



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 3 (2025) pp: 2484-2490

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Klasifikasi Status Stunting Balita Tegal Menggunakan Teknik Smote Pada Metode *Naives Bayes Gaussain*

Sigit Januarto¹, Aang Alim Murtopo², Zaenul Arif³

¹Sistem Informasi, STMIK YMI Tegal

^{2,3}Teknik Informatika, STMIK YMI Tegal

¹sigitjanuarto8@gmail.com, ²aang.alim@stmik-tegal.ac.id, ³zendhunter@gmail.com

Abstrak

Stunting merupakan salah satu masalah kesehatan serius yang memengaruhi tumbuh kembang anak, terutama pada 1.000 Hari Pertama Kehidupan. Kota Tegal termasuk wilayah dengan prevalensi stunting yang cukup tinggi, sehingga diperlukan metode prediksi yang akurat untuk mendukung intervensi gizi tepat sasaran. Penelitian ini menggunakan metode Naive Bayes Gaussian untuk mengklasifikasikan status stunting balita berdasarkan data antropometri. Permasalahan ketidakseimbangan kelas pada dataset diatasi dengan teknik oversampling Synthetic Minority Over-sampling Technique (SMOTE) guna meningkatkan kemampuan model dalam mengenali kelas minoritas. Hasil pengujian menunjukkan bahwa model sebelum penerapan SMOTE memiliki akurasi rata-rata 91,58%. Setelah penerapan SMOTE, akurasi validasi silang meningkat menjadi rata-rata 96,28% dengan presisi 94,03%, recall 91,58%, dan F1-score 92,12%. Peningkatan ini membuktikan bahwa kombinasi Naive Bayes Gaussian dan SMOTE efektif untuk prediksi status stunting. Model yang dihasilkan berpotensi diimplementasikan sebagai sistem pendukung keputusan dalam deteksi dini dan pencegahan stunting di wilayah rawan.

Kata kunci: Klasifikasi, Machine Learning, Naive Bayes Gaussian, Stunting, SMOTE

1. Latar Belakang

Stunting atau gangguan pertumbuhan anak merupakan salah satu masalah kesehatan serius di Indonesia. Kondisi ini terjadi akibat kekurangan gizi kronis yang dialami anak, terutama pada 1.000 Hari Pertama Kehidupan (HPK), yaitu sejak masa kehamilan hingga anak berusia dua tahun [1]. Dampak stunting tidak hanya terlihat dari tinggi badan anak yang lebih rendah dari rata-rata, tetapi juga dapat memengaruhi perkembangan kognitif, daya tahan tubuh, dan produktivitas jangka panjang [2]. Tingginya angka stunting dapat menghambat pembangunan sumber daya manusia dan memperlambat pertumbuhan ekonomi negara [3]. Oleh sebab itu, pemerintah Indonesia menjadikan penurunan stunting sebagai salah satu prioritas nasional dengan berbagai intervensi, seperti peningkatan layanan kesehatan ibu dan anak, perbaikan sanitasi dan penyediaan air bersih, serta edukasi gizi kepada masyarakat [4]. Beragam intervensi telah dilakukan, seperti meningkatkan layanan kesehatan bagi ibu dan anak, memperbaiki sanitasi dan penyediaan air bersih, serta memberikan edukasi gizi kepada masyarakat untuk mengatasi dampak multidimensi stunting [5].

Meskipun upaya penanggulangan stunting terus dilakukan, permasalahan ini masih menjadi tantangan di berbagai daerah, termasuk di Kota Tegal. Data Dinas Kesehatan Kota Tegal menunjukkan bahwa prevalensi stunting masih berada di angka 20,7%, yang menandakan kondisi ini perlu mendapat perhatian khusus Faktor-faktor seperti kondisi sosial ekonomi keluarga, akses air bersih dan sanitasi layak, serta pemberian makan bayi dan anak (PMBA) yang kurang optimal, berkontribusi terhadap tingginya angka stunting [6]. Oleh karena itu, pemantauan dan identifikasi dini kasus stunting sangat penting untuk mencegah dampak jangka panjang pada tumbuh kembang anak.

Maka dari itu, metode yang efektif dan efisien sangat dibutuhkan untuk mengidentifikasi balita yang berisiko stunting. Dalam beberapa tahun terakhir, penggunaan *machine learning* di bidang kesehatan terus berkembang pesat, terutama untuk identifikasi dan prediksi risiko kesehatan. Salah satu algoritma yang populer dan efektif untuk tugas klasifikasi adalah *Naive Bayes*. Algoritma ini berdasarkan pada *Teorema Bayes* diasumsikan bahwa seluruh fitur memiliki sifat independen terhadap fitur lainnya. *Naive Bayes Gaussian*, khususnya, menggunakan asumsi bahwa fitur numerik mengikuti distribusi normal, sehingga sangat cocok untuk data kontinu seperti data antropometri. Algoritma ini memiliki keunggulan dalam kesederhanaan, kecepatan komputasi, serta kemampuan menghasilkan performa yang baik, bahkan dengan dataset yang relatif kecil.

Tantangan utama dalam klasifikasi data stunting adalah ketidakseimbangan kelas, di mana jumlah balita yang mengalami stunting lebih banyak dibandingkan kategori lain seperti *severely stunted*. Ketidakseimbangan ini dapat menyebabkan model bias terhadap kelas mayoritas dan mengurangi akurasi pada kelas minoritas. Untuk mengatasi hal ini, digunakan *Synthetic Minority Over-sampling Technique* (SMOTE), yaitu teknik *oversampling* yang menghasilkan sampel sintetis pada kelas minoritas melalui interpolasi antar data yang ada [7].

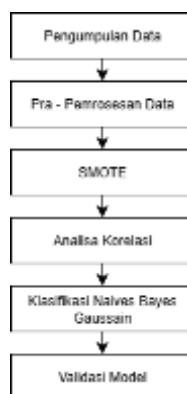
Penelitian oleh [8] menggunakan metode *Naive Bayes* untuk mengklasifikasikan status stunting pada balita di Kota Madiun berbasis website. Model diuji menggunakan pembagian data 70% untuk pelatihan dan 30% untuk pengujian, menghasilkan akurasi rata-rata 58%, presisi 68%, dan *recall* 58% berdasarkan *confusion matrix*. Penelitian ini bertujuan mengembangkan sistem klasifikasi berbasis web guna membantu tenaga kesehatan dalam mengidentifikasi status stunting balita. Penelitian oleh [9] menerapkan algoritma *K-Means* untuk mengelompokkan balita di Desa Tegalwangi menjadi dua *cluster* berdasarkan usia, tinggi badan, dan berat badan. Evaluasi model menggunakan *indeks Davies-Bouldin* (DBI) menghasilkan nilai optimal 0,007 pada *cluster* terbaik. Penelitian ini bertujuan menyediakan informasi status balita, yaitu normal atau stunting, untuk mendukung petugas Posyandu dalam pengolahan data dan intervensi kesehatan. Penelitian oleh [10] mengimplementasikan algoritma *Naive Bayes* dan *K-Nearest Neighbors* (K-NN) untuk mendiagnosis status stunting anak berdasarkan parameter usia, berat badan, dan tinggi badan. Hasilnya, *Naive Bayes* mencapai akurasi 71% dengan F1-score 73%. Penelitian ini bertujuan membandingkan efektivitas kedua algoritma dalam membantu petugas kesehatan melakukan diagnosis stunting dengan cepat dan akurat.

Penelitian ini bertujuan untuk mengembangkan model prediksi status stunting balita di Kota Tegal dengan mengombinasikan algoritma *Naive Bayes Gaussian* dan SMOTE pada data antropometri. Model ini dievaluasi berdasarkan akurasi, presisi, *recall*, F1-score, AUC-ROC, dan *cross-validation*. Hasil penelitian ini dapat menjadi dasar pengembangan sistem pendukung keputusan di bidang kesehatan yang lebih akurat dan responsif, khususnya dalam program deteksi dini dan pencegahan stunting di wilayah rawan.

2. Metode Penelitian

2.1. Desain Penelitian

Penelitian menggunakan pendekatan kuantitatif dengan metode klasifikasi prediktif untuk menganalisis data historis stunting balita. Model dikembangkan dengan mengombinasikan SMOTE (*Synthetic Minority Over-sampling Technique*) pada algoritma *Naive Bayes Gaussian* untuk mengatasi ketidakseimbangan kelas dan meningkatkan akurasi prediksi status stunting. Pendekatan ini dipilih karena kemampuan *Naive Bayes Gaussian* dalam menangani data kontinu serta keunggulan SMOTE dalam menyeimbangkan distribusi kelas minoritas [11].



Gambar 1. Proses Analisa Data

2.2. Pengumpulan Data

Data yang digunakan merupakan data sekunder dalam penelitian ini bersumber dari laman resmi Portal Jateng, khususnya dataset stunting Kota Tegal tahun 2024. Dataset tersebut berisi informasi mengenai balita yang mengalami stunting, mencakup berbagai pengukuran antropometri dan klasifikasi status gizi. Data ini memberikan gambaran mengenai prevalensi stunting di Kota Tegal pada tahun 2024.

No	Nama	JK	Usia Saat Ukur	Berat	Tinggi	LiLA	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB
1.	ATIKA AZZAHRAH	F	1 Tahun - 8 Bulan - 26 Hari	8.5	75.3	0	Kurang	-2.00	Pendek	2.83	Gizi Baik
2.	ARSIAD BIL FAQIH	L	3 Tahun - 0 Bulan - 30 Hari	10.1	85.8	0	Sangat Kurang	-3.00	Pendek	2.97	Gizi Baik
3.	SARA PEATAMA	L	1 Tahun - 0 Bulan - 8 Hari	5.5	70	0	Sangat Kurang	-3.82	Pendek	2.55	Gizi Buruk
4.	ASYAH SHIDDIQIYAH	F	3 Tahun - 9 Bulan - 5 Hari	9.9	90	0	Sangat Kurang	-3.51	Pendek	2.63	Gizi Kurang
	:		:	:	:	:	:	:	:	:	:
295.	AYRA KAMELA	F	1 Tahun - 8 Bulan - 19 Hari	8.9	75	14	Berat Badan Normal	-1.26	Pendek	2.16	Gizi Baik
296.	TASQHA NURHA KAMELA	F	4 Tahun - 3 Bulan - 13 Hari	10.5	91	15	Sangat Kurang	-3.48	Pendek	3.09	Gizi Kurang
297.	M SYAFI ALFARUQ	L	1 Tahun - 2 Bulan - 18 Hari	6.4	73	16	Berat Badan Normal	-1.78	Pendek	2.28	Gizi Baik
298.	HUMAIRA ALIFA RAHMA	F	2 Tahun - 11 Bulan - 0 Hari	11.1	85	0	Berat Badan Normal	-1.67	Pendek	2.49	Gizi Baik

Gambar 2. Data Balita Stunting Kota Tegal 2024

Gambar 2. merupakan data penelitian ini merupakan data stunting Kota Tegal yang memenuhi kriteria kelengkapan informasi pada atribut utama, yaitu usia balita, berat badan (BB), tinggi badan (TB), lingkaran lengan atas (LiLA), serta nilai *Z-Score* untuk BB/U, TB/U, dan BB/TB.

2.3. Pra-Pemrosesan Data

Proses pra-pemrosesan data diawali dengan langkah *cleaning*, meliputi penghapusan data duplikat untuk menghindari bias dan penanganan *missing values* menggunakan metode imputasi rata-rata pada data numerik. Selain itu, variabel kategorikal seperti Status dikonversi ke bentuk numerik (misal: 0 untuk stunting dan 1 untuk *severely stunted*) agar dapat diproses lebih lanjut. Selanjutnya, dilakukan transformasi data dengan melakukan normalisasi pada variabel numerik seperti Berat, Tinggi, dan Umur_Bulan menggunakan *StandardScaler* untuk menyamakan skala data sehingga memperoleh distribusi yang lebih stabil dan memudahkan analisis statistik maupun pemodelan.

2.4. Implementasi SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi ketidakseimbangan kelas dalam dataset stunting, di mana jumlah sampel kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. Dengan menghasilkan sampel sintesis berbasis fitur asli kelas minoritas, SMOTE menyeimbangkan distribusi data tanpa sekadar menduplikasi observasi, sehingga meningkatkan kemampuan model dalam mengenali pola kelas yang awalnya terabaikan. Pendekatan ini membantu mencegah bias prediksi yang menguntungkan kelas mayoritas sekaligus mempertahankan integritas data asli [12].

$$x_{new} = x_1 + \lambda \cdot (x_2 - x_1) \quad (1)$$

Di Mana x_1 adalah Sampel asli kelas minoritas, x_2 adalah Tetangga terdekat x_1 dari kelas minoritas, dan λ adalah ilangan acak untuk interpolasi.

2.5. Analisa Korelasi

Pada tahap ini digunakan untuk melihat keterkaitan antar variabel dalam dataset. Visualisasi ini membantu mengamati pola hubungan, misalnya apakah terdapat kaitan antara tinggi badan dan usia dalam bulan. Selain itu, analisis ini bertujuan untuk memahami struktur data sebelum menentukan variabel mana yang akan digunakan dalam pemodelan lebih lanjut, tanpa melakukan pengujian statistik terlebih dahulu. Fokusnya adalah pada observasi awal terhadap hubungan alami antar fitur dalam data stunting.

2.6. Implementasi Naives Bayes Gaussain

Metode ini digunakan untuk mengklasifikasikan data dengan memanfaatkan probabilitas dari setiap kelas berdasarkan nilai atribut yang dimiliki, sehingga dapat memprediksi kelas suatu sampel baru secara cepat dan efisien. Dengan mengasumsikan independensi antar atribut dan memodelkan distribusinya menggunakan fungsi *Gaussian*, metode ini mampu mengolah data kuantitatif dengan baik serta tetap memberikan hasil yang kompetitif meskipun data tidak sepenuhnya mengikuti distribusi normal. Pendekatan ini sering diterapkan pada berbagai permasalahan klasifikasi, terutama ketika jumlah fitur cukup besar atau ketika diperlukan perhitungan probabilitas yang transparan untuk interpretasi hasil [13].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Dimana, X adalah vektor fitur yang merepresentasikan data, H adalah kelas yang diprediksi. $P(H)$ adalah menunjukkan prior probabilitas kelas, $P(X|H)$ adalah *likelihood* atau peluang munculnya X jika berasal dari H , $P(X)$ adalah *evidence* atau probabilitas keseluruhan X , dan $P(H|X)$ merupakan posterior probabilitas kelas setelah mempertimbangkan X

2.7. Metrik Evaluasi

Model *Naive Bayes Gaussian* adalah salah satu algoritma klasifikasi yang mengasumsikan bahwa setiap fitur dalam dataset mengikuti distribusi normal (*Gaussian*). Untuk menilai performa model ini, beberapa metrik evaluasi digunakan, seperti akurasi, *recall*, *presisi*, F1-Score, serta kurva ROC dan AUC (*Area Under Curve*).

Akurasi untuk mengukur persentase prediksi benar dibandingkan dengan total prediksi [14]. Rumusnya ditunjukkan dalam Persamaan (3):

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (3)$$

Di mana *Number of Correct Predictions* adalah jumlah data yang diprediksi dengan benar, sedangkan *Total Number of Predictions* adalah seluruh data yang diuji.

Selain akurasi, *presisi* dan *recall* juga penting, terutama ketika dataset tidak seimbang. *Presisi* Persamaan (4) untuk menilai akurat model dalam memprediksi kelas positif, sementara *recall* Persamaan (5) mengevaluasi positif yang berhasil diidentifikasi model

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP+Fn)} \quad (5)$$

TP (*True Positive*) adalah Jumlah sampel positif yang diprediksi benar, FP (*False Positive*) adalah Jumlah sampel negatif yang salah diprediksi sebagai positif, FN (*False Negative*) adalah Jumlah sampel positif yang salah diprediksi sebagai negatif.

Karena *presisi* dan *recall* sering kali bertolak belakang, F1-Score Persamaan (6) digunakan sebagai metrik yang menyeimbangkan keduanya:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Selain itu, kurva ROC dan AUC memberikan gambaran kinerja model dalam membedakan kelas pada berbagai ambang batas klasifikasi [15]. Dengan mempertimbangkan berbagai metrik ini dapat memahami kelebihan dan kelemahan model *Naive Bayes Gaussian* secara lebih komprehensif dalam berbagai konteks aplikasi.

3. Hasil dan Diskusi

3.1. Pra Pemrosesan Data

Pada tahap pra-pemrosesan, data stunting dipersiapkan untuk memastikan kualitas dan konsistensi. Langkahnya meliputi penghapusan variabel tidak relevan, *imputasi median* untuk nilai hilang, dan pemeriksaan format data. Ketidakseimbangan kelas diatasi dengan SMOTE untuk menghasilkan data sintetik pada kelas minoritas, kemudian status di numerik agar sesuai dengan asumsi *Gaussian* pada *Naive Bayes*.

Table 1. Data Setelah Pra Pemrosesan

Nama	ID	Umur_Bulan	Berat	Tinggi	TB/U	TB/U	Status	Keterangan
Atika Azzahrah	1	20	8.5	75.5	Pendek	-2.63	0	Stunting
Arsyad Bil Faqih	2	37	10.1	85.6	Pendek	-2.97	0	Stunting
Raka Pratama	3	12	6.3	70	Pendek	-2.55	0	Stunting
Aisyah Shidqiyah	4	45	9.9	90	Pendek	-2.63	0	Stunting
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Naura Elshanum	5	24	7.6	77.7	Pendek	-2.65	0	Stunting
Mutiara Azkia	7	30	8.7	76	Sangat Pendek	-4.31	1	Severely stunted

AMEER AL FARIZI	8	30	10.1	82.4	Pendek	-2.88	0	Stunting
-----------------	---	----	------	------	--------	-------	---	----------

3.2. Imbalance Class

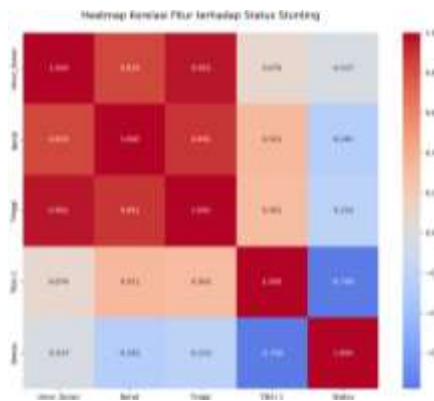
Pada tahap ini digunakan teknik *oversampling* SMOTE untuk menyeimbangkan distribusi kelas pada data latih. Metode ini menciptakan sampel sintetis dari kelas minoritas melalui *interpolasi linier*, sehingga efektif mengurangi ketidakseimbangan. Tabel berikut menampilkan perbandingan jumlah data latih sebelum dan sesudah *oversampling*.

Table 2. Jumlah Data Latih Sebelum dan Sesudah *Oversampling*

Jenis Data	Kelas 0	Kelas 1
Data Sebelum SMOTE	245	56
Data Sesudah SMOTE	245	245

3.3. Analisa Korelasi

Analisis korelasi dilakukan untuk mengidentifikasi hubungan antar fitur numerik dalam dataset, seperti Umur_Bulan, Berat, Tinggi, TB/U.1, dan IMT. Tahap ini membantu memahami keterkaitan antar variabel yang dapat mempengaruhi hasil pemodelan.



Gambar 3. Heatmap Korelasi

Gambar 3. merupakan korelasi di atas menunjukkan adanya hubungan positif yang sangat kuat antara Umur_Bulan dan Tinggi (0,96), serta antara Berat dan Tinggi (0,89). Hal ini menggambarkan bahwa semakin bertambah usia seorang anak, tinggi dan berat badannya juga cenderung meningkat. Meskipun *Naive Bayes Gaussian* mengasumsikan independensi antar fitur, hasil ini memperlihatkan adanya korelasi di beberapa variabel. Namun, algoritma tersebut tetap efektif digunakan dan fitur Umur_Bulan, Berat, serta Tinggi terbukti relevan untuk klasifikasi status stunting.

3.5. Evaluasi Model

Sebagai langkah akhir dalam pengembangan model klasifikasi, dilakukan validasi menyeluruh untuk memastikan keakuratan dan keandalan model *Naive Bayes Gaussian* yang telah menerapkan teknik SMOTE. Evaluasi performa dilakukan menggunakan berbagai metrik statistik dan validasi silang untuk menguji konsistensi model.

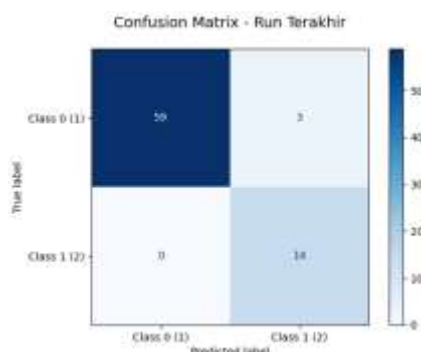
Table 3. Hasil *Multiple Testing*

Run	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
1	92.11	94.47	92.11	92.60
2	85.53	91.89	85.53	86.86
3	92.11	93.48	92.11	92.47
4	89.47	93.20	89.47	90.28
5	96.05	96.21	96.05	96.10
6	86.84	90.95	86.84	87.85
7	93.42	94.28	93.42	93.66
8	92.11	94.47	92.11	92.60
9	92.11	94.47	92.11	92.60
10	96.05	96.75	96.05	96.19

Table 4. Hasil Rata - Rata Evaluasi

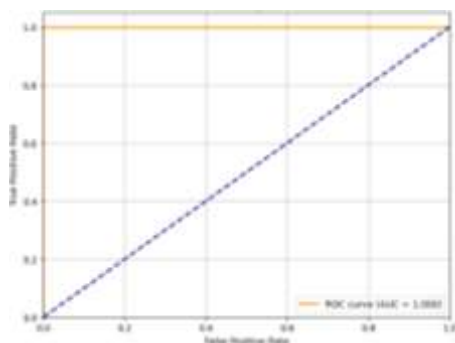
Rata- Rata Evaluasi Model	Hasil (%)
Akurasi	91.58
Presisi	94.03
Recall	91.58
F1-Score	92.12

Berdasarkan Tabel 3. yang menampilkan hasil *multiple testing* selama 10 iterasi, model menunjukkan performa yang cukup stabil dengan akurasi berkisar antara 85.53% hingga 96.05%, presisi 90.95% hingga 96.75%, *recall* 85.53% hingga 96.05%, dan F1-score 86.86% hingga 96.19% Nilai-nilai ini mengindikasikan konsistensi model dalam mengklasifikasikan data. Sementara itu, Tabel 4. merangkum rata-rata hasil evaluasi akurasi 91.58%, presisi 94.03%, *recall* 91.58%, dan F1-score 91.12%, yang mengonfirmasi bahwa model memiliki kemampuan prediktif yang baik tanpa *overfitting*, Hasil ini memperkuat kesimpulan bahwa model *Naive Bayes Gaussian* dengan SMOTE dapat diandalkan untuk prediksi kelas stunting.



Gambar 4. Hasil *Confusion Matrix*

Gambar 5. *Confusion matrix* menunjukkan model klasifikasi status stunting dengan label 0 sebagai stunting dan label 1 sebagai *severely stunted* menunjukkan kinerja yang sangat baik. Dari total 76 data uji, sebanyak 59 data stunting berhasil diprediksi dengan benar dan hanya 3 kasus yang salah diprediksi sebagai *severely stunted*, sedangkan seluruh 14 data *severely stunted* berhasil diklasifikasikan dengan benar tanpa kesalahan prediksi kelas. Performa ini mengindikasikan bahwa kombinasi *Naive Bayes Gaussian* dengan SMOTE mampu menghasilkan model prediksi yang akurat, seimbang, dan efektif dalam membedakan kedua kategori status stunting.



Gambar 5. Kurva ROC

Gambar 6. menunjukkan kurva ROC menggunakan model *Naive Bayes Gaussian* setelah penerapan SMOTE untuk menyeimbangkan data. Kurva ROC ini memperlihatkan hubungan antara *true positive rate* (TPR) dan *false positive rate* (FPR), dengan garis ROC berwarna oranye yang menempel pada sisi kiri dan atas grafik, serta menghasilkan nilai *area under the curve* (AUC) sebesar 1.000. Nilai AUC tersebut menunjukkan bahwa model mampu melakukan klasifikasi dengan sangat baik, tanpa kesalahan dalam membedakan antara dua kelas. Hal ini mencerminkan bahwa model memiliki daya diskriminatif yang sangat tinggi serta efektivitas SMOTE dalam meningkatkan kinerja klasifikasi. Evaluasi dilakukan menggunakan metode *5-fold cross-validation* untuk memastikan stabilitas dan kemampuan generalisasi model terhadap data yang berbeda.

Table 5. Hasil 5-Fold Cross Validation

Fold	Akurasi
1	98.68%
2	96.05%
3	93.33%
4	94.67%
5	98.67%
Rata - Rata	96.28%

Hasil validasi silang menunjukkan performa model yang sangat konsisten, akurasi yang sangat tinggi dan konsisten di setiap *fold*, yaitu sebesar 93.33% hingga 98.68%. Rata-rata akurasi keseluruhan mencapai 96.28%, yang menunjukkan bahwa model memiliki performa klasifikasi yang sangat baik dan stabil di berbagai subset data. Hasil ini mencerminkan kemampuan generalisasi model yang tinggi terhadap data baru, serta memperkuat efektivitas metode SMOTE dalam mengatasi ketidakseimbangan kelas, yang berdampak positif pada peningkatan akurasi prediksi status stunting balita.

4. Kesimpulan

Penelitian ini menunjukkan bahwa model prediksi status stunting balita di Kota Tegal menggunakan algoritma *Naive Bayes Gaussian* yang dikombinasikan dengan teknik SMOTE memiliki performa yang sangat baik dan stabil. Model mampu mencapai akurasi rata-rata 91,58% dengan presisi 94,03%, recall 91,58%, dan F1-score 92,12%, sedangkan validasi silang menghasilkan akurasi rata-rata 96,28%, menandakan ketiadaan *overfitting* serta kemampuan generalisasi yang tinggi. Penerapan SMOTE terbukti efektif dalam mengatasi ketidakseimbangan kelas, sehingga model mampu mengenali seluruh kategori status gizi dengan lebih seimbang. Dengan kinerja yang konsisten dan tingkat akurasi yang tinggi, model ini berpotensi diimplementasikan sebagai sistem pendukung keputusan dalam pemantauan status gizi anak, memungkinkan intervensi gizi yang lebih tepat sasaran. Namun, untuk memastikan validitas eksternal dan memperluas penerapan di lapangan, disarankan penelitian selanjutnya menggunakan data yang lebih beragam, mencakup wilayah berbeda, dan menguji performa pada data real-time.

Referensi

- [1] U. A. Rachmawati et al., "SiCenting+: An Information System for Monitoring the Stunted Growth and Nutritional Status of Children in Pandeglang Regency," in 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), 2022, pp. 1–6. doi: 10.1109/ICATIECE56365.2022.10046915.
- [2] A. Ali, "Brain and Stunting: Integrative Clinical and Community Stunting Prevention Perspectives in Disorders of Brain Development," Int. J. Pap. Public Rev., vol. 4, no. 1, pp. 60–69, 2023, doi: 10.47667/ijppr.v4i1.216.
- [3] R. Maulina, M. B. Qomaruddin, B. Prasetyo, R. Indawati, and R. Alfitri, "The Effect of Stunting on the Cognitive Development in Children: A Systematic Review and Meta-analysis," Stud. Ethno-Medicine, vol. 17, no. 1–2, pp. 19–27, 2023, doi: 10.31901/24566772.2023/17.1-2.661.
- [4] A. Kustanto, O. Rachmat, and S. Setyadi, "The Prevalence of Stunting in Indonesia: An Examination of the Health, Socioeconomic Status, and Environmental Determinants," J. Iran. Med. Coun., vol. 8, no. 1, pp. 67–79, 2025, doi: 10.18502/jimc.v8i1.17062.
- [5] V. Y. Lameky, "Stunting in Indonesia: Current progress and future directions," J. Healthc. Adm., vol. 3, no. 1, pp. 82–90, 2024, doi: 10.33546/joha.3388.
- [6] M. Rosyida and N. R. Herawati, "The Analysis of Stunting Implementation Policy in Tegal Regency," vol. 1, no. 5, pp. 1–6, 2024.
- [7] A. M. A. Rahim, I. Y. R. Pratiwi, and M. A. Fikri, "Klasifikasi penyakit jantung menggunakan metode synthetic minority over-sampling technique dan random forest classifier," Indones. J. Comput. Sci., vol. 12, no. 5, 2023.
- [8] A. Rozaq and A. J. Purnomo, "Classification of Stunting Status in Toddlers Using Naive Bayes Method in the City of Madiun Based on Website," J. Techno Nusa Mandiri, vol. 19, no. 2, pp. 69–76, 2022, doi: 10.33480/techno.v19i2.3337.
- [9] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," Hello World J. Ilmu Komput., vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.
- [10] Wulan Widhari, Agung Triayudi, and Ratih Titi Komala Sari, "Implementation of Naive Bayes and K-NN Algorithms in Diagnosing Stunting in Children," SAGA J. Technol. Inf. Syst., vol. 2, no. 1, pp. 164–174, 2024, doi: 10.58905/saga.v2i1.242.
- [11] Hizbul Izzi, Arief Setyanto, and Anggit Dwi Hartanto, "Optimalisasi Akurasi Algoritma Naive Bayes Dengan Metode Syntetic Minority Oversampling Technique (Smote) Pada Data Numerik," Infotek J. Inform. dan Teknol., vol. 8, no. 1, pp. 217–227, 2025, doi: 10.29408/jit.v8i1.28340.
- [12] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 21, no. 3, pp. 677–690, 2022.
- [13] Y. Shi and H. Xu, "The Iris Cassification Based on Gaussian Naive Bayes Agorithm," in Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering, in EITCE '22. New York, NY, USA: Association for Computing Machinery, 2023, pp. 732–736. doi: 10.1145/3573428.3573559.
- [14] R. E. Putra, "Perbandingan Analisis Sentimen Untuk Prediksi Kepuasan Pelanggan Kedai Kopi Di Kofind Menggunakan Algoritma SVM Dan Naive Bayes," vol. 06, pp. 1039–1048, 2025.
- [15] R. Harahap, M. Irpan, M. A. Dinata, L. Efrizoni, and Rahmadden, "Perbandingan Algoritma Random Forest Dan Xgboost Untuk Klasifikasi Penyakit Paru-Paru Berdasarkan Data Demografi Pasien," J. Ilm. Betrik, vol. 15, no. 02, pp. 130–141, 2024.