

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini umumnya menyerang paru-paru, tetapi dapat menyebar ke organ tubuh lainnya. TBC masih menjadi salah satu penyebab utama kematian akibat penyakit infeksi tunggal di dunia. Penyakit TBC memiliki beban yang tinggi secara global, dan Indonesia memberikan kontribusi kasus yang signifikan. Berdasarkan laporan Kementerian Kesehatan Republik Indonesia, Indonesia menempati peringkat kedua di dunia dalam jumlah kasus TBC tertinggi. Provinsi Jawa Barat merupakan salah satu wilayah dengan jumlah kasus TBC terbesar di Indonesia pada tingkat regional, yaitu sebesar 234.710 kasus pada tahun 2024 (Kementerian Kesehatan RI, 2025).

Tingginya angka kejadian TBC tidak hanya dipengaruhi oleh faktor medis, tetapi juga oleh berbagai indikator determinan sosial kesehatan, seperti kepadatan penduduk, tingkat kemiskinan, sanitasi lingkungan, serta kualitas hunian. Faktor-faktor tersebut berkontribusi terhadap ketimpangan distribusi kasus antarkabupaten/kota, di mana beberapa wilayah menunjukkan prevalensi yang relatif tinggi, sementara wilayah lainnya menunjukkan prevalensi yang lebih rendah (Lönnroth *et al.*, 2009). Berdasarkan kondisi tersebut, penelitian ini memerlukan pendekatan analisis untuk mengelompokkan wilayah berdasarkan kesamaan indikator kesehatan dan sosial sehingga intervensi dan kebijakan penanggulangan TBC dapat dirumuskan secara lebih tepat sasaran dan efektif.

Upaya pengendalian tuberkulosis di tingkat daerah masih menghadapi berbagai kendala dalam penerapan perencanaan berbasis data. Permasalahan tersebut mencakup keterbatasan pelaporan dari fasilitas kesehatan swasta, ketidakkonsistenan pengiriman data dari puskesmas, serta belum optimalnya integrasi sistem informasi kesehatan berbasis digital (WHO Indonesia, 2025). Kondisi tersebut menyebabkan kebijakan kesehatan cenderung bersifat umum dan belum sepenuhnya mempertimbangkan perbedaan karakteristik sosial dan kesehatan antarwilayah. Pemetaan berbasis indikator lokal berperan penting dalam menentukan prioritas sumber daya dan merancang intervensi yang lebih tepat sasaran (Hanson *et al.*, 2017). Berdasarkan kondisi tersebut, analisis pengelompokan wilayah menjadi pendekatan yang relevan untuk mengidentifikasi pola dan kelompok wilayah dengan karakteristik serupa dalam mendukung strategi pengendalian TBC secara lebih terarah. Temuan Islam *et al.* (2025) menegaskan pentingnya identifikasi wilayah prioritas berbasis data dalam pengendalian TBC di tingkat daerah.

Analisis pengelompokan wilayah dapat dilakukan menggunakan metode klusterisasi sebagai salah satu pendekatan dalam analisis data. Metode klusterisasi merupakan pendekatan statistik dan komputasional untuk mengelompokkan objek berdasarkan kemiripan karakteristik tanpa memerlukan label awal. Berbagai metode klusterisasi telah dikembangkan, seperti *hierarchical clustering*, *K-Means*, dan *K-Medoids*. Metode *K-Medoids* memiliki keunggulan karena metode ini lebih tahan terhadap nilai pencilan (*outlier*) dan mampu menghasilkan kluster yang lebih mewakili data, karena pusat kluster berupa objek data aktual (*medoid*), bukan nilai

rata-rata. Metode *K-Medoids* juga meminimalkan ketidakmiripan antarobjek sehingga lebih tahan terhadap gangguan data (*noise*) dan *outlier* dibandingkan metode *K-Means* (Park dan Jun, 2009). Karakteristik tersebut menjadikan metode *K-Medoids* sesuai untuk menganalisis data sosial dan kesehatan yang umumnya memiliki distribusi tidak normal serta variasi antarwilayah yang tinggi (Leis *et al.*, 2023).

Metode lain yang dapat digunakan dalam analisis pengelompokan data adalah *Self Organizing Map* (SOM). Metode *Self Organizing Map* (SOM) merupakan salah satu jenis jaringan saraf tak terawasi (*unsupervised neural network*) yang diperkenalkan oleh Kohonen (1990). Metode ini mampu memproyeksikan data berdimensi tinggi ke dalam peta dua dimensi. Metode SOM melakukan proses pembelajaran dengan menjaga kedekatan topologis antar data sehingga metode ini menempatkan data yang memiliki karakteristik serupa berdasarkan nilai indikator pada posisi yang berdekatan di dalam peta. Proses tersebut memudahkan peneliti dalam memvisualisasikan struktur data yang kompleks serta mengidentifikasi pola atau kluster awal. Metode SOM telah terbukti efektif dalam menganalisis hubungan non-linear antarindikator kesehatan dan lingkungan serta dapat diterapkan dalam studi kesehatan lingkungan yang melibatkan banyak variabel (Basara dan Yuan, 2008).

Penggunaan metode *Self Organizing Map* (SOM) dalam analisis klusterisasi sering dikombinasikan dengan metode partisi untuk meningkatkan hasil pengelompokan. Pendekatan dua tahap (*two-stage clustering*) menggunakan metode SOM untuk menghasilkan prototipe data, kemudian melanjutkan klusterisasi dengan metode partisi seperti *K-Means*, dan pendekatan ini telah

terbukti efektif (Yu *et al.*, 2020). Dalam pendekatan tersebut, metode SOM memetakan data berdimensi tinggi ke dalam ruang dua dimensi serta mengurangi kompleksitas komputasi, kemudian vektor bobot neuron (*codebook vectors*) digunakan sebagai input bagi metode partisi untuk menghasilkan kluster akhir. Pendekatan ini menghasilkan kluster yang lebih stabil karena pusat kluster berasal dari struktur peta SOM yang telah terorganisasi secara topologis. Metode *K-Means* pada tahap kedua memiliki kelemahan karena sensitif terhadap nilai pencilan, sehingga asumsi tersebut tidak selalu sesuai dengan karakter data kesehatan (Kaufman dan Rousseeuw, 1990).

Penelitian ini mengusulkan integrasi metode *Self Organizing Map* (SOM) dan *K-Medoids* untuk mengatasi kelemahan metode sebelumnya. Pendekatan SOM-*K-Medoids* menggabungkan kemampuan SOM dalam memetakan struktur topologi data dengan ketahanan *K-Medoids* terhadap nilai pencilan (Kaufman dan Rousseeuw, 1990). Metode SOM mengekstraksi representasi awal berupa prototipe kluster dari data berdimensi tinggi, sedangkan metode *K-Medoids* menyempurnakan hasil tersebut dengan membentuk kluster akhir yang stabil dan representatif berdasarkan medoid aktual. Integrasi kedua metode tersebut menghasilkan pengelompokan yang terstruktur secara topologis, lebih tahan terhadap *outlier*, serta lebih mudah diinterpretasikan secara analitis.

Pendekatan integratif dua tahap (*two-stage clustering*) telah banyak disarankan dalam literatur modern karena pendekatan ini mampu mengatasi keterbatasan metode tunggal (Kotyrba *et al.*, 2021). Penelitian Imani *et al.* (2023) menerapkan metode *Self Organizing Map* (SOM) dalam klusterisasi kabupaten/kota di Provinsi Nusa Tenggara Timur berdasarkan indikator sosial dan penelitian

tersebut menunjukkan kemampuan SOM dalam memvisualisasikan serta mengidentifikasi pola data yang kompleks. Penelitian Thamrin dan Sahrir (2024) menggunakan metode *K-Medoids* dalam klasterisasi penyebaran Demam Berdarah di Kota Palopo dan penelitian tersebut membuktikan efektivitas metode *K-Medoids* dalam menangani data kesehatan masyarakat yang memiliki distribusi tidak normal serta *outlier*. Berdasarkan temuan tersebut, penelitian ini mengharapkan bahwa integrasi metode SOM-*K-Medoids* dapat menghasilkan klaster yang lebih representatif, tahan terhadap *outlier*, dan mendukung perencanaan kebijakan kesehatan daerah, khususnya dalam identifikasi wilayah prioritas pengendalian TBC di Jawa Barat.

Implementasi penelitian ini difokuskan pada pengelompokan kabupaten/kota di Provinsi Jawa Barat berdasarkan indikator TBC dan determinan sosial kesehatan. Penelitian ini memilih Provinsi Jawa Barat sebagai objek penelitian karena wilayah ini memiliki kompleksitas yang tinggi dan variasi indikator yang besar antarwilayah, terutama dalam hal beban TBC yang termasuk tertinggi di Indonesia (Kementerian Kesehatan RI, 2025). Provinsi Jawa Barat memiliki 27 kabupaten/kota dengan perbedaan mencolok dalam kepadatan penduduk serta akses layanan kesehatan, sehingga kondisi tersebut menjadikan wilayah ini sebagai konteks yang sesuai untuk penerapan model klasterisasi berbasis data. Hasil pengelompokan diharapkan dapat mengidentifikasi kelompok wilayah dengan karakteristik serupa, seperti wilayah dengan prevalensi TBC tinggi dan kondisi sosial kurang mendukung serta wilayah dengan prevalensi rendah dan akses layanan kesehatan yang lebih baik.

Dengan mempertimbangkan kondisi tersebut, penelitian ini bertujuan untuk menerapkan **Integrasi *Self Organizing Map* dan *K-Medoids* dalam Pengelompokan Kabupaten/Kota di Jawa Barat Berdasarkan Indikator TBC dan Determinan Sosial Kesehatan Tahun 2024**. Penelitian ini diharapkan dapat membantu pemerintah daerah dan instansi kesehatan dalam merencanakan program intervensi berbasis data, terutama dalam mengidentifikasi wilayah prioritas pengendalian TBC. Selain itu, penelitian ini memberikan kontribusi akademis berupa penerapan metode integratif yang masih terbatas penggunaannya dalam analisis data kesehatan di Indonesia, sehingga dapat menjadi referensi bagi studi serupa di masa mendatang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang, dapat dirumuskan permasalahan penelitian sebagai berikut:

1. Bagaimana karakteristik dan sebaran indikator TBC dan determinan sosial kesehatan di kabupaten/kota Provinsi Jawa Barat?
2. Bagaimana penentuan konfigurasi optimal *Self Organizing Map* dan *K-Medoids* melalui proses *hyperparameter tuning* berdasarkan indeks validasi klaster?
3. Bagaimana hasil pengelompokan dan profil setiap klaster kabupaten/kota yang terbentuk dari integrasi *Self Organizing Map* dan *K-Medoids*?

### 1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian hanya berfokus pada 27 kabupaten/kota di Provinsi Jawa Barat dengan data tahun 2024 yang bersumber dari Open Data Jawa Barat dan Badan Pusat Statistik (BPS).
2. Variabel yang digunakan terbatas pada indikator TBC dan determinan sosial kesehatan yang lolos uji non-multikolinearitas.
3. Metode yang digunakan dibatasi pada integrasi *Self Organizing Map* (SOM) dan *K-Medoids* dengan pendekatan *two-stage clustering*, tanpa membandingkan dengan metode klasterisasi lainnya.
4. Pemilihan konfigurasi optimal dilakukan menggunakan *Silhouette Coefficient*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CH) melalui proses *hyperparameter tuning*, tanpa membahas hubungan kausal antarindikator atau detail implementasi kebijakan pengendalian TBC.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengidentifikasi karakteristik dan sebaran indikator TBC dan determinan sosial kesehatan di kabupaten/kota Provinsi Jawa Barat.
2. Menentukan konfigurasi optimal *Self Organizing Map* dan *K-Medoids* melalui proses *hyperparameter tuning* berdasarkan indeks validasi kluster (*Silhouette Coefficient*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*).
3. Mendeskripsikan hasil pengelompokan dan profil setiap kluster kabupaten/kota yang terbentuk dari integrasi *Self Organizing Map* dan *K-Medoids* berdasarkan indikator TBC dan determinan sosial kesehatan.