
ANALISIS MODEL KLASIFIKASI DENGAN OPTIMASI PARTICLE SWARM OPTIMIZATION DALAM KLASIFIKASI STATUS GIZI ANAK**Nurwati¹, Rika Nofitri², Diana Selvi³****Universitas Royal, Kisaran**e-mail: ¹nurwati@gmail.com

Abstract: *Nutritional status represents a state of equilibrium in the form of specific variables or the manifestation of nutrition in the form of specific variables. Optimal nutritional status is achieved when there is a balance between nutrient intake and nutrient requirements. Malnutrition not only affects an individual's physical health but also negatively impacts their overall development and well-being. However, data warehouses related to child nutritional status in Asahan Regency from various community health centers (Puskesmas) have not been optimally utilized to generate valuable information. Therefore, the purpose of this study was to analyze child nutrition datasets using a machine learning-based classification model that can predict children's nutritional status early. The classification model was chosen because of its ability to group data or objects based on specific classes, making it suitable for future data prediction. Based on the results, it can be concluded that optimizing the toddler nutritional status classification model using the Particle Swarm Optimization (PSO) method and cross-validation was able to identify the best model, namely the Decision Tree, with an accuracy of 37.50%. Although the overall accuracy of all models was still relatively low. This indicates the need for further data processing, such as data balancing, selecting more relevant features, and further parameter tuning to improve classification performance.*

Keywords: *Classification; Particle_Swarm Optimization; Cross Validation, Nutritional Status*

Abstrak: Status gizi adalah representasi dari keadaan keseimbangan dalam bentuk variabel tertentu atau manifestasi nutrisi dalam bentuk variabel tertentu, di mana status gizi optimal dicapai ketika terjadi keseimbangan antara asupan dan kebutuhan zat gizi. Gizi buruk tidak hanya memengaruhi kesehatan fisik individu, tetapi juga berdampak negatif pada perkembangan dan kesejahteraan mereka secara keseluruhan. Akan tetapi, gudang data yang terkait status gizi anak di Kabupaten Asahan dari berbagai puskesmas belum dimanfaatkan dengan optimal untuk menghasilkan informasi yang berharga. Oleh karena itu, tujuan dari penelitian ini adalah untuk menganalisis dataset gizi anak menggunakan model klasifikasi berbasis machine learning yang dapat memprediksi status gizi anak secara dini. Model klasifikasi dipilih karena kemampuannya dalam mengelompokkan data atau objek berdasarkan kelas tertentu, sehingga cocok untuk prediksi data di masa depan. Berdasarkan hasil penelitian, dapat disimpulkan bahwa optimasi model klasifikasi status gizi balita menggunakan metode Particle Swarm Optimization (PSO) dan cross validation mampu mengidentifikasi model terbaik, yaitu Decision Tree dengan akurasi 37,50%, meskipun secara umum tingkat akurasi seluruh model masih relatif rendah. Hal ini menunjukkan perlunya pengolahan data lebih lanjut, seperti balancing data, pemilihan fitur yang lebih relevan, maupun tuning parameter lanjutan untuk meningkatkan performa klasifikasi.

Kata kunci: Klasifikasi; Particle Swarm Optimization; Cross Validation, Status Gizi

PENDAHULUAN

Perkembangan teknologi saat ini telah merambah ke segala aspek kehidupan. Beberapa bidang teknologi yang dapat dimanfaatkan bagi manusia salah satunya bidang kesehatan seperti pada klasifikasi status gizi. Status gizi mencerminkan kesetimbangan tertentu dalam bentuk variabel dan ketika gizi optimal tercapai, itu menandakan keselarasan antara asupan dan kebutuhan nutrisi. Menurut Riset Kesehatan Dasar (Kemenkes, 2018), prevalensi masalah gizi pada anak usia 5 - 12 tahun di Indonesia mencakup kurang gizi sebesar 9,3%, kegemukan 20,6%, dan kerdil 23,6% (Hasrul, Hamzah, & Hafid, 2021). Kondisi ini dapat menyebabkan gangguan pertumbuhan, penyakit menular, gangguan kognitif, keterlambatan mental, gangguan kesehatan kronis, hingga kematian. Gizi buruk tidak hanya memengaruhi kesehatan fisik individu, tetapi juga berdampak negatif pada perkembangan dan kesejahteraan mereka secara keseluruhan (Haris, Amri, Sunan, & Yogyakarta, 2024).

Status gizi tidak hanya terbatas pada kategori gizi buruk dan gizi baik, tetapi juga meliputi variasi lainnya yang membutuhkan penanganan yang berbeda. Gizi buruk yang menyebabkan kekurangan nutrisi, ada juga kategori gizi kurang yang mengindikasikan asupan nutrisi yang tidak mencukupi, tetapi belum mencapai tingkat gizi buruk. Terdapat gizi lebih yang mengindikasikan asupan nutrisi yang berlebihan. Kategori gizi optimal yang mencerminkan keseimbangan yang baik antara asupan dan kebutuhan nutrisi (Wati, Guru, Anak, & Dini, 2020). Setiap kategori ini memerlukan penanganan yang berbeda-beda (Dita Anggraeni, Nur Rochimah, 2024). Dengan memahami dan menangani setiap kategori secara tepat, pemerintah dapat mengantisipasi dan mengurangi jumlah kasus dengan berbagai kebijakan yang dapat diambil oleh pemerintah dengan penanganan yang tepat.

Teknologi informasi dan komunikasi berbasis machine learning dapat

dimanfaatkan untuk klasifikasi status gizi anak. Akan tetapi, di Kabupaten Asahan data dari berbagai puskesmas belum dimanfaatkan dengan optimal untuk menghasilkan informasi yang berharga. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dataset gizi anak menggunakan model klasifikasi berbasis machine learning guna memprediksi status gizi anak secara dini. Machine learning adalah proses pelatihan atau pembelajaran, yang bergantung pada data pelatihan untuk memperoleh pengetahuan dan pola yang dapat digunakan untuk klasifikasi atau prediksi (Jen & Lin, 2021; Sarker, 2021; Silalahi, Simanullang, & Hutapea, 2023). Metode yang akan dibandingkan dan akan dipilih dengan akurasi terbaik antara lain Decision Tree, Naive Bayes, dan k-Nearest Neighbor karena kecakapannya dalam mengelompokkan data berdasarkan kategori tertentu, yang berguna untuk memprediksi masa depan (Bansal, Goyal, & Choudhary, 2022; Sheth, Tripathi, & Sharma, 2022). Selain itu upaya peningkatan akurasi akan dilakukan dengan algoritma Particle Swarm Optimization (PSO) dan Cross Validation (Zemmal et al., 2020). PSO merupakan algoritma optimasi yang berbasis populasi (Piotrowski, Napiorkowski, & Piotrowska, 2020). Cross-validation adalah teknik pengujian dimana model dilatih dengan data pelatihan dan diuji pada data validasi yang membantu menghindari overfitting dan memberikan klasifikasi yang lebih baik (Misra & Yadav, 2020). Dengan menggunakan teknik optimasi PSO dan Cross Validation, kita dapat meningkatkan akurasi model klasifikasi dengan mencari solusi yang lebih optimal dan efisien secara keseluruhan.

METODE

Penelitian ini menggunakan data primer yang dikumpulkan dari Puskesmas yang terdapat di Kabupaten Asahan antara tahun 2022-2023. Berikut ini langkah-langkah penelitian yang akan dilakukan.

1. Pengumpulan Data

Tahap awal adalah mengumpulkan data gizi balita di puskesmas yang terdapat di Kabupaten Asahan melalui teknik wawancara dan observasi langsung. Ketua peneliti bertanggung jawab untuk melakukan wawancara terkait pemahaman data yang dikumpulkan, sementara anggota tim bertugas untuk mengumpulkan dan mendokumentasikan seluruh proses tersebut.

2. Seleksi Data

Setelah pengumpulan data, dilakukan seleksi data berdasarkan kebutuhan dan pemahaman terhadap analisis model yang akan dilakukan. Proses ini melibatkan ketua dan anggota tim.

3. Pembersihan Data

Langkah selanjutnya adalah membersihkan data dari kesalahan dan duplikat untuk memperbaiki keakuratannya menggunakan python jupyter notebook. Ketua peneliti bertugas menghapus fitur dan rekaman yang bermasalah, sementara anggota tim mengisi data yang hilang

dengan nilai rata-rata dari masing-masing fitur.

4. Optimasi Model Klasifikasi

Dilakukan optimasi terhadap model klasifikasi yang digunakan dengan menggunakan algoritma Particle Swarm Optimization (PSO) untuk meningkatkan akurasi. Proses pengujian akurasi menggunakan cross validation.

Langkah Algoritma Particle Swarm Optimization (PSO)

5. Analisis Data dengan Metode Klasifikasi

Pada tahap ini, data dianalisis menggunakan metode klasifikasi seperti Decision Tree, Naive Bayes, dan k-Nearest Neighbor sebelum di optimasi, untuk dibandingkan dengan hasil optimasi pada langkah berikutnya.

HASIL DAN PEMBAHASAN

Proses pannelitian dimulai dari import data pada Jupyter Notebook dari pengumpulan data Gizi di Kabupaten Asahan.

Usia (bulan)	Jenis Kelamin	Berat Badan (kg)	Tinggi Badan (cm)	Asupan Nutrisi (1-5)	Jumlah Penyakit	Imunisasi	Status Gizi	
0	44	Laki-Laki	9.9	104.9	3	0	Sudah	Gizi Baik
1	57	Perempuan	15.0	68.4	4	0	Sudah	Gizi Lebih
2	34	Perempuan	14.0	85.4	1	2	Sudah	Gizi Baik
3	20	Laki-Laki	9.6	89.9	3	0	Belum	Gizi Kurang
4	48	Laki-Laki	11.2	61.7	1	2	Sudah	Gizi Kurang
...
995	46	Laki-Laki	9.5	100.4	1	0	Belum	Gizi Baik
996	17	Laki-Laki	10.0	103.1	3	2	Sudah	Gizi Baik
997	31	Perempuan	6.7	100.9	4	1	Sudah	Gizi Kurang
998	51	Laki-Laki	11.1	85.6	2	0	Belum	Gizi Kurang
999	30	Perempuan	8.6	81.9	2	2	Sudah	Gizi Kurang

1000 rows x 8 columns

Gambar 1. Import Data

Dataset ini berisi 1.000 baris data dan 8 kolom yang merepresentasikan informasi kondisi gizi balita. Dataset digunakan untuk analisis dan klasifikasi status gizi.

Kolom-kolom dalam dataset:

1. Usia (bulan)

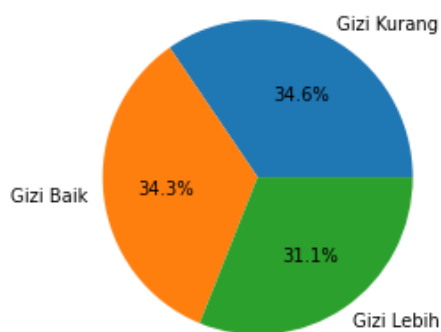
- Tipe data: numerik (integer)
- Keterangan: Usia balita dalam satuan bulan.
- 2. Jenis Kelamin
- Tipe data: kategorikal (Laki-Laki, Perempuan)

- Keterangan: Jenis kelamin balita.
- 3. Berat Badan (kg)
 - Tipe data: numerik (float)
 - Keterangan: Berat badan balita dalam kilogram.
- 4. Tinggi Badan (cm)
 - Tipe data: numerik (float)
 - Keterangan: Tinggi badan balita dalam sentimeter.
- 5. Asupan Nutrisi (1–5)
 - Tipe data: numerik (integer)
 - Keterangan: Skala 1–5 yang menunjukkan tingkat kecukupan asupan nutrisi balita.
- 6. Jumlah Penyakit
 - Tipe data: numerik (integer)
 - Keterangan: Jumlah riwayat penyakit yang diderita balita dalam periode tertentu.
- 7. Imunisasi
 - Tipe data: kategorikal (Sudah, Belum)
 - Keterangan: Status imunisasi balita.
- 8. Status Gizi (*Target Variable*)
 - Tipe data: kategorikal (Gizi Baik, Gizi Lebih, Gizi Kurang)
 - Keterangan: Label status gizi balita yang menjadi target klasifikasi.

Visualisasi Data

Visualisasi data dilakukan menggunakan python melalui jupyter notebook berikut

Persentase Setiap Class pada Dataset



Gambar 2. Visualisasi Dataset

Diagram lingkaran tersebut menggambarkan distribusi status gizi balita dalam dataset. Dari visualisasi terlihat bahwa terdapat tiga kategori status gizi yang dominan dengan proporsi yang relatif seimbang. Kategori pertama memiliki proporsi terbesar yaitu 34,6%, diikuti kategori kedua sebesar 34,3%, dan kategori ketiga sebesar 31,1%. Meskipun perbedaannya tidak terlalu jauh, kategori pertama sedikit lebih banyak dibandingkan dua kategori lainnya.

Secara umum, distribusi ini menunjukkan bahwa data status gizi balita tersebar cukup merata di antara ketiga kategori. Hal ini penting karena keseimbangan distribusi label pada dataset akan mendukung proses pembelajaran model klasifikasi agar tidak bias terhadap salah satu kategori tertentu.

Dari hasil tersebut, dapat disimpulkan bahwa:

1. Model terbaik adalah Decision Tree dengan akurasi 37,50%, yang merupakan nilai akurasi tertinggi dibandingkan model lainnya.
2. Meskipun demikian, akurasi yang diperoleh masih tergolong rendah (di bawah 50%), sehingga perlu dilakukan optimasi parameter (hyperparameter tuning) atau pengolahan data lebih lanjut (misalnya balancing data, feature selection, atau normalisasi yang lebih sesuai).
3. Naive Bayes mendapatkan akurasi paling rendah (30,50%), menunjukkan bahwa asumsi distribusi data yang digunakan oleh algoritma ini kurang sesuai dengan karakteristik dataset.
4. SVM (32,50%) dan KNN (31,50%) memiliki performa menengah, namun tetap lebih rendah daripada Decision Tree.

SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa optimasi model klasifikasi status gizi balita menggunakan metode Particle Swarm Optimization (PSO) dan cross validation mampu mengidentifikasi model terbaik, yaitu Decision Tree dengan akurasi 37,50%, meskipun secara umum tingkat akurasi seluruh model masih relatif rendah. Hal ini menunjukkan perlunya pengolahan data lebih lanjut, seperti balancing data, pemilihan fitur yang lebih relevan, maupun tuning parameter lanjutan untuk meningkatkan performa klasifikasi.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang setinggi-tingginya kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (Kemendikbudristek) atas dukungan pendanaan melalui skema Penelitian Dasar Pemula yang disetujui dalam Program Penelitian Nasional Tahun 2025.

DAFTAR PUSTAKA

- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, 3(May), 100071. <https://doi.org/10.1016/j.dajour.2022.100071>
- Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable selection for Naïve Bayes classification. *Computers and Operations Research*, 135, 105456. <https://doi.org/10.1016/j.cor.2021.105456>
- Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Dita Anggraeni, Nur Rochimah, A. D. (2024). Pengaruh pemberian makanan junk food terhadap kualitas gizi anak usia 4-5 tahun di kecamatan kota baru. *PeTeKa (Jurnal Penelitian Tindakan Kelas Dan Pengembangan Pembelajaran)*, 7(1), 41–48. Retrieved from <http://jurnal.um-tapsel.ac.id/index.php/ptk/article/view/13883>
- Gad, A. G. (2022). *Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. Archives of Computational Methods in Engineering* (Vol. 29). Springer Netherlands. <https://doi.org/10.1007/s11831-021-09694-4>
- Hamed, Y., Ibrahim Alzahrani, A., Shafie, A., Mustaffa, Z., Che Ismail, M., & Kok Eng, K. (2020). Two steps hybrid calibration algorithm of support vector regression and K-nearest neighbors. *Alexandria Engineering Journal*, 59(3), 1181–1190. <https://doi.org/10.1016/j.aej.2020.01.033>
- Haris, A., Amri, M., Sunan, U. I. N., & Yogyakarta, K. (2024). Peran Zakat dalam Mengatasi Stunting dan Gizi Buruk di Kabupaten Brebes. *Jurnal Mabsya*, 6(1), 1–30. <https://doi.org/10.24090/mabsya.v6i1.10706>
- Hasrul, H., Hamzah, H., & Hafid, A. (2021). Pengaruh Pola Makan Terhadap Status Gizi Anak Sekolah Dasar. *Jurnal Keperawatan Muhammadiyah*, 1(1), 54–59. <https://doi.org/10.33096/fmj.v1i1.9>
- Jen, L., & Lin, Y. (2021). A Brief Overview of the Accuracy of Classification Algorithms for Data Prediction in Machine Learning Applications. *Journal of Applied Data Sciences*, 2(3), 84–92. <https://doi.org/10.47738/jads.v2i3.38>

- Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. *International Journal of Engineering Business Management*, 12, 1–9. <https://doi.org/10.1177/1847979020908675>
- Misra, P., & Yadav, A. S. (2020). Improving the Classification Accuracy using Recursive Feature Elimination with Cross-Validation 1. *International Journal on Emerging Technologies* 11(3):, 11(3), 659–665. Retrieved from <http://www.puneetmisra.com/admin/uploads/journals/5f136d202b8ba1.18644117.pdf>
- Permana, B. A. C., Ahmad, R., Bahtiar, H., Sudianto, A., & Gunawan, I. (2021). Classification of diabetes disease using decision tree algorithm (C4.5). *Journal of Physics: Conference Series*, 1869(1). <https://doi.org/10.1088/1742-6596/1869/1/012082>
- Piotrowski, A. P., Napiorkowski, J. J., & Piotrowska, A. E. (2020). Population size in Particle Swarm Optimization. *Swarm and Evolutionary Computation*, 58(March 2019), 100718. <https://doi.org/10.1016/j.swevo.2020.100718>
- Ramon, E., Nazir, A., Novriyanto, N., Yusra, Y., & Oktavia, L. (2022). Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(2), 143–150. <https://doi.org/10.47080/simika.v5i2.2185>
- Sarker, I. H. (2021). Machine Learning : Algorithms , Real - World Applications and Research Directions. *SN Computer Science*, 2(3), 1–21. <https://doi.org/10.1007/s42979-021-00592-x>
- Setiawan, H. B., & Pria Utama, G. (2022). Klasifikasi Status Gizi Balita Menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, (September), 707–715. Retrieved from <http://senafti.budiluhur.ac.id/index.php/senafti/article/view/301>
- Setiawan, R., & Triayudi, A. (2022). Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web. *Jurnal Media Informatika Budidarma*, 6(2), 777. <https://doi.org/10.30865/mib.v6i2.3566>
- Sheth, V., Tripathi, U., & Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose Application Analysis of Machine Learning Algorithms for. *Procedia Computer Science*, 215, 422–431. <https://doi.org/10.1016/j.procs.2022.12.044>