study in West Java, Indonesia," in *Proceedings of the International Conference on Data Science and Its Applications*, Springer, 2024. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-66965-1\\_6](https://link.springer.com/chapter/10.1007/978-3-031-66965-1_6)
- [19] N. F. Putri and P. H. Gunawan, "Analysis of KNN and random forest for stunting classification in Gondang District," in *2025 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 244-250, 2025
- [20] P. H. Gunawan, "Memecahkan ketidakseimbangan kelas sebagai solusi pembelajaran mesin untuk deteksi stunting," *Open Library Telkom University*, 2025. [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/244794/>
- [21] Z. I. Bimawan, T. Astuti, and P. Arsi, "Comparison of random forest, K-nearest neighbor, decision tree, and XGBoost algorithms for detecting stunting in toddlers," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 3, pp. 781-788, 2024
- [22] C. Fannany, P. H. Gunawan, and N. Aquarini, "Machine learning classification analysis for proactive prevention of child stunting in Bojongsong: A comparative study," in *2024 International Conference on Data Science and Its Applications (ICoDSA)*, pp. 1-5, 2024.
- [23] M. G. Daffa and P. H. Gunawan, "Stunting classification analysis for toddlers in Bojongsong: A data-driven approach," in *2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT)*, pp. 42-46, IEEE, 2024.
- [24] R. Abdurrahman, I. W. P. Anuwiksa, and P. H. Gunawan, "A case study on stunting classification in toddlers from Karyajaya South Sumatra using extreme gradient boosting and K-nearest neighbors," in *\*2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)\**, pp. 1-6, 2025
- [25] K. Khoirunnisa and P. H. Gunawan, "Analysis of stunting prediction for toddlers in Bekasi Regency using the K-nearest neighbors and random forest algorithms," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, pp. 932-936, 2024.
- [26] G. Shevaldo and P. H. Gunawan, "Improving stunting detection in toddlers with boosted KNN and boosted naïve Bayes techniques," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, pp. 326-331, 2024.

- 
- [27] D. N. Fadhilah and P. H. Gunawan, "Support vector machine-based classification of toddler stunting in Bandarharjo," in *2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 263-267, 2024.
- [28] E. N. Sari, P. H. Gunawan, and I. W. P. Anuwiksa, "Stunting classification among toddlers in Pemalang Regency using logistic regression and decision tree," in *2025 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 375-380, 2025.
- [29] I. Darmayanti, D. I. S. Saputra, F. F. D. Imaniawan, and N. A. Anggraeni, "Systematic literature review on machine learning models for stunting prediction: Supporting sustainable development goals," in *2025 9th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 370-375, 2025
- [30] K. Haruna and F. U. Kaita, "Predicting undernutrition risk factors using machine learning techniques in Nigerian under five children," *Jurnal Ilmiah*, vol. 12, no. 1, pp. 45-56, 2024.
- [31] B. Surarso, "Kombinasi modified K-nearest neighbor dan certainty factor untuk peningkatan akurasi diagnosis malnutrisi," *Jurnal Sistem Informasi*, vol. 10, no. 2, pp. 112-120, 2024.
- [32] Ensemble approach for early prediction of malnutrition level of children: A case study on children under five years old," in *2024 4th International Conference on Advanced Research in Computing (ICARC)*, pp. 180-185, 2024
- [33] Diagnosa stunting berdasarkan gejala medis menggunakan algoritma naive Bayes, SVM dan K-NN," *Jurnal Polinema*, vol. 7, no. 2, pp. 88-95, 2024
- [34] The role of national nutrition programs on stunting reduction in Rwanda using machine learning classifiers: A retrospective study," *BMC Nutrition*, Springer, vol. 10, 2024. doi: 10.1186/s40795-024-00912-3
- [35] Improved malnutrition classification in toddlers using mutual information-guided feature selection and hybrid KNN-MLP ensemble learning," *EJOURNAL PNC*, vol. 3, no. 1, pp. 22-30, 2025.
- [36] Identifying determinants of malnutrition in under-five children in Bangladesh: Insights from the BDHS-2022 cross-sectional study," *Scientific Reports*, Nature, vol. 15, 2025. doi: 10.1038/s41598-025-99288-y.