

Klasifikasi penentuan status gizi balita dengan metode naive bayes

Alfiyyah 'Ainul Abdillah^{a,1}, Aris Thobirin^{b,2}, Dian Eka Wijayanti^{c,3,*}

^{a, b, c} Universitas Ahmad Dahlan, Indonesia;

¹ alfiyyah1900015020; ² aris.thobi@math.uad.ac.id ³ dian@math.uad.ac.id

*Correspondent Author

Received:

Revised:

Accepted:

KATAKUNCI

Klasifikasi
Naïve bayes
Balita
Status
Gizi

KEYWORDS

Classification
Naïve bayes
Balita
Status
Gizi

ABSTRAK

Klasifikasi merupakan pengelompokan untuk memprediksi suatu kelas berdasarkan data dan data-data tersebut memiliki table atau atribut. Salah satu metode dalam klasifikasi adalah naïve bayes. Metode naïve bayes banyak digunakan dalam berbagai bidang penelitian. Pada bidang kesehatan, metode naïve bayes digunakan dalam penelitian kesehatan anak. Salah satu penelitiannya membahas tentang gizi pada bayi dibawah umur lima tahun. Pada penelitian klasifikasi status gizi balita dengan metode naive bayes digunakan untuk melakukan klasifikasi data pada kelas tertentu. Metode naive bayes diterapkan pada penelitian ini untuk mengidentifikasi data balita. Data balita tersebut kemudian dianalisis untuk pembuatan model. Setelah pembuatan model kemudian menentukan model yang terbaik. Selanjutnya, model tersebut digunakan untuk memprediksi data balita di Puskesmas Ponjong I. Hasil penelitian menunjukkan bahwa pembagian data dengan perbandingan 90% data training dan 10% data testing menghasilkan akurasi sebesar 82,14%. Model klasifikasi ini mampu memprediksi status gizi balita dengan lebih baik daripada pembagian data lainnya. Hasil prediksi menunjukkan bahwa terdapat 14 anak dengan status gizi baik, 2 anak dengan gizi kurang, dan 2 anak dengan gizi lebih. Informasi ini memiliki implikasi penting bagi puskesmas, karena puskesmas dapat melakukan perawatan dan pengawasan lebih fokus terhadap kelima balita yang diklasifikasikan memiliki masalah gizi yang buruk.

Classification of determination of toddler nutritional status using the naive bayes method

Classification is a grouping to predict a class based on data and this data has tables or attributes. One method of classification is naïve Bayes. The naïve Bayes method is widely used in various research fields. In the health sector, the Naïve Bayes method is used in child health research. One of the studies discusses nutrition in babies under five years of age. In research on the classification of nutritional status of toddlers, the Naive Bayes method is used to classify data into certain classes. The Naive Bayes method was applied in this research to identify toddler data. The toddler data was then analyzed to create a model. After making the model, then determine the best model. Next, the model was used to predict data on toddlers at the Ponjong I Community Health Center. The research results show that dividing data with a ratio of 90% training data and 10% testing data produces an accuracy of 82.14%. This classification model is able to predict the nutritional status of toddlers better than other data divisions. The

prediction results show that there are 14 children with good nutritional status, 2 children with malnutrition, and 2 children with over nutrition. This information has important implications for the community health center, because the community health center can provide more focused care and supervision for the five toddlers who are classified as having problems poor nutrition.

This is an open-access article under the [CC-BY-SA](#) license.



Pendahuluan

Saat ini, Indonesia menghadapi tantangan gizi ganda atau *double burden* malnutrition, yaitu kurang gizi dan kelebihan gizi. Penyakit yang bisa disebabkan karena kekurangan nutrisi adalah kekurangan vitamin, anemia, dan gondok[4,5,7]. Kelebihan nutrisi menyebabkan obesitas yang berisiko memicu penyakit jantung, diabetes, dan pembuluh darah[2,8]. Pada Balita, kondisi gizi dapat mencerminkan tingkat kesehatannya. Status gizi adalah ukuran keberhasilan dalam memberikan nutrisi yang memadai kepada balita melalui parameter berat badan dan tinggi badan anak. Status gizi pada anak balita merupakan penanda yang mencerminkan taraf kesejahteraan suatu masyarakat [3, 9,10]. Untuk mendapatkan informasi dan mengawasi gizi pada anak balita, metode yang digunakan adalah melalui suatu evaluasi yang dikenal sebagai penilaian status gizi. Menghitung gizi pada anak balita dalam jumlah yang besar dan dengan cara manual mengindikasikan bahwa proses ini memerlukan waktu yang tidak sedikit dan sering kali mengakibatkan kesalahan saat memasukkan data, mengakibatkan ketidakefisienan dalam proses tersebut. Apalagi angka stunting di Gunungkidul yang tinggi salah satu penyebabnya adalah keterlambatan mengetahui status gizi balita. Angka kelahiran yang tinggi membuat puskesmas-puskesmas di daerah Gunungkidul terutama Puskesmas Ponjong I kewalahan menentukan status gizi balita. Dalam data mining, klasifikasi memiliki potensi untuk mempercepat dan mengoptimalkan proses pengambilan keputusan terkait status gizi balita. Hal ini bisa dicapai dengan menganalisis pola-pola yang terdapat pada data-data sebelumnya [1, 16]. Salah satu metode klasifikasi yang sering digunakan adalah metode naïve bayes. Metode naïve bayes merupakan salah satu metode klasifikasi yang dapat diterapkan dalam sistem pendukung keputusan [17]. Tujuan dari metode naïve bayes adalah melakukan pengklasifikasian data ke dalam kelas (label) yang spesifik, dengan harapan bahwa pola yang teridentifikasi dapat digunakan untuk memprediksi status gizi balita[18]. Dalam metode ini, setiap variabel dianggap memberikan kontribusi masing-masing, dengan asumsi bobot variabel memiliki tingkat penting yang serupa, dan variabel-variabel tersebut dianggap saling bebas[19, 25, 26]. Dengan menggunakan metode naïve bayes, tujuan utamanya adalah

mengembangkan sebuah sistem pendukung keputusan untuk menilai status gizi balita. Metode ini bertujuan untuk memprediksi status gizi balita berdasarkan input data yang ada, dengan mengandalkan informasi yang diperoleh dari data pelatihan yang diambil dari pengalaman masa lalu. Diharapkan hasil prediksi akan akurat, dan proses pemahaman pola-pola dilakukan dengan kecepatan yang relatif tinggi[6, 8,10]. Pada penelitian klasifikasi status gizi balita dengan metode naïve bayes digunakan untuk melakukan klasifikasi data pada kelas tertentu. Metode naïve bayes diterapkan pada penelitian ini untuk mengidentifikasi data balita untuk kemudian dianalisis dan digunakan dalam pembuatan model[9, 20, 21]. Berdasarkan model yang dibuat, ditentukan model yang terbaik yang digunakan untuk memprediksi data balita di Puskesmas Ponjong I[22, 23, 24].

Metode

Penelitian ini merupakan penelitian kuantitatif, menggunakan data berupa angka sebagai alat menganalisis untuk menggali suatu informasi. Dalam penelitian ini, literatur-literatur yang relevan terkait permasalahan, dipelajari dan dijadikan landasan dalam menyelesaikan masalah. Data yang digunakan adalah data primer dan data sekunder. Data primer diambil secara langsung dari wawancara semi terstruktur dengan bidan di Puskesmas Ponjong I dan kader posyandu. Data sekunder diperoleh dari web kaggle (<https://www.kaggle.com/search?q=balita>) dan data balita Puskesmas Ponjong I di Dusun Kerjo I. Metode yang digunakan untuk mengolah data adalah metode naïve bayes, dengan melakukan seleksi data, pembersihan data, transformasi data, split data, klasifikasi, confusion matrix, hasil dan prediksi model terbaik. Dari hasil prediksi model terbaik, dapat disimpulkan prediksi status gizi balita di Puskesmas Ponjong 1.

Hasil dan Pembahasan

1. Persiapan Data

Data yang digunakan pada penelitian ini terdapat dua data yaitu data yang diambil dari kaggle (<https://www.kaggle.com/search?q=balita>) dan Data Balita Puskesmas Ponjong I di Dusun Kerjo I. Data dari kaggle ini berjumlah 1114 dan memiliki 7 atribut yaitu : Nama, Jenis Kelamin, Usia, Berat Badan, Tinggi Badan, Lingkar Lengan, dan Kelas Gizi. Sedangkan, Data Balita Puskesmas Ponjong I di Dusun Kerjo I berjumlah 18 data dan 7 atribut. Data yang dikumpulkan, disusun dalam Tabel 1 yang memuat informasi tentang atribut-atribut balita yang akan dianalisis. Setiap baris dalam tabel mewakili data dari seorang balita, dengan kolom-kolom yang mencakup berbagai atribut seperti nama, jenis kelamin, berat badan, tinggi badan, usia, dan lingkar lengan. Data yang terstruktur dalam tabel ini menjadi dasar bagi model

klasifikasi Naive Bayes untuk memahami pola dan hubungan antara atribut-atribut ini dengan status gizi balita.

Tabel 1. Data Balita dan Atributnya

Nama	Jenis Kelamin	Usia	BB	TB	LiLa	Kelas Gizi
M. ADRIEL	JK1	U5	BB6	TB10	LL4	Gizi Baik
ALISA SALSA	JK2	U4	BB4	TB9	LL3	Gizi Baik
AULIA R	JK2	U4	BB4	TB9	LL3	Gizi Baik
M. AIDIL	JK1	U4	BB5	TB9	LL4	Gizi Baik
RATU AQILAH HSB	JK2	U4	BB4	TB9	LL4	Gizi Baik
SAFANA NAIRA	JK2	U4	BB4	TB9	LL3	Gizi Baik
ARSITA RIFANIA LBS	JK2	U5	BB6	TB10	LL3	Gizi Baik
...
FAIZ HALOMOAN LUBIS	JK1	U2	BB4	TB8	LL3	Gizi Baik

2. Pembuatan Model

a. Pembersihan Data

Pembersihan data (data cleaning) dalam data mining adalah proses mengidentifikasi, menangani, dan memperbaiki masalah atau kecacatan yang ada dalam dataset. Tujuan dari pembersihan data adalah memastikan kualitas data yang baik sebelum dilakukan analisis lebih lanjut. Pada data ke 106 terdapat kesalahan penulisan yang seharusnya TB8 tertulis 88, maka dilakukan pembersihan data. Pada proses pembersihan data atribut yang tidak digunakan dihilangkan. Tabel 2 memuat data balita yang sudah dihilangkan atribut nama.

Tabel 2. Data Balita Tanpa Atribut Nama

Jenis Kelamin	Usia	Berat Badan	Tinggi Badan	Lingkar Lengan	Kelas Gizi
JK1	U5	BB6	TB10	LL4	Gizi Baik
JK2	U4	BB4	TB9	LL3	Gizi Baik
JK2	U4	BB4	TB9	LL3	Gizi Baik
JK1	U4	BB5	TB9	LL4	Gizi Baik
JK2	U4	BB4	TB9	LL4	Gizi Baik
JK2	U4	BB4	TB9	LL3	Gizi Baik
JK2	U5	BB6	TB10	LL3	Gizi Baik
...
JK1	U2	BB4	TB8	LL3	Gizi Baik

b. Transformasi Data

Transformasi data adalah proses pengubahan skala data yang asli ke bentuk lain (skala angka) agar lebih mudah untuk diolah. Tabel 3 menggambarkan proses transformasi data pada

atribut data balita.

Tabel 3. Transformasi Atribut pada Data Balita

Atribut	Jenis atribut	ketentuan transformasi
Jenis Kelamin (x_1)	JK1	0
	JK2	1
Usia (x_2)	U1	0
	U2	1
	U3	2
	U4	3
	U5	4
Berat Badan (x_3)	BB1	0
	BB2	1
	BB3	2
	BB4	3
	BB5	4
	BB6	5
	BB7	6
	BB8	7
	BB9	8
	BB10	9
Tinggi Badan (x_4)	TB1	0
	TB2	1
	TB3	2
	TB4	3
	TB5	4
	TB6	5
	TB7	6
	TB8	7
	TB9	8
	TB10	9
Lingkar Lengan (x_5)	LL1	0
	LL2	1
	LL3	2
	LL4	3
	LL5	4
Kelas Gizi (y)	Gizi Baik	0
	Gizi Kurang	1
	Gizi Buruk	2
	Gizi Lebih	3

Dengan proses transformasi, data pada Tabel 1 dapat disajikan seperti pada Tabel 4

Tabel 4. Hasil Transformasi Data Balita

x_1	x_2	x_3	x_4	x_5	y
0	4	5	9	3	0
1	3	3	8	2	0
1	3	3	8	2	0
0	3	4	8	3	0
1	3	3	8	3	0
1	3	3	8	2	0
1	4	5	9	2	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
0	1	3	7	2	0

c. Split Data

Split data adalah proses membagi dua data menjadi data training dan data testing. Split data ini dilakukan dengan menggunakan google colab dengan pengambilan data 90% data training 10% data testing, 80% data training 20% data testing, dan 70% data training 30% data testing [30]. Pembagian data yang pertama dengan 90% data training 10% data testing menghasilkan 1002 data training dengan 787 balita bergizi baik, 162 bergizi kurang, 36 bergizi buruk, dan 17 bergizi lebih, sedangkan data testingnya menghasilkan. Pembagian data yang kedua dengan 90% data training 10% data testing menghasilkan 1002 data training dengan 787 balita bergizi baik, 162 bergizi kurang, 36 bergizi buruk, dan 17 bergizi lebih, sedangkan data testingnya menghasilkan. Pembagian data yang ketiga dengan 90% data training 10% data testing menghasilkan 1002 data training dengan 787 balita bergizi baik, 162 bergizi kurang, 36 bergizi buruk, dan 17 bergizi lebih.

d. Klasifikasi Naïve Bayes

1) Menghitung Nilai Prior

Menghitung nilai prior adalah menghitung nilai antar class. Pada penelitian ini terdapat 3 split data, maka akan terdapat 3 kali perhitungan prior.

(a) Perhitungan nilai prior dengan split data 90% 10%

Pada data penelitian gizi balita dengan split data 90% data training, terdapat 1002 data balita dengan 787 balita bergizi baik, 162 bergizi kurang, 36 bergizi buruk, dan 17 bergizi lebih.

Nilai prior gizi baik adalah $\frac{787}{1002} = 0,7854$

Nilai prior gizi kurang adalah $\frac{162}{1002} = 0,1617$

Nilai prior gizi buruk adalah $\frac{36}{1002} = 0,0359$

Nilai prior gizi lebih adalah $\frac{17}{1002} = 0,0170$

(b) Perhitungan nilai prior dengan split data 80% 20%

Pada data penelitian gizi balita dengan split data 90% data training, terdapat 891 data balita dengan 699 balita bergizi baik, 144 bergizi kurang, 32 bergizi buruk, dan 16 bergizi lebih.

Nilai prior gizi baik adalah $\frac{699}{891} = 0,7854$

Nilai prior gizi kurang adalah $\frac{144}{891} = 0,1616$

Nilai prior gizi buruk adalah $\frac{32}{891} = 0,0359$

Nilai prior gizi lebih adalah $\frac{16}{891} = 0,0180$

(c) Perhitungan nilai prior dengan split data 70% 30%

Pada data penelitian gizi balita dengan split data 70% data training, terdapat 779 data balita dengan 610 balita bergizi baik, 126 bergizi kurang, 31 bergizi buruk, dan 12 bergizi lebih.

Nilai prior gizi baik adalah $\frac{610}{779} = 0,7831$

Nilai prior gizi kurang adalah $\frac{126}{779} = 0,1617$

Nilai prior gizi buruk adalah $\frac{31}{779} = 0,0359$

Nilai prior gizi lebih adalah $\frac{12}{779} = 0,0154$

2) Menghitung Nilai Mean

Nilai mean adalah jumlah semua nilai dalam kolom dibagi dengan jumlah baris dalam kolom. Nilai akan dihitung perkolom pada 3 split data, sebagai berikut ,

(a) Perhitungan nilai mean pada split data 90% 10%

Mean kelas gizi baik akan dihitung dengan menjumlahkan nilai perbaris di- setiap kolom dibagi dengan jumlah data yaitu 787.

$$\text{Mean } x_1 = \frac{0+1+1+0+\dots+1+0+0}{787} = 0,5070$$

$$\text{Mean } x_2 = \frac{0+3+2+4+\dots+3+1+2}{787} = 2,2274$$

$$\text{Mean } x_3 = \frac{2+3+3+4+\dots+4+3+3}{787} = 3,0089$$

$$\text{Mean } x_4 = \frac{5+8+7+9+\dots+8+6+7}{787} = 7,1245$$

$$\text{Mean } x_5 = \frac{2+2+2+3+\dots+3+3+2}{787} = 2,2185$$

Menggunakan cara yang sama, mean untuk kelas gizi kurang, gizi buruk dan gizi lebih dihitung dengan dibagi jumlah data pada masing-masing kelas. Hasil perhitungan mean per kelas tersebut kemudian dirangkum dalam Tabel 5.

Tabel 5. Hasil perhitungan mean per kelas pada split data 90% 10%

	X ₁	X ₂	X ₃	X ₄	X ₅
0	0,5070	2,2274	3,0089	7,1245	2,2185
1	0,3765	2,0864	2,5802	7,0247	2,0062
2	0,1944	1,4167	1,8056	6,3611	2,0833
3	0,3529	1,1765	2,5882	5,9412	2,1765

(b) Perhitungan nilai mean pada split data 80% 20%

Mean pada split data 80% 20% untuk kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih dihitung dengan cara yang sama, menggunakan jumlah data pada masing-masing kelas sebagai pembagi. Hasil perhitungan mean per kelas tersebut kemudian dirangkum dalam Tabel 6.

Tabel 6. Hasil perhitungan mean per kelas pada split data 80% 20%

	X ₁	X ₂	X ₃	X ₄	X ₅
0	0,5036	2,2346	3,0172	7,1288	2,2160
1	0,3750	2,1042	2,5764	7,0347	2,0069
2	0,2188	1,2813	1,7188	6,2188	2,0000
3	0,3125	1,1250	2,5625	5,8750	2,1250

(c) Perhitungan nilai mean pada split data 70% 30%

Hasil perhitungan mean per kelas pada split data 70% 30% disajikan dalam Tabel 7.

Tabel 7. Hasil perhitungan mean per kelas pada split data 70% 30%

	X ₁	X ₂	X ₃	X ₄	X ₅
0	0,5115	2,2131	3,0115	7,1131	2,2049
1	0,3730	2,1190	2,6032	7,0556	2,0079
2	0,2258	1,2581	1,7097	6,1935	2,0000
3	0,3333	1,3333	2,5833	6,0000	2,1667

3) Menghitung Standar Deviasi

Nilai standar deviasi ini dihitung dengan mengurangkan nilai kolom dengan mean perkolom dan dipangkatkan dua, setelah itu dijumlahkan seluruh kolom dan diakarkan. Perhitungan nilai standar deviasi dilakukan pada 3 split data, sebagai berikut ,

(a) Perhitungan nilai standar deviasi pada split data 90% 10%

Nilai standar deviasi akan dihitung dengan menggunakan persamaan

$$\sigma = \sqrt{(x_{1,1} - \text{mean})^2 + (x_{2,1} - \text{mean})^2 + \dots + (x_{i,1} - \text{mean})^2} \quad (1)$$

Nilai standar deviasi pada kelas gizi baik dihitung sebagai berikut,

$$\sigma(x_1) = \sqrt{(0 - 0,5070)^2 + (1 - 0,5070)^2 + \dots + (0 - 0,5070)^2} = 0,5003$$

$$\sigma(x_2) = \sqrt{(0 - 2,2274)^2 + (3 - 2,2274)^2 + \dots + (2 - 2,2274)^2} = 1,3068$$

$$\sigma(x_3) = \sqrt{(2 - 3,0089)^2 + (3 - 3,0089)^2 + \dots + (3 - 3,0089)^2} = 0,7983$$

$$\sigma(x_4) = \sqrt{(5 - 7,1245)^2 + (8 - 7,1245)^2 + \dots + (7 - 7,1245)^2} = 1,0969$$

$$\sigma(x_5) = \sqrt{(2 - 2,2185)^2 + (2 - 2,2185)^2 + \dots + (2 - 2,2185)^2} = 0,4226$$

Pada kelas gizi kurang, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4860, \sigma(x_2) = 1,0299, \sigma(x_3) = 0,5869, \sigma(x_4) = 0,9120, \sigma(x_5) = 0,0786$$

Pada kelas gizi buruk, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4014, \sigma(x_2) = 1,2042, \sigma(x_3) = 0,7491, \sigma(x_4) = 1,1989, \sigma(x_5) = 0,3684$$

Pada kelas gizi lebih, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4926, \sigma(x_2) = 1,3339, \sigma(x_3) = 1,1213, \sigma(x_4) = 1,4778, \sigma(x_5) = 0,3929$$

Hasil perhitungan nilai standar deviasi dirangkum dalam Tabel 8,

Tabel 8. Hasil perhitungan nilai standar deviasi per kelas pada split data 90% 10%

y	x ₁	x ₂	x ₃	x ₄	x ₅
0	0,5003	1,3068	0,7983	1,0969	0,4226
1	0,4860	1,0299	0,5869	0,9120	0,0786
2	0,4014	1,2042	0,7491	1,1989	0,3684
3	0,4926	1,3339	1,1213	1,4778	0,3929

(b) Perhitungan nilai standar deviasi pada split data 80% 20%

Pada split data 80% 20%, nilai standar deviasi dihitung dengan menggunakan Persamaan (1). Secara sama diperoleh,

Nilai standar deviasi pada kelas gizi baik adalah,

$$\sigma(x_1) = 0,5003, \sigma(x_2) = 1,2999, \sigma(x_3) = 0,7938, \sigma(x_4) = 1,0881, \sigma(x_5) = 0,4187$$

Pada kelas gizi kurang, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4858, \sigma(x_2) = 1,0085, \sigma(x_3) = 0,5981, \sigma(x_4) = 0,8961, \sigma(x_5) = 0,0833$$

Pada kelas gizi buruk, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4200, \sigma(x_2) = 1,1977, \sigma(x_3) = 0,7289, \sigma(x_4) = 1,1841, \sigma(x_5) = 0,0000$$

Pada kelas gizi lebih, nilai standar deviasinya adalah

$$\sigma(x_1) = 0,4787, \sigma(x_2) = 1,3601, \sigma(x_3) = 1,1529, \sigma(x_4) = 1,5000, \sigma(x_5) = 0,3416$$

Hasil perhitungan nilai standar deviasi dirangkum dalam Tabel 9,

Tabel 9. Hasil perhitungan nilai standar deviasi per kelas pada split data 80% 20%

y	x ₁	x ₂	x ₃	x ₄	x ₅
0	0,5003	1,2999	0,7938	1,0881	0,4187
1	0,4858	1,0085	0,5981	0,8961	0,0833
2	0,4200	1,1977	0,7289	1,1842	0,0000
3	0,4787	1,3601	1,1529	1,5000	0,3416

(c) Perhitungan nilai standar deviasi pada split data 70% 30%

Secara sama dengan proses (a) dan (b), pada split data 80% 20%, nilai standar deviasi dirangkum dalam Tabel 10,

Tabel 10. Hasil perhitungan nilai standar deviasi per kelas pada split data 70% 30%

y	x ₁	x ₂	x ₃	x ₄	x ₅
0	0,5003	1,2999	0,7938	1,0881	0,4187
1	0,4858	1,0085	0,5981	0,8961	0,0833
2	0,4200	1,1977	0,7289	1,1842	0,0000
3	0,4787	1,3601	1,1529	1,5000	0,3416

4) Menghitung Likelihood

Setelah nilai mean dan standar deviasi diperoleh, langkah selanjutnya adalah menghitung nilai likelihood, pada semua kolom, menggunakan Persamaan (2)

$$L = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i - \mu_{i,j})^2}{2\sigma^2}} \quad (2)$$

Perhitungan nilai Likelihood ini dilakukan pada 3 data set.

(a) Perhitungan nilai Likelihood pada split data 90% 10% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 11

Tabel 11. Hasil perhitungan nilai Likelihood per kelas pada split data 90% 10%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0.4907	0.3007	0.4997	0.3613	0.8258
	0.4907	0.1217	0.2312	0.2645	0.8258
	⋮	⋮	⋮	⋮	⋮
	0.4772	0.2563	0.4997	0.2645	0.1708
Gizi Kurang	0.3605	0.3860	0.4170	0.4373	5.0621
	0.3605	0.2614	0.5263	0.2469	5.0621
	⋮	⋮	⋮	⋮	⋮
	0.3605	0.2614	0.5263	0.2469	5.0621
Gizi Buruk	0.8839	0.1396	0.1494	0.1307	1.0556
	0.8839	0.3120	0.5149	0.3180	1.0556

	0.8839	0.2946	0.5149	0.2887	1.0556
Gizi Lebih	0.6265	0.1175	0.1610	0.2088	0.1129

(b) Perhitungan nilai Likelihood pada split data 80% 20% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 12

Tabel 12. Hasil perhitungan nilai Likelihood per kelas pada split data 80% 20%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0,4874	0,3019	0,5025	0,3641	0,8340
	0,4874	0,1220	0,2335	0,2661	0,8340
	⋮	⋮	⋮	⋮	⋮
	0,4874	0,1220	0,5025	0,0836	0,1651
Gizi Kurang	0,3589	0,3935	0,4192	0,4449	4,7707
	0,3589	0,2666	0,5190	0,2492	4,7707
	⋮	⋮	⋮	⋮	⋮
	0,6096	0,2172	0,4192	0,2286	4,7707
Gizi Buruk	0,8294	0,1189	0,1167	0,1087	2,0000
	⋮	⋮	⋮	⋮	⋮
	0,8294	0,1189	0,1167	0,1087	2,0000
Gizi Lebih	0,6734	0,1134	0,1590	0,2008	0,0439
	0,2971	0,2385	0,3220	0,2008	0,0439

(c) Perhitungan nilai Likelihood pada split data 70% 30% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 13

Tabel 13. Hasil perhitungan nilai Likelihood per kelas pada split data 70% 30%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0.4950	0.3038	0.5004	0.3641	0.8619
	0.4950	0.1190	0.2320	0.2629	0.8619
	⋮	⋮	⋮	⋮	⋮
	0.4950	0.2560	0.5004	0.2629	0.8619
Gizi Kurang	0.3569	0.4055	0.4006	0.4576	4.4604
	0.3569	0.2720	0.5440	0.2544	4.4604
	⋮	⋮	⋮	⋮	⋮
	0.6117	0.4055	0.5439	0.4576	4.4604
Gizi Buruk	0.8151	0.1170	0.1176	0.1065	2.0000
	⋮	⋮	⋮	⋮	⋮
	0.8151	0.2732	0.4997	0.2659	2.0000
Gizi Lebih	0.6443	0.1416	0.1675	0.2011	0.1036
	⋮	⋮	⋮	⋮	⋮
	0.6443	0.1805	0.2880	0.2011	0.9351

5) Menghitung nilai posterior

Menghitung nilai posterior menggunakan Persamaan (3)

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \quad (3)$$

(a) Perhitungan nilai posterior pada split data 90% 10% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 14

Tabel 14. Hasil perhitungan nilai posterior per kelas pada split data 90% 10%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0.3748	0.2013	0.3925	0.2077	0.1342
	0.3854	0.0956	0.1816	0.2077	0.6486
	⋮	⋮	⋮	⋮	⋮
	0.3748	0.2013	0.3925	0.2077	0.1342
Gizi Kurang	0.0583	0.0624	0.0674	0.0707	0.8184
	0.0583	0.0423	0.0851	0.0399	0.8184
	⋮	⋮	⋮	⋮	⋮
	0.0583	0.0423	0.0851	0.0399	0.8184
Gizi Buruk	0.0318	0.0050	0.0054	0.0047	0.0379
	0.0318	0.0112	0.0185	0.0114	0.0379
	0.0318	0.0106	0.0185	0.0104	0.0379
Gizi Lebih	0.0106	0.0020	0.0027	0.0035	0.0019

(b) Perhitungan nilai posterior pada split data 80% 20% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 15

Tabel 15. Hasil perhitungan nilai posterior per kelas pada split data 80% 20%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0.3824	0.2369	0.3942	0.2856	0.6543
	0.3824	0.0957	0.1832	0.2087	0.6543
	⋮	⋮	⋮	⋮	⋮
	0.3824	0.0957	0.3942	0.0656	0.1295
Gizi Kurang	0.0580	0.0636	0.0678	0.0719	0.7710
	0.0580	0.0431	0.0839	0.04029	0.7710
	⋮	⋮	⋮	⋮	⋮
	0.0985	0.0636	0.0838	0.0403	0.7710
Gizi Buruk	0.0298	0.0043	0.0042	0.0039	0.0718
	⋮	⋮	⋮	⋮	⋮
	0.0298	0.0043	0.0042	0.0039	0.0718
Gizi Lebih	0.0121	0.0020	0.0029	0.0036	0.0008
	0.0053	0.0043	0.0058	0.0036	0.0008

(c) Perhitungan nilai posterior pada split data 70% 30% pada kelas gizi baik, gizi kurang, gizi buruk dan gizi lebih disajikan pada Tabel 16

Tabel 16. Hasil perhitungan nilai posterior per kelas pada split data 70% 30%

atribut Kelas	x ₁	x ₂	x ₃	x ₄	x ₅
Gizi Baik	0.3876	0.2379	0.3919	0.2851	0.6749
	0.3876	0.0931	0.1816	0.2058	0.6749
	⋮	⋮	⋮	⋮	⋮
	0.3876	0.2005	0.3919	0.2058	0.6749
Gizi Kurang	0.0577	0.0656	0.0648	0.0740	0.7214
	0.0577	0.0440	0.0880	0.0411	0.7214
	⋮	⋮	⋮	⋮	⋮
	0.0989	0.0656	0.0880	0.0740	0.7214
Gizi Buruk	0.0324	0.0046	0.0047	0.0042	0.0796
	⋮	⋮	⋮	⋮	⋮
	0.0324	0.0109	0.0199	0.0106	0.0796
Gizi Lebih	0.0099	0.0022	0.0026	0.0031	0.0016
	⋮	⋮	⋮	⋮	⋮
	0.0099	0.0028	0.0044	0.0031	0.0144

6) Membandingkan nilai posterior

Membandingkan nilai posterior yaitu y test (y_t) dan y prediksi (y_p), merupakan langkah penting dalam proses klasifikasi menggunakan metode Naive Bayes. Setelah perhitungan probabilitas posterior selesai dilakukan untuk setiap kategori berdasarkan atribut-atribut yang diberikan, langkah selanjutnya adalah membandingkan nilai-nilai tersebut. Hasil perbandingan ini memberikan wawasan tentang kategori mana yang memiliki probabilitas posterior tertinggi, yang akan menjadi prediksi akhir model untuk data yang diberikan. Tabel 17 memuat perbandingan pada 3 data set berbeda :

Tabel 17. Tabel perbandingan nilai posterior per kelas

90% 10%		80% 20%		70% 30%	
y_t	y_p	y_t	y_p	y_t	y_p
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮
2	0	2	0	2	0
2	0	2	0	2	0

2	0	2	0	2	0
3	0	3	0	3	0
3	0	3	0	3	0

3. Penggunaan Model Terbaik

Confusion matrix adalah alat yang berguna untuk mengevaluasi performa model klasifikasi, terutama dalam konteks klasifikasi dengan lebih dari dua kategori. Dalam penelitian ini, digunakan empat kategori gizi, yaitu Gizi Baik, Gizi Kurang, Gizi Buruk, dan Gizi Lebih. Confusion matrix juga sebagai alat yang membantu mengukur sejauh mana model dapat mengklasifikasikan data dengan benar kedalam keempat kategori ini. Gambar 1 menyajikan hasil confusion matriks pada data set 90%10%, 80 % 20%, dan 70% 30%.

y		prediksi			
		0	1	2	3
t e s t i n g	0	92	0	0	0
	1	16	0	0	0
	2	3	0	0	0
	3	1	0	0	0

(a) 90%10%

y		prediksi			
		0	1	2	3
t e s t i n g	0	180	0	0	0
	1	34	0	0	0
	2	7	0	0	0
	3	2	0	0	0

(b) 80%20%

y		prediksi			
		0	1	2	3
t e s t i n g	0	269	0	0	0
	1	52	0	0	0
	2	8	0	0	0
	3	6	0	0	0

(c) 70%30%

Pada data set 90% 10%, 92 prediksi data benar, 20 salah, pada data set 80% 20%, 180 prediksi data benar, 43 salah dan pada data set 70% 30%, 269 prediksi data benar, 66 salah. Selanjutnya adalah evaluasi model dengan menghitung f1-score, menggunakan Persamaan (4)

$$f1score = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (4)$$

a. Evaluasi model 90% 10%

$$f1score = 2 \times \frac{(82,14\% \times 82,14\%)}{(82,14\% + 82,14\%)} = 82,14\%$$

b. Evaluasi model 80% 20%

$$f1score = 2 \times \frac{(80,72\% \times 80,72\%)}{(80,72\% + 80,72\%)} = 80,72\%$$

c. Evaluasi model 70% 30%

$$f1score = 2 \times \frac{(80,30\% \times 80,30\%)}{(80,30\% + 80,30\%)} = 80,30\%$$

4. Implementasi Model

Setelah peneliti membuat model dengan 3 split data, maka dipilih model yang memiliki akurasi, presisi, recall, dan f1-score yang lebih baik. Maka peneliti menggunakan model dengan split

data 90% data training dan 10% data testing, karena memiliki nilai akurasi, presisi, recall, dan f1score yang lebih baik daripada split data yang lain yaitu 82,14%. Model dengan split data 90% data training dan 10% data testing tersebut akan digunakan untuk memprediksi status gizi balita Dusun Kerjo 1 dengan tabel hasil prediksi berikut :

Tabel 18. Hasil Perdiksi Balita Dusun Kerjo 1

x_1	x_2	x_3	x_4	x_5	yprediksi
1	0	2	5	3	0
0	1	3	7	3	0
0	4	5	9	3	0
1	2	4	8	3	0
1	3	4	8	3	0
1	0	2	5	2	0
1	2	4	7	3	0
1	4	4	9	3	0
0	2	4	8	3	0
1	0	2	6	3	0
1	4	9	9	4	3
1	1	3	6	2	1
1	0	2	5	2	0
0	0	3	6	2	1
1	4	4	8	3	0
0	0	2	5	3	0
1	0	1	4	2	3
0	4	4	9	3	0

Simpulan

Dalam penelitian yang telah dilakukan, peneliti berhasil mengaplikasikan metode klasifikasi Naive Bayes sebagai alat untuk menentukan status gizi pada balita. Melalui rangkaian analisis yang telah dilakukan, dapat diambil kesimpulan dari penelitian ini sebagai berikut:

1. Penerapan metode Naive Bayes untuk mengevaluasi status gizi balita dengan atribut yang ditentukan telah memberikan wawasan yang berharga. Dari analisis akurasi yang dihasilkan, tampak bahwa model yang dikembangkan mampu memberikan hasil prediksi dengan baik.
2. Tingkat akurasi sebesar 82,14% dengan pembagian data 90% data training dan 10 % data testing yang berhasil dicapai oleh model ini menggambarkan kemampuan metode Naive Bayes dalam mengatasi tantangan klasifikasi data gizi balita.
3. Hasil prediksi tersebut, ditemukan bahwa dari total 18 balita yang menjadi subjek penelitian, 14 anak tergolong memiliki status gizi yang baik, 2 anak gizi kurang, dan 2

anak gizi buruk.

Dengan demikian, hasil ini memberikan kontribusi yang berharga dalam mendukung upaya pemantauan dan perbaikan kesejahteraan anak-anak, serta mengilustrasikan potensi penerapan metode analisis yang lebih luas dalam bidang kesehatan anak. Informasi ini memiliki potensi besar dalam memberikan panduan bagi posyandu, orang tua, dan puskes- mas untuk mengambil tindakan lebih lanjut terkait perawatan dan pengawasan gizi balita. Dengan begitu, model ini dapat menjadi alat yang berguna dalam upaya pemantauan dan perbaikan kualitas gizi balita.

Daftar Pustaka

- [1] H. Saidah and R. K. Dewi, "Feeding Rule" Sebagai Pedoman Penatalaksanaan Kesulitan Makan Pada Balita. pertama ed., 2020.
- [2] Y. Supartini, Konsep dasar keperawatan anak. Jakarta: Jakarta : EGC, 2004.
- [3] W. Istiono, H. Suryadi, M. Haris, A. Damardana Tahitoe, M. Adrian Hasdianda, T. Fitria, T. Riris Sidabutar, I. Kesehatan Masyarakat, F. Ugm, P. Samigaluh, and K. Kulonprogo, "Analisis Faktor-Faktor yang Mempengaruhi Status Gizi Balita Analysis of Factors Affecting the Nutritional Status of Under Five Olds," Ber. Kedokt. Masy., vol. 25, no. 3, pp. 150-155, 2009.
- [4] R. Auliana, "Gizi & Pengolahan Pangan," 2001.
- [5] M. Dr.Ir.Yaktiworo Indriani, "Buku Ajar Gizi dan Pangan," 2015.
- [6] A. M. Safitri, D. R. Pangestuti, and R. Aruben, "Hubungan Ketahanan Pangan Keluarga dan Pola Konsumsi dengan Status Gizi Balita Keluarga Petani," J. Kesehat. Masy., vol. 5, no. 3, pp. 120-128, 2017.
- [7] E. Prawoto, "Faktor-faktor yang Berhubungan dengan Status Gizi Balita di Dusun Pangkur," e-Journal Cakra Med., vol. 6, no. 2, p. 16, 2019.
- [8] S. Munawaroh, "Pola Asuh Mempengaruhi Status Gizi Balita Relationship of Parenting Pattern and Toddlers' Nutrititional Status," J. Keperawatan, vol. 6, no. 1, pp. 44-50, 2015.
- [9] O. L. L. Dolwd and H. Dnl, "khazanah informatika," 2015.
- [10] R. F. Putri, D. Sulastri, and Y. Lestari, "Faktor-Faktor yang Berhubungan dengan Status Gizi Anak Balita di Wilayah Kerja Puskesmas Nanggalo Padang," J. Kesehat. Andalas, vol. 4, no. 1, pp. 254-261, 2015.
- [11] A. Reichenbach, A. Bringmann, E. E. Reader, C. J. Pournaras, E. Rungger-Brändle, C. E. Riva, S. H. Hardarson, E. Stefansson, W. N. Yard, E. A. Newman, and D. Holmes, "Hubungan Pengetahuan dan Pendidikan Ibu Terhadap Status Gizi Balita," Prog. Retin. Eye Res., vol. 561, no. 3, pp. S2-S3, 2019.
- [12] Destiadi A, Nindya T S, and Sumarmi S, "Frekuensi Kunjungan Posyandu Dan Riwayat Kenaikan Berat Badan Sebagai Faktor Risiko Kejadian Stunting Pada Anak Usia 3-5 Tahun," Media Gizi Indones., vol. 10, no. 1, pp. 71-75, 2015.
- [13] F. D. Astuti and T. F. Sulistyowati, "Hubungan Tingkat Pendidikan Ibu Dan Tingkat Pendapatan Keluarga Dengan Status Gizi Anak Prasekolah Dan Sekolah Dasar Di Kecamatan Godean," J. Kesehat. Masy. Journal Public Heal., vol. 7, no. 1, 2013.
- [14] F. Hanum, A. Khomsan, and D. G. Masyarakat, "Hubungan Asupan Gizi dan Tinggi Badan Ibu dengan Status Gizi Anak Balita," J. Gizi dan Pangan, vol. 9, no. 1, pp. 1-6, 2014.
- [15] S. P. Astuti, "Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA dan Naïve Bayes," 2020.

-
- [16] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022.
- [17] M. H. Wahyudi, "Sistem Pendukung Keputusan Penentuan Status Gizi Balita Menggunakan Metode Naive Bayes," *Semin. Nas. Teknol. Inf. dan Multi Media*, no. 1, pp. 25-30, 2018.
- [18] A. Fitri Cahyanti, R. Saptono, and S. Widya Sihwi, "Penentuan Model Terbaik pada Metode Naive Bayes Classifier dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter," *J. Teknol. Inf. ITSmart*, vol. 4, no. 1, p. 28, 2016.
- [19] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, p. 134, 2018.
- [20] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020.
- [21] E. Prasetyo, *Data Mining*. 2012.
- [22] E. Priyanti, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Bakteri Gram-Negatif," *J. Tek. Komput.*, vol. III, no. 2, pp. 68-76, 2017.
- [23] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *J. Ris. Komput.*, vol. 8, no. 6, pp. 219-225, 2021.
- [24] N. S. H. Fatimah and B. Wirjatmadi, "Tingkat Kecukupan Vitamin A, Seng dan Zat Besi serta Frekuensi Infeksi pada Balita Stunting dan Non Stunting," *Media Gizi Indones.*, vol. 13, no. 2, p. 168, 2018.
- [25] Y. N. Nanik Rahmawati, Jasmir Jasmir, "Klasifikasi Kondisi Gizi Balita Menggunakan Metode Naive Bayes (Studi Kasus Posyandu Melati IV)," *J. Ilm. Mhs. Tek. Inform.*, vol. 2, no. 3, 2020.
- [26] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *Jatisi*, vol. 1, no. 2, pp. 151161, 2017.