

Optimasi Algoritma SVM dengan Teknik SMOTE dan Tuning Parameter pada Klasifikasi Balita Stunting

Muhammad Al Ghorizmi Muttaqin¹, Gustina Alfa Trisnapidika^{2,*}

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹111202214097@mhs.dinus.ac.id, ^{2,*}gustina.alfa@dsn.dinus.ac.id

Email Penulis Korespondensi: gustina.alfa@dsn.dinus.ac.id

Submitted: 04/09/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Stunting pada balita merupakan masalah gizi kronis yang berdampak jangka panjang terhadap kualitas sumber daya manusia, termasuk perkembangan kognitif dan kerentanan terhadap penyakit. Kabupaten Brebes menjadi salah satu daerah prioritas penanganan stunting di Indonesia. Penelitian ini bertujuan mengoptimalkan performa algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan status stunting balita dengan mengatasi ketidakseimbangan data menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE) dan tuning parameter. Data antropometri sebanyak 9598 sampel dari beberapa puskesmas di Brebes diproses melalui tahapan pembersihan, label encoding, penanganan outlier, standarisasi, dan pembagian kelas dan dibagi menjadi data *train* (80%) dan data *test* (20%). Dua model dibandingkan, yaitu model SVM dasar dan model SVM yang dioptimalkan dengan menggabungkan SMOTE serta tuning parameter melalui GridSearchCV. Hasil menunjukkan bahwa model dasar menghasilkan akurasi 98.31%, namun *recall* untuk kelas stunting hanya 89.19%. Setelah penerapan SMOTE dan tuning parameter, performa model meningkat dengan akurasi 99.78% dan *recall* kelas stunting naik menjadi 98.46%. Peningkatan tersebut membuktikan bahwa penggunaan SMOTE dan tuning parameter sangat efektif dalam meningkatkan sensitivitas model terhadap kelas minoritas. Dengan demikian, penelitian ini menunjukkan bahwa pendekatan optimasi yang komprehensif mampu mendukung deteksi dini stunting secara akurat, sehingga dapat digunakan dalam perencanaan intervensi kesehatan yang lebih tepat sasaran.

Kata Kunci: Stunting; SVM; SMOTE; Data Imbalance; Hyperparameter Tuning

Abstract—Stunting in toddlers is a chronic nutritional problem that has long-term impacts on human resource quality, including cognitive development and vulnerability to diseases. Brebes Regency is one of the priority areas for stunting management in Indonesia. This study aims to optimize the performance of the Support Vector Machine (SVM) algorithm in classifying stunting status among toddlers by addressing data imbalance using the Synthetic Minority Oversampling Technique (SMOTE) and parameter tuning. A total of 9,598 anthropometric samples collected from several community health center in Brebes were processed through stages of data cleaning, label encoding, outlier handling, standardization, and class splitting, and then divided into training data (80%) and testing data (20%). Two models were compared: the baseline SVM model and the optimized SVM model, which integrates SMOTE and parameter tuning through GridSearchCV. The results showed that the baseline model achieved an accuracy of 98.31%, but the recall for the stunting class was only 89.19%. After applying SMOTE and parameter tuning, the model's performance improved, achieving an accuracy of 99.78% and a recall for the stunting class of 98.46%. This improvement demonstrates that the use of SMOTE and parameter tuning is highly effective in enhancing the model's sensitivity toward the minority class. Therefore, this study shows that a comprehensive optimization approach can effectively support early detection of stunting, making it a valuable tool for more targeted health intervention planning.

Keywords: Stunting; SVM; SMOTE; Data Imbalance; Hyperparameter Tuning

1. PENDAHULUAN

Stunting merupakan masalah gizi kronis yang menjadi isu kesehatan masyarakat global yang mendesak, utamanya di negara-negara berkembang [1]. Lebih dari sekadar cerminan ketidakcukupan asupan makanan bergizi, stunting juga berasal pada permasalahan sanitasi lingkungan, praktik pola asuh yang kurang optimal, serta status kesehatan ibu dan anak yang saling terkait [2]. Konsekuensi stunting bersifat jangka panjang dan berdampak signifikan terhadap kualitas sumber daya manusia. Anak yang mengalami stunting memiliki risiko lebih tinggi mengalami gangguan perkembangan fungsi otak, yang dapat berujung pada penurunan kapasitas belajar dan selanjutnya memengaruhi produktivitas mereka saat mencapai usia dewasa [3]. Selain itu, terdapat korelasi antara riwayat stunting dengan peningkatan kerentanan terhadap penyakit tidak menular seperti diabetes melitus dan gangguan kardiovaskular di kemudian hari [4].

Di Indonesia sendiri, data yang diperoleh dari Survei Status Gizi Indonesia (SSGI) tahun 2022 mencatat angka prevalensi stunting sebesar 21,6% [5]. Prevalensi stunting nasional pada tahun 2024 yaitu 19,8% mengalami penurunan sebesar 1,8% dari tahun 2022 [6]. Masih perlu usaha untuk menguranginya setiap tahunnya, sehingga target (14,4%) pada tahun 2029 dapat terealisasi [7].

Secara umum, stunting terjadi akibat kombinasi faktor yang kompleks, termasuk asupan gizi yang tidak mencukupi dalam jangka panjang, seringnya mengalami infeksi berulang, serta paparan terhadap lingkungan fisik dan sosial yang kurang mendukung pertumbuhan optimal [8]. Dalam mengidentifikasi stunting, digunakan berbagai indikator pertumbuhan, diantaranya adalah tinggi badan menurut usia (TB/U), berat badan menurut usia (BB/U), serta berat badan menurut tinggi badan (BB/TB).

Salah satu daerah di Indonesia yang mendapatkan perhatian khusus terkait isu stunting adalah Kabupaten Brebes. Tingginya angka kejadian stunting di kabupaten ini menjadikannya sebagai salah satu lokasi prioritas untuk intervensi yang didukung oleh pemerintah pusat [9]. Identifikasi dan pemahaman mengenai resiko stunting penting

dilakukan untuk mendukung program penanganan yang lebih tepat sasaran. Maka dari itu, diperlukan sebuah pendekatan berbasis teknologi untuk meningkatkan efektivitas pada proses identifikasi.

Pendekatan teknologi yang relevan untuk menjawab tantangan ini adalah pemanfaatan *machine learning*, khususnya melalui penerapan algoritma *machine learning* [10]. *Machine learning* yang merupakan salah satu bidang kecerdasan buatan yang sedang berkembang, memberikan pendekatan inovatif untuk mengklasifikasi stunting [11]. Algoritma *machine learning* memiliki kemampuan untuk mengolah dan menganalisis pola dari data antropometri balita seperti tinggi badan, berat badan, dan usia guna membangun model klasifikasi status stunting. Model yang dihasilkan ini berpotensi besar untuk digunakan sebagai alat bantu deteksi dini balita yang memiliki risiko mengalami stunting, sehingga intervensi dapat direncanakan dan diimplementasikan secara lebih cepat dan tepat sasaran [12].

Permasalahan yang sering terjadi pada model klasifikasi yaitu adanya ketidakseimbangan data (*data imbalance*), yang terjadi karena adanya kelas yang jumlahnya jauh lebih banyak. Dalam membuat model klasifikasi *data imbalance* merupakan suatu masalah [13]. Untuk menangani masalah tersebut dapat dilakukan dengan metode tertentu salah satunya adalah SMOTE. Metode SMOTE bekerja dengan cara membuat sampel sintesis baru untuk kelas minoritas berdasarkan karakteristik sampel minoritas yang telah ada, sehingga distribusi data antar kelas menjadi lebih seimbang [14]. Pada penelitian ini jumlah kelas normal jauh lebih kecil dari jumlah kelas stunting.

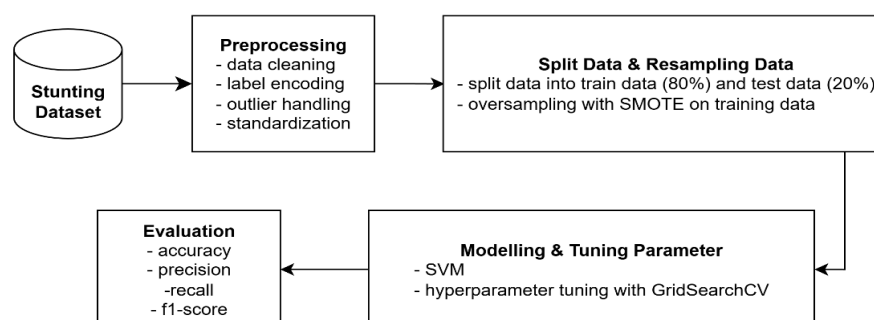
Penelitian yang dilakukan sebelumnya tentang klasifikasi status stunting menggunakan algoritma svm dengan kernel linier dengan kondisi data yang seimbang menghasilkan akurasi 89% [15]. Penelitian selanjutnya untuk klasifikasi penyakit stunting dengan metode SVM. Model dilatih menggunakan kernel Radial Basis Function (RBF) dan Linear, dengan kernel RBF menghasilkan akurasi 95.26% dan dengan kernel Linear menghasilkan akurasi 78.67% [16]. Penelitian lain dalam menangani *data imbalance* pada algoritma SVM untuk klasifikasi penyakit stroke. Pengujian tanpa SMOTE menunjukkan performa terbaik pada kernel Polynomial dengan akurasi tertinggi sebesar 78.86%, sementara setelah SMOTE diterapkan, akurasi model meningkat secara signifikan, terutama pada kernel RBF yang mencapai akurasi optimal 92.12%, presisi 88.65%, recall 97.14% dan F1-Score 89.26% dengan parameter Cost 100 dan Gamma 1 [17]. Penelitian lain juga menggunakan SMOTE dalam klasifikasi penyakit diabetes dengan algoritma SVM yang dievaluasi menggunakan metode *K-Fold cross validation* dengan membagi dataset menjadi 10 subset. Model ini mencapai rata-rata akurasi sebesar 88.77%, presisi 88.50%, dan recall 89.21% [18]. Penelitian yang dilakukan sebelumnya hanya berfokus pada penerapan algoritma SVM pada dataset yang relatif seimbang atau penerapan SMOTE pada penyakit lain.

Berdasarkan penelitian-penelitian tersebut, penelitian ini bertujuan untuk meningkatkan performa dari model SVM dalam mengklasifikasikan status stunting pada balita dengan menangani permasalahan pada dataset penelitian ini yang tidak seimbang menggunakan teknik *oversampling* SMOTE dan dievaluasi. Selain itu, penelitian ini juga akan menginvestigasi dampak dari proses optimasi *tuning parameter* menggunakan GridSearchCV pada model SVM untuk memastikan pemilihan kernel dan parameter lainnya dapat menghasilkan performa klasifikasi yang paling andal. Dengan adanya penanganan ketidakseimbangan pada dataset dan *tuning parameter*, terdapat peningkatan untuk nilai *accuracy*, *precision*, *recall*, dan *f1-Score*. Maka dari itu model SVM dalam mengklasifikasikan status stunting pada balita dapat digunakan untuk membantu mendorong ketepatan klasifikasi stunting balita sehingga bisa memberikan bantuan mencegahnya sejak dini.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan data mining sebagai pendekatannya dengan menggunakan algoritma SVM dalam mengklasifikasi status stunting pada balita untuk mendapatkan hasil yang tepat. Terdapat dua pendekatan untuk dianalisa lebih lanjut, yaitu dengan menggunakan SMOTE dan tanpa SMOTE. Proses pada penelitian ini diawali dengan mengumpulkan data, lalu dilanjutkan dengan preprocessing data, proses berikutnya yaitu membagi dataset menjadi data train (80%) dan data test (20%) disertai penanganan data yang tidak seimbang pada data train, lalu implementasi algoritma SVM dan pencarian parameter terbaik menggunakan GridSearchCV, dan yang terakhir evaluasi untuk mengukur seberapa baik model dalam melakukan klasifikasi balita stunting.



Gambar 1. Metode Penelitian

2.2 Dataset

Pada tahap awal data diambil dari beberapa puskesmas di Kabupaten Brebes diantaranya Puskesmas Brebes, Puskesmas Kaligangsa, Puskesmas Kersana, Puskesmas Kalimati, dan Puskesmas Bulakamba dengan jumlah total data 9598 yang merupakan data pengukuran antropometri balita. Data tersebut berisikan 8 fitur yang terdiri dari JK, Usia Saat Ukur, Berat, Tinggi, ZS BB/U, ZS TB/U, ZS BB/TB, Keterangan dan 2 kelas yang terdiri dari normal dan stunting.

2.3 Preprocessing Data

Setelah tahap pengumpulan data selanjutnya preprocessing data agar data siap sebelum dilakukan klasifikasi. Proses ini merupakan bagian penting dalam penelitian yang menggunakan *data mining* dan *machine learning*, karena kualitas model yang dihasilkan sangat dipengaruhi oleh kualitas data yang digunakan. Data mentah yang diperoleh dari lapangan biasanya masih memiliki berbagai masalah, seperti adanya nilai yang kosong, data yang berulang, distribusi yang tidak seimbang, atau perbedaan skala antar fitur. Jika masalah-masalah tersebut tidak ditangani secara tepat, maka hasil klasifikasi bisa menjadi tidak tepat atau bahkan bias. Tahapan ini meliputi pembersihan data agar tidak ada data kosong dan menghilangkan data duplikat, mengubah data kategori menjadi data numerik (*label encoding*), penanganan outlier menggunakan metode *Inter Quartile Range (IQR)*, dan standarisasi data menggunakan *Standard Scaler*.

2.4 Split Data & Resampling Data

Setelah tahap preprocessing, dataset dibagi menjadi dua bagian utama menggunakan metode train-test split dengan proporsi 80% untuk data train dan 20% untuk data test. Namun pada dataset ini ditemukan adanya ketidakseimbangan antar kelasnya.

Untuk mengatasi masalah tersebut, diterapkan metode SMOTE yang berfungsi untuk menambah data sintetis pada kelas minoritas. Dengan cara ini, model memperoleh lebih banyak contoh dari kelas positif, sehingga dapat belajar secara lebih seimbang dan mengurangi kecenderungan bias terhadap kelas mayoritas.

2.5 Modelling

Tahap selanjutnya yaitu modelling, penelitian ini menggunakan algoritma SVM untuk klasifikasi balita stunting. SVM dikenal memiliki efektivitas tinggi dalam menangani data dengan dimensi yang kompleks dan mampu melakukan pemisahan antar kelas dengan membangun margin yang maksimal, sehingga sangat sesuai untuk problem klasifikasi biner seperti penentuan status stunting [19]. Pada Tabel 1 merupakan nilai-nilai yang diuji dalam tuning parameter pada pelatihan model.

Tabel 1. Parameter model

Parameter	Nilai
Kernel	rbf, linear
C	0.1, 1, 10
Gamma	1, 0.1, 0.01

Pada Tabel 1 menyajikan parameter dan nilai yang diuji pada model saat proses pelatihan menggunakan teknik GridsearchCV dengan menerapkan K fold yaitu 10, meliputi kernel, C, dan Gamma. Penentuan parameter beserta nilai yang digunakan, disesuaikan untuk memaksimalkan kinerja model dalam menyelesaikan tugas klasifikasi.

2.6 Evaluation

Dari dua percobaan yang dilakukan, selanjutnya model akan dievaluasi untuk menilai seberapa baik model dalam melakukan klasifikasi balita stunting. Evaluasi ini sangat penting karena menentukan seberapa baik model mampu mengenali pola dalam data dan memprediksi kondisi stunting dengan tingkat akurasi yang tinggi. Evaluasi dilakukan secara eksklusif menggunakan data test, yang merupakan 20% dari dataset awal dan tidak terlibat sama sekali dalam proses pelatihan. Penggunaan data test terpisah ini penting untuk memberikan estimasi performa yang tidak bias, sehingga hasil evaluasi dapat mencerminkan kemampuan model ketika dihadapkan pada data baru di kondisi nyata. Pendekatan ini juga merupakan langkah krusial untuk menguji kemampuan generalisasi model dan memastikan bahwa performa tinggi yang diperoleh bukan disebabkan oleh overfitting, yaitu kondisi di mana model hanya menghafal data latih.

Dalam proses evaluasi, nilai dari beberapa metrik seperti akurasi, precision, recall, dan F1-score disimpan dan dianalisis. Secara khusus, metrik precision, recall, dan F1-score dihitung untuk masing-masing kelas (Normal dan Stunting). Pendekatan ini penting karena mampu menunjukkan sejauh mana penerapan SMOTE berhasil meningkatkan sensitivitas model terhadap kelas minoritas (stunting), yang seringkali terabaikan jika penilaian hanya berfokus pada akurasi keseluruhan. Hasil perbandingan antara kedua model berdasarkan metrik-metrik tersebut menjadi dasar utama dalam analisis performa dan pembahasan hasil penelitian. Selain metrik kuantitatif tersebut, analisis juga akan diperkuat dengan visualisasi confusion matrix untuk memberikan gambaran detail mengenai jenis-jenis kesalahan yang dibuat oleh setiap model.

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Dataset

Pada penelitian ini berjumlah 9598 dengan 8 fitur. Pada data ini keterangan stunting difokuskan dengan keterangan status Normal yang berjumlah 8185 dan Stunting yang berjumlah 1413. Gambar 2 adalah data stunting yang digunakan dalam model klasifikasi stunting pada balita.

	JK	Berat	Tinggi	Usia Saat Ukur	ZS BB/U	ZS TB/U	ZS BB/TB	Keterangan
0	L	14.7	100.0	54	-1.22	-1.42	-0.54	Normal
1	L	12.0	88.8	33	-1.19	-1.48	-0.61	Normal
2	P	9.4	83.5	26	-1.93	-1.14	-1.85	Normal
3	P	12.2	93.6	47	-1.98	-1.97	-1.19	Normal
4	L	6.5	63.8	6	-1.89	-1.94	-0.87	Normal
...
9593	L	3.2	45.0	0	-0.30	-2.58	2.81	Stunting
9594	L	8.2	62.0	3	1.92	-0.13	2.65	Normal
9595	L	4.9	60.0	1	0.07	2.04	-2.50	Normal
9596	P	3.4	55.0	1	-1.20	0.97	-3.35	Normal
9597	P	3.1	48.0	0	-0.29	-0.62	0.45	Normal

Gambar 2. Data Stunting Balita

Kemudian pada Tabel 2 merupakan keterangan dari setiap fitur pada dataset yang digunakan pada penelitian ini. Dengan adanya penjelasan yang lengkap ini, dataset menjadi lebih rapi dan mudah diproses secara maksimal, sehingga dapat menghasilkan model klasifikasi yang tepat dan dapat dipercaya.

Tabel 2. Keterangan Dataset

Fitur	Keterangan
JK	Jenis Kelamin
Usia Saat Ukur	Usia saat ukur dalam bulan
Berat	Berat badan saat ukur
Tinggi	Tinggi badan saat ukur
ZZ BB/U	Standar deviasi berat badan menurut umur
ZZ TB/U	Standar deviasi tinggi badan menurut umur
ZZ BB/TB	Standar deviasi berat badan menurut tinggi badan
Keterangan	Status stunting

3.2 Preprocessing Data

Pada tahap preprocessing data dilakukan beberapa proses penting dengan tujuan untuk memastikan kualitas dataset sebelum melakukan proses klasifikasi. Proses ini meliputi pembersihan data, *label encoding*, penanganan outlier, pemisahan fitur dan label, dan normalisasi data. Pada proses pembersihan data, dalam penelitian ini data yang kosong dihilangkan karena tidak memberikan informasi yang bermanfaat bagi model. Adanya data kosong dapat menciptakan bias, menurunkan tingkat akurasi, bahkan menyebabkan algoritma gagal saat proses pelatihan. Gambar 3 merupakan data kosong yang ada dalam dataset.

	JK	Berat	Tinggi	Usia Saat Ukur	ZS BB/U	ZS TB/U	ZS BB/TB	Keterangan
3163	L	17.0	109.0	38	NaN	NaN	NaN	Stunting
4739	P	16.6	103.0	15	NaN	NaN	NaN	Stunting
7580	P	17.5	102.0	47	NaN	NaN	NaN	Stunting
7859	L	15.2	101.8	44	NaN	NaN	NaN	Stunting
8170	P	15.7	101.5	19	NaN	NaN	NaN	Stunting
8194	L	16.3	113.0	44	NaN	NaN	NaN	Stunting
8948	P	12.6	98.4	22	NaN	NaN	NaN	Stunting

Gambar 3. Data Kosong

Selain itu, penghapusan data yang duplikat juga sangat penting. Data duplikat biasanya muncul akibat kesalahan dalam pencatatan atau proses input yang berulang di fasilitas kesehatan, dan bila tetap dibiarkan, akan membuat model belajar dari pola-pola yang tidak seimbang. Pada Gambar 4 merupakan data duplikat pada data.

	JK	Berat	Tinggi	Usia Saat Ukur	ZS BB/U	ZS TB/U	ZS BB/TB	Keterangan
138	L	4.6	55.4	2	-1.87	-1.93	-0.15	Normal
728	L	9.2	74.0	13	-0.58	-1.06	-0.13	Normal
753	L	13.4	93.0	45	-1.27	-2.00	-0.17	Normal
1233	L	13.0	97.0	39	-1.03	-0.16	-1.42	Normal

Gambar 4. Data Duplikat

Langkah selanjutnya yaitu mengubah data kategori menjadi data numerik agar dapat diproses oleh algoritma *machine learning*. Hal ini dilakukan karena sebagian besar algoritma klasifikasi, termasuk SVM, hanya dapat mengolah data dalam bentuk numerik dan tidak dapat langsung membaca data dalam format kategorikal. Pada penelitian ini, terdapat dua fitur kategorikal yang harus diubah, yaitu fitur JK dan fitur Keterangan.

Tabel 3. Data sebelum Label Encoding

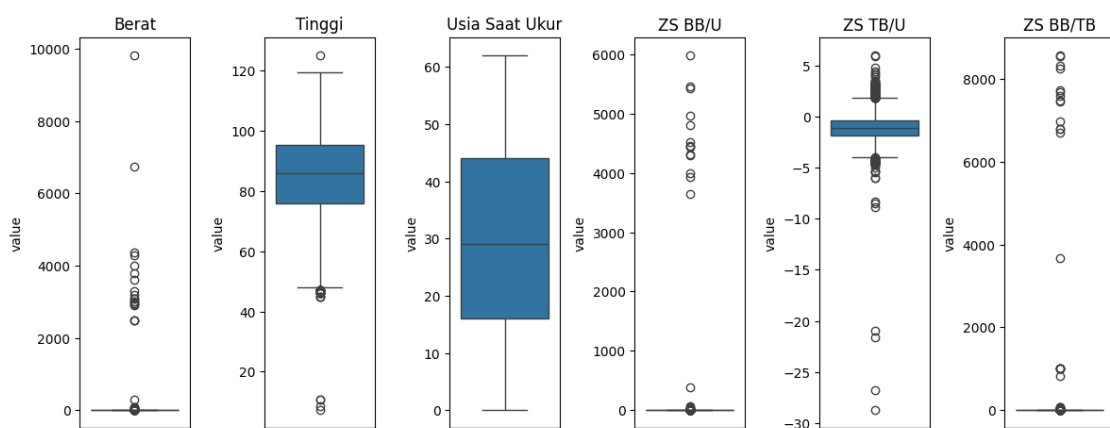
JK	Berat	Tinggi	Usia Saat Ukur	ZS BB/U	ZS TB/U	ZS BB/TB	Keterangan
L	3.2	45.0	0.0	-0.30	-2.58	2.81	Stunting
P	9.4	83.5	26.0	-1.93	-1.14	-1.85	Normal

Pada Tabel 3, fitur JK untuk keterangan L diubah menjadi angka 0 dan P diubah menjadi angka 1, dan pada fitur keterangan untuk normal diubah menjadi angka 0 dan untuk stunting diubah menjadi angka 1. Proses *label encoding* ini penting karena jika kategori tetap dibiarkan dalam bentuk teks, model tidak akan mampu mengenali hubungan antar variabel. Untuk hasil data setelah dilakukan label encoding terdapat pada Tabel 4.

Tabel 4. Data setelah Label Encoding

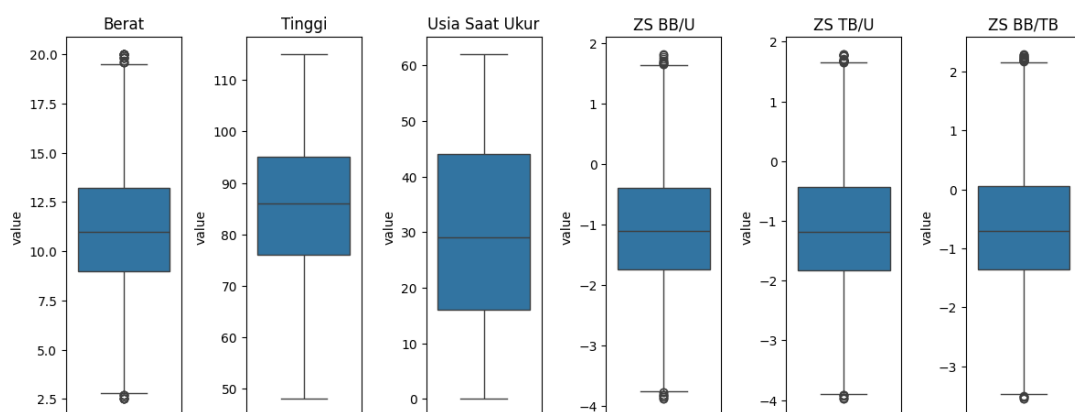
JK	Berat	Tinggi	Usia Saat Ukur	ZS BB/U	ZS TB/U	ZS BB/TB	Keterangan
0	3.2	45.0	0.0	-0.30	-2.58	2.81	1
1	9.4	83.5	26.0	-1.93	-1.14	-1.85	0

Data yang sudah dilabel encoding, selanjutnya dilakukan penanganan outlier. Tahap ini menggunakan metode *Interquartile Range (IQR)* atau dengan analisis visual menggunakan boxplot. Outlier sendiri merupakan nilai yang menyimpang secara signifikan dari mayoritas data pada dataset yang dapat memengaruhi hasil analisis statistik [18]. Dilakukan pemeriksaan pada dataset terkait konsistensi data yang terdiri dari beberapa fitur yang ada seperti Berat, Tinggi, Usia Saat Ukur, ZS BB/U, ZS TB/U, dan ZS BB/TB.



Gambar 5. Hasil Outlier

Dari hasil analisis pada Gambar 5 menggunakan boxplot kolom Berat, ZS BB/U, ZS TB/U, dan ZS BB/TB mempunyai banyak nilai ekstrim sedangkan kolom tinggi lebih sedikit mengandung outlier. Untuk menangani outlier tersebut, dilakukan penghapusan data yang nilainya berada di luar batas normal menggunakan IQR. Jumlah data setelah dilakukan pembersihan data yang mengandung outlier menjadi 9149. Untuk hasil data setelah penanganan outlier dapat dilihat pada Gambar 6.



Gambar 6. Hasil Penanganan Outlier

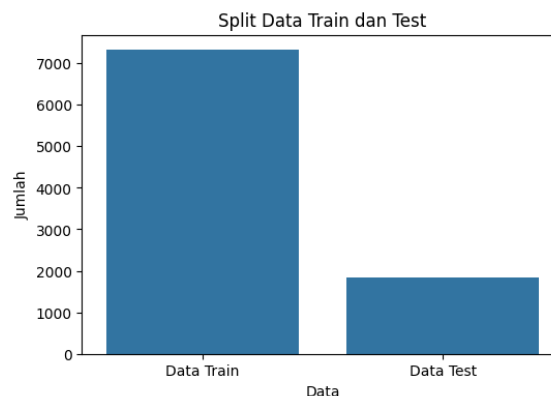
Langkah selanjutnya, dilakukan proses standarisasi data menggunakan *Standard Scaler*. Standarisasi merupakan langkah penting dalam proses *preprocessing* karena data hasil pengukuran antropometri, seperti berat badan, tinggi badan, atau usia, memiliki satuan serta rentang nilai yang berbeda-beda. Perbedaan skala ini bisa menyebabkan bias pada algoritma yang menggunakan jarak, termasuk algoritma SVM, yang sangat rentan terhadap perbedaan skala antar fitur. Oleh karena itu, *Standard Scaler* digunakan untuk mengubah nilai fitur dalam dataset sehingga memiliki nilai rata-rata 0 dan standar deviasi 1 [20]. Dengan dilakukannya hal tersebut, setiap fitur akan memiliki skala distribusi yang sama, sehingga meminimalkan kemungkinan bias yang timbul akibat perbedaan skala antar fitur. Pada Tabel 5 merupakan hasil dari standarisasi pada data.

Tabel 5. Hasil Standarisasi

Fitur	Data Sebelum	Data Sesudah
JK	0	-0.9764494116202096
Berat	14.7	1.1591820458168247
Tinggi	100	1.1542795664916565
Usia saat ukur	1616	1.437133645170635
ZZ BB/U	-1.22	-0.15827210196156344
ZZ TB/U	-1.42	-0.32276523999398377
ZZ BB/TB	-0.54	0.11109540179560867

3.3 Split & Resampling Data

Pada tahap ini, data hasil ekstraksi fitur dibagi menjadi dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji. Tujuan dari pembagian ini adalah agar model dapat mempelajari pola dari sebagian besar data melalui data latih, sementara data uji digunakan untuk menilai kemampuan model dalam mengenali pola pada data baru yang belum pernah dipelajari sebelumnya. Proses pembagian dilakukan secara acak agar distribusi setiap subset tetap representatif terhadap keseluruhan dataset. Gambar 7 merupakan hasil dari pembagian data train yang berjumlah 7319 dan data test yang berjumlah 1830.



Gambar 7. Split data

Namun, hasil analisis menunjukkan adanya ketidakseimbangan kelas antara balita normal (kelas 0) dan balita stunting (kelas 1), di mana jumlah balita normal jauh lebih banyak dibandingkan balita stunting. Kondisi ini berpotensi membuat model lebih condong terhadap kelas mayoritas selama pelatihan.

Untuk mengatasi permasalahan tersebut, digunakan metode SMOTE pada tahap pemodelan, bukan sebelum pembagian data. Pendekatan ini memastikan bahwa proses oversampling hanya diterapkan pada data latih sehingga menghindari kebocoran data ke data uji. Dengan cara ini, evaluasi performa model menjadi lebih objektif dan mencerminkan kemampuan sebenarnya terhadap data baru.

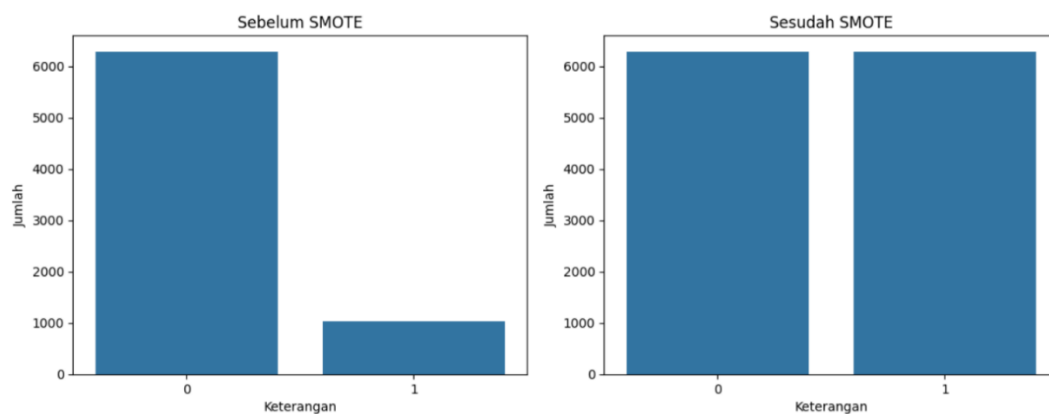
Melalui strategi pembagian data yang seimbang serta penerapan SMOTE pada tahap pemodelan, diharapkan model dapat belajar secara adil dari kedua kelas dan menghasilkan performa klasifikasi yang lebih akurat dan stabil.

3.4 Implementasi Model

Pada penelitian ini model yang digunakan untuk mengklasifikasi balita stunting (kelas 1) dan balita normal (kelas 0) adalah SVM. Algoritma ini dipilih karena SVM dikenal memiliki efektivitas tinggi dalam menangani data yang memiliki dimensi kompleks dan mampu menciptakan *hyperplane* yang optimal sebagai pemisah antar kelas[21]. Dengan sifat tersebut, SVM dapat menghasilkan margin yang maksimal antara dua kelas yang berbeda, sehingga sangat cocok untuk masalah klasifikasi biner seperti penentuan status stunting [19]. Selain itu, SVM juga memiliki kelebihan dalam hal kemampuan generalisasi terhadap data baru karena prinsip kerjanya berfokus pada titik-titik data yang terletak di sekitar batas pemisah, yang membuat model lebih kuat dibandingkan algoritma klasifikasi yang sederhana.

Dalam penelitian ini, digunakan dua metode dalam pelatihan model, yaitu dengan menerapkan SMOTE dengan tuning parameter dan tanpa menerapkan SMOTE dengan parameter *default*. Pendekatan ini dipilih untuk mengetahui pengaruh teknik penyeimbangan data terhadap kinerja SVM dan juga memperoleh parameter terbaik. Jika tidak menggunakan SMOTE, model langsung bekerja pada dataset asli yang memiliki ketidakseimbangan distribusi antara kelas normal dan kelas stunting. Kondisi ini bisa menyebabkan model cenderung bersifat bias terhadap kelas yang jumlah datanya lebih banyak.

Penggunaan SMOTE bertujuan untuk mengatasi ketidakseimbangan data dengan menambahkan jumlah sampel di kelas minoritas [14]. Kedua model dilatih menggunakan data train dan dievaluasi kinerjanya pada data test yang sama. SMOTE diterapkan hanya pada data train untuk mencegah kebocoran data. Selain itu, proses *hyperparameter* tuning melalui GridSearchCV memungkinkan penemuan kombinasi parameter yang optimal untuk SVM termasuk nilai C, gamma, dan jenis kernel yang lebih lanjut menyempurnakan kemampuan model dalam membedakan antara balita stunting dan normal. Proses GridSearchCV menemukan parameter terbaik, yaitu C = 10, gamma = 1, dan kernel = linear, yang menghasilkan rata-rata skor cross-validation terbaik sebesar 98.84%. Kombinasi SMOTE dan tuning parameter ini memastikan kinerja model yang tidak hanya akurat secara keseluruhan, tetapi juga sensitif dan andal dalam mendeteksi kelas minoritas (stunting). Metode ini mampu memberikan gambaran yang lebih lengkap mengenai dampak SMOTE terhadap kinerja model SVM. Pada Gambar 8 merupakan sebaran data sebelum dan sesudah menggunakan SMOTE. Sebelum SMOTE kelas 0 (normal) berjumlah 6282 sampel dan kelas 1 (stunting) berjumlah 1037 sampel. Lalu setelah SMOTE jumlah data menjadi seimbang dengan masing-masing kelas berjumlah 6282 sampel.



Gambar 8. Persebaran data sebelum dan sesudah SMOTE

Dalam proses pelatihan, model dievaluasi dengan menganalisa hasil dari *accuracy*, *precision*, *recall*, dan *f1-score*. *Accuracy* menunjukkan seberapa besar prediksi yang benar dibandingkan dengan seluruh data uji. *Precision* digunakan untuk mengukur ketepatan model dalam mendeteksi kelas positif, dalam hal ini seberapa besar dari semua prediksi positif yang dikeluarkan, seberapa besar di antaranya yang benar. *Recall* mengukur kemampuan model dalam mendeteksi data yang seharusnya positif di dalam kategori positif, seberapa besar kasus positif dijumpai dalam data yang seharusnya di dalam kategori positif. Adapun *f1-score* menunjukkan nilai yang dihasilkan dari rata-rata harmonik antara presisi dan *recall*, di mana *f1-score* sangat berguna dan memberikan gambaran seimbang tentang performa model, terutama pada data dengan distribusi kelas yang tidak seimbang.

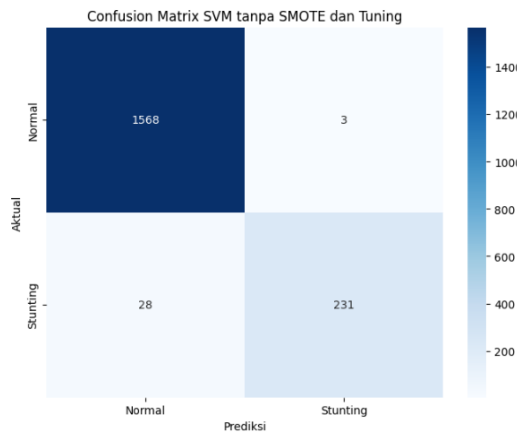
3.5 Evaluasi Model

Model SVM yang dilatih menggunakan data dengan distribusi kelas tidak seimbang dan tanpa tuning parameter menunjukkan akurasi keseluruhan yang tinggi, yaitu sebesar 98.31%, yang menunjukkan model cukup baik dalam memprediksi keseluruhan data. Namun, ketika dianalisis berdasarkan metrik per kelas, terdapat kecenderungan bias.

Tabel 6. Metrik per kelas untuk model tanpa SMOTE dan tuning

Kelas	Precision	Recall	F1-score
Normal (0)	98.25%	99.81%	99.02%
Stunting (1)	98.72%	89.19%	93.71%
Accuracy	98.31%		

Dari Tabel 6, untuk presisi model mencapai 98.25% pada kelas Normal (0) dan 98.72% pada kelas Stunting (1), yang artinya ketika model memprediksi suatu kelas, prediksinya umumnya benar. Meski demikian, nilai *recall* menunjukkan masalah. Model berhasil mendeteksi 99.81% dari kelas Normal (0), tetapi hanya 89.19% dari kelas Stunting (1). Hal ini menunjukkan model sangat baik dalam mengenali anak-anak yang normal, tetapi gagal mendeteksi sekitar 11% dari balita yang sebenarnya mengalami stunting. Di sisi lain, *f1-Score* model menunjukkan 99.02% untuk kelas Normal (0) dan 93.71% untuk kelas Stunting (1), yang menunjukkan keseimbangan antara presisi dan recall yang cukup baik, meski performa recall untuk kelas stunting masih perlu ditingkatkan.



Gambar 9. Confusion matrix setelah SMOTE dan tuning parameter

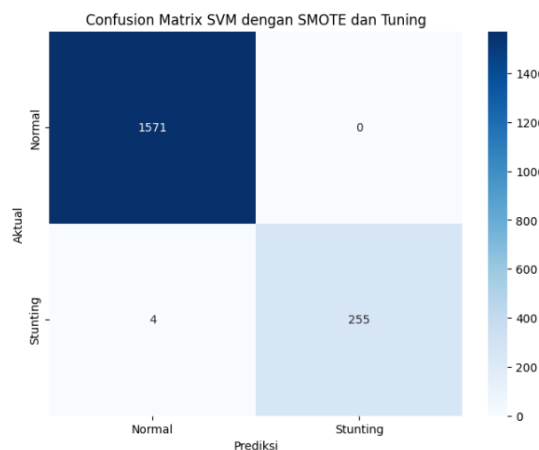
Confusion matrix pada Gambar 9 menunjukkan hasil dari model klasifikasi SVM dasar tanpa menggunakan SMOTE dan penyesuaian parameter. Dari hasil tersebut, terlihat bahwa model memiliki kinerja yang tidak seimbang dalam mengenali sensitivitas antar kelas. Model berhasil memprediksi 1568 kasus sebagai Normal yang sebenarnya Normal (True Negative), menunjukkan kemampuan model yang baik dalam mengenali kelas mayoritas. Namun, terdapat 28 kasus Stunting yang salah diprediksi sebagai Normal (False Negative), yang menunjukkan bahwa model masih mengalami kesulitan dalam mengenali seluruh anggota kelas minoritas. Di sisi lain, 231 kasus berhasil diprediksi sebagai Stunting yang benar (True Positive), sedangkan hanya 3 kasus Normal yang salah dikenali sebagai Stunting (False Positive). Secara umum, meskipun akurasi dan presisi model terlihat tinggi, adanya 28 kasus False Negative menunjukkan bahwa model kurang sensitif terhadap kelas minoritas (stunting). Oleh karena itu, diperlukan penerapan SMOTE dan tuning parameter untuk mengurangi jumlah kasus yang terlewat serta meningkatkan kemampuan model dalam mengenali data dengan distribusi yang tidak seimbang.

Setelah menerapkan SMOTE dan tuning parameter, kinerja model meningkat secara signifikan, mencapai akurasi sebesar 99,78%. Analisis metrik per kelas menunjukkan bahwa model tidak hanya akurat tetapi juga lebih seimbang

Tabel 7. Metrik per kelas dengan SMOTE dan tuning parameter

Kelas	Precision	Recall	F1-score
Normal (0)	99.75%	100%	99.87%
Stunting (1)	100%	98.46%	99.22%
Accuracy	99.78%		

Dari Tabel 7, presisi model mencapai 99.75% untuk kelas Normal (0) dan 100% untuk kelas Stunting (1), yang berarti setiap prediksi stunting yang dibuat model benar. Sementara itu, nilai recall mencapai 100% untuk kelas Normal (0) dan meningkat drastis menjadi 98.46% untuk kelas Stunting (1). Peningkatan recall yang besar ini menunjukkan bahwa model mampu mendeteksi hampir semua kasus stunting, hanya sedikit kasus yang terlewatkan. Di sisi lain, F1-Score model mencapai 99.87% untuk kelas Normal (0) dan 99.22% untuk kelas Stunting (1), menunjukkan keseimbangan yang hampir sempurna antara presisi dan recall di kedua kelas. Secara keseluruhan, peningkatan metrik, terutama recall dan f1-Score untuk kelas stunting, membuktikan bahwa model yang telah dioptimalkan ini jauh lebih baik dan dapat diandalkan dalam mendeteksi dini stunting.



Gambar 10. Confusion matrix sebelum SMOTE

Confusion matrix pada Gambar 10 menunjukkan kinerja model yang jauh lebih baik setelah menerapkan SMOTE dan tuning parameter. Dari hasil yang diperoleh, terlihat bahwa ketidakseimbangan dalam performa telah berhasil diperbaiki secara signifikan. Model berhasil memprediksi 1571 kasus sebagai Normal yang memang benar-benar Normal (True Negative) dan sepenuhnya menghilangkan kesalahan False Positive menjadi 0 kasus. Peningkatan terbesar terlihat pada jumlah False Negative, yang berkurang drastis dari 28 kasus menjadi hanya 4 kasus. Hal ini menunjukkan model kini lebih sensitif dan efektif dalam mengenali kelas minoritas. Selain itu, jumlah kasus yang benar-benar diprediksi sebagai Stunting (True Positive) juga meningkat menjadi 255 kasus. Secara keseluruhan, hasil ini menunjukkan bahwa dengan penerapan SMOTE dan tuning parameter, model SVM menjadi sangat andal dan seimbang. Kemampuannya dalam meminimalkan False Negative secara signifikan membuktikan bahwa model tidak lagi cenderung bias terhadap kelas mayoritas dan sangat efektif dalam mengklasifikasikan kelas minoritas (stunting) secara akurat.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma SVM sangat efektif digunakan untuk mengklasifikasi status stunting pada balita, terutama jika data telah melalui proses penyeimbangan dengan metode SMOTE. Penerapan SMOTE dan tuning parameter terbukti meningkatkan kinerja model, dengan kenaikan *accuracy* dari 98.31% menjadi 99.78% dan peningkatan *recall* untuk kelas minoritas (stunting) dari 89.19% menjadi 98.46%. Peningkatan pada *recall* ini menunjukkan bahwa penggabungan SMOTE dengan tuning parameter bekerja efektif dalam mengatasi masalah ketidakseimbangan antar kelas, yang sebelumnya menjadi hambatan utama dalam kemampuan model untuk mendeteksi kasus stunting. Meskipun hasil yang sangat baik ini bisa memicu kekhawatiran tentang overfitting, pendekatan yang digunakan dalam penelitian ini menunjukkan bahwa model mampu beradaptasi secara baik pada data baru. Uji akhir dilakukan pada data yang benar-benar terpisah dan tidak pernah digunakan dalam proses pelatihan atau penyesuaian parameter. Kinerja yang sangat baik pada data yang belum pernah dilihat membuktikan bahwa model telah belajar pola-pola yang relevan untuk klasifikasi, bukan hanya menghafal data pelatihan. Dengan demikian, hasil yang diperoleh mencerminkan kemampuan prediksi model yang kuat dan tidak bias. Penggunaan teknik ini dapat meningkatkan efektivitas model dalam berbagai aplikasi medis dan prediksi lainnya yang menghadapi masalah distribusi kelas tidak seimbang, seperti dalam kasus prediksi stunting ini. Meskipun penelitian ini berhasil menunjukkan hasil yang baik, ada beberapa keterbatasan pada penelitian ini, yaitu dataset yang digunakan hanya mencakup faktor-faktor antropometri, sedangkan faktor lain seperti kondisi lingkungan, kondisi sosial-ekonomi, maupun cara mendidik anak belum dimasukkan. Hal ini membuat pemahaman tentang risiko stunting belum benar-benar lengkap. Lalu, penelitian ini hanya menggunakan satu jenis algoritma yaitu SVM. Dengan memperhatikan keterbatasan tersebut, penelitian berikutnya bisa fokus pada penggunaan variabel yang lebih beragam, termasuk faktor gizi, sanitasi, dan data sosial-ekonomi agar model yang dihasilkan lebih menyeluruh dalam memprediksi risiko stunting. Selain itu, eksplorasi penggunaan algoritma lain serta penerapan teknik penyeimbangan data yang lebih efektif, seperti ADASYN atau SMOTEENN, bisa menjadi pilihan yang lebih baik.

REFERENCES

- [1] N. Norfai dan A. Abdullah, "Determinan Kejadian Stunting pada Balita di Wilayah Kerja Puskesmas Pekauman Kota Banjarmasin," *Jurnal Akademika Baiturrahim Jambi*, vol. 10, no. 1, hlm. 131–144, Mar 2021, doi: 10.36565/jab.v10i1.292.
- [2] M. L. Tobing, M. Pane, dan E. Harianja, "Pola Asuh Ibu dengan Kejadian Stunting pada Anak Usia 24–59 Bulan di Wilayah Kerja Puskesmas Kelurahan Sekupang Kota Batam," *PREPOTIF*, vol. 5, no. 1, hlm. 448–465, Apr 2021, doi: 10.31004/prepotif.v5i1.1630.
- [3] W. H. Yusuf, "Faktor Resiko Stunting di Nusa Tenggara Barat (NTB), Indonesia," *Religion, Culture, and State Journal*, vol. 2, no. 1, hlm. 34–45, Apr 2022.
- [4] D. T. Rahayu, "Anemia In Pregnancy With Stunting In Gayam Village District Gurah Kediri," *midwifery*, vol. 7, no. 1, hlm. 81–94, Jun 2021, doi: 10.21070/midwifery.v7i1.1319.
- [5] BKKP, "Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022." Diakses: 6 Agustus 2025. [Daring]. Tersedia pada: <https://repository.badankebijakan.kemkes.go.id/id/eprint/4855/>
- [6] BKKP, *Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka*. 2025. Diakses: 6 Agustus 2025. [Daring]. Tersedia pada: <https://repository.badankebijakan.kemkes.go.id/id/eprint/5861/>
- [7] H. BKKP, "SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%," Badan Kebijakan Pembangunan Kesehatan | BKKP Kemenkes. Diakses: 6 Agustus 2025. [Daring]. Tersedia pada: <https://www.badankebijakan.kemkes.go.id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198/>
- [8] S. U. Janah, M. Azinar, dan H. Koesyanto, "Tatanan Keluarga dengan Kejadian Stunting Balita Usia 24–59 Bulan di Dataran Tinggi dan Dataran Rendah," *HIGEIA (Journal of Public Health Research and Development)*, vol. 7, no. 2, Art. no. 2, Apr 2023, doi: 10.15294/higeia.v7i2.64203.
- [9] A. Aseegaf, "Evaluasi Program Percepatan Penurunan Stunting di Kecamatan Bulakamba Kabupaten Brebes," *Public Policy and Management Inquiry*, vol. 8, no. 1, Art. no. 1, Mei 2024, doi: 10.20884/1.ppmi.2024.8.1.10982.
- [10] B. Sapriatin dan F. A. Sianturi, "Penerapan Teorema Bayes Mendeteksi Stunting Pada Balita," *Jurnal Media Informatika*, vol. 3, no. 1, hlm. 24–37, Des 2021, doi: 10.55338/jumin.v3i1.
- [11] S. Lonang, A. Yudhana, dan M. K. Biddinika, "Rancangan Sistem Klasifikasi Kekurangan Gizi Balita Dengan Metode K-Nearest Neighbor," *JINRPL*, vol. 5, no. 1, hlm. 73, Mar 2023, doi: 10.36499/jinrpl.v5i1.7834.



- [12] A. I. Putri, Y. Syarif, P. Jayadi, F. Arrazak, dan F. N. Salisah, “Implementasi Algoritma Decision Tree dan Support Vector Machine (SVM) untuk Prediksi Risiko Stunting pada Keluarga: Implementation of Decision Tree and Support Vector Machine (SVM) Algorithm for Stunting Risk Prediction,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, Art. no. 2, 2023, doi: 10.57152/malcom.v3i2.1228.
- [13] M. I. Anugrah, J. Zeniarja, dan D. S. Setiawan, “Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, Art. no. 1, Jun 2024, doi: 10.29408/edumatic.v8i1.25838.
- [14] T. Yulian dan E. R. Susanto, “Comparative Analysis of Oversampling and SMOTEENN Techniques in Machine Learning Algorithms for Breast Cancer Prediction,” *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 3, Art. no. 3, Mei 2025, doi: 10.32520/stmsi.v14i3.5146.
- [15] A. Jalil, A. Homaidi, dan Z. Fatah, “Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita,” *G-Tech*, vol. 8, no. 3, hlm. 2070–2079, Jul 2024, doi: 10.33379/gtech.v8i3.4811.
- [16] T. R. Dewanti, R. Prathivi, dan S. Susanto, “Implementasi Metode SVM untuk Klasifikasi Penyakit Stunting Bayi,” *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 1, Art. no. 1, Feb 2025, doi: 10.51401/jinteks.v7i1.5070.
- [17] L. Pasiolo, I. Afrianty, E. Budianita, dan R. Abdillah, “Penerapan Teknik SMOTE pada Klasifikasi Penyakit Stroke dengan Algoritma Support Vector Machine,” *Jurnal Sistem Informasi*, vol. 7, no. 1, 2025.
- [18] N. Fitriyani, D. R. Amalia, H. H. Handayani, dan A. F. N. Masruriyah, “Aplikasi Berbasis Web Berdasarkan Model Klasifikasi Algoritma SVM dan Logistic Regression Terhadap Data Diabetes,” *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 7, no. 4, Art. no. 4, Okt 2023, doi: 10.33395/remik.v7i4.13001.
- [19] G. Rahmawati, S. A. Sanmas, E. Nudyawati, N. D. Syaharani, M. R. Annas, dan F. Fauzi, “Studi Perbandingan Performa: Prediksi Status Stunting Pada Anak Berdasarkan Data Antropometri Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN),” *PROSIDING SEMINAR NASIONAL SAINS DATA*, vol. 4, no. 1, Art. no. 1, Okt 2024, doi: 10.33005/senada.v4i1.330.
- [20] T. Z. Jasman, E. Hasmin, Sunardi, C. Susanto, dan W. Musu, “Perbandingan Logistic Regression, Random Forest, dan Perceptron pada Klasifikasi Pasien Gagal Jantung,” *CSRID (Computer Science Research and Its Development Journal)*, vol. 14, no. 3, hlm. 271–286, Des 2022, doi: 10.22303/csrid-.14.3.2022.271-286.
- [21] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, dan A. Christian, “Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes,” *Indonesian Journal on Software Engineering (IJSE)*, vol. 7, no. 1, hlm. 63–68, Jun 2021, doi: 10.31294/ijse.v7i1.10279.