

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan fenomena multidimensional yang tidak hanya menjadi isu kemanusiaan global, tetapi juga menjadi barometer utama keberhasilan pembangunan di tingkat nasional dan regional. Sebagai agenda prioritas, upaya penanggulangan kemiskinan menuntut formulasi kebijakan yang tepat sasaran, yang hanya dapat dicapai melalui pemahaman mendalam terhadap karakteristik wilayah. Dalam konteks Indonesia, Provinsi Jawa Timur memegang peranan vital sebagai salah satu penopang utama perekonomian nasional. Namun, di balik signifikansi ekonominya, provinsi tersebut menghadapi tantangan internal yang kompleks, yaitu adanya disparitas atau kesenjangan sosio-ekonomi yang tajam di antara 38 kabupaten/kota yang ada di wilayahnya.

Realitas kesenjangan tersebut terefleksi dengan jelas dalam rilis data Badan Pusat Statistik (BPS). Data menunjukkan variasi yang signifikan tidak hanya pada angka persentase kemiskinan, tetapi juga pada kedalaman dan keparahan masalah yang dihadapi. Pemerintah daerah dihadapkan pada sebuah tantangan, di mana kebijakan yang bersifat seragam (satu-ukuran-untuk-semua) hampir pasti tidak akan efektif. Setiap wilayah memiliki kontur masalah yang unik. Oleh karena itu, diperlukan sebuah metode pemetaan yang objektif untuk mengidentifikasi wilayah-wilayah dengan karakteristik serupa, sehingga intervensi kebijakan dapat dirancang secara lebih terfokus dan efisien.

Tantangan utama dalam melakukan pemetaan tersebut tidak hanya terletak pada jumlah wilayah yang banyak, tetapi juga pada kompleksitas dan multidimensionalitas data. Untuk memahami kemiskinan secara komprehensif, analisis tidak dapat bergantung pada satu indikator tunggal. Data kemiskinan BPS mencakup sekumpulan variabel krusial, seperti Persentase Penduduk Miskin (P0), Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2).

Ketiga variabel inti tersebut memberikan gambaran utuh: P0 mengukur kuantitas, sementara P1 dan P2 mengukur kualitas (seberapa dalam dan parah) kemiskinan. Sebuah wilayah dengan P0 rendah namun P2 tinggi (sedikit penduduk miskin, tetapi sangat parah) memerlukan intervensi yang berbeda secara fundamental dibandingkan wilayah dengan P0 tinggi namun P2 rendah (banyak penduduk miskin, tetapi tidak terlalu parah).

Meskipun penelitian mengenai pengelompokan kemiskinan telah banyak dilakukan, seperti oleh Habibi dan Nursyanti (2023), sebagian besar masih menggunakan indikator statistik dasar kemiskinan saja. Padahal, kemiskinan merupakan masalah multidimensi yang juga dipengaruhi oleh kualitas sumber daya manusia dan ketersediaan lapangan kerja.

Oleh karena itu, penelitian ini memandang perlu untuk mengembangkan analisis dengan menyertakan variabel Indeks Pembangunan Manusia (IPM) dan Tingkat Pengangguran Terbuka (TPT). Penambahan variabel tersebut membuat struktur data menjadi lebih kompleks, sehingga diperlukan penerapan algoritma *data mining* (K-Means dan Fuzzy C-Means) untuk menemukan pola pengelompokan yang lebih presisi dan representatif dibandingkan penelitian sebelumnya.

Untuk mengatasi tantangan komputasi tersebut, disiplin ilmu penambangan data (*data mining*) menawarkan solusi yang kuat melalui teknik pengelompokan (*clustering*). Pengelompokan adalah sebuah metode *unsupervised learning* yang dirancang untuk mengidentifikasi dan mempartisi sekumpulan objek data ke dalam beberapa grup berdasarkan kesamaan karakteristik intrinsiknya. Manfaat utamanya adalah mentransformasi data mentah yang kompleks menjadi kelompok-kelompok yang bermakna. Hasil pengelompokan yang valid secara statistik memungkinkan pembuat kebijakan untuk merancang intervensi yang spesifik dan relevan bagi setiap klaster, alih-alih memberlakukan kebijakan yang dipukul rata.

Dalam konteks akademik, pendekatan tersebut sejalan dengan konsep Sistem Pendukung Keputusan (SPK) yang menekankan pemanfaatan model dan analisis

data untuk membantu proses pengambilan keputusan yang lebih rasional dan berbasis bukti. SPK tidak bertujuan menggantikan peran pengambil kebijakan, melainkan menyediakan informasi terstruktur, alternatif analisis, serta hasil komputasi yang dapat dijadikan dasar pertimbangan. Hasil clustering yang dihasilkan melalui penerapan algoritma K-Means dan Fuzzy C-Means pada penelitian ini berperan sebagai komponen analitis dalam kerangka Sistem Pendukung Keputusan, khususnya dalam mendukung pemetaan wilayah prioritas penanggulangan kemiskinan di Provinsi Jawa Timur.

Dalam domain pengelompokan, terdapat dua filosofi utama yang sering diperdebatkan, yaitu pendekatan partisi tegas atau *hard clustering* dan partisi kabur *soft clustering*. Pendekatan *hard clustering* yang paling fundamental dan populer adalah algoritma K-Means. K-Means bekerja dengan logika yang lugas: setiap objek data secara eksklusif hanya dapat menjadi anggota dari satu kluster saja. Banyak penelitian telah membuktikan efektivitas K-Means dalam berbagai studi kasus, termasuk pengelompokan wilayah kemiskinan di Banten (Munandar, 2022) dan pemetaan prioritas bantuan (Kurniawan et al., 2023). Keunggulan K-Means terletak pada efisiensi komputasi dan kemudahan interpretasi hasilnya.

Sebagai antitesis dari K-Means, algoritma Fuzzy C-Means (FCM) hadir sebagai representasi utama dari *soft clustering*. FCM beroperasi di atas landasan logika fuzzy, yang mengasumsikan bahwa batas antar kelompok seringkali tidak tegas atau kabur. FCM tidak memaksa sebuah kabupaten/kota untuk masuk ke satu kluster. Sebaliknya, FCM memberikan derajat keanggotaan yaitu sebuah nilai antara 0 dan 1 yang menunjukkan seberapa mirip sebuah wilayah dengan semua kluster yang ada. Sebuah wilayah bisa jadi 70% anggota Kluster A (Miskin Sedang) dan 30% anggota Kluster B (Miskin Rendah). Fleksibilitas tersebut secara teoretis dianggap lebih realistis untuk menangani fenomena sosio-ekonomi seperti kemiskinan, yang pada dasarnya adalah sebuah kontinum, bukan sebuah kondisi biner yang kaku. Efektivitas FCM dalam memetakan data sosial yang kabur telah ditunjukkan dalam studi pengelompokan kemiskinan di Jawa Tengah (Faturahman & Hidayati, 2025)

Adanya dua pendekatan metodologis yang berbeda filosofi antara partisi tegas K-Means versus partisi kabur FCM menimbulkan sebuah dilema. Studi komparatif di berbagai bidang menunjukkan bahwa tidak ada satu algoritma yang superior secara universal, kinerjanya sangat bergantung pada struktur dan karakteristik unik dari dataset yang diuji (Latifa et al., 2021;(Siregar et al., 2022). Hal tersebut mengarah pada kesenjangan penelitian (research gap) yang spesifik belum adanya validasi empiris yang membandingkan secara kuantitatif antara pendekatan partisi tegas (K-Means) dan partisi kabur (FCM) untuk menemukan model pengelompokan yang paling optimal bagi dataset spesifik kemiskinan multidimensi di Provinsi Jawa Timur tahun 2024 yang menambahkan variabel Indeks Pembangunan Manusia (IPM) dan Tingkat Pengangguran Terbuka (TPT).

Memilih algoritma yang tidak optimal dapat menghasilkan struktur kluster yang menyesatkan, yang pada gilirannya dapat berujung pada perancangan kebijakan yang keliru. Oleh karena itu, sekadar menerapkan algoritma tidaklah cukup, diperlukan sebuah proses evaluasi yang objektif untuk memvalidasi hasilnya. Guna mengisi kesenjangan tersebut, penelitian akan melakukan perbandingan kinerja secara kuantitatif.

Untuk mengukur kualitas kluster yang terbentuk, penelitian akan menggunakan dua indeks validasi internal yang telah diakui secara luas. Pertama, Silhouette Coefficient (SC), yang mengukur kualitas pengelompokan berdasarkan dua metrik kohesi (seberapa padat data dalam satu kluster) dan separasi (seberapa jauh jarak antar kluster). Nilai SC berkisar dari -1 hingga 1, di mana nilai yang lebih tinggi atau semakin mendekati 1 menunjukkan struktur kluster yang lebih baik. Kedua, Davies-Bouldin Index (DBI), yang mengukur rasio rata-rata sebaran data di dalam kluster terhadap jarak antar kluster. Berbeda dengan SC, nilai DBI yang lebih rendah atau mendekati 0 mengindikasikan pengelompokan yang lebih optimal. Penggunaan kedua metrik evaluasi tersebut akan memberikan penilaian yang komprehensif.

Meskipun sejumlah penelitian telah melakukan perbandingan antara algoritma K-Means dan Fuzzy C-Means pada data kemiskinan, sebagian besar studi tersebut

masih menggunakan indikator statistik dasar kemiskinan tanpa melibatkan variabel yang merepresentasikan dimensi pembangunan manusia dan kondisi ketenagakerjaan. Padahal, kemiskinan merupakan fenomena multidimensi yang tidak hanya dipengaruhi oleh aspek ekonomi semata, tetapi juga kualitas sumber daya manusia dan tingkat partisipasi kerja.

Dengan demikian, masih terdapat kebutuhan untuk melakukan analisis komparatif pada dataset yang memiliki struktur lebih kompleks melalui penambahan indikator Indeks Pembangunan Manusia (IPM) dan Tingkat Pengangguran Terbuka (TPT). Kompleksitas struktur data tersebut berpotensi menghasilkan karakteristik kluster yang berbeda, sehingga diperlukan validasi kuantitatif untuk menentukan algoritma yang paling optimal dan representatif dalam memetakan kemiskinan multidimensi di Provinsi Jawa Timur.

Selain menghasilkan pengelompokan wilayah, hasil clustering perlu disajikan dalam bentuk yang mudah dipahami oleh pengguna. Penyajian hasil dalam bentuk tabel atau grafik statis sering kali kurang efektif untuk menggambarkan perbedaan karakteristik antar wilayah. Oleh karena itu, diperlukan visualisasi berbasis dashboard yang mampu menampilkan hasil clustering, perbandingan kluster, serta sebaran wilayah secara interaktif.

1.2 Identifikasi Masalah

Berdasarkan dari latar belakang yang telah diuraikan, beberapa permasalahan dapat diidentifikasi sebagai berikut:

1. Pengelompokan wilayah kemiskinan masih menggunakan indikator ekonomi dasar dan belum mempertimbangkan indikator kualitas pembangunan manusia serta tingkat pengangguran secara bersamaan.
2. Belum diketahui algoritma yang lebih optimal antara K-Means atau Fuzzy C-Means ketika diterapkan pada dataset dengan 7 indikator di Jawa Timur.
3. Belum adanya media visualisasi yang mampu menampilkan hasil clustering, perbandingan cluster, sebaran wilayah kemiskinan secara lebih informatif.

1.3 Tujuan Penelitian

Adapun tujuan dari dilaksanakannya penelitian ini adalah sebagai berikut:

1. Melakukan pengelompokan wilayah kemiskinan di Provinsi Jawa Timur menggunakan algoritma K-Means dan Fuzzy C-Means berdasarkan 7 indikator.
2. Menganalisis perbandingan kinerja kedua algoritma menggunakan Silhouette Coefficient (SC) dan Davies-Bouldin Index (DBI) untuk menentukan algoritma yang paling optimal dan valid.
3. Menampilkan hasil clustering wilayah kemiskinan dalam bentuk dashboard visualisasi untuk mempermudah analisis dan pemahaman sebaran kluster.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan merupakan data sekunder tahun 2024 yang bersumber dari publikasi BPS "Jawa Timur dalam Angka 2025" dan rilis data terkait di *website* resmi BPS. Ruang lingkup penelitian terbatas pada 38 kabupaten/kota di Provinsi Jawa Timur.
2. Variabel (atribut) yang dianalisis terbatas pada 7 indikator: Persentase Penduduk Miskin (P0), Indeks Kedalaman Kemiskinan (P1), Indeks Keparahan Kemiskinan (P2), Garis Kemiskinan, Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan Rata-Rata Pengeluaran per Kapita
3. Algoritma *clustering* yang dianalisis dan dibandingkan hanya K-Means dan Fuzzy C-Means (FCM).
4. Sistem visualisasi yang dikembangkan bersifat sebagai prototipe dashboard analitis dan tidak mencakup integrasi dengan sistem pemerintahan atau basis data real-time

1.5 Kontribusi

Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi Penulis Menambah wawasan dan pemahaman dalam menerapkan teknik data mining serta konsep Sistem Pendukung Keputusan, khususnya algoritma K-Means dan Fuzzy C-Means, untuk studi kasus data kemiskinan. Selain itu, penelitian ini menjadi salah satu syarat untuk menyelesaikan pendidikan Sarjana (S1).
2. Bagi Akademik, menyajikan bukti empiris mengenai perbandingan kinerja K-Means dan FCM pada data sosio-ekonomi. Hasil studi dapat menjadi referensi akademis bagi penelitian selanjutnya di bidang data mining dan pengembangan Sistem Pendukung Keputusan yang berfokus pada analisis kemiskinan.
3. Bagi Instansi Pemerintahan, menghasilkan pemetaan klaster kemiskinan di Jawa Timur yang divalidasi secara kuantitatif. Hasil pemetaan dapat menjadi bahan pertimbangan dalam kerangka Sistem Pendukung Keputusan bagi Pemerintah Provinsi Jawa Timur atau BAPPEDA dalam merumuskan kebijakan pengentasan kemiskinan yang lebih terfokus.

