

Optimasi Support Vector Machine Dan K-Nearest Neighbor Untuk Klasifikasi Stunting

Nani Purwati^{*1}, Sri Kiswati², Sri Suharti³, Fransisca Natalia⁴

^{1,2,3,4,5}Universitas Bina Sarana Informatika

Email: ¹nani.npi@bsi.ac.id, ²sri.srk@bsi.ac.id, ³sri.rsh@bsi.ac.id, ⁴fransisca.fia@bsi.ac.id

^{*}Penulis Koresponden

Abstrak

Stunting masih menjadi permasalahan bagi sebagian besar masyarakat di Indonesia. Menurut Kementerian Kesehatan, prevalensi stunting di Indonesia masih cukup tinggi. Sebagai isu yang saat ini menjadi fokus pemerintah Indonesia, stunting menjadi topik penelitian yang sangat penting. Penelitian ini bertujuan untuk membandingkan metode klasifikasi Support Vector System (SVM) dan k-Nearest Neighbor (k-NN) dengan algoritma Particle Swarm Optimization (PSO) untuk mendapatkan model algoritma terbaik dalam klasifikasi stunting pada balita. Dataset penelitian diperoleh dari data posyandu di wilayah pedesaan. Variabel yang digunakan dalam penelitian ini berbeda dengan penelitian sebelumnya, yaitu menggunakan 11 variabel yaitu jenis kelamin, urutan kelahiran, usia, berat badan lahir, tinggi badan lahir, berat badan, tinggi badan, lingkaran lengan atas, lingkaran kepala, pemberian ASI eksklusif, dan status gizi. Hasil pengujian menggunakan algoritma k-NN, SVM, k-NN dengan PSO, dan SVM dengan PSO, menunjukkan hasil akurasi dan durasi tertinggi pada model k-NN sebesar 92,45% dan Kappa sebesar 0,736. Hasil pengujian menggunakan optimasi PSO tidak berubah secara signifikan.

Kata kunci: Klasifikasi, k-Nearest Neighbor, PSO, Stunting, Support Vector Machine.

Abstract

Stunting is still a problem for most people in Indonesia. According to the Ministry of Health, the prevalence of stunting in Indonesia is still quite high. As an issue that is currently the focus of the Indonesian government, stunting is a very important research topic. This study aims to compare the Support Vector System (SVM) and k-Nearest Neighbor (k-NN) classification methods with the Particle Swarm Optimization (PSO) algorithm to get the best algorithm model in the classification of stunting in toddlers. The research dataset was obtained from posyandu data in rural areas. The variables used in this study are different from previous studies, namely using 11 variables: gender, birth order, age, birth weight, birth height, weight, height, arm circumference, head circumference, exclusive breastfeeding, and nutritional status. The test results using the k-NN algorithm, SVM, k-NN with PSO, and SVM with PSO, show the highest accuracy and duration results in the k-NN model of 92.45% and a Kappa of 0.736. The test results using PSO optimization did not change significantly.

Keywords: Classification, k-Nearest Neighbor, PSO, Stunting, Support Vector Machine.

1. PENDAHULUAN

Masalah stunting merupakan salah satu masalah gizi yang menjadi fokus pemerintah Indonesia. Menurut WHO, stunting merupakan masalah gizi kronis yang disebabkan oleh kurangnya asupan gizi dalam jangka panjang (Rahmadhita, 2020). Prevalensi kasus stunting di Indonesia masih tinggi selama satu dekade terakhir (Beal et al., 2018). Indonesia masih menjadi salah satu negara penyumbang masalah stunting global. Prevalensi stunting untuk wilayah Sulawesi Selatan dan Kabupaten Enrekang pada tahun 2019 masing-masing sebesar 30,8%, 35,6%, dan 40% (Sirajuddin et al., 2021). Berbagai upaya pemerintah telah dilaksanakan untuk mengatasi masalah stunting, seperti menyediakan posko gizi, penyuluhan masalah gizi, dan pemberian PMT pada ibu hamil dan balita (Sari & Montessori, 2021). Proyek Gammarana

yang merupakan salah satu upaya pencegahan stunting di masa pandemi tahun 2020 dinilai belum efektif (Sirajuddin et al., 2021).

Beberapa penelitian terdahulu yang berfokus pada klasifikasi stunting, diantaranya uji coba dengan model clustering algoritma K-means menunjukkan hasil akurasi yang rendah dan kurang sesuai untuk analisis kasus stunting (Subayu, 2022). Sedangkan penelitian lain mengenai analisis sentimen dengan menggunakan optimasi PSO menghasilkan akurasi yang meningkat dibandingkan dengan tanpa optimasi PSO (Jatmiko et al., 2022). Penelitian lain dengan menggunakan Metode Decision Tree berbasis PSO juga telah dilakukan dan menghasilkan nilai akurasi yang meningkat dibandingkan dengan sebelum menggunakan optimasi PSO (Rakhman, 2017). Penelitian lain mengenai stunting dengan menggunakan algoritma klasifikasi backpropagation neural network menghasilkan nilai akurasi yang cukup tinggi yaitu sebesar 97,75% (Ristanti et al., 2022). Selain itu, penelitian sebelumnya yang berfokus pada status gizi balita telah dilakukan dengan menggunakan algoritma k-Nearest Neighbor dengan menggunakan variabel data Antropometri atau ukuran tubuh manusia yaitu U (Usia), BB (Berat Badan), TB (Tinggi Badan), LK (Lingkar Kepala) (Saleh et al., 2019).

Belum ada penelitian yang berfokus pada klasifikasi stunting balita menggunakan algoritma Support Vector Machine dan k-Nearest Neighbor berbasis PSO. Oleh karena itu, penelitian ini bermaksud untuk mengklasifikasi kasus stunting balita pasca pandemi Covid-19 menggunakan algoritma Support Vector Machine dan k-Nearest Neighbor berbasis PSO. Variabel yang digunakan dalam penelitian ini berbeda dengan penelitian sebelumnya. Dalam penelitian ini, variabel yang digunakan adalah jenis kelamin, urutan kelahiran, usia, berat badan lahir, tinggi badan lahir, berat badan, tinggi badan, lingkar lengan, lingkar kepala, pemberian ASI eksklusif, dan status gizi. Hasil pengujian algoritma digunakan untuk menentukan metode pengembangan aplikasi deteksi dini kasus stunting pada balita.

2. METODE PENELITIAN

2.1 Penelitian terdahulu

Beberapa penelitian terdahulu untuk melakukan optimasi metode klasifikasi dengan algoritma PSO diantaranya penelitian yang dilakukan oleh (Rustam et al., 2018). Penelitian ini mengusulkan penggunaan optimasi dalam mengidentifikasi daerah endemis penyakit menular dengan menggabungkan algoritma k-means clustering dengan optimasi particle swarm optimization (PSO). Hasil percobaan endemis untuk algoritma k-means dengan iterasi = 10, K-Fold = 2 memiliki koefisien determinasi yang cukup baik yaitu $0,055 > 0,055$. indeks bauldin = 0,169 dan algoritma k-means dengan PSO, iterasi = 10, K-Fold = 5, memiliki davies bouldin index = 0,113. k-fold = 5 mempunyai kinerja yang lebih baik. Penelitian selanjutnya oleh (Wardhani et al., 2018) yaitu melakukan analisis sentimen untuk klasifikasi dokumen dengan text mining. Algoritma yang digunakan pada penelitian ini yaitu Naive Bayes dan Support Vector Machine dengan Particle Swarm Optimization. Penerapan PSO pada penelitian ini terbukti dapat meningkatkan nilai akurasi dan AUC pada masing-masing algoritma. Sedangkan penelitian oleh (Novaldy & Herliana, 2021) mengimplementasikan metode optimasi Particle Swarm Optimization (PSO) pada algoritma Naive Bayes sebagai seleksi fitur, yang bertujuan untuk mengetahui nilai akurasi yang dihasilkan oleh algoritma Naive Bayes dan peran optimasi PSO untuk meningkatkan nilai akurasi prediksi harapan hidup pasien gagal jantung. Setelah dilakukan pengujian, terbukti bahwa optimasi PSO menghasilkan nilai akurasi yang lebih baik yaitu sebesar 92,67% dan nilai AUC sebesar 0,908. Penelitian selanjutnya berfokus pada algoritma AI yang paling tepat untuk menciptakan keseimbangan NPC pada permainan Anagram, menunjukkan bahwa AI dengan algoritma k-NN lebih mendekati kecocokan pemain manusia dan, oleh karena itu, lebih cocok untuk NPC Anagram (Prakosa & Suni, 2022). Sedangkan penelitian oleh (Daldiri & Fitriati, 2023) mengenai perbandingan klasifikasi kanker payudara menggunakan algoritma ID3 Decision Tree dan K-Nearest Neighbor (k-NN), menghasilkan Algoritma ID3 Decision Tree memiliki tingkat akurasi sebesar 93,333%, sedangkan algoritma k-Nearest Neighbor memiliki akurasi yang lebih kecil yaitu sebesar 76,6667%. Penelitian lain yang dilakukan oleh (Muhammadi et al., 2022) mengenai analisis sentimen menggunakan algoritma SVM menunjukkan bahwa SWM berhasil mendapatkan akurasi yang baik yaitu sebesar 81,73%.

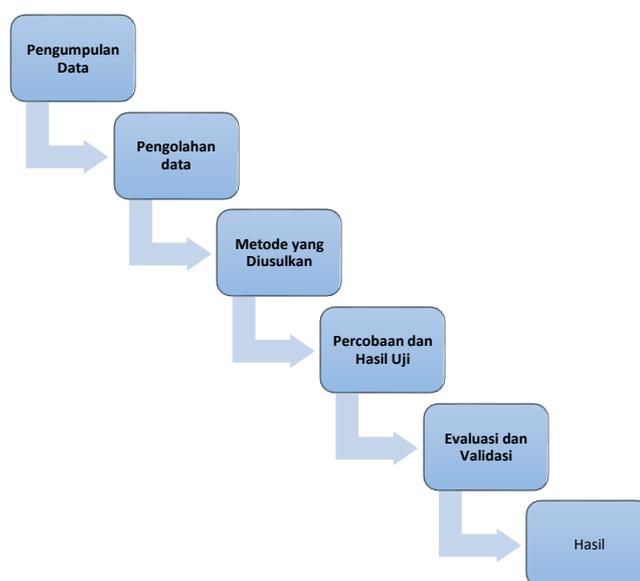
Penelitian ini akan sedikit berbeda dengan penelitian sebelumnya, yaitu penelitian ini akan mengoptimalkan metode Support Vector Machine (SVM) dan k-Nearest Neighbor (k-NN) dengan algoritma PSO untuk mencari nilai akurasi dan AUC tertinggi guna pemilihan model atau metode terbaik.

2.2 Tinjauan Literatur

Data Mining merupakan suatu metode untuk mengekstrak informasi yang berguna dari basis data yang besar. Data Mining juga merupakan bidang ilmu komputer multidisiplin yang melibatkan proses komputasi, teknik statistik, pembelajaran mesin, klasifikasi, pengelompokan, dan penemuan pola (Komi et al., 2017). Klasifikasi merupakan salah satu metode pengambilan keputusan yang paling penting dalam berbagai domain aplikasi. Metode klasifikasi telah berhasil dikembangkan untuk beberapa aplikasi seperti penilaian kredit, diagnosis medis, identifikasi suara, dan lain-lain (Khan et al., 2010). Support Vector Machine merupakan suatu algoritma yang mempelajari klasifikasi dan prediksi numerik. Support Vector Machine juga berguna untuk pengenalan pola yang sangat efektif untuk regresi (Andriyani et al., 2023). Algoritma K-Nearest Neighbor (KNN) merupakan algoritma dengan pendekatan pencarian kasus dengan menghitung kedekatan antara kasus baru dan lama berdasarkan pencocokan bobot beberapa fitur yang ada. Sementara itu, K pada K-NN didefinisikan sebagai jumlah tetangga yang akan diambil yang berguna untuk membuat keputusan yang tepat (Daldiri & Fitriati, 2023). Algoritma PSO pertama kali diusulkan pada tahun 1985 oleh Kennedy dan Eberhard. Algoritma ini merupakan suatu metode simulasi kecerdasan buatan. (Ding, 2017). Algoritma PSO telah mendapat perhatian karena kemampuannya dalam memecahkan masalah optimasi yang kompleks dan sulit (Koohi & Groza, 2014).

2.3 Tahap Penelitian

Dalam penelitian ini terdapat beberapa tahapan, dimulai dari tahap pengumpulan data, kemudian pengolahan data, tahap usulan metode, percobaan dan hasil pengujian, evaluasi dan validasi, kemudian hasil. Dataset diperoleh dari hasil observasi di beberapa titik posyandu di wilayah pedesaan. Data yang dihasilkan sebanyak 331 data balita dengan atribut jenis kelamin, urutan kelahiran, usia, berat badan lahir, tinggi badan lahir, berat badan, tinggi badan, lingkaran lengan atas, lingkaran kepala, pemberian ASI eksklusif, dan status gizi. Metode atau model yang diusulkan dalam penelitian ini adalah metode klasifikasi Support Vector System (SVM) dan k-Nearest Neighbor (k-NN) dengan algoritma Particle Swarm Optimization (PSO). Sedangkan metode penelitian yang diusulkan menggunakan metode CRISP-DM yang bertujuan untuk memperoleh pola atau pengetahuan baru dari data yang digunakan dalam menyelesaikan suatu masalah yang juga digunakan untuk pengambilan keputusan yang tepat. Gambar 1 menunjukkan tahapan penelitian yang dilakukan.



Gambar 1. Tahap Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Tahap *Business Understanding*

Berdasarkan tahapan penelitian, langkah pertama yang dilakukan adalah mengolah dataset balita. Tahap pertama diawali dengan tahap pemahaman usaha. Pada tahap ini, pengetahuan tentang objek penelitian dilakukan dengan cara mengumpulkan informasi tentang stunting dan prediksi yang dilakukan <http://jurnal.bsi.ac.id/index.php/conten>

pada penelitian sebelumnya. Kemudian dilakukan pengolahan data untuk menemukan metode klasifikasi terbaik dengan peningkatan kinerja yaitu pemahaman pada tahap ini.

3.2 Tahap Persiapan Data

Tahap persiapan data dilakukan setelah tahap pemahaman atau tahap pemahaman dataset yang digunakan dalam pengolahan data. Dataset balita diambil dari hasil observasi di beberapa titik posyandu sejak bulan Januari, Februari, Maret, April, dan Mei 2023. Dataset tersebut berisi laporan pengukuran status gizi dan balita stunting sebanyak 425 buah dan 11 atribut. Informasi data beserta atributnya dapat dilihat pada Tabel 1.

Tabel 1. Dataset Balita

Nama Data	Sumber Data	Jumlah Data	Jumlah Atribut	Atribut
Dataset balita	Posyandu	425	11	jenis kelamin, urutan kelahiran, usia, berat badan lahir, tinggi badan lahir, berat badan, tinggi badan, lingkaran lengan, lingkaran kepala, pemberian ASI eksklusif, status gizi

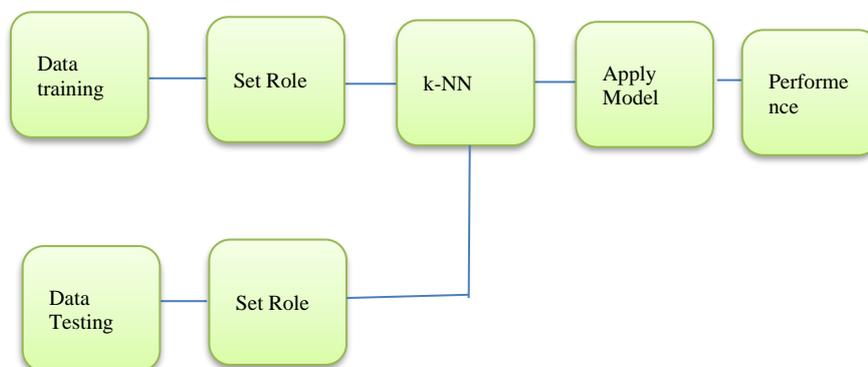
Tabel 2 merupakan data yang telah melalui tahap pengolahan. Angka 1 pada atribut gender berarti laki-laki, angka 2 berarti perempuan, sedangkan angka 1 pada ASI eksklusif berarti ya, dan angka 2 berarti tidak. Pada atribut age menggunakan bulan. Atribut status gizi merupakan atribut yang akan dijadikan kelas, berisi informasi mengenai stunting, normal, dan malnutrisi. Data ini selanjutnya akan diolah menggunakan algoritma k-NN, SVM dengan optimasi PSO.

Tabel 2. Persiapan Data

Jenis kelamin	Urutan kelahiran	usia	Berat lahir	Tinggi badan lahir	berat	tinggi	lingkar lengan	lingkar kepala	pemberian ASI eksklusif	status gizi
2	4	20	3.3	48	9.5	71	14	42	1	pengerdilan
1	2	11	2.5	46	7.4	70.5	14.5	39	1	kekurangan gizi
2	1	10	3.5	51	12	74.5	19	44	1	normal
2	1	9	2.9	50	8.6	66.3	16	41	1	normal
2	2	41	3.4	49	14.7	93	17	47	1	normal
...

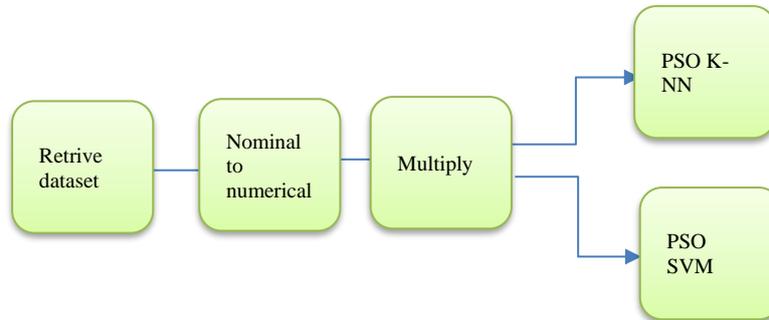
3.3 Tahap Pemodelan

Model perancangan dengan menggunakan tool Rapidminer 10.1 dapat dilihat pada Gambar 2. Pada Gambar 2. dapat dijelaskan bahwa untuk data yang sudah siap dilakukan Import Data kemudian pilih file data. Setelah file data dipilih, dilakukan set role untuk memilih atribut yang akan dijadikan label. Atribut status gizi merupakan atribut yang dijadikan label. Operator Apply Model digunakan untuk pengujian dataset baru yang belum memiliki label.



Gambar 2. Desain Model k-NN Tanpa PSO

Pada pengujian algoritma SVM tanpa PSO, desain modelnya hampir sama. Selanjutnya dilakukan pemodelan dengan optimasi PSO. Gambar 3. merupakan tampilan model dengan optimasi PSO.



Gambar 3. Desain model PSO

3.4 Tahap Evaluasi

Hasil pengukuran kinerja menunjukkan bahwa pengujian menggunakan algoritma k-NN sebelum dioptimasi dengan PSO memperoleh akurasi sebesar 92,42%. Sementara itu, hasil pengujian setelah menggunakan optimasi PSO tidak menunjukkan perubahan yang signifikan dari hasil sebelum dan sesudahnya. Akurasi penggunaan PSO pada algoritma k-NN sebesar 92,14% yang menunjukkan penurunan sekitar 0,28%. Sedangkan pada algoritma SVM sebelum menggunakan optimasi PSO memperoleh akurasi sebesar 86,95%. Setelah menggunakan optimasi PSO turun menjadi 86,27%.

3.5 Pemilihan Model Terbaik

Hasil pengukuran kinerja algoritma k-NN dan SVM ditunjukkan pada Tabel 3. Dari tabel tersebut diketahui bahwa algoritma dengan PSO dari kedua metode yang ada tidak terlalu berbeda secara signifikan. Pada algoritma k-NN nilai akurasi awal sebesar 92,45% dan nilai Kappa sebesar 0,736 menjadi 92,14% dan Kappa sebesar 0,716. Hal tersebut justru mengalami penurunan sebesar 0,28%. Begitu pula pada Algoritma SVM dari semula 86,85% dan Kappa 0,761 menjadi 86,27% dan Kappa 0,741.

Tabel 3. Hasil Pengukuran Kinerja

Sebuah algoritma	Tanpa PSO		Dengan PSO	
	Ketepatan	Kappa	Sebuah keakuratan	Kappa
Bahasa Indonesia: SVM	86,85 %	0.761	86,27 %	0.741
k-NN	92,45 %	0.736	92,14 %	0.716

4. KESIMPULAN

Dari percobaan yang telah dilakukan dapat diambil beberapa simpulan yaitu untuk pengukuran kinerja, pengujian klasifikasi stunting PSO menggunakan algoritma k-NN dan SVM tidak memberikan perubahan yang signifikan. Untuk pengukuran kinerja, nilai akurasi dari tertinggi ke terendah pada percobaan ini menghasilkan akurasi tertinggi dengan menggunakan metode k-NN. Hasil ini dapat dijadikan bahan acuan untuk mengaplikasikan model deteksi dini kasus stunting. Untuk penelitian selanjutnya dapat melibatkan lebih banyak dataset yang berbeda dan metode klasifikasi lainnya.

REFERENSI

- Andriyani, S. Y., Lydia, M. S., & Efendi, S. (2023). Optimization of Support Vector Machine Algorithm Using Stunting Data Classification. *Prisma Sains : Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, 11(1), 164. <https://doi.org/10.33394/j-ps.v11i1.6619>
- Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. *Maternal and Child Nutrition*, 14(4), 1–10. <https://doi.org/10.1111/mcn.12617>
- Daldiri, Z. F., & Fitriati, D. (2023). Comparison of Breast Cancer Classification Using the Decision Tree ID3 Algorithm and K-Nearest Neighbors Algorithm. *Jurnal Riset Informatika*, 5(2), 177–186. <https://doi.org/10.34288/jri.v5i2.406>
- Ding, Z. (2017). Research of improved particle swarm optimization algorithm. *AIP Conference Proceedings*, 1839(May 2017), 1–6. <https://doi.org/10.1063/1.4982513>
- Jatmiko, H. B., Tedi Kurniadi, N., & Maulana, D. (2022). Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta. *Journal Automation Computer Information System*, 2(1), 22–30. <https://doi.org/10.47134/jacis.v2i1.35>
- Khan, N. K., Baig, A. R., & Iqbal, M. A. (2010). A new discrete PSO for data classification. *2010 International Conference on Information Science and Applications, ICISA 2010*. <https://doi.org/10.1109/ICISA.2010.5480366>
- Komi, M., Li, J., Zhai, Y., & Xianguo, Z. (2017). Application of data mining methods in diabetes prediction. *2017 2nd International Conference on Image, Vision and Computing, ICIVC 2017, S Ix*, 1006–1010. <https://doi.org/10.1109/ICIVC.2017.7984706>
- Koohi, I., & Groza, V. Z. (2014). Optimizing Particle Swarm Optimization algorithm. *Canadian Conference on Electrical and Computer Engineering*, 1–5. <https://doi.org/10.1109/CCECE.2014.6901057>
- Muhammadi, R. H., Laksana, T. G., & Arifa, A. B. (2022). Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 8(1), 59–71. <https://doi.org/10.23917/khif.v8i1.15213>
- Novaldy, F., & Herliana, A. (2021). Penerapan Pso Pada Naïve Bayes Untuk Prediksi Harapan Hidup Pasien Gagal Jantung. *Jurnal Responsif: Riset Sains Dan Informatika*, 3(1), 37–43. <https://doi.org/10.51977/jti.v3i1.396>
- Prakosa, Y. A., & Suni, A. F. (2022). Backtracking and k-Nearest Neighbour for Non-Player Character to Balance Opponent in a Turn-Based Role Playing Game of Anagram. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 8(2), 132–138. <https://doi.org/10.23917/khif.v8i2.16902>
- Rahmadhita, K. (2020). Permasalahan Stunting dan Pencegahannya. *Jurnal Ilmiah Kesehatan Sandi Husada*, 11(1), 225–229. <https://doi.org/10.35816/jiskh.v11i1.253>
- Rakhman, A. (2017). Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization (PSO). *Smart Comp :Jurnalnya Orang Pintar Komputer*, 6(1), 193–197. <http://ejournal.poltektegal.ac.id/index.php/smartcomp/article/view/466>
- Ristanti, D. E., Syaquy, D., & Prasetyo, B. H. (2022). *Stunting Early Detection System for Toddlers Based on Height and Weight Using Backpropagation Neural Network Method*. 1(1), 1–9.
- Rustam, S., Santoso, H. A., & Supriyanto, C. (2018). Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. *ILKOM Jurnal Ilmiah*, 10(3), 251–259. <https://doi.org/10.33096/ilkom.v10i3.342.251-259>
- Saleh, H., Faisal, M., & Musa, R. I. (2019). Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor. *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, 4(2), 120–126. <https://doi.org/10.51876/simtek.v4i2.60>
- Sari, R. P. P., & Montessori, M. (2021). Upaya Pemerintah dan Masyarakat dalam Mengatasi Masalah Stunting pada Anak Balita. *Journal of Civic Education*, 4(2), 129–136. <https://doi.org/10.24036/jce.v4i2.491>
- Sirajuddin, S., Thaha, A. R., Razak, A., Thaha, M., Junadi, P., & Ali, P. B. (2021). It has no impact but is useful as a lesson learned from the Gammarana Enrekang stunting prevention project during the COVID-19 pandemic. *Annals Romanian Society Biology Cell*, 25(6), 4723–4728.
- Subayu, A. (2022). Penerapan Metode K-Means untuk Analisis Stunting Gizi pada Balita: Systematic Review. *Jurnal SNATI*, 2, 42–50. <https://journal.uin.ac.id/journalsnati/article/view/24255/14152>
- Wardhani, N. K., et.al. (2018). Sentiment analysis article news coordinator minister of maritime affairs using algorithm naive bayes and support vector machine with particle swarm optimization. *Journal of Theoretical and Applied Information Technology*, 96(24), 8365–8378.