

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesejahteraan merupakan komponen penting dalam kemajuan suatu negara. Kesejahteraan rakyat didefinisikan sebagai kemampuan individu untuk memenuhi kebutuhan hidup yang layak (Yulianto dan Hidayatullah, 2014). Kesejahteraan dapat dilihat dari terpenuhinya kebutuhan dasar dan kemampuan seseorang dalam menjalankan fungsi sosial di masyarakat. Permasalahan kesejahteraan masih menjadi tantangan, terutama bagi masyarakat yang belum mampu memenuhi kebutuhan dasar akibat terhambatnya ekonomi dan sosial (Khairi *et al.*, 2021). Peningkatan kesejahteraan masyarakat telah menjadi prioritas di Indonesia, namun pencapaiannya masih menunjukkan ketimpangan antarwilayah termasuk di Provinsi Jawa Tengah.

Provinsi Jawa Tengah masih menghadapi tantangan dalam pemerataan kesejahteraan rakyat antar kabupaten/kota. Data Badan Pusat Statistik menunjukkan tingkat kemiskinan di beberapa kabupaten di Provinsi Jawa Tengah masih berada di atas 15% pada tahun 2023 (Badan Pusat Statistik, 2024). Kesenjangan tersebut tidak hanya terlihat dalam aspek kemiskinan, tetapi juga pada akses pendidikan, layanan kesehatan, dan infrastruktur dasar yang belum merata. Badan Pusat Statistik (BPS) telah menetapkan indikator untuk mengukur kesejahteraan rakyat yang mencakup delapan indikator, yaitu kependudukan, kesehatan dan gizi, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, perumahan dan lingkungan, kemiskinan, serta sosial lainnya.

Beberapa penelitian terdahulu telah menerapkan metode *clustering* untuk menganalisis kesejahteraan rakyat dengan menggunakan berbagai indikator. Wahyuni dan Wulandari (2022) menggunakan metode analisis *cluster* hierarki dengan tiga komponen indikator utama yaitu ekonomi, kesehatan, dan pendidikan. Indikator ekonomi yang diukur melalui tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, dan persentase penduduk miskin. Indikator kesehatan yang mencakup akses air minum layak, sanitasi layak, dan keluhan kesehatan. Indikator pendidikan meliputi angka partisipasi sekolah dan rata-rata lama sekolah.

Pendekatan indikator yang berbeda dilakukan pada penelitian Ramadhani *et al.* (2018) yang menerapkan metode hierarki. Pada indikator ekonomi menggunakan persentase penduduk miskin, rumah tangga penerima bantuan beras raskin, dan pengeluaran per kapita untuk selain makanan. Pada indikator pendidikan, penelitian tersebut menekankan pada penduduk miskin yang tidak tamat pendidikan dasar dan angka melek huruf. Indikator kesehatan dan kesejahteraan, mereka memperluas cakupan dengan mengukur kepemilikan alat kontrasepsi, luas lantai rumah, akses air bersih, dan kepemilikan jaminan kesehatan. Kedua penelitian tersebut memberikan landasan metodologis yang berharga untuk penerapan analisis *cluster* dalam konteks pengukuran kesejahteraan rakyat dengan berbagai kombinasi indikator.

Berdasarkan penelitian Djumura *et al.* (2022) menjelaskan bahwa Program Keluarga Harapan (PKH) berpengaruh signifikan terhadap kesejahteraan masyarakat melalui bantuan yang diberikan. Julyantari *et al.* (2021) menekankan pentingnya status gizi balita sebagai indikator kesejahteraan rakyat karena status

gizi balita menunjukkan kondisi kesehatan individu yang sejalan dengan indikator kesejahteraan. Berdasarkan penelitian Putra (2022), jumlah kendaraan bermotor berpengaruh terhadap kesejahteraan. Penelitian tersebut mengungkapkan bahwa peningkatan jumlah kendaraan bermotor berkorelasi positif dengan pendapatan masyarakat.

Pada penelitian ini, kesejahteraan rakyat diukur melalui 16 variabel yang terdiri atas 10 variabel numerik dan 6 variabel kategorik. Penentuan pengukuran indikator kesejahteraan didasarkan pada sumber Badan Pusat Statistik dan penelitian terdahulu terkait kesejahteraan. Indikator kependudukan terdiri dari laju pertumbuhan penduduk. Indikator kesehatan mencakup persentase penduduk dengan keluhan kesehatan dan jumlah bayi gizi kurang. Indikator pendidikan meliputi angka harapan lama sekolah dan indeks pembangunan literasi masyarakat. Indikator ketenagakerjaan terdiri dari persentase tingkat pengangguran terbuka, persentase tingkat partisipasi angkatan kerja, dan rata-rata gaji bersih pekerja formal.

Indikator taraf dan pola pengeluaran diukur melalui persentase pengeluaran per kapita sebulan untuk makanan. Indikator perumahan mencakup persentase akses air minum layak, sumber air minum utama, dan persentase rumah tangga dengan sanitasi layak. Indikator kemiskinan terdiri dari persentase penduduk miskin. Indikator sosial lainnya meliputi persentase rumah tangga penerima Program Keluarga Harapan (PKH), persentase penduduk penerima BPJS-PBI, dan jumlah kendaraan bermotor.

Metode *clustering* merupakan pendekatan yang efektif untuk mengelompokkan daerah berdasarkan kesamaan karakteristik (Johnson dan Wichern, 2007). Keunggulan metode ini adalah kemampuan untuk mengelompokkan wilayah berdasarkan karakteristik serupa tanpa memerlukan label awal, sehingga pemerintah dapat menyusun strategi pemerataan yang lebih tepat sasaran. Pada penelitian yang dikembangkan oleh Raman *et al.* (2010), untuk mengelompokkan data campuran, pendekatan transformasi digunakan untuk mengubah data kategorik menjadi numerik sebelum proses *clustering*. Pendekatan transformasi lebih efisien dalam proses komputasi, namun tantangan utama dari metode transformasi yaitu dapat menghilangkan wawasan penting dari data asli.

Penelitian mengenai pengelompokan data campuran sebelumnya dikembangkan oleh He *et al.* (2002), metode *ensemble* untuk data campuran adalah alternatif dari metode transformasi yang mengombinasikan hasil dari beberapa metode *clustering* untuk menghasilkan hasil pengelompokan yang lebih optimal (He *et al.*, 2002). Reddy dan Kavitha (2012) mengembangkan metode *ensemble* untuk menangani data campuran yaitu *Similarity Weight and Filter Method* (SWFM) dengan konsep *similarity weight* untuk mengukur tingkat kemiripan dan *filter method* sebagai parameter jarak. Evaluasi kinerja pengelompokan yang digunakan dalam penelitian Reddy dan Kavitha (2012) adalah dengan pengukuran tingkat kesalahan *cluster*. ander dan Zain (2018) juga melakukan pengelompokan di Provinsi Jawa Timur menggunakan SWFM. Hasil penelitian mengidentifikasi pengelompokan yang baik dengan lima *cluster* wilayah berdasarkan uji validitas simpangan baku.

Metode *Similarity Weight and Filter Method* (SWFM) memiliki beberapa kelebihan dalam pengelompokan data. SWFM mampu menangani data campuran yang terdiri dari variabel numerik dan kategorik secara bersamaan tanpa memerlukan transformasi data. Pendekatan ini menggunakan konsep *similarity weight* yang mengukur tingkat kemiripan antar objek, sehingga dapat mempertahankan karakteristik asli dari kedua jenis data tersebut. Konsep *filter method* dalam SWFM berfungsi sebagai parameter jarak yang efektif untuk menyaring objek yang tidak relevan.

Penelitian ini menggunakan metode *Similarity Weight and Filter Method* (SWFM) dengan kombinasi yang berbeda dari penelitian sebelumnya. Berbeda dengan penelitian Reddy dan Kavitha (2012) yang menggunakan K-Means dan K-Prototype, penelitian ini menggabungkan metode *hierarchical agglomerative clustering* untuk data numerik dan algoritma QROCK untuk data kategorik. Pemilihan *hierarchical agglomerative clustering* dengan pautan *complete linkage*, *average linkage*, *ward's method*, dan jarak *Euclidean* didasarkan pada penelitian Murtagh dan Contreras (2017) yang dapat mengidentifikasi data tanpa jumlah *cluster* di awal dan menghasilkan visualisasi melalui *dendrogram* yang memudahkan interpretasi hasil.

Pengelompokan data kategorik dilakukan dengan menggunakan metode QROCK mengacu pada penelitian Tyagi dan Sharma (2018) yang membuktikan bahwa QROCK lebih efisien dibandingkan algoritma *clustering* kategorik lainnya, serta ketahanannya terhadap *noise* dan *outlier*. Metode QROCK menggunakan *threshold* (θ) 0,3; 0,33; 0,35; 0,38; 0,4; 0,43; 0,45; 0,48; 0,5 untuk mendapatkan

cluster yang optimal. Setelah terbentuk *final cluster* dari data numerik dan data kategorik, kedua *final cluster* digunakan sebagai input pada metode SWFM. Penelitian ini menggunakan *Silhouette Index* sebagai ukuran validitas *cluster*, berbeda dengan penelitian Reddy dan Kavitha (2012) yang menggunakan tingkat kesalahan *cluster*. Nilai *Silhouette Index* dipilih untuk analisis validitas *cluster* karena mampu mengukur kecocokan objek dalam *cluster* tertentu (Hasan, 2024).

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan, rumusan masalah yang dikaji dalam penelitian ini meliputi:

1. Bagaimana hasil pengelompokan yang terbentuk dari kabupaten/kota di Jawa Tengah menggunakan metode *Similarity Weight and Filter Method* (SWFM) berdasarkan indikator kesejahteraan rakyat?
2. Bagaimana *profiling* Kabupaten/Kota di Jawa Tengah berdasarkan indikator kesejahteraan rakyat?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini meliputi:

1. Data yang digunakan adalah data dari Badan Pusat Statistik Jawa Tengah yang mencakup indikator kesejahteraan dari 35 kabupaten/kota di Provinsi Jawa Tengah tahun 2023 yang diunggah pada Tahun 2024.
2. Variabel yang digunakan terbagi menjadi dua, yaitu variabel numerik dan variabel kategorik. Variabel numerik terdiri dari laju pertumbuhan penduduk, persentase penduduk dengan keluhan kesehatan, angka harapan lama sekolah, persentase tingkat pengangguran terbuka, persentase tingkat

partisipasi angkatan kerja, persentase pengeluaran per kapita sebulan untuk makanan, persentase akses air minum layak, persentase penduduk miskin, persentase rumah tangga penerima Program Keluarga Harapan (PKH), dan persentase penduduk penerima BPJS Penerima Bantuan Iuran (BPJS-PBI). Variabel kategorik terdiri dari rata-rata gaji bersih pekerja formal, indeks pembangunan literasi masyarakat, sumber air minum utama, persentase rumah tangga dengan sanitasi layak, jumlah bayi gizi kurang, dan jumlah kendaraan bermotor.

3. Pengelompokan data numerik dilakukan melalui metode pengelompokan hierarki *agglomerative* dengan menggunakan jarak *Euclidean*, dengan tiga jenis pautan yang digunakan, yaitu *complete linkage*, *average linkage*, dan *ward's method*.
4. Pengelompokan data kategorik menggunakan metode *Quick Robust Clustering Using Links* (QROCK) dengan nilai *threshold* (θ) yang digunakan adalah 0,3; 0,33; 0,35; 0,38; 0,4; 0,43; 0,45; 0,48; 0,5.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis hasil pengelompokan kabupaten/kota di Jawa Tengah menggunakan metode *Similarity Weight and Filter Method* (SWFM).
2. Menganalisis *profiling* kabupaten/kota di Jawa Tengah berdasarkan indikator kesejahteraan rakyat.