

Klasifikasi Risiko Stunting pada Balita Berdasarkan Data Kesehatan Menggunakan Algoritma Naïve Bayes

Putri Nabawy^{1*}, Aidil Halim Lubis², Yusuf Ramadhan Nasution³

^{1,2,3}Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1*}putrinabawy@gmail.com, ²aidilhalimlubis@uinsu.ac.id, ³ramadhannst@uinsu.ac.id

(*Email Corresponding Author: putrinabawy@gmail.com)

Received: 28 Agustus 2026 | Revision: 31 Agustus 2026 | Accepted: 10 September 2026

Abstrak

Stunting merupakan permasalahan gizi kronis yang masih menjadi perhatian utama kesehatan masyarakat di Indonesia karena berdampak pada pertumbuhan fisik dan perkembangan kognitif balita. Deteksi dini risiko stunting diperlukan untuk mendukung intervensi yang tepat sasaran. Penelitian ini bertujuan menerapkan algoritma Naïve Bayes untuk mengklasifikasikan risiko stunting pada balita berdasarkan empat atribut antropometri dasar, yaitu usia, jenis kelamin, berat badan, dan tinggi badan, menggunakan data sekunder sebanyak 300 balita dari Puskesmas Labuhan Ruku, Kabupaten Batu Bara. Label kategori stunting dihitung menggunakan indeks Z-score Tinggi Badan menurut Umur (TB/U) mengacu pada standar WHO Child Growth Standards 2006. Tahapan penelitian meliputi pengumpulan data, pra-proses berupa pemeriksaan missing value dan data duplikat, diskretisasi atribut numerik, label encoding, pembagian data latih dan uji dengan rasio 80:20, pembangunan model Naïve Bayes dengan Laplace smoothing, serta evaluasi menggunakan confusion matrix dan cross validation 10-fold. Hasil pengujian pada 60 data uji menghasilkan confusion matrix dengan TP=27, TN=28, FP=5, dan FN=0, sehingga diperoleh akurasi sebesar 91,67%, presisi 92,97%, recall 91,67%, dan F1-score 91,68%. Validasi 10-fold cross validation menghasilkan rata-rata akurasi 92,67% dengan standar deviasi 7,42%. Seluruh kesalahan klasifikasi bersifat false positive sehingga tidak ada balita berisiko stunting yang lolos tidak terdeteksi. Hasil ini menunjukkan bahwa algoritma Naïve Bayes efektif dan stabil digunakan sebagai metode klasifikasi risiko stunting berbasis data antropometri dan berpotensi mendukung deteksi dini oleh tenaga kesehatan.

Kata Kunci: Stunting, Naïve Bayes, Klasifikasi, Data Mining, Antropometri Balita

Abstract

Stunting remains a major public health concern in Indonesia due to its impact on children's physical growth and cognitive development. Early detection of stunting risk is essential to support targeted prevention. This study aims to implement the Naïve Bayes algorithm to classify stunting risk in children under five based on four basic anthropometric attributes: age, gender, weight, and height, using secondary data from 300 toddlers obtained from Labuhan Ruku Public Health Center, Batu Bara Regency. The stunting category label was computed using the Height-for-Age Z-score (HAZ) index based on the WHO Child Growth Standards 2006. The research stages included data collection, preprocessing (missing value and duplicate checking), discretization of numerical attributes, label encoding, an 80:20 train-test split, Naïve Bayes model development with Laplace smoothing, and evaluation using a confusion matrix and 10-fold cross validation. Testing on 60 test samples produced a confusion matrix with TP=27, TN=28, FP=5, and FN=0, yielding an accuracy of 91.67%, precision of 92.97%, recall of 91.67%, and F1-score of 91.68%. The 10-fold cross validation produced an average accuracy of 92.67% with a standard deviation of 7.42%. All misclassifications were false positives, meaning no stunted children were missed. These results indicate that the Naïve Bayes algorithm is effective and stable for classifying stunting risk based on anthropometric data and has strong potential to support early detection by healthcare workers.

Keywords: Stunting, Naïve Bayes, Classification, Data Mining, Toddler Anthropometry

1. PENDAHULUAN

Stunting pada balita masih menjadi salah satu permasalahan kesehatan masyarakat yang signifikan di Indonesia karena berkaitan langsung dengan kualitas pertumbuhan dan perkembangan anak. Kondisi ini terjadi ketika anak mengalami gangguan pertumbuhan akibat kekurangan gizi kronis dalam jangka waktu yang panjang, terutama pada periode seribu hari pertama kehidupan yang merupakan fase paling krusial dalam pembentukan kemampuan fisik dan kognitif. Anak yang mengalami stunting umumnya memiliki tinggi badan di bawah standar usianya, dan kondisi ini juga berpengaruh terhadap perkembangan otak, kemampuan belajar, serta produktivitas individu di masa dewasa sehingga berdampak pada kualitas sumber daya manusia dalam jangka panjang. Berdasarkan Survei Status Gizi Indonesia (SSGI) tahun 2024, prevalensi stunting nasional tercatat sebesar 19,8%, yang berarti hampir satu dari lima balita di Indonesia masih mengalami gangguan pertumbuhan tersebut [1].

Fenomena stunting tidak hanya dipengaruhi oleh faktor biologis atau kondisi kesehatan anak, tetapi juga berkaitan erat dengan kondisi sosial ekonomi keluarga dan lingkungan tempat anak tumbuh, seperti rendahnya tingkat pendidikan orang tua, keterbatasan pendapatan keluarga, serta kurangnya pemahaman mengenai pola asuh dan pemenuhan gizi anak. Kelompok rumah tangga dengan tingkat kesejahteraan terendah dilaporkan memiliki prevalensi stunting yang lebih tinggi dibandingkan kelompok rumah tangga dengan tingkat kesejahteraan yang lebih baik [2], sehingga stunting merupakan permasalahan multidimensional yang tidak hanya berkaitan dengan aspek kesehatan, tetapi juga dipengaruhi oleh kondisi sosial dan ekonomi masyarakat.

Seiring dengan berkembangnya teknologi informasi, pendekatan analisis data berbasis komputasi mulai banyak digunakan untuk mendukung penelitian di bidang kesehatan masyarakat. Salah satu pendekatan yang dapat dimanfaatkan adalah teknik klasifikasi dalam data mining, yaitu metode analisis data yang bertujuan mengelompokkan suatu objek ke dalam kategori tertentu berdasarkan pola data yang tersedia. Salah satu algoritma klasifikasi yang banyak digunakan dalam penelitian data kesehatan adalah Naïve Bayes, yaitu metode probabilistik yang bekerja berdasarkan prinsip Teorema Bayes untuk menghitung kemungkinan suatu kejadian berdasarkan data yang tersedia. Algoritma ini dikenal memiliki proses komputasi yang relatif sederhana, mampu bekerja dengan baik pada dataset yang tidak terlalu besar, serta memiliki tingkat akurasi yang cukup baik dalam berbagai kasus klasifikasi, sebagaimana ditunjukkan pada berbagai penelitian klasifikasi risiko stunting berbasis data kesehatan keluarga [3].

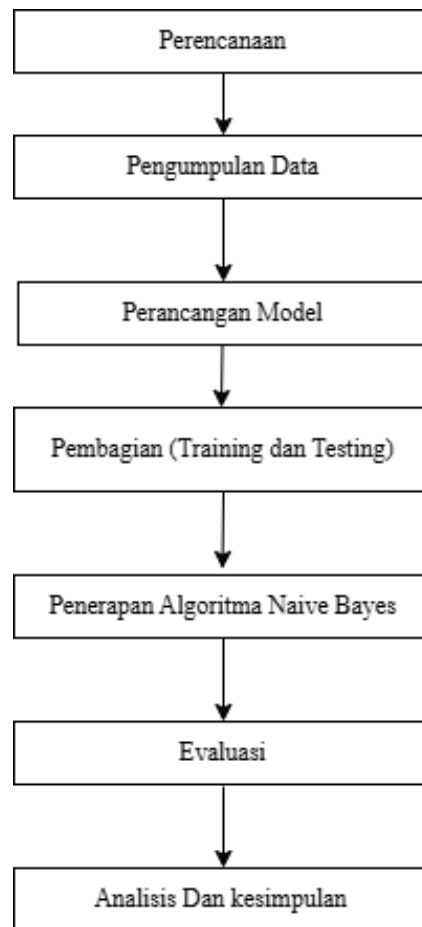
Beberapa penelitian terdahulu telah menerapkan algoritma Naïve Bayes pada permasalahan klasifikasi gizi dan stunting balita dengan hasil yang cukup konsisten. Nilam Wahdiaz Azani et al, menunjukkan bahwa Naïve Bayes efektif digunakan untuk mengklasifikasikan data kesehatan balita melalui tahapan preprocessing, pembagian data, dan evaluasi confusion matrix, dengan akurasi sebesar 81% [4]. Hardiani dan Putri memperoleh akurasi 80% menggunakan tahapan serupa disertai nilai precision dan recall yang baik [5]. Yoshe dan Hadikurniawati melaporkan akurasi 88% dalam mengklasifikasikan status gizi stunting balita menjadi kategori normal dan stunting [6]. Aziz et al, menerapkan Naïve Bayes untuk klasifikasi status gizi balita ke dalam tiga kategori dengan akurasi 90% [7], sementara Abas et al, melaporkan akurasi 95,71% pada klasifikasi status gizi bayi dan balita [8]. Alamin membandingkan KNN, Naïve Bayes, dan SVM untuk klasifikasi status stunting balita dan menemukan bahwa Naïve Bayes memberikan performa terbaik dengan akurasi 99,65%, mengungguli SVM (98,96%) dan KNN (96,36%) [9]. Rumana et al, serta Lutfi et al, menunjukkan Naïve Bayes tetap kompetitif dengan akurasi masing-masing 93,9% dan 95,77% pada klasifikasi keluarga berisiko stunting dan status gizi balita berbasis data antropometri [10], [11].

Berdasarkan penelitian terdahulu tersebut, algoritma Naïve Bayes terbukti efektif dan konsisten digunakan untuk klasifikasi data kesehatan balita dengan akurasi yang relatif tinggi. Namun, sebagian besar penelitian sebelumnya masih menggunakan dataset dari berbagai sumber layanan kesehatan secara umum dan berfokus pada pengujian performa algoritma tanpa menjelaskan secara eksplisit metode perhitungan label kelas yang digunakan, misalnya pada implementasi umum Naïve Bayes untuk aplikasi klasifikasi gizi [12], sehingga proses pelabelan sulit direplikasi. Selain itu, sebagian penelitian menggunakan atribut pendukung yang beragam seperti pola makan dan riwayat MP-ASI [13], atau membandingkan berbagai algoritma tanpa mempertimbangkan efisiensi jumlah atribut yang diperlukan di lapangan [14]. Penelitian ini berbeda dengan penelitian sebelumnya karena secara spesifik menggunakan data kesehatan balita dari Puskesmas Labuhan Ruku, Kabupaten Batu Bara, dengan label kategori stunting yang dihitung secara mandiri dan dapat direplikasi melalui indeks Z-score Tinggi Badan menurut Umur (TB/U) mengacu pada standar WHO Child Growth Standards 2006, serta hanya menggunakan empat atribut antropometri dasar (usia, jenis kelamin, berat badan, dan tinggi badan) yang murah dan mudah diperoleh di lapangan tanpa memerlukan wawancara atau pencatatan data tambahan.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam mengklasifikasikan risiko stunting pada balita berdasarkan data kesehatan yang diperoleh dari Puskesmas Labuhan Ruku, menganalisis tingkat akurasi dan performa model yang dihasilkan menggunakan confusion matrix dan cross validation, serta memberikan informasi yang dapat digunakan sebagai dasar dalam mendukung deteksi dini dan pengambilan keputusan oleh tenaga kesehatan, khususnya di Puskesmas Labuhan Ruku [1], [15].

2. METODOLOGI PENELITIAN

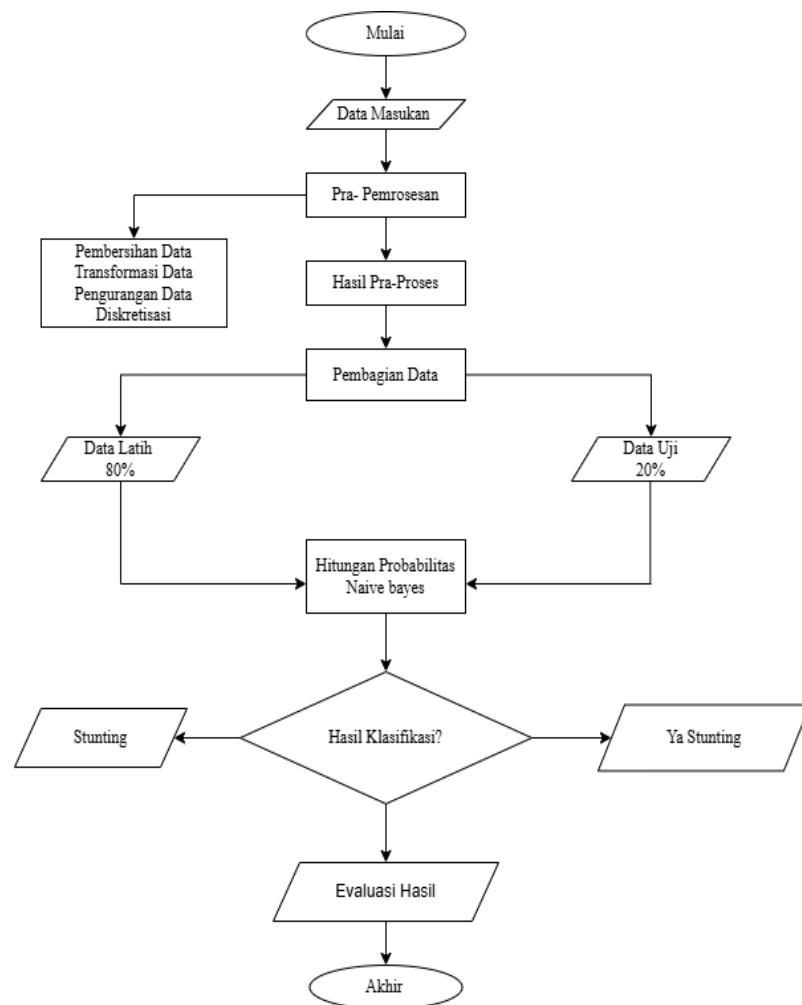
Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen yang bertujuan membangun dan menguji model klasifikasi risiko stunting pada balita berdasarkan data antropometri menggunakan algoritma Naïve Bayes. Alur penelitian secara umum diawali dengan pengumpulan data antropometri balita dari Puskesmas Labuhan Ruku, dilanjutkan dengan tahap praproses data yang meliputi pembersihan data, transformasi data, pengurangan data, dan diskretisasi atribut numerik. Data hasil praproses selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20, kemudian digunakan untuk membangun model klasifikasi Naïve Bayes. Tahap akhir penelitian adalah evaluasi model menggunakan confusion matrix, metrik akurasi, precision, recall, dan F1-score, serta 10-fold cross validation untuk menguji stabilitas model. Rincian setiap tahapan tersebut dijelaskan pada Sub-bab 2.1 dan 2.2 berikut.



Gambar 1. Kerangka Penelitian

2.1 Tahapan Penelitian dan Data

Penelitian ini dilakukan di Puskesmas Labuhan Ruku, Kecamatan Talawi, Kabupaten Batu Bara, Sumatera Utara, menggunakan data sekunder berupa catatan antropometri balita. Tahapan penelitian secara umum meliputi pengumpulan data, praproses data, pembagian data latih dan data uji, penerapan algoritma Naïve Bayes, serta evaluasi model, sebagaimana digambarkan pada alur penelitian pada Gambar 2:



Gambar 2. Alur Tahapan Penelitian Klasifikasi Naïve Bayes

Berdasarkan Gambar 1, proses diawali dengan data masukan berupa atribut antropometri balita, dilanjutkan dengan praproses data yang meliputi pembersihan data, transformasi data, pengurangan data, dan diskretisasi. Data hasil praproses selanjutnya dibagi menjadi data latih (80%) dan data uji (20%). Kedua subset data tersebut digunakan dalam perhitungan probabilitas Naïve Bayes untuk menghasilkan keputusan klasifikasi Stunting atau Tidak Stunting, yang kemudian dievaluasi untuk mengukur performa model.

Dataset yang digunakan berjumlah 300 data balita dengan empat atribut input, yaitu usia (bulan), jenis kelamin, berat badan (kg), dan tinggi badan (cm), sebagaimana dirangkum pada Tabel 1.

Tabel 1. Atribut Data Penelitian

No	Atribut	Tipe Data	Keterangan
1	Usia (Bulan)	Integer	Usia balita dalam satuan bulan
2	Jenis Kelamin	Kategorik	L = Laki-laki, P = Perempuan
3	Berat Badan (kg)	Float	Berat badan balita dalam kilogram
4	Tinggi Badan (cm)	Float	Tinggi badan balita dalam sentimeter

Label kategori stunting tidak diperoleh langsung dari pencatatan puskesmas, melainkan dihitung dari atribut usia, jenis kelamin, dan tinggi badan menggunakan indeks Z-score Tinggi Badan menurut Umur (TB/U) pada Persamaan (1), mengacu pada tabel median dan standar deviasi WHO Child Growth Standards 2006 yang juga menjadi acuan Peraturan Menteri Kesehatan RI Nomor 2 Tahun 2020 tentang Standar Antropometri Anak.

$$Z - score \frac{TB}{U} = \frac{(Tinggi\ Badan\ Terukur - Median \frac{TB}{U} \text{ Baku Rujukan})}{Standar\ Deviasi\ Baku\ Rujukan} \quad (1)$$

Balita dikategorikan Stunting apabila nilai Z-score TB/U < -2 SD, dan Tidak Stunting apabila Z-score TB/U ≥ -2 SD. Perhitungan dilakukan terhadap seluruh 300 data menggunakan Python berdasarkan tabel LMS (Length/Height-for-Age) WHO usia 0–60 bulan yang dipisah menurut jenis kelamin, sehingga diperoleh distribusi kelas sebanyak 133 data (44,33%) Stunting dan 167 data (55,67%) Tidak Stunting.

2.2 Praproses Data, Model Naïve Bayes, dan Evaluasi

Data yang telah diberi label selanjutnya melalui tahap praproses yang meliputi pemeriksaan missing value, pemeriksaan data duplikat, diskretisasi tiga atribut numerik (usia, berat badan, tinggi badan) ke dalam lima interval kategori berdasarkan kuantil data, serta label encoding pada seluruh atribut kategorik menggunakan LabelEncoder scikit-learn untuk keperluan implementasi sistem berbasis Python.

Algoritma Naïve Bayes mengklasifikasikan data berdasarkan Teorema Bayes sebagaimana dirumuskan pada Persamaan (2), dengan asumsi bahwa setiap atribut bersifat independen satu sama lain.

$$P(C|X) = \frac{[P(X|C) \times P(C)]}{P(X)} \quad (2)$$

dengan $P(C|X)$ merupakan probabilitas posterior kelas C setelah mempertimbangkan data X (posterior probability), $P(X|C)$ adalah probabilitas kemunculan data X pada kelas C (likelihood), $P(C)$ adalah probabilitas awal kelas sebelum melihat data (prior probability), dan $P(X)$ adalah probabilitas data yang diamati (evidence). Untuk menghindari nilai probabilitas nol pada fitur kategorik yang tidak muncul pada data latih, diterapkan Laplace smoothing sebagaimana pada Persamaan (3).

$$P(x_i = v | C) = \frac{\text{jumlah data kelas } C \text{ dengan } x_i = v + 1}{(\text{jumlah data kelas } C + k)} \quad (3)$$

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan metode stratified split (random_state=42) agar proporsi kelas tetap terjaga, sehingga diperoleh 240 data latih (106 Stunting, 134 Tidak Stunting) dan 60 data uji (27 Stunting, 33 Tidak Stunting). Model dilatih menggunakan pustaka scikit-learn pada lingkungan Google Colab, kemudian dievaluasi menggunakan confusion matrix serta empat metrik evaluasi, yaitu akurasi (Persamaan 4), precision (Persamaan 5), recall (Persamaan 6), dan F1-score (Persamaan 7) [3], dan divalidasi lebih lanjut menggunakan 10-fold cross validation untuk menguji stabilitas model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan dan Pelabelan Data

Pengumpulan data dilakukan secara langsung di Puskesmas Labuhan Ruku dengan memperoleh 300 record data sekunder berupa catatan antropometri balita yang mencakup empat atribut, yaitu jenis kelamin, usia (bulan), berat badan (kg), dan tinggi badan (cm). Cuplikan lima data pertama dan lima data terakhir dari keseluruhan dataset disajikan pada Tabel 2.

Tabel 2. Cuplikan Dataset Penelitian

No	Jenis Kelamin	Usia (Bulan)	Berat Badan (kg)	Tinggi Badan (cm)
1	L	59	19,6	82,4
2	L	53	6,3	96,3
3	P	51	10,3	63,2
4	P	22	17,8	87,0
5	P	20	10,1	100,9
...	...	...	...	...
297	P	12	11,9	81,5

No	Jenis Kelamin	Usia (Bulan)	Berat Badan (kg)	Tinggi Badan (cm)
298	P	44	6,5	89,8
299	L	58	11,8	89,8
300	P	23	8,0	98,8

Struktur data mentah bersifat konsisten dari baris pertama hingga baris terakhir, dengan format dan tipe nilai yang seragam pada keempat atribut. Label kategori stunting kemudian dihitung menggunakan indeks Z-score TB/U sebagaimana dijelaskan pada Sub-bab 2.1, dengan hasil distribusi kelas disajikan pada Tabel 3.

Tabel 3. Distribusi Kelas Dataset

No	Kategori Stunting	Jumlah Data	Persentase
1	Stunting	133	44,33%
2	Tidak Stunting	167	55,67%
	Total	300	100%

Distribusi kelas yang dihasilkan tergolong cukup seimbang (44,33% Stunting berbanding 55,67% Tidak Stunting) karena label kelas merupakan fungsi langsung dari atribut antropometri balita itu sendiri, sehingga tidak diperlukan teknik penyeimbangan data tambahan seperti oversampling.

3.2 Hasil Prapemroses Data

Pemeriksaan missing value dan data duplikat terhadap seluruh 300 data menunjukkan tidak ditemukan nilai kosong maupun baris data yang identik pada keempat atribut input, sehingga seluruh data dinyatakan siap digunakan tanpa memerlukan imputasi atau penghapusan data. Selanjutnya, tiga atribut numerik (usia, berat badan, tinggi badan) didiskretisasi menjadi lima interval kategori berdasarkan kuantil data agar dapat diproses oleh model Naïve Bayes kategorik, dengan sebaran hasil diskretisasi disajikan pada Tabel 4.

Tabel 4. Distribusi Hasil Diskretisasi Atribut Numerik

Atribut	Interval	Jumlah	Atribut	Interval	Jumlah
Usia	0-12 bln	43	Tinggi Badan	<70 cm	55
Usia	13-24 bln	69	Tinggi Badan	70-84 cm	72
Usia	25-36 bln	61	Tinggi Badan	85-99 cm	74
Usia	37-48 bln	79	Tinggi Badan	100-114 cm	76
Usia	49-59 bln	48	Tinggi Badan	≥115 cm	23
Berat Badan	<8 kg	48	Berat Badan	14-17 kg	62
Berat Badan	8-11 kg	58	Berat Badan	≥17 kg	62
Berat Badan	11-14 kg	70			

Seluruh nilai kategorik, termasuk hasil diskretisasi dan jenis kelamin, selanjutnya dikonversi menjadi kode numerik menggunakan LabelEncoder scikit-learn untuk keperluan implementasi sistem berbasis Python. Dataset kemudian dibagi menggunakan metode stratified split dengan rasio 80:20, menghasilkan 240 data latih (106 Stunting, 134 Tidak Stunting) dan 60 data uji (27 Stunting, 33 Tidak Stunting) sebagaimana dirangkum pada Tabel 5.

Tabel 5. Pembagian Data Latih dan Data Uji

Jenis Data	Proporsi	Jumlah (Stunting / Tidak Stunting)	Rasio
Latih (Training)	80% × 300	240 (106 / 134)	80%
Uji (Testing)	20% × 300	60 (27 / 33)	20%
Total		300	100%

Sebagai ilustrasi proses klasifikasi tingkat atribut, simulasi perhitungan manual juga dilakukan terhadap 10 data sampel (8 data latih dan 2 data uji) menggunakan rumus probabilitas prior dan probabilitas kondisional dengan Laplace smoothing pada Persamaan (3). Hasil simulasi manual berhasil mengklasifikasikan benar 1 dari 2 data uji (akurasi 50%), yang mencerminkan keterbatasan estimasi probabilitas ketika data latih hanya berjumlah 8 baris. Hasil ini mengonfirmasi pentingnya ukuran dataset yang memadai dalam estimasi probabilitas Naïve Bayes, dan menjadi dasar dilakukannya

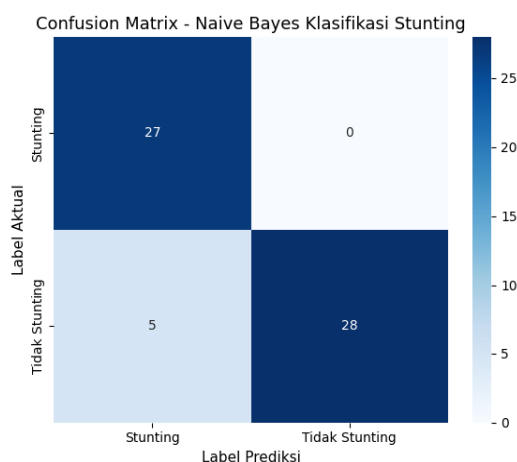
implementasi sistem pada keseluruhan 300 data menggunakan Python dan pustaka scikit-learn di lingkungan Google Colab.

3.3 Hasil Evaluasi Model

Model Naïve Bayes yang dilatih pada 240 data latih diuji pada 60 data uji, menghasilkan confusion matrix sebagaimana disajikan pada Tabel 6 dan divisualisasikan pada Gambar 2.

Tabel 6. Confusion Matrix Hasil Klasifikasi Naïve Bayes

	Prediksi: Stunting	Prediksi: Tidak Stunting
Aktual: Stunting	TP = 27	FN = 0
Aktual: Tidak Stunting	FP = 5	TN = 28



Gambar 3. Visualisasi Confusion Matrix Klasifikasi Stunting

Berdasarkan Tabel 6, model berhasil mengklasifikasikan 27 balita Stunting dengan benar (True Positive) dan 28 balita Tidak Stunting dengan benar (True Negative), tanpa satupun balita Stunting yang salah diprediksi sebagai Tidak Stunting (False Negative = 0). Namun, terdapat 5 balita Tidak Stunting yang salah diprediksi sebagai Stunting (False Positive). Berdasarkan confusion matrix tersebut, keempat metrik evaluasi dihitung menggunakan Persamaan (4)–(7) dan dirangkum pada Tabel 7.

Tabel 7. Ringkasan Metrik Evaluasi Model

No	Metrik Evaluasi	Nilai	Persentase
1	Akurasi	0,9167	91,67%
2	Presisi (Weighted)	0,9297	92,97%
3	Recall (Weighted)	0,9167	91,67%
4	F1-Score (Weighted)	0,9168	91,68%

Untuk melihat performa model pada masing-masing kelas secara terpisah, classification report per kelas disajikan pada Tabel 8.

Tabel 8. Classification Report per Kelas

Kelas	Precision	Recall	F1-Score	Support
Stunting	84,38%	100%	91,53%	27
Tidak Stunting	100%	84,85%	91,80%	33
Weighted Avg	92,97%	91,67%	91,68%	60

Tabel 8 menunjukkan recall kelas Stunting mencapai 100%, yang berarti seluruh balita Stunting pada data uji berhasil terdeteksi tanpa ada yang lolos. Sebaliknya, precision kelas Stunting hanya 84,38% karena 5 balita Tidak Stunting ikut

terprediksi sebagai Stunting. Sebagai validasi tambahan terhadap stabilitas model, dilakukan 10-fold cross validation dengan hasil akurasi tiap fold disajikan pada Tabel 9.

Tabel 9. Hasil Cross Validation 10-Fold

Fold	Akurasi	Fold	Akurasi
1	96,67%	6	93,33%
2	100%	7	83,33%
3	96,67%	8	96,67%
4	86,67%	9	76,67%
5	96,67%	10	100%

Hasil pada Tabel 9 menunjukkan rata-rata akurasi 10-fold cross validation sebesar 92,67% dengan standar deviasi 7,42%. Akurasi tertinggi diperoleh pada Fold 2 dan Fold 10 (100%), sedangkan akurasi terendah diperoleh pada Fold 9 (76,67%).

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan risiko stunting pada balita di Puskesmas Labuhan Ruku dengan performa yang baik hanya menggunakan empat atribut antropometri dasar, yaitu usia, jenis kelamin, berat badan, dan tinggi badan, tanpa memerlukan data tambahan seperti riwayat pemberian ASI eksklusif, status imunisasi, atau kondisi sosial ekonomi keluarga. Akurasi sebesar 91,67% pada data uji, dengan presisi 92,97%, recall 91,67%, dan F1-score 91,68%, menunjukkan bahwa model memiliki performa yang sebanding dengan beberapa penelitian terdahulu, sebagaimana dirangkum pada Tabel 10.

Tabel 10. Perbandingan Akurasi dengan Penelitian Terdahulu

No	Peneliti	Metode	Akurasi
1	Hardiani dan Putri	Naïve Bayes	80%
2	Azani et al.	Naïve Bayes dan SVM	81%
3	Yoshe dan Hadikurniawati	Naïve Bayes	88%
4	Aziz et al.	Naïve Bayes	90%
5	Abas et al.	Naïve Bayes	95,71%
6	Lutfi et al.	Naïve Bayes	95,77%
7	Alamin	Naïve Bayes, SVM, KNN	99,65%
8	Penelitian ini	Naïve Bayes	91,67%

Tabel 10 menunjukkan bahwa akurasi model pada penelitian ini (91,67%) sebanding dengan penelitian Hardiani dan Putri (80%) [5, Azani et al. (81%), Yoshe dan Hadikurniawati (88%) [4], serta Aziz et al. (90%) [7], meskipun sedikit lebih rendah dibandingkan penelitian Abas et al. (95,71%) [8], Lutfi et al. (95,77%) [11], dan Alamin (99,65%) yang menggunakan atribut antropometri dan gizi yang lebih lengkap [9]. Sebagai validasi tambahan terhadap kebenaran fungsional hasil klasifikasi, model juga diuji dengan pendekatan yang sejalan dengan prinsip pengujian black-box, yaitu model diperlakukan sebagai unit yang hanya diamati dari sisi masukan (atribut antropometri data uji) dan keluaran (label prediksi Stunting/Tidak Stunting) tanpa menelusuri proses perhitungan probabilitas internalnya, kemudian keluaran tersebut dicocokkan dengan label aktual pada seluruh 60 data uji dan pada masing-masing fold cross validation. Kesesuaian keluaran model dengan label aktual pada pengujian tersebut, sebagaimana dirangkum pada confusion matrix Tabel 6 dan hasil 10-fold cross validation pada Tabel 9, mengonfirmasi bahwa fungsi klasifikasi model bekerja secara konsisten dan sesuai dengan hasil yang diharapkan.

Karakteristik kesalahan klasifikasi pada penelitian ini seluruhnya bersifat false positive (FP=5), yaitu balita yang sebenarnya Tidak Stunting namun diprediksi Stunting, sementara tidak ada satupun balita Stunting yang lolos tidak terdeteksi (FN=0, recall kelas Stunting 100%). Dari sudut pandang skrining kesehatan, karakteristik ini menguntungkan karena model cenderung lebih waspada (over-detect) dibandingkan melewatkan balita yang benar-benar berisiko stunting, sehingga risiko keterlambatan penanganan dapat diminimalkan meskipun harus diimbangi dengan verifikasi lanjutan pada kasus false positive oleh tenaga kesehatan.

Hasil 10-fold cross validation dengan rata-rata akurasi 92,67% dan standar deviasi 7,42% menunjukkan bahwa model relatif stabil, meskipun variasi akurasi antar fold (terendah 76,67% pada Fold 9, tertinggi 100% pada Fold 2 dan Fold 10) mengindikasikan bahwa dengan hanya empat fitur input, performa model cukup sensitif terhadap komposisi data pada tiap fold. Penerapan Laplace smoothing pada Persamaan (3) tetap terbukti penting untuk menghindari probabilitas nol, khususnya pada kondisi data latih berukuran kecil sebagaimana ditunjukkan pada simulasi manual di Sub-bab 3.2.

Secara praktis, pendekatan klasifikasi berbasis empat atribut antropometri murni ini tergolong murah dan cepat untuk diterapkan di lapangan, karena hanya memerlukan pengukuran usia, jenis kelamin, berat badan, dan tinggi badan balita tanpa memerlukan wawancara atau pencatatan data tambahan lainnya, sehingga berpotensi diterapkan sebagai alat bantu skrining awal risiko stunting oleh tenaga kesehatan di puskesmas maupun posyandu. Meskipun demikian, penelitian ini masih memiliki keterbatasan, yaitu penggunaan data dari satu lokasi puskesmas serta belum melibatkan faktor pendukung lain seperti riwayat ASI eksklusif, status imunisasi, dan kondisi sosial ekonomi maupun pola makan keluarga yang menurut Ratnasari et al. [2] serta Yanti et al. [13] turut berkontribusi terhadap risiko stunting, sehingga perlu dikaji pada penelitian selanjutnya untuk meningkatkan akurasi dan generalisasi model.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes berhasil diterapkan untuk mengklasifikasikan risiko stunting pada balita di Puskesmas Labuhan Ruku menggunakan empat atribut antropometri dasar, yaitu usia, jenis kelamin, berat badan, dan tinggi badan, dengan label kategori stunting yang dihitung secara mandiri dan dapat direplikasi melalui indeks Z-score Tinggi Badan menurut Umur (TB/U) mengacu pada standar WHO Child Growth Standards 2006. Proses penelitian meliputi pengumpulan 300 data sekunder, praproses berupa pemeriksaan missing value dan data duplikat, diskretisasi atribut numerik, label encoding, serta pembagian data latih dan data uji dengan rasio 80:20. Pengujian model pada 60 data uji menghasilkan confusion matrix dengan TP=27, TN=28, FP=5, dan FN=0, sehingga diperoleh akurasi sebesar 91,67%, presisi 92,97%, recall 91,67%, dan F1-score 91,68%. Seluruh kesalahan klasifikasi bersifat false positive, sementara seluruh balita Stunting berhasil terdeteksi dengan recall kelas Stunting mencapai 100%. Hasil 10-fold cross validation menghasilkan rata-rata akurasi 92,67% dengan standar deviasi 7,42%, yang menunjukkan bahwa model cukup stabil dalam mengklasifikasikan risiko stunting menggunakan empat atribut antropometri dasar. Dengan demikian, algoritma Naïve Bayes terbukti efektif, murah, dan cepat untuk diterapkan sebagai alat bantu deteksi dini risiko stunting berbasis data antropometri, serta berpotensi mendukung pengambilan keputusan tenaga kesehatan di puskesmas maupun posyandu. Penelitian selanjutnya disarankan untuk membandingkan performa Naïve Bayes dengan algoritma klasifikasi lain, memperluas cakupan dataset ke berbagai wilayah, serta menambahkan atribut pendukung seperti riwayat pemberian ASI eksklusif, status imunisasi, dan kondisi sosial ekonomi keluarga guna meningkatkan akurasi dan generalisasi model.

REFERENCES

- [1] U. R. Gurning, S. F. Octavia, and D. R. Andriyani, "Prediction of Stunting Risk In Families Using Naïve Bayes Classifier and Chi-Square Prediksi Risiko Stunting pada Keluarga Menggunakan Naïve Bayes Classifier dan Chi-Square," vol. 4, no. January, pp. 172–180, 2024.
- [2] A. J. Wahidin and T. H. Andika, "Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning," pp. 378–387, 2024, doi: 10.33364/algoritma/v.21-2.2122.
- [3] Y. Suliswati, T. O. Mukhti, and A. Salma, "IMPLEMENTASI METODE NAÏVE BAYES DENGAN RANDOM OVERSAMPLING PADA KLASIFIKASI KELUARGA BERISIKO," vol. 5, no. 02, pp. 223–233, 2025.
- [4] N. W. Azani, M. Afdal, R. Novita, U. Islam, N. Sultan, and S. Kasim, "Implementasi Pengklasifikasi Naïve Bayes dan Support Vector Machine untuk Klasifikasi Stunting," vol. 13, no. 4, pp. 5079–5087, 2024.
- [5] T. Hardiani and R. N. Putri, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita," vol. 4, no. 1, pp. 621–627, 2024.
- [6] M. Yoshe and W. Hadikurniawati, "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," vol. 2019, 2021.
- [7] A. Aziz, M. Furqan, and M. Fakhri, "PENERAPAN METODE NAIVE BAYES UNTUK KLASIFIKASI STATUS GIZI BALITA," pp. 617–625, 2024.
- [8] M. I. Abas, R. Lamusu, W. E. Pranata, S. Syahrial, and I. Ibrahim, "BULLETIN OF COMPUTER SCIENCE RESEARCH Penerapan Algoritma Naive Bayes Untuk Sistem Klasifikasi Status Gizi Bayi Balita," vol. 5, no. 5, pp. 1249–1260, 2025, doi: 10.47065/bulletinscr.v5i5.508.
- [9] P. B. Alamin and D. Aryani, "PERBANDINGAN ALGORITMA KNN , NAIVE BAYES DAN SVM UNTUK KLASIFIKASI STATUS STUNTING ANAK BALITA DI PUSKESMAS," vol. 4, no. 4, pp. 360–369, 2025.
- [10] S. A. Rumanama, M. S. N. Van Delsen, and N. S. Laamena, "An Application of the Naïve Bayes Algorithm Method for Classification of Families at Risk of Stunting (Case Study: Waeapo District, Buru Regency)," vol. 5, no. September, 2025.
- [11] M. Lutfi et al., "KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN NAÏVE BAYES CLASSIFICATION DI KELURAHAN," vol. 3, no. 1, pp. 1–11, 2025.
- [12] F. Bunda et al., "IMPLEMENTASI ALGORITMA NAIVE BAYES PADA APLIKASI WEB," pp. 333–340, 2023.

- [13] Z. V. Yanti, M. Furqan, and I. N. Muhdar, "HUBUNGAN POLA MAKAN DAN PERILAKU MP-ASI DENGAN STATUS GIZI ANAK USIA 6-12 BULAN DI DESA CIGEULIS PROVINSI BANTEN," vol. 1, no. 2, pp. 110–117, 2025, doi: 10.36590/akg.v2i1.1895.
- [14] A. M. Nisfi, A. S. Bachtiar, and C. I. Agustyaningrum, "Analisis Perbandingan Algoritma Klasifikasi Machine Learning dalam Pendeteksian Risiko Stunting pada Balita di Desa Sukahati," vol. 8, no. September, pp. 10562–10568, 2025.
- [15] R. D. Hartana and E. I. Sela, "Nutritional Status Classification Of Stunting In Toddlers Using Naive Bayes Classifier Method," vol. 3, no. 1, 2024.