

Penerapan Algoritma *Naïve Bayes Classifier* Untuk Klasifikasi Status Gizi Stunting Pada Balita

Fajar Maula Hidayat¹, Kusriani¹, Ainul Yaqin²

¹PJJ Informatika, Universitas AMIKOM Yogyakarta, Jalan Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281, Indonesia

²Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, Jalan Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281, Indonesia

ARTICLE INFO

Article history:

Received March 4, 2024

Revised March 20, 2024

Accepted August 31, 2024

Keywords:

Stunting;

Machine Learning;

Naïve Bayes Classifier;

ABSTRACT

Child stunting is a major public health concern in Indonesia. This study uses the Naïve Bayes classification algorithm to assess the nutritional condition of stunted children based on demographic and anthropometric characteristics. The information used comes from the Toddler Weighing Month (Bulan Penimbangan Balita - BPB) in Majalengka Regency. Data type conversion, separating data into training and testing sets, and data normalization are all examples of preprocessing steps. The model's evaluation results reveal an accuracy of 94.65%, with precision and recall for each category of stunted nutritional status. This study makes a substantial contribution to early diagnosis and mitigation of stunting in Indonesia, as well as providing the framework for future development of more powerful predictive models.

Stunting pada anak merupakan masalah kesehatan masyarakat yang signifikan di Indonesia. Penelitian ini bertujuan untuk mengaplikasikan algoritma klasifikasi *Naïve Bayes Classifier* dalam mengidentifikasi status gizi stunting pada balita berdasarkan atribut demografis dan antropometri. Data yang digunakan berasal dari Bulan Penimbangan Balita (BPB) Kabupaten Majalengka. Langkah-langkah *preprocessing data* meliputi perubahan tipe data, pembagian data latih dan uji, serta standarisasi data. Hasil evaluasi model menunjukkan akurasi sebesar 94.65%, dengan *precision* dan *recall* untuk setiap kategori status gizi stunting. Penelitian ini memberikan kontribusi penting dalam upaya deteksi dini dan penanggulangan stunting di Indonesia, serta menyediakan landasan untuk pengembangan model prediksi yang lebih canggih di masa depan.

Corresponding Author:

Fajar Maula Hidayat, Program S2 PJJ Informatika, Pascasarjana, Universitas AMIKOM Yogyakarta, Jalan Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281
Email: fajarmaulahidayat@students.amikom.ac.id

1. PENDAHULUAN

Stunting adalah jenis kegagalan pertumbuhan yang disebabkan oleh kekurangan nutrisi yang berkelanjutan [1]. Kondisi ini menyebabkan anak-anak mengalami keterbelakangan pertumbuhan tinggi badan pada usia tertentu [2]. Selain meningkatkan morbiditas dan mortalitas, kekurangan gizi juga mengganggu perkembangan anak dalam aspek kognitif, motorik, dan bahasa, serta meningkatkan beban biaya perawatan medis [3]. Dalam jangka panjang, dampak kekurangan gizi juga mempengaruhi kesehatan reproduksi, kemampuan konsentrasi belajar, dan produktivitas kerja [4]. Dampak negatif ini dapat menciptakan kerugian yang lebih besar bagi negara, terutama pada anak-anak yang mengalami stunting dan tumbuh dewasa dengan tingkat pendidikan yang rendah, kemiskinan, kesehatan yang buruk, serta rentan terhadap penyakit tidak menular seperti obesitas dan penyakit jantung [5].

Indonesia, sebagai salah satu negara berpendapatan menengah, memiliki tingkat prevalensi stunting tertinggi [6]. Meskipun angka tersebut lebih rendah daripada Myanmar (35%), namun tetap lebih tinggi jika dibandingkan dengan Vietnam (23%), Malaysia (17%), Thailand (16%), dan Singapura (4%) [7]. Menurut Kementerian Kesehatan, hasil Survei Status Gizi Indonesia (SSGI) menunjukkan penurunan dari 24,4% pada

tahun 2021 menjadi 21,6% pada tahun 2022, seperti yang diumumkan dalam Rapat Kerja Nasional BKKBN pada tahun 2023. Standar WHO menetapkan bahwa prevalensi stunting seharusnya berada di bawah 20%, dan Pemerintah menetapkan target untuk menurunkan tingkat stunting hingga 14% pada tahun 2024 [8].

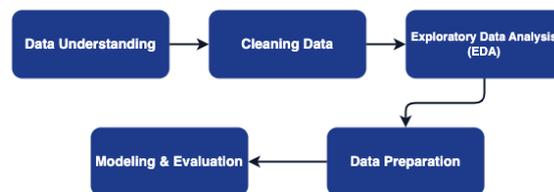
Hasil SSGI tahun 2022 menunjukkan bahwa ada 20,2% balita di Jawa Barat mengalami stunting. Kabupaten Majalengka berada di antara sepuluh kota dan kabupaten dengan status gizi tertinggi di Jawa Barat, dengan prevalensi mencapai 24,3% [9]. Bulan Penimbangan Balita (BPB) adalah kegiatan rutin yang dilakukan oleh posyandu setiap bulan Agustus, dimana balita usia 0-59 bulan ditimbang berat badannya dan diukur panjang atau tingginya. Data Rekapitulasi Hasil BPB tahun 2022 di Kabupaten Majalengka, yang dikumpulkan oleh Dinas Kesehatan Kabupaten Majalengka menunjukkan bahwa kecamatan Sumberjaya memiliki tingkat balita stunting tertinggi.

Antropometri adalah metode langsung yang digunakan untuk menilai status gizi seseorang, khususnya terkait masalah kekurangan energi dan protein (KEP). Faktor genetik dan lingkungan, seperti pola makan dan kondisi kesehatan, turut memengaruhi hasil antropometri [10]. Meskipun antropometri memiliki beberapa keuntungan, seperti prosedur yang mudah, aman, dan dapat dilakukan dalam sampel besar, serta memungkinkan penapisan kelompok rawan gizi, namun terdapat juga beberapa kelemahan. Kelemahan tersebut antara lain ketidakmampuan untuk memberikan informasi status gizi dengan cepat, kemungkinan dipengaruhi oleh faktor-faktor yang tidak berhubungan dengan gizi, dan adanya kesalahan pengukuran yang dapat mengurangi ketepatan dan akurasi hasil pengukuran [11].

Metode yang lebih maju seperti *machine learning*, harus digunakan untuk mengatasi masalah ini. Dengan menggunakan teknik *data mining* yang melibatkan *machine learning*, statistik matematika, dan kecerdasan buatan (AI), kita dapat memprediksi kondisi status gizi balita dalam jangka panjang seperti stunting. Penggunaan *machine learning* memungkinkan untuk menemukan pola-pola yang tersembunyi dalam data gizi dan memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi status gizi balita. *Machine learning* dapat membantu mendeteksi status gizi dengan lebih akurat dan tepat. Ini dapat membantu dalam pencegahan dan intervensi yang lebih efektif untuk meningkatkan kesehatan balita dan mencegah masalah gizi buruk di masa depan [12].

2. METODE

Metode penelitian yang digunakan dalam menerapkan algoritma *Naïve Bayes Classifier* untuk mengklasifikasi status gizi stunting pada balita mencakup beberapa tahap. 1. *Data Understanding*; 2. *Data Cleaning*; 3. *Exploratory Data Analysis*; 4. *Data Preparation*; 5. Penerapan model algoritma *Naïve Bayes Classifier* dan evaluasi.



Gambar 1. Tahapan *machine learning*

2.1. Data Understanding

Penelitian ini menggunakan *dataset* yang diperoleh dari Bulan Penimbangan Balita (BPB), yang disediakan oleh Dinas Kesehatan Kabupaten Majalengka. *Dataset* ini terdiri dari 1000 entri data dengan 7 atribut. Dimana 6 atribut berupa *input* dan 1 atribut berupa target. Target disini yaitu mengindikasikan status stunting pada balita, yang dikategorikan normal, pendek, atau sangat pendek. Berikut merupakan deskripsi dari masing-masing atribut pada *dataset*:

Tabel 1. Deskripsi masing-masing atribut

No.	Atribut	Deskripsi
1.	JK	Jenis Kelamin
2.	Usia	Usia balita dari 1 bulan sampai 60 bulan
3.	Berat	Berat badan balita
4.	Tinggi	Panjang atau tinggi badan balita
5.	LiLA	Lingkar lengan atas balita
6.	ZS TB/U	Z-Score tinggi badan per umur
7.	TB/U	Tinggi badan per umur

2.2. Cleaning Data

Untuk memastikan kualitas data, dilakukan tiga proses pada tahap *cleaning data*. Pertama, dilakukan pengecekan terhadap nilai yang hilang (*missing value*) pada setiap kolom data guna memastikan ketiadaan kekosongan data. Kedua, dilakukan pengecekan terhadap adanya data duplikat untuk mencegah duplikasi yang tidak diinginkan serta memastikan konsistensi data. Terakhir, dilakukan pengecekan terhadap data yang memiliki nilai 0 (nol) untuk menilai keberadaan nilai tersebut rasional serta memastikan tidak mengganggu terhadap analisis yang dilakukan.

2.3. Exploratory Data Analysis (EDA)

Analisis data eksploratif (EDA) adalah proses untuk memahami *dataset* dengan menggunakan visualisasi data. EDA fokus pada penemuan pola, keterkaitan, anomali, dan wawasan penting dalam dataset [13]. Fokus analisis meliputi distribusi tinggi badan, usia balita, proporsi jenis kelamin, status gizi stunting, dan distribusi status tinggi badan per umur (TB/U) menurut jenis kelamin.

2.4. Data Preparation

Pada tahap *data preparation*, dilakukan 3 tahapan yaitu *label encoding*, *split data*, dan standarisasi data.

a. Label Encoding

Label encoding adalah teknik dalam *machine learning* dan *preprocessing data* yang mengubah data kategorikal, yang ditunjukkan dengan label teks, menjadi format numerik. Dengan kata lain, memberi nilai numerik yang unik untuk setiap kategori atau label yang terdapat dalam variabel kategori [14].

b. Split Data

Data latih dan data uji dipisahkan dari *dataset* saat pemisahan data. Tujuannya adalah untuk menguji data dengan titik data yang belum diketahui daripada dengan titik data yang sama dengan model yang dilatih. Proses pemisahan dilakukan secara acak. Data latih akan digunakan untuk melatih klasifikasi dalam mengenali karakteristik balita yang mengalami status gizi stunting berdasarkan kategori. Data uji digunakan untuk menguji model klasifikasi yang telah dibuat, dengan cara membandingkan hasil klasifikasi model dengan label sebenarnya dari setiap data dalam data uji [15].

c. Standarisasi Data

Proses standarisasi data dilakukan melalui metode *preprocessing* yang disebut *Standard Scaler*, metode ini bertujuan untuk menghilangkan nilai rata-rata dan menyesuaikannya dengan unit varians. Setiap fitur dalam sampel data menjalani proses ini untuk memastikan konsistensi dalam skala data. Standarisasi ini dilakukan untuk mengatasi ketidakseimbangan skala yang dapat mengganggu proses pelatihan model. Rumus matematis dari *standar scaler*, didefinisikan dalam persamaan (1), dimana X merupakan nilai rata-rata sampel dan σ adalah standar deviasi [16].

$$z = \frac{X - \mu}{\sigma} \quad (1)$$

2.5. Penerapan *machine learning* (Naïve Bayes Classifier)

Metode pengklasifikasian probabilitas sederhana seperti *Naïve Bayes Classifier* telah terbukti efektif dalam banyak kasus penelitian, termasuk dalam situasi dimana terdapat hubungan yang kompleks antara variabel-variabel yang diamati. Keunggulan *naïve bayes classifier* terletak pada kemampuannya untuk menangani data dimensi tinggi dan jumlah sampel relative kecil [17]. Thomas Bayes juga menciptakan metode klasifikasi berdasarkan probabilitas statistik yang memprediksi peluang dimasa depan berdasarkan pengalaman dari masa yang lalu, dibandingkan dengan beberapa algoritma klasifikasi lainnya, *naïve bayes classifier* memungkinkan pemrosesan yang cepat dan efisien serta tidak memerlukan parameter yang kompleks untuk disesuaikan [18]. Teorema Bayes memiliki bentuk umum pada persamaan (2) [19]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Dimana,

- X : Data *class* yang belum diketahui
- H : Hipotesis Data
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X
- P(H) : Probabilitas hipotesis H
- P(X|H) : Probabilitas X berdasar kondisi pada hipotesis H
- P(X) : Probabilitas X

2.6. Evaluasi menggunakan *confusion matrix*

Evaluasi kinerja model klasifikasi menggunakan *confusion matrix* bertujuan untuk mengukur efektivitas model pasca pelatihan, dengan mengidentifikasi akurasi dan metrik lainnya seperti, *precision*, *recall*, dan *f-1 score*. *Confusion matrix* memungkinkan pengambilan informasi secara rinci mengenai performa model melalui analisis hasil klasifikasi aktual, yang memfasilitasi perhitungan metrik evaluatif seperti *accuracy*, *precision*, *recall*, dan *f-1 score*, berdasarkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

		Predicted Class	
		False (0)	True (1)
Actual Class	False (0)	TN True Negative	FP False Positif
	True (1)	FN False Negative	TP True Positive

Gambar 2. *Confusion matrix*

Dimana,

True Positive (TP) : Jumlah dari data aktual positif yang diprediksi sebagai positif

False Positive (FP) : Jumlah dari data aktual negatif yang diprediksi sebagai positif

True Negative (TN) : Jumlah dari data aktual negatif yang diprediksi sebagai negatif

False Negative (FN) : Jumlah dari data aktual positif yang diprediksi sebagai negatif

Accuracy

Accuracy merupakan nilai rasio dari jumlah prediksi yang bernilai positif dengan jumlah dari sampel. *Accuracy* dirumuskan dengan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision

Precision merupakan nilai rasio dari jumlah prediksi yang bernilai positif dengan jumlah dari prediksi yang bernilai positif. *Precision* dirumuskan dengan:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Recall

Recall merupakan nilai rasio dari hasil prediksi yang benar positif dengan jumlah dari nilai aktual positif. *Recall* dirumuskan dengan:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F-1 Score

F-1 Score merupakan nilai rasio *harmonic mean* dari *Precision* dan *Recall*. *F-1 Score* dirumuskan dengan:

$$F - 1 \text{ Score} = \frac{2(Precision \times Recall)}{Precision + Recall} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1. *Data Understanding*

Langkah awal dalam melakukan analisis *naïve bayes classifier* adalah dengan menyiapkan data. *Dataset* berformat file *.xls* diimport ke *Google Colab* menggunakan *library pandas*. *Dataset* status gizi stunting yang digunakan dalam penelitian ini terdiri dari 1.000 data dengan 6 atribut *input* dan 1 atribut target. *Dataset* yang telah diimport dapat dilihat dalam gambar berikut:

	JK	Usia	Berat	Tinggi	LiLA	ZS TB/U	TB/U
0	P	30	12.4	86.8	16.0	-0.98	Normal
1	L	58	12.8	99.5	15.0	-1.98	Normal
2	P	55	14.0	103.8	15.5	-0.58	Normal
3	L	54	12.2	96	14.0	-2.34	Pendek
4	L	53	14.2	98.3	16.0	-1.65	Normal
...
995	P	50	10.4	90	13.5	-3.13	Sangat Pendek
996	P	47	15.0	99	16.8	-0.73	Normal
997	P	44	12.5	97.5	14.3	-0.55	Normal
998	L	43	10.8	89	14.3	-2.78	Pendek
999	P	42	11.2	92.4	13.6	-1.58	Normal

1000 rows x 7 columns

Gambar 3. Dataset status gizi stunting

Dataset yang digunakan terdiri dari tiga tipe data, yaitu *object*, *float*, dan *integer*. Atribut jenis kelamin (JK) dan tinggi badan per umur (TB/U) memiliki tipe data *object*, sementara atribut Usia memiliki tipe data *integer*, sedangkan atribut lainnya memiliki tipe data *float*. Informasi selengkapnya dapat dilihat pada gambar berikut:

```

JK          object
Usia       int64
Berat      float64
Tinggi     float64
LiLA       float64
ZS TB/U    float64
TB/U       object
dtype: object
    
```

Gambar 4. Tipe dataset status gizi stunting

3.2. Cleaning Data

Setelah memahami karakteristik data, tahapan selanjutnya adalah melakukan proses *cleaning data* untuk memastikan kualitas data yang digunakan dalam analisis *naïve bayes classifier*. Berikut adalah langkah-langkah yang dilakukan dalam proses *cleaning data*:

1. Cek nilai yang hilang (*Missing value*)
 Pada tahap ini, dilakukan pemeriksaan terhadap nilai yang hilang pada tiap kolom *dataset*. Ditemukan bahwa terdapat nilai yang hilang pada kolom TB/U. Mengingat pentingnya informasi tinggi badan untuk usia balita dalam menentukan status gizi stunting, maka baris data yang memiliki nilai *missing value* pada kolom TB/U dihapus.
2. Cek data duplikat
 Data duplikat dapat menyebabkan bias dalam analisis. Oleh karena itu, dilakukan pemeriksaan terhadap data duplikat pada *dataset*. Hasilnya, tidak ditemukan data duplikat dalam *dataset*.
3. Cek data bernilai 0
 Nilai 0 pada beberapa atribut dapat memiliki interpretasi yang berbeda. Pada atribut LiLA, nilai 0 dapat diindikasikan bahwa pengukuran lingkaran lengan atas tidak dilakukan. Oleh karena itu, baris data yang memiliki nilai 0 pada atribut LiLA dihapus.

Tabel 2. Cleaning data

No.	Atribut	Missing Value	Duplikat	Bernilai 0
1.	JK	0	0	0
2.	Usia	0	0	0
3.	Berat	0	0	0
4.	Tinggi	6	0	0
5.	LiLA	0	0	91
6.	ZS TB/U	6	0	0
7.	TB/U	6	0	0

4. Hasil *Cleaning Data*

Setelah melakukan proses *cleaning data*, *dataset* yang dihasilkan memiliki 908 data dengan 6 atribut *input* dan 1 atribut target. Berikut adalah tabel ringkasan hasil *cleaning data*:

	JK	Usia	Berat	Tinggi	LiLA	ZS TB/U	TB/U
0	P	30	12.4	86.8	16.0	-0.98	Normal
1	L	58	12.8	99.5	15.0	-1.98	Normal
2	P	55	14.0	103.8	15.5	-0.58	Normal
3	L	54	12.2	96.0	14.0	-2.34	Pendek
4	L	53	14.2	98.3	16.0	-1.65	Normal
...
995	P	50	10.4	90.0	13.5	-3.13	Sangat Pendek
996	P	47	15.0	99.0	16.8	-0.73	Normal
997	P	44	12.5	97.5	14.3	-0.55	Normal
998	L	43	10.8	89.0	14.3	-2.78	Pendek
999	P	42	11.2	92.4	13.6	-1.58	Normal

908 rows × 7 columns

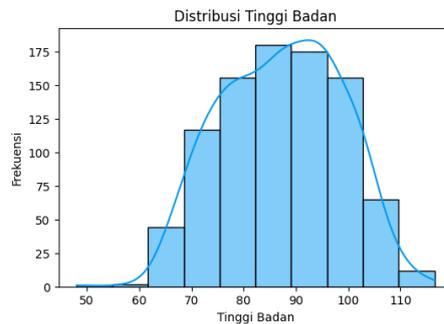
Gambar 5. *Dataset* status gizi setelah *cleaning data*

3.3. *Exploratory Data Analysis*

Setelah proses *cleaning data*, tahapan selanjutnya adalah melakukan *Exploratory Data Analysis* (EDA) untuk memahami karakteristik data dan mendapatkan gambaran awal mengenai hubungan antar variabel. Berikut adalah visualisasi yang dilakukan pada tahap EDA:

1. Distribusi Tinggi Badan

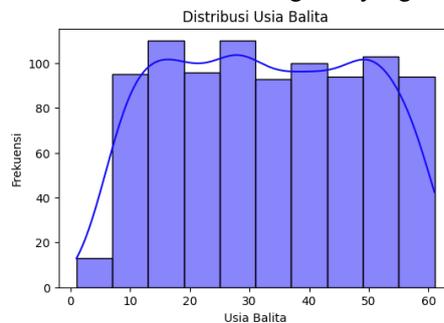
Visualisasi distribusi tinggi badan balita disajikan dalam bentuk histogram. Hasilnya menunjukkan bahwa distribusi tinggi badan balita rata-rata pada 80–100 cm dengan *value* sebanyak 150–175 balita.



Gambar 6. Distribusi tinggi badan

2. Distribusi Usia Balita

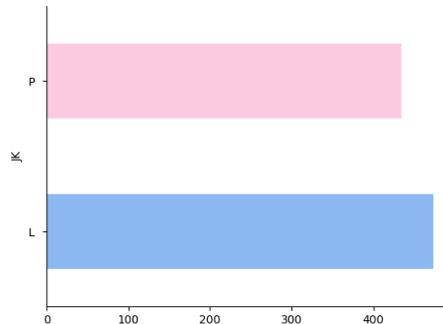
Distribusi usia balita juga disajikan dalam bentuk histogram. Hasilnya menunjukkan bahwa distribusi usia balita relatif merata di seluruh rentang usia yang diteliti.



Gambar 7. Distribusi usia balita

3. Proporsi Jenis Kelamin antara Laki-laki dan Perempuan

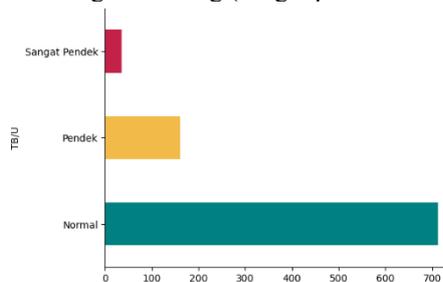
Proporsi jenis kelamin balita divisualisasikan dengan menggunakan *pie chart*. Hasilnya menunjukkan bahwa terdapat proporsi yang hampir sama antara balita laki-laki dan perempuan. Dimana jenis kelamin laki-laki berjumlah 474 balita dan jenis kelamin perempuan berjumlah 434 balita.



Gambar 8. Proporsi jenis kelamin

4. Proporsi Status Gizi Stunting: Sangat Pendek, Pendek, dan Normal

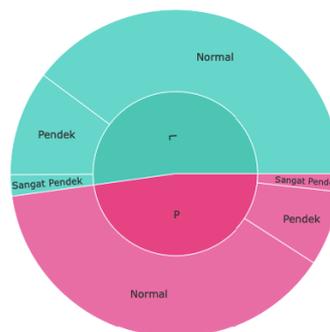
Proporsi status gizi balita dikategorikan berdasarkan klasifikasi WHO (sangat pendek, pendek, dan normal) dan disajikan dalam bentuk *pie chart*. Hasilnya menunjukkan bahwa terdapat proporsi balita yang cukup tinggi dengan status gizi stunting (sangat pendek dan pendek).



Gambar 9. Proporsi status gizi stunting

5. Distribusi Status Tinggi Badan per Umur (TB/U) menurut Jenis Kelamin

Distribusi status gizi tinggi badan per umur (TB/U) menurut jenis kelamin divisualisasikan dengan menggunakan *boxplot*. Hasilnya menunjukkan bahwa terdapat perbedaan distribusi TB/U antara balita laki-laki dan perempuan. Balita laki-laki memiliki median TB/U yang lebih tinggi dibandingkan dengan balita perempuan.



Gambar 10. Distribusi status TB/U menurut jenis kelamin

3.4. Data Preparation

Dalam *preparation data* akan dilakukan 1. Perubahan tipe data atribut jenis kelamin dari *object* menjadi integer; 2. *Split data* menjadi data latih dan data uji; dan 3. Melakukan Standarisasi data.

1. *Label encoding*

	JK	Usia	Berat	Tinggi	LiLA	ZS	TB/U	TB/U
0	1	30	12.4	86.8	16.0	-0.98	Normal	
1	0	58	12.8	99.5	15.0	-1.98	Normal	
2	1	55	14.0	103.8	15.5	-0.58	Normal	
3	0	54	12.2	96.0	14.0	-2.34	Pendek	
4	0	53	14.2	98.3	16.0	-1.65	Normal	

Gambar 11. *Label encoding* kolom jenis kelamin (JK)

- Proses *split data* membagi data menjadi dua bagian: data latih (*training set*) dan data uji (*test set*). Data dibagi dengan rasio 65% untuk data latih dan 35% untuk data uji. Sehingga menghasilkan:
 Data Latih : 590 sampel
 Data Uji : 318 sampel
- Standard Scaler* digunakan untuk menstandarkan data dengan mengurangi *mean* dan membagi dengan standar deviasi. Tujuannya adalah membuat semua fitur memiliki skala yang sama, sehingga memudahkan proses pelatihan model.

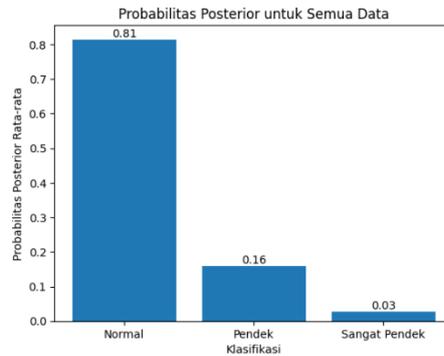
	JK	Usia	Berat	Tinggi	LiLA	ZS	TB/U
0	1.048639	-0.114437	-0.087829	0.349727	-0.144263	1.165260	
1	1.048639	-0.241950	-0.423031	-0.341013	-0.144263	-0.554895	
2	1.048639	0.714395	1.009194	1.302773	0.977149	1.525502	
3	-0.953617	-1.134538	-0.666813	-1.154162	-0.704969	-0.554895	
4	1.048639	-1.198294	-0.849651	-1.145419	-0.704969	0.291674	
...	
585	-0.953617	-0.624488	-0.575395	-0.489653	-0.368545	-0.257695	
586	1.048639	-0.688244	-0.880124	-0.577089	-0.424616	0.030498	
587	-0.953617	-0.178193	-0.179248	-0.358500	-0.256404	-1.149294	
588	1.048639	-1.262051	-1.520054	-1.486417	-1.321745	-0.653961	
589	1.048639	-0.369462	-0.575395	-0.708242	0.080019	-1.248360	

[590 rows x 6 columns]

Gambar 12. Standarisasi data menggunakan *standard scaler*

3.5. Penerapan Algoritma *Naïve Bayes Classifier*

Penerapan algoritma *naïve bayes classifier* akan dilakukan pada data latih dan data uji yang sudah dibuat sebelumnya. Pada penerapan *naïve bayes classifier* akan dicari nilai akurasi menggunakan persamaan (2). Dengan bantuan Bahasa pemrograman Python dengan *tools* Google Colab akan didapatkan nilai klasifikasi.



Gambar 13. Probabilitas posterior untuk semua data

Nilai probabilitas posterior rata-rata pada grafik menunjukkan proporsi data dalam setiap kelas yang diklasifikasikan sebagai kelas tersebut. Berikut adalah interpretasi nilai pada grafik:

Klasifikasi Normal

Probabilitas posterior rata-rata: 0.81

Interpretasi: **81%** data dalam kategori **Normal** diklasifikasikan dengan benar sebagai **Normal**

Klasifikasi Pendek

Probabilitas posterior rata-rata: 0.16

Interpretasi: **16%** data dalam kategori **Pendek** diklasifikasikan dengan benar sebagai **Pendek**

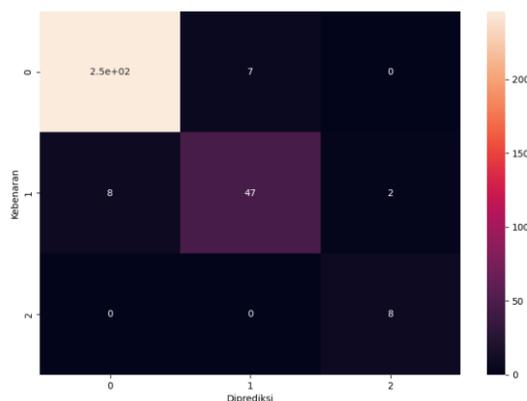
Klasifikasi Sangat Pendek

Probabilitas posterior rata-rata: 0.03

Interpretasi: **3%** data dalam kategori **Sangat Pendek** diklasifikasikan dengan benar sebagai **Sangat Pendek**

3.6. Evaluasi Model

Setelah menerapkan algoritma *naïve bayes classifier* pada *dataset* yang digunakan, langkah selanjutnya akan dilakukan evaluasi model. Evaluasi dilakukan dengan bahasa pemrograman Python dengan *tools* Google Colab untuk menghasilkan *confusion matrix*. Selanjutnya, dengan *confusion matrix* kita dapat menghitung nilai *accuracy*, *precision*, *recall*, dan *f-1 score*. Dengan nilai klasifikasi yang didapatkan hasil evaluasi model sebagai berikut:



Gambar 14. Confusion matrix

Dengan hasil penerapan algoritma *naïve bayes classifier* pada klasifikasi status gizi stunting, diperoleh:

Table 3. Hasil *confusion matrix*

MetriK	Normal	Pendek	Sangat Pendek
Accuracy	0.95	0	0
Precision	0.97	0.87	0.80
Recall	0.97	0.82	1.00
F-1 Score	0.97	0.85	0.89

Simulasi perhitungan performa algoritma *naïve bayes classifier* dapat dihitung dengan menggunakan formula pada persamaan (3), (4), (5), dan (6) sebagai berikut:

$$Accuracy : \frac{246+47+8}{246+7+0+8+47+2+0+0+8} = \frac{301}{318} = 0.95$$

$$Precision \text{ Normal} : \frac{246}{246+8+0} = \frac{246}{254} = 0.97$$

$$Precision \text{ Pendek} : \frac{47}{47+7+0} = \frac{47}{54} = 0.87$$

$$Precision \text{ Sangat Pendek} : \frac{8}{8+2+0} = \frac{8}{10} = 0.80$$

$$Recall \text{ Normal} : \frac{246}{246+7+0} = \frac{246}{253} = 0.97$$

$$Recall \text{ Pendek} : \frac{47}{8+47+2} = \frac{47}{57} = 0.82$$

$$Recall \text{ Sangat Pendek} : \frac{8}{0+0+8} = \frac{8}{8} = 1.00$$

$$F-1 \text{ Score Normal} : \frac{2 \times (0.97 \times 0.97)}{0.97 + 0.97} = \frac{1.88}{1.94} = 0.97$$

$$F-1 \text{ Score Pendek} : \frac{2 \times (0.87 \times 0.82)}{0.87 + 0.82} = \frac{1.44}{1.69} = 0.85$$

$$F-1 \text{ Score Sangat Pendek} : \frac{2 \times (0.80 \times 1.00)}{0.80 + 1.00} = \frac{1.60}{1.80} = 0.89$$

3.7. Perbandingan Model

Beberapa penelitian sebelumnya telah mengeksplorasi algoritma klasifikasi untuk status gizi stunting, termasuk penelitian oleh Reza dan Rohman [20] menggunakan *Random Forest* dan penelitian lainnya Lonang dan Normawati [21] menggunakan *K-Nearest Neighbor*. Penelitian ini menerapkan *Naïve Bayes Classifier* dan mencapai tingkat akurasi 94.65%. Penting untuk mencatat bahwa kinerja algoritma dapat dipengaruhi oleh karakteristik *dataset* yang digunakan. Meskipun *dataset* dalam penelitian ini memiliki kesamaan variabel yang serupa dengan penelitian sebelumnya, seperti jenis kelamin, usia, berat badan, tinggi badan, dan status stunting. Perbedaan sumber data dapat menyebabkan variasi dalam karakteristik *dataset*.

Table 4. Perbandingan model

Peneliti	Model Klasifikasi	Akurasi	Populasi
Achmad Aria Reza dan Muhammad Syaifur Rohman [20]	<i>Random Forest Algorithm and Random Search Optimization</i>	96.33%	10.000
Syahrani Lonang dan Dwi Normawati [21]	<i>K-Nearest Neighbor dengan Feature Selection Backward Elimination</i>	92.2%	1.000
Penelitian ini	<i>Naïve Bayes Classifier</i>	94.65%	1.000

4. KESIMPULAN

Dalam evaluasi model menggunakan algoritma *Naïve Bayes Classifier* untuk klasifikasi status gizi stunting pada balita, hasil menunjukkan kinerja yang memuaskan meskipun perbandingan dengan dua penelitian sebelumnya yang menggunakan *feature* optimasi. Meskipun hasilnya bukan yang terbaik karena sumber data yang berbeda, algoritma *naïve bayes classifier* tetap menjadi pilihan yang sesuai untuk klasifikasi status gizi stunting. Akurasi model yang diperoleh sebesar 94.65% menunjukkan kemampuan model dalam mengklasifikasikan data secara tepat. Hasil evaluasi juga menunjukkan presisi yang tinggi untuk kategori "Normal", "Pendek", dan "Sangat Pendek", masing-masing sebesar 97%, 87%, dan 80%. Presisi yang tinggi menunjukkan bahwa model cenderung memberikan sedikit *false positive* dalam mengklasifikasikan data. Selain itu, *recall* yang tinggi untuk semua kategori menunjukkan bahwa model mampu mendeteksi sebagian besar data yang relevan. *F-1 Score*, yang merupakan harmonisasi antara presisi dan recall, juga menghasilkan nilai yang cukup tinggi, menunjukkan keselarasan antara kedua metrik tersebut. Dengan demikian, hasil evaluasi ini menunjukkan bahwa model *Naïve Bayes Classifier* memiliki kinerja yang baik dalam mengklasifikasikan status gizi stunting pada balita berdasarkan kategori yang telah ditentukan. Untuk penelitian selanjutnya mempertimbangkan penggunaan teknik *feature* optimasi untuk meningkatkan kinerja model klasifikasi. Selain itu, mempertimbangkan variasi data dan melakukan eksperimen dengan berbagai algoritma klasifikasi lainnya juga dapat menjadi langkah yang bermanfaat untuk memperdalam pemahaman kinerja klasifikasi status gizi stunting pada balita.

REFERENSI

- [1] C. G. Victora, L. Adair, C. Fall, P. C. Hallal, R. Martorell, L. Richter, and S. Sachdev, "Maternal and Child Undernutrition 2 Maternal and child undernutrition: consequences for adult health and human capital," *www.thelancet.com*, vol. 371, 2008, doi: 10.1016/S0140-6736(07)61692-4.
- [2] M. de Onis and F. Branca, "Childhood stunting: A global perspective," *Maternal and Child Nutrition*, vol. 12. Blackwell Publishing Ltd, pp. 12–26, May 01, 2016. doi: 10.1111/mcn.12231.
- [3] J. K. Das, R. A. Salam, M. Saeed, H. Bilal, and Z. A. Bhutta, "PROTOCOL: Effectiveness of interventions to manage acute malnutrition in children under five years of age in low- and middle-income countries: a systematic review," *Campbell Systematic Reviews*, vol. 14, no. 1, pp. 1–26, Jan. 2018, doi: 10.1002/CL2.193.
- [4] A. Nshimiyiryo, B. Hedt-Gauthier, C. Mutaganzwa, C. M. Kirk, K. Beck, A. Ndayisaba, J. Mubiligi, F. Kateera, and Z. El-Khatib, "Risk factors for stunting among children under five years: A cross-sectional population-based study in Rwanda using the 2015 Demographic and Health Survey," *BMC Public Health*, vol. 19, no. 1, Feb. 2019, doi: 10.1186/s12889-019-6504-z.
- [5] Trihono, Atmarita, Dwi Hapsari Tjandrarini, Anies Irawati, Nur Handayani Utami, Teti Tejayanti, and Iin Nurlinawati, *Pendek (stunting) di Indonesia : masalah dan solusinya*. Lembaga Penerbit Balitbangkes, 2015. Accessed: Feb. 19, 2024. [Online]. Available: <https://repository.badankebijakan.kemkes.go.id/id/eprint/3513/1/Stunting%20in%20Indonesia.pdf>

- [6] N. Akseer, H. Tasic, M. N. Onah, J. Wigle, R. Rajakumar, D. Sanchez-Hernandez, J. Akuoku, R. E. Black, B. L. Horta, N. Nwuneli, R. Shine, K. Wazny, N. Japra, M. Shekar, and J. Hoddinott, "Economic costs of childhood stunting to the private sector in low- and middle-income countries," *eClinical Medicine*, Mar. 2022, doi: 10.1016/j.eclinm.2022.101320.
- [7] Martin W. Bloem, Saskia de Pee, Thi Le Hop, Nguyen Cong Khan, Arnaud Laillou, Regina Moench-Pfanner, Damayanti Soekarjo, J. Antonio Solon, Chan Theary, and Emorn Wasantwisut, "Key strategies to further reduce stunting in Southeast Asia: Lessons from the ASEAN countries workshop," *Food and Nutrition Bulletin (FNB)*, Jun. 2013, doi: 10.1177/15648265130342S103.
- [8] Kementerian Kesehatan RI, "Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%." Accessed: Jul. 01, 2023. [Online]. Available: <https://www.kemkes.go.id/article/view/23012500002/prevalensi-stunting-di-indonesia-turun-ke-21-6-dari-24-4-.html>
- [9] B. K. P. K. Kementerian Kesehatan RI, "BUKU SAKU Hasil Survei Status Gizi Indonesia (SSGI) 2022," 2022. Accessed: Jul. 01, 2023. [Online]. Available: <https://kesmas.kemkes.go.id/assets/uploads/contents/attachments/09fb5b8ccfd088080f2521ff0b4374f.pdf>
- [10] I. Aritonang, *Memantau dan Menilai Status Gizi Anak*. Yogyakarta: Leutika Books, 2013.
- [11] A. Istiany and Rusilanti, *Gizi Terapan*. Bandung: Remaja Rosdakarya, 2013.
- [12] D. P. Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, pp. 437–444, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [13] A. Páez and G. Boisjoly, *Exploratory Data Analysis*. 2023. doi: 10.1007/978-3-031-20719-8_2.
- [14] R. LaRose and B. Coyle, "Robust data encodings for quantum classifiers," *Phys Rev A (Coll Park)*, Mar. 2020, doi: 10.1103/PhysRevA.102.032420.
- [15] A. B. Wibisono and A. Fahrurrozi, "PERBANDINGAN ALGORITMA KLASIFIKASI DALAM PENGKLASIFIKASIAN DATA PENYAKIT JANTUNG KORONER," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.
- [16] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 1, no. 3, pp. 117–122, Jan. 2020, doi: <https://doi.org/10.29207/resti.v4i1.1517>.
- [17] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp. 427–434, Oct. 2018, doi: 10.25126/jtiik.201854773.
- [18] S. Aji and F. Fatma Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *Jurnal Sains dan Manajemen*, vol. 8, no. 1, 2020, Accessed: Feb. 20, 2024. [Online]. Available: <https://ejournal.bsi.ac.id/ejournal/index.php/evolusi/article/view/7535>
- [19] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES," 2018. doi: <https://doi.org/10.33096/ilkom.v10i2.303.160-165>.
- [20] A. A. R. Reza and Muhammad Syaifur Rohman, "Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 7, no. 2, pp. 534–544, Jan. 2024, doi: 10.31289/jite.v7i2.10628.
- [21] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 49, Jan. 2022, doi: 10.30865/mib.v6i1.3312.

BIOGRAFI PENULIS

Fajar Maula Hidayat, mahasiswa di Program Jarak Jauh (PJJ) Magister Teknik Informatika di Universitas AMIKOM Yogyakarta. Untuk korespondensi lebih lanjut, dapat dihubungi melalui email: fajarmaulahidayat@students.amikom.ac.id



Kusriani, Dosen di Universitas AMIKOM Yogyakarta sejak 2003, saat ini menjabat sebagai Direktur Pascasarjana dan Ketua Program Studi S2 Teknik Informatika. Menyelesaikan studi S3 di Universitas Gadjah Mada Tahun 2010. Aktif sebagai koordinator PJJ dan MOOC di APTIKOM Pusat, sebagai Sekretaris Jenderal CORIS dan Ketua Bidang Kerjasama IndoCEISS. Senang meneliti bidang Artificial Intelligence, Computer Vision, Decision Support System, Data Mining dan Database. Terlibat aktif sebagai konsultan IT baik untuk pemerintah maupun swasta. Aktif melakukan publikasi. Id Scopus : 36057015500 dan Id Sinta 153621. Email: kusriani@amikom.ac.id



Ainul Yaqin, Dosen di Universitas AMIKOM Yogyakarta, menyelesaikan studi S2 di STMIK AMIKOM Yogyakarta Tahun 2014. Senang meneliti bidang Artificial Intelligence, Natural Language Processing, dan Machine Learning. Email: ainulyaqin@amikom.ac.id