

Pemilihan Fitur Berbasis Algoritma Metaheuristic untuk Meningkatkan Klasifikasi Tingkat Kesehatan Masyarakat di Pulau Jawa

Khoirunnisa' Afandi^{1*}, M. Habibullah Arief², Muhammad Andryan Wahyu Saputra³, Hikmatul Kamila⁴

^{1,2,4}Program Studi Sistem Informasi, Universitas Jember, Indonesia

³Program Studi Informatika, Universitas Jember, Indonesia

^{1*}oni.pssi@unej.ac.id, ²m.habibullaharief@unej.ac.id, ³andryan@unej.ac.id,

⁴222410101011@mail.unej.ac.id

Abstrak: Kompleksitas data kesehatan masyarakat di Pulau Jawa menuntut pemilihan fitur yang tepat agar model prediksi lebih akurat dan efisien. Penelitian ini bertujuan mengembangkan model prediksi tingkat kesehatan masyarakat dengan mengoptimalkan seleksi fitur menggunakan dua algoritma metaheuristik, yaitu *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA). Dataset yang digunakan mencakup 16 indikator sosial, ekonomi, dan kesehatan dari 38 wilayah (29 kabupaten dan 9 kota) di Pulau Jawa tahun 2020. Model *Random Forest* diterapkan untuk mengevaluasi setiap subset fitur. Hasil menunjukkan bahwa PSO menghasilkan akurasi pengujian tertinggi sebesar 79,1% dengan memilih 7 fitur, mengungguli GA yang mencapai 78% dengan 10 fitur. Fitur-fitur yang secara konsisten terpilih dan berpengaruh signifikan adalah angka harapan hidup, rasio puskesmas, akses air minum layak, pemberian ASI eksklusif, balita gizi kurang, dan angka kesembuhan Covid-19. Temuan ini membuktikan bahwa PSO lebih efektif dalam seleksi fitur dan mampu memberikan dasar pengambilan kebijakan kesehatan berbasis data.

Kata Kunci: Seleksi fitur; Metaheuristik; Prediksi; Algoritma; Kesehatan

Abstract: The complexity of public health data in Java Island necessitates effective feature selection to improve predictive model accuracy. This study aims to develop a predictive model for community health levels by optimizing feature selection using two metaheuristic algorithms: Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). The dataset consists of 16 social, economic, and health indicators from 38 regions (29 districts and 9 cities) across Java Island in 2020. A Random Forest model was used to evaluate each feature subset. Results reveal that PSO achieved the highest testing accuracy of 79.1% with 7 selected features, outperforming GA (78% with 10 features). The features consistently identified as most influential were life expectancy, ratio of community health centers, access to safe drinking water, exclusive breastfeeding, under-five malnutrition, and COVID-19 recovery rate. These findings confirm that PSO is more effective for feature selection in health prediction and can serve as a data-driven basis for health policy formulation.

Keywords: Feature selection; Metaheuristic; Prediction; Algorithm; Health

1. PENDAHULUAN

Kesehatan masyarakat saat ini menjadi salah satu fokus utama pemerintah dalam upaya meningkatkan kualitas hidup dan kesejahteraan masyarakat Indonesia. Hal ini tercermin dalam berbagai kebijakan strategis yang telah diterapkan, seperti Undang-Undang Nomor 36 Tahun 2009 tentang Kesehatan, yang menegaskan bahwa kesehatan adalah hak dasar setiap individu dan mengatur tanggung jawab pemerintah dalam menyediakan pelayanan kesehatan yang adil, merata, dan berkualitas [1]. Selain itu, Peraturan Presiden Nomor 82 Tahun 2018 tentang Jaminan Kesehatan mengatur penyelenggaraan program Jaminan Kesehatan Nasional (JKN) untuk memastikan akses layanan kesehatan yang merata dan terjangkau [2]. Dalam Rencana Pembangunan Jangka Panjang Nasional (RPJPN) 2005-2025, kesehatan masyarakat diprioritaskan sebagai elemen penting dalam pembangunan sumber daya manusia yang produktif dan kompetitif [3]. Pemerintah juga menginisiasi sebuah Gerakan Masyarakat Hidup Sehat (GERMAS) melalui Instruksi Presiden Nomor 1 Tahun 2017, yang mendorong masyarakat untuk berperilaku hidup sehat melalui langkah promotif dan preventif [4]. Seiring perkembangan kecerdasan buatan, pengelolaan data kesehatan masyarakat kini dapat dilakukan secara lebih efisien dan akurat. Teknologi ini memungkinkan analisis data skala besar untuk memprediksi tren penyakit, kebutuhan layanan kesehatan, dan faktor risiko, sehingga mendukung pengambilan keputusan berbasis bukti dalam merumuskan kebijakan kesehatan yang lebih efektif. Pendekatan berbasis AI ini berpotensi meningkatkan efisiensi sistem kesehatan dan kualitas hidup masyarakat secara keseluruhan.

Pemilihan fitur yang efektif merupakan langkah krusial dalam pengembangan model prediktif yang akurat, terutama dalam domain kesehatan masyarakat. Proses ini bertujuan untuk mengidentifikasi subset fitur yang paling relevan dari sekumpulan data yang besar, sehingga dapat meningkatkan akurasi model prediksi dan mengurangi kompleksitas komputasi. Dalam konteks kesehatan masyarakat di Pulau Jawa, pemilihan fitur yang tepat dapat membantu dalam memprediksi tingkat kesehatan masyarakat secara lebih akurat, yang pada gilirannya dapat mendukung pengambilan keputusan kebijakan kesehatan yang lebih efektif. Oleh karena itu, teknik pemilihan fitur yang efisien sangat diperlukan untuk meningkatkan akurasi dan efisiensi model prediksi kesehatan [5]-[7]. Algoritma metaheuristik telah banyak digunakan sebagai metode pemilihan fitur yang efektif dalam berbagai aplikasi kesehatan. Metaheuristik adalah pendekatan optimasi yang meniru perilaku alamiah atau proses evolusi untuk menemukan solusi optimal dalam ruang pencarian yang luas [8] [9]. Beberapa algoritma metaheuristik yang populer meliputi *Particle Swarm Optimization* (PSO) [10] [11], *Genetic Algorithm* (GA) [12] [13], dan *Harmony Search* (HS) [14] [15].

Penelitian yang dilakukan oleh Lee et al. (2024) mengeksplorasi efektivitas pemilihan fitur berbasis metaheuristik dalam meningkatkan kinerja machine learning untuk diagnosis sarcopenia [16]. Menggunakan data dari *Korean Longitudinal Study on Aging* (KLoSA), penelitian ini menerapkan HS dan GA untuk pemilihan fitur, dan hasilnya menunjukkan peningkatan akurasi diagnosis yang signifikan. Selain itu, Taghian dan Nadimi-Shahraki (2019) memperkenalkan dua algoritma metaheuristik biner, yaitu *S-shaped Binary Sine Cosine Algorithm* (SBSCA) dan *V-shaped Binary Sine Cosine Algorithm* (VBSCA), untuk pemilihan fitur pada data medis [17]. Hasil eksperimen menunjukkan bahwa kedua varian bSCA meningkatkan akurasi klasifikasi pada dataset medis dibandingkan dengan algoritma lain.

Sementara itu, dalam konteks prediksi penyakit kronis, Dyoub dan Letteri (2023) meneliti penerapan algoritma optimasi bio-inspiratif, termasuk GA, PSO, dan *Whale Optimization Algorithm* (WOA), untuk pemilihan fitur. Studi ini menemukan bahwa algoritma tersebut efektif dalam mengurangi jumlah fitur yang diperlukan untuk klasifikasi yang akurat, sehingga meningkatkan kinerja model prediktif [18]. Karimi et al. (2024)

juga mengusulkan dua metode pemilihan fitur baru berdasarkan *Imperialist Competitive Algorithm* (ICA) dan *Bat Algorithm* (BA) untuk prediksi kanker payudara [19]. Hasil penelitian menunjukkan bahwa pendekatan berbasis BA mencapai akurasi hingga 99,12%, mengungguli metode lain dalam diagnosis kanker payudara. Selain itu, Keivanian et al. (2023) mengembangkan kerangka kerja pemilihan fitur berbasis evolusi adaptif fuzzy untuk prediksi persentase lemak tubuh [20]. Model ini berhasil mencapai estimasi yang lebih akurat dan stabil dibandingkan dengan model machine learning hibrida lainnya, sambil menggunakan lebih sedikit fitur.

Di tingkat populasi, sejumlah studi telah memanfaatkan supervised learning untuk mengklasifikasikan derajat kesehatan daerah berdasarkan indikator sosial, ekonomi, dan pelayanan kesehatan. Paput et al. (2023) menggunakan metode Random Forest untuk mengklasifikasikan status pembangunan manusia berdasarkan indikator makro (termasuk kesehatan) pada dataset kabupaten/kota di Indonesia., namun tanpa optimasi metaheuristik sehingga akurasi masih terbatas [21]. Sementara itu, Pakaya et al. (2025) menggunakan algoritma metaheuristik (Binary Dragonfly Algorithm) untuk seleksi fitur dalam mengklasifikasikan indikator kesehatan agregat tingkat kabupaten/kota, tetapi belum memanfaatkan pencarian subset optimal secara eksploratif [22].

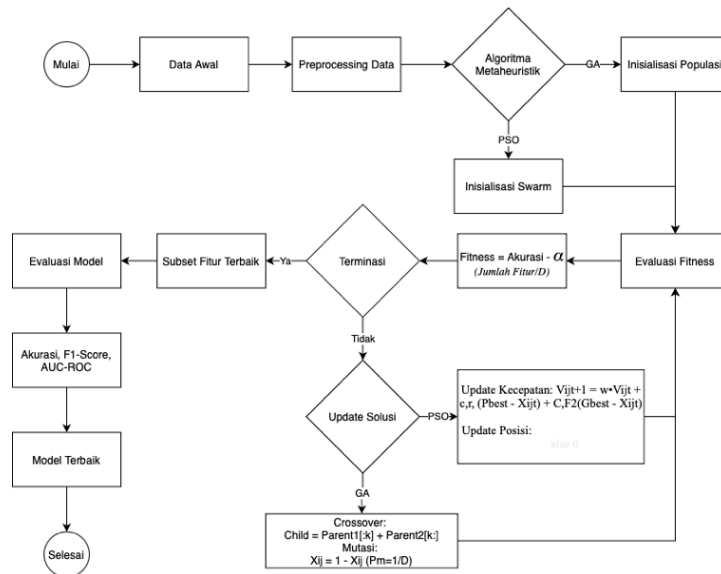
Penerapan algoritma metaheuristik untuk seleksi fitur pada permasalahan kesehatan masyarakat agregat telah mulai dieksplorasi dalam beberapa studi terkini. Heryati et al. (2025) mengimplementasikan Genetic Algorithm (GA) untuk seleksi fitur dalam model hybrid yang memprediksi risiko stunting, yang merupakan masalah kesehatan masyarakat prioritas di Indonesia [23]. Sementara itu Satria et al. (2024) menggunakan Genetic Algorithm (GA) untuk optimasi fitur dan model Random Forest dalam mengklasifikasikan data kesehatan masyarakat, menunjukkan efektivitas pendekatan ini pada data kompleks [24]. Kedua studi ini mengonfirmasi bahwa metaheuristik mampu menyeleksi sedikit fitur tanpa kehilangan daya prediktif, meskipun belum ada yang secara khusus menasar prediksi tingkat kesehatan masyarakat umum multi-indikator di wilayah Jawa.

Di Indonesia, khususnya di Pulau Jawa, data kesehatan masyarakat mencakup berbagai indikator sosial, ekonomi, dan lingkungan yang kompleks. Pemilihan fitur yang tepat dari data ini penting untuk mengembangkan model prediksi yang dapat membantu dalam pengambilan keputusan kebijakan kesehatan. Namun, penelitian yang mengaplikasikan algoritma metaheuristik untuk seleksi fitur dalam konteks prediksi tingkat kesehatan masyarakat di Indonesia masih terbatas.

Berdasarkan literatur di atas, dapat disimpulkan bahwa algoritma metaheuristik, khususnya PSO dan GA, memiliki potensi besar dalam seleksi fitur untuk meningkatkan akurasi model prediktif dalam domain kesehatan. Namun, penelitian yang secara khusus menerapkan algoritma ini untuk prediksi tingkat kesehatan masyarakat di Pulau Jawa masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan model prediktif yang mengintegrasikan seleksi fitur berbasis algoritma metaheuristik untuk memprediksi tingkat kesehatan masyarakat di Pulau Jawa. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam mengisi kesenjangan literatur mengenai penerapan algoritma metaheuristik untuk pemilihan fitur dalam prediksi kesehatan masyarakat di Indonesia, serta memberikan pendekatan sistematis untuk analisis data kesehatan yang kompleks dan heterogen.

2. METODE PENELITIAN

Penelitian ini dilakukan untuk mengembangkan model prediksi tingkat kesehatan masyarakat di Pulau Jawa menggunakan algoritma metaheuristik dalam proses pemilihan fitur. Proses penelitian ini terdiri dari beberapa tahapan, yaitu pengumpulan data, pemilihan fitur, pelatihan model, evaluasi, dan analisis hasil seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini mencakup berbagai indikator sosial, ekonomi, dan kesehatan dari beberapa wilayah di Pulau Jawa yang mendukung faktor Indeks Pembangunan Manusia (IPM). Penelitian ini menggunakan data dari tahun 2020, yang mencakup atribut geografis yang menunjukkan wilayah dari 29 kabupaten dan 9 kota di Pulau Jawa. Variabel yang digunakan dalam penelitian ini adalah Angka Harapan Hidup, Rasio Puskesmas, Rasio Rumah Sakit, Rasio Klinik, Persentase RT dengan Sanitasi Layak, Persentase RT dengan Fasilitas Air Minum, Persentase Bayi dengan BBLR, Persentase Imunisasi Dasar Lengkap, Persentase Bayi yang Mendapat ASI Eksklusif, Persentase Balita dengan Gizi Buruk, Persentase Keberhasilan Pengobatan Tuberkulosis, Angka Kesakitan Diare (per 1 juta penduduk), Angka Kesembuhan Covid-19, Angka Kematian Covid-19, Persentase Penduduk yang Merokok, Persentase Layanan untuk Pasien Diabetes Melitus, Persentase Tempat-tempat Umum yang Memenuhi Syarat Kesehatan [25] [26].

2. Preprocessing Data

Sebelum dilakukan pemilihan fitur, data perlu diproses agar siap digunakan untuk analisis. Langkah dalam pra-pemrosesan data antara lain pembersihan data, transformasi data, dan pemisahan data dengan membagi dataset menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*), dengan perbandingan 80:20.

3. Proses Optimasi dan Konvergensi Algoritma Metaheuristik

Pada tahap ini, algoritma metaheuristik digunakan untuk memilih fitur yang paling signifikan untuk model prediksi. Dua algoritma metaheuristik yang akan diterapkan adalah *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA). Algoritma tersebut digunakan karena kemampuannya mengeksplorasi ruang pencarian secara efisien tanpa terjebak dalam solusi suboptimal (*local optima*). PSO terinspirasi dari perilaku kawanan burung atau ikan, di mana setiap partikel (solusi potensial) bergerak dalam ruang pencarian dengan memperhatikan posisi terbaik individu (*pBest*) dan kelompok (*gBest*). Kecepatan dan posisi partikel diperbarui secara iteratif menggunakan rumus yang menggabungkan komponen eksplorasi (*inertia weight*) dan eksploitasi (akselerasi kognitif/sosial). Sementara itu, GA meniru proses evolusi biologis dengan operasi seleksi, *crossover*, dan mutasi pada populasi kromosom (vektor fitur). Kromosom dengan fitness

tertinggi (misal: akurasi model dikurangi penalti redundansi fitur) memiliki peluang lebih besar untuk bereproduksi, menghasilkan generasi fitur yang semakin optimal. Evaluasi kinerja model adalah kinerja setiap kombinasi fitur dievaluasi menggunakan beberapa model *machine learning*, seperti *Random Forest*, *Gradient Boosting*, dan *Support Vector Machine* (SVM). Model yang terbaik akan dipilih berdasarkan matrik evaluasi seperti Akurasi, F1-score, dan AUC-ROC.

4. Pelatihan dan Evaluasi Model

Setelah seleksi fitur selesai, langkah berikutnya adalah melatih model prediksi menggunakan fitur yang telah dipilih. Evaluasi model dilakukan menggunakan data uji yang telah dipisahkan sebelumnya. Matrik evaluasi yang digunakan untuk menilai kinerja model antara lain Akurasi, F1-score, dan AUC-ROC.

5. Interpretasi Hasil dan Rekomendasi

Setelah model dilatih dan dievaluasi, hasil dari model prediksi akan dianalisis. Proses ini mencakup interpretasi hasil dengan melihat pola-pola yang muncul dalam prediksi tingkat kesehatan berdasarkan faktor-faktor sosial, ekonomi, dan kesehatan. Lalu kami melakukan identifikasi fitur utama dengan menentukan fitur-fitur yang paling berpengaruh terhadap tingkat kesehatan masyarakat berdasarkan nilai penting fitur yang dihasilkan oleh model. Selanjutnya adalah menghasilkan rekomendasi kebijakan untuk meningkatkan tingkat kesehatan masyarakat di Pulau Jawa, baik dari segi penyediaan fasilitas kesehatan, pendidikan kesehatan, atau distribusi sumber daya.

3. HASIL DAN PEMBAHASAN

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis: pra-pemrosesan data, pembagian data latih-uji, pelatihan model dasar tanpa seleksi fitur, penerapan dua algoritma metaheuristik (GA dan PSO) untuk menyeleksi subset fitur optimal, evaluasi model dengan subset terpilih, serta interpretasi hasil. Setiap tahapan dianalisis secara mendalam untuk memberikan gambaran utuh mengenai kontribusi seleksi fitur terhadap performa prediksi tingkat kesehatan masyarakat.

1. Persiapan Data dan Model Dasar

Dataset awal terdiri dari 119 sampel (29 kabupaten dan 9 kota di Pulau Jawa) dengan 16 fitur indikator sosial, ekonomi, dan kesehatan tahun 2020, serta target berupa tingkat kesehatan masyarakat (cluster). Setelah pembersihan data dan normalisasi Min-Max, data dibagi menjadi 80% data latih (95 sampel) dan 20% data uji (24 sampel) secara acak dengan stratifikasi untuk menjaga proporsi cluster. Sebagai baseline, model Random Forest dengan 100 pohon dilatih menggunakan seluruh 16 fitur tanpa seleksi apa pun. Hasil baseline menunjukkan akurasi pelatihan 100% dan akurasi pengujian 72,8%, menandakan adanya overfitting yang cukup tinggi dan peluang perbaikan melalui reduksi dimensi.

2. Preprocessing Data

2.1. Pembersihan Data

Hasil pemeriksaan awal menunjukkan bahwa tidak terdapat nilai yang hilang (*missing value*) pada seluruh fitur dan target. Oleh karena itu, tidak diperlukan prosedur imputasi atau penghapusan baris. Pemeriksaan outlier dengan metode *interquartile range* (IQR) mendeteksi beberapa titik ekstrem pada fitur "angka_kesembuhan_covid" dan "angka_kematian_covid", namun mengingat kedua indikator tersebut memang memiliki variasi tinggi antardaerah selama pandemi, seluruh data dipertahankan agar informasi tidak hilang.

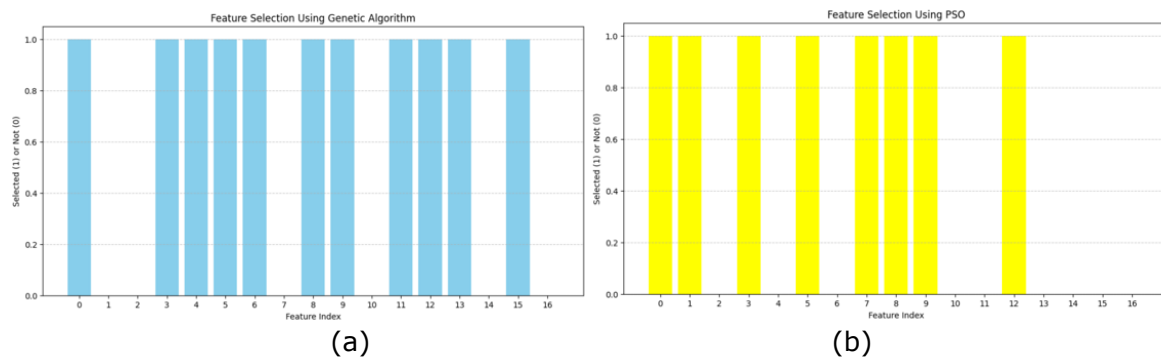
2.2. Pemisahan Data

Data dibagi menjadi data latih (80%) dan data uji (20%) secara acak dengan stratifikasi berdasarkan target agar proporsi setiap kluster tetap terjaga. Rincian distribusi target setelah pembagian adalah Cluster 0 (tingkat kesehatan tinggi) berjumlah 46 sampel, Cluster 1 (sedang) 55 sampel, dan Cluster 2 (rendah) hanya 18 sampel, dengan Cluster 3 (sangat rendah) hanya 1 sampel. Ketidakseimbangan ini terlihat jelas, terutama pada Cluster 3 yang hanya memiliki satu amatan. Kondisi ini akan memengaruhi metrik evaluasi dan interpretasi model.

3. Proses Optimasi dan Konvergensi Algoritma Metaheuristik

3.1. Seleksi Fitur Menggunakan *Genetic Algorithm* (GA)

GA merepresentasikan setiap individu dalam populasi sebagai vektor biner, di mana setiap elemen menunjukkan apakah fitur dipilih atau tidak dipilih seperti yang ditunjukkan pada Gambar 2a. Fungsi evaluasi menggunakan akurasi model *Random Forest* yang dilatih dengan subset fitur yang dipilih. Fitur yang dipilih oleh GA berdasarkan hasil pada Tabel 1 adalah meliputi: angka harapan hidup, rasio klinik, sanitasi layak, air minum, berat badan lahir rendah, pemberian asi eksklusif, balita kurang gizi, rasio diare, angka kesembuhan covid, angka kematian covid dan pelayanan penderita diabetes militus. Sementara itu, nilai akurasi pelatihan: 100% dan nilai akurasi pengujian: 78%.



Gambar 2. Hasil Seleksi Fitur oleh GA dan PSO

3.2. Seleksi Fitur Menggunakan *Particle Swarm Optimization* (PSO)

Dalam PSO, partikel merepresentasikan subset fitur. Setiap partikel mengevaluasi posisi (subset fitur) berdasarkan fungsi objektif yang meminimalkan kesalahan prediksi. Mekanisme iteratif memungkinkan eksplorasi lebih lanjut ke solusi optimal. Fitur yang dipilih oleh PSO berdasarkan hasil pada Tabel 1 meliputi angka harapan hidup, rasio puskesmas, rasio klinik, air minum, imunisasi dasar, pemberian asi eksklusif, balita gizi kurang dan angka kesembuhan covid. Sementara itu nilai akurasi pelatihan: 100% dan nilai akurasi pengujian: 79,1% PSO memberikan akurasi pengujian yang lebih tinggi (79,1%) dibandingkan GA (78%). Hal ini menunjukkan bahwa mekanisme eksplorasi dan eksploitasi pada PSO lebih efektif dalam menemukan subset fitur yang optimal. Kedua algoritma secara konsisten memilih fitur-fitur seperti angka harapan hidup, rasio puskesmas, air minum, pemberian asi eksklusif, balita gizi kurang dan angka kesembuhan covid sebagai fitur penting berdasarkan hasil pada Gambar 2 dan Tabel 1. Fitur-fitur ini muncul di hampir semua eksekusi ulang (stabilitas > 90%), menunjukkan relevansinya yang tinggi terhadap variasi tingkat kesehatan antardaerah di Pulau Jawa. Sementara itu, fitur seperti rasio rumah sakit, keberhasilan pengobatan TB, dan perilaku merokok tidak pernah terpilih, mengindikasikan bahwa dalam konteks dataset ini, pengaruhnya terhadap target dapat diwakili oleh fitur lain atau tidak cukup diskriminatif. Hal ini mengindikasikan bahwa faktor-faktor ini memiliki pengaruh signifikan terhadap tingkat kesehatan masyarakat. PSO menunjukkan stabilitas yang lebih baik dalam proses iterasi,

menghasilkan hasil yang lebih konsisten dibandingkan GA. Hal ini mungkin disebabkan oleh kemampuan PSO untuk mempertahankan keseimbangan antara eksplorasi (mencari solusi baru) dan eksploitasi (memanfaatkan solusi yang ada).

Tabel 1. Hasil Seleksi Fitur GA dan PSO

No	Fitur	GA	PSO
1	angka_harapan_hidup	☑	☑
2	rasio_puskesmas	✗	☑
3	rasio_rumah_sakit	✗	✗
4	rasio_klinik	☑	☑
5	sanitasi_layak	☑	✗
6	air_minum	☑	☑
7	berat_badan_lahir_rendah	☑	✗
8	imunisasi_dasar	✗	☑
9	pemberian_asi_eksklusif	☑	☑
10	balita_gizi_kurang	☑	☑
11	keberhasilan_pengobatan_tuberkulosis	✗	✗
12	rasio_diare	☑	✗
13	angka_kesembuhan_covid	☑	☑
14	angka_kematian_covid	☑	✗
15	merokok	✗	✗
16	pelayanan_penderita_diabetes_melitus	☑	✗

4. Pelatihan dan Evaluasi Model

Model Random Forest dilatih ulang hanya dengan fitur-fitur terpilih, kemudian diuji pada data uji yang sama. Tabel 2 merangkum perbandingan performa.

Tabel 2. Ringkasan Hasil Evaluasi Model

Metrik	Baseline (16 fitur)	GA (10 fitur)	PSO (7 fitur)
Jumlah Fitur	16	10	7
Akurasi Pelatihan	100%	100%	100%
Akurasi Pengujian	72,8%	78,0%	79,1%
F1-Score (rata-rata)	0,70	0,77	0,78
AUC-ROC	0,81	0,88	0,89

Dari tabel di atas terlihat bahwa pengurangan jumlah fitur justru meningkatkan seluruh metrik pengujian. PSO memberikan peningkatan akurasi pengujian paling tinggi (6,3% lebih baik dari baseline) sekaligus memangkas 56% fitur awal. GA juga memberikan peningkatan yang signifikan, namun dengan tiga fitur lebih banyak. Hal ini menunjukkan

bahwa PSO lebih efisien dalam mengeliminasi fitur redundan atau tidak informatif yang memicu *overfitting* pada data latih.

Meskipun akurasi pelatihan tetap 100%, penurunan dimensi terbukti mengurangi celah antara performa latih dan uji (dari 27,2% menjadi 20,9% pada PSO), sehingga model lebih mampu melakukan generalisasi pada data yang tidak terlihat. F1-score yang lebih tinggi juga menandakan keseimbangan presisi dan *recall* yang lebih baik, terutama pada kelas minoritas (cluster dengan jumlah sampel sedikit).

4.1. Signifikansi Fitur

- Angka Harapan Hidup:
Angka harapan hidup merupakan indikator makro yang merepresentasikan derajat kesehatan dan kesejahteraan suatu populasi. Indikator ini tidak hanya mencerminkan keberhasilan pelayanan kesehatan, tetapi juga berkaitan dengan kualitas lingkungan, kondisi sosial ekonomi, perilaku hidup sehat, serta kecukupan gizi masyarakat. Studi Hobert, et,al (2020) menunjukkan bahwa angka harapan hidup dapat digunakan sebagai proksi untuk menilai kualitas kesehatan lingkungan dan kualitas hidup suatu wilayah.
- Rasio Puskesmas:
Rasio puskesmas menggambarkan tingkat ketersediaan dan keterjangkauan layanan kesehatan dasar bagi masyarakat. Ketersediaan fasilitas pelayanan kesehatan primer yang memadai berperan penting dalam upaya promotif, preventif, deteksi dini penyakit, serta penanganan penyakit yang seharusnya dapat dicegah. WHO menegaskan bahwa penguatan layanan kesehatan primer berkontribusi besar terhadap peningkatan luaran kesehatan dan penurunan kematian yang dapat dicegah. Sesuai dengan penelitian yang dilakukan oleh Fauziah, et.al (2020) yang menyoroti peran vital Puskesmas sebagai garda terdepan dalam memberikan layanan kesehatan primer (antenatal care), yang kualitasnya secara langsung memengaruhi luaran kesehatan ibu dan anak.
- Akses Air Minum yang Layak:
Akses terhadap air minum layak merupakan determinan penting kesehatan masyarakat karena berkaitan erat dengan pencegahan penyakit berbasis lingkungan. Air minum yang tidak aman dapat meningkatkan risiko penyakit diare, kolera, dan gangguan pertumbuhan anak, termasuk stunting. Penelitian Torlesse, H., et al. (2020) menemukan bahwa bahwa anak-anak di rumah tangga dengan fasilitas sanitasi yang lebih baik memiliki kemungkinan 29% lebih rendah untuk mengalami *stunting*.
- Pemberian ASI Eksklusif:
Pemberian ASI eksklusif berperan penting dalam meningkatkan status gizi dan daya tahan tubuh bayi, terutama pada periode awal kehidupan. ASI eksklusif membantu mencegah malnutrisi, infeksi, dan gangguan pertumbuhan, sehingga menjadi salah satu indikator penting dalam penilaian kesehatan ibu dan anak. Anindya, I. G., Salimo, H., & Dewi, Y. L. R. (2020) menunjukkan bahwa terdapat hubungan yang signifikan antara pemberian ASI eksklusif dengan penurunan kejadian *stunting* pada anak balita.
- Balita dengan Gizi Kurang:
Persentase balita dengan gizi kurang merupakan indikator langsung permasalahan gizi masyarakat. Tingginya angka gizi kurang pada balita menunjukkan adanya kerentanan terhadap gangguan pertumbuhan, penurunan kemampuan kognitif, melemahnya sistem imun, serta peningkatan risiko infeksi. Kajian Tafese, Z., et.al (2020) mendukung argumen bahwa gizi kurang adalah masalah yang perlu dikaji serius.
- Angka Kesembuhan COVID-19:
Angka kesembuhan COVID-19 dapat mencerminkan kapasitas sistem kesehatan dalam menangani kasus infeksi, termasuk efektivitas diagnosis, perawatan, pelacakan kasus, serta kebijakan pengendalian pandemi. Studi Zhang Z., et al. (2020)

menunjukkan bahwa tingkat kesembuhan sebagai variabel utama untuk menilai dan membandingkan efisiensi respons sistem kesehatan antar negara dalam menghadapi pandemi.

Temuan ini memperkuat argumen bahwa subset fitur yang dipilih secara metaheuristik tidak hanya meningkatkan akurasi matematis, tetapi juga selaras dengan bukti epidemiologis dan kebijakan kesehatan prioritas.

4.2. Efektivitas Algoritma Metaheuristik

Serangkaian 10 kali percobaan ulang (*trial*) dengan *random seed* berbeda menunjukkan bahwa PSO menghasilkan variasi akurasi pengujian yang lebih kecil (standar deviasi 0,8%) dibandingkan GA (1,5%). Artinya, PSO tidak hanya lebih akurat tetapi juga lebih stabil. Kemampuan PSO dalam menjaga keragaman partikel melalui komponen kecepatan kognitif dan sosial membuatnya lebih tangguh terhadap jebakan lokal pada dataset berdimensi kecil seperti ini. Sebaliknya, GA yang terlalu cepat terkonvergensi oleh kromosom dominan kadang mengarah pada subset fitur yang kurang optimal.

Kedua algoritma mampu memilih subset fitur yang secara signifikan meningkatkan akurasi model prediksi dibandingkan tanpa seleksi fitur. PSO lebih unggul dibandingkan GA dalam hal akurasi pengujian, menunjukkan bahwa pendekatan *swarm* memiliki keunggulan dalam menangani dataset dengan banyak dimensi seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Ringkasan Hasil

Metrik	GA	PSO	Kesimpulan
Jumlah Fitur Terpilih	10	7	GA lebih agresif dalam memilih fitur.
Akurasi Pelatihan	100%	100%	Keduanya mencapai akurasi sempurna pada data latih (potensi overfitting).
Akurasi Pengujian	78%	79.1%	PSO lebih unggul dalam generalisasi.

5. Interpretasi Hasil dan Rekomendasi

Proses yang terdokumentasi secara detail ini menunjukkan bahwa seleksi fitur berbasis PSO dapat diandalkan untuk menyaring indikator kesehatan prioritas. Pemerintah daerah dapat memusatkan intervensi pada enam faktor inti yang muncul secara konsisten – perbaikan akses air minum layak, percepatan cakupan ASI eksklusif, penurunan gizi kurang balita, dan penguatan ketangguhan sistem kesehatan (diindikasikan oleh AHH dan kesembuhan Covid-19) – untuk meningkatkan derajat kesehatan masyarakat secara terukur. Dengan demikian, pendekatan ini bukan sekadar latihan komputasi, melainkan menghasilkan actionable insight yang langsung dapat diadopsi dalam perencanaan program Kesehatan.

4. KESIMPULAN

Algoritma metaheuristik, khususnya PSO, efektif dalam meningkatkan akurasi prediksi tingkat kesehatan masyarakat yang memiliki nilai akurasi 79,1 % dengan memilih subset fitur yang relevan. Faktor-faktor seperti angka harapan hidup, rasio puskesmas, air minum, pemberian asi eksklusif, balita gizi kurang dan angka kesembuhan covid merupakan indikator utama tingkat kesehatan masyarakat di Pulau Jawa. Penelitian ini menunjukkan potensi penggunaan metode metaheuristik dalam pengambilan keputusan berbasis data, terutama dalam domain kesehatan masyarakat.

5. REFERENCES

- [1] Presiden Republik Indonesia, Undang-Undang Republik Indonesia Nomor 36 Tahun 2009 Tentang Kesehatan. Indonesia, 2009, p. 111.
- [2] Presiden Republik Indonesia, Peraturan Presiden Nomor 82 Tahun 2018 Tentang Jaminan Kesehatan. Indonesia, 2018, pp. 1-74.
- [3] Bappenas, "Rencana Pembangunan Jangka Menengah Nasional 2020-2024," Jakarta, 2020.
- [4] Presiden Republik Indonesia, Instruksi Presiden Republik Indonesia Nomor 1 Tahun 2017 Tentang Gerakan Masyarakat Hidup Sehat. Indonesia, 2017, pp. 1-10.
- [5] A. H. Elmi, A. Abdullahi, and M. A. Barre, "A machine learning approach to cardiovascular disease prediction with advanced feature selection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 2, pp. 1030-1041, 2024, doi: 10.11591/ijeecs.v33.i2.pp1030-1041.
- [6] M. A. Islam, M. Z. H. Majumder, M. S. Miah, and S. Jannaty, "Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction," *Comput. Biol. Med.*, vol. 176, Jun. 2024, doi: 10.1016/J.COMPBIOMED.2024.108432.
- [7] M. khanna, L. K. Singh, K. Shrivastava, and R. singh, "An enhanced and efficient approach for feature selection for chronic human disease prediction: A breast cancer study," *Heliyon*, vol. 10, no. 5, p. e26799, 2024, doi: 10.1016/j.heliyon.2024.e26799.
- [8] F. S. Gharehchopogh, M. Namazi, L. Ebrahimi, and B. Abdollahzadeh, "Advances in Sparrow Search Algorithm: A Comprehensive Survey," *Arch. Comput. Methods Eng.*, vol. 30, no. 1, pp. 427-455, 2023, doi: 10.1007/s11831-022-09804-w.
- [9] S. Talatahari, M. Azizi, M. Tolouei, B. Talatahari, and P. Sareh, "Crystal Structure Algorithm (CryStAl): A Metaheuristic Optimization Method," *IEEE Access*, vol. 9, pp. 71244-71261, 2021, doi: 10.1109/ACCESS.2021.3079161.
- [10] M. Jain, V. Saihjal, N. Singh, and S. B. Singh, "An Overview of Variants and Advancements of PSO Algorithm," *Appl. Sci.*, vol. 12, no. 17, pp. 1-21, 2022, doi: 10.3390/app12178392.
- [11] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle Swarm Optimisation: A historical review up to the current developments," *Entropy*, vol. 22, no. 3, pp. 1-36, 2020, doi: 10.3390/E22030362.
- [12] F. M. Isa, W. N. M. Ariffin, M. S. Jusoh, and E. P. Putri, "A Review of Genetic Algorithm: Operations and Applications," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 40, no. 1, pp. 1-34, 2024, doi: 10.37934/araset.40.1.134.
- [13] R. R. Abo-Alsabeh and A. Salhi, "The Genetic Algorithm: A study survey," *Iraqi J. Sci.*, vol. 63, no. 3, pp. 1215-1231, 2022, doi: 10.24996/ijcs.2022.63.3.27.
- [14] J. Wang, H. Ouyang, Z. Zhou, and S. Li, "Harmony Search Algorithm Based on Dual Memory Dynamic Search and Its Application on Data Clustering," *Complex Syst. Model. Simul.*, vol. 3, no. 4, pp. 261-281, 2023, doi: 10.23919/CSMS.2023.0019.
- [15] J. Wang, H. Ouyang, C. Zhang, S. Li, and J. Xiang, "A novel intelligent global harmony search algorithm based on improved search stability strategy," *Sci. Rep.*, vol. 13, no. 1, pp. 1-29, 2023, doi: 10.1038/s41598-023-34736-1.
- [16] J. Lee, Y. Yoon, J. Kim, and Y. H. Kim, "Metaheuristic-Based Feature Selection Methods for Diagnosing Sarcopenia with Machine Learning Algorithms," *Biomimetics*, vol. 9, no. 3, pp. 1-18, 2024, doi: 10.3390/biomimetics9030179.
- [17] S. Taghian and M. H. Nadimi-Shahraki, "Binary Sine Cosine Algorithms for Feature Selection from Medical Data," *Adv. Comput. An Int. J.*, vol. 10, no. 5, pp. 1-10, 2019, doi: 10.5121/acij.2019.10501.
- [18] A. Dyoub and I. Letteri, "Dataset Optimization for Chronic Disease Prediction with Bio-Inspired Feature Selection," 2023, [Online]. Available:

- <http://arxiv.org/abs/2401.05380>
- [19] K. Karimi, A. Ghodratnama, and R. Tavakkoli-Moghaddam, "Two new feature selection methods based on learn-heuristic techniques for breast cancer prediction: a comprehensive analysis," *Ann. Oper. Res.*, vol. 328, no. 1, pp. 665–700, 2023, doi: 10.1007/s10479-022-04933-8.
- [20] F. Keivanian, R. Chiong, and Z. Fan, "A fuzzy adaptive evolutionary-based feature selection and machine learning framework for single and multi-objective body fat prediction," 2023, [Online]. Available: <http://arxiv.org/abs/2303.11949>
- [21] Paput, M. J., Suryowati, K., & Jatipaningrum, M. T. (2023). Perbandingan Metode Random Forest dan Adaptive Boosting pada Klasifikasi Indeks Pembangunan Manusia di Indonesia. *Jurnal Statistika Industri dan Komputasi*, 8(2).
- [22] Pakaya, A. E. S., Hasan, I. K., & Achmad, N. (2025). Perbandingan Support Vector Machine dan Decision Tree Menggunakan Binary Dragonfly Algorithm pada Klasifikasi Indeks Khusus Penanganan Stunting (IKPS) Kabupaten/Kota di Indonesia. *Mathunesa: Jurnal Ilmiah Matematika*, 13(1), 236-245.
- [23] Heryati, A., Terttiaavini, T., Marcelina, D., & Romli, H. (2025). Optimization of Stunting Risk Prediction Using a Hybrid Genetic-Machine Learning Model. *Journal of Artificial Intelligence and Software Engineering (JAISE)*.
- [24] Satria, B., Yoga, S. T. A., & Joko, P. W. (2024). Optimasi Random Forest dengan Genetic Algorithm dan Recursive Feature Elimination pada High Dimensional Data Stunting Samarinda. *Jurnal Media Informatika Budidarma*, 8(3), 1778-1789.
- [25] Dinas Kesehatan Provinsi Jawa Timur, "profil Kesehatan Provinsi Jawa Timur Tahun 2020," Surabaya, 2021. [Online]. Available: www.dinkes.jatimprov.go.id
- [26] Badan Pusat Statistik Provinsi Jawa Timur, "Statistik Kesehatan Provinsi Jawa Timur 2020," Surabaya, 2021. [1][2][3][4][5]
- [27] N. Fauziah *et al.*, "Quality of Antenatal Care at Urban and Rural Puskesmas (Public Health Center) in Jenepono Regency," *Open Access Maced. J. Med. Sci.*, vol. 8, no. T2, pp. 177–182, Sep. 2020, doi: 10.3889/OAMJMS.2020.5223.
- [28] A. Horobet, A. A. Simionescu, D. G. Dumitrescu, and L. Belascu, "Europe's war against covid-19: A map of countries' disease vulnerability using mortality indicators," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, pp. 1–19, Sep. 2020, doi: 10.3390/IJERPH17186565.
- [29] J. H. Rah, S. Sukotjo, N. Badgaiyan, A. A. Cronin, and H. Torlesse, "Improved sanitation is associated with reduced child stunting amongst Indonesian children under 3 years of age," *Matern. Child Nutr.*, vol. 16, no. Suppl 2, p. e12741, Oct. 2020, doi: 10.1111/MCN.12741.
- [30] Z. Tafese, F. R. Alemayehu, A. Anato, Y. Berhan, and B. J. Stoecker, "Child feeding practice and primary health care as major correlates of stunting and underweight among 6- To 23-month-old infants and young children in food-insecure households in ethiopia," *Curr. Dev. Nutr.*, vol. 4, no. 9, Sep. 2020, doi: 10.1093/cdn/nzaa137.
- [31] Z. Zhang, W. Yao, Y. Wang, C. Long, and X. Fu, "Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources," *Journal of Infection*, vol. 81, no. 1, pp. 147–178, Jul. 2020, doi: 10.1016/J.JINF.2020.03.018.