

Optimasi Metode BMR dan *Simulated Annealing* untuk Algoritma SVM dalam Mengatasi *Imbalanced* dan *High Dimensional Data Stunting*

Mukminatul Munawaroh¹, Taghfirul Azhima Yoga Siswa², Wawan Joko Pranoto³

^{1,2,3}Program Studi Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Indonesia

2011102441064@umkt.ac.id¹, tay758@umkt.ac.id², wjp337@umk.ac.id³

ABSTRACT; *Stunting cases in Samarinda city since 2022 have experienced an increase in prevalence by 23.9%. Handling of these cases remains suboptimal as in 2023, the stunting rate remained high. The objective of this research is to enhance the accuracy of stunting classification or prediction in Samarinda city using the Support Vector Machine (SVM) algorithm, optimized with Simulated Annealing (SA) and Boundary Margin Relief (BMR) as the feature selection method, to identify features influencing the classification of stunting in children. This research process involved data collection from the Samarinda City Health Office in 2023, encompassing 26 health centers, 20 attributes, and a total of 150,465 data rows. The division of test and training data used 10-fold cross-validation. The study found that BMR identified nine significant features in determining or classifying stunting conditions in children: ZS TB/U, ZS BB/U, ZS BB/TB, Height, Weight, Weight Gain, LiLA, BB/TB, and BB/U. This resulted in identifying dominant features consistent with previous research, namely Weight for Age (BB/U) and Weight for Height (BB/TB). Performance evaluation was presented using a confusion matrix. Thus, the final result showed the average accuracy of SVM with RBF kernel using Simulated Annealing optimization and Boundary Margin Relief feature selection was 54.90%, compared to 88.00% accuracy without optimization, indicating an increase in accuracy by 33.10%. This improvement in accuracy demonstrates the effectiveness of this method in addressing stunting with large datasets and class imbalance.*

Keywords: *Classification, Stunting, SVM, BMR, Simulated Annealing*

ABSTRAK; Kasus Stunting di kota Samarinda dari tahun 2022 masih mengalami peningkatan prevalensi sebesar 23,9%, penanganannya kasus ini masih kurang optimal karena tahun 2023 tingkat stunting masih tinggi. Tujuan penelitian ini meningkatkan akurasi klasifikasi atau prediksi stunting di Kota Samarinda menggunakan algoritma Support Vector Machine (SVM) yang akan di optimasi dengan metode Simulated Annealing (SA) dan metode Bunday Margin Relief (BMR) sebagai seleksi fitur, guna mengetahui fitur yang berpengaruh dalam menentukan atau mengklasifikasikan kondisi stunting pada anak. Proses penelitian ini melibatkan pengumpulan data dari Dinas Kesehatan Kota Samarinda tahun 2023 yang melibatkan 26 puskesmas, 20 atribut dan total 150,465 baris data. Pembagian

data uji dan data latih menggunakan cross validation k-fold 10, Hasil penelitian ini ditemukan bahwa BMR menghasilkan sembilan fitur yang berpengaruh dalam menentukan atau mengklasifikasikan kondisi stunting pada anak-anak yaitu ZS TB/U, ZS BB/U, ZS BB/TB, Tinggi Badan, Berat Badan, Naik Berat Badan, LiLA, BB/TB, dan BB/U. Sehingga mendapat fitur dominan yang sama dengan penelitian sebelumnya yaitu atribut Berat Badan/Umur (BB/U) dan Berat Badan/Tinggi Badan (BB/TB). Evaluasi kinerja di tampilkan dalam confusion matrix, sehingga hasil akhirnya berupa rata-rata akurasi SVM dengan kernel RBF menggunakan metode optimasi Simulated Annealing dan seleksi fitur Boundary Margin Relief sebesar 54.90%, akurasi tanpa optimasi sebesar 88.00% sehingga menunjukkan kenaikan akurasi sebesar 33.10%. Peningkatan akurasi menunjukkan metode ini baik dalam mengatasi stunting dengan dataset besar dan ketidakseimbangan kelas.

Kata Kunci: Klasifikasi, Stunting, SVM, BMR, *Simulated Annealing*

PENDAHULUAN

Stunting adalah masalah penelitian nasional yang masih trending sebagai riset nasional tahun 2020 hingga 2024, Sumber [1]. Prevalensi balita stunting di Kalimantan Timur tahun 2022 adalah 23,9%. Stunting menjadi masalah kesehatan yang mendesak, karena dampaknya yang berkelanjutan dapat mengganggu perkembangan kognitif, produktivitas, dan kualitas hidup anak di masa dewasa. Oleh karena itu, penggunaan teknologi informasi dan metode klasifikasi data mining menjadi penting dalam upaya untuk memberikan diagnosis dini dan intervensi yang tepat guna, sehingga dapat mencegah dampak yang lebih serius pada pertumbuhan dan perkembangan anak. Salah satu metode *machine learning* dalam klasifikasi atau prediksi stunting adalah SVM [2], karena dalam situasi di mana ada perbedaan kelas yang jelas dan ruang berdimensi besar, SVM cocok untuk masalah dengan banyak fitur sehingga performa metode tidak menurun drastis ketika jumlah fitur melebihi jumlah sampel. Pada penelitian terdahulu SVM+PSO menghasilkan akurasi 78% [3]. C4.5 menghasilkan akurasi 61,82% [4][5][6] dan *Naïve Bayes* menghasilkan akurasi 75% dengan data yang berdimensi rendah [7][8].

Adapun masalah *high dimension* berdasarkan penelitian dari [9]–[13] menurut penelitian tersebut, metode *Relief* adalah algoritma yang baik dalam menangani data berdimensi tinggi. Menurut [14] untuk memilih fitur yang berkualitas dari dataset yang besar, dalam penelitiannya menyarankan, untuk menggunakan algoritma pembobotan fitur inovatif yang disebut *Boundary Margin Relief* (BMR). Konsep utama BMR adalah memprediksi bobot fitur

melalui pengukuran *hyperplane* lokal, yang umumnya dilakukan dalam teknik *I-Relief*. Bobot fitur dalam metode tersebut sangat efektif dalam mengeliminasi fitur yang berlebihan. Terbukti dapat meningkatkan akurasi dengan eksperimennya pada data *Leukemia* menghasilkan 91,2% akurasinya. Oleh karena itu, penulis ingin menerapkan metode tersebut dalam penelitian data stunting di Kota Samarinda guna membantu dalam proses seleksi fitur.

Selain itu, masalah ketidakseimbangan kelas juga menambah tantangan dalam klasifikasi karena dapat menyebabkan tumpang tindih antara label atau kelas [15] sehingga penelitian ini bertujuan untuk meningkatkan kinerja SVM dengan menangani data yang tidak seimbang, menggunakan seleksi fitur dengan metode *Boundary Margin Relief* (BMR) untuk data berdimensi tinggi, dan metode *Simulated Annealing* (SA) untuk meningkatkan akurasi model SVM. SA Terbukti dapat meningkatkan akurasi SVM, berdasarkan hasil eksperimen dari [16], dalam penelitiannya, menguji metode SVM (MLP karnel) dan SA dengan data *Portuguese course* menghasilkan akurasi sebesar 78.35% tanpa SA setelah di optimasi dengan SA menjadi 90.72%. sehingga penulis ingin menerapkan metode SA sebagai metode Optimasi SVM dalam klasifikasi atau prediksi data Stunting di Kota Samarinda.

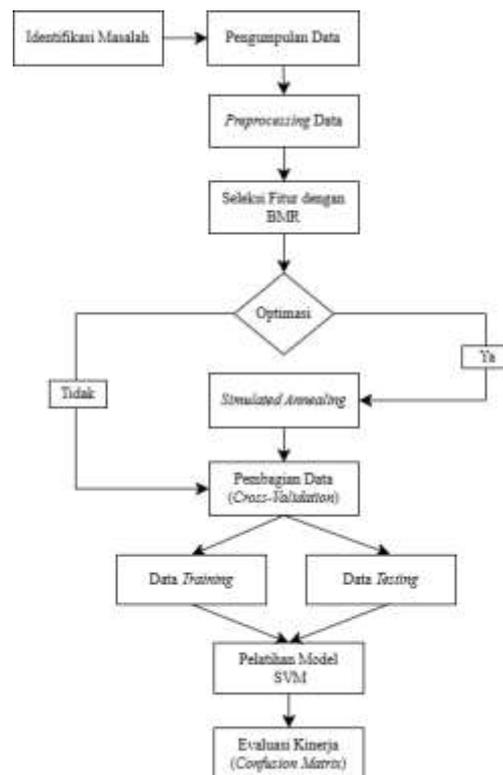
METODE PENELITIAN

A. Objek Penelitian

Data yang digunakan dalam penelitian ini adalah data stunting yang bersumber dari Dinas Kesehatan Jl. Milono No.1, Bugis, Kecamatan Samarinda Kota, Kota Samarinda, Kalimantan Timur 76112. Stunting adalah kondisi dimana pertumbuhan fisik seseorang, terutama anak-anak, terhambat sehingga tinggi badan tidak sesuai dengan usia mereka. Data diperoleh dari hasil pemeriksaan pasien setiap harinya baik yang mengidap penyakit stunting atau tidak. Dari 26 Puskesmas di Kota Samarinda.

B. Prosedur Penelitian

Proses penelitian yang dilakukan secara terstruktur untuk merencanakan, melaksanakan, dan mengevaluasi suatu penelitian berdasarkan model yang digunakan, untuk memastikan bahwa penelitian berjalan lancar dan menghasilkan hasil yang dapat dipercaya. Alur penelitian ini tertera pada gambar diagram berikut:



Gambar 1. Diagram Alur Penelitian

C. Identifikasi Masalah

Penanganan masalah klasifikasi data stunting dengan dimensi tinggi dan ketidakseimbangan kelas memerlukan pendekatan yang cermat. Metode SVM telah terbukti efektif dalam menangani masalah ini [2], tetapi untuk meningkatkan kinerjanya, penggunaan *Simulated Annealing* dan *Boundary Margin Relief* dalam seleksi fitur. *Simulated Annealing* digunakan untuk mencari konfigurasi parameter SVM yang optimal, sementara *Boundary Margin Relief* membantu mengidentifikasi fitur-fitur yang paling penting. Kombinasi ketiga metode ini diharapkan dapat meningkatkan akurasi klasifikasi data stunting dan memberikan dasar yang kuat untuk merancang metodologi penelitian yang efektif.

D. Pengumpulan Data

Teknik pengumpulan data berupa data sekunder, yaitu penggunaan data yang sudah ada dan dikumpulkan oleh Dinas Kesehatan Kota Samarinda dari lembaga lain untuk tujuan tertentu. Data sekunder dapat berupa data dari *survey*, laporan, basis data, rekaman administratif, atau sumber data lainnya yang tersedia untuk umum. Data yang diperoleh terdiri dari 20 kolom dengan total 150,465 *record*.

Tabel 1. Informasi Atribut

No	Atribut	Tipe Data	Keterangan
1	Nik	String	Nomor Induk Kependudukan
2	Nama	String	Nama
3	JK	String	Jenis Kelamin
4	Tel Lahir	Date	Tanggal Lahir
5	Nama Ortu	String	Nama Orang Tua
6	Provinsi	String	Provinsi
7	Kab/Kota	String	Kabupaten atau Kota
8	Kec	String	Kecamatan
9	Puskesmas	String	Lokasi Puskesmas
10	Posyandu	String	Lokasi Posyandu
11	Total Pengukuran	Integer	Total Pengukuran
12	Tanggal Pengukuran	Integer	Tanggal Pengukuran
13	Berat	Integer	Berat Badan
14	Tinggi	Integer	Tinggi Badan
15	BB/U	Numeric	Berat Badan Menurut Umur
16	ZS BB/U	Numeric	Z Score Berat Badan menurut Umur
17	TB/U	Numeric	Tinggi Badan menurut Umur
18	ZS TB/U	Numeric	Z Score Tinggi Badan menurut Umur
19	BB/TB	Numeric	Berat Badan menurut Tinggi Badan
20	ZS BB/TB	Numeric	Z Score Berat Badan menurut Tinggi Badan

Namun, fitur yang di gunakan ada 14 fitur yaitu Nama, JK, Berat, Tinggi, LiLA, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB, Naik Berat Badan, Jml Vit A, dan Tanggal Pengukuran.

E. Data Preprocessing

Pada tahapan ini adalah serangkaian langkah yang dilakukan untuk membersihkan, menyiapkan, dan mengatur data mentah menjadi format yang sesuai untuk analisis lebih lanjut. Tahapan ini penting untuk memastikan data yang digunakan dalam analisis memiliki kualitas yang baik dan siap digunakan. Berikut adalah beberapa tahapannya.

1) Pembersihan data

Pada dataset stunting di Kota Samarinda, dilakukan pembersihan data yang mencakup penanganan data yang tidak wajar, misalnya tinggi badan atau berat badan yang ekstrem. Penghapusan data yang memiliki nilai #N/A (nilai tidak tersedia) atau tidak memiliki nilai sama sekali, serta penghapusan data yang terduplikasi. Proses ini penting untuk memastikan bahwa data yang digunakan dalam analisis adalah valid dan akurat.

Sebanyak 150,465 *record*. sebagai data terduplikasi dan dihapus dari dataset. Data yang terduplikasi dapat menyebabkan kesalahan dalam analisis karena informasi yang sama muncul lebih dari sekali. Terdapat 34,199 *record* yang memiliki nilai #N/A, yang juga dihapus karena tidak memberikan informasi yang berguna. Data yang memiliki nilai #N/A berarti informasi tersebut tidak tersedia atau hilang, sehingga tidak dapat digunakan untuk mendapatkan hasil analisis yang tepat. Setelah melalui proses pembersihan yang teliti ini, jumlah total data yang tersisa dan siap digunakan dalam analisis stunting di Kota Samarinda adalah 18,396 *record*.

2) Transformasi Data

Transformasi variabel kategorikal menjadi numerik dilakukan dengan menggunakan fungsi *LabelEncoder* dari *library scikit-learn*. *LabelEncoder* adalah fungsi yang mengubah variabel kategorikal, seperti jenis kelamin dan berat badan berdasarkan umur, menjadi angka yang dapat dimengerti oleh algoritma *machine learning*.

3) Data *Balanced*

Ketidakseimbangan kelas ini dapat menyebabkan model menjadi cenderung tidak adil, di mana model lebih cenderung memprediksi kelas dengan lebih banyak sampel dan mengabaikan kelas dengan lebih sedikit sampel [17]. Akibatnya, performa model secara keseluruhan menurun, terutama dalam hal prediksi kelas yang kurang terwakili. Salah satu teknik yang efektif untuk menangani masalah ini adalah SMOTE (*Synthetic Minority Over-sampling Technique*), yang bekerja dengan menghasilkan sampel sintetis untuk kelas dengan lebih sedikit sampel, sehingga meningkatkan jumlah dan keberagaman kelas tersebut dalam dataset stunting.

F. Seleksi Fitur Menggunakan BMR

Salah satu metode yang dapat digunakan untuk seleksi fitur adalah *Boundary Margin Relief*. *Boundary Margin Relief* adalah teknik yang mengukur kepentingan setiap fitur berdasarkan margin batas antara kelas-kelas yang berbeda dalam dataset. Teknik ini membantu dalam mengidentifikasi fitur-fitur yang paling berpengaruh dalam klasifikasi, dengan mempertimbangkan jarak antara sampel dari kelas yang sama dan kelas yang berbeda.

G. Optimasi dengan *Simulated Annealing*

Simulated Annealing akan digunakan untuk mengoptimasi pemilihan fitur dalam proses pembelajaran mesin. Tujuan utamanya adalah untuk menemukan kumpulan fitur yang memberikan kinerja terbaik pada model SVM. *Simulated Annealing* adalah sebuah metode yang terinspirasi dari proses annealing dalam fisika, untuk mencari solusi optimal dalam ruang fitur dengan mengizinkan beberapa solusi yang lebih buruk diterima pada awalnya untuk menghindari terjebak di dalam daerah yang tidak optimal. Fungsi ini mengoptimalkan pemilihan fitur dengan mengevaluasi performa model pada data uji, mengubah suhu (T) dan faktor penurunan suhu (α) secara iteratif, dan memperbarui solusi terbaik berdasarkan skor akurasi.

H. Pembagian Data dengan *Cross Validation*

Data stunting akan dibagi menjadi data pelatihan (*training*) dan data uji (*testing*) menggunakan metode *cross-validation*. Pembagian data dengan *cross-validation* yaitu pembagian dataset menjadi subset yang disebut (*folds*), di mana model dilatih pada *fold* ke-10 dan diuji pada *fold* lainnya secara bergantian [2].

I. Pelatihan Model SVM

Support Vector Machine (SVM) adalah algoritma pembelajaran yang digunakan untuk tugas-tugas klasifikasi dan regresi. Tujuan utama dari SVM adalah untuk menemukan *hyperplane* terbaik yang memisahkan dua kelas dalam ruang fitur sedemikian rupa sehingga jarak antara *hyperplane* dan *instance* terdekat dari masing-masing kelas (yang disebut sebagai *support vectors*) sebesar mungkin. Berikut persamaan yang dapat digunakan untuk mendapatkan *hyperplane* pada SVM.

$$(w \cdot x_i) + b = 0 \quad (1)$$

Data x_i yang termasuk dalam kelas -1 dapat dirumuskan sebagai berikut:

$$(w \cdot x_i) + b < 1, y_i = -1 \quad (2)$$

Data x_i yang termasuk dalam kelas +1 dapat dirumuskan sebagai berikut:

$$(w \cdot x_i) + b > 1, y_i = 1 \quad (3)$$

Keterangan:

x_i = Data ke-i

w = Nilai bobot *support vector* yang tegak lurus dengan *hyperplane* (*weight vector*)

b = Nilai bias (merujuk pada parameter yang menentukan posisi *hyperplane* terhadap titik asal (*origin*) dalam ruang fitur)

y_i = Kelas/label data ke-i

Dalam proses klasifikasi dengan SVM, seringkali terjadi kondisi di mana kernel linear tidak bekerja dengan baik, yang menyebabkan hasil klasifikasi data yang buruk. Kondisi ini dapat diatasi dengan menggunakan kernel *Radial Basis Function* (RBF) [18], yang dirumuskan sebagai berikut:

$$K(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2) \quad (4)$$

Keterangan:

$K(x_1, x_2)$ = fungsi kernel RBF yang mengukur kesamaan antara dua sampel x_1 dan x_2

$\|x_1 - x_2\|$ = Jarak Euclidean kuadrat antara dua vektor fitur x_1 dan x_2

γ (*gamma*) = Parameter yang menentukan "lebar" fungsi RBF

exp = Fungsi eksponensial

Pelatihan model SVM akan menggunakan karnel RBF, $\gamma = 10$, $\text{cost} = 5$ untuk membuktikan teori dari [19]. Model SVM akan di latih menggunakan data *training* setelah penanganan ketidakseimbangan kelas, *high dimension* dengan BMR dan optimasi dengan *Simulated Annealing*.

J. Evaluasi Kinerja Algoritma

Kinerja model SVM yang dioptimalkan akan dievaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* adalah sebuah tabel yang digunakan dalam evaluasi klasifikasi untuk menampilkan performa model dengan membandingkan hasil prediksi dengan nilai sebenarnya dari data yang diuji.

Tabel 2. *Confusion Matrix*

Prediction	True values	
	True	False
True	TP Correct result	FP Unexpected result

<i>False</i>	FN <i>Missing result</i>	TN <i>Correct absence of result</i>
--------------	-----------------------------	--

Dalam *confusion matrix*, terdapat beberapa istilah yang digunakan dalam kasus klasifikasi seperti yang dijelaskan oleh [17].

Keterangan :

True Positive (TP) = Data positif benar.

False Positive (FP) = Data negatif sebagai data positif.

False Negative (FN) = Data positif sebagai data negatif.

True Negative (TN) = Data negatif benar

Evaluasi kinerja model yang akan dilakukan dengan mengukur nilai akurasi, berikut rumusnya

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Keterangan :

Accuracy = Mengukur sejauh mana model klasifikasi memberikan prediksi yang benar secara keseluruhan.

Performa model akan dibandingkan dengan metode SVM dengan metode SVM + BMR + *Simulated Annealing* untuk menilai efektivitas kombinasi model.

Berikut adalah kategori akurasi berdasarkan persentase:

Sangat Baik : 90% - 100%

Baik : 80% - 89%

Cukup : 70% - 79%

Kurang : 60% - 69%

Sangat Kurang: < 60%

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Hasil yang diperoleh dari setiap tahapan akan dijelaskan secara rinci untuk memberikan gambaran lengkap mengenai efektivitas metode yang digunakan dalam penelitian ini. Berikut adalah hasil dari tahapan-tahapan tersebut:

1) Hasil *Preprocessing* Data

Pada bagian ini akan menampilkan hasil dari setiap tahapan penelitian yang dilakukan. Tahapan tersebut meliputi preprocessing data, pembagian data, penerapan metode *Simulated Annealing* untuk optimasi parameter BMR pada algoritma SVM, serta evaluasi kinerja model yang telah dioptimasi. Hasil yang diperoleh dari setiap tahapan akan dijelaskan secara rinci untuk memberikan gambaran lengkap mengenai efektivitas metode yang digunakan dalam penelitian ini. Pada tahap ini ada beberapa proses yang akan di tampilkan yaitu hasil data cleaning, transformasi data dan penanganan ketidakseimbangan kelas.

1) Hasil Data Cleaning

Data sampel yang berjumlah 150,465 baris didalamnya masih banyak data terduplikat, data *missing* dan data *error* (NaN). Setelah proses penghapusan duplikat dari 150,466 menjadi 34,199 *record* dan penghapusan data missing dari 34,199 menjadi 18396 *record*.

2) Hasil Transformasi Data

Berikut adalah hasil transformasi data yang menunjukkan perubahan signifikan dalam pola dan struktur data, sehingga mempermudah analisis lebih lanjut

Tabel 3. Data Sebelum di Transformasi

No	JK	BB/U	Naik Berat Badan	BB/TB
0	L	Berat Badan Normal	-	Gizi Baik
1	L	Berat Badan Normal	O	Gizi Baik
2	L	Berat Badan Normal	T	Gizi Baik
3	L	Berat Badan Normal	O	Gizi Baik
4	L	Berat Badan Normal	N	Gizi Baik
...
18391	P	Berat Badan Normal	O	Gizi Baik
18392	P	Kurang	O	Gizi Baik
18393	L	Berat Badan Normal	T	Gizi Baik
18394	L	Berat Badan Normal	O	Gizi Baik
18395	L	Berat Badan Normal	N	Gizi Baik

Pada Tabel 3 aribut yang akan di ubah adalah isi pada fitur JK, BB/U, Naik Berat Badan, dan BB/TB. Dalam Fitur JK 'L' akan di ubah menjadi 0 dan 'P' menjadi 1. Pada Fitur BB/U 'Berat Badan Normal' di ubah menjadi 0 dan berat badan yang 'Kurang' di ubah menjadi 1. Selanjutnay pada fitur Naik Berat Badan, setiap huruf akan di ubah menjadi numerik

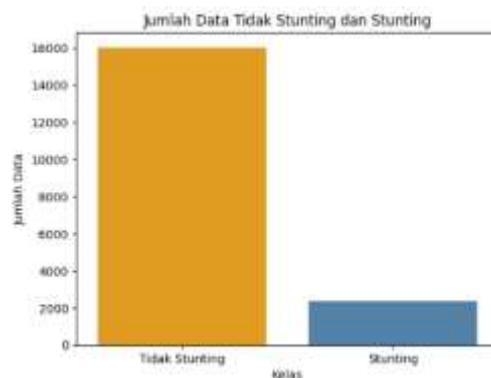
berdasarkan bobot, seperti huruf O (Obesitas) menjadi 10, N (Normal) menjadi 11 dan T (Tinggi) menjadi 12. Pada fitur BB/TB ‘Gizi Baik’ di ubah menjadi 0 dan ‘Gizi Buruk’ menjadi 1.

Tabel 4 Hasil Transformasi

No	JK	BB/U	Naik Berat Badan	BB/TB
0	0	0	0	0
1	0	0	11	0
2	0	0	12	0
3	0	0	11	0
4	0	0	10	0
...
18391	1	0	11	0
18392	1	1	11	0
18393	0	0	12	0
18394	0	0	11	0
18395	0	0	10	0

3) Hasil Penanganan Ketidakseimbangan Kelas

Berikut adalah hasil penanganan ketidakseimbangan data yang menunjukkan kelas yang lebih merata



Gambar 2. Kelas Data Tidak Seimbang

Pada Gambar 2, Grafik pertama menunjukkan distribusi awal data dengan ketidakseimbangan yang signifikan antara jumlah data "Tidak Stunting" sebanyak 16046 baris data dan "Stunting" sebanyak 2350 baris data, di mana data "Tidak Stunting" jauh lebih banyak dibandingkan data "Stunting".

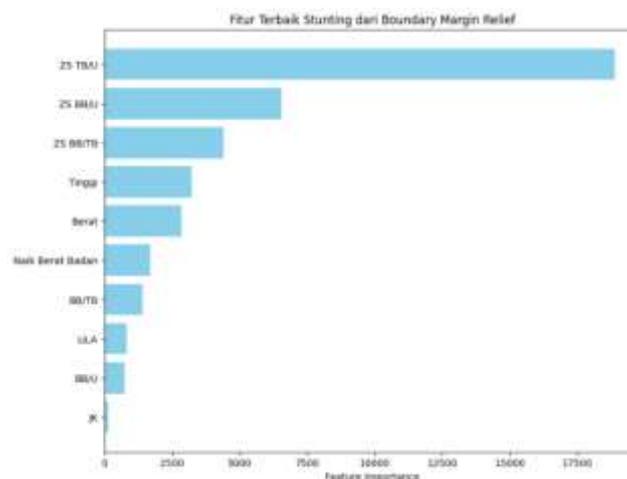


Gambar 3. Hasil Penanganan Ketidakseimbangan Data

Pada Gambar 3, memperlihatkan hasil setelah penanganan ketidakseimbangan data, di mana jumlah data antara "Stunting" dan "Tidak Stunting" telah disamakan, yaitu 16046 baris data. Penanganan ketidakseimbangan data ini dilakukan untuk memastikan bahwa model pembelajaran tidak berat sebelah terhadap kelas dominan dan dapat memprediksi dengan lebih akurat untuk kedua kelas tersebut.

B. Hasil Seleksi Fitur dengan BMR

Seleksi fitur menggunakan *Boundary Margin Relief* menghasilkan beberapa fitur terbaik yang memiliki nilai bobot tinggi. Terlihat pada gambar grafik di bawah ini:



Gambar 4. Hasil Fitur Terbaik BMR

Berdasarkan hasil implementasi, fitur yang mempengaruhi kinerja BMR dan *Support Vector Machine* (SVM) terlihat pada tabel berikut berdasarkan nilai bobotnya:

Tabel 5. Bobot Fitur

No	Fitur	Bobot	Peringkat
1	ZS TB/U	18890	1
2	ZS BB/U	6543	2
3	ZS BB/TB	4396	3
4	Tinggi	3212	4
5	Berat	2836	5
6	Naik Berat Badan	1698	6
7	LiLA	1409	7
8	BB/TB	817	8
9	BB/U	742	9
10	JK	128	10

Berdasarkan Tabel 3.5 fitur yang dipilih adalah ZS TB/U, ZS BB/U, ZS BB/TB, Tinggi, Berat, Naik Berat Badan, LiLA, BB/TB, dan BB/U.

C. Hasil Optimasi SVM dengan *Simulated Annealing*

Berikut adalah tabel untuk menampilkan hasil SVM dan optimasi SVM dengan *Simulated Annealing* berdasarkan 10 *fold*:

Tabel 6. Perbandingan Akurasi SVM

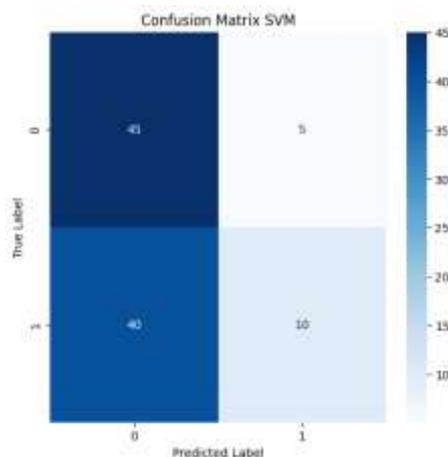
Fold	Akurasi SVM	Akurasi SVM Dengan SA	Kenaikan Akurasi
1	61.00%	87.00%	26.00%
2	54.00%	81.00%	27.00%
3	51.00%	89.00%	38.00%
4	58.00%	87.00%	29.00%
5	55.00%	93.00%	38.00%
6	54.00%	88.00%	34.00%
7	51.00%	88.00%	37.00%
8	50.00%	95.00%	45.00%
9	58.00%	83.00%	33.00%
10	57.00%	89.00%	32.00%
Rata-rata	54.90%	88.00%	33.10%

Rata-rata akurasi SVM adalah 54.90%, sedangkan rata-rata akurasi SVM dengan *Simulated Annealing* adalah 87.80%. Terdapat kenaikan rata-rata akurasi sebesar 33.10% ketika menggunakan *Simulated Annealing* untuk optimasi SVM serta karnel RBF dengan nilai gamma = 10 dan cost = 5.

D. Hasil Evaluasi Performa dengan *Confusion Matrix*

Menggunakan *confusion matrix* untuk mengevaluasi kinerja model klasifikasi yang dikembangkan. *Confusion matrix* adalah tabel yang memperlihatkan perbandingan antara prediksi yang benar dan salah yang dilakukan oleh model terhadap data yang telah diketahui kelasnya. Dari *confusion matrix* ini, dapat dilihat seberapa baik model dalam mengklasifikasikan data ke dalam kategori yang benar, serta mengidentifikasi jenis-jenis kesalahan prediksi yang dilakukan oleh model tersebut.

a. Model SVM



Gambar 5. Hasil *Confusion Matrix* Model SVM

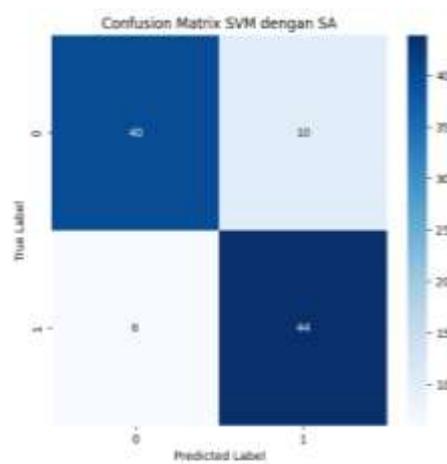
Untuk menghitung akurasi secara manual dari *confusion matrix* dapat menggunakan rumus akurasi sebagai berikut:

Tabel 7. Perhitungan Rata-rata *Convusion Matriks* SVM

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	45	<i>FP</i>	5
<i>FN</i>	40	<i>TP</i>	10

$$ccuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{10 + 45}{10 + 45 + 5 + 10} = 54.90\%$$

b. Model SVM dan SA



Gambar 6. Hasil *Confusion Matrix* Model SVM dengan SA

Untuk menghitung akurasi secara manual dari *confusion matrix* dapat menggunakan rumus akurasi sebagai berikut:

Tabel 8. Perhitungan Rata-rata *Convusion Matriks* SVM dengan SA

<i>True Label</i>		<i>Predicted Label</i>	
<i>TN</i>	40	<i>FP</i>	10
<i>FN</i>	6	<i>TP</i>	44

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} = \frac{44 + 40}{44 + 40 + 10 + 6} = 88.00\%$$

Pembahasan

Berikut pembahasan penelitian berdasarkan hasil dan rumusan masalah yang di cantumkan:

a. Fitur yang Mempengaruhi Kinerja BMR dan SVM

Berdasarkan hasil yang terlihat pada Tabel 6, maka ditentukan bahwa atribut yang memiliki ranking 1-9 yaitu ZS TB/U, ZS BB/U, ZS BB/TB, Tinggi, Berat, Naik Berat Badan, LiLA, BB/TB, dan BB/U digunakan sebagai atribut dalam pemodelan SVM karena memiliki nilai bobot yang tinggi. Sedangkan pada penelitian yang di lakukan oleh [20] menggunakan algoritma SVM dan *Chi-Square* menghasilkan 6 fitur yakni ZS TB/U, BB/U, ZS BB/U, Tinggi, LiLA dan Berat. Pada penelitian [21] menggunakan *Random Forest* dan PSO, fitur yang paling berpengaruh yaitu BB/TB. Penelitian lainnya oleh [22] menggunakan metode KNN dan ANN menghasilkan 3 atribut yaitu Berat Badan/Umur, Berat Badan/Tinggi Badan, dan Tinggi

Badan/Umur. Sehingga persamaan fitur yang dominan dari penelitian tersebut adalah atribut Berat Badan/Umur (BB/U) dan Berat Badan/Tinggi Badan (BB/TB).

b. Performa Algoritma

Perbandingan performa akurasi dengan penelitian [21] menghasilkan akurasi 78.33% menggunakan metode *Random Forest* dan *Cross Validation* tanpa optimasi ataupun seleksi fitur. Penelitian [20] menggunakan data stunting SVM dengan Chi-Square dan SMOTE untuk *High Dimension Data Stunting* dengan *Cross Validation* $k\text{-fold}=10$ menghasilkan akurasi 96,6% namun tanpa metode optimasi. Penelitian lainnya dilakukan oleh [19] menggunakan data stunting, metode SVM dan kernel RBF untuk data stunting dengan nilai $\gamma = 10$ dan $\text{cost} = 5$ menghasilkan akurasi 100% namun, saat di terapkan dalam penelitian ini, kernel RBF dengan nilai $\gamma = 10$ dan $\text{cost} = 5$ menghasilkan rata-rata akurasi 54.90% mengalami penurunan. Sedangkan saat penambahan seleksi fitur BMR dan optimasi *Simulated Annealing* akurasinya mengalami peningkatan sebanyak 33.10% sehingga akurasinya menjadi 88.00%. dengan demikian implementasi seleksi fitur menggunakan BMR dan optimasi menggunakan *Simulated Annealing* dapat meningkatkan akurasi klasifikasi atau prediksi dari model SVM dengan kernel RBF pada data Stunting.

KESIMPULAN DAN SARAN

Kesimpulan

Berikut kesimpulan berdasarkan hasil penelitian ini:

- a. Fitur yang memiliki pengaruh dominan terhadap stunting dari metode seleksi fitur *Boundary Margin Relief* (BMR), yaitu ZS TB/U, ZS BB/U, ZS BB/TB, tinggi badan, berat badan, naik berat badan, LiLA, BB/TB, dan BB/U. Fitur dominannya yaitu Berat Badan/Umur (BB/U) dan Berat Badan/Tinggi Badan (BB/TB).
- b. Hasil akurasi SVM kernel RBF dengan metode optimasi *Simulated Annealing* dan seleksi fitur *Boundary Margin Relief* sebesar 54.90% tanpa optimasi dan sebesar 88.00% setelah di optimasi dengan SA, sehingga menunjukkan kenaikan akurasi sebesar 33.10%.

Saran

Berikut beberapa saran untuk penelitian kedepannya:

- a. Penelitian selanjutnya disarankan untuk penyesuaian nilai pada parameter *gamma* dan *C* (*Cost*) kernel RBF, agar akurasi yang di dapat benar-benar akurat. Karena pada penelitian ini saat menambahkan nilai *gamma* dan nilai *C* pada kernel RBF akurasi SVM tanpa optimasi SA menurun sangat jauh, sehingga memerlukan penyesuaian lebih lanjut. Selain itu untuk penelitian lanjutan, dapat mencoba kernel SVM lainnya seperti *linear*, *sigmoid* atau *polynomial* untuk membandingkan tingkat kinerja kernel yang lebih baik.
- b. Disarankan untuk melakukan penelitian lanjutan dengan mempertimbangkan lebih banyak fitur dan mencoba metode klasifikasi yang berbeda seperti *Random Forest* guna mendapatkan hasil yang lebih akurat dalam mendiagnosis stunting pada anak.

DAFTAR PUSTAKA

- Ministry of Research and Technology of the republic of Indonesia, "Prioritas riset nasional," no. July, pp. 1–23, 2020.
- J. R. Khan, "Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh," *Informatics Heal. Soc. Care*, vol. 46, no. 4, pp. 425–442, 2021, doi: 10.1080/17538157.2021.1904938.
- S. Y. Andriyani, M. S. Lydia, and S. Efendi, "Optimization of Support Vector Machine Algorithm Using Stunting Data Classification," *Prism. Sains J. Pengkaj. Ilmu dan Pembelajaran Mat. dan IPA IKIP Mataram*, vol. 11, no. 1, p. 164, 2023, doi: 10.33394/j-ps.v11i1.6619.
- M. Yunus, M. K. Biddinika, and A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *J. Online Inform.*, vol. 8, no. 1, pp. 99–106, 2023, doi: 10.15575/join.v8i1.1062.
- M. Ula, A. F. Ulva, M. A. Ali, and Y. Rilasmi, "APPLICATION OF MACHINE LEARNING IN DETERMINING THE CLASSIFICATION OF CHILDREN ' S NUTRITION WITH DECISION TREE PENERAPAN MACHINE LEARNING DALAM PENENTUAN KLASIFIKASI GIZI," vol. 3, no. 5, pp. 1457–1465, 2022.
- R. Anggriawan and H. W. Nugroho, "Komparasi Algoritma C4.5 dan Naive Bayes Dalam Prediksi Penderita Penyakit Gagal Jantung," *J. SIMADA (Sistem Inf. dan Manaj. Basis Data)*, vol. 6, no. 1, pp. 82–91, 2023, doi: 10.30873/simada.v6i1.3425.

- E. R. Arumi, Sumarno Adi Subrata, and Anisa Rahmawati, "Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 2, pp. 201–207, 2023, doi: 10.29207/resti.v7i2.4438.
- V. Herliansyah, R. Latuconsina, and A. Dinimaharawati, "Prediksi Stunting Pada Balita Dengan Menggunakan Algoritma Klasifikasi Random Forest," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 6643–6649, 2021.
- D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, "Machine Learning in Nutrition Research," *Adv. Nutr.*, vol. 13, no. 6, pp. 2573–2589, 2022, doi: 10.1093/advances/nmac103.
- X. Fu *et al.*, "Water quality prediction of copper-molybdenum mining-beneficiation wastewater based on the PSO-SVR model," ... *Sci. & Engineering*, 2023, doi: 10.1007/s11783-023-1698-9.
- P. Migoń, "Disentangling polygenetic relief of low mountains at the margin of inland glaciation – Upper Nysa Szalona drainage basin, Sudetes, Central Europe," *Catena*, vol. 204, 2021, doi: 10.1016/j.catena.2021.105383.
- A. Yaghoubzadeh-Bavandpour, M. Rajabi, and ..., "Support Vector Machine Applications in Water and Environmental Sciences," ... *Intell. Water ...*, 2022, doi: 10.1007/978-981-19-2519-1_14.
- F. Yun, H. Dong, C. Liang, T. Weimin, and T. Chao, "Feature Selection of XLPE Cable Condition Diagnosis Based on PSO-SVM," *Arab. J. Sci. ...*, 2023, doi: 10.1007/s13369-022-07175-9.
- D. M. D. Raj and R. Mohanasundaram, "An Efficient Filter-Based Feature Selection Model to Identify Significant Features from High-Dimensional Microarray Data," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 2619–2630, 2020, doi: 10.1007/s13369-020-04380-2.
- H. I. Hussein, S. A. Anwar, and M. I. Ahmad, "Imbalanced Data Classification Using SVM Based on Improved Simulated Annealing Featuring Synthetic Data Generation and Reduction," 2023, doi: 10.32604/cmc.2023.036025.
- E. A. Mahareek, A. S. Desuky, and H. A. El-Zhni, "Simulated annealing for svm parameters optimization in student's performance prediction," *Bull. Electr. Eng. Informatics*, vol. 10, no. 3, pp. 1211–1219, 2021, doi: 10.11591/eei.v10i3.2855.

- A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," vol. 13, no. 2, 2023, doi: 10.18178/ijiet.2023.13.2.1806.
- I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, "The effect of gamma value on support vector machine performance with different kernels," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5497–5506, 2020, doi: 10.11591/IJECE.V10I5.PP5497-5506.
- I. Rahmi, M. Susanti, H. Yozza, and F. Wulandari, "Classification of Stunting in Children Under Five Years in Padang City Using Support Vector Machine," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 3, pp. 771–778, 2022, doi: 10.30598/barekengvol16iss3pp771-778.
- T. Azhima, Y. Siswa, and N. Azmi, "Optimasi Algoritma SVM dengan Chi-Square dan SMOTE untuk High Dimension Data Stunting Kota Samarinda," vol. xx, no. 200, pp. 1–10, 2021.
- Fadellia Azzahra, N. Suarna, and Y. Arie Wijaya, "Penerapan Algoritma Random Forest Dan Cross Validation Untuk Prediksi Data Stunting," *Kopertip J. Ilm. Manaj. Inform. dan Komput.*, vol. 8, no. 1, pp. 1–6, 2024, doi: 10.32485/kopertip.v8i1.238.
- Gina Purnama Insany, Indra Yustiana, and Sri Rahmawati, "Penerapan KNN dan ANN pada klasifikasi status gizi balita berdasarkan indeks antropometri," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 2, pp. 385–393, 2023, doi: 10.37859/coscitech.v4i2.5079.