BAB I
PENDAHULUAN

1.1 Latar Belakang

Indonesia sebagai negara kepulauan terbesar di dunia dengan 34 provinsi memiliki kompleksitas demografis yang sangat tinggi dan unik. Setiap provinsi memiliki karakteristik demografis, sosial, ekonomi, dan geografis yang berbeda-beda, menciptakan tantangan tersendiri dalam perumusan kebijakan pembangunan nasional yang efektif dan berkeadilan (Bimbim Oktaviandi et al., 2024). Keragaman ini tercermin dalam berbagai indikator kependudukan seperti jumlah penduduk, laju pertumbuhan, tingkat kepadatan, struktur umur, tingkat pendidikan, dan kondisi kesejahteraan yang sangat bervariasi antar wilayah. Dinamika kependudukan yang kompleks ini mempengaruhi berbagai aspek pembangunan, mulai dari perencanaan infrastruktur, alokasi anggaran pembangunan, distribusi pelayanan publik, hingga strategi pemberdayaan masyarakat, yang pada akhirnya berdampak pada pencapaian target pembangunan nasional (Aliabit et al., 2023).

Ketimpangan demografis antar provinsi menciptakan berbagai tantangan dalam implementasi program pembangunan nasional. Pemerintah menghadapi kesulitan dalam mengalokasikan sumber daya secara optimal, merancang program yang tepat sasaran, dan memastikan pemerataan pembangunan di seluruh wilayah Indonesia (Hilmiah et al., 2025). Kondisi ini diperparah oleh kurangnya pemahaman mendalam tentang pola dan karakteristik demografis yang spesifik di setiap wilayah, sebagaimana tercermin dalam disparitas capaian Indeks Pembangunan Manusia (IPM) yang masih menunjukkan kesenjangan signifikan antar wilayah selama dua dekade terakhir. Data migrasi dan perpindahan penduduk seringkali tersebar dan sulit dianalisis tanpa alat yang tepat, padahal informasi ini sangat penting untuk memahami dinamika sosial-ekonomi yang mempengaruhi pergerakan penduduk. Penelitian terbaru menunjukkan bahwa pendekatan data-driven dalam pengembangan kebijakan pembangunan menjadi semakin penting untuk mengatasi kompleksitas demografis Indonesia, karena

analisis berbasis data dapat memberikan *insight* yang lebih objektif dan akurat dalam mengidentifikasi kebutuhan spesifik setiap wilayah (Sabrina & Setiawan, 2025).

Analisis *clustering* telah terbukti sebagai metode yang *powerful* untuk memahami pola dan struktur data demografis yang kompleks. Teknik ini memungkinkan pengelompokan wilayah berdasarkan kemiripan karakteristik demografis, sehingga dapat mengidentifikasi wilayah-wilayah dengan kebutuhan pembangunan yang serupa. *Clustering* membantu pemerintah dalam menentukan prioritas alokasi infrastruktur di wilayah padat penduduk atau peningkatan kesejahteraan di wilayah dengan pertumbuhan rendah, sekaligus memungkinkan perumusan kebijakan berbasis klaster yang lebih tepat sasaran. Selain itu, metode *clustering* juga memungkinkan pemantauan progress pembangunan secara lebih sistematis berdasarkan karakteristik klaster dan dapat mengidentifikasi wilayah dengan karakteristik demografis yang unik atau ekstrem yang memerlukan perhatian khusus (Bintang & Allo, 2024).

Oleh karena itu dibutuhkan suatu ilmu yang mampu memproses data tersebut secara efesien dan akurat. *Data mining* menjadi solusi dari proses eksplorasi dan analisis data dengan tujuan mengidentifikasi pola-pola yang bermanfaat, menemukan hubungan tersembunyi, serta memperoleh informasi yang dapat diambil dari kumpulan data besar. Melalui berbagai teknik dan algoritma, *data mining* membantu dalam mengenali tren, mengklasifikasikan data, dan membuat prediksi yang berguna untuk mendukung pengambilan keputusan (Azia & Shaib, 2024).

Salah satu algoritma *clustering* yang terbukti efektif dalam mengelompokkan data dengan distribusi yang tidak merata adalah DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). DBSCAN bekerja dengan cara mengelompokkan data berdasarkan kepadatan data, sehingga dapat mendeteksi *cluster* dengan bentuk yang tidak teratur, serta mampu mengidentifikasi *outlier*. Algoritma ini memiliki keunggulan dibandingkan dengan metode *clustering* lainnya, seperti *K-Means*, yang membutuhkan

penentuan jumlah *cluster* sebelumnya dan kurang efektif dalam menangani data yang tidak terstruktur dengan baik (Miftahurrahmi et al., 2024).

Pada penelitian yang dilakukan oleh Yanto, Ahmad Homaidi, Ahmad Luthi pada tahun 2024, yang berjudul ”Implementasi Metode *clustering* dengan Algoritma DBSCAN Untuk Identifikasi Sentra Industri Berbasis *Google Maps*”. Penelitian ini menunjukkan hasil analisis bahwa algoritma DBSCAN berhasil mengelompokkan lokasi industri ke dalam beberapa *cluster* dengan densitas tinggi, yang diidentifikasi sebagai sentra industri. Analisis lebih lanjut menggunakan aglomerasi mengungkapkan adanya potensi sentra industri baru dengan komoditi kerupuk poli di wilayah Kecamatan Asembagus. Hasil ini dapat digunakan oleh pemerintah daerah Kabupaten Situbondo sebagai salah satu dasar untuk pengembangan wilayah dan perencanaan ekonomi yang lebih baik (Yanto et al., 2024).

Penelitian yang serupa juga dilakukan oleh Ihsan Bagus Fahad Arafat, Mokhamad Amin Hariyadi, Irwan Budi Santoso, Cahyo Crysdian pada tahun 2023, yang berjudul ”*Clustering* Gempabumi Di Wilayah Regional Vii Menggunakan Pendekatan DBSCAN”. Pada penelitian ini digunakan pendekatan algoritma Unsupervised Learning *clustering* berbasis kepadatan yaitu, *Density Based Spatial Clustering of Application with Noise* atau DBSCAN, algoritma ini membutuhkan parameter input *epsilon* (ε) dan MinPts. Data yang digunakan pada penelitian ini adalah data gempabumi wilayah Regional VII tahun 2017 hingga 2021 yang diperoleh dari BMKG. Selanjutnya, proses *clustering* dilakukan dengan membagi data gempabumi berdasarkan periode yaitu periode tahunan dan periode lima tahun dengan tujuan untuk mengetahui pola *cluster* berdasarkan periode waktu. Hasil yang terbentuk selanjutnya dievaluasi menggunakan *silhouette coefficient* serta dibandingkan dengan peta seismisitas Jawa yang telah ada dari katalog PuSGeN 2017. Hasil *clustering* menggunakan DBSCAN diperoleh jumlah *cluster* sebanyak 2 hingga 6 *cluster* dengan nilai *Silhouette coefficient* terendah sebesar 0.270 untuk periode T5_2017-2021 dan tertinggi sebesar 0.499 untuk periode T1_2020 (Bagus et al., 2023).

Meskipun demikian, penerapan algoritma DBSCAN dalam konteks *clustering* provinsi di Indonesia berdasarkan karakteristik demografis masih sangat terbatas. Penelitian yang secara khusus menerapkan DBSCAN untuk *clustering* provinsi berdasarkan karakteristik demografis komprehensif belum banyak dilakukan, padahal algoritma ini memiliki potensi besar untuk memberikan hasil *clustering* yang lebih akurat dalam mengidentifikasi pola demografis. Penelitian terbaru (2025) mengeksplorasi parameter tuning pada algoritma DBSCAN untuk optimalisasi hasil *clustering*, menekankan pentingnya metodologi yang sistematis dalam penerapan DBSCAN, namun belum ada yang fokus pada aplikasi demografis tingkat provinsi. Algoritma DBSCAN dapat membentuk *cluster* secara bebas dan tidak beraturan sehingga proses pembentukan *cluster* dapat lebih fleksibel dalam menangani karakteristik demografis yang beragam.

Gap penelitian yang teridentifikasi adalah kurangnya penerapan algoritma DBSCAN secara komprehensif dan sistematis untuk *clustering* provinsi di Indonesia berdasarkan karakteristik demografis. Padahal, algoritma DBSCAN memiliki keunggulan signifikan dalam menangani data demografis yang seringkali memiliki distribusi kompleks, kepadatan yang bervariasi, dan keberadaan *outlier* demografis yang perlu diidentifikasi secara khusus. Kemampuan DBSCAN dalam mendeteksi *cluster* dengan bentuk tidak beraturan juga sangat relevan untuk data demografis yang tidak selalu mengikuti pola distribusi uniform. Selain itu, kemampuan deteksi *outlier* otomatis pada DBSCAN dapat membantu mengidentifikasi provinsi-provinsi dengan karakteristik demografis yang unik atau ekstrem yang memerlukan pendekatan pembangunan yang berbeda.

Penelitian ini diharapkan dapat memberikan kontribusi signifikan berupa penerapan algoritma DBSCAN dalam konteks analisis demografis provinsi di Indonesia yang lebih komprehensif dan sistematis. Hasil penelitian dapat mengidentifikasi pola dan kelompok provinsi berdasarkan karakteristik demografis yang dapat mendukung perumusan kebijakan pembangunan daerah yang lebih efektif dan berbasis data objektif. Selain itu, penelitian ini diharapkan

dapat mendorong pembangunan yang lebih inklusif dan berkelanjutan melalui pemahaman yang lebih baik tentang karakteristik demografis setiap wilayah, sehingga program pembangunan dapat dirancang sesuai dengan kebutuhan spesifik masing-masing klaster provinsi. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan metodologi *clustering* untuk analisis demografis, tetapi juga memberikan basis empiris yang kuat untuk pengambilan keputusan kebijakan pembangunan regional yang lebih tepat sasaran dan efektif.

Berdasarkan penjelasan di atas, maka penelitian yang akan dilakukan adalah "*Clustering* Provinsi Di Indonesia Berdasarkan Karakteristik Demografis Menggunakan Algoritma *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN)".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, maka permasalahan yang penulis rumuskan adalah :

1. Bagaimana cara mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik demografis mereka menggunakan algoritma DBSCAN untuk menghasilkan pola distribusi penduduk yang lebih akurat dan relevan ?
2. Bagaimana merancang dan membangun sistem berbasis *website* yang dapat mengimplementasikan algoritma DBSCAN dalam pengelompokan provinsi berdasarkan karakteristik demografis di Indonesia ?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas maka tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik demografis mereka menggunakan algoritma DBSCAN, guna mengidentifikasi pola distribusi penduduk yang lebih akurat dan relevan.
2. Untuk merancang dan membangun sistem berbasis *website* yang dapat mengimplementasikan algoritma DBSCAN dalam pengelompokan provinsi

berdasarkan karakteristik demografis, yang dapat digunakan sebagai alat bantu untuk analisis data demografis dan pengambilan keputusan.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Penelitian ini diharapkan dapat menambah wawasan dalam bidang ilmu *data mining*, khususnya dalam penerapan metode DBSCAN untuk klasterisasi data demografis. Penelitian ini juga dapat berkontribusi pada pengembangan literatur mengenai pemanfaatan *data mining* untuk pengelompokan wilayah berdasarkan karakteristik demografis.
2. Penggunaan metode DBSCAN dalam penelitian ini diharapkan dapat menjadi acuan bagi pengembangan sistem berbasis yang lebih canggih untuk keperluan analisis dan pengelolaan data demografis, sehingga dapat mempermudah visualisasi dan interpretasi data bagi para pengambil kebijakan.
3. Hasil penelitian ini dapat digunakan oleh pemerintah dan para pemangku kepentingan untuk mendukung pengambilan keputusan yang lebih tepat dalam distribusi sumber daya, serta kebijakan sosial dan ekonomi. Pengelompokan provinsi berdasarkan karakteristik demografis akan membantu dalam memahami kondisi masing-masing wilayah sehingga kebijakan yang diterapkan dapat lebih tepat sasaran.
4. Sebagai sarana untuk menerapkan pengetahuan peneliti yang diperoleh selama menempuh masa studi Strata-1 Teknik Informatika Universitas Malikussaleh.

1.5 Ruang Lingkup dan Batasan Penelitian

Adapun batasan masalah yang ditetapkan pada penelitian ini adalah sebagai berikut :

1. Metode yang digunakan dalam penelitian ini dibatasi pada algoritma *Density Based Spatial Clustering of Applications with Noise* (DBSCAN).
2. Data yang digunakan berupa data historis 2020-2023 dengan jumlah 136 data.

3. Sumber data penelitian ini berasal dari *www.bps.go.id*.
4. Variabel yang digunakan adalah jumlah penduduk, laju pertumbuhan penduduk per tahun, persentase penduduk, kepadatan penduduk, dan rasio jenis kelamin penduduk.
5. Jumlah *cluster* dibagi menjadi 3: tinggi, sedang, dan rendah dengan memberikan perbedaan setiap *cluster* berdasarkan warna, merah untuk *cluster* tinggi, kuning untuk *cluster* sedang, dan hijau untuk *cluster* rendah.