

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan hal yang sangat penting untuk kemajuan sosial dan ekonomi suatu negara karena menjadi dasar utama untuk meningkatkan kualitas orang-orang yang ada di dalamnya. Namun, di Indonesia masalah dalam meratakan akses dan kualitas pendidikan masih menjadi masalah besar, terutama di tingkat menengah atas. Menurut data dari BPS (2024) pada Statistik Pendidikan 2024, rata-rata lama pendidikan untuk orang yang berusia 15 tahun ke atas baru mencapai 9,22 tahun, yang artinya sama dengan lulus dari SMP. Hal ini menunjukkan bahwa banyak orang yang masih belum bisa menyelesaikan pendidikan menengah atas, padahal tingkat pendidikan itu sangat penting untuk menentukan kualitas pekerja dan daya saing suatu wilayah.

Kondisi ini semakin terlihat di Provinsi Kalimantan Timur, yang dalam beberapa tahun terakhir mengalami masalah setiap tahun terkait penerimaan siswa baru. Menurut laporan DPRD Kalimantan Timur tahun 2024, masalah utama timbul karena jumlah tempat di sekolah negeri tidak cukup untuk menampung semua lulusan Sekolah Menengah Pertama (SMP). Banyak calon siswa tidak bisa masuk sekolah dekat rumah dan harus mencari sekolah yang jauh atau berpindah ke sekolah swasta dengan biaya yang lebih mahal. Ketidakmerataan ini disebabkan oleh penyebaran fasilitas yang tidak sama, beberapa sekolah di pusat kota kelebihan murid, sementara di daerah pinggiran justru kurang murid dan tenaga pengajar (Muadin, 2024).

Provinsi Kalimantan Timur merupakan daerah yang berkembang pesat didorong oleh industri minyak dan gas, pembangunan Ibu Kota Nusantara (IKN), dan urbanisasi besar-besaran. Provinsi Kalimantan Timur mengalami peningkatan jumlah siswa sekolah jenjang menengah atas yang tidak sebanding dengan ketersediaan fasilitas pendidikan. Data dari Badan Pusat Statistik (BPS) dan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) tahun 2024 menunjukkan bahwa jumlah siswa per kelas di sekolah jenjang menengah atas Negeri Kalimantan Timur sering kali lebih dari 40 siswa, jauh lebih banyak daripada standar ideal yang seharusnya 32 hingga 36 siswa per kelas (Permendikbud No. 111 Tahun 2014). Keadaan ini menunjukkan bahwa ada kelebihan kapasitas yang berdampak pada kualitas pembelajaran, beban kerja guru, dan kesejahteraan siswa.

Sejalan dengan itu, artikel yang ditulis Muadin (2024) menyatakan bahwa kurangnya infrastruktur dan distribusi sekolah yang tidak merata masih menjadi masalah besar di wilayah Provinsi Kalimantan Timur. Perbedaan dalam akses pendidikan antara daerah kota dan desa membuat beberapa siswa di lokasi terpencil kesulitan untuk melanjutkan ke sekolah jenjang menengah atas karena masalah jarak, biaya, dan kurangnya fasilitas. Masalah ini tidak hanya mempengaruhi kesempatan belajar yang sama, tetapi juga berdampak pada kualitas sumber daya manusia yang dihasilkan. Oleh karena itu, penting untuk melakukan pemetaan kondisi pendidikan menengah di Kalimantan Timur untuk mendukung kebijakan pembangunan yang adil.

Pada penelitian statistik, masalah ini menarik untuk dilihat dari segi lokasi karena distribusi sekolah, kapasitas ruang kelas, dan jumlah siswa adalah data yang

terkait dengan geografi. Analisis lokasi bisa membantu untuk menemukan pola ketimpangan antar area berdasarkan posisi sekolah. Pendekatan pengelompokan data geografis merupakan sebuah solusi yang menjanjikan untuk menemukan pola penyebaran sekolah negeri jenjang menengah atas di Provinsi Kalimantan Timur. Akan tetapi, metode pengelompokan tradisional seperti *K-Means* kurang tepat untuk data spasial karena menganggap bahwa *cluster* memiliki bentuk bulat atau konveks dan memiliki variasi yang seragam. Sementara itu, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) klasik membutuhkan penentuan parameter ϵ dan minPts secara menyeluruh, yang membuatnya tidak efektif dalam menangani *cluster* dengan kepadatan yang sangat beragam dan rentan menghasilkan banyak *noise* pada data yang kompleks (Campello et al., 2013). Salah satu metode yang cocok untuk analisis seperti ini adalah *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN).

Penggunaan algoritma HDBSCAN telah terbukti efektif untuk *clustering* data geospasial berbasis *latitude* dan *longitude*, sebagaimana diterapkan oleh Mou et al. (2020) dalam mengekstrak *Area of Interest* (AOI) wisata dari data *geotagged* Flickr di Shanghai, yang memungkinkan identifikasi *cluster* dengan densitas bervariasi dan integrasi kendala spasial untuk analisis alur wisata *inbound*. Penelitian itu menunjukkan bahwa HDBSCAN mampu menangani pola *cluster* yang tidak teratur, berbagai dimensi, serta jumlah *cluster* yang tidak tetap. Metode ini dianggap paling tepat untuk menganalisis data lokasi yang rumit, termasuk data sebaran sekolah yang memiliki perbedaan besar antar daerah. Pada penelitian ini akan memakai metode HDBSCAN untuk mengelompokkan data sekolah negeri

jenjang menengah atas di Provinsi Kalimantan Timur berdasarkan data spasial berupa *longitude* dan *latitude*.

Penelitian ini mengadopsi metode HDBSCAN dengan merujuk pada studi Amalina dan Fauzan (2024) yang berhasil memetakan potensi wilayah di Kabupaten Buleleng menggunakan data indikator fasilitas publik dan kependudukan. Pemilihan metode ini didasarkan pada keunggulan HDBSCAN dalam mengidentifikasi pola spasial yang kompleks pada berbagai tingkat kepadatan tanpa memerlukan penentuan jumlah *cluster* di awal penelitian. Hal ini sangat relevan untuk memetakan sekolah di Kalimantan Timur yang memiliki sebaran tidak merata, sehingga pembentukan *cluster* dapat dilakukan secara adaptif berdasarkan variasi kerapatan lokal di setiap wilayah. Kemudian ditegaskan bahwa algoritma berbasis kepadatan memberikan lebih banyak manfaat dibandingkan teknik konvensional seperti *K-Means*. Fokus utamanya adalah fleksibilitas HDBSCAN dalam mengelola dataset dengan ukuran dan tingkat kepadatan *cluster* yang bervariasi, sehingga memberikan solusi analisis spasial yang lebih akurat untuk mengidentifikasi wilayah potensial atau sebaran fasilitas.

Selain itu, penelitian ini juga akan melakukan evaluasi hasil pengelompokan dengan menggunakan *Density-Based Clustering Validation* (DBCV). Moulavi et al. (2014), menyatakan bahwa DBCV adalah ukuran yang paling tepat untuk metode yang berbasis kepadatan. Berbeda dengan metode validasi lain seperti *Silhouette* atau *Calinski-Harabasz* yang lebih sesuai untuk metode berdasarkan jarak, DBCV dapat mengukur kualitas pengelompokan berdasarkan seberapa padat titik data dalam tiap kelompok, sehingga hasil evaluasi yang dihasilkan akan lebih akurat dan cocok dengan cara kerja HDBSCAN.

Sejalan dengan penelitian Riayah dan Mahdy (2025), penerapan DBCV terbukti efektif dalam memvalidasi struktur *cluster* yang stabil, dimana dalam konteks pemetaan indikator wilayah, metode ini mampu menghasilkan nilai validitas sebesar 0,5506. Capaian nilai tersebut mengindikasikan bahwa objek-objek di dalam *cluster* memiliki kepadatan yang tinggi dan setiap kelompok terpisah secara signifikan. Evaluasi ini didukung oleh penggunaan jarak *Euclidean* untuk mengukur kedekatan antar titik data guna memastikan ketepatan geometri dalam ruang spasial.

Provinsi Kalimantan Timur terpilih dalam penelitian karena wilayah ini memiliki banyak sekolah yang tidak tersebar secara merata, selain itu juga menjadi menjadi tempat pembangunan IKN yang baru. Data yang digunakan dalam penelitian ini diambil dari *website* Kementerian Pendidikan dan Menengah berupa jumlah data satuan pendidikan yang ada di Kalimantan Timur yang kemudian dilakukan pengecekan letak geografis menggunakan *Google Maps*. Penelitian ini melakukan penyaringan data guna menjaga homogenitas karakteristik objek. Fokus penelitian dibatasi pada sekolah yang terdiri dari SMA Negeri dan SMK Negeri. Sekolah kategori khusus seperti SLB Negeri dan MAN sengaja tidak diikutsertakan karena memiliki karakteristik operasional serta segmentasi siswa yang spesifik. Dengan pendekatan yang berbasis kuantitatif dan spasial menggunakan metode HDBSCAN serta evaluasi DBCV, penelitian ini diharapkan dapat memberikan pemahaman yang lebih jelas tentang ketidakmerataan kapasitas sekolah menengah atas di Kalimantan Timur dan menjadi saran bagi pemerintah daerah dalam merencanakan pemerataan fasilitas pendidikan di masa depan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pola persebaran sekolah negeri jenjang menengah atas di Provinsi Kalimantan Timur berdasarkan data koordinat geografis *latitude* dan *longitude*?
2. Bagaimana hasil *clustering* lokasi sekolah menggunakan metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN)?
3. Bagaimana hasil evaluasi performa *cluster* dengan menggunakan metode validasi *Density-Based Clustering Validation* (DBCV) untuk menilai kualitas hasil *cluster* berbasis kepadatan?
4. Bagaimana wilayah yang tergolong memiliki kepadatan sekolah tinggi dan wilayah mana yang relatif rendah, berdasarkan hasil *clusterisasi* metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN)?

1.3 Batasan Masalah

Agar penelitian ini terfokus dan terarah, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan terbatas pada data spasial sekolah negeri jenjang menengah atas di Provinsi Kalimantan Timur, yang terdiri atas koordinat *latitude* dan *longitude* pada setiap sekolah.
2. Metode *clustering* yang digunakan dalam penelitian ini adalah *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN).

3. Evaluasi hasil *clustering* dilakukan menggunakan metode *Density-Based Clustering Validation* (DBCV) yang sesuai untuk model berbasis kepadatan.
4. Penelitian ini bersifat eksploratif deskriptif, dan mengidentifikasi pola persebaran spasial sekolah di Provinsi Kalimantan Timur.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis pola persebaran spasial sekolah negeri jenjang menengah atas di Provinsi Kalimantan Timur berdasarkan data koordinat geografis yaitu *latitude* dan *longitude* untuk mengidentifikasi wilayah dengan kepadatan sekolah tinggi maupun rendah.
2. Menerapkan metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) dalam proses *clustering* data spasial sekolah negeri jenjang menengah atas di Provinsi Kalimantan Timur dengan tujuan mengelompokkan lokasi sekolah Negeri jenjang menengah atas yang memiliki kedekatan geografis dan kepadatan tertentu.
3. Mengevaluasi hasil *clusterisasi* menggunakan metode *Density-Based Clustering Validation* (DBCV) sebagai ukuran validasi yang sesuai dengan metode berbasis kepadatan, sehingga diperoleh penilaian yang objektif terhadap kualitas hasil *cluster*.
4. Menginterpretasikan hasil *clustering* secara deskriptif untuk memberikan gambaran mengenai distribusi spasial sekolah Negeri jenjang menengah atas di Provinsi Kalimantan Timur yang dapat digunakan sebagai bahan pertimbangan dalam perencanaan pemerataan fasilitas pendidikan.