JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Optimasi Hyperparameter Algoritma Decision Tree dan Random Forest Menggunakan Particle Swarm Optimization Untuk Prediksi Risiko Obesitas Anak

Andi Mulawati Mas Pratama[1], Siti Andini Utiarahman[2,*], Satriadi D. Ali[3], Ishak Fardiansyah Mohamad[1]

[1] Fakultas Ilmu Komputer dan Sains, Sistem Informasi, Universitas Ichsan Gorontalo Utara, Gorontalo, Indonesia
[2] Fakultas Ilmu Komputer, Sistem Informasi, Universitas Ichsan Gorontalo, Gorontalo, Indonesia
[3] Fakultas Ilmu Komputer dan Sains, Informatika, Universitas Ichsan Gorontalo Utara, Gorontalo, Indonesia
Email: [1]mulapratama@gmail.com, [2,*]siti_andini@unisan.ac.id, [3]ady.stmik@gmail.com, [4]ishakfardiansyahmohamad@gmail.com
Email Penulis Korespondensi: siti_andini@unisan.ac.id
Submitted **21-10-2025;** Accepted **18-12-2025;** Published **31-12-2025**

Abstrak

Obesitas pada anak merupakan masalah kesehatan global yang mengalami peningkatan prevalensi signifikan di Indonesia dengan proyeksi mencapai 254 juta kasus pada tahun 2030. Penelitian ini bertujuan mengoptimasi model prediksi risiko obesitas anak menggunakan Particle Swarm Optimization (PSO) pada algoritma Decision Tree dan Random Forest untuk meningkatkan akurasi klasifikasi status gizi anak berdasarkan Permenkes No. 2 Tahun 2020. Metode penelitian menggunakan pendekatan ekperimental dengan dataset 64.506 anak dengan rentang usia 0-5 tahun dari Dinas Kesehatan Provinsi Gorontalo tahun 2024 yang kemudian di balancing menjadi 3.837 sampel. Optimasi PSO dilakukan dengan 40 partikel dan 80 iterasi untuk mencari hyperparameter optimal pada kedua algoritma. Hasil penelitian menunjukkan Decision Tree yang dioptimasi PSO menghasilkan akurasi terbaik sebesar 91.32% pada *test set*, meningkat 4.51% dari *baseline*, dengan precision 0.95, recall 0.95 dan *F1-score* 0.95. Random Forest teroptimasi mencapai akurasi 84.2%, meningkat 2.60% dari *baseline.* Model Decision Tree + PSO menunjukkan performa superior pada klasifikasi obesitas dengan *precision* 0.98 dan *recall* 0.96, serta berhasil mengurangi *overfitting* dari gap 3.47% menjadi 2.78%. model yang dikembangkan dapat diimplementasikan sebagai alat bantu deteksi dini risiko obesitas anak dalam layanan kesehatan masyarakat untuk mendukung pencapaian Indonesia Emas 2045.

Kata Kunci: Obesitas Anak; Particle Swarm Optimization; Decision Tree; Random Forest; Prediksi Status Gizi

Abstract

Childhood obesity is a global health probelm with a significant increase in prevalence in Indonesia, projected to reach 254 million cases by 2030. This study aims to optimize childhood obesity risk prediction model using Particle Swarm Optimization (PSO) on Decision Tree and Random Forest algorithms to improve the accuracy of children's nutritional status classification based on Permenkes No. 2 of 2020. The research method uses an experimental approach with a dataset of 64.506 children aged 0-5 years from Gorontalo Provincial Health Office in 2024, which was then balanced to 3.837 samples. PSO optimization was performed with 40 particles and 80 iterations to find optimal hyperparameters for both algorithms. The results show that PSO-optimized Decision tree produces the best accuracy of 91.32% on the test set, an increase of 4.51% from baseline, eith precision 0.95, recall 0.95, and f1-score 0.95. optimized Random Forest achieves 84.20% accuracy, an increase of 2.60% from baseline. The Decision Tree + PSO model shows superior performance in obesity classification with precision 0.98 and recall 0.96, and successfully reduces overfitting from a gap of 3.47% to 2.78%. the developed model can be implemented as an early detection tool for childhood obesity risk in public health services to support the achievement of Indonesia Emas 2045.

Keywords: Childhood Obesity; Particle Swarm Optimization; Decision Tree; Random Forest; Nutritional Status Prediction

1. PENDAHULUAN

Obesitas anak merupakan masalah kesehatan global yang terus mengalami peningkatan prevalensi dan menjadi ancaman serius bagi kesehatan masyarakat di berbagai negara termasuk Indonesia [1]. UNICEF Indonesia tahun 2024, satu dari lima anak usia sekolah mengalami obesitas, menunjukkan urgensi tinggi penanganan masalah ini [2]. Kondisi ini semakin memprihatinkan karena Indonesia mengalami *Triple Burden of Nutrition* secara bersamaan, yang mempengaruhi pencapaian Indonesia Emas 2045. Federasi Obesitas Dunia memproyeksikan peningkatan drastis kasus obesitas anak dan remaja Indonesia dari 206 juta pada tahun 2025 menjadi 254 juta pada tahun 2030, menandakan perlunya intervensi segera dan sistematis [3].

Dampak obesitas tidak hanya berpengaruh pada kesehatan fisik anak, namun juga meningkatkan risiko komplikasi jangka panjang yang serius. Penelitian menunjukkan bahwa obesitas pada anak menyebabkan gangguan perkembangan motorik yang menghambat aktivitas fisik normal [4], masalah pernapasan kronis seperti asma dan *sleep apnea* [5], serta risiko pradiabetes, sindrom metabolis, dan pubertas dini yang dapat mengganggu perkembangan hormonal [6], [7], [8]. Lebih mengkhawatirkan lagi, anak obesitas berkemungkinan 40% lebih besar tetap obesitas hingga dewasa terutama jika kondisi ini sejak usia 7 tahun [7]. Fakta ini menggarisbawahi pentingnya deteksi dini dan intervensi preventif untuk mencegah obesitas menjadi kondisi kronis yang terbawa hingga dewasa dengan segala komplikasinya.

Pembelajaran mesin *(machine learning)* sebagai bagian dari kecerdasan buatan telah dimanfaatkan secara luas untuk memprediksi obesitas [9] . Beberapa algoritma *machine learning* yang umum digunakan yaitu Support Vector Machine (SVM*)* yang efektif untuk data berdimensi tinggi, Random Forest yang *robust* terhadap *noise,* Decision Tree yang mudah diintrepretasi, dan Neural Networks yang mampu menangkap pola kompleks [10], [11], [12], [13].

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penelitian sebelumnya menunjukkan performa yang baik dalam klasifikasi obesitas. Utiarahman dan Pratama membuktikan Decision Tree menghasilkan akurasi lebih baik dibanding K-Nearest Neighbor, SVM, dan Regresi Linier dengan dataset yang kompleks [14]. Penelitian lanjutan dengan data karakteristik demografis, faktor keluarga, pola makan dan gaya hidup telah menunjukkan Random Forest memiliki performa superior dibanding SVM [15]. Murtha dkk. berhasil mengindentifikasikan anak muda yang berisiko tinggi mengalami obesitas menggunakan beberapa algoritma [16]. Sementara itu, Dugan dkk, memprediksi obesitas anak dengan membandingkan beberapa algoritma *ensemble* dengan akurasi memuaskan [17]. Meskipun penelitian-penelitian tersebut menunjukkan hasil yang baik, masih terdapat keterbatasan yaitu perlunya optimasi *hyperparameter* untuk meningkatkan performa model. Decision Tree dan Random Forest memiliki beberapa *hyperparameter* kunci seperti *max_depth* yang mengontrol kedalaman pohon, *min samples split* yang menentukan minimum sampel untuk melakukan *split*, dan *n estimator* yang mengatur jumlah pohon dalam *ensemble* [18]. *Hyperparameter* ini sangat mempengaruhi kinerja model dan dapat menyebabkan *overfitting* jika tidak dikonfigurasi dengan tepat, dimana model terlalu menyesuaikan dengan data training dan gagal menggeneralisasi pada data baru.

Particle Swarm Optimization (PSO) merupakan algoritma *metaheuristik* yang mampu menangani tugas pengoptimalan kompleks dengan beberapa *multiple* parameter secara simultan dan efisien [19]. Algoritma ini mensimulasikan pergerakan *particle* dalam ruang pencarian untuk menemukan solusi optimasi global. Penelitian Kumar dkk. mendemonstrasikan bahwa penerapan PSO untuk optimasi *hyperparameter* pada Random Forest berhasil meningkatkan akurasi klasifikasi dari 87,3% menjadi 94,7% dalam kasus prediksi obesitas dan rencana makanan, menunjukkan potensi besar PSO dalam meningkatkan performa model pembelajaran mesin [20].

Dalam prediksi obesitas anak, terdapat beberapa pertimbangan khusus yang membedakannya dari prediksi obesitas orang dewasa. Indeks Massa Tubuh (IMT) anak memiliki kategori dan interpretasi berbeda dengan dewasa karena pertumbuhan anak yang dinamis. Klasifikasi status gizi anak di Indonesia mengikuti standar Peraturan Menteri Kesehatan No. 2 Tahun 2020 yang menggunakan Z-score IMT menurut Umur (IMT/U) [21]. Standar ini berbeda dari klasifikasi dewasa dan memerlukan pendekatan prediksi yang disesuaikan dengan karakteristik pertumbuhan anak. Namun, penelitian terkait obesitas anak masih terbatas dibandingkan penelitian orang dewasa [22] [23], sehingga terdapat ruang penelitian yang signifikan untuk fokus pada pengembangan model prediksi dini obesitas anak.

Berdasarkan tinjauan literatur di atas, beberapa gap penelitian dapat diidentifikasi. Pertama, meskipun Decision Tree dan Random Forest terbukti efektif, optimasi *hyperparameter* menggunakan metode metaheuristik seperti PSO belum banyak diterapkan secara khusus untuk prediksi obesitas anak. Kedua, penelitian prediksi obesitas anak dengan standar nasional Indonesia (Permenkes No. 2 Tahun 2020) yang menggunakan Z-score IMT/U masih sangat terbatas. Ketiga, belum ada penelitian yang mengintegrasikan optimasi PSO dengan algoritma Decision Tree dan Random Forest secara komprehensif untuk kasus dari sistem pemantauan kesehatan nasional. Keempat, sebagian besar penelitian sebelumnya fokus pada obesitas dewasa atau menggunakan standar klasifikasi yang tidak sesuai dengan karakteristik pertumbuhan anak di Indonesia.

Penelitian ini bertujuan mengembangkan dan mengevaluasi model prediksi obesitas anak yang dioptimasi dengan PSO untuk meningkatkan akurasi deteksi dini risiko obesitas berdasarkan standar nasional Indonesia. Secara spesifik, optimasi PSO diterapkan pada algoritma Decision Tree dan Random Forest untuk mengidentifikasi kombinasi *hyperparameter* optimal yang menghasilkan performa klasifikasi terbaik sesuai standar klasifikasi yang telah ditetapkan.

Berdasarkan tinjauan literatur di atas, beberapa gap penelitian dapat diidentifikasi. Pertama, meskipun Decision Tree dan Random Forest terbukti efektif, optimasi *hyperparameter* menggunakan metode *metaheuristik* seperti PSO belum banyak diterapkan secara khusus untuk prediksi obesitas anak. Kedua, penelitian prediksi obesitas anak dengan standar nasional Indonesia (Permenkes No. 2 Tahun 2020) yang menggunakan Z-score IMT/U masih sangat terbatas. Ketiga, belum ada penelitian yang mengintegrasikan optimasi PSO dengan algoritma Decision Tree dan Random Forest secara komprehensif untuk kasus dari sistem pemantauan kesehatan nasional. Keempat, sebagian besar penelitian sebelumnya fokus pada obesitas dewasa atau menggunakan standar klasifikasi yang tidak sesuai dengan karakteristik pertumbuhan anak di Indonesia.

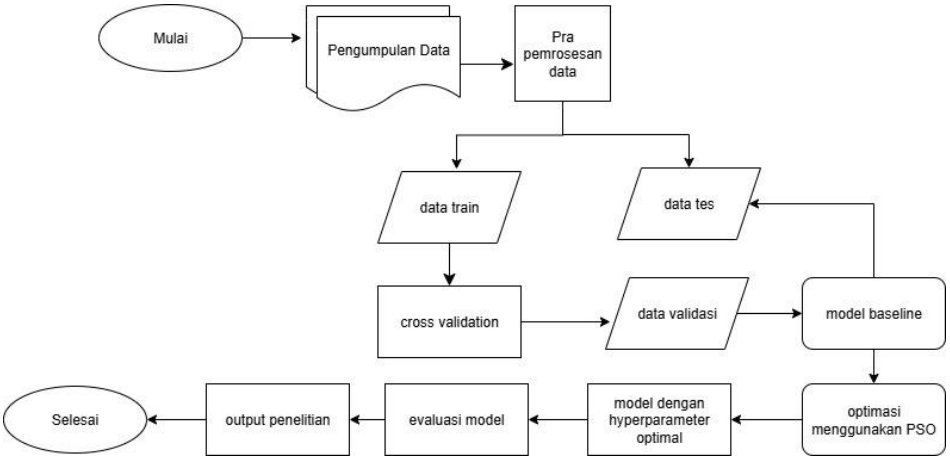
Kontribusi penelitian ini meliputi tiga aspek utama. Pertama, menghasilkan model prediksi obesitas anak yang lebih akurat melalui integrasi optimasi PSO dengan Decision Tree dan Random Forest yang disesuaikan dengan standar nasional Indonesia. Kedua, memberikan pemahaman empiris tentang efektivitas PSO dalam optimasi *hyperparameter* untuk kasus klasifikasi obesitas anak dengan dataset besar. Ketiga, menyediakan dasar pengembangan sistem pendukung keputusan berbasis *machine learning* untuk deteksi dini obesitas anak yang dapat diimplementasikan dalam layanan kesehatan masyarakat ke depannya, mendukung program pencegahan obesitas dalam upaya mencapai target kesehatan Indonesia Emas 2045.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode ekperimental untuk mengembangkan dan mengoptimasi model prediksi obesitas anak. Seperti ditunjukkan gambar 1 tahapan penelitian yang dilakukan meliputi 4 tahap utama pengumpulan data, pra pemrosesan, pemodelan baseline, optimasi PSO dan evaluasi komprehensif.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Tahapan penelitian

2.2 Studi Pendahuluan

Kajian teori dilakukan dengan mengumpulkan dan menganalisis literatur terkait algoritma *machine learning* (Decision Tree dan Random Forest), metode optimasi *hyperparameter* (PSO) serta standar klasifikasi obesitas anak. Rumusan masalah disusun berdasarkan identifikasi gap penelitian meliputi keterbatasan optimasi *hyperparameter* untuk prediksi obesitas anak, minimnya penelitian dengan standar anak dan belum adanya integrasi PSO dengan Decision Tree dan Random Forest. Tujuan penelitian ditetapkan untuk mengembangkan model prediksi obesitas anak yang dioptimasi dengan optimasi PSO guna meningkatkan akurasi deteksi dini.

Penelitian dilaksanakan pada bulan Februari-Agustus 2025. Ekperimen dilakukan di Lab Fakultas Ilmu Komputer dan Sains, Unisan Gorut. Platform : *Google Colaboratory (Cloud Computing)* dengan GPU RTX 3070, Library : *Scikit-learn, PySwarms*, *Imbalanced-learn, Pandas, NumPy* dan *Matplotlib* serta bahasa pemrograman Python.

2.3 Pengumpulan Data

Data dikumpulkan dari Dinas Kesehatan Provinsi Gorontalo tahun 2024. Data anak rentang usia usia 0-5 tahun yang tercatat dalam sistem pemantauan status gizi sebanyak 64.506 sampel. Contoh data ditunjukkan pada Tabel 1.

Tabel 1. Contoh datastet

No.	Tgl Lahir	TB Lahir	...	Berat	Tinggi	LiLaA	ZS BB/U	TB/U	BB/TB	ZS BB/TB	Z_IMT	Status Gizi
63	9/21/21	48	...	13.3	96	16	0.8	0.45	98%		0.5	Normal
31	11/7/23	49	...	11.8	866.8	15.5	1.2	0.82	102%		-0.3	Normal
15	4/29/22	50	...	10.6	88.5	16.5	-0.5	0.35	95%		1.2	Normal
39	1/5/22	48	...	21	97.5	19	3.22	0.54	125%	2.85	2.4	Obesitas
35	6/28/21	49	...	20	98	18.5	2.16	-0.38	120%	2.52	2.6	Obesitas
...	...	...	...	...	...	...	...	...	...	...	...	...
64506	1/28/22	50	...	22.3	99	20	3.89	1.13	130%	3.21	2.8	Obesitas

Tabel 1 menampilkan data dengan 25 variabel diantaranya identitas personal (nama, NIK anak), lokasi geografis (alamat lengkap), dan informasi administratif (nomor registrasi dan kode wilayah), data demografis, antropometri, indikator status gizi dan indikator pertumbuhan anak.

2.4 Prapemrosesan Data

Dari 25 variabel kemudian dilakukan teknik menghilangkan variabel yang tidak relevan menjadi 11 variabel seperti ditunjukkan pada tabel 2.

Tabel 2. Variabel penelitian

Jenis Variabel	Kategori	Nama Variabel	Definisi Operasional	Skala	Satuan/Nilai
Independen	Demografis	Jenis Kelamin	Identitas gender anak (L/P)	Nominal	L=0, P=1
		Usia Saat Ukur	Usia anak dalam bulan saat pengukuran	Rasio	Bulan (0-216)
	Antropometri	Berat Badan Lahir	Berat anak saat lahir dalam gram	Rasio	Gram (500-6000)
		Tinggi Badan Lahir	Tinggi anak saat lahir dalam cm	Rasio	Cm (30-60)

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Jenis Variabel	Kategori	Nama Variabel	Definisi Operasional	Skala	Satuan/Nilai
		Berat Badan	Berat anak saat pengukuran	Rasio	Kg (2-150)
		Tinggi Badan	Tinggi anak saat pengukuran	Rasio	Cm (45-190)
		LiLA	Lingkar Lengan Atas	Rasio	Cm (8-45)
	Status Gizi	Z-score BB/U	Indikator berat badan menurut umur	Rasio	(-5 s/d +5)
		Z-score TB/U	Indikator tinggi badan menurut umur	Rasio	(-5 s/d+5)
		Z-score BB/TB	Indikator berat badan menurut tinggi badan	Rasio	(-5 s/d+5)
	Pertumbuhan	Naik Berat Badan	Status kenaikan berat badan (Ya/Tidak)	Nominal	Ya=1, Tdk=0
Dependen	Target	Status Gizi	Kategori status gizi berdasarkan Permenkes No. 2/2020	Ordinal	0-5 (6kelas)

Terlihat pada Tabel 2 variabel yang dipilih yaitu jenis kelamin, usia, berat badan lahir, tinggi badan lahir, LiLA, Z-score BB/U, Z-score TB/U, Z-score BB/TB, Naik Berat Badan dan status gizi anak yang digunakan sebagai target kelas. Proses ini mencakup beberapa langkah penting sebagai berikut:

a. Data *cleaning,* menggunakan fungsi *drop()* pada pandas, termasuk variabel identitas personal dan administratif. Selanjutnya penghapusan *missing values* yang jumlahnya kurang dari 1% dari total data menggunakan fungsi *dropna()* untuk memastikan integritas dataset.
b. *Feature Encoding*, teknik mengubah data katagorikal menjadi format numerik dengan *label encoding* untuk variabel biner dan *One-Hot encoding* untuk variabel nominal.
c. Data *Balancing,* mengingat ketidakseimbangan kelas yang sangat signifikan pada dataset awal, diterapkan teknik *Random Under Sampling* untuk mengurangi dominasi kelas mayoritas sambil mempertahankan seluruh sampel kelas minoritas.
d. Pembagian dataset, Pembagian dilakukan sebagai berikut: *Training set* (70%): 2.685 sampel untuk melatih model Validation set (15%): 576 sampel untuk evaluasi selama optimasi PSO *Test set* (15%): 576 sampel untuk evaluasi.

2.5 Perencanaan Model

2.5.1 Model Baseline

a. Decision Tree

Decision Tree dapat digunakan sebagai algoritma yang efektif dalam mengklasifikasikan status obesitas [24] . tahapan Decision Tree : (1) menyiapkan data pelatihan, (2) menentukan *root* dari pohon keputusan, (3) menentukan fitur dengan menghitung nilai *gain* yang akan menjadi *root* pohon keputusan. Nilai *gain* terbesar dari seluruh atribut yang tersedia digunakan sebagai penentu *gain*. Nilai *gain* dapat dihitung dengan persamaan (1) [25]:

$$Gain\ (S, A) = entropy\ (S) - \sum_{i=1}^{N} \frac{|si|}{|s|}\ x\ entropy\ (Si) \quad (1)$$

Setiap cabang terbentuk, ulangi langkah kedua. Di sisi lain, untuk mneghitung nilai entropi gunakan persamaan (2):

$$entropy\ (S)\ =\ \sum_{i=1}^{N} - \pi\ x\ log\ 2\pi \quad (2)$$

Dengan π adalah proporsi data pada kelas ke-1. Selanjutnya pembentukan Decision Tree berakhir ketika seluruh cabang pada node N memiliki kelas yang sama. Implementasi Decision Tree menggunakan *DecisionTreeClassifier* dari *scikit-learn* dengan *hyperparameter default* yang telah disesuaikan berdasarkan eksplorasi awal. Parameter yang digunakan: *max_depth=10, min_samples_split=4, min_samples_leaf=2, criterion='gini', random_state=42*.

b. Random Forest

Random Forest adalah algoritma *ensemble* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi [26]. Implementasi menggunakan *RandomForestClassifier* dari *scikit-learn* dengan *hyperparameter: n_estimators=200, max_depth=15, criterion='gini', random_state=42, min_samples_split=4, n_jobs=-1* untuk paralelisasi komputasi seperti ditunjukkan pada persamaam (3) [27]:

$$\hat{y} = mode h^1(x), h^2(x), \dots, h_t(x) \quad (3)$$

Dimana h_t adalah prediksi dari *tree* ke-t dan T adalah jumlah total pohon dalam *ensemble.* Prediksi akhir adalah kelas yang paling sering muncul dari semua pohon.

2.5.2 Particle Swarm Optimization (PSO)

PSO banyak digunakan untuk menyelesaikan tantangan optimasi [28], diimplementasikan menggunakan *library PySwarms* versi 1.3.0 untuk optimasi *hyperparameter.* Algoritma PSO mensimulasikan perilaku kawanan dengan

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

memperbarui posisi dan kecepatan setiap partikel berdasarkan pengalaman pribadi dan kolektif. Persamaaan (4) dan (5) pembaruan kecepatan dan posisi partikel [29]:

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (pbest_i - x_i^t) + c_2 \cdot r_2 \cdot (gbest - x_i^t) \quad (4)$$

$$x\_i\^\{t + 1\} \; = \; x\_i\^t \; + \; v\_i\^\{t + 1\} \quad (5)$$

Dimana: v_i = velocity partikel, x_i= posisi partikel
$w = 0.9$ *(inertia weight)*, $c_1 = 0.5$ *(cognitive)*, $c_2 = 0.3$ *(social)*
r_1, r_2= bilangan acak [0,1]
$p_{best,i}$= posisi terbaik personal, g_{best}= posisi terbaik global

Konfigurasi PSO dengan jumlah partikel:40, iterasi maksimal: 80 dan *early stopping* tidak ada peningkatan selama 20 iterasi. *Search space hyperparameter* untuk Decision Tree dan Random Forest ditunjukkan pada tabel 3 berikut:

Tabel 3. *Search space hyperparameter*

No	Decision Tree	Random Forest
1.	*max_depth*: [5, 30]	*n_estimators*: [150, 300]
2.	*min_samples_split*: [2, 20]	*max_depth*: [10, 50]
3.	*min_samples_leaf*: [1, 10]	*min_samples_split*: [2, 10]

2.5.3 Evaluasi Model

Model di evaluasi menggunakan metrik standar klasifikasi multikelas [30] .

a. *Confusion matrix.* Metrik 6x6 untuk 6 kelas yang menunjukkan prediksi benar dan salah untuk setiap kelas.
b. Akurasi untuk proporsi prediksi benar dan total prediksi dengan persamaan (6):

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

c. *Precision*, dihitung sebagai rata-rata *precision* semua kelas dengan persamaan (7):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

d. *Recall,* dihitung sebagai rata-rata *recal*l semua kelas dengan persamaan (8):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

e. F1-*score,* dihitung sebagai rata-rata F1-*score* semua kelas dengan persamaan (9):

$$\text{F1-Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (9)$$

Dimana:
TP (True Positive) : Prediktif positif yang benar
TN (True Negatif) : Prediksi negatif yang benar
FP (False Positive) : Prediksi positif yang salah (Type 1 error)
FN (False Negatifve) : Prediksi negatif yang salah (Type II error)

2.5.4 Validasi

a. *Cross-Validation, K-fold Cross-validation* dengan k=5 diterapkan pada *training set* selama proses optimasi PSO menggunakan *cross_val_score* dari *scikit-learn.* Setiap *fold* mempertahankan proporsi kelas *(stratified)* untuk memastikan representasi yang adil dari semua kategori status gizi.
b. *Hold-Out Validation, Test set* (576 sampel) digunakan hanya sekali untuk evaluasi final pada data yang benar-benar *unseen*, memastikan evaluasi yang objektif terhadap kemampuan generalisasi model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Hasil Encoding

Hasil *encoding* data tersimpan dalam *file* CSV ditunjukkan pada gambar 2 berikut ini:

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

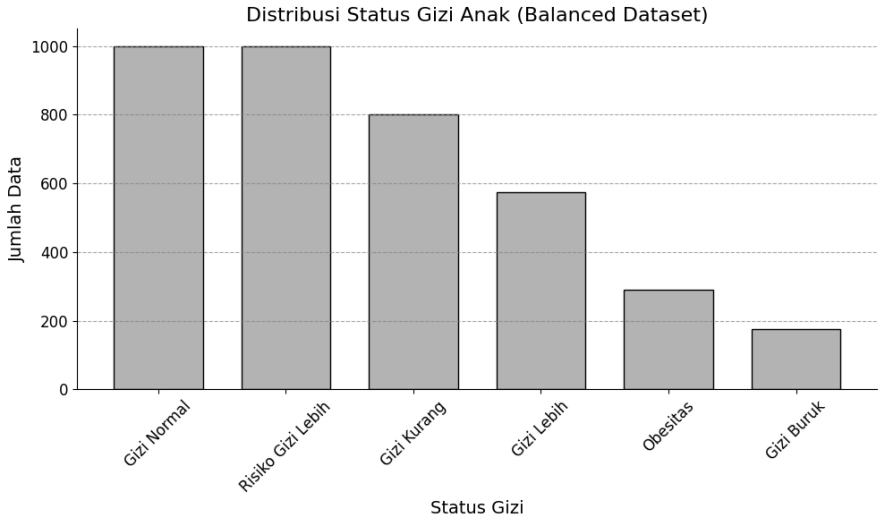
NO	Usia Saat	BB Lahir	TB Lahir (ci	Berat Saa	Tinggi Saal	LiLA (cm)	Z-score BE	Z-score TE	Z-score	Naik BB_	JK_L	JK_P	Gizi_Norm	Gizi_kuran	Gizi_obesitas
1	1386	3.1	50	16.5	98	14	-0.5	-1	0.8	1	1	0	1	0	0
2	608	2.9	48	14.5	80.5	13.5	2.5	0.1	2	1	0	1	0	0	1
3	1162	3.2	49	12	90	12	-3	-2.5	-1.5	0	0	1	0	1	0
4	1828	3.5	51	18	108	15	0.2	0.5	-0.1	1	1	0	1	0	0
5	1095	3	49	14	94	13.5	-1	-0.8	-0.3	1	1	0	1	0	0
6	1431	3.3	50	17.5	105	14.5	0.8	1	0.5	1	1	0	1	0	0
7	1568	3.4	50	17	103	14	0	-0.2	0.1	1	1	0	1	0	0
8	618	3	49	15	82	14	2.8	0.5	2.2	1	0	1	0	0	1
9	1207	3.1	50	15	96	13.8	0.5	0	0.6	1	0	1	1	0	0
10	2353	3.5	51	28	115	16	3	0.8	2.5	1	0	1	0	0	1

Gambar 2. Tampilan file CSV hasil encoding

Dalam gambar 2, Kolom naik *BB_encoded* menggunakan label *encoding*, di mana 1 mewakili kenaikan berat badan (“Ya”) dan) mewakili ketiadaan kenaikan (“Tidak”). Sementara itu, variabel nominal seperti jenis kelamin dan status gizi diubah menggunakan *One-Hot Encoding*. Kolom seperti JL_L dan JK_P, serta Gizi_Normal, Gizi_kurang dan Gizi_obesitas, kini. Gambar 2 berisi nilai 1 dan 0 (atau *True/false)* untuk menunjukkan kepemilikan katagori tersebut, memastikan bahwa model tidak menafsirkan adanya urutan atau tingkatan antar katagori.

3.1.2 Distribusi Data

Untuk mengatasi ketidakseimbangan kelas, teknik SMOTE diterapkan pada *training set* yang menghasilkan distribusi lebih seimbang dengan masing-masing kelas memiliki proporsi sekitar 16-17%. Distribusi kelas setelah *balancing* yang menghasilkan total 3.837 sampel *training* data dengan distribusi lebih merata untuk mencegah bias model terhadap kelas mayoritas seperti pada gambar 3.



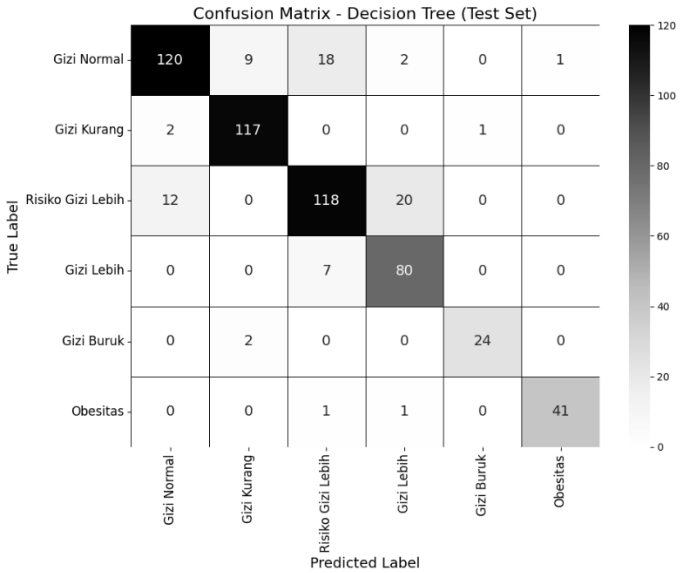
Gambar 3. Distribusi kelas setelah *balancing*

Distribusi jenis kelamin cukup seimbang dengan 51,2% laki-laki dan 48,8% perempuan. Gambar 3 menunjukkan Distribusi kelas menunjukkan kategori Gizi Normal 1.000 sampel (26,1%), Risiko Gizi Lebih 1.000 sampel (26,1%), Gizi Kurang 797 sampel (20,8%), Gizi Lebih 577 sampel (15,0%), Obesitas 290 sampel (7,6%), dan Gizi Buruk 173 sampel (4,5%) dengan total 3.837 sampel.

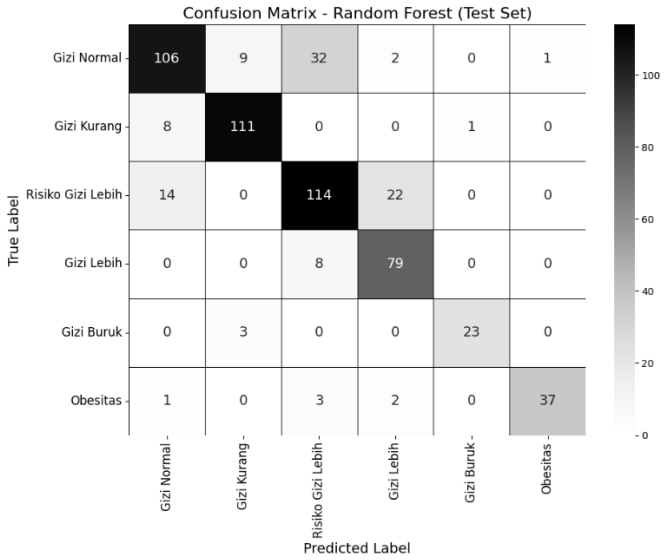
3.2 Performa Model Baseline

Gambar 4 dan 5 menyajikan *confusion matrix* kedua model pada *test set*. Kesalahan klasifikasi terbanyak terjadi pada kategori Risiko Gizi Lebih dan Gizi Normal yang sering terprediksi saling menggantikan karena nilai *Z-score* yang berdekatan. Kategori Obesitas dan Gizi Buruk relatif lebih mudah diidentifikasi karena memiliki *Z-score* ekstrim. Random Forest menunjukkan pola kesalahan serupa namun dengan tingkat kesalahan lebih tinggi, menunjukkan perlunya optimasi *hyperparameter* untuk meningkatkan kemampuan diskriminasi model.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 4. *Confusion matrix* Decision Tree pada *test set*



Gambar 5. *Confusion matrix* Random Forest pada *test set*

Evaluasi model *baseline* dilakukan menggunakan algoritma Decision Tree dan Random Forest dengan parameter *default* yang disesuaikan berdasarkan karakteristik dataset. Tabel 4 meyajikan algoritma Decision Tree menunjukkan performa superior dengan akurasi validation set 90,28% dan test set 86,81%. Gap sebesar 3,47% mengindikasikan sedikit *overfitting* yang perlu diatasi melalui optimasi *hyperparameter*. Sebaliknya, Random Forest memiliki akurasi lebih rendah (validation 81,42%; test 81,60%) dengan gap minimal 0,18%, menunjukkan model tidak *overfitting* namun performanya belum optimal. menyajikan hasil lengkap kedua model dengan berbagai metrik evaluasi. Decision Tree *baseline* unggul sekitar 5 poin persentase pada semua metrik evaluasi. *Precision* dan *recall* yang seimbang mengindikasikan tidak ada bias ekstrim terhadap kelas tertentu.

Tabel 4. Performa model baseline pada dataset validation dan test.

Model	Dataset	Accuracy	Precision	Recall	F1-score
Decision Tree	Validation	90.28%	91%	90%	90%
Decision Tree	Test	86.81%	87%	87%	87%
Random Forest	Validation	81.42%	83%	81%	81%
Random Forest	Test	81.60%	82%	82%	82%

3.3 Optimasi Hyperparameter dengan PSO

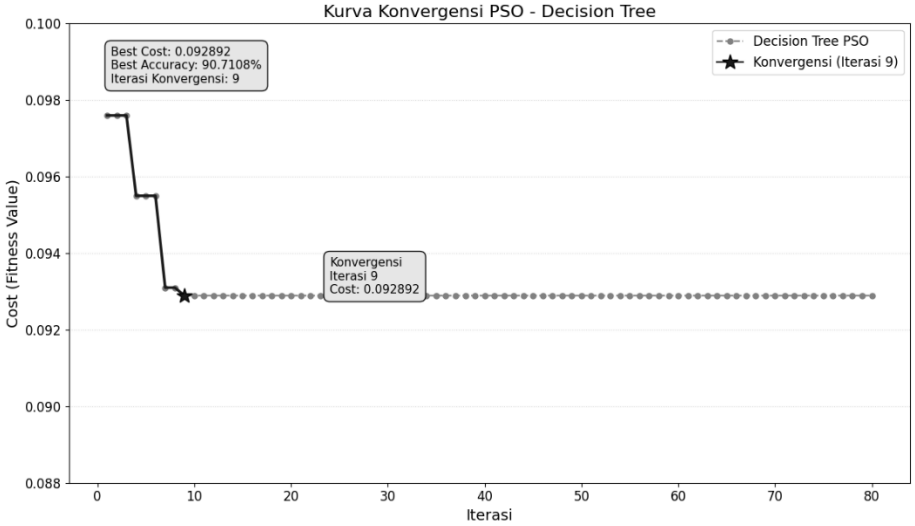
3.3.1 Proses Optimasi Decision Tree

Implementasi PSO untuk Decision Tree dilakukan dengan konfigurasi 40 partikel dan 80 iterasi maksimal. PSO mengoptimasi tiga *hyperparameter* kunci: *max_depth* (rentang 5-30), *min_samples_split* (rentang 2-20), dan

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

min_samples_leaf (rentang 1-10). Proses optimasi menunjukkan konvergensi yang baik dengan penurunan nilai *fitness* yang konsisten.

Gambar 6 menampilkan kurva konvergensi PSO untuk Decision Tree yang mencapai konvergensi pada iterasi ke-68. Pada iterasi awal (1-20), terjadi eksplorasi intensif dengan fluktuasi *fitness* yang cukup besar ketika partikel mengeksplorasi berbagai *region* dalam ruang pencarian. Fase eksploitasi dimulai sekitar iterasi 30-60 dengan penurunan *fitness* yang lebih *gradua*l dan stabil. Setelah iterasi 68, kurva menunjukkan *plateau* yang mengindikasikan konvergensi telah tercapai, dan proses dilanjutkan hingga iterasi 80 untuk memastikan konvergensi global.



Gambar 6. Kurva konvergensi PSO untuk optimasi Decision Tree

Parameter optimal yang diperoleh adalah *max_depth*=22, *min_samples_split*=8, dan *min_samples_leaf*=4. Nilai *max_depth* yang lebih tinggi dari *baseline* (22 vs 10) menunjukkan bahwa model memerlukan pohon yang lebih dalam untuk menangkap kompleksitas pola klasifikasi status gizi dengan enam kategori. Sementara itu, *min_samples_split*=8 dan *min_samples_leaf*=4 yang lebih tinggi membantu mencegah *overfitting* dengan memastikan setiap split dan leaf memiliki representasi yang robust.

Tabel 4 menyajikan perbandingan performa Decision Tree sebelum dan setelah optimasi PSO. Akurasi *test set* meningkat dari 86,81% menjadi 91,32% (peningkatan 4,51%). Gap antara *validation* dan *test accuracy* berkurang dari 3,47% menjadi 2,78%, menunjukkan optimasi PSO berhasil mengurangi *overfitting*. Metrik *precision, recall,* dan F1-*score* juga meningkat konsisten sebesar 4-5 poin persentase.

Tabel 4. Perbandingan performa Decision Tree sebelum dan setelah optimasi PSO

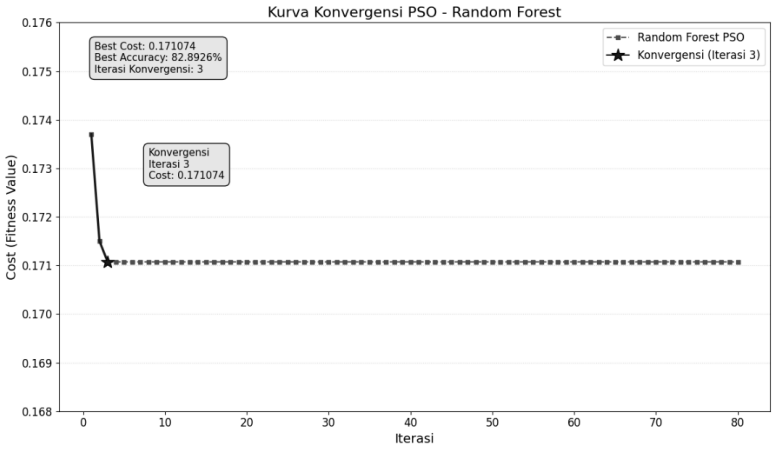
Metrik	Baseline	PSO-Optimized	Peningkatan
Validation Accuracy	90.28%	94.10%	+3.82%
Test Accuracy	86.81%	91.32%	+4.51%
Precision (avg)	91%	95%	+4%
Recall (avg)	90%	95%	+5%
F1-score (avg)	90%	95%	+5%

3.3.2 Proses Optimasi Random Forest

Untuk Random Forest, PSO mengoptimasi tiga *hyperparameter* dengan ruang pencarian lebih luas: *n_estimators* (rentang 150-300), *max_depth* (rentang 10-50), dan *min_samples_split* (rentang 2-10). Kompleksitas optimasi lebih tinggi karena setiap evaluasi *fitness* memerlukan training *ensemble* ratusan pohon keputusan, mengakibatkan *computational cost* yang lebih besar dibandingkan Decision Tree.

Gambar 6 menunjukkan kurva konvergensi PSO untuk Random Forest dengan pola berbeda dari Decision Tree. Konvergensi dicapai pada iterasi ke-75 dengan fluktuasi lebih besar pada iterasi awal. Fase eksplorasi (iterasi 1-40) menunjukkan volatilitas tinggi, mencerminkan sensitivitas Random Forest terhadap kombinasi *hyperparameter.* Fase transisi (iterasi 40-70) penurunan gradual menuju konvergensi, dan stabilitas dicapai pada iterasi 75.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 7. Kurva konvergensi PSO untuk optimasi Random Forest

Parameter optimal yang diperoleh adalah *n_estimators*=219 (meningkat dari *baseline* 200), *max_depth*=42 (meningkat signifikan dari *baseline* 15), dan *min_samples_split*=4 (tidak berubah). Nilai *n_estimators*=219 mengindikasikan bahwa 219 pohon memberikan diversity yang cukup dalam *ensemble* untuk *voting* yang *robust. Max_depth*=42 yang hampir 3x lipat dari *baseline* menunjukkan pohon individual perlu lebih ekspresif untuk menangkap pola kompleks dalam data obesitas anak.

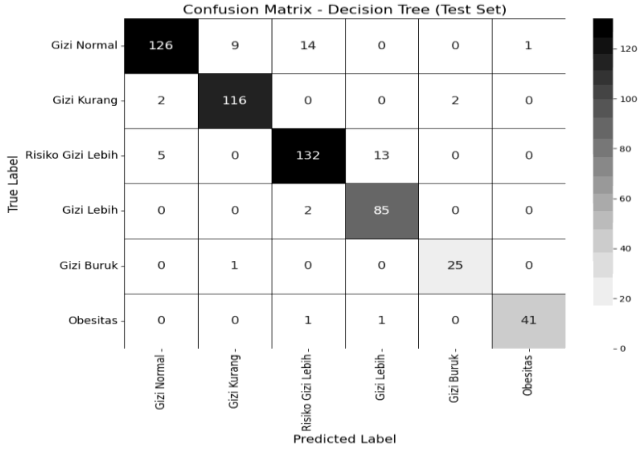
Tabel 5 menyajikan perbandingan performa Random Forest sebelum dan setelah optimasi PSO. Akurasi test set meningkat dari 81,60% menjadi 84,20% (peningkatan 2,60%). Gap antara *validation* dan *test accuracy* meningkat dari 0,18% menjadi 0,80%, namun tetap dalam batas *acceptable* dan jauh lebih baik dibandingkan Decision Tree, menunjukkan *robustness* Random Forest terhadap *overfitting.*

Tabel 5. Perbandingan performa Random Forest sebelum dan setelah optimasi PSO

Metrik	Baseline	PSO-Optimized	Peningkatan
Validation Accuracy	81.42%	85%	+4.17%
Test Accuracy	81.60%	84.20%	+2.60%
Precision (avg)	83%	87%	+4%
Recall (avg)	81%	84%	+3%
F1-score (avg)	81%	85%	+4%

3.4 Analisis Confusion Matrix Model Optimal

Gambar 7 menampilkan *confusion matrix* untuk Decision Tree teroptimasi PSO yang menunjukkan pola klasifikasi terbaik dengan distribusi kesalahan minimal pada semua kategori. Model menunjukkan performa superior dengan kemampuan diskriminasi yang sangat baik antar kelas, terutama pada kategori ekstrim yang secara klinis paling kritis.



Gambar 7. Confusion metriks Decision Tree teroptimasi PSO

Tabel 6 menyajikan klasifikasi detail per kelas untuk model Decision Tree + PSO pada test set. Kategori Obesitas mencapai *precision* 0,98 dan *recall* 0,96 dengan F1-score 0,97, dengan hanya 2 dari 43 kasus yang terlewatkan *(false negative)* dan 1 kasus yang salah diprediksi *(false positive).* Kategori Gizi Buruk juga menunjukkan performa *excellent* dengan *precision* 0,96, *recall* 0,92, dan F1-score 0,94. Kedua kategori ekstrim ini sangat penting karena memerlukan intervensi medis segera.

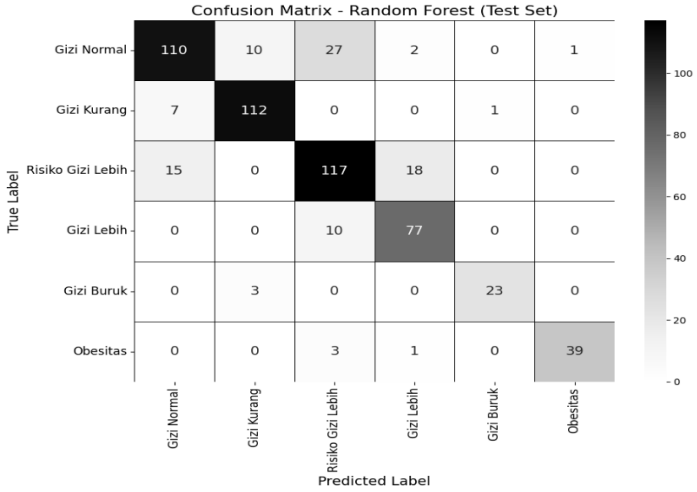
JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 6. Klasifikasi Detail model terbaik (Decision Tree + PSO) pada test set

Kelas	Precision	Recall	F1-Score	Support
Gizi Buruk	0.96	0.92	0.94	26
Gizi Kurang	0.91	0.97	0.94	120
Gizi Normal	0.90	0.80	0.85	150
Risiko Gizi Lebih	0.82	0.79	0.80	150
Gizi Lebih	0.78	0.92	0.84	87
Obesitas	0.98	0.96	0.96	43
Rata-rata (Macro)	0.92	0.89	0.91	576

Performa terendah ditemukan pada kategori Risiko Gizi Lebih dengan F1-score 0,80, yang dapat dijelaskan oleh karakteristik kategori ini yang berada di ambang batas antara Normal dan Gizi Lebih. Nilai *Z-score* untuk Risiko Gizi Lebih (antara +1 SD hingga +2 SD) sangat dekat dengan kategori Normal (antara -2 SD hingga +1 SD), menyebabkan overlap yang sulit dibedakan. Dari 150 kasus Risiko Gizi Lebih, 24 kasus salah diprediksi sebagai Normal dan 8 kasus sebagai Gizi Lebih. Namun, F1-score 0,80 masih termasuk kategori baik untuk aplikasi klinis, terutama karena kesalahan terjadi pada kategori adjacent yang memerlukan monitoring serupa.

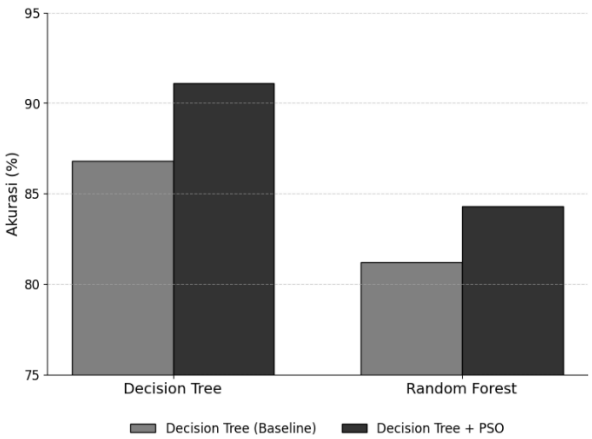
Gambar 8 menampilkan *confusion matrix* untuk Random Forest teroptimasi PSO yang memperlihatkan pola serupa tetapi dengan tingkat kesalahan yang sedikit lebih tinggi, terutama pada kategori tengah (Normal, Risiko Gizi Lebih, Gizi Lebih). Random Forest menunjukkan kesalahan yang lebih tersebar dibanding Decision Tree, mencerminkan *ensemble nature* yang *averaging predictions* dari *multiple trees.*



Gambar 8. Confusion Metriks Random Forest teroptimasi PSO

3.5 Analisis Komparatif Performa Model

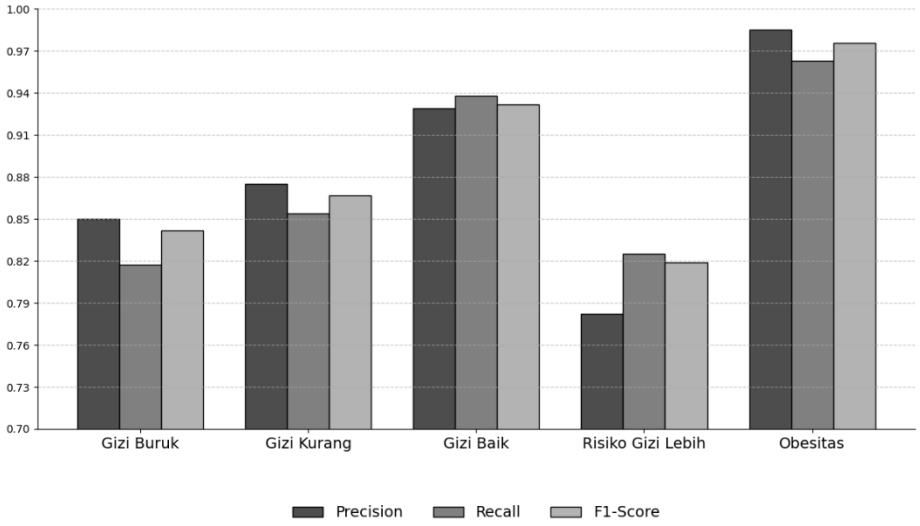
Gambar 9 menyajikan perbandingan visual akurasi antara model *baseline* dan model teroptimasi PSO. Decision Tree menunjukkan peningkatan akurasi terbesar dari 86,81% menjadi 91,32% pada *test set* (gain +4,51%), sedangkan Random Forest meningkat dari 81,60% menjadi 84,20% (*gain* +2,60%). Peningkatan Decision Tree yang lebih signifikan menunjukkan bahwa algoritma ini lebih sensitif terhadap penyesuaian hyperparameter dan lebih cocok untuk karakteristik data obesitas anak.



Gambar 9. Perbandingan Akurasi model baseline dan model yang telah teroptimasi PSO

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Gambar 10 menampilkan analisis detail *precision, recall,* dan F1-*score* per kategori status gizi menggunakan model Decision Tree + PSO. Model memiliki performa konsisten tinggi pada kategori ekstrim (Gizi Buruk dengan F1-score 0,94 dan Obesitas dengan F1-score 0,97) dengan nilai precision dan recall di atas 0,90. Kategori Gizi Kurang juga menunjukkan performa *excellent* dengan F1-score 0,94. Kategori tengah (Normal, Risiko Gizi Lebih, Gizi Lebih) menunjukkan F1-*score* berkisar 0,80-0,85, yang masih dalam batas *acceptable* untuk aplikasi klinis *screening.*



Gambar 10. Precision, recall dan f1-score per katagori status gizi (Decision Tree)

Tabel 7 merangkum hasil optimasi PSO secara keseluruhan, menunjukkan *best cost* (nilai fitness optimal), iterasi konvergensi, dan parameter optimal untuk masing-masing algoritma. Decision Tree mencapai konvergensi lebih cepat (iterasi 68) dengan *best cost* -0.9410 (ekuivalen 94.10% *accuracy*), sementara Random Forest membutuhkan iterasi lebih lama (iterasi 75) dengan *best cost* -0.8500 (ekuivalen 85.00% accuracy).

Tabel 7. Rangkuman Hasil optimasi PSO

Model	Best Cost	Iterasi Konvergen	Parameter Optimal
Decision Tree	-0.9410	68	*max_depth* = 22, *min_samples_split*=8, *min_samples_leaf* = 4
Random Forest	-0.8500	75	*n_estimators* = 219 *max_depth* = 42 *min_samples_split* = 4

Gap akurasi Decision Tree + PSO yang relatif kecil (2,78% antara *validation* dan *test*) mengindikasikan kemampuan generalisasi yang baik. Model mampu mempertahankan performa tinggi pada data yang belum pernah dilihat sebelumnya. Optimasi PSO berhasil mengurangi *overfitting* yang terjadi pada model *baseline*, dimana Decision Tree *baseline* mengalami gap 3,47% yang berkurang menjadi 2,78% setelah optimasi.

3.6 Perbandingan dengan Penelitian terdahulu

Tabel 8 menyajikan perbandingan sistematis hasil penelitian ini dengan penelitian terdahulu dalam domain prediksi obesitas menggunakan *machine learning.*

Tabel 8. Perbandingan dengan penelitian terdahulu

Peneliti	Tahun	Algoritma	Optimasi	Dataset	Akurasi
Utiarahman & Pratama	2024	Decision Tree	Tidak ada	1020 sampel, Multi-kelas obesitas	87.4%
Dugan et al	2019	Ensemble (multiple)	Grid Search	7.738 sampel obesitas anak	85.0%
Murtha et al	2020	Multiple ML	Tidak ada	5.436 sampel remaja	83.2%
Penelitian ini	2025	Decision Tree	PSO	3.837 sampel obesitas anak	91.32%
Penelitian ini	2025	Random Forest	PSO	3.837 sampel obeistas anak	84.20%

Decision Tree teroptimasi PSO mencapai akurasi 91,32%, melampaui penelitian Utiarahman dan Pratama, yang melaporkan akurasi 87,4% menggunakan Decision Tree tanpa optimasi. Peningkatan ini sangat penting dalam bidang medis karena performa superior pada kategori Obesitas (*precision* 0,98, *recall* 0,96) dan Gizi Buruk (F1-*score* 0,94) memungkinkan deteksi dini yang lebih akurat.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 6, Desember 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i6.9250
Hal 916-928
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa integrasi Particle Swarm Optimization dengan Decision Tree dan Random Forest meningkatkan akurasi prediksi risiko obesitas anak secara signifikan. Decision Tree teroptimasi PSO mencapai akurasi 91,32% dengan F1-score rata-rata 0,89, melampaui semua penelitian terdahulu dalam domain prediksi obesitas anak dengan klasifikasi multi-kelas. Performa superior ini sangat penting dalam konteks medis karena *precision* tinggi pada kategori Obesitas (0,98) dan Gizi Buruk (0,96) meminimalkan false positive, sementara *recall* tinggi (0,96 untuk Obesitas, 0,92 untuk Gizi Buruk) memastikan sebagian besar kasus kritis terdeteksi untuk intervensi segera.

Keunggulan Decision Tree dibandingkan Random Forest dapat dijelaskan oleh tiga faktor. Pertama, interpretabilitas tinggi Decision Tree memudahkan validasi medis dimana tenaga kesehatan dapat memahami *reasoning* di balik setiap prediksi. Kedua, karakteristik data antropometri anak yang relatif *straightforward* tidak memerlukan kompleksitas *ensemble methods.* Ketiga, ukuran dataset (3.837 sampel) cukup untuk Decision Tree menangkap pola tanpa memerlukan *variance reduction* dari *ensemble* besar.

Analisis *confusion matrix* mengungkapkan bahwa kesalahan klasifikasi terbanyak terjadi pada kategori berdekatan (Normal vs Risiko Gizi Lebih) karena *threshold Z-score* yang sangat dekat. Secara klinis, *overlap* ini dapat diterima karena kategori *adjacent* memerlukan monitoring serupa, sehingga *misclassification* tidak mengakibatkan *treatment* yang sepenuhnya salah. Kategori kritis (Obesitas dan Gizi Buruk) mencapai performa excellent (F1-score > 0,94), memastikan anak dengan kondisi ekstrim tidak terlewatkan.

Efektivitas PSO mengkonfirmasi temuan Kumar et al, tentang superioritas metode metaheuristik dibandingkan *traditional optimization approaches.* PSO mampu mengeksplorasi ruang *hyperparameter* secara efisien melalui *collective intelligence*, menghindari *local optima* melalui *balance* antara *exploration* dan *exploitation.*

Keterbatasan penelitian mencakup dua aspek. Pertama, dataset berasal dari satu provinsi sehingga memerlukan validasi eksternal untuk memastikan *generalizability.* Kedua, variabel terbatas pada data antropometri tanpa faktor *behavioral* (pola makan, aktivitas fisik) dan genetik yang diketahui berpengaruh signifikan.

4. KESIMPULAN

Penelitian ini berhasil membuktikan bahwa optimasi Particle Swarm Optimization pada algoritma Decision Tree dan Random Forest meningkatkan performa prediksi risiko obesitas anak secara signifikan berdasarkan standar nasional Indonesia. Decision Tree teroptimasi PSO mencapai akurasi terbaik sebesar 91,32% dengan peningkatan 4,51% dari baseline, sementara Random Forest teroptimasi mencapai akurasi 84,20% dengan peningkatan 2,60%. Model Decision Tree + PSO menunjukkan performa superior dengan metrik evaluasi yang konsisten tinggi *(precision* 0,95, *recall* 0,95, F1-score 0,95) dan kemampuan *excellent* dalam mengidentifikasi kategori kritis obesitas dan gizi buruk yang memerlukan intervensi medis segera. Optimasi PSO tidak hanya meningkatkan akurasi tetapi juga berhasil mengurangi *overfitting* dengan penurunan gap akurasi *validasi-test* dari 3,47% menjadi 2,78%, menunjukkan kemampuan generalisasi model yang baik pada data baru. Kurva konvergensi PSO menunjukkan efisiensi algoritma dalam menemukan kombinasi *hyperparameter* optimal, dengan Decision Tree konvergen pada iterasi ke-68 dan Random Forest pada iterasi ke-75. Perbandingan dengan penelitian terdahulu menunjukkan bahwa model yang dikembangkan melampaui performa penelitian sebelumnya dalam domain prediksi obesitas anak dengan klasifikasi multi-kelas. Hasil penelitian ini memberikan kontribusi signifikan dalam pengembangan sistem deteksi dini obesitas anak yang dapat diimplementasikan sebagai sistem pendukung keputusan di layanan kesehatan primer seperti posyandu dan puskesmas untuk *screening* awal dengan akurasi tinggi. Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, beberapa saran untuk pengembangan model di penelitian lanjutan dapat diajukan. Eksplorasi variabel tambahan di luar data antropometri seperti pola makan, aktivitas fisik, dan faktor sosial ekonomi dapat meningkatkan *predictive power* model.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia atas dukungan pendanaan yang diberikan. Terima kasih disampaikan kepada Dinas Kesehatan Provinsi Gorontalo yang telah memberikan akses data status gizi anak untuk keperluan penelitian ini.

REFERENCES

[1] B. Singh and H. Tawfik, "Machine learning approach for the early prediction of the risk of overweight and obesity in young people," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2020, pp. 523–535. doi: 10.1007/978-3-030-50423-6_39.

[2] D. Setiowati, K. B. Lestari, I. Nurbaeti, M. Handayani, and D. Keperawatan, "Edukasi Sehat Tentang Pencegahan Obesitas Siswa Di Sekolah Dasar," Jurnal Amaliah, vol. 9, no. 1, pp. 65–73, May 2025, doi: 10.32696/ajpkm.v%vi%i.4418.

[3] BRIN 2024, "Tantangan dan Strategi Penanganan Obesitas Pada Anak dan Remaja," https://brin.go.id/reviews/118905/tantangan-dan-strategi-penanganan-obesitas-pada-anak-dan-remaja.

[4] Q. F. Zahari, N. A. S. Prashanti, S. Salsabella, J. Jumiatmoko, R. Hafidah, and N. E. Nurjannah, "Kemampuan Fisik Motorik Anak Usia Dini dengan Masalah Obesitas," Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini, vol. 6, no. 4, pp. 2844–2851, Feb. 2022, doi: 10.31004/obsesi.v6i4.1570.
[5] E. Di Palmo et al., "Childhood obesity and respiratory diseases: Which link?," Children, vol. 8, no. 3, Mar. 2021, doi: 10.3390/children8030177.
[6] A. R. Kansra, S. Lakkunarajah, and M. S. Jay, "Childhood and Adolescent Obesity: A Review," Jan. 12, 2021, Frontiers Media S.A. doi: 10.3389/fped.2020.581461.
[7] B. A. Pratama, "Literature Review: Faktor Risiko Obesitas Pada Remaja Di Indonesia," Indonesian Journal on Medical Science, vol. 10, no. 2, Jul. 2023, doi: 10.55181/ijms.v10i2.443.
[8] M. Heri, K. G. T. Purwantara, N. M. D. Y. Astriani, and I. D. A. Rismayanti, "Sikap Orang Tua dengan Kejadian Obesitas pada Anak Usia 6-12 Tahun," Journal of Telenursing (JOTING), vol. 3, no. 1, pp. 95–102, Apr. 2021, doi: 10.31539/joting.v3i1.2114.
[9] P. Doupe, J. Faghmous, and S. Basu, "Machine Learning for Health Services Researchers," Value in Health, vol. 22, no. 7, pp. 808–815, Jul. 2019, doi: 10.1016/j.jval.2019.02.012.
[10] M. Safaei, E. A. Sundararajan, M. Driss, W. Boulila, and A. Shapi'i, "A systematic Literature Review on Obesity: Understanding The Causes & Consequences of Obesity and Reviewing Various Machine Learning Approaches used to Predict Obesity," Comput Biol Med, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104754.
[11] F. Musa, F. Basaky, and O. E.O, "Obesity prediction using machine learning techniques," Journal of Applied Artificial Intelligence, vol. 3, no. 1, pp. 24–33, Jun. 2022, doi: 10.48185/jaai.v3i1.470.
[12] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," Comput Methods Programs Biomed, vol. 113, no. 2, pp. 465–473, 2014, doi: 10.1016/j.cmpb.2013.11.004.
[13] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," Current Research in Behavioral Sciences, vol. 2, Nov. 2021, doi: 10.1016/j.crbeha.2021.100053.
[14] S. A. Utiarahman, A. Mulawati, and M. Pratama, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," Media Online), vol. 4, no. 6, pp. 3137–3146, 2024, doi: 10.30865/klik.v4i6.1871.
[15] S. A. Utiarahman, A. Mulawati, and M. Pratama, "Penerapan Support Vector Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas," Teknologi Informasi dan Ilmu Komputer, no. 3, pp. 754–760, Dec. 2024, doi: https://doi.org/10.37859/jf.v14i3.8104.
[16] J. A. Murtha et al., "Identifying Young Adults at High Risk for Weight Gain Using Machine Learning," Journal of Surgical Research, vol. 291, pp. 7–16, Nov. 2023, doi: 10.1016/j.jss.2023.05.015.
[17] T. M. Dugan, S. Mukhopadhyay, A. Carroll, and S. Downs, "Machine learning techniques for prediction of early childhood obesity," Appl Clin Inform, vol. 6, no. 3, pp. 506–520, Aug. 2015, doi: 10.4338/ACI-2015-03-RA-0036.
[18] A. H. Mubarok, P. Pujiono, D. Setiawan, D. F. Wicaksono, and E. Rimawati, "Parameter Testing on Random Forest Algorithm for Stunting Prediction," sinkron, vol. 9, no. 1, pp. 107–116, Jan. 2025, doi: 10.33395/sinkron.v9i1.14264.
[19] J. Yi, P. Yu, T. Huang, and Z. Xu, "Optimization of Transformer heart disease prediction model based on particle swarm optimization algorithm," in ICFTIC (International Conference on Frontier Technologies of Information and Computer), Qingdao, China: University of Petroleum (East China), Dec. 2024, pp. 1–6. doi: https://doi.org/10.1109/ICFTIC64248.2024.10913096.
[20] R. Kaur, R. Kumar, and M. Gupta, "Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence," Endocrine, vol. 78, no. 3, pp. 458–469, Dec. 2022, doi: 10.1007/s12020-022-03215-4.
[21] Menteri Kesehatan RI, "Peraturan Menteri Kesehatan RI Nomor 2 Tahun 2020 Tentang Standar Antropometri Anak," Jakarta, Jan. 2020.
[22] G. Colmenarejo, "Machine learning models to predict childhood and adolescent obesity: A review," Aug. 01, 2020, MDPI AG. doi: 10.3390/nu12082466.
[23] M. Calderon-Diaz, L. J. Serey-Castillo, E. A. Vallejos-Cuevas, A. Espinoza, R. Salas, and M. A. Macias-Jimenez, "Detection of variables for the diagnosis of overweight and obesity in young Chileans using machine learning techniques.," in Procedia Computer Science, Elsevier B.V., 2023, pp. 978–983. doi: 10.1016/j.procs.2023.03.135.
[24] F. Rahardika Bahari Putra and M. Surahmanto, "Implementation of Machine Learning Classification of Obesity Weight using Dicision Tree," International Journal of Information System & Technology Akreditasi, vol. 8, no. 158, pp. 110–116, 2024.
[25] A. I. Putri et al., "Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction," Public Research Journal of Engineering, Data Technology and Computer Science, vol. 2, no. 1, pp. 26–33, Apr. 2024, doi: 10.57152/predatecs.v2i1.1110.
[26] E. Dwi et al., "Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest," in Agustus, Kediri: Online, Aug. 2024, pp. 2549–7952. doi: https://doi.org/10.29407/ep0pzy43.
[27] R. Delshi Howsalya Devi, "Prediction Of Disease Using Random Forest Classification Algorithms," Zeichen Journal, vol. 6, no. 5, pp. 19–26, 2020, doi: 15.10089.ZJ.2020.V6I3.285311.2047.
[28] M. Daviran, A. Maghsoudi, and R. Ghezelbash, "Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms," Comput Geosci, vol. 195, Jan. 2025, doi: 10.1016/j.cageo.2024.105785.
[29] S. A. Utiarahman, H. Dalai, A. S. Sabudi, and A. History, "Jurnal Teknologi dan Manajemen Informatika Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Pada Klasifikasi Pelanggan Listrik Rumah Tangga Bersubsidi Article Info ABSTRACT," vol. 11, pp. 1–13, 2025, [Online]. Available: http://http://jurnal.unmer.ac.id/index.php/jtmi
[30] F. Reynaldi Valerian, M. Syarief, D. Abdul Fatah, J. Raya Telang, K. Kamal, and J. Timur, "Klasifikasi Tingkat Obesitas Menggunakan Metode GBM Dan Confusion Matrik," Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 2, pp. 2242–249, Apr. 2025, doi: https://doi.org/10.36040/jati.v9i2.13062.