

Classification of Toddler Stunting Status Using Naïve Bayes Classifier with K-Fold Cross Validation

Vania Riski Affah, Zilrahmi*, Syafriandi, dan Tessy Octavia Mukhti

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

Corresponding author: zilrahmi@fmipa.unp.ac.id

Submitted : 14 Juli 2026
Revised : 11 Agustus 2026
Accepted : 08 September 2026

ABSTRACT

The growth of toddlers that doesn't meet age standards can affect their quality of life in the future. In Indonesia, one of the nutritional problems that continues to receive significant attention is stunting, which is generally indicated by a mismatch between a child's height and age (height-for-age, HAZ/TB/U). Considering that determining stunting status requires a high level of accuracy, a data-driven approach is needed to support the identification and evaluation of children's nutritional conditions in a more systematic manner. This study aims to classify stunting status among toddlers using the Naïve Bayes Classifier (NBC) algorithm with a 10-Fold Cross Validation method. NBC was selected because it is simple, efficient, and suitable for probability-based classification of health data. The data were obtained from the Health Office of Pasaman Regency in 2022. The variables used include gender, birth weight, birth height, age at measurement, body weight, body height, Mid-Upper Arm Circumference (MUAC), and height-for-age (HAZ) status. Stunting status was divided into two categories: stunted and severely stunted. The analysis showed that 53.91% of toddlers were classified as stunted, while 46.10% were classified as severely stunted. Based on the evaluation using 10-Fold Cross Validation, the NBC model showed good performance with an accuracy of 86.26%, precision of 73.33%, recall of 68.75%, and an F1-score of 70.97%. The analysis also showed that height, weight, and MUAC at the time of measurement were the characteristics that most clearly distinguished the stunted and severely stunted categories. These findings indicate that anthropometric indicators can support the detection and monitoring of stunting among toddlers. Overall, this study is expected to help health workers identify toddlers who require further nutritional assessment and support more targeted stunting management efforts.

Keywords: Stunting, Naïve Bayes Classifier, Gaussian Naive Bayes, 10-Fold Cross Validation.

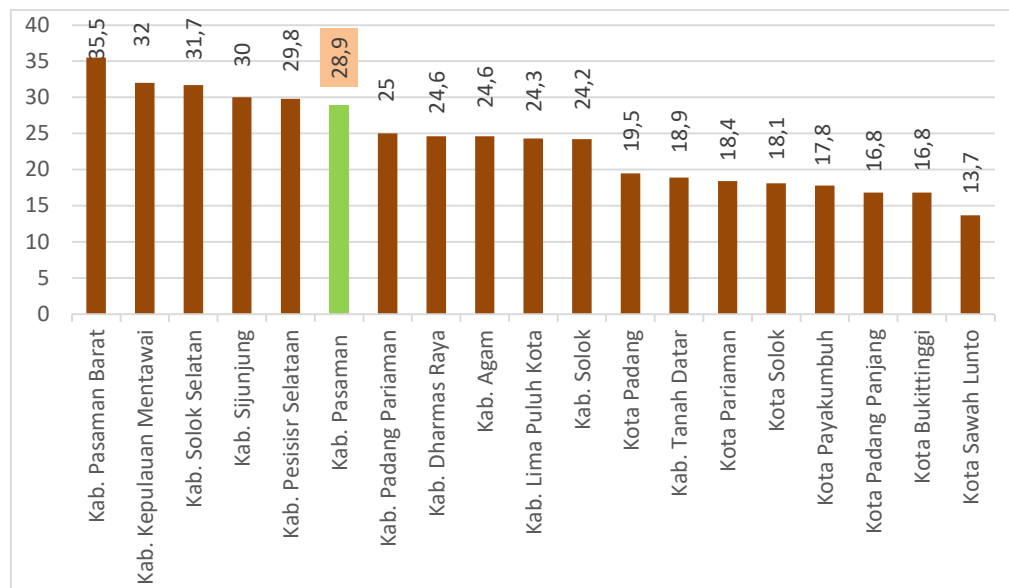


This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Mewujudkan tujuan “Indonesia Emas 2045” dalam mengembangkan SDM yang berkualitas tinggi merupakan hal yang sangat penting dalam agenda Indonesia saat ini. Salah satu tantangan utama dalam mewujudkan hal tersebut adalah permasalahan gizi, khususnya stunting pada balita. Ketika balita mengalami malnutrisi kronis, mereka tidak dapat mencapai potensi penuh mereka sebagaimana balita seusianya; kondisi ini dikenal sebagai stunting. Gangguan ini meningkatkan risiko penyakit di masa dewasa serta menghambat perkembangan fisik, pematangan kognitif, dan produktivitas. (Kementerian Kesehatan RI, 2023).

Berdasarkan hasil “Survei Status Gizi Indonesia” (SSGI) tahun 2022 prevalensi stunting nasional sebesar 21,6%. Pada tahun yang sama, prevalensi *stunting* di Provinsi Sumatera Barat mencapai 25,2%, lebih besar dibandingkan angka nasional. Di tingkat kabupaten/kota, prevalensi *stunting* di Provinsi Sumatera Barat menampilkan variasi yang cukup besar. Kabupaten Pasaman tercatat memiliki prevalensi stunting sebesar 28,9%, lebih tinggi dibandingkan prevalensi Provinsi Sumatera Barat (25,2%) maupun prevalensi nasional (21,6%). Gambar 1 menyajikan 19 Kabupaten/Kota dengan prevalensi *stunting* di Provinsi Sumatera Barat tahun 2022.



Sumber : Kemenkes RI, 2022

Gambar 1. Grafik Prevelensi *Stunting* di Sumatera Barat Tahun 2022

Berlandaskan Gambar 1, dapat diketahui bahwa prevalensi *stunting* di Kabupaten Pasaman tahun 2022 meraih 28,9%, lebih tinggi dibandingkan prevalensi Provinsi Sumatera Barat (25,2%) maupun prevalensi nasional (21,6%). Angka tersebut menunjukkan bahwa Kabupaten Pasaman termasuk sebuah kabupaten di Sumatera Barat yang mempunyai prevalensi *stunting* relatif tinggi. Jika dibandingkan dengan target nasional penurunan *stunting* yang ditetapkan pemerintah, yaitu sebesar 14% pada tahun 2024, maka prevalensi *stunting* di Kabupaten Pasaman masih lebih dari 2 x lipat dari target yang ingin dicapai. Hal ini mengindikasikan perlunya upaya intervensi yang lebih intensif dan terintegrasi untuk mempercepat penurunan prevalensi *stunting* di wilayah tersebut. Tingginya angka *stunting* dipengaruhi oleh berbagai faktor seperti ekonomi, pola asuh, sanitasi, dan akses layanan kesehatan yang masih perlu ditingkatkan (Dinkes Kabupaten Pasaman, 2022). Karena itu, dibutuhkan metode yang mampu mengidentifikasi status *stunting* secara cepat dan akurat. Salah satu strategi yang dapat diterapkan ialah teknik klasifikasi penambangan data.

Metode klasifikasi dalam penambangan data dapat membangun model yang mengelompokkan informasi ke dalam kelas-kelas yang relevan berdasarkan atributnya; hal ini menjadikan metode tersebut sebagai alat yang berharga bagi sistem pendukung keputusan di bidang kesehatan. Suatu algoritma klasifikasi yang banyak dipakai ialah Algoritma “*Naïve Bayes Classifier*” (NBC). Algoritma ini menerapkan Teorema Bayes pada asumsi bahwasannya tiap atribut saling bebas (independent). Keunggulan Algoritma *Naïve Bayes Classifier* (NBC) adalah sederhana, efisien, efektif diterapkan pada dataset besar dan data yang beragam, serta memiliki biaya komputasi yang rendah (Chen, 2021). Untuk meningkatkan performa model, digunakan metode *K-Fold Cross Validation* yang dapat menerapkan evaluasi model yang lebih stabil.

Berdasarkan uraian tersebut, studi berikut tujuannya untuk menerapkan “*Algoritma Naïve Bayes Classifier*” pada klasifikasi status *stunting* dan mengevaluasi performanya menggunakan “*K-Fold Cross Validation*.” Selanjutnya, penelitian berikut difokuskan guna mengkaji performa model menggunakan ukuran *accuracy*, *precision*, *recall*, serta *F1-score*. Perolehan studi harapannya jadi salah satu alternatif pendukung pengambilan keputusan dalam identifikasi dini status *stunting* sehingga intervensi yang diberikan dapat dilakukan secara lebih tepat sasaran.

II. METODE PENELITIAN

A. Jenis dan Sumber Data

Data yang diterapkan dalam studi berikut ialah data sekunder yang didapatkan dari Dinas Kesehatan Kabupaten Pasaman tahun 2022. Data yang digunakan hanya mencakup balita dengan kategori *stunting*, yaitu pendek dan sangat pendek. Variabel yang digunakan meliputi:

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala
Status Stunting Balita TB/U (Y)	Kategori: Pendek Sangat Pendek	Ordinal
Jenis Kelamin (X_1)	Kategori: 0 = Perempuan 1 = Laki-laki	Nominal
Berat Badan Lahir (X_2)	Kilogram	Rasio
Tinggi Badan Lahir (X_3)	Centimeter	Rasio
Usia Saat di Ukur (X_4)	Bulan	Rasio
Berat Badan Balita (X_5)	Kilogram	Rasio
Tinggi Badan Balita (X_6)	Centimeter	Rasio
Lingkar Lengan Atas (X_7)	Centimeter	Rasio

Sumber: Badan Pusat Statistik Kabupaten Pasaman

B. Teknik Analisis Data

Teknik analisis data dalam studi berikut menerapkan algoritma “*Naïve Bayes Classifier*” pada metode uji *K-Fold Cross Validation*. Adapun langkah-langkah penerapan algoritma *Naïve Bayes Classifier* menggunakan *K-Fold Cross Validation* sebagai berikut.

1. Mengumpulkan data balita stunting Kabupaten Pasaman dengan total data sebanyak 3.253 data.
2. Membersihkan data untuk menghapus atau memperbaiki data yang tidak valid, menangani nilai hilang, dan menyesuaikan format agar sesuai dengan kebutuhan analisis.
3. Transformasi data, yaitu mengubah format, skala, atau satuan variabel, termasuk mengonversi data kategorikal menjadi numerik atau membuat variabel baru yang relevan.
4. Pembagian data dengan *K-Fold Cross Validation*.
 - a. Menetapkan nilai K pada *K-Fold Cross Validation*, yaitu K=10.
 - b. *Dataset* dibagi secara acak mejadi K bagian (partisi), yaitu sebanyak 10 bagian.
 - c. Membagi setiap partisi data menjadi data pelatihan dan data pengujian secara bergantian sebanyak K kali,yaitu 10 kali.
5. Penerapan Algoritma *Naïve Bayes Classifier* (NBC).
 - a. Menghitung probabilitas prior untuk setiap kelas target (Pendek dan Sangat Pendek) menggunakan persamaan (1).

$$P(y) = \frac{n_i}{N} \quad (1)$$

- b. Menghitung *probabilitas kondisional* untuk setiap nilai fitur (baik numerik maupun kategorikal) terhadap masing-masing kelas target. Untuk fitur numerik menggunakan persamaan (2).

$$P(x_k|C_i) = \frac{1}{\sqrt{2\pi\sigma_{c_i}}} e^{-\frac{(x_k-\mu_{c_i})^2}{2\sigma_{c_i}^2}} \quad (2)$$

sedangkan fitur kategorikal dihitung menggunakan persamaan (3).

$$P(x_k|y) = \frac{n_{ij}}{n_j} \quad (3)$$

- c. Menghitung *probabilitas posterior* dari setiap data uji menggunakan persamaan (4).
$$C_{map} = \operatorname{argmax}_{y \in Y} P(y) \prod_{k=1}^n P(x_k|y) \quad (4)$$
 - d. Menentukan kelas prediksi berdasarkan nilai *probabilitas posterior* tertinggi. Kelas dengan nilai probabilitas terbesar dipilih sebagai hasil klasifikasi.
6. Melakukan Evaluasi Model.
 - a. Membentuk *Confusion Matrix*.
Confusion Matrix berbentuk tabel guna mengkaji kinerja model klasifikasi. *Confusion Matrix* disajikan dalam Tabel 2.

Tabel 2. Confusion Matrix

Prediction	Actual		
	Yes	No	
Yes	TP (True Positive)	FP (False Positive)	
No	FN (False Negative)	TN (True Negative)	

- b. Menilai matrik evaluasi misalnya *Accuracy*, *Precision*, *Recall*, serta *F1 Score*. Adapun persamaan untuk setiap iterasi K-Fold adalah sebagai berikut:

- Akurasi merupakan rasio antara jumlah prediksi yang benar terhadap banyaknya keseluruhan data yang di evaluasi. Berikut rumus menghitung akurasi disajikan pada persamaan (5).

$$\text{Akurasi (Acc)} = \frac{TN+TP}{FN+FP+TN+TP} \quad (5)$$

- Presisi merupakan rasio antara banyaknya *True Positive* serta keseluruhan prediksi positif (*True Positive* dan *False Positive*). Berikut rumus menghitung presisi yang disajikan pada persamaan (6).

$$\text{Presisi (Pre)} = \frac{TP}{TP+FP} \quad (6)$$

- Recall* merupakan rasio antara banyaknya *True Positive* serta keseluruhan instance yang sebenarnya positif. Berikut rumus menghitung *recall* disajikan pada persamaan (7).

$$\text{Recall (Re)} = \frac{TP}{TP+FN} \quad (7)$$

- F1-Score* menunjukkan tingkat ketepatan model untuk melaksanakan klasifikasi dengan benar. Berikut rumus hitung *F1-score* yang disajikan pada persamaan (8).

$$F1 - score = \frac{2 \times pre \times rec}{pre+rec} \quad (8)$$

III. HASIL DAN PEMBAHASAN

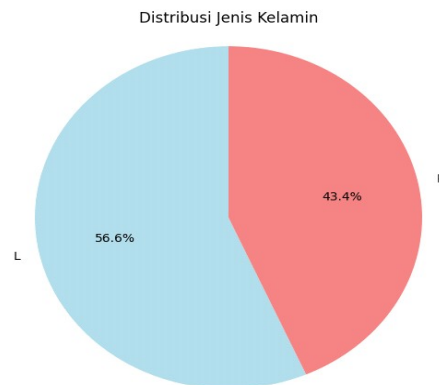
A. Analisis Deskriptif

Tabel 3. Analisis Deskriptif Data Kontinu

	BB Lahir	TB Lahir	Usia Saat Diukur	Berat	Tinggi	LiLA
Min.	1,3	38	1	1,4	41	9
1st Qu.	3	48	20	9	75,5	14
Median	3	49	30	10,4	82,3	15
Mean	3,11	48,77	41	10,47	81,66	15,04
3rd Qu.	3,3	50	41	12	89	16
Max.	6,7	68	59	19	100	50

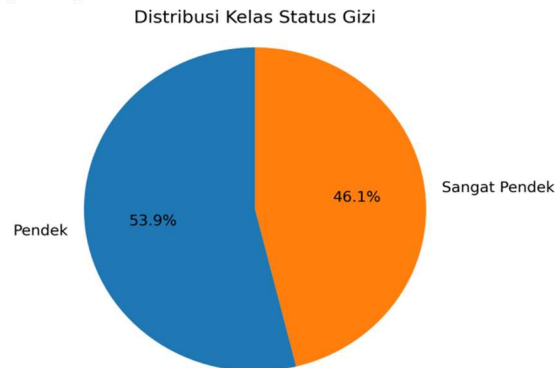
Berdasarkan Tabel 3, pada kolom usia saat diukur terdapat nilai minimum 1. Hal ini karena terdapat data bayi berusia 0 tahun 0 bulan 2 hari yang dibulatkan menjadi 1 bulan. Pembulatan dilakukan agar tidak terdapat nilai 0, sehingga data lebih konsisten dan memudahkan proses analisis. Berat badan dan tinggi badan balita menunjukkan variasi yang cukup besar, mulai dari nilai terendah hingga tertinggi dengan rentang yang lebar. Hal ini wajar karena setiap balita memiliki pertumbuhan yang berbeda-beda.

Sementara itu, pada atribut lingkaran lengan atas (LiLA), nilai yang diperoleh relatif lebih mendekati rata-rata sehingga perbedaannya antar balita tidak terlalu jauh. Dengan kata lain, LiLA cenderung lebih stabil dibandingkan berat dan tinggi badan. Terkait data hilang, dilakukan pengecekan pada seluruh atribut sebelum proses pemodelan. Apabila ditemukan nilai hilang pada variabel seperti usia, berat, tinggi badan, atau LiLA, nilai tersebut ditangani dengan menggunakan nilai pengganti yang sesuai untuk data numerik, misalnya median, agar jumlah sampel tetap dapat digunakan. Namun, jika terdapat nilai hilang pada variabel target TB/U, maka data tersebut tidak dimasukkan ke tahap pemodelan karena label *stunting* diperlukan untuk membentuk model dan evaluasi.



Gambar 2. Diagram Lingkaran Jenis Kelamin

Berdasarkan Gambar 2 terlihat banyaknya balita laki-laki lebih banyak daripada balita perempuan. Banyaknya balita laki-laki mencapai 56,6% (754 dari 1.334 populasi), sedangkan balita perempuan sebesar 43,4% (579 dari 1.334 populasi). Hal ini menunjukkan bahwa pada data yang dianalisis, balita laki-laki lebih mendominasi dibandingkan balita perempuan.



Gambar 3. Diagram Lingkaran Status *Stunting* Balita

Berdasarkan Gambar 3 didapatkan, mayoritas balita berada pada kategori stunting pendek. Dari total 1.334 balita, sebanyak 718 balita (53,9%) tergolong pendek, sedangkan 615 balita (46,14%) tergolong sangat pendek. Diagram tersebut menunjukkan bahwa jumlah balita pada kelas Pendek sedikit lebih banyak dibandingkan kelas Sangat Pendek, dengan selisih sekitar 7,72 poin persentase atau 103 data. Kondisi ini mengindikasikan bahwa dataset mengalami ketidakseimbangan kelas (*class imbalance*), meskipun tingkat ketidakseimbangannya relatif ringan (*slight imbalance*) karena proporsi kedua kelas masih cukup berdekatan.

B. Kinerja Klasifikasi

Tabel 4. Hasil Akurasi Klasifikasi

Fold ke-	Akurasi	Precision	Recall	F1-score
1	85,61%	72,41%	65,62%	68,85%
2	82,58%	63,64%	65,62%	64,62%
3	82,58%	62,16%	71,88%	66,67%
4	83,33%	67,86%	59,38%	63,33%
5	75,76%	50,00%	62,50%	55,56%
6	84,85%	70,00%	65,62%	67,74%
7	84,85%	62,50%	93,75%	75,00%

8	79,55%	56,41%	68,75%	61,97%
9	81,68%	60,00%	67,74%	63,64%
10	86,26%	73,33%	68,75%	70,97%

Berdasarkan tabel diatas, hasil akurasi menggunakan *Naïve Bayes Classifier* dapat dilihat bahwa akurasi tertinggi berada pada iterasi ke-10 dengan akurasi 86,26%, sedangkan akurasi terendah berada pada iterasi ke-5 dengan akurasi sebesar 75,76%. Secara keseluruhan rata-rata akurasi pada setiap iterasi didapat sebesar 82,70%.

Evaluasi kinerja dari proses klasifikasi juga dapat diukur dengan menggunakan *confusion matrix* untuk melihat ketepatan *Naïve Bayes Classifier* dalam menggolongkan data pada setiap kelas yang akan diklasifikasi. Berikut merupakan salah satu *confusion matrix* pada *fold* ke-10 dapat dilihat pada tabel 5.

Tabel 5. *Confusion Matrix Fold 10*

Aktual	Prediksi	
	Pendek	Sangat Pendek
Pendek	91	8
Sangat Pendek	10	22

Pada Tabel 5 terlihat bahwa model Algoritma *Naïve Bayes Classifier* (NBC) memprediksi balita dengan status *stunting* pendek sebanyak 91 dan sangat pendek sebanyak 22. Sementara itu, pada data aktual terdapat 99 balita dengan status *stunting* pendek dan 32 balita dengan status *stunting* sangat pendek. Berdasarkan perbandingan antara hasil prediksi dan data aktual tersebut, dapat dihitung nilai presisi, *recall*, dan *f1-score* dengan perhitungan sebagai berikut.

1. Presisi

$$pre = \frac{TP}{TP+FP} = \frac{22}{22+8} = 0,7333 \text{ (73,33\%)}$$

2. Recall

$$re = \frac{TP}{TP+F} = \frac{22}{22+2} = 0,6875 \text{ (68,75\%)}$$

3. *F1-Score*

$$f1\text{-score} = \frac{2 \times pre \times re}{pre + re} = \frac{2 \times 0,7333 \times 0,6875}{0,7333 + 0,6875} = 0,7097 \text{ (70,97\%)}$$

Berdasarkan hasil perhitungan, dapat disimpulkan bahwa model memiliki tingkat ketepatan yang sangat tinggi dalam memprediksi kategori sangat pendek, yang ditunjukkan oleh nilai presisi sebesar 73,33%, sehingga kemungkinan kesalahan berupa balita yang sebenarnya tidak sangat pendek tetapi diprediksi sebagai sangat pendek relatif kecil. Namun, nilai recall sebesar 68,75% menunjukkan bahwa masih terdapat sekitar 31,25% balita yang sebenarnya termasuk kategori sangat pendek tetapi belum berhasil terdeteksi oleh model, sehingga masih ada risiko keterlambatan penanganan pada sebagian kasus. Sementara itu, *F1-score* sebesar 70,97% menunjukkan keseimbangan yang cukup baik antara presisi dan recall, sehingga secara keseluruhan model dapat dikatakan cukup andal dalam mengklasifikasikan kasus *stunting* kategori sangat pendek.

C. Karakteristik *Stunting* (Pendek dan Sangat Pendek)

Tabel 6. Rata-Rata Variabel

Variabel	Pendek	Sangat Pendek
Jenis Kelamin	0,564	0,567
Berat Badan Lahir	3,125	3,100
Tinggi Badan Lahir	48,806	48,735
Usia Saat diukur	32,025	29,164
Berat Badan	10,832	10,049
Tinggi Badan	83,430	79,614
LiLA	15,203	14,864

Berdasarkan Tabel 6, hasil pelatihan model, ciri-ciri balita yang mengarah pada kelas pendek adalah:

- a. Memiliki tinggi badan saat diukur sekitar 83,43 cm.
 - b. Memiliki berat badan saat diukur sekitar 10,83 kg.
 - c. Memiliki LiLA sekitar 15,20 cm.
 - d. Memiliki tinggi badan lahir sekitar 48,81 cm.
 - e. Memiliki berat badan lahir sekitar 3,13 kg.
 - f. Diukur pada usia sekitar 32 bulan.
 - g. Jenis kelamin tidak menjadi pembeda yang kuat karena proporsi laki-laki dan perempuan hampir sama.
- Dari seluruh variabel, tinggi badan saat diukur, berat badan saat diukur, dan LiLA merupakan karakteristik yang paling menggambarkan kelas Pendek karena mempunyai skor rata-rata yang lebih besar daripada kelas Sangat Pendek.

Balita akan memiliki probabilitas lebih besar diklasifikasikan ke dalam kelas Sangat Pendek apabila memiliki karakteristik sebagai berikut.

- a. Memiliki tinggi badan saat diukur sekitar 79,61 cm.
- b. Memiliki berat badan saat diukur sekitar 10,05 kg.
- c. Memiliki LiLA sekitar 14,86 cm.
- d. Memiliki tinggi badan lahir sekitar 48,73 cm.
- e. Memiliki berat badan lahir sekitar 3,10 kg.
- f. Diukur pada usia sekitar 29 bulan.
- g. Jenis kelamin juga tidak menunjukkan perbedaan yang berarti dibandingkan kelas Pendek.

Karakteristik tersebut menunjukkan bahwa balita dengan tinggi badan, berat badan, dan LiLA yang lebih rendah memiliki kecenderungan lebih besar untuk masuk ke dalam kelas Sangat Pendek.

Hasil penelitian ini menunjukkan bahwa *Naïve Bayes Classifier* dapat digunakan untuk mengklasifikasikan status stunting balita ke dalam kategori Pendek dan Sangat Pendek. Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini sejalan dengan penelitian Arisandi et al. (2022) yang menerapkan metode *Naïve Bayes Classifier* dengan *K-Fold Cross Validation* untuk klasifikasi status gizi balita stunting. Penelitian tersebut memperoleh rata-rata akurasi sebesar 88,46%, sedangkan penelitian ini memperoleh rata-rata akurasi sebesar 82,70%. Perbedaan hasil tersebut dapat dipengaruhi oleh karakteristik data dan variabel yang digunakan pada masing-masing penelitian. Meskipun demikian, hasil tersebut menunjukkan bahwa metode *Naïve Bayes Classifier* dapat memberikan kinerja klasifikasi yang baik dalam mengidentifikasi kategori stunting.

Penelitian ini juga masih memiliki keterbatasan karena data yang digunakan hanya berasal dari Kabupaten Pasaman tahun 2022, sehingga hasil penelitian belum dapat menggambarkan kondisi pada wilayah atau periode yang berbeda. Selain itu, penelitian ini hanya menggunakan metode *Naïve Bayes Classifier* sehingga belum diketahui perbandingan kinerjanya dengan metode klasifikasi lainnya. Oleh karena itu, penelitian selanjutnya diharapkan dapat menggunakan data dengan cakupan wilayah dan periode yang lebih luas serta membandingkan beberapa metode klasifikasi, sehingga diperoleh hasil yang lebih komprehensif dan dapat mendukung pengembangan penelitian mengenai klasifikasi status stunting pada balita.

IV. KESIMPULAN

Hasil klasifikasi menggunakan metode *Naïve Bayes Classifier* (NBC) pada data *stunting* balita di Kabupaten Pasaman tahun 2022 yang diperoleh dari Dinas Kesehatan menunjukkan bahwa model memiliki kinerja yang baik. Evaluasi dilakukan menggunakan metode *10-Fold Cross Validation* dan diukur melalui *confusion matrix* dengan indikator akurasi, presisi, *recall*, dan *F1-score*. Berdasarkan hasil evaluasi, model menghasilkan nilai akurasi sebesar 86,26%, presisi sebesar 73,33%, *recall* sebesar 68,75%, dan *F1-score* sebesar 70,97%. Hasil tersebut menunjukkan bahwa model memiliki performa klasifikasi yang baik dalam membedakan kelas *stunting* balita. Tingginya nilai akurasi dan presisi mengindikasikan bahwa model mampu menghasilkan prediksi yang tepat dengan tingkat kesalahan yang relative rendah, sehingga layak digunakan untuk mendukung proses klasifikasi data *stunting*.

Selain mengevaluasi performa klasifikasi, analisis karakteristik kelas berdasarkan *Naïve Bayes Classifier* (NBC) menunjukkan bahwa tinggi badan saat diukur, berat badan saat diukur, dan lingkaran lengan atas (LiLA) merupakan karakteristik utama yang membedakan kelas Pendek dan Sangat Pendek. Balita dengan tinggi badan sekitar 79,61 cm, berat badan sekitar 10,05 kg, dan LiLA sekitar 14,86 cm memiliki probabilitas yang lebih tinggi untuk diklasifikasikan ke dalam kelas Sangat Pendek. Sebaliknya, balita dengan tinggi badan sekitar 83,43 cm, berat badan sekitar 10,83 kg, dan LiLA sekitar 15,20 cm cenderung diklasifikasikan ke dalam kelas Pendek. Hasil

penelitian ini memberikan manfaat dalam mengidentifikasi karakteristik *antropometri* yang paling berperan dalam membedakan tingkat keparahan *stunting* pada balita. Informasi tersebut dapat dimanfaatkan sebagai pegangan tenaga kesehatan mendeteksi dini serta pengelompokan tingkat *stunting* yang lebih akurat berdasarkan indikator TB, BB, serta lingkaran lengan atas (LiLA).

Saran untuk penelitian berikutnya adalah memperbanyak jumlah dataset dan lebih bervariasi, baik dari segi jumlah data maupun cakupan wilayah. Dengan demikian, model yang dihasilkan diharapkan memiliki kemampuan generalisasi yang lebih baik dalam mengklasifikasikan status *stunting* pada beragam karakteristik populasi.

DAFTAR PUSTAKA

- Arisandi, R. R. R., Warsito, B., & Hakim, A. R. (2022). Aplikasi Naïve Bayes Classifier (NBC) pada Klasifikasi Status Gizi Balita Stunting dengan Pengujian K-Fold Cross Validation. *Jurnal Gaussian*, 11(1), 130–139.
- Badan Pusat Statistik Provinsi Sumatera Barat. (2023). *Sumatera Barat dalam Angka 2023*. Sumatera Barat: Badan Pusat Statistik Provinsi Sumatera Barat. <https://sumbar.bps.go.id>.
- Chen, H. (2021). Improved Naive Bayes Classification Algorithm for Traffic Risk Prediction. *EURASIP Journal on Advances in Signal Processing*, 2021(1), 2-12. <https://doi.org/10.1186/s13634-021-00769-3>
- Dinas Kesehatan Kabupaten Pasaman. (2022). *Profil Kesehatan Kabupaten Pasaman Tahun 2022*. Pasaman: Dinas Kesehatan Kabupaten Pasaman.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques. Third Edition*. Morgan Kaufmann.
- Iskandar, S. (2024). Trends in Stunting Prevalence Reduction: An Examination of Data Toward Achieving the 2024 Target in Indonesia. *The Social Perspective Journal*, 3(1), 51–68. <https://doi.org/10.53947/tspj.v3i1.795>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer.
- Kementerian Kesehatan Republik Indonesia. (2023). *Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Jakarta:Kementerian Kesehatan RI.
- Musu, dkk. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5. *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, 10(1), 186 – 195.
- Nurainun, N., Haerani, E., Syafria, F., & Oktavia, L. (2023). Penerapan Algoritma Naïve Bayes Classifier dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation. *Journal of Computer System and Informatics (JoSYCI)*, 4(3), 578–586. <https://doi.org/10.47065/josyc.v4i3.3414>
- Nwanganga, F., & Chapple, M. (2020). *Practical Machine Learning in R*. Wiley.
- Raschka, S., & Mirjalili, V. (2020). *Python Machine Learning (3rd Ed.)*. Packt Publishing.
- Usnah, A., Hasan, F. N., & Kuntoro, A. Y. (2025). Analisis Kepuasan Pengguna Aplikasi GoPay menggunakan Algoritma Naïve Bayes dan K-Fold Cross Validation. *Jurnal Ilmiah Informatika*, 13(2), 128–133. <https://doi.org/10.33884/jif.v13i02.10276>
- VanderPlas, J. (2022). *Python Data Science Handbook* (2nd ed.). O'Reilly Media.