Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

Implementasi Algoritma Gaussian Naïve Bayes Dalam Klasifikasi Status Gizi Pada Balita

Hery Kurniawan, Abdul Rahim*, Taghfirul Azhima Yoga Siswa

Fakultas Sains Dan Teknologi, Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia
Email: 2011102441015@umkt.ac.id, ar622@umkt.ac.id, tay@umkt.ac.id
Email Penulis Korespondensi: ar622@umkt.ac.id
Submitted: **05/07/2024**; Accepted: **06/09/2024**; Published**: 07/09/2024**

Abstrak−Status gizi merupakan suatu kondisi terkait gizi yang bisa diukur dan merupakan hasil dari adanya keseimbangan kebutuhan gizi pada tubuh dengan asupan gizi dari makanan. Di Indonesia masalah gizi masih banyak ditemukan seperti gizi buruk, dan masalah gizi lainnya. Dalam hal ini, penggunaan teknik dan alat pembelajaran machine learning (ML) dan data mining (DM) bisa sangat membantu dalam menghadapi tantangan di bidang manufaktur. Oleh karena itu, Pada penelitian ini akan menggunakan algoritma Naïve Bayes Classifier dengan model Gaussian. Data yang digunakan adalah data status gizi balita pada bulan Januari-Juli 2023 di Kota Samarinda. Atribut pada penelitian ini diantaranya Jenis Kelamin, Berat Badan Lahir, Tinggi Badan Lahir, Usia Saat Ukur, Berat Badan, Tinggi Badan, ZS BB/U, BB/U, ZS TB/U, dan TB/U. Penentuan status gizi balita pada penelitian ini berdasarkan indeks BB/TB yang terdiri dari 6 kelas, yaitu gizi buruk, gizi kurang, gizi baik, risiko gizi lebih, gizi lebih, dan obesitas. Dari penelitian yang dilakukan didapatkan bahwa algoritma Naïve Bayes Classifier dengan model gaussian bisa mengklasifikasikan status gizi balita secara tepat. Dari pengolahan data yang dilakukan, diketahui bahwa nilai akurasi model Gaussian yaitu sebesar 8l,85%.

Kata Kunci: Status Gizi, Algoritma Naive Bayes, Gaussian, Akurasi

Abstract− Nutritional status is a condition related to nutrition that can be measured and results from the balance between the body's nutritional needs and nutrient intake from food. In Indonesia, nutritional problems such as malnutrition and other nutritional issues are still prevalent. In this context, the use of machine learning (ML) and data mining (DM) techniques and tools can be very helpful in tackling challenges in the manufacturing sector. Therefore, this study will use the Naïve Bayes Classifier algorithm with a Gaussian model. The data used is the nutritional status data of toddlers from January to July 2023 in Samarinda City. The attributes in this study include Gender, Birth Weight, Birth Height, Age at Measurement, Body Weight, Body Height, ZS BW/A, BW/A, ZS BH/A, and BH/A. The determination of toddlers' nutritional status in this study is based on the BW/BH index, which consists of 6 classes: severe malnutrition, undernutrition, good nutrition, risk of overnutrition, overnutrition, and obesity. From the study conducted, it was found that the Naïve Bayes Classifier algorithm with the Gaussian model can accurately classify toddlers' nutritional status. From the data processing performed, it was found that the accuracy value of the Gaussian model is 81.85%.

Keywords: Nurtritional Status; Naïve Bayes Algorithm; Gaussian; Accuracy;

1. PENDAHULUAN

Stunting pada anak-anak merupakan masalah kesehatan masyarakat yang signifikan dan menjadi salah satu penyebab utama morbiditas dan gangguan perkembangan global. Di Indonesia, masalah stunting pada anak-anak dipicu oleh berbagai faktor yang kompleks dan saling terkait. Salah satu faktor utama adalah kualitas kesehatan sumber daya manusia (SDM) yang rendah, baik dalam aspek peningkatan maupun penurunan. Rendahnya kualitas SDM ini sebagian besar dipengaruhi oleh konsumsi makanan yang tidak seimbang, yang mengakibatkan anak-anak mengalami gangguan fisik dan mental akibat kurangnya asupan nutrisi yang diperlukan tubuh, yang pada gilirannya menyebabkan stunting [1].

Stunting memiliki dampak yang luas, tidak hanya pada kesehatan fisik anak, tetapi juga pada perkembangan kognitif dan emosional mereka. Kondisi ini sangat merugikan karena menghambat potensi anak-anak untuk berkembang secara optimal. Sayangnya, stunting merupakan salah satu penyebab utama gangguan perkembangan di negara berkembang, termasuk Indonesia. Oleh karena itu, penanganan stunting menjadi prioritas nasional yang memerlukan perhatian serius dan berkelanjutan [2].

Prevalensi stunting di Indonesia masih cukup tinggi. Data dari Kementerian Kesehatan Indonesia menunjukkan bahwa pada tahun 2022, prevalensi stunting pada balita mencapai 24,4% [3]. Berdasarkan Survey Status Gizi Indonesia (SSGI) tahun 2021, angka prevalensi stunting di Indonesia pada tahun 2021 sebesar 24,4%, yang menunjukkan penurunan dari 30,8% di tahun 2018 [4]. Meskipun demikian, angka ini masih menunjukkan bahwa stunting adalah masalah yang memerlukan intervensi lebih lanjut dan berkelanjutan. Pemerintah Indonesia melalui Rencana Pembangunan Jangka Menengah Nasional (RPJMN) menargetkan untuk menurunkan prevalensi stunting menjadi 14% pada tahun 2024 [5].

Penilaian status gizi anak dapat dilakukan melalui berbagai metode, salah satunya adalah dengan pengukuran tubuh atau antropometri, yang meliputi berat badan, tinggi badan, lingkar lengan atas, lingkar kepala, lingkar dada, dan lapisan lemak bawah kulit [6]. Stunting pada balita umumnya disebabkan oleh kurangnya asupan energi dan protein dalam jangka waktu yang lama . Dalam upaya untuk menangani masalah ini, Kementerian Kesehatan Republik Indonesia telah mengumpulkan data kesehatan dari 270 juta penduduk, yang mencakup rekam medis individu yang disimpan dalam basis data kesehatan perorangan dan dikelola oleh berbagai aplikasi [7]. Data ini sangat berharga

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

dalam mendukung pengambilan keputusan strategis di bidang kesehatan, terutama melalui penggunaan metode pemrosesan data prediktif seperti algoritma C4.5, KNN, dan Naïve Bayes [8].

Penelitian ini menggunakan algoritma Naïve Bayes untuk memprediksi status gizi balita, karena metode ini mampu memperhitungkan probabilitas hasil dan memberikan informasi yang akurat meskipun dengan jumlah data yang relatif sedikit [9]. Naïve Bayes adalah salah satu metode klasifikasi yang populer dalam data mining, yang bekerja dengan menghitung probabilitas dari setiap kelas berdasarkan atribut yang ada. Untuk klasifikasi status gizi balita, atribut yang digunakan meliputi berat badan, tinggi badan, umur, dan indeks antropometri lainnya [10]. Penelitian oleh Cahyanti menunjukkan bahwa Naïve Bayes efektif dalam memproses data dengan variabel numerik dan kategori yang besar, sehingga cocok untuk aplikasi pada data status gizi [11].

Berbagai penelitian sebelumnya telah menunjukkan bahwa algoritma Naïve Bayes dapat memberikan hasil yang akurat dalam klasifikasi status gizi balita. Misalnya, penelitian di Posyandu Anggrek di Limo, Depok, menunjukkan bahwa algoritma ini dapat mencapai akurasi hingga 75% dalam mengklasifikasikan status gizi balita berdasarkan data berat badan dan tinggi badan [12]. Penelitian lain menggunakan metode K-Fold Cross Validation menemukan bahwa algoritma Naïve Bayes dapat mencapai akurasi rata-rata sebesar 75.47% dalam klasifikasi status gizi balita di Puskesmas Rambah Samo I [13]. Selain itu, penelitian di Desa Tunjungtirto menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk mengklasifikasikan status gizi balita berdasarkan tiga indeks antropometri, yaitu berat badan menurut umur (WFA), tinggi badan menurut umur (HFA), dan berat badan menurut tinggi badan (WFH). Hasil penelitian ini menunjukkan bahwa akurasi klasifikasi untuk indeks WFA mencapai 88%, sedangkan untuk indeks HFA dan WFH masing-masing mencapai 64% dan 68% [14].

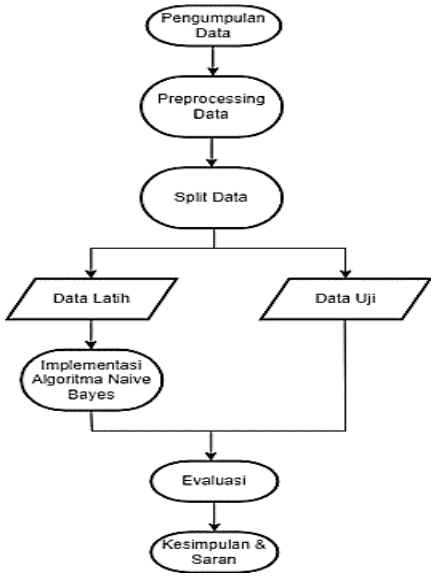
Di Kecamatan Baros, Kota Cimahi, algoritma Naïve Bayes membantu kader posyandu dalam menentukan status kesehatan dan gizi balita dengan lebih akurat, menggunakan indeks berat badan menurut umur (WFA) [15]. Penelitian lain yang fokus pada kasus stunting di Indonesia menunjukkan bahwa algoritma Naïve Bayes dengan pengujian K-Fold Cross Validation dapat memberikan hasil yang baik dalam klasifikasi status gizi balita stunting, dengan akurasi tertinggi mencapai 94.39% pada iterasi ke-6 dan akurasi rata-rata sebesar 88.46% [16].

Untuk memanfaatkan data prediktif dalam bidang kesehatan, artikel ini membahas prediksi tingkat penyebaran balita yang mengalami stunting di Kota Samarinda menggunakan metode Naïve Bayes selama periode 2022–2023. Metode Naïve Bayes dipilih karena kemampuannya dalam melakukan klasifikasi probabilitas dan evaluasi nilai kontinu terkait dengan fitur numerik [17]. Prediksi dilakukan dengan faktor antropometri untuk menganalisis nilai akurasi dan validasi yang diperoleh, sehingga dapat membantu orang tua dan Dinas Kesehatan Kota Samarinda dalam mengontrol dan menurunkan angka stunting [18].

Penelitian ini bertujuan untuk melakukan perbandingan dari model lain, yang dimana termasuk model Gaussian Naïve Bayes adalah termasuk yang tertinggi, untuk memprediksi tingkat penyebaran balita stunting di Kota Samarinda berdasarkan data dari tahun 2022–2023[19]. Penelitian ini diharapkan dapat membantu Dinas Kesehatan Kota Samarinda dalam mengetahui penyebaran stunting sehingga dapat melakukan upaya pengendalian dan sosialisasi kepada orang tua balita untuk menurunkan angka stunting[20].

2. METODOLOGI PENELITIAN

Penelitian ini terdiri dari beberapa tahapan yang dirancang untuk mencapai tujuan penelitian. Pelaksanaan penelitian dimulai dari tahap identifikasi masalah, pengumpulan data, analisis data, hingga tahap evaluasi [21]. Alur penelitian dapat dilihat pada gambar 1 dibawah ini:



Gambar 1 Alur Penelitian

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

2.1 Pengumpulan Data

Data pada penelitian merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kota Samarinda dengan rentang waktu dari tanggal 3 agustus 2022 hingga 31 juli 2023. Data yang digunakan yaitu data kuantitatif berupa data status gizi balita stunting. Pada proses pengumpulan data ini diperoleh sejumlah 15.593 dengan 17 atribut dan 1 kelas target, dan hanya menggunakan data paling mutakhir yaitu data pada tahun 2023 dari januari hingga juli dengan total data 9.494.

2.2 Preprocessing Data

Data yang didapat dari Dinas Kesehatan Kota Samarinda harus diolah lebih lanjut sebelum memasuki tahap permodelan untuk menghindari bagian data yang tidak diperlukan. Langkah-langkah pemprosesan data tersebut meliputi cleaning data dan transformation data.

a. Data Cleaning
   data cleaning menggunakan metode '.isna)' untuk mengidentifikasi missing values dalam dataset. Setiap elemen dalam dataset yang bernilai 'NaN' atau 'none' akan dianggap sebagai missing values. Kemudian metode '.sum()' digunakan untuk menghitung jumlah missing dalam setiap kolom. Selanjutnya pada metode penghapusan menggunakan '.dropna()' dengan parameter 'inplace=True', yang menghapus semua baris yang mengandung missing values dan pada metode '.datadrop_duplicates()' menghapus semua baris duplikat pada dataframe.
b. Data Transformation
   Data transformation digunakan untuk mengubah data kedalam tipe yang sesuai dalam data mining. Pada penelitian ini data yang digunakan adalah data dalam tipe numerik, sehingga data yang masih dalam tipe kategorial harus diubah terlebih dahulu. Seperti pada penelitian ini, atribut jenis kelamin akan diubah kedalam bentuk biner yaitu 0 dan 1, dimana 0 menunjukan jenis kelamin perempuan dan 1 menunjukan jenis kelamin laki-laki. Begitu juga atribut lainnya 'BB//U', 'TB/U', 'BB/TB', Naik Berat Badan'.
c. Pembagian Data
   Pembagian data menggunakan fungsi train_test_split dari library scikit-learn dalam Python, yang sangat berguna dalam membagi dataset menjadi dua subset: data latih dan data uji. Dalam contoh tersebut, variabel x menyimpan dataset fitur, sedangkan variabel y berisi label yang sesuai dengan setiap fitur. Dengan menentukan test_size = 0.2, kode tersebut membagi dataset sehingga 20% dari data akan digunakan untuk pengujian (x_test dan y_test), sementara 80% sisanya akan digunakan untuk pelatihan (x_train dan y_train). Pengaturan random_state = 42 memastikan bahwa pembagian dataset bersifat acak tetapi dapat direproduksi dengan hasil yang konsisten.
d. Permodelan
   Dalam melakukan perbandingan dan evaluasi performa dengan Gaussian Naive Bayes dalam melakukan klasifikasi pada dataset yang berbeda karakteristiknya. Data dibagi menjadi data pelatihan dan data pengujian, kemudian menguji model Gaussian Naive Bayes untuk jenis dataset tertentu, adapun rumus dari algoritma dibawah ini

$$P(C\ |X) = \frac{\mathrm{P(X}\mid C)\ .\ P(C)}{\mathrm{P(X)}} \quad (1)$$

   Rumus diatas menggunakan beberapa variabel kunci: X adalah sampel data yang memiliki kelas (label) yang tidak diketahui; C adalah hipotesis bahwa X adalah data dari kelas C; P(C) adalah probabilitas awal dari hipotesis C; P(X) adalah probabilitas keseluruhan dari data sampel yang diamati; dan P(X|C) adalah probabilitas X terjadi berdasarkan kondisi hipotesis C. Teorema Bayes memungkinkan kita untuk menghitung probabilitas P(C|X), yaitu probabilitas bahwa sampel X termasuk dalam kelas C dengan mempertimbangkan informasi yang diketahui. Setelah melakukan Gaussian Naive Bayes adalah varian dari algoritma Naive Bayes yang digunakan ketika fitur-fitur mengikuti distribusi normal (Gaussian). Rumus untuk menghitung probabilitas dari fitur x_i yang mengikuti distribusi normal adalah:

$$P(x_i|yk) = \frac{1}{\sqrt{2\pi\sigma_{Yk}^2}} exp(-\frac{\left(X_{i-\mu_{Yk}}\right)^2}{2\sigma_{yk}^2} \quad (2)$$

   Dimana rumus diatas adalah fungsi distribusi normal atau Gaussian yang menggambarkan probabilitas variabel x_i diberikan kelas y_k. Disini μ_yk adalah mean dari kelas y_k, σ_yk adalah varians dari kelas y_k, dan exp adalah fungsi eksponensial. Rumus ini digunakan untuk menghitung probabilitas suatu nilai dalam distribusi normal berdasarkan mean dan varians tertentu, membantu dalam klasifikasi probabilistik.
e. Evaluasi
   Evaluasi klasifikasi Naive Bayes menggunakan metode akurasi mengukur seberapa dekat prediksi model dengan nilai sebenarnya. Akurasi dihitung dengan membagi jumlah prediksi yang benar (True Positive dan True Negative) dengan jumlah total prediksi (semua True Positive, True Negative, False Positive, dan False Negative), kemudian dikalikan dengan 100% untuk mendapatkan nilai persentase. Metrik ini memberikan Gambaran keseluruhan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

tentang seberapa baik model Naive Bayes dapat mengklasifikasikan data dengan tepat, mencakup kemampuan untuk mengidentifikasi baik kelas positif maupun kelas negatif dengan akurat.

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} x100\% \quad (3)$$

Rumus diatas digunakan untuk mengukur kinerja suatu model klasifikasi. Disini TP (True Positive) adalah jumlah sampel yang benar diklasifikasikan sebagai positif, TN (True Negative) adalah jumlah sampel yang benar diklasifikasikan sebagai negatif, FP (False Positive) adalah jumlah sampel yang salah diklasifikasikan sebagai positif, dan FN (False Negatives) adalah jumlah sampel yang salah diklasifikasikan sebagai negatif. Rumus ini menghitung persentase prediksi yang benar dari total prediksi yang dilakukan oleh model.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini membaca dataset stunting Dinas Kesehatan Kota Samarinda, terdapat 9 fitur dan untuk keseluruhan terdapat 9494 records.

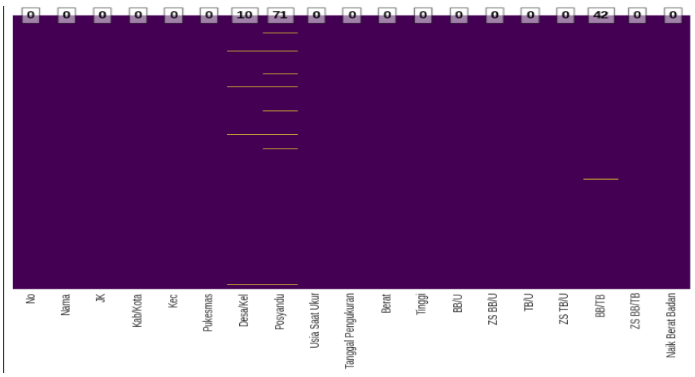
Tabel 1. Informasi Dataset

No	Kecamatan	Gizi Baik	Gizi Buruk	Gizi Kurang	Gizi Lebih	Obesitas	Risiko Gizi Lebih	Blanks	Jumlah
1	Loa Janan Ilir	1723	44	303	53	19	129	10	2281
2	Palaran	665	8	87	18	14	55	1	848
3	Samarinda Ilir	270	14	104	19	9	38	4	458
4	Samarinda Kota	270	11	53	6	4	17	1	362
5	Samarinda Seberang	565	15	84	19	7	52	8	750
6	Samarinda Ulu	835	7	123	33	10	89	2	1099
7	Samarinda Utara	494	4	46	12	5	30	3	594
8	Sambutan	553	16	72	8	10	36	3	698
9	Sungai Kunjang	1102	26	147	52	32	130	7	1496
10	Sungai Pinang	682	8	119	11	9	76	3	908
	Total	**7159**	**153**	**1138**	**231**	**119**	**652**	**42**	**9494**

3.2 Preprocessing Data

Data preprocessing adalah langkah awal dan krusial dalam analisis data serta pengembangan model machine learning yang mencakup pembersihan data (data cleaning), transformasi data (data transformation), dan seleksi data (normalitation data).

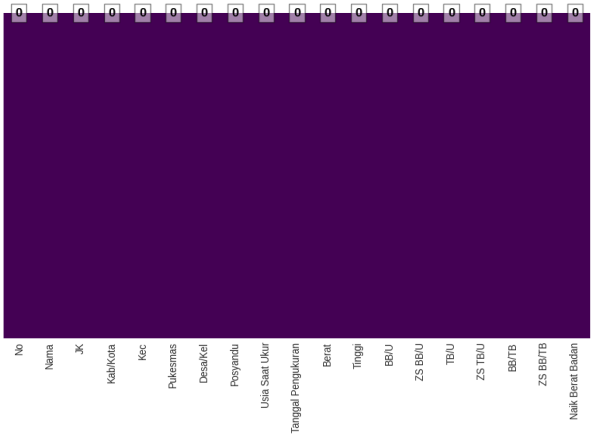
a. Data Cleaning
Proses pertama yang dilakukan yaitu penghapusan atau perbaikan nilai-nilai hilang, tidak valid, serta deteksi dan penanganan outlier untuk memastikan integritas dan kualitas dataset yang digunakan dalam analisis. Proses tersebut ditunjukkan pada gambar 2



Gambar 2. Data Sebelum dibersihkan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

Gambar di atas menunjukkan visualisasi missing values dalam sebuah DataFrame menggunakan heatmap. Warna ungu menunjukkan data yang tersedia, sementara garis kuning menunjukkan missing values. Di bagian atas setiap kolom, terdapat angka yang menunjukkan jumlah missing values dalam kolom tersebut. Dari gambar ini, terlihat bahwa kolom 'Puskesmas', 'Desa/Kel', 'Posyandu', dan 'Naik Berat Badan' memiliki missing values, dengan jumlah masing-masing 10, 71, 42, dan 1. Kolom lainnya tidak memiliki missing values. Visualisasi ini membantu untuk cepat mengidentifikasi dan menangani missing values dalam dataset. Dibawah adalah gambar 3 data yang telah dibersihkan



Gambar 3. Data setelah dibersihkan

Pada gambar 3 adalah hasil data yang sudah dibersihkan, data yang mengandung missing value sudah tidak ada. Peneliti juga melakukan penghapusan terhadap atribut yang tidak diperlukan seperti “No”, “Nama” dan “Tanggal Pengukuran”.

b. Data Transformation

Data transformation digunakan untuk mengubah data kedalam tipe yang sesuai dalam data mining. Seperti data yang masih terdapat objek maka akan diubah ke tipe numerik seperti yang di tunjukkan pada gambar 4 dibawah

	No	Nama	JK	Kab/Kota	Kec	Puskesmas	Desa/Kel	Posyandu	Usia Saat Ukur	Tanggal Pengukuran	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	Naik Berat Badan
0	1	JANUARSIH GORIA ELIORA	P	SAMARINDA	SUNGAI KUNJANG	WONOREJO	KARANG ANYAR	HARAPAN KITA	0 Tahun - 0 Bulan - 0 Hari	2023-01-01	3.2	45.0	Berat Badan Normal	-0.07	Pendek	-2.23	Gizi Lebih	2.77	-
1	2	SITI AISYAH	P	SAMARINDA	SUNGAI PINANG	REMAJA	TEMINDUNG PERMAI	PULAU INDAH	4 Tahun - 0 Bulan - 16 Hari	2023-01-02	12.0	94.0	Kurang	-2.25	Pendek	-2.09	Gizi Baik	-1.46	O
2	3	RAYYAN RAMADHAN	L	SAMARINDA	SAMBUTAN	SAMBUTAN	SAMBUTAN	AGLO NEMA	2 Tahun - 7 Bulan - 24 Hari	2023-01-02	11.0	85.0	Berat Badan Normal	-1.81	Pendek	-2.35	Gizi Baik	-0.73	O
3	4	MUHAMAD RAZZAN ARKANZA	L	SAMARINDA	SAMBUTAN	SAMBUTAN	SAMBUTAN	KEMUNING	0 Tahun - 11 Bulan - 5 Hari	2023-01-02	8.0	70.0	Berat Badan Normal	-1.52	Pendek	-2.03	Gizi Baik	-0.63	T
4	5	HAZZIMA RENNA QANITA	P	SAMARINDA	SAMBUTAN	SAMBUTAN	SAMBUTAN	KEMUNING	1 Tahun - 11 Bulan - 15 Hari	2023-01-02	9.0	75.0	Berat Badan Normal	-1.96	Sangat Pendek	-3.43	Gizi Baik	-0.18	O
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9487	9488	AFSHA	P	SAMARINDA	SUNGAI KUNJANG	LOA BAKUNG	LOA BAKUNG	MERAK	2 Tahun - 7 Bulan - 12 Hari	2023-07-31	9.4	82.0	Kurang	-2.68	Pendek	-2.90	Gizi Baik	-1.28	O
9488	9489	SHABILA NUR ZAHRA	P	SAMARINDA	SAMARINDA KOTA	SAMARINDA KOTA	SUNGAI PINANG LUAR	MELATI	4 Tahun - 8 Bulan - 19 Hari	2023-07-31	11.4	96.0	Sangat Kurang	-3.21	Pendek	-2.51	Gizi Kurang	-2.50	N
9489	9490	Shanum Alesha S	P	SAMARINDA	SAMARINDA KOTA	SAMARINDA KOTA	BUGIS	BUGIS	3 Tahun - 0 Bulan - 28 Hari	2023-07-31	10.9	87.4	Kurang	-2.02	Pendek	-2.15	Gizi Baik	-1.10	T
9490	9491	MUHAMMAD RAFASSYA HAFIZ	L	SAMARINDA	SAMARINDA SEBERANG	MANGKUPALAS	MESJID	LAIS	1 Tahun - 0 Bulan - 28 Hari	2023-07-31	8.2	71.2	Berat Badan Normal	-1.68	Pendek	-2.33	Gizi Baik	-0.72	O
9493	9494	BY NY SENNI	P	SAMARINDA	SAMARINDA SEBERANG	MANGKUPALAS	MESJID	LELE	0 Tahun - 0 Bulan - 0 Hari	2023-07-31	2.6	46.0	Sangat Kurang	-3.34	Sangat Pendek	-3.61	Gizi Baik	-0.63	-

9382 rows × 19 columns

Gambar 4. Sebelum di Transformasi

	JK	Kec	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB
0	1	8	0	3.2	45.0	0	-0.07	0	-2.23	3	2.77
1	1	9	1476	12.0	94.0	1	-2.25	0	-2.09	0	-1.46
2	0	7	964	11.0	85.0	0	-1.81	0	-2.35	0	-0.73
3	0	7	335	8.0	70.0	0	-1.52	0	-2.03	0	-0.63
4	1	7	710	9.0	75.0	0	-1.96	1	-3.43	0	-0.18
...	...	...	...	...	...	...	...	...	...	...	...
9487	1	8	952	9.4	82.0	1	-2.68	0	-2.90	0	-1.28
9488	1	3	1719	11.4	96.0	3	-3.21	0	-2.51	2	-2.50
9489	1	3	1123	10.9	87.4	1	-2.02	0	-2.15	0	-1.10
9490	0	4	393	8.2	71.2	0	-1.68	0	-2.33	0	-0.72
9493	1	4	0	2.6	46.0	3	-3.34	1	-3.61	0	-0.63

9382 rows × 11 columns

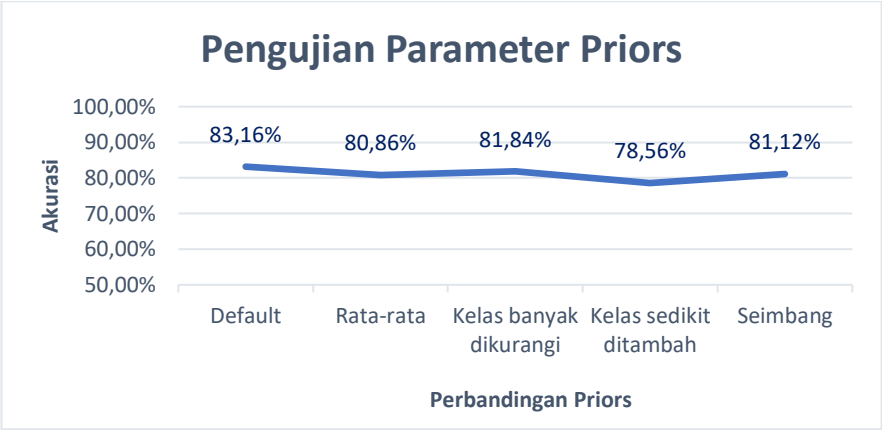
Gambar 5. Setelah di Transformasi

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

Gambar 5 diatas tampilan data pada kolom 'JK', 'BB/U', 'TB/U', 'ZS TB/U', 'BB/TB', 'ZS BB/TB' setelah dilakukan transformasi data dimana data yang sebelumnya berupa String di ubah menjadi Integer untuk memudahkan proses klasifikasi.

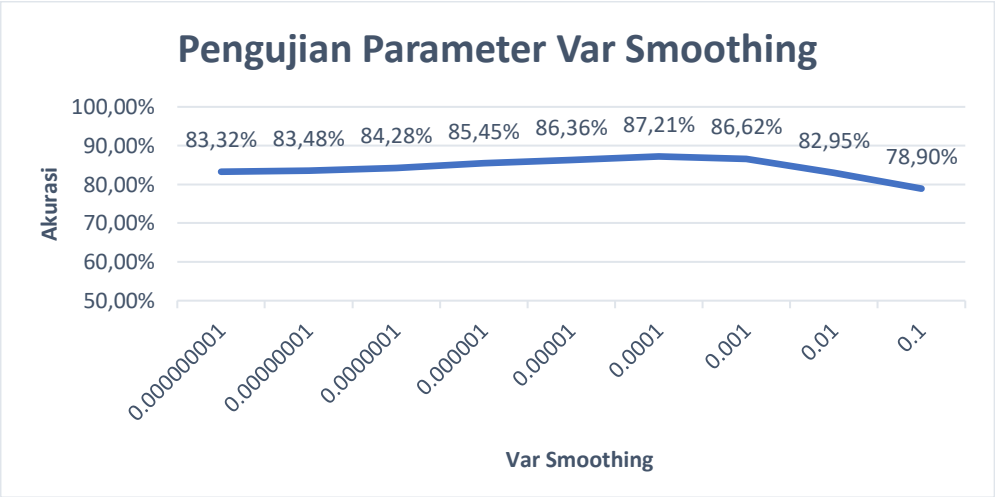
3.3 Permodelan

Metode klasifikasi Naive Bayes menjadi kunci dalam pengembangan model untuk menentukan status gizi balita. Dalam penelitian ini, Naive Bayes digunakan dengan model Gaussian untuk memprediksi label klasifikasi dari data uji.. Dalam upaya mencapai model dengan akurasi terbaik, penting untuk memilih parameter yang optimal, seperti pemilihan model distribusi dan perhitungan probabilitas yang sesuai dengan karakteristik dataset yang digunakan. Bisa dilihat pada gambar 6 dibawah ini



Gambar 6. Pengujian Parameter Priors

Hasil pengujian menunjukkan bahwa model dengan priors default (tanpa penyesuaian) mencapai akurasi tertinggi sebesar 83.16%. Ketika priors dibagi rata, akurasi menurun menjadi 80.86%. Mengurangi prior kelas dengan data terbanyak dan menambah prior kelas dengan data sedikit menghasilkan akurasi masing-masing 81.84% dan 78.56%. Sementara itu, pengaturan priors yang seimbang memberikan akurasi sebesar 81.12%. Sehingga dari pengujian ini nilai priors default “None” diterapkan untuk pengujian selanjutnya. Setelah dilakukannya pengujian priors langkah berikutnya adalah melakukan pengujian parameter var smoothing bisa di lihat pada gambar 7 dibawah ini



Gambar 7. Pengujian Var Smoothing

Gambar tersebut menunjukkan hasil pengujian parameter var smoothing pada model klasifikasi, di mana sumbu horizontal mewakili nilai var smoothing yang diuji, dan sumbu vertikal menunjukkan akurasi model yang dihasilkan. Dari grafik terlihat bahwa peningkatan nilai var smoothing dari 0.00000001 hingga 0.0001 meningkatkan akurasi model dari 83.85% menjadi 88.65%. Namun, peningkatan lebih lanjut nilai var smoothing hingga 0.1 menyebabkan penurunan akurasi hingga 79.54%. Ini menunjukkan bahwa ada nilai optimal var smoothing di sekitar 0.0001 untuk mendapatkan akurasi terbaik.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

3.4 Pengujian Rasio Data

Hasil pengujian model Gaussian Naïve Bayes dengan berbagai rasio pembagian data menunjukkan variasi performa pada data training dan data testing. Berikut adalah tabel hasil pengujian dengan rasio yang berbeda-beda:

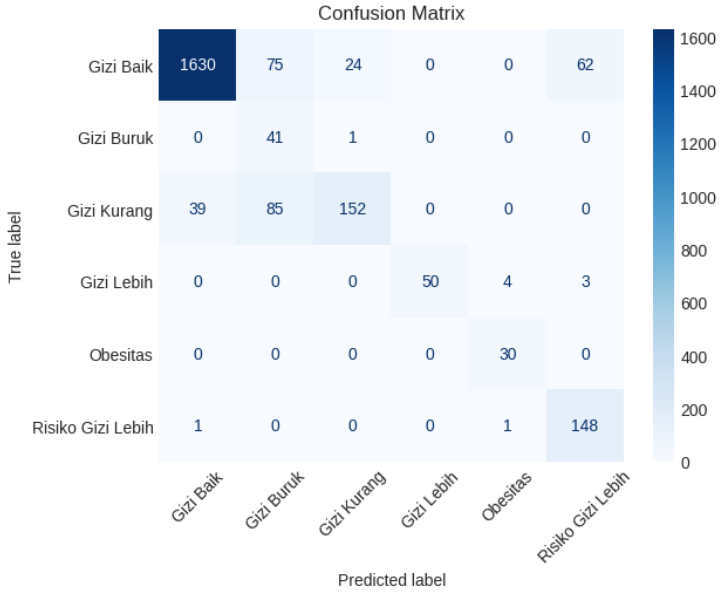
Tabel 2. Hasil Rasio Data

Persentase Rasio Data	Jumlah Data *Training*	Jumlah Data *Testing*	Akurasi
60% : 40%	5629	3753	87.21%
65% : 35%	6098	3284	86.84%
70% : 30%	6567	2815	87.35%
75% : 25%	7036	2346	87.42%
80% : 20%	7505	1877	87.21%
85% : 15%	7974	1408	86.57%
90% : 10%	8443	939	86.79%

Dari hasil tabel di atas, model Gaussian Naïve Bayes menunjukkan akurasi tertinggi sebesar 87,42% pada rasio data 75%:25%. Akurasi terendah tercatat sebesar 86,57% pada rasio data 85%:15%, sedangkan pada rasio data 70%:30%, akurasi mencapai 87,35%. Ini menunjukkan bahwa rasio 75%:25% memberikan hasil terbaik dalam hal akurasi model.

3.5 Hasil dan Validasi

Untuk mengevaluasi performa model Naïve Bayes Classifier dalam mengklasifikasikan status gizi balita, dilakukan pengujian menggunakan Confusion Matrix seperti yang ditunjukkan pada gambar 8 dibawah ini:



Gambar 8. Hasil Confusion Matrix

Gambar 8 di atas adalah matriks kebingungan (confusion matrix) yang menunjukkan kinerja suatu model klasifikasi dalam mengidentifikasi status gizi dari data sampel. Matriks ini menunjukkan perbandingan antara label sebenarnya (True label) dan prediksi model (Predicted label) untuk beberapa kategori gizi: Gizi Baik, Gizi Buruk, Gizi Kurang, Gizi Lebih, Obesitas, dan Risiko Gizi Lebih. Dari matriks ini, kita dapat melihat bahwa model paling akurat dalam mengklasifikasikan kategori "Gizi Baik" dengan 1630 prediksi benar, tetapi memiliki beberapa kesalahan klasifikasi dalam kategori lainnya seperti "Gizi Kurang" dan "Gizi Buruk". Misalnya, ada 85 sampel yang sebenarnya termasuk "Gizi Kurang" tetapi diprediksi sebagai "Gizi Baik". Hal ini menunjukkan area di mana model dapat ditingkatkan untuk akurasi yang lebih baik.

$$Accuracy = \frac{Correct\ Prediction}{Total\ Instance}$$

$$Accuracy = \frac{(1630 + 41 + 152 + 50 + 30 + 148)}{2.346}$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

$$Accuracy = \frac{2051}{2346}$$

$$Accuracy = \ 0{,}8742 = 87.42\%$$

Hasil perhitungan di atas menunjukkan bahwa model klasifikasi yang digunakan memiliki akurasi sebesar 87,42%. Akurasi dihitung dengan membagi jumlah prediksi yang benar (jumlah dari 1630, 41, 152, 50, 30, dan 148) dengan total instance yang ada, yaitu 2346. Ini berarti bahwa dari 2346 instance, 2085 instance diklasifikasikan dengan benar oleh model, menunjukkan bahwa model tersebut cukup baik dalam memprediksi dengan benar sebagian besar instance yang diuji.

4. KESIMPULAN

Hasil dari pengujian menggunakan Gaussian Naive Bayes menunjukkan bahwa akurasi tertinggi dicapai pada rasio 75% data training dan 25% data testing dengan akurasi 87.42%. Akurasi ini sedikit lebih tinggi dibandingkan dengan rasio 70%:30% yang mencapai 87.35%, dan menunjukkan tren kenaikan akurasi hingga titik tersebut. Akurasi terendah terjadi pada rasio 85% data training dan 15% data testing dengan akurasi 86.57%. Secara keseluruhan, meskipun ada variasi dalam rasio data, akurasi tetap berada dalam kisaran yang relatif sempit, dengan fluktuasi kecil di sekitar angka 87%.

REFERENCES

[1] N. Rismayanti, A. Naswin, U. Zaky, M. Zakariyah, and D. A. Purnamasari, 'Evaluating Thresholding-Based Segmentation and Humoment Feature Extraction in Acute Lymphoblastic Leukemia Classification using Gaussian Naive Bayes', *Int. J. Artif. Intell. Med. Issues*, vol. 1, no. 2, pp. 74–83, 2023, doi: 10.56705/ijaimi.v1i2.99.

[2] M. Nesca, A. Katz, C. K. Leung, and L. M. Lix, 'A scoping review of preprocessing methods for unstructured text data to assess data quality', *Int. J. Popul. Data Sci.*, vol. 7, no. 1, pp. 1–15, 2022, doi: 10.23889/ijpds.v7i1.1757.

[3] O. S. Ads, M. M. Alfares, and M. A. M. Salem, 'Multi-limb Split Learning for Tumor Classification on Vertically Distributed Data', *Proc. - 2021 IEEE 10th Int. Conf. Intell. Comput. Inf. Syst. ICICIS 2021*, pp. 88–92, 2021, doi: 10.1109/ICICIS52592.2021.9694163.

[4] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, 'A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data', *Front. Energy Res.*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/fenrg.2021.652801.

[5] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, *Data preprocessing in predictive data mining*, vol. 34, no. January. 2019. doi: 10.1017/S026988891800036X.

[6] D. Jeevaraj, B. Karthik, T. Vijayan, and M. Sriram, 'Feature Selection Model using Naive Bayes ML Algorithm for WSN Intrusion Detection System 179 Original Scientific Paper', *Int. J. Electr. Comput. Eng. Syst.*, vol. 14, no. November 2, pp. 179–185, 2023.

[7] C. K. Tan, C. P. Tan, and N. Shaun Wes, *Information Technology Students' Preferences on Blended Learning*, vol. 724. 2021. doi: 10.1007/978-981-33-4069-5_9.

[8] F. D. Pratama, I. Zufria, and T. Triase, 'Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar', *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 77–84, 2022, doi: 10.36341/rabit.v7i1.2217.

[9] Nugroho Arif Sudibyo, Ardymulya Iswardani, Kartika Sari, and Siti Suprihatiningsih, 'Penerapan Data Mining Pada Jumlah Penduduk Miskin Di Indonesia', *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 3, pp. 199–207, 2020, doi: 10.46306/lb.v1i3.42.

[10] G. Gusnedi *et al.*, 'Risk factors associated with childhood stunting in Indonesia: A systematic review and meta-analysis', *Asia Pac. J. Clin. Nutr.*, vol. 32, no. 2, pp. 184–195, 2023, doi: 10.6133/apjcn.202306_32(2).0001.

[11] M. Abbas, K. Ali, A. Jamali, K. Ali Memon, and A. Aleem Jamali, 'Multinomial Naive Bayes Classification Model for Sentiment Analysis', *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, pp. 62–67, 2019, doi: 10.13140/RG.2.2.30021.40169.

[12] R. Nailuvar and I. Laily Hilmi, 'Analysis of Factors Affecting Stuting Incidence in Indonesia : Literature riview', *J. eduhealth*, vol. 13, no. 02, p. 1099, 2022, [Online]. Available: http://ejournal.seaninstitute.or.id/index.php/healt

[13] S. Zaleha and H. Idris, 'Implementation of Stunting Program in Indonesia: a Narrative Review', *Indones. J. Heal. Adm.*, vol. 10, no. 1, pp. 143–151, 2022, doi: 10.20473/jaki.v10i1.2022.143-151.

[14] D. Simbolon, Asmawati, B. Battbual, I. D. R. Ludji, and Eliana, 'Pendampingan Gizi Spesifik Pada Ibu Hamil Upaya Menuju Kampung KB Bebas Stunting', *Edukasi Masy. Sehat Sejah. J. Pengabdi. Kpd. Masy.*, vol. 3, no. 2, pp. 112–121, 2021.

[15] Harliana and D. Anggraini, 'Penerapan Algoritma Naïve Bayes Pada Klasifikasi Status Gizi Balita di Posyandu

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 627−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/ bits.v6i2.5493

Desa Kalitengah', *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 21, no. 2, pp. 38–45, 2023, doi: 10.61805/fahma.v21i2.16.

[16] E. Bukhari, 'Pengaruh Dana Desa dalam Mengentaskan Kemiskinan Penduduk Desa', *J. Kaji. Ilm.*, vol. 21, no. 2, pp. 219–228, 2021, doi: 10.31599/jki.v21i2.540.

[17] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, 'Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation', *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 578–586, 2023, doi: 10.47065/josyc.v4i3.3414.

[18] UNICEF, 'Mengatasi tiga beban malnutrisi di Indonesia', *unicef*, 2023. https://www.unicef.org/indonesia/id/gizi?gad_source=1&gclid=Cj0KCQjwsuSzBhCLARIsAIcdLm4VJrL8AgOdYCjtDtyE9qcXJsHdHkBNFevVNAxGy7QWz0rza7aQD_oaAte6EALw_wcB (accessed Jun. 26, 2024).

[19] R. Nurida, E. Sugiharti, and A. Alamsyah, 'Implementation of Fuzzy K-Nearest Neighbor Method in Decision Support System for Identification of Under-five Children Nutritional Status Based on Anthropometry Index', *J. Adv. Inf. Syst. Technol.*, vol. 1, no. 1, pp. 83–89, 2019.

[20] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, 'Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation', *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.

[21] kemenkes, *Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi, dan Kabupaten/Kota Tahun 2021*. Kementerian Kesehatan Republik Indonesia, 2021. [Online]. Available: https://www.badankebijakan.kemkes.go.id/buku-saku-hasil-studi-status-gizi-indonesia-ssgi-tahun-2021/