Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 6861-6868
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Perbandingan Algoritma K-Means dan Algoritma K-Medoids untuk Klasterisasi Pasien TBC dalam Analisis Berdasarkan Tingkat Risiko Penyebaran

Shofie Muthia Putri Hamdani, Royan Habibie Sukarna, Febriyanti Darnis*
Informatika, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa
333210041@untirta.ac.id, royan@untirta.ac.id, *febriyanti.darnis@untirta.ac.id

Abstrak

Tuberkulosis (TBC) masih menjadi permasalahan kesehatan masyarakat di Provinsi Banten dengan pola penyebaran yang berbeda antar wilayah. Variasi kondisi geografis dan demografis menyebabkan perlunya pendekatan analisis spasial untuk memahami distribusi kasus TBC secara lebih mendalam. Penelitian ini bertujuan untuk menganalisis pola penyebaran TBC menggunakan metode klasterisasi serta mengelompokkan wilayah berdasarkan tingkat penyebaran TBC ke dalam kategori tinggi dan rendah. Data yang digunakan merupakan data sekunder pasien TBC yang telah melalui proses preprocessing dan rekayasa fitur, meliputi geocoding fasilitas pelayanan kesehatan untuk memperoleh koordinat latitude dan longitude, serta penambahan variabel usia pasien dan kepadatan penduduk. Evaluasi kualitas klaster dilakukan menggunakan nilai silhouette score. Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan nilai silhouette sebesar 0,475, sedangkan algoritma K-Medoids memperoleh nilai sebesar 0,472. Nilai tersebut menunjukkan bahwa kedua algoritma mampu membentuk klaster dengan kualitas yang cukup baik. Selain itu, visualisasi hasil klaster pada peta digital membantu mengidentifikasi wilayah dengan tingkat penyebaran tinggi sehingga mempermudah proses pemantauan dan pengambilan keputusan oleh instansi kesehatan daerah. Hasil klasterisasi ini memberikan informasi wilayah prioritas yang dapat dimanfaatkan sebagai dasar perencanaan dan penentuan strategi intervensi penanggulangan TBC di Provinsi Banten. Penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian selanjutnya terkait pemanfaatan data spasial dan teknik data mining dalam mendukung kebijakan kesehatan masyarakat yang lebih efektif, terarah, dan berbasis data untuk menekan angka penyebaran TBC secara berkelanjutan di masa mendatang. Pendekatan ini turut membantu peningkatan efisiensi alokasi sumber daya kesehatan serta pelaksanaan program pencegahan penyakit pada daerah rentan.

Kata kunci: Tuberkulosis, K-Means, K-Medoids.

1. Latar Belakang

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis* dan ditularkan melalui udara dari individu yang terinfeksi (Sari et al., 2022). Penyakit ini umumnya menyerang paru-paru, namun juga dapat menyerang organ lain di luar paru-paru (ekstraparu). Infeksi yang ditimbulkan oleh bakteri tersebut memiliki dampak yang besar dan berpotensi mengancam kesehatan masyarakat secara luas, terutama di wilayah dengan akses layanan kesehatan yang terbatas (Kemenkes RI, 2023; Kusuma & Firmansyah, 2022).

Berdasarkan data terbaru dari *World Health Organization* (WHO), TBC masih tercatat sebagai salah satu penyebab kematian tertinggi akibat penyakit menular di dunia. Pada tahun 2022, lebih dari 10 juta kasus baru TBC dilaporkan secara global, yang menunjukkan tingginya urgensi bagi pemerintah dan masyarakat untuk meningkatkan upaya penanggulangan secara lebih efektif (*World Health Organization* (WHO), 2022, 2023). Data tersebut mengindikasikan bahwa meskipun telah terdapat kemajuan dalam pengobatan dan pencegahan, TBC tetap menjadi ancaman serius, khususnya di negara-negara berkembang dengan keterbatasan infrastruktur kesehatan (Latifah et al., 2022; Purwanto et al., 2022) .

Tingginya angka kasus tuberkulosis (TBC) di Indonesia memerlukan penanganan serius melalui kerja sama lintas sektor antara pemerintah dan masyarakat. Pemerintah berperan dalam memperkuat program kesehatan, sementara masyarakat didorong untuk meningkatkan kesadaran dan partisipasi dalam pencegahan serta pengobatan TBC

(Pralambang & Setiawan, 2021; Ula et al., 2023). SR Konsorsium Penabulu–STPI Banten sebagai pelaksana hibah Global Fund Indonesia berperan dalam mendukung program eliminasi TBC melalui penguatan kinerja komunitas, advokasi kebijakan, peningkatan kapasitas pemangku kepentingan, dan kemitraan lintas sektor di tujuh kabupaten/kota di Provinsi Banten. Program ini berfokus pada pelacakan kasus melalui Investigasi Kontak dan penemuan kasus secara aktif. Mengingat Indonesia menempati peringkat kedua dengan kasus TBC tertinggi di dunia, keterlibatan Organisasi Masyarakat Sipil (OMS) menjadi sangat penting dalam mendukung upaya promotif, preventif, dan percepatan eliminasi TBC pada tahun 2030 (Bakrie *Center Foundation*, n.d.; Sub Recipient (SR) Konsorsium Penabulu - STPI Provinsi Banten, 22 C.E.).

Pola penyebaran TBC tidak bersifat *homogen* antar wilayah. Faktor geografis, demografis, kepadatan penduduk, serta akses terhadap fasilitas pelayanan kesehatan berperan penting dalam menentukan tingkat penularan penyakit ini (Putri et al., 2024; Sari et al., 2022). Oleh karena itu, analisis spasial menjadi pendekatan yang relevan untuk mengidentifikasi pola distribusi kasus TBC secara geografis dan mengelompokkan wilayah berdasarkan tingkat risiko penyebaran (Handayani, 2022; Munandar, 2022).

Pendekatan klasterisasi merupakan salah satu metode analisis data yang banyak digunakan untuk mengelompokkan objek berdasarkan kemiripan karakteristik tertentu (Rahmah et al., 2022). Dalam konteks kesehatan masyarakat, klasterisasi dapat dimanfaatkan untuk mengelompokkan wilayah berdasarkan tingkat kejadian penyakit sehingga dapat membantu pengambil kebijakan dalam menentukan prioritas intervensi (Abdullah et al., 2022; Riswana, 2023). Penelitian ini menggunakan algoritma K-Means dan K-Medoids untuk menganalisis pola penyebaran TBC di Provinsi Banten, dengan mempertimbangkan variabel spasial dan demografis yang relevan.

2. Metode Penelitian

Penelitian ini menggunakan data sekunder pasien TBC yang diperoleh dari PR Konsorsium Penabulu STPI dengan jumlah data sebanyak 23.929. Data tersebut mencakup informasi lokasi fasilitas pelayanan kesehatan, usia pasien, serta data kepadatan penduduk. Seluruh data yang digunakan telah melalui proses pembersihan dan validasi untuk memastikan kualitas data yang baik sebelum dilakukan analisis. Tahapan praproses data meliputi pembersihan data dari nilai hilang dan duplikasi, data yang telah dibersihkan berkurang menjadi 23.390.

Rekayasa fitur dilakukan untuk memperkaya informasi yang digunakan dalam analisis klasterisasi. Proses geocoding diterapkan pada data fasilitas pelayanan kesehatan untuk mengonversi alamat menjadi koordinat geografis berupa latitude dan longitude. Koordinat tersebut digunakan sebagai representasi spasial wilayah. Selain variabel spasial, ditambahkan variabel usia pasien dan kepadatan penduduk sebagai indikator demografis yang berpotensi memengaruhi tingkat penyebaran TBC.

Metode klasterisasi yang digunakan dalam penelitian ini adalah algoritma K-Means dan K-Medoids. K-Means mengelompokkan data berdasarkan jarak rata-rata terhadap pusat klaster (centroid), sedangkan K-Medoids menggunakan objek *representatif* (medoid) sebagai pusat klaster sehingga lebih robust terhadap outlier. Jumlah klaster ditentukan berdasarkan pertimbangan kebutuhan analisis dan evaluasi kualitas klaster. Evaluasi kualitas klaster dilakukan menggunakan silhouette score, yang mengukur tingkat kohesi dan separasi antar klaster. Nilai silhouette berada pada rentang -1 hingga 1, di mana nilai yang lebih mendekati 1 menunjukkan kualitas klaster yang semakin baik.

3. Hasil dan Diskusi

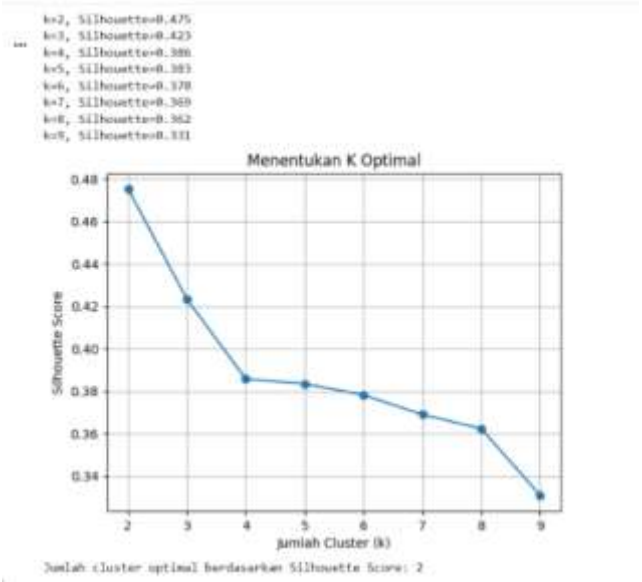
Hasil

Sebelum menentukan jumlah klaster optimal (nilai K), data terlebih dahulu melalui tahap *preprocessing*. Tahapan ini diawali dengan memuat dataset pasien tuberkulosis (TBC) Provinsi Banten yang tersimpan pada Google Drive. Selanjutnya dilakukan tahap *geocoding* fitur fasyankes menggunakan titik koordinat fasyankes pasien. Kemudian, menambahkan atribut kepadatan penduduk dan menormalisasikan data menggunakan StandarScaler. Tahap akhir preprocessing adalah penentuan jumlah klaster optimal dengan menggunakan *Silhouette Score*, sehingga klaster yang dihasilkan memiliki tingkat pemisahan dan struktur yang baik sebelum dilakukan penerapan algoritma K-Means dan K-Medoids.

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Penentuan jumlah klaster (nilai *k*) dilakukan menggunakan *Silhouette Score*, yang berfungsi untuk mengevaluasi tingkat kohesi data dalam satu klaster serta kejelasan pemisahan antar klaster. Pada tahap ini, pengujian nilai *k* dilakukan pada rentang 2 hingga 10. Meskipun algoritma K-Means memungkinkan perhitungan *Silhouette Score* secara langsung pada dataset berukuran besar, penelitian ini menerapkan teknik pengambilan sampel acak sebanyak 2.000 data untuk menjaga kesetaraan proses evaluasi dengan algoritma K-Medoids. Pendekatan serupa juga diterapkan pada K-Medoids guna mengurangi kompleksitas perhitungan matriks jarak. Selanjutnya, *Silhouette Score* dihitung untuk setiap nilai *k*, dan nilai dengan skor tertinggi ditetapkan sebagai jumlah klaster optimal.

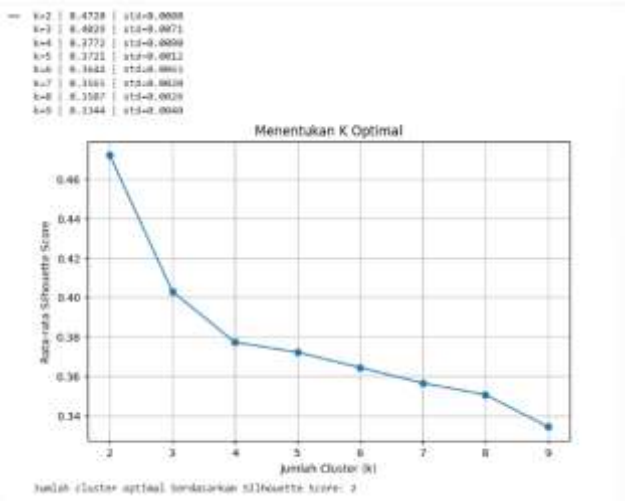
Berdasarkan hasil perhitungan *Silhouette Score* untuk K-Means, nilai tertinggi diperoleh pada jumlah klaster *k* = 2 dengan skor sebesar 0,475. Nilai tersebut lebih tinggi dibandingkan dengan konfigurasi jumlah klaster lainnya, seperti *k* = 3 yang menghasilkan skor 0,423, *k* = 4 sebesar 0,386, dan *k* = 6 sebesar 0,378. Temuan ini juga diperkuat oleh hasil visualisasi grafik hubungan antara jumlah klaster dan nilai *Silhouette Score*, yang menunjukkan puncak nilai pada *k* = 2. Oleh karena itu, dapat disimpulkan bahwa jumlah klaster optimal untuk algoritma K-Means pada dataset yang digunakan adalah dua klaster. Pemilihan *k* = 2 menunjukkan bahwa data dapat dikelompokkan secara paling optimal ke dalam dua kelompok utama dengan tingkat kohesi internal dan separasi antar klaster yang lebih baik dibandingkan konfigurasi jumlah klaster lainnya. Nilai *Silhouette Score* yang berada pada kisaran 0,3 hingga 0,4 mengindikasikan bahwa struktur klaster yang terbentuk tergolong cukup baik, meskipun masih terdapat tumpang tindih antar klaster yang kemungkinan dipengaruhi oleh karakteristik data spasial serta heterogenitas distribusi kasus tuberkulosis.



Gambar 1. Silhouette Score K-Means

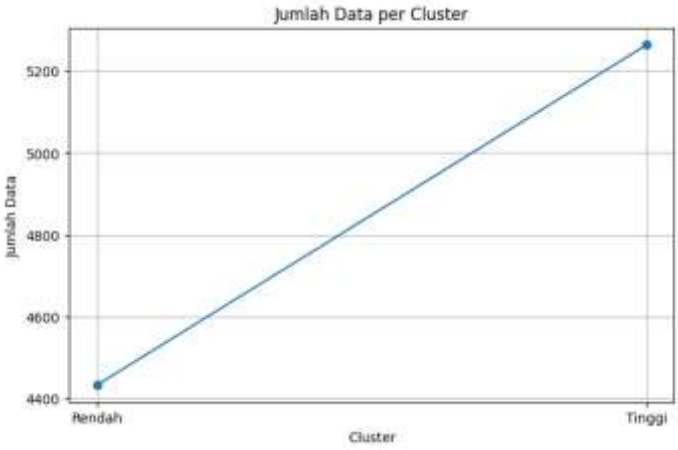
Evaluasi penentuan jumlah klaster optimal pada algoritma K-Medoids disajikan melalui grafik *Silhouette Score* untuk nilai *k* pada rentang 2 hingga 9. Guna meningkatkan kestabilan hasil evaluasi, perhitungan *Silhouette Score* dilakukan secara berulang sebanyak sejumlah pengujian (*n_run*), kemudian dihitung nilai rata-rata (*mean*) dan standar deviasi untuk setiap nilai *k*. Pendekatan ini diterapkan untuk meminimalkan pengaruh inisialisasi medoid yang bersifat acak. Hasil evaluasi menunjukkan bahwa nilai rata-rata *Silhouette Score* tertinggi diperoleh pada *k* = 2 dengan nilai sebesar 0,472 dan standar deviasi 0,0000, yang mengindikasikan bahwa konfigurasi klaster tersebut tidak hanya memiliki kualitas terbaik, tetapi juga tingkat stabilitas yang tinggi antar pengulangan. Sementara itu, nilai *Silhouette Score* pada jumlah klaster lainnya cenderung lebih rendah, seperti pada *k* = 3 dengan nilai 0,4029 (std = 0,0071), *k* = 4 sebesar 0,3772 (std = 0,0000), dan *k* = 5 sebesar 0,3721 (std = 0,0012).

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 2. Silhouette Score K-Medoids

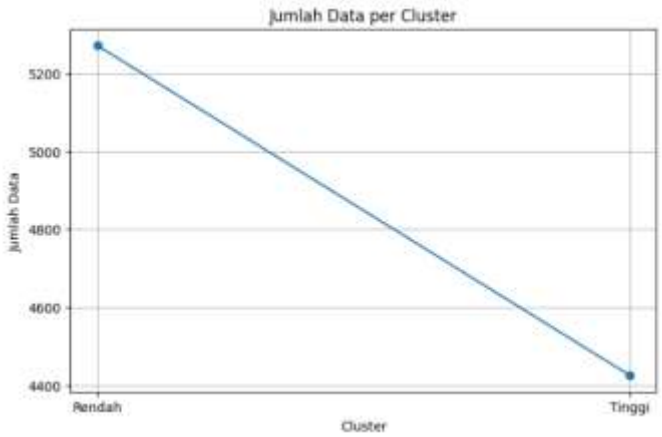
Implementasi algoritma K-Means dalam penelitian ini dilakukan dengan menggunakan jumlah klaster optimal yang telah ditentukan pada tahap sebelumnya. Model K-Means dibangun menggunakan perintah KMeans(n_clusters=optimal_k, random_state=42), di mana parameter *random_state* digunakan untuk menjaga konsistensi hasil pembentukan centroid pada setiap proses eksekusi, dengan nilai *optimal_k* ditetapkan sebesar 2. Selanjutnya, proses klasterisasi dilakukan melalui fungsi fit_predict(data_scaled), yang memungkinkan algoritma melakukan iterasi untuk menentukan posisi centroid yang paling representatif serta mengelompokkan setiap data ke dalam klaster terdekat berdasarkan jarak Euclidean.



Gambar 3, Grafik data Tiap Klaster K-Means

Gambar 3 menampilkan grafik distribusi jumlah data pada masing-masing klaster hasil klasterisasi K-Means. Setiap titik pada grafik merepresentasikan jumlah pasien TBC yang tergolong ke dalam kategori klaster Rendah dan Tinggi. Hasil visualisasi menunjukkan bahwa klaster Rendah memiliki sebanyak 4.433 data pasien, sedangkan klaster Tinggi mencakup 5.264 data pasien. Distribusi ini memperlihatkan bahwa sebagian besar pasien tergolong dalam klaster Tinggi, yang mengindikasikan adanya ketidakseimbangan jumlah anggota antar klaster. Kondisi tersebut mencerminkan dominasi karakteristik pasien dengan tingkat risiko atau intensitas kasus yang lebih tinggi dalam dataset. Secara analitis, perbedaan jumlah anggota klaster ini memberikan gambaran awal mengenai struktur hasil klasterisasi, di mana klaster Rendah merepresentasikan kelompok pasien dengan karakteristik kasus yang relatif lebih rendah dan cenderung homogen, sedangkan klaster Tinggi menggambarkan kelompok pasien dengan tingkat kompleksitas dan intensitas kasus yang lebih tinggi. Dengan demikian, visualisasi distribusi ini tidak hanya menunjukkan proporsi jumlah pasien dalam setiap klaster, tetapi juga menjadi dasar untuk analisis lanjutan terkait pola penyebaran kasus tuberkulosis.

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



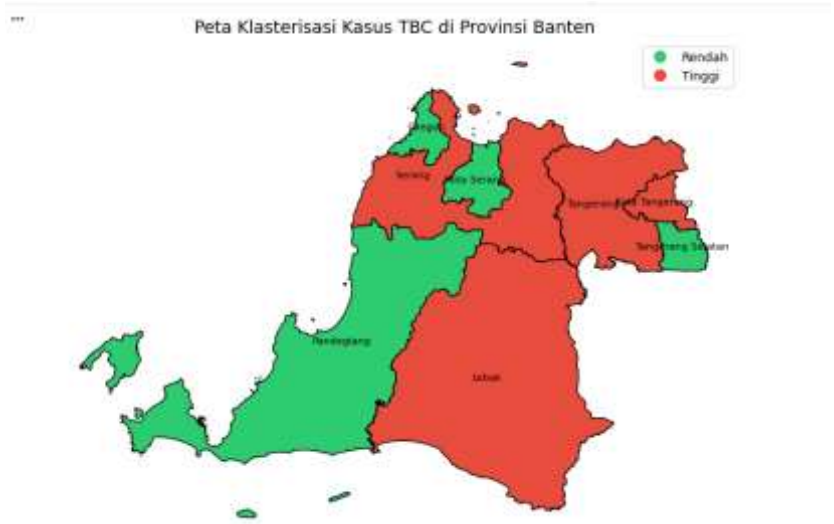
Gambar 4. Grafik data Tiap Klaster K-Medoids

Setelah jumlah klaster optimal ditentukan, proses klasterisasi akhir dilakukan menggunakan seluruh data yang telah distandarisasi, bukan hanya data sampel. Algoritma K-Medoids dijalankan hingga mencapai konvergensi, dan hasil klasterisasi akhir menunjukkan bahwa data pasien TBC terbagi ke dalam dua klaster utama, yaitu Klaster 0 dengan 5.272 data dan Klaster 1 dengan 4.425 data. Distribusi ini menunjukkan pembagian data yang relatif seimbang. Berdasarkan karakteristik distribusinya, klaster hasil K-Medoids kemudian diinterpretasikan ke dalam kategori risiko, di mana Klaster 0 dikategorikan sebagai Risiko Rendah dan Klaster 1 sebagai Risiko Tinggi. Interpretasi ini dilakukan untuk mempermudah analisis spasial dan mendukung proses visualisasi pemetaan wilayah. Visualisasi jumlah data pada masing-masing klaster menunjukkan bahwa klaster dengan kategori Risiko Tinggi memiliki jumlah data yang lebih besar, yang mengindikasikan bahwa sebagian besar pasien TBC dalam dataset berada pada wilayah atau fasilitas pelayanan kesehatan dengan tingkat risiko yang lebih tinggi.

Pada tahap klasifikasi tingkat kasus berdasarkan kota/kabupaten, data terlebih dahulu dipersiapkan agar sesuai untuk pemodelan berbasis numerik. Dataset pasien TBC terdiri atas variabel numerik dan kategorikal, sehingga variabel kategorikal yang relevan, yaitu *kota_kab* dan *tipe_pasien*, ditransformasikan ke dalam bentuk numerik menggunakan teknik *One-Hot Encoding* dengan fungsi pd.get_dummies(). Sementara itu, variabel numerik diidentifikasi berdasarkan tipe data *int64* dan *float64*. Seluruh hasil transformasi kemudian digabungkan menjadi dataset baru yang siap dianalisis. Selanjutnya, seluruh fitur yang telah diencoding dan variabel numerik asli distandarisasi menggunakan metode *StandardScaler* untuk menyamakan skala data dan mencegah dominasi fitur tertentu dalam proses klasterisasi. Setelah proses standarisasi, algoritma K-Means dan K-Medoids diterapkan dengan jumlah klaster $k = 2$, sesuai hasil evaluasi sebelumnya menggunakan *silhouette score*. Pada K-Means, parameter *random_state* dan *n_init* digunakan untuk menjaga konsistensi hasil, sedangkan pada K-Medoids proses dilakukan secara iteratif dengan inisialisasi medoid acak dan perhitungan jarak Euclidean hingga mencapai konvergensi. Hasil klasterisasi kemudian ditambahkan ke dalam dataset sebagai label klaster dan dianalisis untuk melihat distribusi jumlah data pada setiap klaster. Analisis ini memberikan gambaran awal mengenai karakteristik dan dominasi tingkat kasus TBC pada masing-masing kelompok wilayah.

Selanjutnya dilakukan visualisasi hasil klasterisasi kasus tuberkulosis (TBC) di Provinsi Banten dalam bentuk peta tematik (*choropleth map*) untuk menggambarkan distribusi spasial tingkat risiko TBC pada tingkat kota/kabupaten. Data kasus TBC mentah terlebih dahulu diagregasikan berdasarkan wilayah administratif untuk memperoleh jumlah kasus per kota/kabupaten, kemudian diklasifikasikan ke dalam empat kategori tingkat risiko, yaitu Rendah, Sedang, Tinggi, dan Sangat Tinggi berdasarkan rentang jumlah kasus. Selanjutnya, data batas administrasi Provinsi Banten dalam format *shapefile* diintegrasikan dengan data kasus menggunakan pustaka *geopandas*. Penyesuaian nama wilayah dilakukan untuk memastikan kesesuaian antara data spasial dan data atribut. Hasil penggabungan ini menghasilkan data geospasial yang memuat informasi wilayah beserta kategori tingkat risiko TBC. Visualisasi peta dilakukan dengan pewarnaan berbeda untuk setiap kategori risiko, yaitu hijau (Rendah), dan merah (Tinggi). Label nama wilayah ditambahkan pada titik representatif untuk mempermudah interpretasi. Hasil visualisasi menunjukkan adanya variasi distribusi kasus TBC antar wilayah, dengan konsentrasi kasus tinggi hingga sangat tinggi pada beberapa daerah tertentu. Peta tematik ini memberikan gambaran pola spasial penyebaran TBC yang dapat dimanfaatkan sebagai dasar perencanaan kebijakan dan intervensi kesehatan masyarakat yang lebih terarah.

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 5. Peta Klasterisasi Kasus TBC di Banten

Selain visualisasi statis menggunakan Matplotlib dan GeoPandas, penelitian ini juga menghasilkan peta interaktif berbasis Folium untuk menyajikan persebaran kasus tuberkulosis (TBC) di Provinsi Banten secara lebih dinamis. Peta dasar dibangun menggunakan folium.Map dengan titik pusat Provinsi Banten dan tingkat *zoom* yang disesuaikan agar seluruh wilayah dapat ditampilkan dengan jelas. Data spasial hasil penggabungan *shapefile* dan data agregat kasus TBC ditambahkan ke peta sebagai layer GeoJSON, dengan pewarnaan wilayah berdasarkan kategori tingkat kasus, yaitu Rendah, Sedang, Tinggi, dan Sangat Tinggi.

Setiap wilayah dilengkapi dengan *tooltip* interaktif yang menampilkan informasi nama daerah, jumlah kasus, dan kategori risiko, serta legenda warna untuk memudahkan interpretasi. Selain itu, lokasi fasilitas pelayanan kesehatan ditampilkan sebagai titik pada peta, namun hanya Fasyankes yang memiliki koordinat geografis valid yang dapat divisualisasikan. Peta interaktif ini memungkinkan eksplorasi data spasial secara lebih intuitif dan mendukung analisis distribusi kasus TBC secara lebih komprehensif.



Gambar 6. Peta interaktif menggunakan folium

PEMBAHASAN

Hasil klasterisasi menggunakan algoritma K-Means dan K-Medoids menghasilkan dua kelompok, yaitu Klaster Rendah dan Klaster Tinggi. Klaster Tinggi didominasi oleh wilayah dengan jumlah kasus TBC besar, seperti Kabupaten Tangerang, Kota Tangerang, Kabupaten Serang, dan Kabupaten Lebak, yang umumnya memiliki kepadatan penduduk tinggi, urbanisasi dan mobilitas yang intens, serta aktivitas ekonomi yang padat. Faktor-faktor tersebut meningkatkan potensi penularan TBC, sementara akses layanan kesehatan yang lebih baik turut

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

berkontribusi pada tingginya kasus terlapor. Sebaliknya, Klaster Rendah mencakup wilayah dengan jumlah kasus relatif lebih sedikit, seperti Kabupaten Pandeglang, yang didominasi wilayah pedesaan dengan kepadatan penduduk dan mobilitas rendah. Namun, keterbatasan akses layanan kesehatan di beberapa daerah berpotensi menyebabkan rendahnya angka kasus terlapor. Secara spasial, wilayah utara Provinsi Banten cenderung tergabung dalam Klaster Tinggi, sedangkan wilayah selatan didominasi Klaster Rendah. Hasil ini menegaskan bahwa kepadatan penduduk, urbanisasi, dan mobilitas berperan penting dalam distribusi kasus TBC dan dapat dimanfaatkan sebagai dasar penentuan prioritas intervensi.

Tabel 1. Perbandingan K-Means dan K-Medoids

	K-Means	**K-Medoids**
Optimal K	2	2
Silhouette Score	0,475	0,472
Waktu Komputasi	1 detik (data sampling)	3 menit (data sampling)
Banyak Data	2000 (sampling)	2000 (sampling)
Stabilitas Klaster	Tinggi	Sedang

Berdasarkan pengujian dengan jumlah klaster optimal $k = 2$, algoritma K-Means dan K-Medoids menunjukkan perbedaan karakteristik dalam pengelompokan wilayah kasus TBC di Provinsi Banten. K-Means menghasilkan nilai *Silhouette Score* sebesar 0,475, sedikit lebih tinggi dibandingkan K-Medoids, sehingga menunjukkan pemisahan klaster yang lebih baik. Dari sisi komputasi, K-Means jauh lebih efisien dengan waktu pemrosesan sekitar 1 detik, sedangkan K-Medoids membutuhkan waktu hingga 3 menit meskipun menggunakan data sampling yang sama. Selain itu, hasil klaster K-Means juga lebih stabil pada beberapa iterasi percobaan. Meskipun demikian, kedua metode secara konsisten mampu mengidentifikasi pola penyebaran TBC dan membedakan wilayah berdasarkan tingkat risiko. Temuan ini menegaskan bahwa klasterisasi dapat digunakan sebagai pendekatan kuantitatif untuk mendukung penentuan wilayah prioritas dan perencanaan strategi pengendalian TBC yang lebih terarah di Provinsi Banten.

4. Kesimpulan

Hasil klasterisasi menghasilkan dua kategori risiko penyebaran TBC di Provinsi Banten, yaitu Rendah dan Tinggi. Kabupaten Tangerang, Kota Tangerang, Kabupaten Lebak, dan Kabupaten Serang termasuk dalam kategori risiko Tinggi, sedangkan Kota Serang, Kota Cilegon, Kota Tangerang Selatan, dan Kabupaten Pandeglang tergolong risiko Rendah. Algoritma K-Means dan K-Medoids sama-sama mampu mengelompokkan wilayah berdasarkan penyebaran kasus TBC dengan konfigurasi klaster optimal $k = 2$, di mana K-Means memperoleh nilai *Silhouette Score* sebesar 0,475 dan K-Medoids sebesar 0,472. Meskipun K-Medoids lebih tahan terhadap *outlier*, K-Means menunjukkan keunggulan dari sisi efisiensi komputasi dengan waktu pemrosesan yang jauh lebih cepat, sehingga lebih sesuai untuk pemetaan risiko TBC berskala besar. Visualisasi spasial menggunakan peta statis dan peta interaktif Folium berhasil menyajikan distribusi klaster TBC secara jelas, meskipun pemetaan titik fasilitas kesehatan masih terbatas akibat keterbatasan layanan geocoding OpenStreetMap. Selain itu, meskipun Kabupaten Pandeglang termasuk dalam kategori risiko Rendah, rendahnya kasus terlapor berpotensi dipengaruhi oleh keterbatasan akses layanan kesehatan, skrining, dan pelaporan, sehingga tetap diperlukan kewaspadaan dan upaya pemantauan berkelanjutan.

Referensi

1. Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). *The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. Quality & Quantity*, 56(3), 1283–1291. https://doi.org/10.1007/s11135-021-01176-w
2. Bakrie Center Foundation. Retrieved December 29, 2024, from https://bcf.or.id/
3. Handayani, F. (2022). *Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar*. Jurnal Teknologi Dan Informasi, 12(1), 46–63. https://doi.org/10.34010/jati.v12i1.6733
4. Kemenkes RI. (2023). Profil Kesehatan Indonesia 2023. https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2023
5. Kusuma, P. A., & Firmansyah, A. U. (2022). *Deteksi Penyebaran Penyakit Tuberkulosis dengan Algoritma K-Means Clustering Menggunakan Rapid Miner. Jurnal Teknologi Informatika Dan Komputer*, 8(2), 41–54. https://doi.org/10.37012/jtik.v8i2.1173
6. Latifah, A., Kurniasih, D., Muslina, & Armizan, E. W. (2022). *Sosialisasi Cara Pencegahan Dan Penularan Penyakit Tb Serta Upaya Peningkatan Mikroelemen Tubuh Bagi Penderita Tb (The Socialization On Methods To Prevent And To Transmist Of Tb Disease And To Improve Body Microelements For Tb Patients)*. Jurnal Abdikemas, 4(2).
7. Munandar, Tb. A. (2022). *Penerapan algoritma clustering untuk pengelompokan tingkat kemiskinan provinsi banten*. JSiI (Jurnal Sistem Informasi), 9(2), 109–114. https://doi.org/10.30656/jsii.v9i2.5099

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

8. Pralambang, S. D., & Setiawan, S. (2021). *Faktor Risiko Kejadian Tuberkulosis di Indonesia*. Jurnal Biostatistik, Kependudukan, Dan Informatika Kesehatan, 2(1), 60. https://doi.org/10.51181/bikfokes.v2i1.4660
9. Purwanto, B., Nilogiri, A., & Wardoyo. Ari Eko. (2022). *Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Penyakit Tbc (Studi Kasus: Puskesmas Di Kabupaten Jember)*. Jurnal Smart Teknologi, 3(3), 273–285.
10. Putri, R., Riana, F., & Wulandari, B. (2024). *Implementasi K-Medoids Dalam Pengelompokan Fasilitas Pelayanan Kesehatan Pada Kasus Tuberculosis. Jurnal Informatika*, 11(1), 17–24. https://doi.org/10.31294/inf.v11i1.20044
11. Rahmah, E., Haerani, E., Nazir, A., & Ramadhani, S. (2022). *Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Srategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)*. Jurnal Nasional Komputasi Dan Teknologi Informasi, 5(3).
12. Riswana, D. (2023). *Analisis segmentasi konsumen menggunakan algoritma k-means clustering berdasarkan model rfm sebagai rekomendasi strategi pemasaran (studi kasus di pt. Xyz).*
13. Sari, G. K., Sarifuddin, & Setyawati, T. (2022). *Tuberkulosis paru post wodec pleural efusion: laporan kasus pulmonary tuberculosis post wodec pleural effusion: case report*. Jurnal Medical Profession (MedPro), 4(2), 174–182.
14. Sub Recipient (SR) Konsorsium Penabulu - STPI Provinsi Banten. (22 C.E.). PROGRAM ELIMINASI TBC SUB RECIPIENT (SR) KONSORSIUM PENABULU - STPI PROVINSI BANTEN. https://tbckomunitas.id/wp-content/uploads/2022/06/Profil-SR-Banten.pdf
15. Ula, M., Zulfikri, A., Ulva, A. F., & Rizal, R. A. (2023). *Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru*. The Indonesian Journal of Computer Science, 12(1). https://doi.org/10.33022/ijcs.v12i1.3162
16. World Health Organization (WHO). (2022). GLOBAL TUBERCULOSIS REPORT 2022.
17. World Health Organization (WHO). (2023). GLOBAL TUBERCULOSIS REPORT 2023.

DOI: https://doi.org/10.31004/riggs.v5i2.8870
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)