

Klasifikasi Risiko Stunting Pada Balita Menggunakan Algoritma Naive Bayes

Putri Navizah^{1*}, Nur Azise², Fajriyanto³

^{1,2,3} Program Studi Sistem Informasi Fakultas Sains Dan Teknologi, Universitas Ibrahimy, Situbondo

*Korespondensi Penulis. E-mail: putrnavizah2@gmail.com

Abstrak

UPTD Puskesmas Pujer yang terletak di Kabupaten Bondowoso menghadapi kesulitan dalam pengelolaan serta pemantauan data mengenai stunting pada anak-anak balita dengan masih mengandalkan metode pencatatan manual dan semi-digital. Situasi ini menyebabkan efisiensi yang sangat terbatas, risiko kesalahan input yang relatif tinggi, dan proses analisis yang lambat untuk secara akurat mengidentifikasi kebutuhan intervensi. Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes dalam mengembangkan sistem yang dapat mengklasifikasikan risiko stunting pada balita, dengan kemampuan untuk secara cepat dan otomatis mengenali status risiko berdasarkan rekam medis yang telah dikumpulkan. Dalam pengembangan model klasifikasi berbasis data mining ini, pendekatan Gaussian Naive Bayes digunakan bersama pemrograman Python, serta melalui langkah-langkah preprocessing data, normalisasi menggunakan Min-Max Scaling, dan pemisahan data menjadi training set (80%) dan testing set (20%), yang kemudian dievaluasi menggunakan metrik akurasi dari Confusion Matrix. Variabel yang dianalisis di dalam sistem ini mencakup jenis kelamin, usia, tinggi badan, berat badan, dan alamat. Hasil evaluasi menunjukkan tingkat akurasi mencapai 92.5%. Temuan penelitian mengindikasikan bahwa algoritma Naive Bayes dapat mengelompokkan status risiko stunting dengan efisien, otomatis, dan akurat, sehingga model ini sangat layak digunakan sebagai sistem pendukung keputusan bagi puskesmas dalam melakukan deteksi awal, pemetaan distribusi wilayah, serta pengoptimalan program pencegahan dan edukasi gizi kepada masyarakat.

Kata kunci: Stunting, Data Mining, Klasifikasi, Naive Bayes, Sistem Pendukung Keputusan

Abstract

UPTD Puskesmas Pujer, located in Bondowoso Regency faces challenges in managing and monitoring data on stunting in toddlers, relying on manual and semi-digital recording methods. This situation results in very limited efficiency, a relatively high risk of input errors, and a slow analysis process to accurately identify intervention needs. This study aims to apply the Naive Bayes algorithm to develop a system that can classify stunting risk in toddlers, with the ability to quickly and automatically identify risk status based on collected medical records. In developing this data mining-based classification model, the Gaussian Naive Bayes approach is used in conjunction with Python programming, as well as through data preprocessing steps, normalization using Min-Max Scaling, and data separation into training sets (80%) and testing sets (20%), which are then evaluated using the Confusion Matrix accuracy metric. The variables analyzed in this system include gender, age, height, weight, and address. The evaluation results show an accuracy level of 92.5%. The research findings indicate that the Naive Bayes algorithm can classify stunting risk status efficiently, automatically, and accurately, making this model very suitable for use as a decision support system for community health centers in conducting early detection, mapping regional distribution, and optimizing prevention programs and nutrition education for the community.

Keyword: Stunting, Data Mining, Classification, Naive Bayes, Decision Support System

PENDAHULUAN

Stunting merupakan masalah gizi yang serius dan memiliki dampak langsung terhadap pertumbuhan fisik dan perkembangan mental anak, biasanya diakibatkan oleh kekurangan gizi yang berlangsung secara terus-menerus selama seribu hari pertama kehidupan (Wahidin & Andika, 2024). Di tingkat nasional, data dari Riset Kesehatan Dasar (Riskesdas) menunjukkan prevalensi stunting

pada balita di Indonesia mencapai 37,2% pada tahun 2013 dan mengalami penurunan menjadi 30,8% pada tahun 2018(Yuda et al., 2022). Meskipun ada penurunan, angka tersebut masih belum memenuhi target nasional untuk pengurangan stunting sebesar 14% pada tahun 2024. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI) tahun 2024, prevalensi stunting di Kabupaten Bondowoso menunjukkan penurunan signifikan dari 17% pada tahun 2023 menjadi 11,2% pada tahun 2024. Walaupun fluktuasi angka ini lebih baik dibandingkan dengan rata-rata Provinsi Jawa Timur (14,7%) maupun tingkat nasional (19,8%), penanggulangan stunting masih memerlukan perhatian serius melalui metode pemantauan yang menyeluruh, berkelanjutan, dan berbasis data.

Pada penelitian terdahulu menunjukkan bahwa penggunaan data mining memiliki efektivitas dalam mengklasifikasikan status gizi. Penelitian yang dilakukan oleh Mulyanto dan rekan-rekannya di Kabupaten Poto Tano menerapkan algoritma Naive Bayes pada data sebanyak 5.829 balita dengan tingkat akurasi mencapai 95,08%(Mulyanto et al., 2024). Di sisi lain, Hardiani dan Putri menerapkan Naive Bayes Classifier menggunakan kerangka kerja CRISP-DM dengan 7.573 entri data, menghasilkan akurasi 80%, presisi 85%, dan recall 91%(Hardiani & Putri, 2024). Perbandingan antara kedua penelitian tersebut mengindikasikan bahwa Naive Bayes efisien dalam pengolahan data lapangan, tetapi cakupan variabel yang diamati masih terbatas pada aspek antropometri standar. Oleh karena itu, penelitian ini mengembangkan atribut analisis dengan menambah indikator jenis kelamin, umur, tinggi badan, berat badan, serta alamat spesifik di sekitar fasilitas kesehatan.

Kondisi nyata di lapangan menunjukkan bahwa manajemen dan pencatatan informasi tentang layanan kesehatan anak di UPTD Puskesmas Pujer saat ini masih banyak dilakukan dengan metode manual dan semi-digital. Ketidakhadiran sistem komputerisasi yang mumpuni dapat menyebabkan ketidakcocokan data, memperlambat pengolahan informasi, dan menyulitkan penyajian data visual yang jelas untuk para pengambil keputusan dalam mengevaluasi program intervensi gizi(Aisyah & Baijuri, n.d.). Untuk mengatasi keterbatasan dalam operasional ini, perlu diterapkan teknik pembelajaran mesin yang efisien dan tepat. Penelitian ini bertujuan untuk menerapkan algoritma Naive Bayes berbasis web di UPTD Puskesmas Pujer untuk secara otomatis, cepat, dan tepat menilai risiko stunting pada anak balita.

METODE PENELITIAN

Studi ini merupakan penelitian eksperimental yang berbasis pada sistem informasi data mining. Pengumpulan data sekunder dilakukan di UPTD Puskesmas Pujer, Kabupaten Bondowoso, Jawa Timur. Subjek dalam kajian ini terdiri dari seluruh data rekam medis anak-anak yang tercatat dalam sistem layanan kesehatan UPTD Puskesmas Pujer.

Metode pengumpulan data

Metode pengumpulan data dilakukan melalui gabungan empat teknik utama:

1. Wawancara merupakan proses komunikasi yang melibatkan dua orang atau lebih yang dapat berlangsung secara langsung, di mana satu orang bertindak sebagai penginterview dan yang lainnya sebagai narasumber dengan maksud tertentu, seperti untuk memperoleh informasi atau mengumpulkan data(Rahmawati et al., n.d.). Pada penelitian ini, wawancara dilakukan dengan pihak UPTD Puskesmas Pujer yang terlibat dalam pengelolaan data kesehatan balita, khususnya yang berkaitan dengan pencatatan dan pemantauan data stunting,
2. Observasi adalah cara untuk mengumpulkan informasi yang dilakukan melalui melihat dan merekam secara teratur terhadap kondisi atau perilaku objek yang sedang diteliti, baik dengan cara langsung maupun tidak langsung(Hasibuan et al., 2023). Observasi dilakukan dengan cara mengamati dan merekam secara teratur langkah-langkah pencatatan, pengelolaan, dan pelaporan data mengenai stunting pada balita di UPTD Puskesmas Pujer. Pengamatan ini berfokus pada fenomena dan proses kerja yang sedang berlangsung, dimulai dari pengumpulan data balita hingga pembuatan laporan, untuk mengidentifikasi masalah serta kebutuhan sistem yang perlu dikembangkan.

3. Tinjauan pustaka dilaksanakan oleh penulis dengan cara mencari dan menelaah buku, artikel, serta sumber-sumber ilmiah lain yang berkaitan dengan topik penelitian sebagai dasar teori, referensi, dan pedoman dalam penyusunan skripsi ini. Studi pustaka dalam penelitian ini mencakup berbagai teori yang berhubungan dengan Klasifikasi data, pengolahan data stunting, serta algoritma pembelajaran mesin seperti Naive Bayes yang diterapkan untuk menganalisis dan klasifikasi data (Chandra & ., 2023).
4. Dalam studi ini, salah satu cara untuk mengumpulkan data adalah melalui dokumentasi, di mana penulis mengumpulkan beragam dokumen yang berhubungan dengan stunting dan pengelolaan gizi di Puskesmas. Dokumentasi tersebut mencakup data catatan medis balita, laporan keadaan gizi, buku catatan pemantauan stunting, formulir untuk mengukur tinggi dan berat badan, serta pedoman atau kebijakan Puskesmas mengenai program pencegahan stunting (Ardiansyah et al., 2023).

Alur Proses

Sangkaian langkah penelitian dilaksanakan secara sistematis mengikuti alur data mining yang terdiri dari:

1. Pengumpulan data: Data ini berupa data balita dengan karakteristik berikut: Jenis kelamin, umur, berat badan, tinggi badan, tempat tinggal
2. Preprocessing Data: Melakukan pembersihan data dengan memeriksa nilai yang hilang menggunakan fungsi *isnull()*, *sum()* dan menghapus data duplikat dengan fungsi *duplicated()*, *sum()* di *Python*.
3. Seleksi Data: Memilih baris dan kolom atribut yang penting, valid dan relevan untuk di analisis.
4. Transformasi/ Normalisasi Data : Proses ini dilakukan untuk mengubah atribut yang masih dalam bentuk kategorikal (nominal) menjadi numerik (jenis kelamin: laki-laki=1, Perempuan=2; label target status : Normal = 1, Pendek = 2, sangat pendek = 3) sedangkan variable kontinu seperti Usia, TB, dan BB diseragamkan skalanya menggunakan pendekatan *Min- Max Scaling* kedalam rentang nilai 1 hingga 10 menggunakan rumus berikut:

$$Norm = \frac{dvalue - X_{min}}{X_{MAX} - X_{min}} \times 9 + 1$$

5. Metode Naïve bayes: Setelah dilakukan normalisasi, tahap selanjutnya adalah implementasi data mining menggunakan algoritma *Naive Bayes Gaussian*. Algoritma ini digunakan untuk melakukan proses klasifikasi berdasarkan probabilitas dari setiap atribut terhadap kelas tertentu. Sedangkan untuk menghitung evaluasi Tingkat akurasi menggunakan *Confusion Matrix*

Teknik Analisis Data

Proses klasifikasi menggunakan algoritma *Naive Bayes Gaussian*, memanfaatkan *library* *pandas* dan *scikit-learn* dengan *Python*. Pada tahap pertama, Impor perpustakaan yang diperlukan *Pandas* untuk menangani *data tabular (DataFrames)*, *train_test_split* untuk membagi data menjadi set pelatihan dan pengujian, *GaussianNB* sebagai model *Naive Bayes*, dan *akurasi_score* untuk mengukur seberapa akurat model tersebut. Selanjutnya dataset dibuat sebagai *DataFrame* dengan kolom JK, Norm_Umur, Norm_TB, Norm_BB dan Status. Data ini berasal dari langkah normalisasi sebelumnya, yang memastikan semua nilai numerik berada pada skala yang sama. Setelah dataset dibuat, data dibagi menjadi dua bagian: variabel X sebagai fitur (JK, Norm_Umur, Norm_TB, Norm_BB) dan variabel y sebagai label (Status) yang akan diprediksi oleh model.

Langkah selanjutnya adalah membagi data menjadi set pelatihan dan pengujian menggunakan fungsi *train_test_split*, mengalokasikan 80% untuk pelatihan dan 20% untuk pengujian. Parameter *random_state=0* memastikan pemisahan data tetap konsisten setiap kali program dijalankan. Setelah itu, model *Naive Bayes* dibangun menggunakan *GaussianNB* dan kemudian dilatih dengan data pelatihan melalui fungsi *fit* sehingga model dapat mempelajari pola yang menghubungkan fitur ke label.

Setelah model dilatih, kami menggunakan fungsi prediksi untuk membuat prediksi pada data pengujian. Hasil prediksi kemudian dibandingkan dengan data yang benar untuk mengukur seberapa baik performa model. Langkah terakhir adalah menghitung tingkat akurasi model menggunakan *akurasi_score*, yang melibatkan perbandingan nilai prediksi dan nilai sebenarnya).

HASIL DAN PEMBAHASAN

Berdasarkan hasil pemeriksaan nilai yang hilang pada dataset pengukuran gizi balita, diketahui bahwa sebagian besar atribut tidak memiliki nilai kosong sehingga data dapat dianggap cukup lengkap. Namun, ada beberapa atribut yang masih memiliki nilai yang hilang. Atribut dengan jumlah nilai yang hilang tertinggi terdapat pada kolom yang tidak bernama (*Unnamed*), di mana sebagian besar datanya tidak terisi. Selain itu, ada beberapa atribut lain yang memiliki nilai yang hilang dalam jumlah kecil. Kondisi ini menunjukkan bahwa masih perlu melakukan tahap pembersihan data, seperti menghapus atribut yang tidak relevan atau menangani nilai yang hilang, agar kualitas data menjadi lebih baik dan hasil analisis yang diperoleh lebih tepat. Berikut tabel 1 menunjukkan hasil *cek missing values*:

Tabel 1 Hasil Cek *Missing Values*

No	Nama Atribut	Jumlah <i>Missing Values</i>
1.	No	0
2.	NIK	0
3.	Nama	0
4.	JK	0
5.	Tgl Lahir	0
6.	BB Lahir	0
7.	TB Lahir	0
8.	Nama Ortu	0
9.	Prov	0
10.	Kab/Kota	0
11.	Kec	0
12.	Puskesmas	0
13.	Desa/Kel	0
14.	Posyandu	61
15.	RT	0
16.	RW	0
17.	Alamat	0
18.	Usia Saat Ukur	0
19.	Tgl Pengukuran	0

Tabel 1 Hasil Cek *Missing Values*

No	Nama Atribut	Jumlah <i>Missing Values</i>
20.	Berat	0
21.	Tinggi	0
22.	Cara Ukur	0
23.	LiLA	2199
24.	BB/U	0
25.	ZS BB/U	0
26.	TB/U	0
27.	ZS TB/U	0
28.	BB/TB	0
29.	ZS BB/TB	0
30.	Naik BB	0
31.	Jml Vit A	2205
32.	KPSP	70
33.	KIA	5
34.	Kelas Ibu Balita	0
35.	MBG	0

Berikut segmen program Pengecekan nilai yang hilang atau data kosong, pengecekan data duplikat dan perbaikan data yang tidak konsisten. Untuk memastikan bahwa data yang digunakan dalam proses analisis memiliki kualitas yang baik. Sebagai berikut:

Segmen Porgram 1 cek *Missing Values* dan *Duplicate*

```
#chek
print("\n=== JUMLAH MISSING VALUE
===")
print(df.isnull().sum())
#CEK JUMLAH DATA DUPLIKAT
jumlah_duplikat =
df.duplicated().sum()
```

Tahap selanjutnya adalah tahap seleksi data, atribut yang paling penting untuk dianalisis lebih lanjut dipilih. Atribut yang digunakan dalam penelitian ini terdiri dari Nama, Jenis Kelamin, TB (Tinggi Badan), BB (Berat Badan), Umur dan Alamat, sebagaimana ditunjukkan pada tabel berikut:

Tabel 2 Tampilan Atribut yang Dipilih

No	Nama	JK	Umur	TB	BB	Alamat/ desa	Status
1.	Misbahul Munir	L	2th, 8bln, 3hr	87	11.0	Desa Sukodono	Gizi Baik
2.	Zalifatul Mukarromah	P	5 thn, 0bln, 5hr	101,4	15.0	Alassumur	Gizi Baik
3.	Jihan Aisyah	P	4thn, 11bln, 12 hr	102,5	13.6	Alassumur	Gizi Baik

Tabel Lanjutan

No	Nama	JK	Umur	TB	BB	Alamat/ desa	Status
4.	Mirza	L	4thn, 11bln, 10 hr	101,4	13.7	Alassumur	Gizi Baik
5.	M. Stabit	L	4thn, 10bln, 2 hr	102,5	14.0	Alassumur	Gizi Baik
6.	Kafiya inta bela	P	4thn, 7bln, 17 hari	101,4	13.7	Alassumur	KurangBaik
7.	Balyan	L	4thn, 5bln, 13 hari	97	12.2	Alassumur	Gizi Baik
8.	Dira	P	4thn, 3bln, 26 hari	98,7	13.2	Alassumur	Gizi Baik
9.	Dinara	P	4thn, 1bln, 4 hari	98	13	Alassumur	Gizi Baik
10.	Keenan	L	4thn, 0bln,6 hari	97,5	13.5	Alassumur	KurangBaik
11.	Viona	P	3thn, 10bln, 15 hari	97,5	12.2	Alassumur	Gizi Baik
12.	Raisa	P	3thn, 10bln, 14 hari	96,5	13.4	Alassumur	Gizi Baik
13.	Ainun	P	3thn, 10bln, 15 hari	94	13.6	Alassumur	Gizi Baik
14.	Kafil Mubarak	L	3thn, 9bln, 11 hari	98,6	13.5	Alassumur	Gizi Baik

Seleksi atribut dilakukan dengan menggunakan aplikasi *Visual Studio Code(VSC)* dengan menggunakan kode berikut:

Segmen Program 3 kode Seleksi Atribut

```
print("\n=== DATA AWAL ===")
print(
    df[
        ['Nama', 'JK', 'Usia Saat Ukur',
'Tinggi', 'Berat', 'Alamat', 'TB/U']
    ].head(15)
)
```

Kode di atas berfokus pada pemilihan atribut dari DataFrame data pengukuran gizi balita untuk analisis lebih lanjut. Atribut-atribut yang digunakan yaitu nama, jenis_kelamin, umur, tb, bb, Alamat dan status Gizi. Selanjutnya, data dibersihkan dengan menghapus baris yang memiliki nilai kosong (*missing values*) pada atribut-atribut tersebut. Setelah itu, DataFrame dipilih sehingga hanya mencakup atribut yang telah dipilih dan disimpan dalam variabel data. Tahap akhir adalah menampilkan beberapa baris pertama dari dataset untuk memberikan gambaran awal terhadap data yang sudah diseleksi. Pada penelitian ini, atribut yang diubah adalah: Jenis Kelamin dan Status Gizi. berikut Aturan Transformasi:

- Jenis Kelamin:**
Laki-laki (L) = 1
Perempuan (P) = 2
- Status Gizi:**

Normal = 1
 Stunting = 2

Tabel 4 Perubahan Atribut Jenis Kelamin dan Status Gizi dari Nominal ke Numerik

No	Nama	JK	Umur	TB	BB	Status
1.	Misbahul Munir	1	32	87	11.0	1.0
2.	Zalifatul Mukarromah	2	60	101.4	15.0	1.0
3.	Jihan Aisyah	2	59	102.5	13.6	1.0
4.	Mirza	1	59	101.4	13.7	1.0
5.	M. Stabit	1	58	102.5	14.0	1.0
6.	Kafiya inta bela	2	55	101.4	13.7	1.0
7.	Balyan	1	53	97.0	12.2	2.0
8.	Dira	2	51	98.7	13.2	1.0
9.	Dinara	2	51	98.0	13.0	1.0
10.	Keenan	1	49	97.5	13.5	1.0
11.	Viona	2	48	97.5	12.2	1.0
12.	Raisa	2	46	96.5	13.4	1.0
13.	Ainun	2	46	94.0	13.6	1.0
14.	Kafil Mubarak	1	46	98.6	13.5	1.0
15.	Aluna Farzana	2	45	94.5	13.0	1.0

Normalisasi data dilakukan untuk menyamakan nilai antar atribut agar tidak ada nilai yang lebih dominan dalam proses klasifikasi. Pada penelitian ini, normalisasi dilakukan menggunakan metode *Min-Max Scaling* dengan rentang nilai 1–10. Rumus normalisasi yang digunakan adalah:

$$Norm = \frac{dvalue - X_{min}}{X_{MAX} - X_{min}} \times 9 + 1 \quad (3.0)$$

Berikut hasil dari Normalisasi menggunakan Excel:

Tabel 5 hasil Normalisasi excel

No	Norm JK	Norm Umur	Norm TB	Norm BB
1.	1	1.00	1.00	1.00
2.	2	10.00	9.36	10.00
3.	2	9.70	10.00	6.85
4.	1	9.70	9.36	7.08
5.	1	9.40	10.00	7.75
6.	2	8.40	9.36	7.08
7.	1	7.70	3.32	3.95
8.	2	7.00	4.47	6.18
9.	2	7.00	4.32	5.62
10.	1	6.30	3.63	6.81
11.	2	6.00	3.63	3.95
12.	2	5.40	3.05	6.73
13.	2	5.40	1.45	6.85
14.	1	5.40	4.681	6.81
15.	2	5.10	1.77	5.62

Sedangkan dalam *Visual Studio Code* (VSC) dengan mengimplementasikan kode untuk melakukan proses normalisasi data menggunakan metode *Min-Max Scaling* dengan rentang nilai 1 sampai 10. Berikut tampilan dari hasil Normalisasi dengan *Visual Studio Code*:

Tabel 6 Hasil dari Kode Hasil Normalisasi

	JK	Norm-Umur	Norm-TB	Norm-BB	Status
0.	1	5.80	6.263158	5.135135	1
1.	2	10.00	8.157895	7.081081	1
2.	2	9.85	8.302632	6.400000	1
3.	1	9.85	8.289474	6.448649	1
4.	1	9.79	8.302632	6.594595	1
5.	2	9.25	8.157895	6.448649	1
6.	1	8.95	7.578947	5.718919	2
7.	2	8.65	7.802632	6.205405	1
8.	2	8.65	7.710526	6.108108	1
9.	1	8.35	7.644737	6.351351	1
10.	2	8.20	7.644737	5.718919	1
11.	2	7.90	7.513158	6.302703	1
12.	2	7.90	7.184211	6.400000	1
13.	1	7.90	7.789474	6.351351	1
14.	2	7.75	7.250000	6.108108	1

Dataset dalam penelitian ini terbagi menjadi dua kategori, yaitu data pelatihan dan data pengujian, menggunakan fungsi *train_test_split* yang disediakan oleh pustaka *Scikit-Learn*. Fitur atau variabel independen (x) dan variabel dependen atau target (y) dipisahkan dengan proporsi 80:20. Penyesuaian ini diterapkan melalui parameter *test_size=0.2*, sehingga 80% dari keseluruhan data digunakan oleh model untuk mengenali pola selama proses pembelajaran, sedangkan 20% yang tersisa disimpan untuk menguji kemampuan model dalam kinerjanya. Selain itu, parameter *random_state=0* diterapkan sebagai benih acak untuk menjamin bahwa hasil pembagian data tetap sama setiap kali program dijalankan. Pada tahap akhir, fungsi *shape* dipanggil untuk memeriksa dimensi pada variabel X_{train} dan X_{test} . Proses ini bertujuan untuk memastikan dengan baik bahwa jumlah baris data di setiap kategori sudah sesuai dengan rencana pembagian yang telah ditentukan. Berikut segmen program dalam menentukan pembagian data training dan data testing.

Segmen Program 6 Pembagian Data *Training* dan Data *Testing*

```
#split
X_train, X_test, y_train, y_test
= train_test_split(
    X,
    y,
    test_size=0.2,
    random_state=0
)
print("Jumlah Data Training:",
X_train.shape)
print("Jumlah Data Testing:",
X_test.shape)
```

Proses penilaian hasil prediksi dalam studi ini tidak hanya bergantung pada label kategori keluaran, tetapi juga mencakup analisis nilai probabilitas yang diciptakan oleh model tersebut. Berikut segmen program untuk menghitung probabilitas, Dalam pelaksanaannya, setiap baris data dari fitur yang telah dinormalkan, seperti Jenis Kelamin (JK), Usia (Norm_Umur), Tinggi (Norm_TB) dan Berat (Norm_BB), dipetakan untuk menganalisis distribusi peluang di masing-masing kelas sasaran. Penentuan label akhir dilakukan dengan memilih nilai probabilitas tertinggi (*maximum a posteriori*) dari distribusi yang ada. Hasil perhitungan ini kemudian dipadukan ke dalam sebuah *dataframe* yang terstruktur demi memudahkan proses pemastian antara hasil prediksi sistem dengan kondisi nyata di lapangan, sehingga tingkat ketidakpastian model dalam mengklasifikasikan data tertentu dapat diungkapkan dengan jelas.

Setelah melalui tahap prediksi dan perhitungan probabilitas menggunakan model Naïve Bayes, langkah selanjutnya adalah menganalisis secara mendalam kinerja klasifikasi dengan menggunakan *Confusion Matrix*. *Confusion Matrix* adalah sebuah alat evaluasi yang memperlihatkan tabel yang membandingkan antara kelas sebenarnya (kondisi yang benar) dengan kelas yang diprediksi oleh model. Dengan memanfaatkan matriks ini, kita dapat tidak hanya mengetahui tingkat akurasi secara keseluruhan, tetapi juga dapat mengidentifikasi jenis kesalahan yang dibuat oleh model, seperti jumlah data yang salah dikategorikan ke dalam klasifikasi yang berbeda. Berikut adalah segmen program yang digunakan untuk menampilkan dan memvisualisasikan *Confusion Matrix* agar lebih mudah dipahami. berikut hasil akurasi Confussion Matrix, ditunjukkan oleh gambar 1:

Kelas Sebenarnya	Normal	408	0	0
	Pendek	27	0	0
	Stunting	6	0	0
		Normal	Pendek	Stunting
		Kelas Prediksi		

Gambar 1 Confusion Matrix

Pada hasil confusion matrix menunjukkan jumlah prediksi Normal 408, pendek 27 dan stunting 6 dengan total 441 data uji secara keseluruhan. Tabel 7 menunjukkan hasil *confusion matrix* menggunakan naïve bayes:

Tabel 7 Hasil Confussion Matrix Naïve bayes

Prediksi	Normal	Pendek	Stunting
Normal	408	0	0
Pendek	27	0	0
Stunting	6	0	0

KESIMPULAN

Berdasarkan keseluruhan proses analisis, perancangan, pelaksanaan, dan pengujian model yang telah dilakukan dalam penelitian ini, dapat dikatakan bahwa algoritma *Naive Bayes* dengan

baik dapat diterapkan dalam sistem yang berbasis teknologi informasi untuk mendeteksi tingkat risiko stunting pada balita di UPTD Puskesmas Pujer. Melalui serangkaian tahapan penambangan data yang sistematis, mulai dari pembersihan data, normalisasi fitur numerik menggunakan metode *Min-Max scaling*, dan perhitungan probabilitas, pengelolaan informasi kesehatan anak dapat dilaksanakan dengan lebih teratur. Selain itu, evaluasi yang dilakukan menggunakan alat ukur *Confusion Matrix* menunjukkan bahwa penerapan metode *Gaussian Naive Bayes* sangat efektif dalam mengidentifikasi status risiko gizi anak karena dapat memberikan prediksi yang tepat, sehingga mengurangi keterlambatan informasi dan kesalahan yang serius yang sering terjadi pada sistem pencatatan konvensional.

DAFTAR PUSTAKA

- A. J. Wahidin and T. H. Andika, “Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning,” pp. 378–387, 2024, doi: 10.33364/algoritma/v.21-2.2122.
- A. P. Yuda, Z. Septina, A. Maharani, and Y. Nurdiatami, “Tinjauan Literatur : Perkembangan Program Penanggulangan Stunting di Indonesia,” vol. 6, no. 2, pp. 53–58, 2022.
- D. K. K. Bondowoso, “Hasil Survei Status Gizi Indonesia (SSGI) 2024 Kabupaten Bondowoso.” [Online]. Available: <https://dinkes.bondowosokab.go.id/berita/hasil-survei-status-gizi-indonesia-ssgi-2024-kabupaten-bondowoso>
- Oslida Martony, “No Title,” vol. 5, pp. 1734–1745, 2023.
- Y. Mulyanto, F. Idifitriani, A. Wati, U. T. Sumbawa, D. Mining, and K. P. Tano, “Vol 7 No 2 , September 2024 KLASIFIKASI DATA MINING UNTUK PENENTUAN STUNTING,” vol. 7, no. 2, pp. 129–135, 2024.
- A. Rahmawati *et al.*, “Optimalisasi Teknik Wawancara Dalam Penelitian Field Research Melalui Pelatihan Berbasis Participatory Action Research Pada Mahasiswa Lapas Pemuda Kelas IIA Tangerang,” pp. 135–142.
- M. P. Hasibuan, R. Azmi, D. B. Arjuna, S. U. Rahayu, U. Islam, and N. Sumatera, “Analisis Pengukuran Temperatur Udara Dengan Metode Observasi,” vol. 1, 2023.
- H. D. Chandra and . A., “Pengembangan sistem Informasi Absensi Berbasis Radio Frequency Identification (RFID) Terintegrasi dengan sistem Informasi Akademik,” *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 6, no. 1, pp. 76–86, 2023, doi: 10.47324/ilkominfo.v6i1.168.
- Ardiansyah, Risnita, and M. S. Jailani, “Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif,” *J. IHSAN J. Pendidik. Islam*, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.61104/ihsan.v1i2.57.