

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemampuan komputasi modern untuk menyimpan, mengolah, dan mengekstrak pola dari data berskala masif secara fundamental mengubah lanskap analisis data, mendorong pengembangan teknik *data mining* yang canggih untuk menggali wawasan strategis dalam pengambilan keputusan (Liu dkk., 2023). Teknik yang banyak digunakan adalah *clustering* atau pengelompokan menempati posisi sentral. *Clustering* termasuk dalam metode *unsupervised learning* yang paling fundamental dengan kapabilitas untuk mengorganisasi data tanpa label ke dalam grup-grup berdasarkan kesamaan intrinsik (Bouabdallaoui dkk., 2024; Lins Galdino & Da Silva, 2024; Singh dkk., 2024).

Algoritma partisi, khususnya K-Means, telah diadopsi secara luas di berbagai domain mulai dari analisis medis hingga logistik karena efisiensi komputasi dan kesederhanaannya (Liu dkk., 2024; Surampudi & Kumudham, 2024; Wang W. dkk., 2025; Wang X. L. dkk., 2024). Namun, performa K-Means secara inheren sangat dipengaruhi oleh penentuan *centroid* awal (Seo dkk., 2025), di mana inisialisasi yang suboptimal dapat menyebabkan konvergensi prematur ke solusi optima lokal, sehingga tidak bisa menemukan pembagian kluster terbaik secara global karena pemilihan *centroid* awal yang kurang tepat (Latief dkk., 2025). Hal ini terjadi karena proses awal K-Means bergantung pada posisi *centroid* yang dipilih secara acak, ketika *centroid* tersebut tidak merepresentasikan distribusi data secara merata, maka perhitungan jarak antar titik data menjadi bias. Akibatnya, algoritma cepat berhenti pada titik konvergensi yang tampak optimal secara lokal, padahal masih terdapat konfigurasi kluster lain yang lebih baik secara keseluruhan (Jia dkk., 2025). Konvergensi merupakan kondisi ketika nilai fungsi objektif tidak berubah secara signifikan antar iterasi (Mohammed & Ali, 2024). Fenomena ini menjadi semakin krusial pada dataset yang kompleks, berdimensi tinggi, atau memiliki struktur kluster non-sferis (Saifullah & Drezewski, 2024).

Meskipun perbaikan seperti K-Means++ telah diusulkan untuk melakukan inisialisasi secara lebih cerdas, metode ini tidak selalu menjamin solusi optima

global, sehingga pengembangan strategi inisialisasi baru tetap menjadi area riset aktif (Zaghdoud dkk., 2024). Kelemahan fundamental K-Means terletak pada proses acak dalam pemilihan *centroid* awal, yang dapat memicu pemilihan *centroid* kurang representatif sehingga algoritma berisiko terjebak pada solusi minimum lokal (*local minimum solution*) dan gagal mencapai solusi optimal global (Latief dkk., 2025). Berdasarkan hal tersebut, algoritma optimasi metaheuristik berbasis populasi, seperti *Genetic Algorithm* (GA) dan *Particle Swarm Optimization* (PSO) menawarkan kemampuan pencarian global dan lokal yang kuat sehingga menjadi kandidat yang sangat potensial untuk mengatasi permasalahan inisialisasi *centroid* pada K-Means (Li dkk., 2024; Saifullah & Dreżewski, 2024).

GA yang terinspirasi dari evolusi Darwin, unggul dalam eksplorasi (*exploration*), memanfaatkan operator genetik seperti *crossover* dan mutasi untuk menjelajahi ruang solusi secara luas dan menjaga keragaman populasi sehingga efektif menghindari optima lokal (Chen dkk., 2025; Karishma & Kumar, 2024). Sebaliknya, PSO yang terinspirasi dari perilaku sosial kawanan dikenal karena kecepatan konvergensi dan kemampuan eksploitasi (*exploitation*) lokalnya yang efisien, di mana partikel bergerak menuju solusi terbaik yang pernah ditemukan (Wang dkk., 2024; Wang dkk., 2024). Akan tetapi, kedua algoritma ini juga memiliki kelemahan inheren jika digunakan secara terpisah. GA cenderung lambat dalam konvergensi (Karishma & Kumar, 2024), sedangkan PSO rentan terhadap konvergensi prematur pada optima lokal (Hassan dkk., 2025; Wang dkk., 2024).

Melihat keunggulan dan kelemahan komplementer ini, pendekatan hybrid yang menggabungkan GA dan PSO telah terbukti sangat efektif dalam berbagai masalah optimasi yang kompleks (Gálvez dkk., 2025; Karishma & Kumar, 2024). Hibridisasi ini bertujuan untuk mensinergikan kekuatan eksplorasi global GA dengan kemampuan eksploitasi lokal dan konvergensi cepat PSO (Azevedo dkk., 2025). Melalui kemampuan eksplorasi GA untuk mengidentifikasi wilayah solusi yang menjanjikan kemudian disempurnakan oleh kemampuan eksploitasi PSO dengan mekanisme *swarm intelligence* untuk menemukan solusi optimal secara cepat, model hybrid ini mampu mencapai keseimbangan yang superior antara pencarian global dan lokal.

Penerapan algoritma tersebut akan diaplikasikan untuk menganalisis Data Gizi Indonesia: Prevalensi dan Konsumsi pada tahun 2023 yang diperoleh dari Badan Pangan Nasional dan Badan Pangan Nasional dan Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional (Bappenas). Dataset yang akan digunakan mencakup berbagai indikator ekonomi, kesehatan, dan sosial seperti konsumsi protein, *prevalence of undernourishment*, jumlah penduduk, skor pola pangan harapan, indeks ketahanan pangan, hingga konsumsi energi. Kesalahan dalam mengelompokkan wilayah berdasarkan profil kecukupan gizi dapat berimplikasi serius, mulai dari alokasi sumber daya, bantuan, dan program yang tidak efisien hingga program intervensi yang salah sasaran.

Pemetaan klaster gizi yang presisi ini memiliki urgensi tinggi untuk mendukung target *Sustainable Development Goals* (SDGs) poin kedua mengenai pengentasan kelaparan, serta selaras dengan agenda strategis nasional dalam menekan prevalensi stunting di berbagai daerah. Oleh karena itu, meningkatkan akurasi *clustering* melalui inisialisasi *centroid* yang superior bukan hanya merupakan tantangan akademis, tetapi juga memiliki implikasi praktis yang signifikan untuk mendukung pengambilan keputusan berbasis bukti dalam upaya peningkatan kecukupan gizi nasional.

Penelitian ini mengusulkan penerapan model Hybrid *Genetic Algorithm* dan *Particle Swarm Optimization* (GAPSO) untuk mengoptimalkan inisialisasi *centroid* pada K-Means *clustering* dengan tujuan meningkatkan akurasi, stabilitas, dan kualitas hasil klasterisasi. Penentuan jumlah cluster dalam penelitian ini dilakukan dengan menguji nilai klaster yaitu $n = 2-10$, mengikuti pendekatan empiris yang umum digunakan dalam studi optimasi K-Means, seperti yang dilakukan oleh Maryam dkk., (2023). Rentang ini dipilih untuk mengeksplorasi secara luas kemungkinan struktur alami data sekaligus menjaga keseimbangan antara representativitas hasil dan efisiensi komputasi. Hasil tersebut akan dibandingkan dengan metode inisialisasi K-Means dan K-Means + GA pada variasi jumlah klaster yaitu $n = 2-10$ serta tingkat K-Means GAPSO *refinement* 0.1–0.9. Pendekatan ini diharapkan mampu menghasilkan *centroid* yang lebih representatif dan mendekati

optimum global sehingga memberikan segmentasi yang lebih akurat terhadap profil kecukupan gizi di Indonesia.

1.2. Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, penelitian ini memiliki tujuan spesifik sebagai berikut:

1. Mengimplementasikan model hybrid *Genetic Algorithm–Particle Swarm Optimization* (GA–PSO) untuk mengoptimalkan inisialisasi *centroid* pada K-Means.
2. Mengevaluasi efektivitas algoritma hibrida yang diusulkan dalam meningkatkan akurasi, stabilitas, dan kualitas *clustering* dibandingkan K-Means standar dan K-Means + GA pada variasi jumlah kluster yaitu $n = 2-10$ serta tingkat *refinement* K-Means GAPSO adalah 0,1–0,9 menggunakan Data Gizi Indonesia tahun 2023.

1.3. Manfaat Penelitian

Hasil penelitian ini diharapkan memberikan kontribusi yang signifikan, baik dari sisi pengembangan ilmu pengetahuan (teoretis) maupun dari sisi penerapan praktis di bidang teknologi dan sistem informasi yaitu sebagai berikut:

1.3.1. Manfaat Teoritis

1. Memperkaya literatur dengan mengusulkan model hybrid GA–PSO untuk mengoptimalkan inisialisasi *centroid* pada K-Means.
2. Kelemahan K-Means konvensional dapat diminimalkan melalui pendekatan optimasi metaheuristik.
3. Memvalidasi metode *hybrid* pada dataset kecukupan gizi multidimensional yang kompleks, sehingga memperluas generalisasi hasil penelitian.

1.3.2. Manfaat Praktis

1. Model *hybrid* yang dikembangkan dapat meningkatkan akurasi dan stabilitas hasil analisis data pada berbagai domain aplikasi.
2. Hasil penelitian menjadi dasar dalam pengembangan kebijakan kecukupan gizi di Indonesia dan dukungan terhadap pemantauan capaian *Sustainable Development Goals* (SDGs), khususnya Tujuan ke-2 (*Zero Hunger*) dan Tujuan ke-3 (*Good Health and Well-Being*).