



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 3 Tahun 2024 Page 8401-8417

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

## Penggunaan Algoritma K-Nearest Neighbor (KNN) Untuk Mendeteksi Stunting Pada Anak

Natalia Situmeang<sup>1✉</sup>, Eliena Artamezia Komarsyah<sup>2</sup>, Ahmad Fauzi<sup>3</sup>, Jepri Sagala<sup>4</sup>

Fakultas Teknologi dan Ilmu Komputer, Universitas Satya Terra Bhinneka

Email: [nataliasitumeang942@gmail.com](mailto:nataliasitumeang942@gmail.com)<sup>1✉</sup>

### Abstrak

Stunting merupakan salah satu masalah gizi kronis yang masih tinggi di Indonesia dan berdampak serius pada pertumbuhan serta perkembangan anak. Penelitian ini bertujuan untuk membangun sistem klasifikasi status stunting pada balita menggunakan algoritma *K-Nearest Neighbors* (KNN). Data yang digunakan berasal dari platform Kaggle, mencakup variabel usia, jenis kelamin, tinggi, dan berat badan anak. Penelitian dilakukan melalui tahapan pengolahan data, normalisasi, pembagian data latih dan uji, serta evaluasi performa model. Hasil pengujian menunjukkan bahwa model KNN mampu mengklasifikasikan status stunting dengan akurasi tinggi mencapai 97%, serta nilai precision, recall, dan F1-score yang seimbang. Penerapan model ini diharapkan dapat membantu tenaga medis dan pihak terkait dalam mendeteksi risiko stunting secara dini, sehingga dapat dilakukan intervensi yang lebih cepat dan tepat untuk menekan angka stunting di Indonesia.

Kata kunci: *stunting, klasifikasi, K-Nearest Neighbors, machine learning, deteksi dini*

## Abstract

Stunting is one of the chronic nutritional problems that remains high in Indonesia and has a serious impact on children's growth and development. This study aims to develop a classification system for stunting status in toddlers using the K-Nearest Neighbors (KNN) algorithm. The data used comes from the Kaggle platform and includes variables such as age, gender, height, and weight of the child. The research was conducted through data processing, normalization, training and testing data division, and model performance evaluation. The test results showed that the KNN model was able to classify stunting status with high accuracy reaching 97%, as well as balanced precision, recall, and F1-score values. The application of this model is expected to assist medical personnel and relevant parties in detecting the risk of stunting early, enabling faster and more accurate interventions to reduce stunting rates in Indonesia.

*Keywords: stunting, classification, K-Nearest Neighbors, machine learning, early detection*

## PENDAHULUAN

Pertumbuhan dan perkembangan balita merupakan aspek penting dalam kesehatan masyarakat (Aminah et al., 2024). Pada Tahun 2021 Hasil Survei Status Gizi Indonesia (SSGI) menunjukkan bahwa angka stunting masih cukup tinggi sebesar 24.4%, tetapi pada tahun 2022 mengalami penurunan dimana tingkat stunting menurun menjadi 21.6%. Meskipun mengalami penurunan, hal ini masih jauh dari target yang hendak di capai yaitu sebesar 14%. Oleh karena itu, pencapaian ini harus dimaksimalkan. Daerah-daerah di mana tingkat stunting masih tinggi Contohnya adalah provinsi Nusa Tenggara Timur (NTT), Nusa Tenggara Barat (NTB), Sulawesi Barat (Sulbar), Sulawesi Tengah (Sulteng), Kalimantan Selatan (Kalsel), Kalimantan Barat (Kalbar), dan Aceh Desa Kesetnana adalah salah satu dari 278 desa di Kabupaten TTS yang memiliki prevalensi stunting tinggi. Bahkan, prevalensi stunting di Kabupaten TTS menurut Studi Status Gizi Indonesia (SSGI) mencapai 48,3%, yang paling tinggi di NTT dan bahkan di Indonesia (Rahmadhita, 2020). Berdasarkan data tersebut diperlukan alternatif sebagai solusi terhadap untuk mengatasi melalui implementasi dalam pemahaman pencegahan stunting dalam mendukung Indonesia emas di tahun 2045. (Hastuti & Dulame, 2024)

Seiring dengan perkembangan teknologi, pemanfaatan kecerdasan buatan dan machine learning dalam bidang kesehatan semakin berkembang. Salah satu algoritma yang efektif dalam klasifikasi data adalah K-Nearest Neighbors (K-NN), yang mampu mengidentifikasi pola berdasarkan data historis. (Aminah et al., 2024). Pencapaian yang diinginkan dalam penelitian ini ialah untuk mengembangkan dan menerapkan model klasifikasi menggunakan algoritma KNN. Implementasi algoritma K-Nearest Neighbor (KNN) ini memungkinkan identifikasi kondisi stunting pada balita secara lebih sistematis dan

berbasis data, yang pada akhirnya mendukung pengambilan keputusan klinis yang berbobot serta menjamin pemberian perawatan gizi yang terarah kepada kelompok anak yang berisiko. Serta memberikan wawasan, pemahaman dan penerapan teknologi dalam pemantauan kesehatan anak, khususnya dalam upaya pencegahan stunting dalam upaya mengurangi angka stunting pada balita. Data balita seperti usia, berat badan, tinggi badan dan status gizi dikumpulkan dari sampel yang relevan. Teknik analisis data seperti pemrosesan, pembagian dan penerapan algoritma KNN. Pendekatan ini dapat mendeteksi stunting anak sejak dini dan memberikan wawasan yang lebih luas dengan menggabungkan teknik informatika dengan pengetahuan sektor kesehatan.(Akbarollah et al., 2023).

Berdasarkan uraian latar belakang yang telah disampaikan, diperlukan perumusan masalah secara lebih terarah guna memberikan fokus yang jelas dalam penelitian ini. Permasalahan utama yang hendak dikaji adalah bagaimana mekanisme kerja algoritma *K-Nearest Neighbor* (KNN) dalam memprediksi risiko stunting pada anak. Penelitian ini juga berupaya memahami proses klasifikasi yang dilakukan oleh KNN dalam mengenali anak-anak yang berpotensi mengalami stunting melalui analisis kemiripan data. Di samping itu, perlu diidentifikasi fitur-fitur atau variabel yang relevan serta memiliki pengaruh signifikan dalam proses prediksi tersebut, seperti usia, berat dan tinggi badan, pola konsumsi gizi, serta faktor sosial ekonomi. Selain itu, penting untuk mengevaluasi sejauh mana tingkat akurasi algoritma KNN dalam memprediksi risiko stunting berdasarkan data yang tersedia, sebagai dasar penilaian efektivitas penerapannya di lapangan.

Selaras dengan rumusan masalah tersebut, tujuan dari penulisan ini adalah untuk memberikan penjelasan mengenai prinsip kerja algoritma *K-Nearest Neighbor* (KNN) dalam konteks prediksi risiko stunting pada anak. Penelitian ini juga ditujukan untuk mengenali pola-pola serta karakteristik anak yang memiliki risiko stunting tinggi berdasarkan hasil klasifikasi yang dihasilkan oleh algoritma KNN. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mendukung upaya pencegahan stunting, khususnya melalui pemanfaatan teknologi kecerdasan buatan sebagai alat bantu dalam sistem pengambilan keputusan berbasis data.

## Pengertian Stunting

Stunting merupakan situasi dimana terjadinya kekurangan gizi pada anak, adanya infeksi kronis, dan umumnya dilihat dari tinggi badan anak sesuai umur pada anak balita yang kurang normal. Kekurangan gizi dapat menghambat pertumbuhan dan perkembangan, menyebabkan kelainan pada bentuk tubuh, meskipun gen yang ada di dalam sel mampu berkembang secara normal. (Fauziah et al., 2023). Menurut Kementerian

Kesehatan, balita dianggap mengalami stunting jika skor z-nya kurang dari  $-2,00$  SD atau standar deviasi (stunting) atau kurang dari  $-3,00$  SD (stunting parah). Akibatnya, stunting dapat digambarkan sebagai kondisi pertumbuhan yang memengaruhi balita dan mengakibatkan keterlambatan perkembangan yang tidak sesuai dengan standar, yang dapat berdampak langsung maupun jangka panjang.(Hastuti & Dulame, 2024)

Kekurangan gizi pada anak usia ini dapat mempengaruhi perkembangan kognitif, pertumbuhan fisik, serta daya tangkap, respons, dan sistem motorik mereka. Kekurangan nutrisi juga dapat memengaruhi koordinasi mata dan tangan, sistem sensorik, serta keterampilan motorik halus dan kasar anak. Dalam konteks respons dan daya tangkap, kekurangan gizi dapat berdampak pada pemrosesan sensorik dan respons emosional, seperti kecemasan dan depresi.(Azis et al., 2024).

Faktor-faktor yang dapat menyebabkan seorang anak mengalami stunting melibatkan aspek ekonomi, pola asuh, riwayat infeksi penyakit, riwayat imunisasi, asupan protein, serta aspek kesehatan dan pengetahuan orang tua. Keluarga dengan pendapatan rendah mungkin mengalami keterbatasan dalam memenuhi kebutuhan gizi anak mereka. Pemberian ASI (Air Susu Ibu) tidak dapat diabaikan, karena ASI memberikan nutrisi esensial dan perlindungan dari infeksi yang mendukung pertumbuhan optimal anak. Kurangnya alat kesehatan seperti ultrasonografi (USG) yang diharapkan dapat mengetahui lebih dini kondisi janin, serta kurang tersedianya alat antropometrik berstandar Kemenkes di posyandu juga menjadi masalah. Pengetahuan ibu tentang gizi menentukan perilaku ibu dalam menyediakan makanan untuk anaknya. Ibu dengan pengetahuan gizi yang baik dapat menyediakan makanan yang tepat untuk mendukung pertumbuhan dan perkembangan balita. Pendidikan yang rendah dan kurangnya promosi pengetahuan tentang pencegahan stunting dapat menghambat upaya orang tua dalam memberikan asuhan gizi yang baik.(Azis et al., 2024)

### K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) merupakan salah satu metode klasifikasi yang digunakan untuk menentukan kelas suatu data berdasarkan kedekatannya dengan data lain yang telah diketahui kelasnya. Tetangga terdekat diartikan sebagai data dengan tingkat kemiripan paling tinggi atau jarak paling kecil. Pemilihan nilai  $k$  sangat penting; nilai  $k$  yang terlalu kecil dapat membuat hasil mudah terpengaruh oleh data yang tidak representatif (noise), sementara nilai  $k$  yang terlalu besar dapat mengaburkan batas antara satu kelas dengan kelas lainnya. (Dayera, Musa Bundaris Palungan, 2024). Secara umum, proses kerja algoritma KNN adalah sebagai berikut

1. Menentukan jumlah tetangga (K) yang akan digunakan untuk mengklasifikasi titik data baru.
2. Menghitung jarak antara titik data baru dengan setiap titik data dalam dataset.
3. Memilih K titik data terdekat dan menentukan kelas titik data baru berdasarkan tetangga tersebut.

Untuk menghitung jarak antara dua titik pada algoritma KNN digunakan metode

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots}$$

Euclidean Distance dengan rumus sebagai berikut:

Dimana:

Dis : Jarak antara dua titik (Distance)

i : Jumlah data set

x : dataset uji

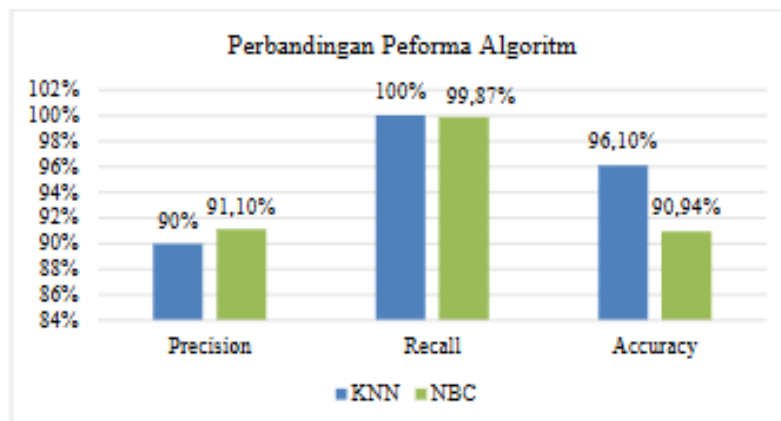
y : dataset pelatihan

#### Hasil Penelitian Terdahulu

Berdasarkan penelitian terdahulu (Dwinanto et al., 2024), model K-Nearest Neighbors (KNN) digunakan untuk memprediksi risiko stunting pada balita dengan memperhitungkan kedekatan data. KNN dipilih karena kemudahan instalasinya dan manfaatnya dalam mengelola data non-linier. Model KNN memiliki kinerja yang baik, dengan nilai akurasi rata-rata sebesar 0,87, berdasarkan hasil k-fold cross validation dengan nilai k = 5. Selanjutnya, nilai f1-score, precision, dan recall yang diperoleh masing-masing adalah 0,86, 0,64, dan 0,73. Dimana Akurasi 0.87 ini menunjukkan bahwa 87% prediksi yang dilakukan menggunakan metode KNN adalah benar. Dimana model ini baik dalam mengenali anak yang beresiko stunting dan tidak. *Precision* 0.86 menunjukkan bahwa model jarang salah menandai anak sehat sebagai beresiko stunting, sehingga kesalahan positif palsu (*false positive*) cukup rendah, ini juga menunjukkan bahwa anak yang diprediksi beresiko stunting oleh model, sebanyak 86% memang benar-benar beresiko. *Recall* 0.64 menunjukkan bahwa dari seluruh anak yang sebenarnya beresiko stunting, hanya 64% yang berhasil terdeteksi oleh model. Ini juga berarti bahwa masih ada 36% anak yang beresiko stunting

namun tidak terdeteksi oleh model (*false negative*). *Recall* ini sedikit rendah, menunjukkan bahwa model masih kehilangan sejumlah kasus penting. *F1-score* adalah rata-rata harmonis dari *precision* dan *recall*. Nilai 0.73 menunjukkan bahwa model memiliki keseimbangan yang cukup baik antara kemampuan mengenali kasus stunting (*recall*) dan ketepatan dalam menandai anak yang memang berisiko (*precision*). Secara keseluruhan, performa tergolong baik (*F1-score* 73%), tapi perlu perbaikan agar lebih sensitif terhadap kasus nyata stunting.

Pada penelitian lain (Loka & Marsal, 2023) yang menunjukkan perbandingan penggunaan model K-Nearest Neighbor (KNN) dan Naïve Bayes Classifier (NBC) untuk



Gambar 3. Perbandingan Peforma Algoritma

memprediksi resiko stunting pada anak. Pada penelitian ini diperoleh hasil sebagai berikut:

Pada gambar diatas didapatkan nilai *accuracy*, *precision*, dan *recall*. Pada algoritma KNN *accuracy* 96,10% menunjukkan bahwa prediksi model adalah benar, baik untuk anak yang berisiko maupun yang tidak. Ini juga menunjukkan angka performa umum model yang sangat baik. Untuk nilai *precision* dari seluruh anak yang berisiko stunting sebanyak 90% memang benar-benar berisiko, dan 10% sisanya tidak berisiko (*false positive*). Dan untuk *recall* menunjukkan hasil 100% semua anak yang benar-benar berisiko stunting berhasil terdeteksi oleh model, tanpa ada kasus stunting yang terlewat. Ini sangat penting dalam konteks stunting, karena mendeteksi seluruh kasus nyata adalah prioritas utama. Pada penelitian ini, peneliti menyimpulkan bahwa klasifikasi data penimbangan masal balita menggunakan algoritma KNN lebih baik dari pada algoritma NBC.

## METODE PENELITIAN

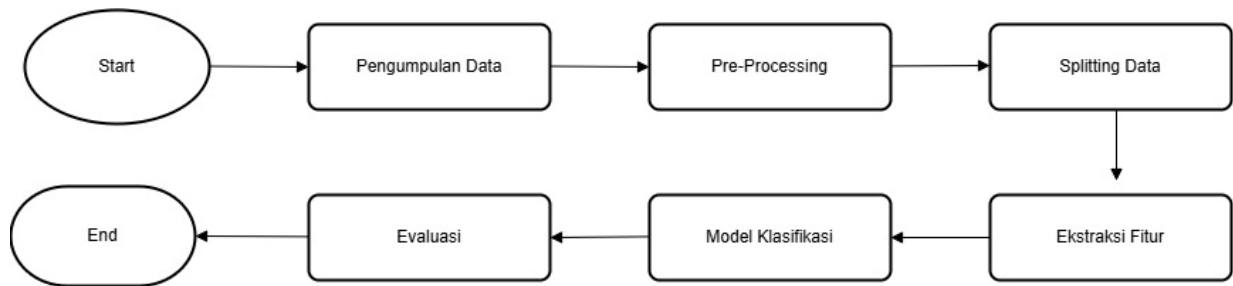
### 1. Jenis Pengumpulan Data

Jenis data yang dipakai berdasarkan sumbernya adalah data sekunder. Data ini diperoleh dari sebuah database bernama kaggle. Database tersebut mendata jenis kelamin,

usia, berat badan, dan tinggi badan. Data tersebut juga sudah membagi menjadi 2 (dua) kelompok yaitu stunting dan tidak stunting. Pada data tersebut juga mengelompokan berat badan menjadi 4 (empat) kelompok yaitu sangat kekurangan berat badan, kekurangan berat badan, resiko kelebihan berat badan, dan juga berat badan yang normal normal. Tetapi pembagian berat badan tersebut tidak digunakan karena penelitian ini berfokus pada kondisi stunting dan tidak stunting seseorang.

## 2. Tahapan Pengolahan Data

Penelitian ini akan dilakukan dengan tahapan sebagai berikut:



(Ramadhani, 2024)

1. Berdasarkan gambar diatas tahapan yang dilakukan pertama adalah pengumpulan data. Pada tahap pengumpulan data ini peneliti mengambil data dari sebuah situs database bernama Kaggle.
2. Setelah memperoleh data, tahapan kedua adalah dilakukannya *pre-processing* pada data sehingga model machine learning dapat menanganinya dengan lebih efisien.(Ramadhani, 2024). Tahapan *pre-processing* yang dilakukan antara lain menghapus kolom yang tidak digunakan, melakukan encoding label menjadi numerik untuk kolom "Jenis Kelamin" dan "Stunting", serta menampilkan kembali dataset sebelum dan sesudah diperbaharui.
3. Pada tahapan ketiga yaitu splitting data dimana data dibagi menjadi dua subset yaitu data latih dan data uji. Data latih digunakan untuk melatih model algoritma dan data uji digunakan untuk melihat kinerja model yang telah dibuat. Menggunakan library Scikit Learn pada Python untuk secara otomatis membagi dataset menjadi data uji dan data latih.(Ramadhani, 2024)
4. Tahapan keempat adalah melakukan ekstraksi fitur. Pada penelitian ini ekstraksi fitur yang dilakukan dengan melakukan normalisasi data. Normalisasi akan mengubah data sehingga menghasilkan data dengan nilai rata-rata(mean) adalah 0. Hal ini dilakukan agar fitur yang memiliki nilai besar tidak mendominasi jarak antar data dan membuat model semakin akurat. Normalisasi menggunakan rumus sebagai berikut:

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

$\mu$  = Mean

$\sigma$  = Standard Deviation

5. Tahapan kelima adalah model klasifikasi. Pada tahap ini menentukan k atau jumlah tetangga terdekat untuk menentukan kelas dari sebuah data baru.
6. Pada tahapan terakhir adalah Evaluasi. Evaluasi dilakukan untuk menguji apakah algoritma atau system yang digunakan. Evaluasi terhadap akurasi sistem diperlukan untuk mengetahui tingkat ketepatan dalam proses klasifikasi data. Salah satu pendekatan yang banyak diterapkan dalam proses evaluasi ini adalah *confusion matrix*, yang berfungsi untuk memetakan hasil klasifikasi sistem terhadap nilai klasifikasi yang sebenarnya, sehingga memungkinkan analisis kesalahan dan performa model secara menyeluruh. (Sholikhin & Atmojo, 2022).

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

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<b>TP</b> (True Positive)	<b>FP</b> (False Positive) <i>Type I Error</i>
	0 (Negative)	<b>FN</b> (False Negative) <i>Type II Error</i>	<b>TN</b> (True Negative)

Dimana:

TP (True Positive) = data positif di golongkan dengan benar oleh sistem.

TN (True Negative) = data negative di digolongkan benar oleh sistem.

FP (False Positive) = data negative namun di digolongkan salah oleh sistem.

FN (False Negative) = data positif namun di golongkan salah oleh sistem.

## HASIL DAN PEMBAHASAN

Hasil

### 1. Pengumpulan Data

Penulis memperoleh data sekunder melalui situs web basis data *Kaggle*. Total data yang digunakan dalam penelitian ini sebanyak 100.000 data. Berikut ditampilkan sebagian data sebagai ilustrasi, sedangkan data lengkapnya dapat diakses melalui tautan berikut:

<https://www.kaggle.com/datasets/jabirmuktabir/stunting-wasting-dataset>

Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Stunting	Wasting
Laki-laki	19	91,6	13,3	Tall	Risk of Overweight
Laki-laki	20	77,7	8,5	Stunted	Underweight
Laki-laki	10	79	10,3	Normal	Risk of Overweight
Perempuan	2	50,3	8,3	Severely Stunted	Risk of Overweight
Perempuan	5	56,4	10,9	Severely Stunted	Risk of Overweight
Laki-laki	11	76,3	5,9	Normal	Severely Underweight
Laki-laki	16	80,7	9,9	Normal	Normal weight
Perempuan	15	72,6	6,5	Normal	Severely Underweight
Perempuan	18	78,4	15,6	Normal	Risk of Overweight
Laki-laki	2	63,4	7	Normal	Risk of Overweight
Laki-laki	6	60,4	11,5	Severely Stunted	Risk of Overweight
Perempuan	17	77,5	6,2	Normal	Severely Underweight
Perempuan	19	76,6	12,9	Normal	Risk of Overweight
Perempuan	7	59,8	11,4	Severely Stunted	Risk of Overweight
Perempuan	17	82,8	9	Normal	Normal weight
Perempuan	21	84,7	8,4	Normal	Underweight
Laki-laki	17	86,2	14,7	Normal	Risk of Overweight

Perempuan	3	61,2	9	Normal	Risk of Overweight
Perempuan	13	69,8	5,8	Stunted	Severely Underweight
Perempuan	20	80,5	9,2	Normal	Normal weight
Laki-laki	23	84,1	10	Normal	Normal weight
Perempuan	12	67,8	12,4	Stunted	Risk of Overweight
Laki-laki	24	97,4	14,6	Tall	Risk of Overweight
Perempuan	7	59,4	10,7	Severely Stunted	Risk of Overweight
Laki-laki	7	72,6	10,7	Normal	Risk of Overweight
Laki-laki	9	69,8	6,3	Normal	Severely Underweight
Laki-laki	8	72,5	7,7	Normal	Normal weight
Perempuan	22	80,7	10,1	Normal	Normal weight
Laki-laki	11	76,7	12,8	Normal	Risk of Overweight
Laki-laki	2	53.1	7.2	Stunted	Risk of Overweight
Laki-laki	4	64.7	9.1	Normal	Risk of Overweight
Laki-laki	8	73.8	6	Normal	Severely Underweight
Perempuan	12	68.8	13.2	Stunted	Risk of Overweight
Laki-laki	19	84.4	13	Normal	Risk of Overweight

Laki-laki	22	88,1	12,2	Normal	Normal weight
Laki-laki	3	67,8	7,4	Tall	Risk of Overweight
Laki-laki	21	92,5	12,7	Normal	Risk of Overweight
Perempuan	24	77,6	11,7	Stunted	Normal weight
Perempuan	11	71,3	12,8	Normal	Risk of Overweight
Perempuan	22	92,8	10	Normal	Normal weight
Laki-laki	0	46,3	3,2	Normal	Normal weight
Perempuan	12	80,3	5,4	Normal	Severely Underweight
Perempuan	8	68,5	10,1	Normal	Risk of Overweight
Perempuan	11	68,4	6,6	Normal	Underweight
Perempuan	23	89,8	10,6	Normal	Normal weight
Laki-laki	19	78,6	13	Normal	Risk of Overweight
Laki-laki	16	74,1	15	Stunted	Risk of Overweight
Laki-laki	16	88,6	10	Tall	Normal weight
Laki-laki	4	61,6	8,2	Normal	Risk of Overweight
Perempuan	1	54,1	4,2	Normal	Normal weight
....	....	....	....	....	....

## 2. Pra-Processing Data

Adapun pre-processing pada penelitian ini adalah sebagai berikut:

```
# Hapus fitur 'Wasting' yang tidak diperlukan
if 'Wasting' in df.columns:
    df.drop(columns=['Wasting'], inplace=True)
```

a) Menghapus kolom yang tidak digunakan yaitu kolom 'Wasting'

Maka akan dihasilkan table sebagai berikut :

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	\
0	Laki-laki	19	91.6	13.3	
1	Laki-laki	20	77.7	8.5	
2	Laki-laki	10	79.0	10.3	
3	Perempuan	2	50.3	8.3	
4	Perempuan	5	56.4	10.9	
	Stunting				
0	Tall				
1	Stunted				
2	Normal				
3	Severely Stunted				
4	Severely Stunted				

b) Encoding 'Jenis Kelamin' dan Status 'Stunting' ke dalam data Numerik Pada penelitian ini pra-processing meliputi konversi data. Konversi data yang dimaksud adalah mengonversi data kategori ke data numerik. Hal ini perlu dilakukan karena K-NN menggunakan rumus

```
# Encode 'Jenis Kelamin' dan 'Stunting'
df['Jenis Kelamin'] = le_gender.fit_transform(df['Jenis Kelamin']) # Encode Laki-laki/Perempuan
df['Stunting'] = le_stunting.fit_transform(df['Stunting']) # Encode label Stunting

# Menampilkan mapping hasil encoding
print("\nMapping Jenis Kelamin:")
for index, label in enumerate(le_gender.classes_):
    print(f"{label} -> {index}")

print("\nMapping Stunting:")
stunting_mapping = {index: label for index, label in enumerate(le_stunting.classes_)}
for index, label in stunting_mapping.items():
    print(f"{label} -> {index}")

# Lihat dataset setelah encoding
print("\nDataset (setelah encoding):")
print(df.head())
```

euclidian distance.(Malikussaleh, 2024)

Maka didapatkan hasil:

```

Mapping Jenis Kelamin:
Laki-laki -> 0
Perempuan -> 1

Mapping Stunting:
Normal -> 0
Severely Stunted -> 1
Stunted -> 2
Tall -> 3

Dataset (setelah encoding):

```

	Jenis Kelamin	Umur (bulan)	Tinggi Badan (cm)	Berat Badan (kg)	Stunting
0	0	19	91.6	13.3	3
1	0	20	77.7	8.5	2
2	0	10	79.0	10.3	0
3	1	2	50.3	8.3	1
4	1	5	56.4	10.9	1

### 3. Splitting Data

Splitting data membagi menjadi 2 subset data yaitu data latih (*training set*) dan data uji (*testing set*). 80% data untuk melatih model (train), 20% untuk menguji model (test). Random-state nya adalah 42 agar hasil pembagian data tetap sama setiap kali dijalankan.

```

# Split data untuk training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

```

### 4. Ekstraksi Fitur

Normalisasi data proses mengubah skala nilai fitur agar semua fitur memiliki rentang yang seragam. Ini penting karena K-Nearest Neighbor perlu menggunakan jarak antar data untuk membuat keputusan. Jika skala tidak seimbang, fitur dengan nilai besar bisa mendominasi.

```

# Normalisasi fitur
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

```

### 5. Model Klasifikasi

Pada K-Nearest Neighbor akan ditentukan terlebih dahulu jumlah k, dimana k adalah jumlah tetangga terdekat. Data baru akan dihitung jarak nya dengan semua data latih. Hal ini disebut Euclidian Distance, dimana ini adalah cara KN menentukan sebuah jarak.

```
# Membuat dan melatih model KNN
k = 3 # Pilih jumlah tetangga
knn = KNeighborsClassifier(n_neighbors=k)
knn.fit(X_train_scaled, y_train)
```

## 6. Evaluasi Model

Dalam tahap evaluasi, penulis menggunakan 20.000 data uji guna mengetahui seberapa baik model dalam melakukan klasifikasi. Pengujian ini menghasilkan metrik performa yang mencakup *accuracy*, *precision*, *recall*, serta *f1-score*.

```
# Evaluasi model
y_pred = knn.predict(X_test_scaled)
class_counts = df['Stunting'].value_counts()
print("\nJumlah data di setiap kelas:")
print(class_counts)
```

```
# Menghitung metrik evaluasi
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
conf_matrix = confusion_matrix(y_test, y_pred)
```

```
# Menampilkan hasil evaluasi
print(f"\nEvaluasi Model dengan")
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")
print("Confusion Matrix:")
print(conf_matrix)

print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

Maka didapatkan hasil sebagai berikut:

```

Evaluasi Model dengan
Accuracy: 0.97
Precision: 0.97
Recall: 0.97
F1-Score: 0.97
Confusion Matrix:
[[14198  0  97  111]
 [  0 1033 113  0]
 [ 106 113 3073  0]
 [ 137  0  0 1019]]

Classification Report:

```

	precision	recall	f1-score	support
0	0.98	0.99	0.98	14406
1	0.90	0.90	0.90	1146
2	0.94	0.93	0.93	3292
3	0.90	0.88	0.89	1156
accuracy			0.97	20000
macro avg	0.93	0.93	0.93	20000
weighted avg	0.97	0.97	0.97	20000

## Pembahasan

Pada penelitian ini, penulis menerapkan metode *K-Nearest Neighbors* (KNN) yang dievaluasi menggunakan *confusion matrix* sebagai alat ukur kinerja klasifikasi. Proses pengujian dilakukan dengan menggunakan 20% dari total data yang tersedia, yaitu sebanyak 20.000 data uji dari keseluruhan 100.000 data. Sementara itu, 80% sisanya, yaitu sebanyak 80.000 data, digunakan sebagai data latih (*training data*) untuk melatih model agar dapat mengenali pola dan melakukan klasifikasi dengan lebih akurat. Pada hasil pengujian confusion matrix didapatkan 4 (empat)

Model *K-Nearest Neighbors* (KNN) yang diterapkan menunjukkan performa yang sangat baik, dengan nilai *Accuracy* sebesar 97%. Hal ini menunjukkan bahwa model mampu mengklasifikasikan data uji dengan tingkat ketepatan yang tinggi secara keseluruhan. Selain *Accuracy*, evaluasi juga dilakukan menggunakan metrik *Precision*, *Recall*, dan *F1-score*. Ketiga metrik tersebut masing-masing memperoleh nilai sebesar 0,97, yang menunjukkan bahwa model tidak hanya akurat, tetapi juga seimbang dalam mengenali kelas positif maupun negatif. Nilai *precision* yang tinggi menandakan bahwa prediksi positif yang dihasilkan oleh model sebagian besar benar, sedangkan nilai *recall* yang tinggi mengindikasikan bahwa sebagian besar data aktual yang tergolong positif berhasil dikenali oleh model. Nilai *F1-score* yang tinggi menunjukkan keseimbangan antara *precision* dan *recall*.

Lebih lanjut, laporan klasifikasi (*classification report*) memperlihatkan bahwa nilai *Precision*, *Recall*, dan *F1-score* pada setiap kelas berada di atas 0,90. Kelas 0 memiliki nilai *precision* tertinggi, yaitu 0,98, dan *recall* sebesar 0,99, yang menunjukkan bahwa model sangat efektif dalam mengenali kelas tersebut. Sementara itu, nilai rata-rata makro (*macro*

*average*) dan rata-rata tertimbang (*weighted average*) dari seluruh metrik berada di kisaran 0,93 hingga 0,97, yang mengindikasikan bahwa model memiliki performa yang stabil dan tidak bias terhadap kelas dengan jumlah data yang lebih besar maupun lebih kecil.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model KNN yang dibangun mampu melakukan klasifikasi dengan tingkat akurasi dan konsistensi yang sangat baik. Oleh karena itu, model ini dapat dianggap layak untuk digunakan dalam tugas klasifikasi data yang serupa, khususnya dalam konteks identifikasi status stunting pada data antropometri anak. Meskipun demikian, untuk memastikan keandalan model secara umum, disarankan untuk melakukan evaluasi lanjutan melalui teknik validasi silang (*cross-validation*) atau pengujian terhadap data uji eksternal yang belum pernah dilibatkan dalam proses pelatihan model.

## SIMPULAN

Metode K-NN mengelompokkan data gejala stunting pada balita berdasarkan kemiripan karakteristik. (Azis et al., 2024). Penelitian ini menerapkan algoritma klasifikasi K-Nearest Neighbors (KNN) untuk mengklasifikasikan status stunting pada anak berdasarkan data antropometri yang mencakup jenis kelamin, umur, tinggi badan, dan berat badan. Melalui serangkaian proses, mulai dari pra-pemrosesan data, normalisasi, pemisahan data latih dan data uji, hingga evaluasi performa model, diperoleh hasil yang menunjukkan bahwa model KNN mampu memberikan kinerja yang sangat baik. Model menghasilkan *Accuracy* sebesar 97%, serta nilai *Precision*, *Recall*, dan *F1-score* yang konsisten tinggi, yaitu masing-masing sebesar 0,97. Hasil ini mengindikasikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga mampu melakukan klasifikasi secara seimbang pada seluruh kelas target.

Meskipun demikian, peneliti menyadari adanya beberapa keterbatasan, terutama pada cakupan atribut yang digunakan. Oleh karena itu, untuk penelitian selanjutnya disarankan untuk memperluas jenis data yang digunakan dengan menambahkan atribut tambahan seperti lingkaran kepala, lingkaran lengan atas, atau variabel antropometri lainnya guna lebih meningkatkan akurasi dan ketepatan prediksi status stunting pada anak.

## DAFTAR PUSTAKA

Akbarollah, M. F., Wiyanto, W., Ardiatma, D., & Zy, A. T. (2023). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung. *Journal of Computer System and Informatics (JoSYC)*, 4(4), 850–860. <https://doi.org/10.47065/josyc.v4i4.4071>

- Aminah, S. T., Ghani, D., Salman, N., Liklikwatil, R. D., & Hajratul, S. (2024). *APLIKASI PREDIKSI STATUS PERTUMBUHAN BALITA MENGGUNAKAN ALGORITMA*. 2(1), 31–42.
- Azis, M. F., Kaesmetan, Y. R., Kupang, S. U., I, J. P. K., Putih, K., & Oebobo, K. (2024). *Penerapan K-NN ( K-Nearest Neighbors ) Pada Sistem Pakar Diagnosa Gejala Stunting Pada Balita Menggunakan Naïve Bayes Classifier*. 1(1), 75–91.
- Dayera, Musa Bundaris Palungan, F. O. (2024). G-Tech : Jurnal Teknologi Terapan. *G-Tech : Jurnal Teknologi Terapan*, 8(1), 186–195. <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- Dwinanto, R. W., A, A. S. S., & Ardianto, R. (2024). *KLASIFIKASI BERISIKO STUNTING PADA BALITA : PERBANDINGAN K- NEAREST NEIGHBOR , NAÏVE BAYES , SUPPORT VECTOR MACHINE*. 8(2), 264–273.
- Fauziah, J., Trisnawati, K. D., Rini, K. P. S., & Putri, S. U. (2023). Stunting: Penyebab, Gejala, dan Pencegahan. *Jurnal Parenting Dan Anak*, 1(2), 11. <https://doi.org/10.47134/jpa.v1i2.220>
- Hastuti, W., & Dulame, I. M. (2024). Penyuluhan Masalah Stunting Terkait Pencegahan Masalah Dan Masa Depan Indonesia. *Jurnal Abdi Citra*, 1(2), 130–136. <https://doi.org/10.62237/jac.v1i2.151>
- Loka, S. K. P., & Marsal, A. (2023). Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 8–14. <https://doi.org/10.57152/malcom.v3i1.474>
- Malikussaleh, U. (2024). *SENASTIKA Universitas Malikussaleh*. 1–6.
- Ramadhani, D. H. (2024). *Implementasi Algoritma K-Nearest Neighbors ( KNN ) untuk Prediksi Gizi Buruk*. 14(2), 326–336.
- Sholikhin, N. A., & Atmojo, S. (2022). Aplikasi Web Untuk Klasifikasi Stunting Pada Balita Dengan Menggunakan Metode K-Nearest Neighbours (Studi Kasus Posyandu Jawa Kidul). *Journal of System Engineering and Technological Innovation (JISTI)*, 1(02), 44–47. <https://doi.org/10.38156/jisti.v1i02.23>