

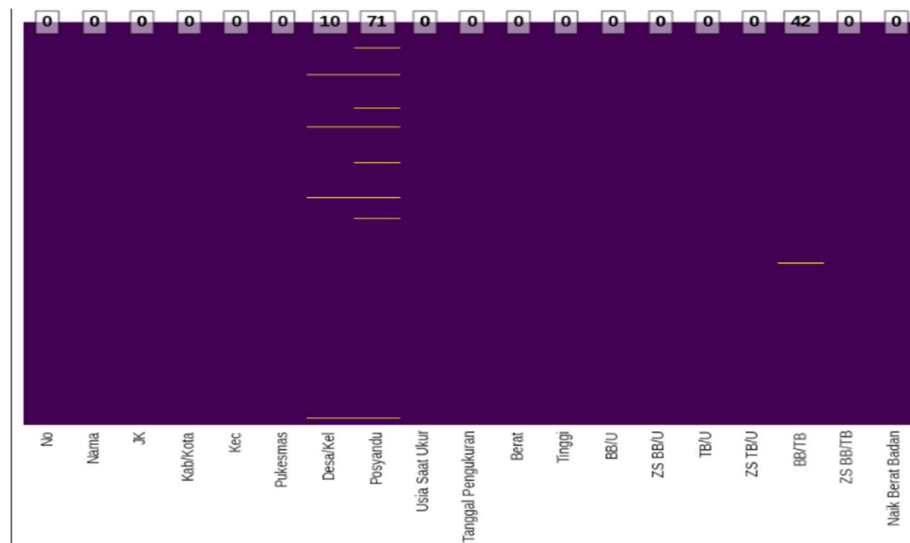
## BAB III HASIL PEMBAHASAN

### 3.1 Hasil Penelitian

Penelitian ini bertujuan untuk mengimplementasikan algoritma *Naïve Bayes Classifier* dalam klasifikasi status gizi pada balita di Kota Samarinda. Proses penelitian melibatkan beberapa tahapan, yaitu pengumpulan data, *preprocessing* data, implementasi algoritma *Gaussian Naïve Bayes* dan evaluasi kinerja model.

#### 3.1.1 Data Cleaning

Pada tabel 3.1 merupakan pengecekan terhadap *missing value* atau nilai yang kosong dalam dataset. Data yang mengandung *missing value* dihapus agar tidak mengurangi hasil akurasi.



**Gambar 3.1** Data sebelum dibersihkan

Pada Gambar 3.2, data yang mengandung *missing value* dihapus agar tidak mengurangi hasil akurasi. atribut yang memiliki *missing value* adalah “Desa/Kel”, “Posyandu” dan “BB/TB”, selanjutnya dilakukan penghapusan dengan menggunakan perintah `.dropna(inplace=True)`. Dan adapun hasil dari data yang telah dibersihkan bisa dilihat pada gambar 3.2 dibawah

No	Nama	JK	Kab/Kota	Kec	Pukesmas	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	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**Gambar 3.2** Data setelah dibersihkan

Pada Gambar 3.2 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”.

### 3.1.2 Data Transformation

Beberapa atribut yang ditransformasi dalam dataset ini meliputi: JK (Jenis Kelamin), Kec (Kecamatan), Usia Saat Ukur, Berat, Tinggi, BB/U (Berat Badan per Umur), ZS BB/U (Z-Score Berat Badan per Umur), TB/U (Tinggi Badan per Umur), ZS TB/U (Z-Score Tinggi Badan per Umur), BB/TB (Berat Badan per Tinggi Badan), ZS BB/TB (Z-Score Berat Badan per Tinggi Badan). Data ini telah ditransformasi dengan menghitung Z-Score untuk setiap pengukuran berat dan tinggi, yang berguna untuk mengidentifikasi penyimpangan dari rata-rata populasi. Data yang sudah diproses ini siap digunakan untuk implementasi model analisis kesehatan anak atau pemodelan statistik lebih lanjut. Berikut adalah data sebelum dan sesudah transformasi dapat dilihat pada gambar 3.3 dan 3.4

No	Nama	JK	Kab/Kota	Kec	Pukesmas	Desa/Ke1	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	JANJARSH GORIA ELIORA	P	SAMARINDA	SUNGGAI 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	SUNGGAI PINANG	REMAJA	TEMUNDUNG 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 RAMACHAN	L	SAMARINDA	SAMBUTAN	SAMBUTAN	SAMBUTAN	AGLO NEIMA	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	SUNGGAI 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	SUNGGAI 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	-

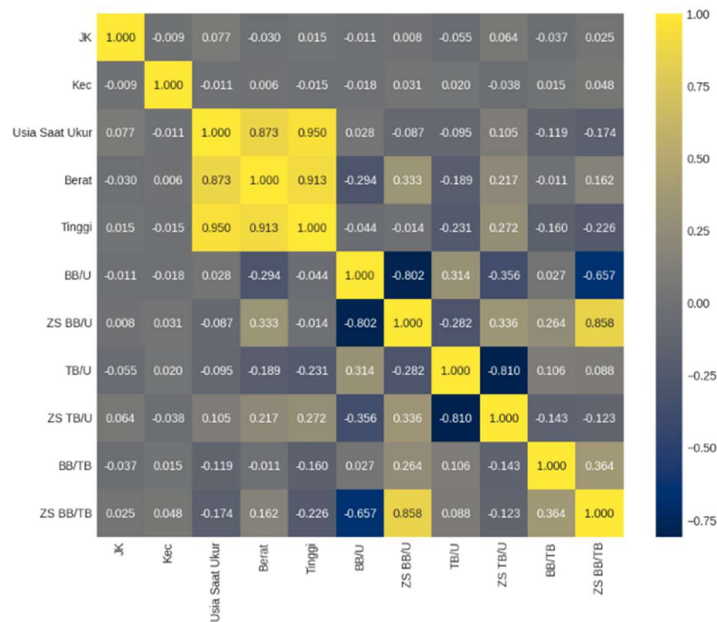
Gambar 3.3 Data sebelum di transformasi

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

Gambar 3.4 Setelah di Transformasi

Pada gambar 3.4 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.

Selain dari atribut yang dihapus tersebut, peneliti juga melakukan perhitungan statistik dengan melihat hubungan antar atribut dengan menggunakan *correlation matrix*. Penerapan perintah ini peneliti dapat melihat dan menganalisa hubungan korelasi kuat antar atribut. Dari hasil *correlation matrix*, peneliti akan menimbangakan beberapa atribut yang perlu dan tidak perlu yang digunakan untuk tahapan selanjutnya seperti ditunjukkan pada Gambar 3.5



Gambar 3.5 Korelasi data

Berdasarkan Gambar 3.5, menjelaskan terkait hasil dari tahapan *correlation matrix* yang menampilkan hubungan antara atribut dengan kelas target. Dari hubungan ini peneliti dapat melakukan analisa singkat terkait atribut yang kurang berpengaruh terhadap kelas target. Terdapat dua hasil hubungan yang dihasilkan dari proses ini yaitu hubungan linear positif dan linear negatif. Semakin dekat dengan nilai 0 maka atribut tersebut memiliki korelasi rendah dengan kelas target. Beberapa atribut yang dihapus dalam tahapan ini adalah *Puskesmas* (0,002), *Desa/Kel* (-0,003), *Posyandu* (0,001), dan *Naik Berat Badan* (0,008).

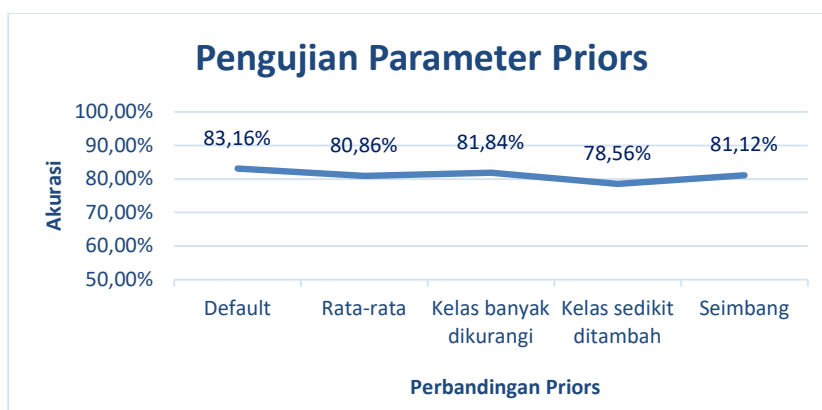
Data selanjutnya dipisahkan menjadi dua variabel yang berbeda yaitu x dan y. variabel x merupakan kolom yang hanya berisikan atribut saja dan variabel y berisi target atau kelas. Untuk variabel x, selanjutnya dilakukan proses normalisasi data untuk menstabilkan range data dan menyelaraskan skala dari masing masing atribut guna meningkatkan kinerja dari model klasifikasi. Normalisasi dibuat dengan menggunakan skala 0-1, hal ini akan mengubah nilai dari setiap atribut menjadi range 0 sampai 1 saja. Normalisasi data akan memberikan pengaruh yang seimbang terhadap model klasifikasi tanpa memberikan bobot yang berlebihan terhadap atribut yang memiliki skala yang lebih besar.

### 3.1.3 Pembagian Data

Data yang sudah dilakukan proses data *Preprocessing* selanjutnya mulai dilakukan pembagian data untuk dibagi menjadi data *training* dan data *testing*. Proses ini bertujuan untuk melakukan pengujian model terhadap data yang belum pernah dilihat sebelumnya. Rasio pembagian data yang digunakan dalam proses ini adalah sebesar 75% digunakan untuk data *training* dengan total data sebanyak 7036 dan 25% digunakan untuk data *testing* dengan total data sebanyak 2346. Pelatihan model akan diterapkan terlebih dahulu di data *training* sebelum diaplikasikan ke data *testing* maupun data baru yang ingin diuji.

### 3.1.4 Implementasi Naïve Bayes

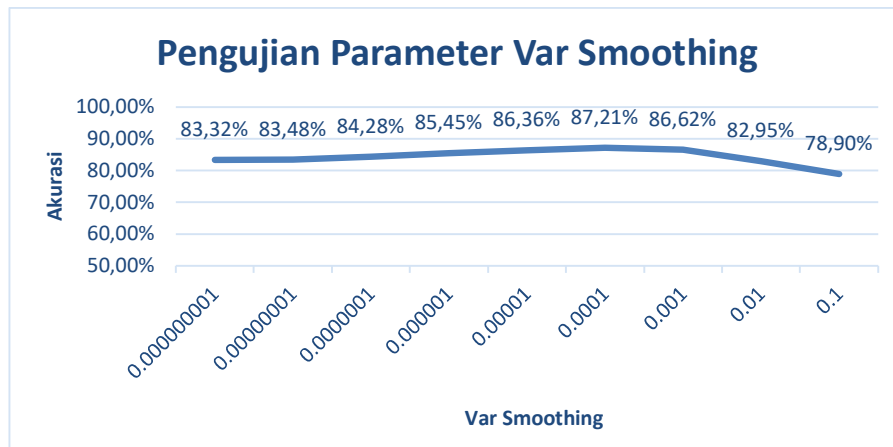
Metode klasifikasi Naive Bayes menjadi kunci dalam pengembangan model untuk menentukan status gizi balita. Dalam penelitian ini, *Naive Bayes* digunakan dengan model *Gaussian* untuk memprediksi label klasifikasi dari data uji.. Dalam upaya mencapai model dengan akurasi terbaik, penting untuk memilih parameter yang optimal, seperti pemilihan model distribusi dan perhitungan probabilitas yang sesuai dengan karakteristik dataset yang digunakan. Bisa dilihat pada gambar 3.6 dibawah ini



Gambar 3. 6 Pengujian Parameter Priors

Grafik di atas menggambarkan hasil pengujian akurasi dari model *Gaussian Naive Bayes* dengan berbagai pengaturan *priors* pada dataset yang digunakan. Lima jenis pengaturan *priors* diuji dalam eksperimen ini: *priors* default (None) yang menggunakan prior bawaan dari model berdasarkan distribusi data pelatihan; *priors* dibagi rata, di mana setiap kelas memiliki prior yang sama sebesar 1/6 atau sekitar 0.1667; *priors* di mana kelas dengan data terbanyak dikurangi, memberikan *prior* yang lebih rendah pada kelas tersebut; *priors* di mana kelas dengan data paling sedikit ditambah, memberikan prior yang lebih tinggi pada kelas tersebut; dan *priors* yang dibagi seimbang tetapi tidak merata, memberikan variasi untuk setiap kelas.

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 3.7 dibawah ini



**Gambar 3. 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.

### 3.1.5 Evaluasi Model

Setelah melakukan pengujian menggunakan *Gaussian Naïve Bayes*, tahapan selanjutnya melakukan evaluasi hasil pengujian model *Gaussian Naïve Bayes* dengan berbagai rasio pembagian data. Berikut adalah tabel hasil pengujian dengan rasio yang berbeda-beda:

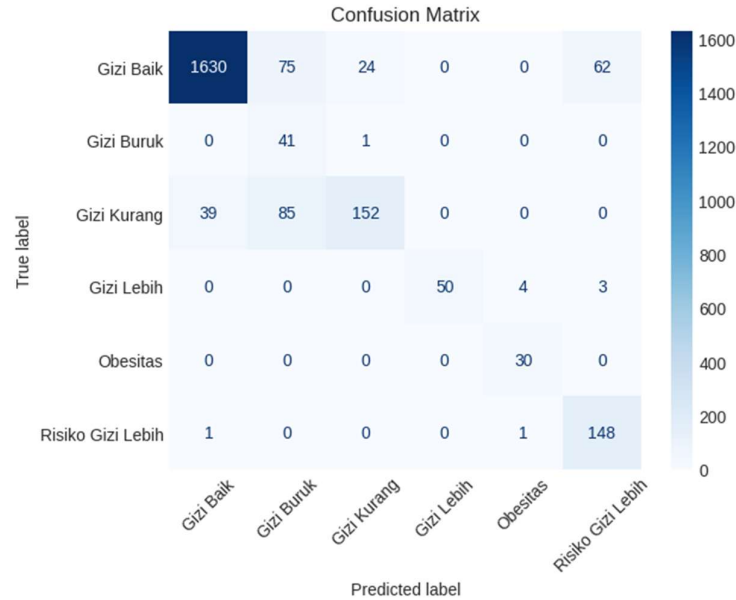
**Tabel 3. 1** Hasil Pengujian Rasio Data

Persentase Rasio Data	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>	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.

### 3.1.6 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 3.8 dibawah ini:



**Gambar 3. 8** *Confusion Matrix*

Gambar 3.8 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.

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

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

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

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

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, 2051 instance diklasifikasikan dengan benar oleh model, menunjukkan bahwa model tersebut cukup baik dalam memprediksi dengan benar sebagian besar instance yang diuji.