

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembangunan berkelanjutan telah menjadi pilar utama dalam agenda global sejak *Sustainable Development Goals* (SDGs) diadopsi sebagai pengganti *Millennium Development Goals* (MDGs) pada tahun 2015 (Putra *et al.*, 2024). Kerangka kerja ini terdiri atas 17 tujuan yang mencakup aspek ekonomi, sosial, dan lingkungan, serta 169 target dengan lebih dari 230 indikator yang harus dipantau secara ketat untuk mewujudkan prinsip tidak ada yang tertinggal (*no one left behind*). SDGs merupakan respons global terhadap tantangan pembangunan yang semakin kompleks, mencakup isu perubahan iklim, ketimpangan sosial-ekonomi, serta pengelolaan sumber daya alam secara berkelanjutan. SDGs digunakan sebagai acuan nasional di Indonesia untuk mengukur kemajuan pembangunan secara keseluruhan, tetapi implementasinya di tingkat regional masih menghadapi hambatan berupa keterbatasan sumber daya dan kecenderungan pembangunan yang bersifat parsial.

SDGs memang dirancang sebagai agenda pembangunan dengan target global, tetapi implementasinya di tingkat subnasional seringkali menghadapi tantangan yang lebih kompleks (Meuleman dan Niestroy, 2015). Farida *et al.* (2024) melaporkan bahwa riset terkait SDGs di Indonesia menunjukkan peningkatan yang signifikan selama periode 2016-2024, tetapi beberapa isu penting seperti pasokan air bersih, sanitasi, dan infrastruktur masih belum mendapat perhatian yang memadai. Temuan ini mengindikasikan adanya peluang untuk analisis yang lebih mendalam, terutama dalam hal analisis spasial dan pengelompokan indikator SDGs di tingkat provinsi.

Indonesia masih menghadapi ketimpangan kesejahteraan rakyat antarprovinsi dalam konteks pemerataan pembangunan (Siregar *et al.*, 2025). Indikator seperti ketersediaan pangan, layanan kesehatan, pendidikan, akses air minum, dan sanitasi menunjukkan ketimpangan yang signifikan antarwilayah. Data dari Badan Pusat Statistik (BPS) tahun 2024 menunjukkan bahwa hanya 30,64 persen rumah tangga di Papua Pegunungan yang memiliki akses ke air minum yang layak, jauh di bawah DKI Jakarta sebesar 99,96 persen dan rata-rata nasional sebesar 92,64 persen (BPS, 2024b). Data BPS tahun 2023 pada bidang pendidikan menunjukkan bahwa persentase anak yang tidak bersekolah meningkat dari 0,67 persen pada jenjang SD menjadi 6,93 persen pada jenjang SMP, dan mencapai 21,61 persen pada jenjang SMA, yang mencerminkan adanya disparitas pendidikan yang semakin jelas pada tingkat yang lebih tinggi (BPS, 2023).

Putra *et al.* (2024) menekankan bahwa integrasi norma-norma SDGs ke dalam perencanaan pembangunan regional masih bersifat normatif dan tidak sepenuhnya mencerminkan perbedaan kondisi antarprovinsi. Situasi ini menimbulkan tantangan dalam memastikan pembangunan yang adil dan seimbang di seluruh wilayah Indonesia. Pengelompokan provinsi berdasarkan indikator kesejahteraan merupakan langkah strategis untuk merumuskan kebijakan yang lebih kontekstual, terarah, dan relevan bagi setiap wilayah.

Analisis *cluster* merupakan teknik multivariat yang digunakan untuk mengelompokkan objek dengan karakteristik serupa sehingga pola perbedaan antarkelompok dapat dilihat dengan lebih jelas (Hair *et al.*, 2019). Analisis *cluster* dibagi menjadi dua pendekatan utama, yaitu metode hierarki dan non-hierarki. Metode non-hierarki banyak diterapkan pada data berukuran besar dan beragam karena lebih fleksibel dalam membentuk kelompok dengan jumlah *cluster* yang telah ditentukan sebelumnya (Jain, 2010).

Metode *K-Means* merupakan metode non-hierarki yang populer dalam analisis *cluster*. Metode ini sensitif terhadap *outlier* sehingga dapat menghasilkan *cluster* yang kurang stabil pada data yang mengandung *outlier*. Kaufman dan Rousseeuw (1990) mengembangkan metode *K-Medoids* atau *Partitioning Around Medoids* (PAM) untuk mengatasi kelemahan tersebut dengan menggunakan objek aktual sebagai pusat *cluster* (*medoid*). Pendekatan ini membuat pengelompokan lebih tahan terhadap nilai *outlier* dan menghasilkan representasi *cluster* yang lebih akurat dibandingkan *K-Means*. Metode PAM sering digunakan ketika penelitian memprioritaskan hasil pengelompokan yang lebih stabil terhadap *outlier* dibandingkan kecepatan komputasi.

Metode non-hierarki seperti PAM menggunakan beberapa teknik validasi untuk menentukan jumlah *cluster* optimal, antara lain *Silhouette*, *Elbow*, dan *Gap Statistic*. Penelitian ini menggunakan *Gap Statistic* karena data yang dianalisis bersifat numerik dan berpotensi mengandung *outlier*, sehingga pendekatan statistik yang lebih objektif diperlukan dalam penentuan jumlah *cluster* optimal (Khan *et al.*, 2025). Tibshirani *et al.* (2001) menyatakan bahwa *Gap Statistic* adalah suatu pendekatan statistik untuk mengestimasi jumlah *cluster* optimal dengan cara membandingkan dispersi (tingkat variasi) dalam *cluster* pada data aktual dengan dispersi yang diharapkan dari data acak tanpa pola pengelompokan. Jumlah *cluster* optimal ditentukan pada nilai *K* yang memaksimalkan selisih antara kedua dispersi tersebut.

Penelitian terdahulu telah mengkaji penerapan algoritma *K-Medoids* dalam berbagai sektor pembangunan. Fialine *et al.* (2021) menggunakan metode ini untuk mengelompokkan provinsi di Indonesia berdasarkan delapan indikator pendidikan, yang menghasilkan tiga *cluster* dengan karakteristik capaian pendidikan tinggi, sedang, dan rendah. Penelitian tersebut memberikan kontribusi yang signifikan,

tetapi belum membahas penentuan jumlah *cluster* optimal secara sistematis, sehingga validitas pemilihan jumlah *cluster* masih berpotensi bersifat subjektif dan belum didukung oleh indeks evaluasi internal.

Aspek penentuan jumlah *cluster* optimal secara sistematis kemudian dibahas oleh Rindiawan *et al.* (2025) melalui penerapan algoritma *K-Medoids* pada data penjualan produk kosmetik. Penelitian tersebut menerapkan tiga metode evaluasi, yaitu *Elbow*, *Silhouette*, dan *Gap Statistic* untuk memperoleh jumlah *cluster* optimal. Hasil penelitian menunjukkan bahwa $K = 3$ merupakan jumlah *cluster* yang paling tepat, sebagaimana terlihat dari titik *Elbow* pada $K = 3$ ketika *Sum of Squared Errors* mulai menunjukkan stabilitas, koefisien *Silhouette* sebesar 0,58 yang menunjukkan pemisahan *cluster* baik, dan nilai *Gap* tertinggi pada $K = 3$, yang memperkuat hasil dari dua metode sebelumnya. Ketiga metode tersebut secara konsisten menunjukkan keefektifan algoritma *K-Medoids* dalam mengidentifikasi jumlah *cluster* optimal.

Dyaherawati *et al.* (2025) menerapkan algoritma *K-Medoids* yang dioptimalkan menggunakan *Gap Statistic* untuk mengklasifikasikan wilayah rawan kriminalitas di Indonesia. Penelitian ini menghasilkan empat *cluster* optimal dengan nilai *Gap Statistic* sebesar 0,65, yang mencakup kategori provinsi dari tingkat kriminalitas sangat rawan hingga tidak rawan. Hasil penelitian tersebut menegaskan bahwa integrasi *Gap Statistic* pada algoritma *K-Medoids* efektif dalam menentukan jumlah *cluster* optimal dan meningkatkan kualitas pengelompokan, sehingga pendekatan ini dapat diandalkan untuk aplikasi pengelompokan berbasis data.

Penelitian sebelumnya menunjukkan bahwa kombinasi *K-Medoids* dan *Gap Statistic* memiliki potensi dalam menghasilkan pengelompokan yang reliabel. Penerapan metode tersebut pada indikator SDGs, khususnya terkait aspek

kesejahteraan rakyat, masih terbatas. Indikator kesejahteraan rakyat yang mencakup konsumsi pangan, kesehatan, pendidikan, dan perumahan merupakan dimensi penting dalam menilai capaian pembangunan berkelanjutan (BPS, 2024a). Penelitian ini bertujuan mengatasi keterbatasan penelitian sebelumnya dengan menerapkan metode PAM yang divalidasi menggunakan *Gap Statistic* untuk mengelompokkan provinsi di Indonesia berdasarkan delapan indikator SDGs pada aspek kesejahteraan rakyat. Pendekatan ini menghasilkan pemetaan provinsi yang lebih akurat sebagai dasar perumusan kebijakan pembangunan yang tepat sasaran, efisien, dan berorientasi pada pencapaian SDGs.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil pengelompokan provinsi di Indonesia menggunakan metode PAM dan berapa jumlah *cluster* optimal yang diperoleh berdasarkan validasi *Gap Statistic*?
2. Bagaimana karakteristik setiap *cluster* berdasarkan hasil pengelompokan tersebut sehingga dapat diidentifikasi provinsi yang menjadi prioritas intervensi pemerintah dalam upaya pencapaian indikator SDGs pada aspek kesejahteraan rakyat?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan untuk memperjelas ruang lingkup pembahasan, yaitu sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data indikator SDGs pada aspek kesejahteraan rakyat tahun 2024 yang bersumber dari publikasi Badan Pusat

Statistik (BPS) berjudul *Indikator SDGs Kesejahteraan Rakyat 2024*, dengan unit analisis seluruh provinsi di Indonesia.

2. Variabel yang dianalisis dalam penelitian ini merupakan indikator SDGs pada aspek kesejahteraan rakyat, meliputi prevalensi ketidakcukupan konsumsi pangan, prevalensi kerawanan pangan, persalinan ditolong tenaga kesehatan, *unmet need* pelayanan kesehatan, angka anak tidak sekolah jenjang SMP/ sederajat, Angka Partisipasi Kasar (APK) kuintil terbawah/teratas tingkat SMA/SMK sederajat, akses air minum aman, dan akses sanitasi layak.
3. Analisis pengelompokan dilakukan menggunakan metode PAM dengan validasi *Gap Statistic* untuk menentukan jumlah *cluster* optimal pada rentang $K = 2$ hingga $K = 6$.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan yang disusun untuk menjawab permasalahan yang telah dirumuskan sebelumnya, yaitu sebagai berikut:

1. Menentukan hasil pengelompokan provinsi di Indonesia menggunakan metode PAM serta memperoleh jumlah *cluster* optimal berdasarkan validasi *Gap Statistic*.
2. Mendeskripsikan dan menginterpretasikan karakteristik setiap *cluster* berdasarkan hasil pengelompokan tersebut untuk mengidentifikasi provinsi yang menjadi prioritas intervensi pemerintah dalam upaya pencapaian indikator SDGs pada aspek kesejahteraan rakyat.