Jurnal_Nasional-2022-Gaussian_Vol_11_No_1_Aplikasi_NBC.pdf

by

Submission date: 12-Apr-2023 12:28PM (UTC+0700)
Submission ID: 2062296793
File name: Jurnal_Nasional-2022-Gaussian_Vol_11_No_1_Aplikasi_NBC.pdf (941.26K)
Word count: 3690
Character count: 20757

ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 11, Nomor 1, Tahun 2022, Halaman 130 - 139
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

APLIKASI *NAÏVE BAYES CLASSIFIER (NBC)* PADA KLASIFIKASI STATUS GIZI BALITA *STUNTING* DENGAN PENGUJIAN *K-FOLD CROSS VALIDATION*

Riza Rizqi Robbi Arisandi[1], Budi Warsito[2*], Arief Rachman Hakim[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***email:** budiwrst2@gmail.com

ABSTRACT

The case of *stunting* in Indonesia is a problem that has been discussed for a long time. One of many efforts to overcome this problem is through an accelerated *stunting* reduction program to improve the nutritional status of the community and also to reduce the prevalence of *stunting* or stunted toddlers. Generally, the index used to determine the nutritional status of *stunting* toddlers height compared to age. This study aims to identify the classification results, evaluate the model, and predict the nutritional status of *stunting* toddlers using the *Naïve Bayes Classifier* algorithm with *K-Fold Cross Validation* testing. The data processing system used is the GUI-R (Graphical User Interface) in order to facilitate the analysis process by implementing the Shiny Package in the Rstudio program. The results of accuracy using *Naïve Bayes Classifier* with 10-Fold Cross Validation test obtained the highest accuracy on the 6th iteration with an accuracy 94.39%, while the lowest accuracy on the 8th iteration with an accuracy 82.08%. Overall, the average accuracy in each iteration is 88.46%, so it can be concluded that *Naïve Bayes Classifier* model considered good enough to classified data on the nutritional status of *stunting* toddlers.
Keywords: *Stunting*, *Data Mining, Naïve Bayes Classifier*, *K-Fold Cross Validation*, *Shiny Package*

1. PENDAHULUAN

Stunting adalah kondisi ketika balita memiliki tinggi badan di bawah rata-rata. Hal ini diakibatkan asupan gizi yang diberikan dalam waktu yang panjang tidak sesuai dengan kebutuhan. *Stunting* berpotensi memperlambat perkembangan otak dengan dampak jangka panjang berupa keterbelakangan mental, rendahnya kemampuan belajar, dan risiko serangan penyakit kronis seperti diabetes, hipertensi, hingga obesitas (Kemenkes RI, 2011). Kasus *stunting* di Indonesia merupakan masalah yang sudah mengakar sejak dahulu. Masalah *stunting* penting untuk diselesaikan, karena berpotensi mengganggu potensi sumber daya manusia dan berhubungan dengan tingkat kesehatan, bahkan kematian anak. Pada awal tahun 2021, Pemerintah Indonesia menargetkan angka *stunting* turun menjadi 14 persen di tahun 2024 melalui program percepatan penurunan *stunting* sebagai upaya untuk meningkatkan status gizi masyarakat dan juga untuk penurunan prevalensi *stunting* atau balita pendek.

Penelitian mengenai klasifikasi status gizi balita *stunting* perlu dilakukan untuk mengetahui bagaimana pengklasifikasian status gizi balita *stunting* tersebut agar dapat diketahui secara mendetail dengan akurasi yang tepat. Salah satu algoritma yang digunakan untuk klasifikasi yaitu algoritma *Naïve Bayes Classifier (NBC)*. Pada penelitian ini akan dilakukan klasifikasi status gizi balita *stunting* dengan algoritma *Naïve Bayes Classifier (NBC)* dengan menambahkan metode *K-Fold Cross Validation*. *K-Fold Cross Validation* adalah salah satu dari jenis pengujian Cross Validation yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K (Rohani et al., 2018). Pengujian menggunakan *K-Fold Cross Validation* memiliki kelebihan dapat melihat model yang memiliki akurasi terbaik karena data secara random dibagi menjadi K-partisi sehingga dapat diketahui komposisi model terbaik. Pada penelitian ini akan dilakukan klasifikasi status gizi balita *stunting* dengan *Naïve Bayes Classifier* dengan pengujian *K-Fold Cross Validation*. Penelitian ini diharapkan dapat mengklasifikasikan status gizi balita *stunting*

dengan maksimal, mengevaluasi model dengan akurasi terbaik, dan mengaplikasikan model untuk memprediksi status gizi balita *stunting*.

2. TINJAUAN PUSTAKA

2.1. Status Gizi Balita *Stunting*

Menurut Supariasa (2016) Status gizi adalah perwujudan dari keadaan keseimbangan di dalam bentuk variabel tertentu, atau perwujudan dari nutriture dalam bentuk variabel tertentu. Salah satu metode untuk mengukur status gizi adalah dengan menggunakan pengukuran antropometri. Pengukuran antropometri yang digunakan pada penelitian ini yaitu indeks tinggi badan menurut umur. Indeks Tinggi badan menurut umur adalah tinggi badan anak yang dicapai pada umur tertentu. Indeks tinggi badan menurut umur memberikan indikasi masalah gizi yang sifatnya kronis sebagai akibat dari keadaan yang berlangsung lama. Keadaan tersebut bisa berupa kemiskinan, perilaku hidup tidak sehat, dan asupan makanan kurang dalam waktu yang lama sehingga mengakibatkan anak menjadi pendek. Untuk menentukan nilai indeks tersebut digunakan *Z-Score*. *Z-Score* untuk tinggi badan menurut umur adalah nilai simpangan TB normal menurut baku pertumbuhan WHO. Rumus yang digunakan untuk menghitung *Z-Score* TB/U dirumuskan sebagai berikut:

$$Z - Score\frac{TB}{U} = \frac{TB\ Anak - TB\ Standar}{Std\ TB\ Standar} \quad (1)$$

Menurut World Health Organization (2006) *stunting* adalah kondisi dimana keadaan tubuh yang pendek hingga melampaui defisit 2 SD (*Z-Score* = < -2 SD) di bawah standar yang ditetapkan yaitu median panjang atau tinggi badan populasi yang menjadi referensi internasional.

2.2. *Naïve Bayes Classifier*

Naïve Bayes Classifier merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve Bayes Classifier* mengasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Berikut persamaan teorema bayes.

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A) \quad (2)$$

Keterangan:

A	: Sampel data yang label kelasnya tidak diketahui.
B	: Kelas hasil klasifikasi
$A\|B)$	: Probabilitas terjadinya A jika B diketahui
$P(B\|A)$	: Probabilitas terjadinya B jika A diketahui
$P(A)$	: Probabilitas prior A
$P(B)$	: Probabilitas prior B dan bertindak sebagai

Diasumsikan bahwa kumpulan data berisi n *instance* (kasus) x_i, i=1,2,..., n, yang terdiri dari atribut p, yaitu $x_i=(x_{1i}, x_{2i}, \dots, x_{pi})$. Setiap *instance* diasumsikan milik satu (dan hanya satu) kelas y∈{ $y_1, y_2, \dots, y_c$}. Pengklasifikasian *Naïve Bayes* sederhana menggunakan probabilitas ini untuk menetapkan sebuah *instance* ke kelas. Menerapkan teorema Bayes (persamaan 2) dan menyederhanakan notasi sedikit, maka didapat persamaan:

$$P(y_j|x_i) = \frac{P(x_i|y_j)P(y_j)}{P(x_i)} \quad (3)$$

Catatan bahwa pembilang dalam persamaan (3) adalah probabilitas gabungan dari x_i dan y_j. Di sini hanya akan menggunakan x dengan menghilangkan indeks i untuk disederhanakan. Oleh karena itu, pembilangnya dapat ditulis ulang sebagai berikut:

$$\begin{aligned} \mathrm{P}(\mathrm{x}|\mathrm{y_j})\mathrm{P}(\mathrm{y_j}) &= \mathrm{P}(\mathrm{x}, \mathrm{y_j}) \\ &= \mathrm{P}(\mathrm{x_1}, \mathrm{x_2}, \dots, \mathrm{x_p}, \mathrm{y_j}) \end{aligned}$$

Karena $\mathrm{P(a, b)} = \mathrm{P(a|b)P(b)}$, maka

$$\begin{aligned} &= \mathrm{P}(\mathrm{x_1}|\mathrm{x_2}, \mathrm{x_3}, \dots, \mathrm{x_p}, \mathrm{y_j})\mathrm{P}(\mathrm{x_2}, \mathrm{x_3}, \dots, \mathrm{x_p}, \mathrm{y_j}) \\ &= \mathrm{P}(\mathrm{x_1}|\mathrm{x_2}, \mathrm{x_3}, \dots, \mathrm{x_p}, \mathrm{y_j})\mathrm{P}(\mathrm{x_2}|\mathrm{x_3}, \mathrm{x_4}, \dots, \mathrm{x_p}, \mathrm{y_j})\mathrm{P}(\mathrm{x_3}, \mathrm{x_4}, \dots, \mathrm{x_p}, \mathrm{y_j}) \\ &= \mathrm{P}(\mathrm{x_1}|\mathrm{x_2}, \mathrm{x_3}, \dots, \mathrm{x_p}, \mathrm{y_j})\mathrm{P}(\mathrm{x_2}|\mathrm{x_3}, \mathrm{x_4}, \dots, \mathrm{x_p}, \mathrm{y_j}) \dots \mathrm{P}(\mathrm{x_p}|\mathrm{y_j})\mathrm{P}(\mathrm{y_j}) \end{aligned}$$

Diasumsikan bahwa x_i tidak bergantung satu sama lain. Asumsi tersebut merupakan asumsi yang jelas-jelas dilanggar di sebagian besar aplikasi praktis dan sebab disebut dengan "naif". Hal tersebut yang membedakan Bayesian dengan *Naïve Bayes*. Asumsi ini menyiratkan bahwa $\mathrm{P}(\mathrm{x_1}|\mathrm{x_2}, \mathrm{x_3}, \dots, \mathrm{x_p}, \mathrm{y_j}) = \mathrm{P}(\mathrm{x_1}|\mathrm{y_j})$, misalnya. Jadi, probabilitas gabungan dari x dan $\mathrm{y_j}$ adalah:

$$\mathrm{P}(\mathrm{x}|\mathrm{y_j})\mathrm{P}(\mathrm{y_j}) = \mathrm{P}(\mathrm{x_1}|\mathrm{y_j}).\mathrm{P}(\mathrm{x_2}|\mathrm{y_j}) \dots \mathrm{P}(\mathrm{x_p}|\mathrm{y_j}).\mathrm{P}(\mathrm{y_j})$$

$$= \prod_{k=1}^{p} \mathrm{P}(\mathrm{x_k}|\mathrm{y_j}) \,.\, \mathrm{P}(\mathrm{y_j}) \quad 4)$$

Jika persamaan (4) dimasukkan ke persamaan (3) maka didapatkan:

$$\mathrm{P}(\mathrm{y_j}|x) = \frac{\prod_{k=1}^{p} \mathrm{P}(\mathrm{x_k}|\mathrm{y_j}) \,.\, \mathrm{P}(\mathrm{y_j})}{\mathrm{P}(\mathrm{x})} \quad 5)$$

Keterangan:

$\mathrm{P}(\mathrm{y_j}|x)$: Probabilitas data dengan vektor x pada kelas y

$P(\mathrm{y_j})$: Probabilitas awal kelas y (*prior probability*)

$\prod_{k=1}^{p} \mathrm{P}(\mathrm{x_k}|\mathrm{y_j})$: Probabilitas independen kelas y dari semua fitur dalam vektor x

$P(x)$: Probabilitas dari x, tidak bergantung pada kelasnya.

Penentuan kelas yang cocok bagi suatu sampel dilakukan dengan cara membandingkan nilai *posterior* untuk masing-masing kelas dan mengambil kelas dengan nilai *posterior* yang tinggi. Kelas terbaik dalam klasifikasi *Naïve Bayes* ditentukan dengan mencari *Maximum a Posteriori* (MAP) kelas $\mathrm{C_{map}}$ melalui persamaan 8 sebagai berikut:

$$C_{MAP} = argmax_{y \epsilon Y} P(y) \prod_{k=1}^{n} P(x_k|y) \quad 6)$$

Argmax merupakan fungsi untuk mengambil nilai terbesar sub kelas y dari kelas Y. Pada persamaan 6 tidak memuat nilai $\mathrm{P(x)}$, faktor tersebut dapat dihilangkan karena memiliki nilai yang positif dan tetap untuk semua kelas sehingga tidak mempengaruhi perbandingan nilai *posterior*. Perlu menjadi perhatian pula bahwa metode *Naïve Bayes*

Classifier ini dapat digunakan bila sebelumnya telah tersedia data yang dijadikan acuan untuk melakukan klasifikasi.

Menurut Han & Kamber (2006) jika A_k nilai kontinu, maka perhitungan nilai posterior dilakukan dengan cara yang berbeda. Atribut dengan nilai kontinu diasumsikan memiliki distribusi gaussian dengan *mean* μ dan standar deviasi σ, didefinisikan sebagai berikut.

$$P(x_k|C_i) = \frac{1}{\sqrt{2\pi}\sigma_{c_i}} e^{-\frac{(x_k-\mu_{c_i})^2}{2\sigma_{c_i}^2}} \quad (8)$$

Keterangan:

$P(x_k|C_i)$: Peluang x_k bersyarat C_i

x_k : Nilai data k yang akan diklasifikasi

k : data yang akan diklasifikasi; k = 1, 2, 3, ..., m; m : jumlah data yang akan diklasifikasi

C_i : Sub kelas C yang dicari ($C_{normal}, C_{pendek}, C_{sangat\ pendek}$)

i : kelas yang diklasifikasikan (normal, pendek, dan sangat pendek); i = 1, 2, 3

μ_{c_i} : *Mean*, menyatakan rata-rata atribut dengan kelas ci

σ_{c_i} : Standar deviasi atribut dengan kelas ci

2.3. *K-Fold Cross Validation*

Cross Validation adalah sebuah metode dari teknik *data mining* yang bertujuan untuk memperoleh hasil akurasi maksimum ketika data dibagi menjadi dua subset (data latih dan data uji). Salah satu dari jenis pengujian *Cross Validation* adalah *K-Fold Cross Validation* yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K pada *K-Fold*. Pada pendekatan metode *K-Fold Cross Validation*, *dataset* dibagi menjadi sejumlah buah partisi secara acak. Data partisi tersebut diolah sejumlah K kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke-K sebagai *data testing* dan menggunakan sisa partisi lainnya sebagai *data training* (Kurniawan, 2017).

2.4. Evaluasi Model

Kriteria evaluasi yang dipertimbangkan adalah akurasi, *standard deviation, F1 Score, Recall, Precision dan specificity* (Attal et al., 2015). Pada penelitian ini, evaluasi yang dilakukan adalah dengan menghitung akurasi dan *F1 Score*. Pengukuran kinerja klasifikasi juga dihitung dengan *Confusion Matrix* dengan *Precision* dan *Recall*.

Confusion Matrix merupakan alat pengukuran yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran proses klasifikasi. Penggunaan *Confusion Matrix* dapat dianalisa seberapa baik *classifier* dapat mengenali *record* dari kelas-kelas yang berbeda. *Confusion Matrix* terdiri dari tabel yang berisi jumlah data yang tepat diklasifikasikan dan jumlah data yang tidak tepat diklasifikasikan (Indriani, 2014).

Akurasi merupakan metode pengujian berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Melalui hasil analisis berupa jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi (Kabir & Hasan, 2017). Persamaan akurasi seperti pada persamaan berikut.

$$Accuracy = \frac{N_{benar}}{N} \times 100\% \quad (9)$$

Presisi merupakan metode pengujian dengan melakukan perbandingan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi yang terambil

oleh sistem baik yang relevan maupun tidak (Kabir & Hasan, 2017). Persamaan presisi ditunjukkan pada persamaan berikut.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Recall merupakan metode pengujian yang membandingkan jumlah informasi relevan yang didapatkan sistem dengan jumlah seluruh informasi relevan yang ada dalam koleksi informasi (baik yang terambil atau tidak terambil oleh sistem) (Kabir & Hasan, 2017). Persamaan *Recall* ditunjukkan pada persamaan berikut.

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

F-measure merupakan *harmonic mean* dari *Precision* dan *Recall* (Tripathy et al., 2016). Persamaaan *F1 Score* ditunjukkan pada persamaan berikut:

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

3. METODE PENELITIAN

3.1. Jenis dan Sumber Data

Data pada penelitian merupakan data sekunder yang diperoleh dari dari Puskesmas Padangsari, Kec. Banyumanik, Kota Semarang pada Tahun 2021. Data yang digunakan yaitu data kuantitatif berupa data status gizi balita *stunting* dengan variabel yang digunakan yaitu jenis kelamin, tinggi badan, berat badan, lingkar lengan atas dan usia. Pada proses pengumpulan data ini diperoleh sejumlah 1066 data.

3.2. Isi dari metode penelitian

Metode analisis yang digunakan untuk mencapai tujuan penelitian dalam penulisan Tugas Akhir ini diuraikan sebagai berikut:

1. Mengumpulkan data kemudian diinput ke dalam sebuah file bertipe csv (*microsoft excel)*
2. Melakukan proses *prepocesing* data untuk menyiapkan data agar bisa diolah. Proses yang dilakukan pada tahap ini berupa *Data Cleaning* dan *Data Transformation*
3. Mengidentifikasi data dengan melakukan analisis deskriptif pada data
4. Hasil *prepocesing* data di*input* ke dalam database program *RStudio*
5. Klasifikasi dengan *Naïve Bayes Classifier*
    a. Menentukan nilai K untuk *K-Fold Cross Validation*, ditentukan K=10
    b. *Dataset* dibagi menjadi sejumlah K buah partisi secara acak, dalam hal ini dibagi menjadi 10 partisi
    c. Membagi partisi data menjadi *data testing* dan *data training* secara bergantian sejumlah K yaitu 10 kali.
    d. Melakukan klasifikasi dengan *Naïve Bayes Classifier*
    e. Iterasi dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai *data testing*. Hal ini berarti dilakukan 10 kali eksperimen dan masing-masing partisi digunakan sebagai data testing secara bergantian
    f. Data yang tidak digunakan sebagai data testing digunakan sebagai data training

6. Melakukan evaluasi model dengan melihat pengujian *K-Fold Cross Validation* dan *Confusion Matrix*
    a. Membentuk *Confusion Matrix Multiclass* yaitu dengan tiga kelas
    b. Menghitung nilai *Accuracy*, nilai *Precision*, nilai *Recall*, dan *F1 Score*
    c. Menentukan model terbaik dari setiap iterasi yang dilakukan
7. Melakukan prediksi untuk data baru dengan menggunakan model sebelumnya yang telah ditentukan sebagai model terbaik.
8. Interpretasi dan Kesimpulan

4. HASIL DAN PEMBAHASAN

4.1. *Prepocesing Data*

Data yang digunakan untuk penelitian merupakan data yang belum siap untuk diolah. Data masih memiliki variabel yang tidak digunakan, *missing value*, *noise*, dan data masih memiliki jenis data yang belum sesuai dengan penelitian yang dilakukan sehingga perlu dilakukan *prepocesing data*. Variabel yang akan digunakan yaitu variabel jenis kelamin, tinggi badan, berat badan, nilai *Z-Score*, lingkar lengan atas, dan usia balita yang akan dihitung dengan tanggal pengukuran berdasarkan tanggal lahir balita. Data kemudian dilakukan *prepocesing* berdasarkan penyesuaian yang dibutuhkan melalui *data cleaning* dan *data transformation*. Berikut merupakan *layout data* setelah *prepocesing* dapat dilihat pada tabel 1.

Tabel 1. *Layout Data Prepocesing*

Usia	JK	Berat	Tinggi	LiLA	Status
48	P	14	97	15	NORMAL
55	L	18,7	112,2	15	NORMAL
46	P	25	105,5	23	NORMAL
54	L	16,3	105	17	NORMAL
38	P	11,4	89,2	14,5	NORMAL
38	P	13	94	17	NORMAL
41	P	14	102,5	17	NORMAL
44	P	11	87,5	20	SANGAT PENDEK
...	...	...	...	...	...
49	L	10,9	84,7	17	SANGAT PENDEK

Sumber : Puskesmas Padangsari, Kec. Banyumanik, Kota Semarang

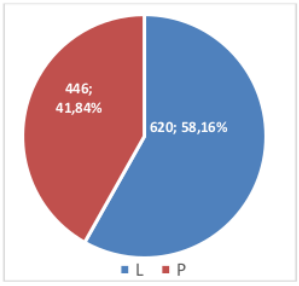
4.2. Analisis Deskriptif

Tabel 2. Analisis Deskriptif Data Kontinyu

	USIA	BB	TB	LiLA
Min.	1	2,7	45	11,4
1st Qu.	21	10	79,5	14,3
Median	32	12,2	88	15,4
Mean	32,22	12,57	87,76	15,7
3rd Qu.	45	14,9	97,5	17
Max.	59	28,85	115,5	25

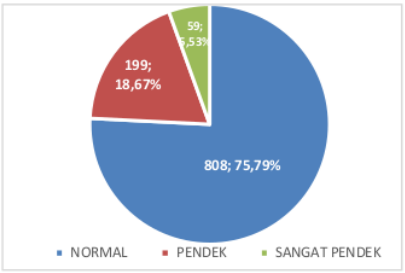
Sumber : Data diolah dengan RStudio GUI RShiny

Berdasarkan tabel 2 dapat dilihat pada kolom usia bahwa usia balita menyebar rata dari 1 bulan sampai 59 bulan, ukuran penyebaran data kuantil juga tersebar dari kuantil pertama 21 bulan dan kuantil ketiga pada 45 bulan. Berat badan balita dan tinggi badan balita sangat bervariasi dari terendah sampai tertinggi memiliki rentang yang cukup jauh. Pada variabel lingkar lengan atas rentang masih terbilang dekat dengan rata-rata.



Gambar 1. Diagram Lingkaran Jenis Kelamin

Berdasarkan gambar 1 dapat dilihat bahwa kecenderungan balita laki-laki lebih banyak daripada perempuan. Balita dengan jenis kelamin laki-laki didapat 605 dari 1037 populasi atau 58,34%, sedangkan balita dengan jenis kelamin perempuan berjumlah 432 dari 1037 populasi atau 41,66%.



Gambar 2. Diagram Lingkaran Status Gizi Balita *Stunting*

Berdasarkan gambar 2 dapat dilihat bahwa secara keseluruhan mayoritas masih normal tetapi masih ada balita *stunting* yang jumlahnya tidak juga sedikit. Balita dengan status gizi normal didapat 808 dari 1037 populasi atau 75,6%, balita dengan status gizi pendek berjumlah 195 dari 1037 populasi atau 18,8%, sedangkan balita dengan status gizi sangat pendek berjumlah 58 dari 1037 populasi atau 5,5%.

4.3. Klasifikasi

Pengujian ini bertujuan untuk mengeidentifikasi hasil klasifikasi, mengevaluasi kinerja dari algoritma *Naïve Bayes Classifier* dalam mengklasifikasi data ke dalam kelas yang telah ditentukan, dan menentukan model terbaik untuk melakukan prediksi. Data diacak menggunakan *random seed* kemudian data dibagi menjadi data training dan data testing menggunakan metode *K-Fold Cross Validation*. Berikut merupakan salah satu hasil klasifikasi pada *fold* ke-8 dapat dilihat pada tabel 3.

Tabel 3. Hasil Klasifikasi *Fold* 8

ID	Usia	JK	TB	BB	LiLA	Status Aktual	Status Prediksi
234	13	L	11	76	15	NORMAL	NORMAL
501	39	L	19,5	96,2	16,7	NORMAL	NORMAL
664	14	L	8,9	77,5	14	NORMAL	NORMAL
781	9	L	7,4	69,2	15,5	NORMAL	NORMAL
969	29	L	11,8	87,5	14,2	NORMAL	NORMAL
669	34	L	16	96	17	NORMAL	NORMAL

772	31	L	12,3	85	16,3	PENDEK	PENDEK
99	21	P	13	79	17	NORMAL	NORMAL
565	48	L	15,2	107,2	17	NORMAL	NORMAL
170	41	L	13,4	87,9	12,4	PENDEK	PENDEK
...	...	...	...	...	...	...	...
463	46	P	16.3	103	18	NORMAL	NORMAL

Sumber : Data diolah dengan RStudio GUI RShiny

Berdasarkan hasil klasifikasi pada tabel 3 dapat dilihat bahwa sebanyak 101 data diklasifikasikan dengan tepat dan sebanyak 6 data salah diklasifikasikan.

4.4. Kinerja Klasifikasi

Tabel 4. Hasil Akurasi Klasifikasi

Fold ke-	Akurasi
1	0,89423
2	0,88462
3	0,81731
4	0,87500
5	0,84615
6	0,91346
7	0,87500
8	0,95146
9	0,89320
10	0,90291

Sumber : Data diolah dengan Rstudio GUI RShiny

Berdasarkan hasil akurasi menggunakan *Naïve Bayes Classifier* dapat dilihat bahwa akurasi tertinggi berada pada iterasi ke-8 dengan akurasi sebesar 95,14%, sedangkan akurasi terendah berada pada iterasi ke-3 dengan akurasi sebesar 81,73%. Secara keseluruhan rata-rata akurasi pada setiap iterasi didapat sebesar 88,53%.

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-8 dapat dilihat pada tabel 5.

Tabel 5. *Confusion Matrix Fold* 8

	Aktual		
Prediksi	**Normal**	**Pendek**	**Sangat Pendek**
Normal	81	0	0
Pendek	2	14	0
Sangat Pendek	0	3	3

Sumber : Data diolah dengan RStudio GUI RShiny

Pada tabel 5 dilihat bahwa model Naïve Bayes Classifier memprediksi status gizi balita normal sebanyak 81, status gizi balita pendek sebanyak 14, dan status gizi balita sangat pendek sebanyak 3; sedangkan pada data aktual status gizi balita normal sebanyak 81, status gizi balita pendek sebanyak 16, dan status gizi balita sangat pendek sebanyak 6. Berikut merupakan salah satu *statistics by class* pada *fold* ke-6 dapat dilihat pada tabel 10.

Tabel 6. *Statistics by Class Fold 6*

	Class: Normal	Class: Pendek	Class: Sangat Pendek
Precision	1	0.875	0.5
Recall	0.9759	0.8235	1
F1	0.9878	0.8485	0.66667

Pada tabel 6 dapat dilihat bahwa pada kelas normal didapat nilai presisi sebesar 100%, nilai *recall* sebesar 97,59%, dan nilai *F1-Score* sebesar 98,78%; selanjutnya pada kelas

pendek didapat nilai presisi sebesar 87,5%, nilai *recall* sebesar 82,35%, dan nilai *F1-Score* sebesar 84,85%; sedangkan pada kelas sangat pendek didapat nilai presisi sebesar 50%, nilai *recall* sebesar 100%, dan nilai *F1-Score* sebesar 66,67%.

4.5. Prediksi

Pada tahap sebelumnya telah dilakukan klasifikasi dengan menggunakan model *Naïve Bayes Classifier*. Pada analisis tersebut didapat model akurasi terbaik pada iterasi ke-8 sehingga akan dilakukan prediksi dengan *data training* menggunakan model iterasi ke-8 sebagai model prediksi. Data yang akan diprediksi merupakan data yang belum pernah digunakan sebelumnya. Berikut merupakan hasil prediksi dapat dilihat pada tabel 7.

Tabel 7. *Layout Data* Prediksi

ID	USIA	JK	TB	BB	LiLA	STATUS PREDIKSI
1	26	P	13.3	94	14.8	NORMAL
2	52	L	14.1	99.5	16.5	NORMAL
3	44	L	15	102	17	NORMAL
4	27	L	10.7	81.7	12	PENDEK
5	46	L	12.7	96.5	13	NORMAL

Sumber : Data diolah dengan RStudio GUI RShiny

Model *Naïve Bayes Classifier* dapat dilakukan untuk melakukan prediksi dengan hasil pada tabel 7 dan juga dapat dilakukan untuk melakukan prediksi lain dengan menggunakan variabel yang sama.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan klasifikasi status gizi balita *stunting* didapatkan beberapa kesimpulan sebagai berikut:

1. Hasil klasifikasi status gizi balita *stunting* dapat teridentifikasi dan diklasifikasikan dengan baik pada masing-masing iterasinya.
2. Hasil kinerja klasifikasi menggunakan *10-Fold Cross Validation* pada algoritma *Naïve Bayes Classifier* memperoleh akurasi tertinggi pada iterasi ke-8 dengan akurasi sebesar 95,14%, sedangkan akurasi terendah berada pada iterasi ke-3 dengan akurasi sebesar 81,73%. Secara keseluruhan rata-rata akurasi pada setiap iterasi didapat sebesar 88,53%.
3. Hasil klasifikasi dapat digunakan menjadi model untuk melakukan prediksi dengan karakteristik atau variabel data yang sama, sehingga jika terdapat data dengan variabel yang sama GUI-R dapat digunakan untuk mengklasifikasi data dengan lebih cepat dan efisien.

DAFTAR PUSTAKA

Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical human activity recognition using wearable sensors. *Sensors*, *15*(12), 31314–31338.

Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode NaÃ¯ ve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, *1*(1).

Kabir, A., & Hasan, A. (2017). *Analisis sentimen data kritik dan saran pelatihan aplikasi teknologi informasi (pati) menggunakan algoritma support vector machine*. University of Muhammadiyah Malang.

Kemenkes RI. (2011). KEPMENKES RI Tentang Standar Antropometri Penilaian Status Gizi Anak. In *Jornal de Pediatria* (Vol. 95, Issue 4, p. 41).

Kurniawan, T. (2017). *Implementasi Text Mining pada Analisis Sentimen Pengguna*

Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier dan Support Vector Machine. Institut Teknologi Sepuluh Nopember.

Organization, W. H. (2006). *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. World Health Organization.

Putri, R. A., Sendari, S., & Widiyaningtyas, T. (2018). Classification of toddler nutrition status with anthropometry calculation using Naïve Bayes Algorithm. *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, 66–70.

Rohani, A., Taki, M., & Abdollahpour, M. (2018). A novel soft computing model (Gaussian process regression with *K-Fold Cross Validation*) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, *115*, 411–422.

Supariasa, I. D. M., Bakri, B., & Fajar, I. (2016). *Penilaian Status Gizi*. *jakarta: Buku Kedokteran EGC*.

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, *57*, 117–126.

Jurnal_Nasional-2022-Gaussian_Vol_11_No_1_Aplikasi_NBC.pdf

ORIGINALITY REPORT

24%	24%	20%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	edi-ismanto.blogspot.com Internet Source	2%
2	ejournal.bsi.ac.id Internet Source	2%
3	repositori.unsil.ac.id Internet Source	1%
4	www.voaindonesia.com Internet Source	1%
5	media.neliti.com Internet Source	1%
6	Submitted to University of Ghana Student Paper	1%
7	Submitted to University of Mauritius Student Paper	1%
8	journal.uniga.ac.id Internet Source	1%
9	luthfipedia.xyz Internet Source	1%

10	repository.ub.ac.id Internet Source	1 %
11	repository.uin-suska.ac.id Internet Source	1 %
12	www.kominfo.go.id Internet Source	1 %
13	Submitted to Universitas Lancang Kuning Student Paper	1 %
14	Submitted to Universitas Negeri Surabaya The State University of Surabaya Student Paper	1 %
15	repository.unhas.ac.id Internet Source	1 %
16	es.scribd.com Internet Source	1 %
17	repository.widyatama.ac.id Internet Source	1 %
18	kc.umn.ac.id Internet Source	1 %
19	Submitted to Management & Science University Student Paper	1 %
20	repository.usd.ac.id Internet Source	1 %

21	adoc.pub Internet Source	1 %
22	e-journal.unair.ac.id Internet Source	1 %
23	www.pengalaman-edukasi.com Internet Source	1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		