

Strategi Intervensi untuk Peserta BPJS Tuberkulosis Berdasarkan Segmentasi Usia dan Perilaku Akses Fasilitas Kesehatan Menggunakan K-Means Clustering

Tamara Yunika Sianipar¹, Jessica A. Silburian¹, Krisnia Calysta Siahaan¹, Ruth Septiana Simanullang¹

¹ Institut Teknologi Del

[Diserahkan: 13 Mei 2025, Direvisi: 12 Mei 2023, Diterima: x bulan 2025]

Corresponding Author: XX (email: xxx@institusi.com)

ABSTRAK

Tuberkulosis (TB) merupakan salah satu masalah kesehatan utama di Indonesia, yang memerlukan penanganan jangka panjang dan distribusi layanan yang merata. BPJS Kesehatan, sebagai penyedia layanan kesehatan, menghadapi tantangan dalam memberikan perawatan yang tepat sasaran kepada peserta yang terdiagnosis TB, mengingat keberagaman karakteristik peserta dan akses mereka terhadap fasilitas kesehatan. Penelitian ini bertujuan untuk mengelompokkan peserta BPJS Tuberkulosis berdasarkan karakteristik demografi dan perilaku akses fasilitas kesehatan menggunakan teknik K-Means Clustering, dengan tujuan memberikan rekomendasi untuk kebijakan intervensi yang lebih relevan dan efisien. Data yang digunakan dalam penelitian ini mencakup data kepesertaan dan pelayanan fasilitas kesehatan dari tahun 2015 hingga 2021, yang kemudian dianalisis melalui beberapa tahap, mulai dari pemahaman bisnis hingga evaluasi klaster. Hasil analisis menunjukkan bahwa jumlah klaster optimal adalah tiga, dengan masing-masing klaster memiliki karakteristik yang berbeda: Klaster 0 (usia muda dengan kunjungan rendah), Klaster 1 (dewasa dengan kunjungan lebih tinggi), dan Klaster 2 (lebih tua dengan kebutuhan perawatan intensif). Evaluasi menggunakan Silhouette Coefficient dan visualisasi dengan Principal Component Analysis (PCA) menunjukkan adanya perbedaan antar klaster, meskipun kualitas pemisahan klaster masih dapat ditingkatkan. Temuan ini menunjukkan pentingnya pendekatan berbasis data dalam merancang strategi intervensi TB yang lebih tepat sasaran dan efisien, dengan mempertimbangkan karakteristik demografi dan perilaku peserta. Penelitian ini juga membuka peluang untuk pengembangan lebih lanjut dalam penggunaan teknik data mining dan penggabungan variabel sosial-ekonomi untuk meningkatkan hasil segmentasi.

KATA KUNCI : Tuberkulosis, BPJS Kesehatan, K-Means Clustering, Segmentasi, Evaluasi, Silhouette Coefficient, Principal Component Analysis (PCA).

I. PENDAHULUAN

Indonesia masih menghadapi masalah kesehatan serius terkait Tuberkulosis (TB), yang memiliki angka kejadian cukup tinggi. Salah satu pihak yang berperan penting dalam upaya penanggulangan TB adalah BPJS Kesehatan. BPJS Kesehatan bertugas untuk memberikan layanan kesehatan secara merata kepada seluruh masyarakat Indonesia, termasuk bagi penderita TB. Namun, dengan semakin banyaknya jumlah peserta BPJS setiap tahunnya, pengelolaan layanan kesehatan, terutama untuk penyakit jangka panjang seperti TB, semakin kompleks. Oleh karena itu, dibutuhkan pendekatan berbasis data yang lebih efisien dan efektif untuk memahami karakteristik peserta yang menderita TB.

Dalam penelitian ini, kami mengusulkan penggunaan metode K-Means clustering untuk melakukan segmentasi peserta BPJS berdasarkan dua faktor utama: usia dan perilaku mereka dalam mengakses fasilitas kesehatan. Dengan

segmentasi ini, BPJS Kesehatan diharapkan bisa mengidentifikasi klaster-klaster peserta dengan kebutuhan yang serupa, sehingga intervensi kesehatan yang diberikan bisa lebih terfokus dan tepat sasaran.

Selain itu, evaluasi klasterisasi dilakukan menggunakan metode Silhouette Coefficient (SC) untuk memastikan bahwa klaster yang terbentuk memiliki kualitas yang baik. Evaluasi ini penting agar hasil segmentasi tidak hanya bisa membedakan klaster dengan jelas, tetapi juga dapat diterapkan dalam kebijakan kesehatan yang lebih tepat.

Dengan penerapan analisis ini, diharapkan BPJS Kesehatan dapat meningkatkan efisiensi dalam pengelolaan layanan kesehatan bagi peserta TB, serta dapat merancang strategi distribusi pelayanan yang lebih efektif dan relevan berdasarkan karakteristik klaster peserta. Intervensi yang lebih tepat dan berbasis data dapat mencakup penyuluhan kesehatan, edukasi pencegahan, dan pengelolaan penyakit yang lebih terarah.

II. LITERATURE REVIEW

Penanganan Tuberkulosis (TB) di Indonesia masih menjadi tantangan besar dalam sistem kesehatan, terutama dalam konteks BPJS Kesehatan yang harus mengelola semakin banyaknya jumlah peserta. Berbagai pendekatan berbasis data dan analitik digunakan untuk membantu dalam pengelolaan layanan kesehatan, termasuk dalam penanganan penyakit TB. Pada bagian ini, kami akan membahas beberapa studi dan teori yang relevan terkait dengan penggunaan metode klusterisasi, khususnya K-Means, dalam analisis data kesehatan.

1. Penggunaan K-Means Clustering dalam Analisis Data Kesehatan

Metode K-Means Clustering telah banyak diterapkan dalam berbagai bidang, termasuk analisis data kesehatan, untuk mengelompokkan data yang memiliki karakteristik serupa. Sebagai contoh, penelitian oleh Lilis Setyowati et al. (2018) mengaplikasikan K-Means untuk menganalisis data kasus Tuberkulosis di Indonesia, dengan tujuan untuk mengidentifikasi pola distribusi geografis dan karakteristik peserta. Hasil dari penelitian tersebut menunjukkan bahwa K-Means dapat digunakan untuk segmentasi peserta kesehatan berdasarkan faktor-faktor tertentu seperti usia dan status kesehatan.

Selain itu, K-Means Clustering juga digunakan dalam penelitian kesehatan lainnya untuk mengidentifikasi pola dalam data kesehatan, seperti pemanfaatan layanan kesehatan, pengelompokan penyakit, dan segmentasi pasien. Menurut Bustamam et al. (2012), K-Means sangat efektif dalam mengelompokkan pasien berdasarkan karakteristik medis yang relevan, yang memungkinkan identifikasi kelompok yang membutuhkan perhatian khusus atau intervensi lebih lanjut.

2. Evaluasi Klusterisasi Menggunakan Silhouette Coefficient

Evaluasi klusterisasi merupakan aspek penting dalam analisis data berbasis kluster. Salah satu metode yang banyak digunakan untuk mengevaluasi kualitas kluster adalah Silhouette Coefficient (SC). Silhouette Coefficient mengukur sejauh mana suatu objek berada di dalam kluster yang benar dan sejauh mana objek tersebut berbeda dengan kluster lain. Nilai SC berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan klusterisasi yang lebih baik.

Rousseeuw (1987) memperkenalkan metode Silhouette untuk mengevaluasi hasil klusterisasi. Dalam konteks kesehatan, SC digunakan untuk mengevaluasi apakah kluster yang terbentuk memiliki pemisahan yang jelas antar kluster, yang sangat penting agar hasil kluster dapat diterapkan dalam kebijakan kesehatan yang tepat. Penelitian lain yang relevan adalah oleh Sainz et al. (2016), yang menggunakan SC untuk menilai klusterisasi pasien berdasarkan data medis, dan menemukan bahwa klusterisasi yang efektif dapat membantu dalam perencanaan strategi intervensi.

3. Segmentasi Berdasarkan Usia dan Perilaku Kesehatan

Segmentasi berdasarkan faktor demografis seperti usia dan perilaku kesehatan menjadi salah satu pendekatan yang relevan dalam pengelolaan data kesehatan. Beberapa studi menunjukkan bahwa karakteristik peserta, seperti usia, frekuensi kunjungan ke fasilitas kesehatan, dan perilaku akses layanan, sangat mempengaruhi kebutuhan perawatan dan

intervensi yang diperlukan. Misalnya, penelitian oleh Rahardianto et al. (2022) yang menggunakan K-Means untuk mengelompokkan peserta berdasarkan usia dan frekuensi kunjungan ke fasilitas kesehatan, menemukan bahwa segmentasi ini dapat membantu dalam merancang intervensi yang lebih spesifik dan berbasis data.

Selain itu, segmentasi berdasarkan usia juga memberikan wawasan yang lebih mendalam mengenai bagaimana faktor usia berperan dalam pengelolaan penyakit menular seperti TB. Peserta yang lebih tua mungkin memerlukan perhatian medis yang lebih intensif, sedangkan peserta yang lebih muda mungkin membutuhkan lebih banyak edukasi pencegahan. Segmentasi ini penting untuk memastikan bahwa intervensi yang dilakukan oleh BPJS Kesehatan dapat lebih tepat sasaran dan relevan dengan kondisi peserta.

4. Aplikasi K-Means dalam Kebijakan Kesehatan

Implementasi K-Means dalam kebijakan kesehatan dapat membantu pemerintah dan lembaga kesehatan untuk merancang strategi intervensi yang lebih efisien dan sesuai dengan kondisi peserta. Sebagai contoh, studi oleh Alhadi Bustamam et al. (2017) mengaplikasikan K-Means untuk mengelompokkan wilayah berdasarkan tingkat prevalensi penyakit, yang memungkinkan pengalokasian sumber daya yang lebih efisien. Dalam konteks TB, penerapan K-Means untuk segmentasi peserta dapat membantu BPJS Kesehatan dalam merencanakan distribusi layanan kesehatan dan menentukan wilayah yang membutuhkan perhatian lebih.

III. METODOLOGI

Metodologi Penelitian

Metodologi dalam penelitian ini dibagi ke dalam beberapa tahap utama yang mengarah pada tujuan pengelompokan peserta BPJS Tuberkulosis berdasarkan karakteristik pelayanan dan demografi menggunakan K-Means clustering.

1. Pemahaman Bisnis (Business Understanding)

Pada tahap ini, tujuan utama adalah untuk meningkatkan efisiensi dan ketepatan sasaran dalam pengambilan keputusan terkait penanganan Tuberkulosis (TB) di Indonesia. Dengan menggunakan data peserta BPJS, peneliti bertujuan untuk mengidentifikasi karakteristik wilayah yang membutuhkan intervensi berdasarkan data demografi dan perilaku layanan kesehatan yang diperoleh.

2. Pemahaman Data (Data Understanding)

Tahap ini mencakup eksplorasi data dari berbagai sumber, termasuk data kepesertaan, pelayanan fasilitas kesehatan tingkat pertama (FKTP), dan fasilitas kesehatan rujukan tingkat lanjut (FKRTL). Peneliti melakukan analisis mendalam terhadap struktur data, seperti jenis kelamin, usia, wilayah tempat tinggal, dan jenis fasilitas yang digunakan oleh peserta. Data yang digunakan meliputi tahun 2015 hingga 2021.

3. Persiapan Data (Data Preparation)

Tahap ini melibatkan langkah-langkah pembersihan data, termasuk:

- Penggabungan Data: Menggabungkan data dari berbagai sumber untuk menghasilkan dataset yang lebih lengkap.
- Pembersihan Data: Menangani nilai yang hilang dengan mengisi nilai numerik menggunakan rata-rata dan nilai kategorikal dengan modus. Data duplikat juga dihapus untuk memastikan kualitas analisis.
- Normalisasi Data: Untuk memastikan setiap fitur memiliki skala yang seragam, normalisasi dilakukan agar fitur seperti usia dan frekuensi kunjungan memiliki pengaruh yang setara dalam model.

4. Modelling (Pembuatan Model)

Pemilihan Fitur: Fitur yang relevan untuk segmentasi wilayah meliputi usia, jenis kelamin, provinsi peserta, frekuensi kunjungan ke FKTP dan FKRTL.

- Penggunaan K-Means Clustering: Metode K-Means digunakan untuk melakukan segmentasi berdasarkan karakteristik tersebut. Untuk menentukan jumlah kluster yang optimal, digunakan metode Elbow yang memplot nilai inertia. Dari hasil analisis, ditemukan bahwa jumlah kluster yang optimal adalah 3.
- Penerapan Model: Model K-Means diterapkan dengan menggunakan tiga kluster, yang selanjutnya diinterpretasikan untuk melihat karakteristik tiap kluster berdasarkan variabel yang ada.

5. Evaluasi (Evaluation)

Evaluasi dilakukan dengan menggunakan Silhouette Coefficient (SC), yang mengukur kualitas kluster. Nilai SC yang lebih besar dari 0.55 menunjukkan pemisahan yang lebih jelas antara kluster. Selain itu, visualisasi dilakukan menggunakan Principal Component Analysis (PCA) untuk mereduksi dimensi data dan mempermudah pemahaman hasil kluster. Hasil visualisasi menunjukkan perbedaan yang jelas antara ketiga kluster yang terbentuk.

6. Deployment

Setelah evaluasi, tahap akhir adalah penyusunan laporan yang menyajikan hasil segmentasi wilayah dan memberikan rekomendasi untuk strategi intervensi berdasarkan tipologi wilayah yang terbentuk. Laporan ini dapat digunakan oleh BPJS Kesehatan untuk merencanakan distribusi layanan yang lebih tepat sasaran sesuai dengan kebutuhan masing-masing kluster.

7. Relevansi Kluster dengan Kebijakan Kesehatan

Setiap kluster memiliki karakteristik yang berbeda dan membutuhkan pendekatan yang disesuaikan:

- Kluster 0 (Warna Kuning): Mungkin membutuhkan pendekatan preventif seperti penyuluhan kesehatan.
- Kluster 1 (Warna Biru Muda): Menunjukkan kebutuhan akan perawatan lanjutan dengan penyuluhan lebih intensif.
- Kluster 2 (Warna Ungu): Memerlukan perawatan intensif dan perhatian lebih dari fasilitas kesehatan tingkat lanjut.

Metodologi ini berfokus pada penggunaan data untuk mendukung kebijakan yang lebih efektif dalam penanganan Tuberkulosis di Indonesia dengan pendekatan berbasis data yang lebih terstruktur

IV. HASIL DAN PEMBAHASAN

A. HASIL

1) ANALISIS DESKRIPTIF DATA

Sebelum melakukan segmentasi, dilakukan pembersihan data yang mencakup penggabungan data kepesertaan dan pelayanan, penanganan *missing value*, serta penghapusan duplikat data. Berikut adalah analisis deskriptif yang dilakukan terhadap data peserta dan pelayanan yang digunakan dalam penelitian ini:

- Data Kepesertaan (2019 - 2021): Data ini mencakup lebih dari 2 juta peserta dengan atribut seperti jenis kelamin, usia, provinsi tempat tinggal, dan status perkawinan. Kolom penting seperti *Tanggal Lahir Peserta* digunakan untuk menghitung usia, dan kolom *Provinsi Peserta* serta *Kabupaten/Kota* memberikan informasi lokasi yang digunakan dalam segmentasi wilayah.
- Pelayanan FKTP dan FKRTL: Data ini menunjukkan jumlah kunjungan peserta ke fasilitas kesehatan tingkat pertama (FKTP) dan tingkat lanjut (FKRTL). Data mencakup jenis layanan yang diterima oleh peserta, yang menjadi dasar dalam pembentukan kluster. Frekuensi kunjungan ke FKTP dan FKRTL adalah fitur kunci yang digunakan untuk segmentasi wilayah.

2) HASIL SEGMENTASI MENGGUNAKAN K-MEANS

Model K-Means digunakan untuk mengelompokkan peserta BPJS Tuberkulosis berdasarkan karakteristik demografi dan perilaku akses fasilitas kesehatan. Setelah melakukan normalisasi data, peneliti menggunakan Metode Elbow untuk menentukan jumlah kluster optimal. Hasilnya, jumlah kluster optimal adalah 3 kluster.

Hasil kluster:

- Kluster 0 (Warna Kuning): Kluster ini terdiri dari peserta dengan usia lebih muda dan frekuensi kunjungan ke fasilitas kesehatan yang lebih rendah. Peserta di kluster ini kemungkinan besar memerlukan perhatian preventif, seperti penyuluhan kesehatan untuk mencegah TB.
- Kluster 1 (Warna Biru Muda): Kluster ini berisi peserta dengan usia dewasa dan memiliki frekuensi kunjungan lebih tinggi ke FKTP dan FKRTL. Ini menunjukkan bahwa mereka membutuhkan perawatan lanjutan, dan intervensi seperti edukasi mengenai pengelolaan penyakit sangat dibutuhkan.
- Kluster 2 (Warna Ungu): Peserta dalam kluster ini biasanya lebih tua dan sering mengakses fasilitas kesehatan tingkat lanjut. Mereka memerlukan perawatan intensif dan perhatian lebih dari fasilitas kesehatan untuk menangani TB yang lebih serius.

3) EVALUASI HASIL KLASSTER

Untuk mengevaluasi kualitas kluster yang terbentuk, dilakukan perhitungan Silhouette Coefficient (SC). Hasilnya menunjukkan nilai SC sebesar 0.2745, yang mengindikasikan bahwa pemisahan antar kluster masih kurang jelas. Idealnya, nilai SC di atas 0.55 menunjukkan pemisahan yang lebih baik dan kualitas kluster yang lebih tinggi.

Visualisasi Hasil Klasterisasi:

Penggunaan Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data menjadi dua komponen utama. Hasil visualisasi yang ditampilkan pada grafik scatter plot menunjukkan pemisahan yang jelas antar klaster, meskipun kualitas klaster belum optimal. Warna pada plot mewakili masing-masing klaster, dengan klaster 0 (kuning) di sisi kiri atas, klaster 1 (biru muda) di tengah, dan klaster 2 (ungu) di bagian bawah

4) RELEVANSI KLASSTER DENGAN KEBIJAKAN KESEHATAN

Setelah melakukan evaluasi hasil klasterisasi, dilakukan analisis relevansi klaster-klaster yang terbentuk dengan kebijakan kesehatan yang ada. Masing-masing klaster memiliki karakteristik yang berbeda, yang berhubungan dengan kebijakan intervensi yang diperlukan:

- Klaster 0 (Warna Kuning):
Memerlukan pendekatan preventif, seperti edukasi tentang pencegahan TB, pemeriksaan rutin, dan promosi kesehatan untuk mengurangi risiko infeksi.
- Klaster 1 (Warna Biru Muda):
Memerlukan perawatan lanjutan, termasuk pengelolaan penyakit TB yang lebih intensif. Strategi intervensi melibatkan penyuluhan lebih lanjut mengenai pengelolaan kesehatan dan akses ke layanan lebih lanjut.
- Klaster 2 (Warna Ungu):
Memerlukan perawatan intensif dan perhatian lebih dari fasilitas kesehatan tingkat lanjut, karena peserta dalam klaster ini mungkin mengalami TB yang lebih serius dan memerlukan intervensi medis yang lebih mendalam.

B. PEMBAHASAN

Hasil klasterisasi menunjukkan bahwa segmentasi peserta BPJS Tuberkulosis berdasarkan usia dan frekuensi kunjungan ke fasilitas kesehatan adalah pendekatan yang efektif untuk mengidentifikasi kebutuhan intervensi yang berbeda-beda di setiap wilayah. Evaluasi dengan Silhouette Coefficient menunjukkan adanya potensi untuk meningkatkan pemisahan klaster dengan penggabungan fitur tambahan atau menggunakan model lain yang lebih sesuai.

Meskipun evaluasi visualisasi menunjukkan adanya perbedaan yang cukup jelas antar klaster, hasil klasterisasi masih dapat ditingkatkan dengan memperbaiki kualitas data atau menambahkan variabel lain yang lebih relevan. Penelitian lebih lanjut dapat melibatkan penggunaan teknik yang lebih kompleks atau penggabungan variabel spasial dan sosial ekonomi untuk memperkaya segmentasi.

V. KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan peserta BPJS yang terdiagnosis Tuberkulosis (TB) berdasarkan segmentasi usia dan perilaku akses fasilitas kesehatan menggunakan metode K-Means Clustering. Hasil segmentasi menunjukkan bahwa klaster-klaster yang terbentuk memiliki karakteristik yang berbeda, yang masing-masing memerlukan pendekatan intervensi yang disesuaikan.

Klaster pertama (Klaster 0) terdiri dari peserta yang lebih muda dengan frekuensi kunjungan rendah ke fasilitas kesehatan. Mereka

membutuhkan pendekatan preventif seperti penyuluhan kesehatan dan promosi pencegahan TB. Klaster kedua (Klaster 1) menunjukkan peserta dewasa dengan frekuensi kunjungan lebih tinggi, yang membutuhkan perawatan lanjutan dan pengelolaan penyakit yang lebih intensif. Klaster ketiga (Klaster 2) berisi peserta yang lebih tua dengan kebutuhan perawatan intensif dan perhatian lebih lanjut dari fasilitas kesehatan tingkat lanjut.

Meskipun hasil segmentasi sudah memberikan gambaran yang jelas mengenai kebutuhan intervensi di masing-masing klaster, kualitas klasterisasi masih dapat ditingkatkan. Evaluasi menggunakan Silhouette Coefficient menunjukkan bahwa pemisahan antar klaster dapat diperbaiki dengan penambahan variabel atau penggunaan model yang lebih kompleks. Selain itu, penelitian lebih lanjut dapat melibatkan faktor spasial atau sosial ekonomi untuk meningkatkan ketepatan dalam segmentasi.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam pengembangan strategi intervensi berbasis data untuk penanganan Tuberkulosis yang lebih efisien dan terarah sesuai dengan karakteristik masing-masing wilayah dan kelompok peserta.

REFERENSI

- [1] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987. doi: 10.1016/0377-0427(87)90125-7.
- [2] Hanniva, A. Kurnia, S. Rahardianto, and A. A. Mattjik, "Penggerombolan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan Metode K-Means dan Fuzzy C-Means," *Xplore: Journal of Statistics*, vol. 11, no. 1, pp. 36-47, 2022. doi: 10.29244/xplore.v11i1.855.
- [3] L. Anggilia, "Analisa Clustering Penyebaran Penyakit Tuberkulosis Menggunakan Algoritma K-Means di Provinsi Sumatera Selatan," *Skripsi*, Universitas Bina Sarana Informatika, Jakarta, 2024.
- [4] A. Bustamam, K. Burrage, and N. A. Hamilton, "Fast Parallel Markov Clustering in Bioinformatics Using Massively Parallel Computing on GPU with CUDA and ELLPACK-R Sparse Format," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 3, pp. 679-688, May/June 2012. doi: 10.1109/TCBB.2011.68.