

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

keberhasilan pembangunan ekonomi tidak sekadar diukur dari tingginya angka pertumbuhan ekonomi, melainkan dalam meningkatkan taraf hidup masyarakat secara nyata melalui perbaikan kualitas hidup, pengentasan kemiskinan, dan pemerataan distribusi pendapatan (Hasanah et al., 2025:95). Dalam konteks pembangunan di Provinsi Jawa Timur, ketimpangan capaian antar-daerah masih menjadi masalah struktural yang sulit diselesaikan. Perbedaan sumber daya alam dan kondisi demografis antarwilayah memicu ketimpangan pembangunan, yang berpengaruh langsung pada tingkat kesejahteraan masyarakat serta arah kebijakan pembangunan pemerintah daerah (Ravi et al., 2024). Selama ini, analisis wilayah kurang optimal akibat proses pengolahan data dari portal *open-data* indikator daerah yang cenderung disamaratakan secara manual, tanpa mempertimbangkan bahwa karakteristik komparatif serta variasi kondisi daerah merupakan gabungan multidimensi dari berbagai indikator yang saling mengikat.

Memastikan integritas dan konsistensi data merupakan hal krusial saat menghubungkan berbagai sumber data. Menjaga keselarasan antar sumber data menjadi tantangan besar, terutama akibat kendala seperti ketidaksesuaian kosa kata, tidak adanya pengidentifikasi *dataset*, serta minimnya informasi (Flausino et al., 2024). Penyajian data wilayah saat ini masih didominasi oleh tabel angka dan

masih sangat jarang ditampilkan dalam bentuk peta yang sesuai dengan kaidah kartografi (Putri et al., 2025). Format data mentah yang berupa tabel angka padat ini membuat masyarakat umum, akademisi, peneliti, maupun analis kebijakan pemula membutuhkan waktu lebih lama untuk menarik kesimpulan. Mereka mengalami keterbatasan dalam memetakan dan membandingkan posisi capaian antar daerah secara cepat apabila data tersebut tidak disajikan dalam bentuk visualisasi spasial.

Untuk mengatasi masalah tersebut, diperlukan penyediaan *website open-data analytics* yang fleksibel dan terbuka untuk umum. Sistem yang dibangun dalam penelitian ini menerapkan teknik *data mining* berbasis *website* yang dirancang secara dinamis. Algoritma *K-Means Clustering* diintegrasikan sebagai mesin pengolah utama pada *backend website* untuk mengelompokkan wilayah di Provinsi Jawa Timur secara otomatis berdasarkan data dari portal *open-data* indikator daerah yang diunggah oleh pengguna. *Website open-data analytics* ini dirancang secara fleksibel agar mampu menerima pengunggahan data numerik dari *dataset open-data*, dengan syarat minimal menyertakan 2 variabel numerik dalam satu aspek atau indikator yang sejenis (seperti deret waktu PDRB atau perbandingan PDRB antarwilayah) dan tidak dapat digabungkan dengan aspek yang berbeda.

Pengelompokan objek data ke dalam *cluster* terdekat umumnya dihitung menggunakan metode *Euclidean Distance* (Nisa & Kurniawati, 2025). Jika pengguna hanya mengunggah 1 variabel saja, maka pemetaan koordinat posisi

objek tidak dapat membentuk ruang dimensi, sehingga proses perhitungan matrik jarak antar daerah tidak dapat berjalan secara akurat. Dengan mewajibkan minimal 2 variabel numerik, sistem dapat membandingkan capaian indikator secara berpasangan (*head-to-head*), serta memastikan proses *clustering* menghasilkan pembagian wilayah yang valid, objektif, dan terukur. Pengguna cukup mengunggah file Excel (.xlsx/.xls) atau CSV melalui antarmuka *website*, dan sistem secara otomatis akan memisahkan kolom identitas daerah serta mengekstraksi seluruh kolom numerik untuk diproses secara *real-time*.

Proses komputasi pada *backend* mengelola seluruh tahapan pra-pemrosesan data secara otomatis untuk menjamin validitas perhitungan sebelum algoritma dijalankan. Untuk menentukan jumlah *cluster* (k) yang paling ideal secara objektif, sistem mengalokasikan data ke dalam titik pusat *cluster* (*centroid*) pada perulangan rentang nilai k_{min} sampai k_{max} , lalu mengevaluasi performa tiap model menggunakan metrik *Davies-Bouldin Index* (DBI) dan *Silhouette Score*. Jumlah *cluster* optimal (k) kemudian dipilih berdasarkan skor evaluasi terbaik tersebut, untuk menghindari kegagalan pencocokan data akibat kemiripan nama daerah, sistem menerapkan standarisasi Kode Wilayah sebagai kunci relasi utama.

Melalui arsitektur yang terintegrasi ini, seluruh perhitungan rumit algoritma dikelola sepenuhnya oleh sistem secara otomatis, sehingga hasil analisis dapat langsung disajikan kepada pengguna secara *real-time*. *Website* memanfaatkan bahasa pemrograman *Python* dengan *framework FastAPI* karena kemampuannya

yang sangat responsif dalam mengeksekusi fungsi matematika algoritma *K-Means* serta menyediakan *endpoint* API secara *real-time*. Sementara pada tampilan *website*, dirancang menggunakan *React (Vite)* dan *Tailwind CSS* untuk menciptakan antarmuka yang modern, responsif, serta rapi saat menyajikan visualisasi data kepada pengguna.

Website ini mengubah *dataset* mentah yang diunggah pengguna menjadi peta interaktif Provinsi Jawa Timur berbentuk peta gradasi warna menggunakan *library react-leaflet*. Warna peta akan langsung berubah otomatis membentuk gradasi warna sesuai dengan pengelompokan daerah berdasarkan *centroid* terdekat dari jumlah *cluster* optimal yang terbentuk. Sistem juga akan menampilkan grafik DBI & *Silhouette Score* sebagai dasar analisis validitas jumlah cluster yang terpilih.

Untuk mengeksplorasi data, visualisasi peta dirancang ketika pengguna mengklik langsung salah satu wilayah pada peta interaktif, lalu sistem akan mengarahkan pengguna ke halaman detail wilayah. Halaman ini akan menampilkan peta wilayah terpilih yang langsung diperbesar secara otomatis. Secara bersamaan, *frontend* akan merender panel detail wilayah dengan memunculkan *Grouped Bar Chart*. Grafik ini dirancang untuk menyajikan perbandingan *head-to-head* antara nilai asli data mentah dengan nilai rata-rata (*centroid*) dari keseluruhan anggota di dalam kelompok *clusternya*. Melalui visualisasi dua batang berdampingan untuk setiap indikator yang diunggah, pengguna dapat langsung mengidentifikasi dan menganalisis alasan mengapa suatu wilayah dikelompokkan ke dalam *cluster* tertentu.

Melalui pendekatan ini, penelitian yang dilakukan diharapkan bisa memberikan manfaat baik secara teori maupun praktik. Dari sisi akademis, studi ini diharapkan dapat menambah referensi pemanfaatan *sains data* dalam menghubungkan ilmu komputer dengan perencanaan wilayah sosial-ekonomi melalui implementasi sistem informasi *website open-data analytics* yang adaptif, dinamis, dan fleksibel. Sementara dari sisi praktis, *website* ini dapat digunakan secara luas oleh masyarakat umum, akademisi, dan peneliti sebagai alat bantu strategis untuk mempercepat identifikasi karakteristik wilayah, mengeksplorasi potensi serta tantangan daerah, serta mendemokratisasi data dari portal *open-data* indikator daerah menjadi informasi spasial yang mudah dipahami demi mendukung kebijakan pembangunan yang lebih inklusif di Provinsi Jawa Timur.

B. Pembatasan Masalah

Agar penelitian ini tetap berjalan sesuai dengan perencanaan yang telah ditetapkan, maka batasan masalah dirumuskan sebagai berikut:

1. *Dataset* yang digunakan merupakan data dari portal *open-data* indikator wilayah Jawa Timur berformat .csv, .xlsx, atau .xls. Sistem dirancang secara fleksibel agar mampu menerima pengunggahan data numerik dari *dataset open-data*, dengan syarat minimal menyertakan 2 variabel numerik dalam satu aspek atau indikator yang sejenis (seperti deret waktu PDRB atau perbandingan PDRB antarwilayah) dan tidak digabungkan dengan aspek yang berbeda.

2. Penelitian ini dibatasi pada 29 kabupaten dan 9 kota di Jawa Timur agar sesuai dengan ketersediaan format peta wilayah spasial (*TopoJSON*) yang diintegrasikan dalam sistem.
3. Metode pengelompokan wilayah menggunakan algoritma *K-Means Clustering* pada *backend (FastAPI)*, di mana penentuan anggota *cluster* dihitung berdasarkan kedekatan jarak spasial terkecil ke titik pusat *cluster (centroid)*. Adapun pemilihan jumlah *cluster (k)* optimal ditentukan secara otomatis pada rentang nilai $K = 2$ sampai $K = 5$ menggunakan metrik evaluasi *Davies-Bouldin Index (DBI)* dan *Silhouette Score*.
4. *Website open-data analytics* dikembangkan menggunakan *React (Vite)* dan *Tailwind CSS*. Hasil pemetaan disajikan melalui peta spasial interaktif dengan gradasi warna berbasis *react-leaflet*, grafik evaluasi performa model (*DBI & Silhouette Score*), serta *Grouped Bar Chart* untuk membandingkan capaian nilai indikator tiap daerah terhadap nilai rata-rata (*centroid*) *clusternya*.

C. Perumusan Masalah

Berdasarkan latar belakang dan pembatasan masalah diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan membangun arsitektur *website open-data analytics* untuk memproses data dari portal *open-data* indikator daerah secara dinamis?

2. Bagaimana mengimplementasikan algoritma *K-Means Clustering* ke dalam *backend* sistem agar mampu mengelompokkan 38 kabupaten/kota di Provinsi Jawa Timur secara otomatis berdasarkan kedekatan jarak ke *centroid*?
3. Bagaimana mengevaluasi performa model *clusterisasi* menggunakan metrik *Davies-Bouldin Index* (DBI) dan *Silhouette Score* untuk menentukan jumlah *cluster* optimal pada rentang nilai $K = 2$ sampai $K = 5$ secara otomatis?

D. Tujuan Penelitian

Berdasarkan perumusan masalah yang telah disusun, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk merancang dan membangun arsitektur *website open-data analytics* untuk memproses data dari portal *open-data* indikator daerah secara dinamis.
2. mengimplementasikan algoritma *K-Means Clustering* ke dalam *backend* sistem agar mampu mengelompokkan 38 kabupaten/kota di Provinsi Jawa Timur secara otomatis berdasarkan kedekatan jarak ke *centroid*.
3. Untuk mengevaluasi performa model *clusterisasi* menggunakan metrik *Davies-Bouldin Index* (DBI) dan *Silhouette Score* untuk menentukan jumlah *cluster* optimal pada rentang nilai $K = 2$ sampai $K = 5$ secara otomatis.

E. Kegunaan Penelitian

1. Kegunaan Teoritis

Secara teoritis, penelitian ini diharapkan dapat berkontribusi pada perkembangan ilmu pengetahuan serta menjadi referensi bagi peneliti lain dalam mengembangkan sistem pengelompokan wilayah menggunakan algoritma *K-Means Clustering*. Penelitian ini juga diharapkan dapat menjadi acuan akademik dalam merancang *website open-data analytics* yang fleksibel, adaptif, dan dapat digunakan secara luas oleh masyarakat umum.

2. Kegunaan Praktis

a. Bagi Mahasiswa

Penelitian ini bermanfaat dalam meningkatkan keterampilan dalam mengolah data mentah, menyelaraskan data spasial berdasarkan Kode Wilayah pada peta interaktif, serta menyajikan *Grouped Bar Chart* untuk visualisasi perbandingan detail indikator tiap daerah terhadap rata-rata (*centroid*) *clusternya*. Selain itu, penelitian ini juga mengasah kemampuan dalam merancang antarmuka pengguna yang responsif guna menyajikan hasil analisis data publik agar lebih mudah dipahami oleh masyarakat luas.

b. Bagi Fakultas/Program Studi

Penelitian ini diharapkan dapat menjadi bahan referensi akademik bagi fakultas maupun program studi, khususnya dalam bidang pemanfaatan *data sains* dan rekayasa perangkat lunak. Hasil dari penelitian ini juga berfungsi sebagai contoh studi kasus mengenai cara mengintegrasikan

konsep analisis statistik, pemodelan algoritma *clustering*, dan pengembangan *website open-data* dalam satu kesatuan sistem yang utuh.

c. Bagi Masyarakat

Penelitian ini diharapkan dapat membantu masyarakat umum, akademisi, serta analis kebijakan pemula dalam membaca dan memahami data dari portal *open-data* indikator daerah dengan mudah melalui visualisasi peta. Melalui *website* ini, pengguna dapat mengidentifikasi karakteristik dan pola penyebaran indikator daerah di Jawa Timur secara mandiri, cepat, dan objektif melalui tampilan visualisasi peta dan grafik yang interaktif.

d. Bagi Peneliti Selanjutnya

Penelitian ini diharapkan dapat menjadi rujukan bagi peneliti lain yang ingin mengembangkan sistem *open-data analytics* serupa. Peneliti selanjutnya dapat memperluas kajian ini dengan membandingkan algoritma pengelompokan lain selain K-Means atau memperbarui fitur interaktif pada *website* agar dapat menyajikan visualisasi data yang lebih mendalam.