

## Kombinasi Algoritma *Naive Bayes* dan *K-Means* dalam Penentuan Status Gizi Stunting

### *Combining Naive Bayes Algorithm and K-Means to determine Stunting Nutritional Status*

Ariastuti Rahman\*<sup>1</sup>, Astuti Abdullah <sup>2</sup>, Salmawati<sup>3</sup>, Nurmarifah<sup>4</sup>

<sup>1,3,4</sup>Sistem Informasi, Universitas Al Asyariah Mandar, Indonesia

<sup>2</sup>Kesehatan Masyarakat, Universitas Al Asyariah Mandar, Indonesia

E-mail: \*<sup>1</sup>ariastuti.rahman@gmail.com

---

<b>Article Info:</b>	Received 07 Mei 2026	Revised 08 Mei 2026	Accepted 29 Juni 2026	Published: 29 Juni 2026
----------------------	-------------------------	------------------------	--------------------------	----------------------------

---

#### **Abstrak**

Stunting tetap menjadi masalah kesehatan masyarakat yang kritis, ditandai dengan malnutrisi kronis pada anak di bawah lima tahun, yang menyebabkan gangguan pertumbuhan, peningkatan morbiditas, dan risiko kematian yang lebih tinggi. Di Indonesia, khususnya di Kabupaten Polewali Mandar, prevalensi stunting yang tinggi memerlukan pendekatan yang lebih akurat dan berbasis data untuk mendukung identifikasi dan intervensi dini. Namun, studi yang ada sebagian besar berfokus pada pendekatan metode tunggal dan jarang memberikan bukti komparatif antara teknik machine learning supervised and unsupervised untuk klasifikasi stunting, sehingga membatasi kokoh sistem pendukung keputusan. Studi ini bertujuan untuk mengevaluasi dan membandingkan efektivitas teknik klasifikasi dan pengelompokan dalam menentukan status gizi stunting dan untuk mengusulkan pendekatan pendukung keputusan berbasis data bagi praktisi kesehatan. Penelitian ini menggunakan pendekatan data mining dengan mengintegrasikan Naive Bayes untuk klasifikasi dan K-Means untuk pengelompokan, menggunakan indikator antropometri dari dataset 173 anak di bawah lima tahun. Kinerja kedua metode dievaluasi menggunakan metrik akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa Naive Bayes secara signifikan mengungguli K-Means, mencapai accuracy 97%, precision 100%, dan recall 94%, dibandingkan dengan 54%, 58%, dan 60% secara berturut-turut. Temuan ini menyoroti keunggulan klasifikasi probabilistik dibandingkan pengelompokan tanpa pengawasan dalam konteks ini dan menunjukkan potensi Naive Bayes sebagai metode yang andal untuk mendukung klasifikasi stunting. Studi ini berkontribusi dengan mengatasi kesenjangan penelitian melalui analisis komparatif langsung dari pendekatan supervised dan unsupervised, menawarkan wawasan empiris dan landasan praktis untuk mengembangkan sistem pendukung keputusan cerdas di bidang kesehatan masyarakat.

**Kata Kunci:** K-Means, Naive Bayes, Stunting

#### **Abstract**

Stunting remains a critical public health issue characterized by chronic malnutrition in children under five, leading to impaired growth, increased morbidity, and higher mortality risk. In Indonesia, particularly in Polewali Mandar Regency, the prevalence of stunting necessitates more accurate and data-driven approaches to support early identification and intervention. However, existing studies predominantly focus on single-method approaches and rarely provide comparative evidence between supervised and unsupervised machine learning techniques for stunting classification, limiting the robustness of decision support systems. This study aims to evaluate and

compare the effectiveness of classification and clustering techniques in determining stunting nutritional status and to propose a data-driven decision support approach for health practitioners. This research employs a data mining approach by integrating the Naive Bayes algorithm for classification and K-Means for clustering, using anthropometric indicators from a dataset of 173 children under five. The performance of both methods is evaluated using accuracy, precision, and recall metrics. The results indicate that Naive Bayes significantly outperforms K-Means, achieving accuracy 97%, precision 100%, and recall 94%, compared to 54%, 58%, and 60%, respectively. These findings highlight the superiority of probabilistic classification over unsupervised clustering in this context and demonstrate the potential of Naive Bayes as a reliable method for supporting stunting classification. This study contributes by addressing the research gap through a direct comparative analysis of supervised and unsupervised learning approaches, offering empirical insights and a practical foundation for developing intelligent decision support systems in public health.

**Keywords:** K-Means, Naïve Bayes, Stunting.

This is an open access article under the CC BY-SA license.



## 1. PENDAHULUAN

Stunting telah dikenal sebagai salah satu permasalahan kesehatan yang dibahas ditingkat global dan termasuk dalam fokus utama *World Health Organization* (WHO) karena berdampak signifikan terhadap kualitas sumber daya manusia di masa depan. Stunting yang didefinisikan sebagai kondisi gagal tumbuh pada anak usia di bawah lima tahun akibat kekurangan gizi kronis dalam jangka panjang, yang dapat mengakibatkan gangguan metabolisme [1], [2] dan gangguan perkembangan fisik dan kognitif [3], [4]. Dampak stunting tidak hanya bersifat jangka pendek seperti peningkatan morbiditas, keterlambatan perkembangan motorik, dan gangguan kognitif, tetapi juga berdampak jangka panjang seperti penurunan produktivitas, peningkatan risiko penyakit tidak menular, serta siklus kemiskinan antar generasi [5].

Penentuan status gizi stunting dilakukan melalui pendekatan teori antropometri, yaitu pengukuran dimensi tubuh manusia seperti tinggi badan menurut umur (TB/U), berat badan menurut umur (BB/U), dan indikator lainnya yang mengacu pada standar WHO Child Growth Standards [4] [6]. Selain itu, faktor sosial ekonomi, lingkungan, dan pola asuh juga berperan penting dalam menentukan status gizi anak [7], [8]. Namun demikian, pendekatan konvensional yang dilakukan tenaga kesehatan masih bersifat manual dan deskriptif, sehingga kurang optimal dalam menangkap pola kompleks dari data multidimensi.

Seiring dengan perkembangan teknologi, pendekatan berbasis teori *machine learning* dalam data mining mulai banyak digunakan dalam bidang kesehatan masyarakat termasuk untuk prediksi status gizi dan stunting pada balita menggunakan data antropometri maupun faktor sosial ekonomi [2], [6], [9]. *Machine learning* memungkinkan sistem untuk belajar dari data historis guna menghasilkan prediksi atau klasifikasi secara otomatis [2], [10]. Penggunaan *machine learning* untuk melakukan analisis gizi pada anak dapat meningkatkan akurasi prediksi [11], [12]. Selain termasuk *undernutrition* pada balita dibandingkan metode statistik tradisional seperti regresi logistik [13].

Dalam konteks metode, algoritma klasifikasi seperti *Naïve Bayes* didasarkan pada teori probabilitas Bayes yang mengasumsikan independensi antar variabel dan dikenal memiliki kinerja tinggi pada dataset dengan dimensi besar serta efisiensi komputasi yang baik [14], [15]. Sementara itu, algoritma K-Means merupakan metode klusterisasi berbasis partisi yang bertujuan mengelompokkan data ke dalam beberapa cluster berdasarkan kedekatan jarak (*distance-based clustering*), yang banyak digunakan dalam eksplorasi pola data kesehatan [16].

Beberapa penelitian terkini menunjukkan bahwa algoritma *Naïve Bayes* mampu memberikan akurasi tinggi dalam klasifikasi status gizi anak [17], sedangkan K-Means efektif dalam mengidentifikasi pola kelompok risiko stunting berdasarkan karakteristik data [16]. Namun demikian, terdapat kesenjangan pada kurangnya studi yang secara langsung membandingkan kinerja metode *supervised learning* (klasifikasi) dan *unsupervised learning* (klasterisasi) dalam konteks penentuan status gizi stunting secara empiris dan terintegrasi [11], [12].

Berdasarkan hal tersebut, tujuan penelitian ini adalah untuk membandingkan kinerja algoritma *Naïve Bayes* (klasifikasi) dan K-Means (klasterisasi) dalam menentukan status gizi stunting pada balita berdasarkan indikator antropometri, serta mengevaluasi metode yang paling efektif sebagai dasar pengembangan sistem pendukung keputusan di bidang kesehatan masyarakat. Secara metodologis, penelitian ini menggunakan pendekatan data mining dengan menerapkan algoritma *Naïve Bayes* untuk klasifikasi dan K-Means untuk klasterisasi pada data antropometri balita, kemudian mengevaluasi kinerja model menggunakan metrik akurasi, *precision*, dan *recall*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi empiris dalam pemilihan metode machine learning yang optimal untuk deteksi stunting serta mendukung pengambilan keputusan berbasis data di sektor kesehatan.

## 2. METODE

### 2.1 Pendekatan penelitian

Penelitian ini menggunakan pendekatan data mining berbasis machine learning dengan algoritma *Naïve Bayes* untuk klasifikasi status gizi dan K-Means untuk klasterisasi data antropometri. Kinerja kedua metode dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan Receiver Operating Characteristic (ROC) untuk menentukan metode yang paling efektif dalam identifikasi status gizi stunting.

### 2.2. Data Penelitian

Sebanyak 173 data balita digunakan dan status gizi dilihat berdasarkan indikator antropometri menggunakan parameter umur, berat badan dan tinggi badan. Data akan dikelompokkan kedalam dua cluster, cluster 1 (Tidak beresiko) dan cluster 2 (Beresiko stunting) menggunakan metode K-Means *clustering* sementara metode *Naive Bayes* melakukan dua klasifikasi yaitu Presence (Beresiko stunting), Absence (Tidak beresiko).

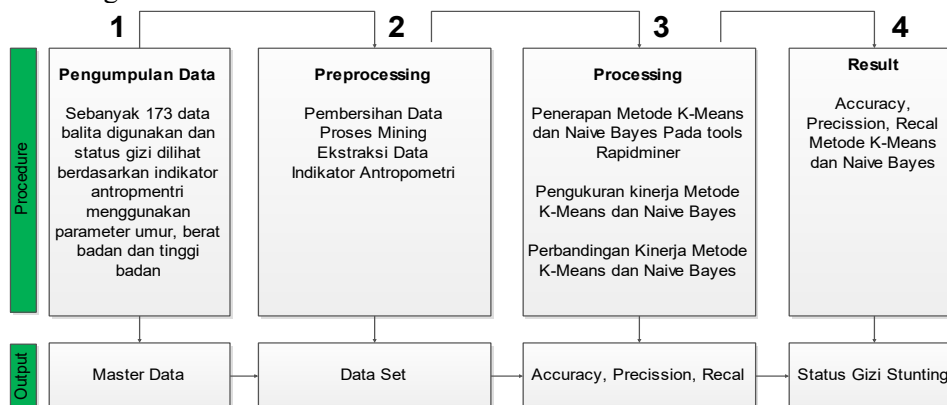
Proses klasifikasi algoritma *Naïve Bayes* menentukan data latih dan data uji menggunakan metode split validation dengan rasio 60:40. Sebanyak 103 data digunakan sebagai data latih dan 70 data digunakan sebagai data uji. Sementara itu, algoritma K-Means menggunakan seluruh data. Status gizi stunting terjadi jika pertumbuhannya berada pada -2 sampai < -3 (Berat badan: kurang, tinggi badan; pendek), dan dibawah < -3 (Berat badan: sangat kurang, tinggi badan: sangat pendek) selain dari kondisi tersebut maka termasuk normal atau tidak beresiko stunting [3]. Pengukuran antropometri, status gizi ideal berdasarkan pada ambang batas nilai z-score pada tabel 1 [18]. Tabel 1 menunjukkan standar pengukuran antropometri, status gizi ideal berdasarkan pada ambang batas nilai z-score [18] sebagai berikut:

Tabel 1. Batas ambang nilai z-score dalam status gizi anak [18]

Indeks	Kategori status gizi anak	Ambang batas (z-score)
Berat Badan Menurut Umur(BB/U) Anak umur 0-60 bulan	Berat badan sangat kurang Berat badan kurang Berat badan normal Resiko berat badan lebih	< -3 SD -3 sampai dengan < -2 SD -2 sampai dengan +1SD > +1 SD
Panjang Badan Menurut Umur (PB/U) atau Tinggi Badan Menurut Umur (TB/U) Anak umur 0-60 bulan	Sangat pendek Pendek Normal Tinggi	< -3 SD -3 sampai dengan < -2 SD -2 sampai dengan +3 SD > +3 SD

### 2.3. Tahapan Penelitian

Adapun tahapan yang menunjukkan prosedur penelitian yang dikerjakan disajikan pada Gambar 1. sebagai berikut:



Gambar 1. Prosedur Penelitian

Berdasarkan Gambar 1, penelitian dilakukan melalui tahapan pengumpulan data, *preprocessing*, *processing*, dan evaluasi hasil. Data antropometri balita diproses menggunakan algoritma *Naïve Bayes* untuk klasifikasi dan K-Means untuk clustering pada aplikasi RapidMiner [19]. Selanjutnya, kinerja kedua metode dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan ROC untuk menentukan metode yang paling efektif dalam identifikasi status gizi stunting [19].

### 2.4. Naïve Bayes Klasifikasi

*Naïve Bayes* merupakan metode klasifikasi probabilistik berbasis Teorema Bayes yang efisien pada dataset kecil maupun besar dan banyak digunakan pada klasifikasi kesehatan [11], [15], [35]. Secara matematis dirumuskan sebagai:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(x)} \quad (1)$$

Dimana:

$P(H|X)$  = probabilitas posterior

$P(H)$  = prior

$P(X|H)$  = likelihood

$P(X)$  = evidence

Model ini akan memilih kelas dengan probabilitas maksimum  $H^* = \arg \max P(H|X)$

### 2.5. K-Means Klustering

Algoritma K-Means mengelompokkan data berdasarkan kedekatan jarak menggunakan pendekatan *centroid-based clustering*. Proses dimulai dengan menentukan jumlah cluster dan centroid awal, kemudian menghitung jarak *Euclidean* untuk mengelompokkan data hingga centroid konvergen.

1. Menentukan jumlah cluster K dan Inisialisasi centroid awal
2. Menghitung jarak menggunakan *Euclidean Distance*:

$$\sqrt{\sum (X_i - \mu_j)^2} \quad (2)$$

3. Mengelompokkan data ke cluster terdekat kemudian memperbarui centroid:

$$\mu_j^{(t+1)} = \frac{1}{N_s} \sum x \quad (3)$$

4. Iterasi hingga konvergen

K-Means banyak digunakan dalam analisis data kesehatan untuk menemukan pola tersembunyi dan segmentasi risiko.

## 2.6. Evaluasi Kinerja Metode

Evaluasi model machine learning pada penelitian kesehatan umumnya menggunakan confusion matrix, *accuracy*, *precision*, *recall*, ROC, dan AUC untuk mengukur kemampuan prediksi model [20]. Evaluasi dilakukan menggunakan confusion matrix merujuk padata Tabel 2. [12], [21]:

**Tabel 2.** Confusion Matrix

	Aktual Positif	Aktual Negatif
Prediksi Positif	TP	FP
Prediksi Negatif	FN	TN

*Accuracy* menggunakan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

*Precision* menggunakan:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

*Recall* menggunakan:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

## 2.7. Analisis ROC

Kurva ROC (*Receiver Operating Characteristic*) digunakan untuk mengevaluasi kemampuan model dalam membedakan kelas *presence* dan *absence* berdasarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). Semakin mendekati sudut kiri atas, maka semakin baiki performa model [20], [22]. Sebaliknya, kurva yang mendekati garis diagonal 45° (FPR = TPR) menunjukkan performa yang mendekati *random classifier* atau kurang akurat. Selain itu, *Confusion Matrix* juga digunakan untuk membandingkan hasil prediksi dengan kondisi aktual . Melalui *Confusion Matrix* dapat dihitung nilai *accuracy*, *precision*, dan *recall*.

## 3. HASIL

### 3.1 Representasi Dataset

Penelitian ini menggunakan 173 data balita yang diperoleh dari puskesmas berdasarkan indikator antropometri berupa umur, berat badan, dan tinggi badan. Dataset terdiri dari 98 data kategori *absence* dan 75 data kategori *presence*. Distribusi data tersebut digunakan sebagai dasar proses klasifikasi menggunakan algoritma *Naïve Bayes* dan clustering menggunakan algoritma K-Means.

### 3.2. Hasil Naïve Bayes Klasifikasi

#### 3.2.1. Data Atribut dan Kategori

Tahap awal klasifikasi dilakukan dengan mengubah atribut numerik menjadi kategori diskrit untuk mempermudah proses probabilistik pada algoritma *Naïve Bayes*.

**Tabel 3.** Kategori Atribut

Atribut	Kategori
Usia	1. 0-23 Bulan = baduta (anak usia dibawah dua tahun)
	2. 23-59 Bulan = balita(anak usian dibawah lima tahun)
Berat Badan	1. Berat badan sangat kurang
	2. Berat badan kurang
Tinggi Badan	1. Sangat pendek
	2. Pendek

Klasifikasi status gizi mengacu pada standar antropometri Kementerian Kesehatan RI berdasarkan nilai z-score.

### 3.2.2. Hasil Klasifikasi

Hasil klasifikasi menghasilkan dua kelas atribut status gizi yaitu kategori status *Absence* dan *Presence*. Adapun sistem keputusan yang telah diproses disajikan pada tabel 4.

Tabel 4. Hasil Sistem Keputusan *Naïve Bayes*

NAMA	UMUR	BB	TB	STATUS
Balita 1	Baduta	Berat Badan Normal	Pendek	Presence
Balita 2	Baduta	Berat Badan Normal	Pendek	Presence
Balita 3	Baduta	Berat Badan Normal	Normal	Absence
Balita 4	Baduta	Berat Badan Normal	Sangat Pendek	Presence
Balita 5	Baduta	Berat Badan Normal	Normal	Absence

Tabel 4 menampilkan 5 data sebagai representasi dataset. Selanjutnya dilakukan perhitungan probabilitas setiap atribut antropometri.

### 3.2.3. Analisis Probabilitas

Setelah proses klasifikasi dilakukan, probabilitas masing-masing atribut dihitung menggunakan Teorema Bayes. Adapun tabel 5 menyajikan probabilitas atribut antropometri sebagai berikut:

Tabel 5. Nilai Probabilitas Atribut Antropometri

Atribut	Kategori	<i>Absence</i>	<i>Presence</i>
Usia	Baduta	38%	40%
	Balita	62%	60%
Berat Badan	Berat Badan Sangat Kurang	0%	3%
	Berat Badan Kurang	2%	25%
	Berat Badan Normal	97%	69%
	Resiko Berat Badan Lebih	1%	3%
	Sangat Pendek	0%	21%
Tinggi Badan	Pendek	0%	53%
	Normal	99%	25%
	Tinggi	1%	0%
Total		100%	100%

Tabel 5 menunjukkan distribusi probabilitas setiap atribut antropometri terhadap kelas absence dan presence yang digunakan sebagai dasar perhitungan posterior pada algoritma *Naïve Bayes*.

### 3.2.4. Pengujian *Naïve Bayes*

Pengujian dilakukan menggunakan *probabilitas posterior* berdasarkan kombinasi atribut antropometri. Berdasarkan nilai probabilitas pada Tabel 5, probabilitas kelas *absence* dan *presence* dihitung sebagai contoh proses klasifikasi satu data uji:

$$Absence = 38\% \times 97\% \times 0\% = 0$$

$$Presence = 40\% \times 69\% \times 53\% = 0.146$$

Karena probabilitas kelas *presence* lebih tinggi dibandingkan *absence*, maka balita tersebut diklasifikasikan sebagai balita stunting (*presence*). Secara keseluruhan, hasil pengujian data balita disajikan pada Tabel 6.

Tabel 6. Hasil Uji Data Balita

Nama	<i>Absence</i>	<i>Presence</i>	Prediksi Kelas
Balita 1	0.00%	6.40%	Presence
Balita 2	0.00%	6.40%	Presence
Balita 3	20.50%	3.00%	Absence
Balita 4	20.50%	3.00%	Absence
Balita 5	20.50%	3.00%	Absence

### 3.3. Hasil K-Means Clustering

#### 3.3.1. Centroid Awal

Proses clustering dimulai dengan menentukan centroid awal secara acak untuk Inisiasi nilai pusat cluster sebagaimana disajikan pada tabel 7.

Tabel 7. Centroid Awal

Klaster	Umur	BB	TB	Keterangan
C1	19	10.8	79.9	Absence
C2	20	9.6	77.2	Presence

#### 3.3.2. Proses Clustering

Pengelompokan data dilakukan menggunakan Euclidean Distance sebagaimana disajikan pada Tabel 8.

Tabel 8. Hasil Cluster dengan Jarak Terdekat

Nama	C1	C2	Jarak Terdekat	Kelompok Data
Balita 1	2.85	1.42	1.42	C2
Balita 2	3.12	0.00	0.00	C2
Balita 3	0.00	3.12	0.00	C1
Balita 4	1.99	3.16	1.99	C1
Balita 5	1.57	2.87	1.57	C1

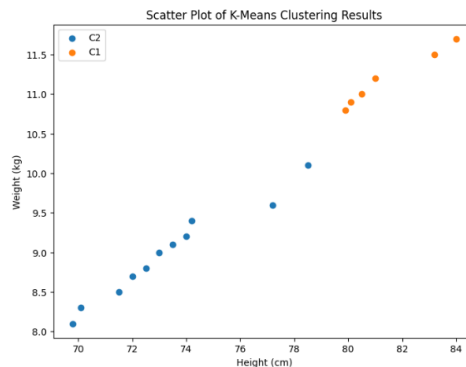
#### 3.3.3. Penentuan Centroid Baru

Setelah iterasi clustering dilakukan, centroid baru yang diperoleh disajikan pada Tabel 9.

Tabel 9. Centroid Baru

Cluster	Umur	BB	TB	Keterangan
C1	42.3	13.1	91.2	Absence
C2	13.6	8.6	71.9	Presence

Tabel 9 merupakan hasil pengelompokan menunjukkan bahwa balita di Klaster 2 (C2) cenderung memiliki nilai berat dan tinggi badan yang lebih rendah dibandingkan dengan Klaster 1 (C1).



Gambar 3. Scatter Plot Hasil Clustering K-Means

Scatter plot pada Gambar 3 memperlihatkan distribusi dua cluster berdasarkan atribut antropometri. Terlihat terdapat beberapa titik yang saling berdekatan antar cluster, yang mengindikasikan adanya overlap karakteristik data.

### 3.4. Evaluasi Kinerja Model

Hasil pengujian kedua algoritma menggunakan data yang sama menunjukkan hasil klasifikasi dan cluster yang berbeda pula. Penelitian ini menggunakan parameter *Confusion Matrix*, *Accuracy*, *Precision*, *Recall* untuk mengukur kinerja algoritma *Naïve Bayes* dan *K-Means*.

#### 3.4.1. Confusion Matrix

1. Naive Bayes dan *K-Means*

Tabel 10 merupakan tabel *Confusion Matrix* dengan menggunakan algoritma *Naive Bayes*.

**Tabel 10. Confusion Matrix Naive Bayes**

Algoritma	TP	TN	FP	FN
Naive Bayes	31	37	0	2
K-Means	57	37	41	38

Keterangan :

TP = *True Positive*      TN = *True Negative*      FP = *False Positive*      FN = *False Negative*  
 Berdasarkan Tabel 10 menunjukkan jumlah TP, TN, FP dan FN masing-masing algoritma.

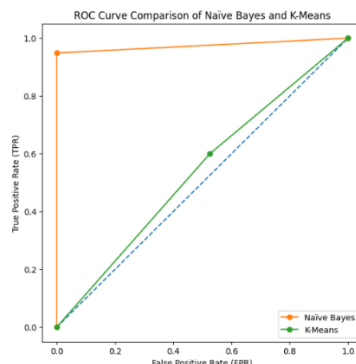
### 3.4.2. Pengujian

Perbandingan Hasil Pengujian Algoritma pada Tabel 11 dimaksudkan untuk melihat kinerja pengujian masing-masing algoritma.

Tabel 11. Perbandingan Hasil Pengujian Algoritma

Algoritma	Accuracy	Precision	Recall
Naive Bayes	97%	100%	94%
K-Means	54%	58%	60%

Perbandingan hasil pengujian algoritma disajikan pada Tabel 11. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki nilai *accuracy*, *precision*, dan *recall* yang lebih tinggi dibandingkan K-Means. Gambar 4 menunjukkan kurva ROC untuk menunjukkan seberapa baik model membedakan dua kelas.



**Gambar 4.** Kurva ROC Perbandingan Kinerja Naive Bayes dan K-Means

Berdasarkan Gambar 4, algoritma *Naive Bayes* menunjukkan kinerja klasifikasi yang lebih baik dibandingkan K-Means, ditunjukkan oleh posisi kurva yang lebih mendekati sudut kiri atas pada ruang ROC [23]. *Naive Bayes* memiliki nilai TPR sebesar 0.95 dan FPR sebesar 0.00, Nilai tersebut menunjukkan bahwa model memiliki kemampuan membedakan kelas yang lebih baik dibandingkan K-Means. Sebaliknya, K-Means memiliki nilai TPR sebesar 0.60 dan FPR sebesar 0.53, yang menunjukkan kemampuan diskriminasi kelas yang lebih rendah, dan cenderung mendekati *random classifier*.

## 4. PEMBAHASAN

Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki kinerja lebih baik dibandingkan K-Means dalam menentukan status gizi stunting, dengan nilai *accuracy* 97%, *precision* 100%, dan *recall* 94%. Kinerja tersebut diduga dipengaruhi oleh kemampuan *Naive Bayes* dalam memanfaatkan probabilitas antar atribut berdasarkan data training dan label kelas aktual, sehingga lebih efektif pada kasus klasifikasi dengan label yang jelas. Selain itu, jumlah *false positive* dan *false negative* yang lebih rendah mengindikasikan bahwa *Naive Bayes* mampu mengurangi kesalahan klasifikasi dibandingkan K-Means.

Sebaliknya, K-Means menghasilkan *accuracy* yang lebih rendah karena menggunakan pendekatan *unsupervised learning* yang mengelompokkan data berdasarkan kemiripan atribut

tanpa mempertimbangkan label kelas. Kondisi ini diduga menyebabkan beberapa data dengan karakteristik antropometri yang serupa tetapi memiliki status gizi berbeda berada pada cluster yang sama. Temuan tersebut juga didukung oleh *scatter plot* hasil clustering yang memperlihatkan adanya *overlap* antar cluster sehingga batas pemisahan kelas menjadi kurang jelas. Hasil ini mengisyaratkan bahwa K-Means masih berpotensi mengidentifikasi pola umum karakteristik antropometri, namun kurang sesuai untuk klasifikasi status gizi dibandingkan Naïve Bayes.

Interpretasi tersebut sejalan dengan penelitian Titimeidara dan Hadikurniawati [24], Tadesse *et al.* [11] serta Talukder *et al.* [12] yang menunjukkan bahwa pendekatan supervised learning memiliki kinerja lebih baik dalam klasifikasi status gizi dan prediksi kesehatan berbasis antropometri. Selain itu, atribut tinggi badan menunjukkan kontribusi paling dominan terhadap klasifikasi stunting, terutama pada kategori "pendek" dan "sangat pendek", yang konsisten dengan standar antropometri WHO [25].

Meskipun Naïve Bayes dan K-Means berasal dari pendekatan **machine learning** yang berbeda, perbandingan keduanya tetap relevan untuk mengevaluasi efektivitas pendekatan *supervised* dan *unsupervised* learning dalam identifikasi status gizi stunting [26]. Dari sisi implementasi, *Naïve Bayes* berpotensi digunakan sebagai dasar pengembangan sistem pendukung keputusan untuk deteksi dini stunting pada layanan kesehatan masyarakat.

## 5. KESIMPULAN

Penelitian ini membandingkan algoritma *Naïve Bayes* dan *K-Means* dalam klasifikasi status gizi stunting menggunakan data antropometri dari 173 balita. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki kinerja lebih baik dengan *accuracy* 97%, *precision* 100%, dan *recall* 94%, sedangkan K-Means memperoleh *accuracy* 54%, *precision* 58%, dan *recall* 60%. Temuan ini menunjukkan bahwa pendekatan supervised learning lebih efektif dibandingkan clustering dalam identifikasi status stunting, dengan atribut tinggi badan sebagai indikator paling dominan. Metode yang diusulkan berpotensi mendukung pengembangan sistem deteksi dini stunting pada layanan kesehatan masyarakat. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menambahkan variabel non-antropometrik, serta membandingkan algoritma machine learning lainnya untuk meningkatkan keandalan model. Penelitian selanjutnya juga dapat mengembangkan pendekatan AutoML dan optimasi hyperparameter untuk meningkatkan performa klasifikasi pada data antropometri balita [25].

## KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak terdapat konflik kepentingan antara para penulis maupun dengan objek penelitian dalam makalah ini.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas Al Asyariah Mandar, Dinas Kesehatan, dan Puskesmas atas dukungan data dan fasilitas penelitian.

## DAFTAR PUSTAKA

- [1] M. R. M. Budu, "Characterizing the link between stunting and metabolic syndrome in developing nations," *Cell. Signal.*, vol. 137, p. 112193, 2026, doi: 10.1016/j.cellsig.2025.112193.
- [2] M. Islam *et al.*, "Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh," *Int. J. Cogn. Comput. Eng.*, vol. 3, no. February 2021, pp. 46–57, 2022, doi: 10.1016/j.ijcce.2022.02.002.
- [3] A. J. Prendergast and J. H. Humphrey, "The stunting syndrome in developing countries," *Paediatr. Int. Child Health*, vol. 34, no. 4, pp. 250–265, 2014, doi:

- 10.1179/2046905514Y.0000000158.
- [4] WHO (World Health Organization), *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*, vol. 66. 2018. [Online]. Available: <https://iris.who.int/handle/10665/272596>
  - [5] C. G. Victora, P. Christian, L. P. Vdaletti, G. Gatica-Domínguez, P. Menon, and R. E. Black, “Revisiting Maternal and Child Undernutrition in Low-Income and Middle-Income Countries,” *Lancet*, vol. 397, no. 10282, pp. 1388–1399, 2021, doi: 10.1016/S0140-6736(21)00394-9.
  - [6] Mundirin, Idawati, and I. Latief, “Klasifikasi Status Gizi Balita Berbasis Data Antropometri menggunakan Random Forest,” *J. Comput. Sci. Informatics Eng.*, vol. 04, no. 4, pp. 324–333, 2025, doi: 10.55537/cosie.v4i4.1202.
  - [7] B. J. Akombi, K. E. Agho, J. J. Hall, N. Wali, A. M. N. Renzaho, and D. Merom, “Stunting and Severe Stunting among Children Under-5 Years in Nigeria,” *BMC Pediatr.*, vol. 17, no. 1, pp. 1–16, 2017, doi: 10.1186/s12887-016-0770-z.
  - [8] S. M. J. Rahman *et al.*, “Investigate the Risk Factors of Stunting, Wasting, and Underweight among Under-Five Bangladeshi Children and its Prediction based on Machine Learning Approach,” *PLoS One*, vol. 16, no. 6, pp. 1–11, 2021, doi: 10.1371/journal.pone.0253172.
  - [9] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, “Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques,” *J. Prev. Med. Public Heal.*, pp. 41–49, 2023, doi: 10.3961/jpmp.22.388.
  - [10] A. Beam and I. Kohane, “Big data and machine learning in health care,” *JAMA*, vol. 325, no. 13, pp. 1317–1318, 2021, doi: 10.1001/jama.2021.3117.
  - [11] G. A. Tadesse *et al.*, “Forecasting acute childhood malnutrition in Kenya using machine learning and diverse sets of indicators,” *PLoS One*, vol. 20, no. 5, pp. 1–15, 2025, doi: 10.1371/journal.pone.0322959.
  - [12] A. Talukder and B. Ahammed, “Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh,” *Nutrition*, vol. 78, p. 110861, 2020, doi: 10.1016/j.nut.2020.110861.
  - [13] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, “Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia,” *Public Health Nutr.*, vol. 25, no. 2, pp. 269–280, 2021, doi: 10.1017/S1368980021004262.
  - [14] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. MIT Press, 2022.
  - [15] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
  - [16] L. A. F. Maharani, Purwadi, and D. U. Hidayah, “Clustering and Classification of Toddler Stunting Risk Using K-Means and Naive Bayes: A Case Study at Kembaran 1 Community Health Center,” *J. Inf. Technol. Informatics*, vol. 7, no. 2, 2026, doi: 10.52436/1.jutif.2026.7.2.5420.
  - [17] R. D. Hartana and E. I. Sela, “Nutritional Status Classification of Stunting in Toddlers Using Naive Bayes Classifier Method,” *J. Technol. Informatics Eng.*, vol. 3, no. 1, 2024, doi: 10.51903/jtie.v3i1.154.
  - [18] M. K. RI, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 Tentang Standar Antropometri Anak*, vol. 21, no. 1. Indonesia: Menteri Kesehatan, 2020, pp. 1–9. [Online]. Available: <https://peraturan.bpk.go.id/Details/152505/permenkes-no-2-tahun-2020>
  - [19] T. Prasetya, I. Ali, C. L. Rohmat, and O. Nurdiawan, “Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor,” *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 1, p. 93, 2020, doi: 10.51211/itbi.v5i1.1431.
  - [20] T. Gneiting and E. M. Walz, “Receiver operating characteristic (ROC) movies, universal

- ROC (UROC) curves, and coefficient of predictive ability (CPA),” *Mach. Learn.*, vol. 111, no. 8, pp. 2769–2797, 2022, doi: 10.1007/s10994-021-06114-3.
- [21] D. Chicco and G. Jurman, “The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation,” *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.
- [22] A. M. Carrington *et al.*, “Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, no. August, 2022, doi: 10.1109/TPAMI.2022.3145392.
- [23] P. Riesthuis and H. Otgaar, “On the use of receiver operating characteristic area under the curve in eyewitness memory research,” *Leg. Criminol Psychol*, vol. 30, no. June 2024, pp. 212–230, 2025, doi: 10.1111/lcrp.12300.
- [24] M. Y. Titimeidara and W. Hadikurniawati, “Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita,” *J. Ilm. Inform.*, vol. 9, no. 01, pp. 54–59, 2021, doi: 10.33884/jif.v9i01.3741.
- [25] Joharini and A. Subekti, “Comparative Analysis of Automated Machine Learning Methods for Multiclass Stunting Prediction Using Anthropometric Data,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 10, no. 2, pp. 991–1002, 2026, doi: 10.33395/sinkron.v10i2.15886.
- [26] E. Indrisari, H. Febiansyah, and B. Adiwino, “A Systematic Literature Review on the Application of Machine Learning for Predicting Stunting Prevalence in Indonesia ( 2020 – 2024 ),” *J. SISFOKOM (Sistem Inf. dan Komputer)*, vol. 14, no. 03, pp. 277–283, 2025, doi: 10.32736/sisfokom.v14i3.2366.