

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan merupakan pilar utama dalam membangun kemajuan suatu bangsa serta merupakan investasi strategis dalam peningkatan kualitas sumber daya manusia. Di Indonesia, pendidikan menjadi salah satu fokus utama dalam pembangunan nasional, sebagaimana tercermin dalam komitmen terhadap *Sustainable Development Goals* (SDGs), khususnya pada tujuan keempat yang menekankan pemerataan akses terhadap pendidikan berkualitas sepanjang hayat. Untuk mendukung hal tersebut, pemerintah telah menerapkan program wajib belajar 12 tahun guna memperluas akses sekaligus meningkatkan mutu pendidikan [1].

Mutu pendidikan sangat berpengaruh terhadap kualitas sumber daya manusia dan daya saing suatu negara. Namun, hingga kini pendidikan di Indonesia masih menghadapi beragam persoalan, terutama dalam hal pemerataan kualitas dan perluasan akses [2]. Jika ditinjau berdasarkan jenjang pendidikan, tantangan tersebut lebih terlihat pada jenjang Sekolah Menengah Atas (SMA) yang merupakan tahap akhir dalam program wajib belajar 12 tahun. Berdasarkan data tahun 2024, Angka Partisipasi Sekolah (APS) usia 16–18 tahun masih berada pada kisaran sekitar 70–85 persen, lebih rendah

dibandingkan kelompok usia 7–12 tahun dan 13–15 tahun yang telah mendekati atau mencapai 95 persen. Selain itu, Angka Partisipasi Murni (APM) pada jenjang SMA di sejumlah provinsi masih berada pada kisaran sekitar 60–78 persen. Hal ini menunjukkan bahwa partisipasi pendidikan pada jenjang SMA masih belum setinggi jenjang pendidikan sebelumnya dan terdapat variasi capaian antarprovinsi [3].

Untuk memperoleh gambaran yang lebih mendalam mengenai variasi partisipasi pendidikan jenjang SMA antarprovinsi, diperlukan metode analisis yang mampu mengelompokkan provinsi berdasarkan kesamaan indikator pendidikan. Meskipun perbandingan nilai APM, APK, dan APS secara deskriptif dapat menunjukkan kategori tinggi, sedang, atau rendah, pendekatan tersebut masih bersifat parsial karena setiap indikator dianalisis secara terpisah, sedangkan setiap provinsi memiliki kombinasi karakteristik yang berbeda. Oleh karena itu, digunakan metode *clustering* untuk menganalisis kemiripan beberapa variabel secara simultan, sehingga diperoleh pengelompokan yang lebih objektif sekaligus mampu mengidentifikasi data yang tidak mengikuti pola umum [2].

Berkaitan dengan hal tersebut, *clustering* adalah suatu metode yang digunakan untuk mengelompokkan sekumpulan data ke dalam beberapa kelompok (*cluster*) [4]. Salah satu metode *clustering* yang banyak digunakan adalah *K-Medoids*. *K-Medoids* adalah teknik pengelompokan partisi yang mengelompokkan kumpulan  $n$  objek ke dalam  $k$  *cluster* dengan memanfaatkan objek pada kumpulan objek guna mewakili sebuah *cluster* yang dinamakan

*medoid* [5]. Penelitian yang dilakukan oleh Firdawanti, Ahmad, dan Agustiani tentang *analisis perbandingan kinerja algoritma K-Means dan K-Medoids pada indikator kesehatan dan sosial* menunjukkan bahwa algoritma *K-Medoids* memiliki kemampuan yang lebih baik dalam menghadapi keberadaan *outlier* dibandingkan dengan *K-Means* [6]. Pengelompokan dengan metode *K-Medoids* juga digunakan oleh Asmiatun, dkk dalam pengelompokan kondisi jalan di Kota Semarang [5]. Hardiyanti, dkk menerapkan metode *K-Medoids* pada penanganan kasus diare di Indonesia [7]. Fialine, dkk juga menerapkan metode *K-Medoids* untuk pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan [8].

Selain metode *K-Medoids*, metode *clustering* lain yang sering digunakan adalah *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Metode DBSCAN merupakan bagian dari metode *non-hierarki* yang membentuk *cluster* berdasarkan tingkat kerapatan data. Metode ini memiliki keunggulan dalam mengidentifikasi *outlier* tanpa perlu menentukan jumlah *cluster* sejak awal [1]. Pengelompokan dengan metode DBSCAN telah banyak digunakan dalam berbagai penelitian terdahulu, di antaranya Sari, dkk melakukan pengelompokan penyakit pasien menggunakan metode DBSCAN [9]. Kristiano menerapkan metode DBSCAN untuk pengelompokan minat mahasiswa pascapandemi [10]. Gusti, dkk juga menerapkan DBSCAN dalam pengelompokan kabupaten/kota di Sulawesi Tenggara berdasarkan indikator pendidikan [1].

Penelitian terdahulu menunjukkan bahwa metode *K-Medoids* dan DBSCAN telah banyak digunakan dalam proses *clustering*. Namun, sebagian besar penelitian tersebut masih menerapkan kedua metode secara terpisah atau membandingkannya dengan algoritma *clustering* lain. Sebagai contoh, Adha, dkk membandingkan metode DBSCAN dengan metode *K-Means* dalam pengelompokan kasus COVID-19 di dunia [11]. Sementara itu, Firdawanti, dkk membandingkan metode *K-Medoids* dengan metode *K-Means* dalam menganalisis indikator kesehatan dan sosial [6].

Sementara itu, kajian yang secara khusus membandingkan keduanya dalam konteks pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan, terutama indikator partisipasi pendidikan jenjang SMA, masih sangat terbatas. Kondisi ini menimbulkan *research gap* yang penting untuk diisi, karena perbandingan kedua metode tersebut berpotensi memberikan pemahaman yang lebih komprehensif terkait variasi partisipasi pendidikan SMA antarprovinsi.

Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja metode *K-Medoids* dan DBSCAN dalam mengelompokkan provinsi di Indonesia berdasarkan Angka Partisipasi Murni (APM), Angka Partisipasi Kasar (APK), dan Angka Partisipasi Sekolah (APS) pada jenjang SMA tahun 2024. Evaluasi kualitas hasil pengelompokan dilakukan menggunakan *Silhouette Coefficient* untuk mengukur kesesuaian (*cohesion*) dan pemisahan (*separation*) antar *cluster*. Selain itu, hasil pengelompokan juga dianalisis lebih lanjut dengan melihat karakteristik dan komposisi anggota pada masing-masing metode.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang masalah yang telah dijelaskan sebelumnya, maka dalam penelitian ini akan digunakan rumusan masalah sebagai berikut :

1. Bagaimana hasil pengelompokan provinsi di Indonesia berdasarkan indikator partisipasi pendidikan menggunakan metode *K-Medoids*?
2. Bagaimana hasil pengelompokan provinsi di Indonesia berdasarkan indikator partisipasi pendidikan menggunakan metode DBSCAN?
3. Bagaimana perbandingan hasil pengelompokan antara metode *K-Medoids* dan DBSCAN dalam menggambarkan karakteristik provinsi di Indonesia berdasarkan indikator partisipasi pendidikan?

## 1.3 Batasan Masalah

Penelitian ini dibatasi pada pengelompokan provinsi di Indonesia berdasarkan indikator partisipasi pendidikan jenjang SMA tahun 2024 dengan menggunakan metode *K-Medoids* dan DBSCAN. Pada metode *K-Medoids*, jumlah *cluster* dibatasi pada  $k = 2, 3$ , dan 4 karena berdasarkan evaluasi menggunakan *Silhouette Coefficient*, penambahan jumlah *cluster* di atas nilai tersebut menunjukkan penurunan kualitas pengelompokan. Adapun pada metode DBSCAN, parameter *MinPts* ditetapkan pada variasi 3, 4, 5, dan 6. Pemilihan rentang tersebut mempertimbangkan prinsip kepadatan dalam DBSCAN, di mana nilai *MinPts* yang terlalu kecil dapat menghasilkan

*cluster* yang terlalu longgar, sedangkan nilai yang terlalu besar berpotensi meningkatkan jumlah data yang teridentifikasi sebagai *noise*.

## 1.4 Tujuan Penelitian

Berdasarkan permasalahan yang telah dirumuskan sebelumnya, penelitian ini bertujuan untuk:

1. Mengelompokkan provinsi di Indonesia berdasarkan indikator partisipasi pendidikan menggunakan metode *K-Medoids*.
2. Mengelompokkan provinsi di Indonesia berdasarkan indikator partisipasi pendidikan menggunakan metode DBSCAN.
3. Membandingkan hasil pengelompokan antara metode *K-Medoids* dan DBSCAN dalam menggambarkan karakteristik provinsi di Indonesia berdasarkan indikator partisipasi pendidikan.

## 1.5 Sistematika Penulisan

Penulisan ini terdiri dari lima bab. BAB I Pendahuluan memuat latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan. BAB II Landasan Teori berisi konsep dan teori yang menjadi dasar penelitian. BAB III Metode Penelitian menjelaskan sumber data, variabel penelitian, serta teknik analisis data. BAB IV menyajikan hasil analisis data, dan BAB V memuat kesimpulan serta saran untuk penelitian selanjutnya.