IMPLEMENTASI METODE K-MEANS UNTUK MENGELOMPOKKAN RUMAH SAKIT DI INDONESIA BERDASARKAN KAPASITAS PELAYANAN

Aufa Fahmi Azhar S, April Lia Hananto, Bayu Priyatna, Shofa Shofiah Hilabi
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jl. Ronggo Waluyo Sinarbaya, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
si22.aufas@mhs.ubpkarawang.ac.id

ABSTRAK

Perbedaan kapasitas pelayanan antar rumah sakit di Indonesia dapat menyebabkan ketimpangan akses layanan kesehatan. Permasalahan tersebut perlu dianalisis agar diperoleh gambaran distribusi kapasitas rumah sakit secara objektif. Penelitian ini bertujuan untuk mengelompokkan rumah sakit di Indonesia berdasarkan kapasitas pelayanan. Metode yang digunakan adalah algoritma K-Means dengan variabel jumlah tempat tidur, total layanan, dan jumlah tenaga kerja. Tahapan penelitian meliputi preprocessing data, normalisasi, penerapan K-Means, serta evaluasi menggunakan Silhouette Coefficient. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster dengan nilai Silhouette sebesar 0,5367. Cluster yang terbentuk terdiri dari kategori kapasitas rendah, sedang, dan tinggi. Visualisasi menggunakan PCA menunjukkan adanya overlap antar cluster, yang mengindikasikan bahwa kapasitas pelayanan rumah sakit bersifat kontinu. Dengan demikian, metode K-Means efektif digunakan sebagai alat segmentasi untuk memahami distribusi kapasitas rumah sakit di Indonesia.

Kata kunci : *K-Means, Clustering, kapasitas pelayanan, rumah sakit, silhouette score.*

1. PENDAHULUAN

Rumah sakit memiliki peran strategis dalam sistem pelayanan kesehatan karena menjadi fasilitas utama dalam penyediaan layanan medis bagi masyarakat. Kemampuan rumah sakit dalam melayani pasien umumnya tercermin dari kapasitas pelayanan yang dimiliki, seperti jumlah tempat tidur, jumlah tenaga kerja, dan total layanan yang tersedia. Kapasitas tersebut menjadi indikator penting untuk menilai kesiapan fasilitas kesehatan dalam memenuhi kebutuhan masyarakat secara optimal.

Pada kondisi nyata, kapasitas pelayanan antar rumah sakit di Indonesia menunjukkan perbedaan yang cukup besar. Sebagian rumah sakit menghadapi beban pelayanan yang tinggi akibat tingginya jumlah pasien, sedangkan rumah sakit lainnya masih belum dimanfaatkan secara maksimal. Ketimpangan tersebut berpotensi memengaruhi kualitas layanan kesehatan dan pemerataan akses masyarakat terhadap fasilitas kesehatan. Oleh karena itu, diperlukan pendekatan berbasis data untuk memahami pola distribusi kapasitas rumah sakit secara objektif.

Salah satu pendekatan yang dapat digunakan adalah data mining melalui teknik clustering. Teknik ini memungkinkan data dikelompokkan berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label awal [1].

Dalam penelitian ini digunakan algoritma K-Means karena memiliki proses yang sederhana, efisien, dan sesuai untuk pengolahan data numerik dalam jumlah besar. Selain itu, metode ini telah banyak diterapkan pada berbagai penelitian pengelompokan data.

Agar hasil pengelompokan lebih representatif, penelitian ini diawali dengan tahap preprocessing data yang meliputi pembersihan data, penanganan nilai hilang, normalisasi, serta penambahan fitur rasio. Setelah itu, dilakukan proses clustering menggunakan algoritma K-Means. Kualitas cluster yang dihasilkan kemudian dievaluasi menggunakan Silhouette Coefficient untuk menentukan jumlah cluster optimal [2].

Hasil pengelompokan selanjutnya divisualisasikan menggunakan Principal Component Analysis (PCA) agar distribusi cluster dapat diamati dengan lebih jelas.

Merujuk pada permasalahan tersebut, penelitian ini bertujuan untuk mengelompokkan rumah sakit di Indonesia berdasarkan kapasitas pelayanan, menentukan jumlah cluster yang paling optimal, serta menganalisis karakteristik setiap cluster yang terbentuk. Hasil penelitian diharapkan dapat memberikan gambaran distribusi kapasitas rumah sakit di Indonesia dan menjadi bahan pertimbangan dalam mendukung pemerataan serta pengembangan layanan kesehatan.

2. TINJAUAN PUSTAKA

2.1. Rumah Sakit dan Kapasitas Pelayanan

Rumah sakit merupakan lembaga pelayanan kesehatan yang menyediakan layanan medis secara komprehensif, meliputi rawat inap, rawat jalan, dan pelayanan gawat darurat [3].

Dalam sistem kesehatan nasional, rumah sakit berperan sebagai indikator utama dalam mengukur tingkat kesiapan suatu wilayah dalam menyediakan layanan kesehatan yang mencukupi [4].

Kapasitas rumah sakit dapat diukur melalui beberapa indikator utama, yaitu jumlah tempat tidur, jumlah tenaga kerja kesehatan, serta total layanan yang disediakan.

Perbedaan kapasitas layanan antar rumah sakit menunjukkan adanya variasi kemampuan dalam memberikan layanan kesehatan kepada masyarakat [5].

Variasi ini sering menyebabkan ketidakseimbangan layanan, sehingga diperlukan analisis berbasis data untuk memahami pola dan karakteristik kapasitas rumah sakit secara tujuan [6].

2.2. Data Mining

Data *Mining* merupakan proses ekstraksi pola atau informasi yang berguna dari analisis dataset dalam jumlah besar menggunakan teknik tertentu [7].

Proses data *Mining* umumnya meliputi beberapa tahapan, yaitu pembersihan data, transformasi data, seleksi data, pemodelan, dan evaluasi [8].

Dalam bidang kesehatan, data *Mining* digunakan untuk berbagai tujuan, seperti analisis epidemiologi, prediksi penyakit, serta segmentasi fasilitas kesehatan. Dengan menggunakan data *Mining*, informasi yang tersembunyi di dalam dataset dapat diidentifikasi sehingga membantu dalam pengambilan keputusan berdasarkan data.

2.3. Clustering

Klasterisasi data merupakan instrumen dalam data *Mining* yang mengadopsi prinsip pemelajaran tanpa pengawasan (unsupervised learning). Teknik ini bekerja dengan cara mengorganisir sekumpulan informasi ke dalam beberapa kelompok berdasarkan kesamaan karakteristiknya tanpa memerlukan panduan label atau kategori yang telah ditentukan sebelumnya [9].

Fokus utama dari proses klasterisasi terletak pada upaya meningkatkan keseragaman internal dengan cara memperkecil jarak antar data di dalam satu kelompok yang sama (intra-cluster), sembari mempertegas batasan dengan menciptakan jarak sejauh mungkin antar kelompok yang berbeda (inter-cluster) [10]. Dalam konteks penelitian ini, *Clustering* digunakan untuk mengelompokkan rumah sakit berdasarkan kemampuan pelayanan yang dimiliki.

2.4. Algoritma K-Means

Dalam dunia pengolahan data, K-Means menonjol sebagai salah satu teknik pengelompokan (*Clustering*) yang paling diandalkan karena efisiensi dan kemudahan pengaplikasiannya pada data numerik [11].

Algoritma ini beroperasi dengan membagi sekumpulan data ke dalam beberapa kelompok berdasarkan kedekatan jaraknya terhadap titik pusat kelompok yang disebut centroid [12].

Mekanismenya dimulai dengan menetapkan jumlah klasternya terlebih dahulu, lalu setiap data akan ditarik menuju centroid terdekat. Posisi pusat kelompok ini kemudian diperbarui secara berulang mengikuti nilai rata-rata dari anggota di dalamnya hingga mencapai kondisi stabil atau tidak lagi mengalami perubahan yang berarti [13].

Kelebihan utama dari K-Means terletak pada proses komputasinya yang relatif cepat serta fleksibilitasnya dalam menangani basis data berskala besar [14].

Pendekatan algoritma tersebut diimplementasikan untuk mengklasifikasikan berbagai rumah sakit di Indonesia ke dalam beberapa tingkatan kapasitas. Proses segmentasi ini didasarkan pada tiga variabel utama, yaitu kapasitas tempat tidur, volume aktivitas layanan, serta jumlah energi kesehatan yang tersedia. Untuk menentukan tingkat kedekatan atau kedekatan posisi data terhadap pusat kelompok (centroid), penelitian ini menggunakan teknik perhitungan Euclidean Distance dengan persamaan sebagai berikut:

$$d(x,y) = \sqrt{\Sigma(x_i - y_i)}^2 \quad (1)$$

Variabel d(x,y) dalam persamaan tersebut berfungsi sebagai parameter untuk mengukur tingkat kedekatan atau jarak antara dua entitas data yang dibandingkan. Simbol xi dan yi masing-masing merepresentasikan nilai pada dimensi atau fitur ke-i dari kedua data tersebut, sementara variabel n menunjukkan total jumlah atribut yang dilibatkan dalam seluruh proses perhitungan jarak. Melalui komponen-komponen ini, algoritma dapat menentukan seberapa besar kemiripan antara satu objek data dengan objek lainnya secara matematis.

2.5. Silhouette Coefficient

Untuk memvalidasi efektivitas pengelompokan yang dihasilkan, penelitian ini menggunakan *Silhouette Coefficient* sebagai metrik evaluasi kualitas klaster. Metode ini bekerja dengan membandingkan tingkat kepadatan data dalam satu kelompok (cohesion) terhadap jaraknya dengan kelompok lain yang paling mirip (separation). Skor yang dihasilkan berada pada interval -1 hingga 1, di mana nilai yang mendekati 1 mencerminkan bahwa data telah terbagi secara akurat dengan pemisahan yang jelas antar klaster [15].

Sebaliknya, nilai di sekitar 0 menandakan adanya tumpang tindih pada perbatasan kelompok, dan nilai negatif mengindikasikan adanya kekeliruan dalam penempatan data. Dalam konteks riset ini, *Silhouette Coefficient* berperan krusial untuk menentukan jumlah klaster (K) yang paling tepat agar model mampu merepresentasikan kapasitas rumah sakit secara optimal, dengan perhitungan matematis sebagai berikut:

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (2)$$

Variabel s(i) dalam rumus tersebut bertindak sebagai parameter evaluasi untuk mengukur sejauh mana sebuah data entitas selaras dengan klaster yang ditempatinya. Nilai a(i) mencerminkan tingkat kepadatan internal atau jarak rata-rata data terhadap anggota lain pada kelompok asalnya, sedangkan b(i) mewakili jarak rata-rata data tersebut ke klaster

terdekat yang bukan kelompoknya. Dengan membandingkan kedua aspek ini, peneliti dapat menentukan kualitas pemisahan antar kelompok serta memastikan bahwa setiap data telah terbagi secara optimal sesuai dengan struktur alaminya.

2.6. Python

Python telah menjadi standar utama dalam ranah kecerdasan buatan, pemelajaran mesin, serta analisis data berkat ekosistemnya yang sangat kaya. Bahasa ini menyediakan berbagai pustaka spesifik untuk mempermudah alur kerja data, seperti Pandas yang digunakan untuk mengorganisir informasi secara efisien dan NumPy untuk menangani kebutuhan komputasi angka yang kompleks. Selain itu, implementasi algoritma pemelajaran mesin, khususnya teknik pengelompokan K-Means, didukung secara optimal oleh pustaka Scikit-Learn yang menawarkan stabilitas dan efisiensi tinggi dalam proses pemodelan data [16].

Keunggulan Python terletak pada kemudahan sintaks, fleksibilitas, serta dukungan komunitas yang luas sehingga memudahkan dalam pengembangan aplikasi berbasis data [17].

Seluruh rangkaian analisis dalam riset ini dikerjakan menggunakan bahasa pemrograman Python sebagai fondasi teknis utamanya. Alur kerja yang dijalankan mencakup tahapan prapemrosesan data, penyelarasan skala melalui normalisasi, hingga penerapan algoritma K-Means untuk pembentukan klaster. Proses tersebut ditutup dengan tahap evaluasi menggunakan Silhouette Score untuk mengukur sejauh mana model yang dihasilkan mampu mengelompokkan data secara akurat dan optimal.

2.7. Google Colab

Aksesibilitas dalam pengembangan model pada penelitian ini didukung oleh penggunaan Google Colab, sebuah lingkungan pemrograman berbasis awan yang memungkinkan eksekusi skrip Python secara langsung melalui peramban web. Keunggulan utamanya terletak pada kemudahan penggunaan tanpa memerlukan konfigurasi atau instalasi perangkat lunak di komputer pribadi, sehingga proses analisis menjadi lebih praktis. Selain telah terintegrasi dengan berbagai ekosistem pustaka sains data dan pemelajaran mesin, platform ini juga menyediakan dukungan infrastruktur komputasi yang mumpuni, termasuk akses ke GPU serta TPU untuk mengoptimalkan kecepatan pemrosesan data yang kompleks [18].

Penggunaan Google Colab bertujuan untuk mempermudah proses implementasi algoritma K-Means dan pengolahan data secara efisien [19].

Selain itu, Colab juga memungkinkan penyimpanan dan integrasi data dengan Google Drive, sehingga memudahkan dalam pengelolaan dataset serta replikasi eksperimen penelitian.

2.8. Penelitian Terdahulu

Kajian terhadap penelitian terdahulu diperlukan untuk melihat perkembangan penerapan metode clustering pada bidang kesehatan sekaligus mengidentifikasi posisi penelitian yang dilakukan. Beberapa studi sebelumnya menunjukkan bahwa algoritma K-Means banyak digunakan karena mampu mengelompokkan data numerik secara efektif dan efisien. Ringkasan penelitian terdahulu yang relevan disajikan pada Tabel 1.

Tabel 1. Ringkasan Penelitian Terdahulu

No	Penelitian	Metode	Hasil Utama
1	Lutfiannisa dkk. (2024)	K-Means	Menghasilkan 2 cluster pasien berdasarkan usia untuk mendukung analisis kebutuhan layanan kesehatan.
2	Pahlevi dkk. (2024)	K-Means	Mengelompokkan wilayah di Indonesia berdasarkan distribusi tenaga kesehatan untuk mendukung pemerataan layanan.
3	Ziqrullah dkk. (2024)	K-Means	Menghasilkan cluster karakteristik penyakit pasien sebagai dasar pengambilan keputusan pelayanan medis.
4	Shepyantoni dkk. (2024)	K-Means	Membentuk cluster pasien rawat inap BPJS berdasarkan intensitas layanan dan lama rawat inap.
5	Nengsih dkk. (2025)	K-Means + Silhouette	Menghasilkan segmentasi pasien berdasarkan pola kunjungan dengan cluster optimal.
6	Penelitian ini	K-Means + Silhouette	Mengelompokkan rumah sakit di Indonesia berdasarkan kapasitas pelayanan.

Herlin Lutfiannisa dkk. menerapkan algoritma K-Means untuk mengelompokkan data pasien berdasarkan usia di puskesmas. Hasil penelitian tersebut menunjukkan bahwa pembagian data ke dalam dua cluster dapat membantu memahami profil pasien dan kebutuhan layanan kesehatan. Temuan ini membuktikan bahwa metode clustering dapat digunakan sebagai alat segmentasi pada data pelayanan kesehatan.

Pada skala yang lebih luas, Pahlevi dkk. menggunakan K-Means untuk menganalisis distribusi tenaga kesehatan di Indonesia. Hasil penelitian menunjukkan adanya beberapa kelompok wilayah berdasarkan kondisi persebaran tenaga kesehatan. Penelitian ini menegaskan bahwa metode K-Means dapat dimanfaatkan untuk mendukung kebijakan pemerataan layanan kesehatan antar wilayah.

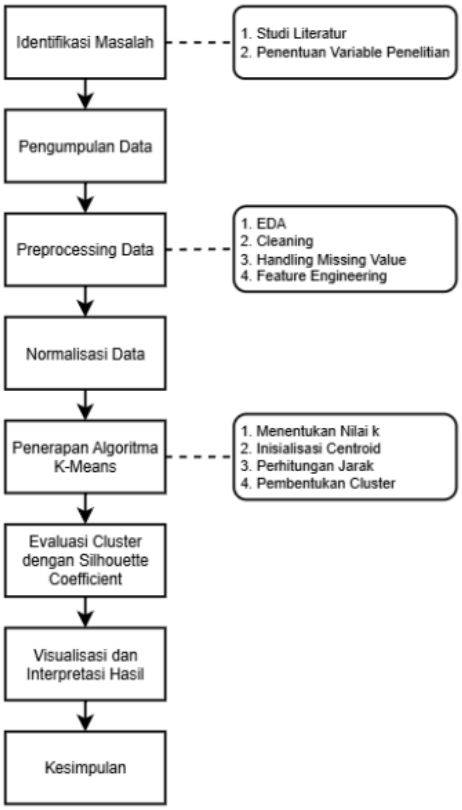
Penelitian lain dilakukan oleh Ziqrullah dkk. yang menerapkan K-Means untuk mengelompokkan data penyakit pasien di puskesmas. Selain itu, Shepyantoni dkk. juga menggunakan metode yang sama untuk menganalisis data pasien rawat inap

peserta BPJS di rumah sakit. Kedua penelitian tersebut menunjukkan bahwa K-Means efektif digunakan dalam mengidentifikasi pola data pasien maupun pemanfaatan fasilitas layanan kesehatan.

Selanjutnya, Nengsih dkk. mengombinasikan algoritma K-Means dengan Silhouette Coefficient untuk melakukan segmentasi pasien rumah sakit berdasarkan pola kunjungan. Hasil evaluasi menunjukkan bahwa penggunaan Silhouette mampu membantu menentukan jumlah cluster yang paling optimal sehingga kualitas pengelompokan menjadi lebih baik.

Berdasarkan beberapa penelitian terdahulu tersebut, metode K-Means terbukti relevan untuk pengelompokan data kesehatan, baik pada tingkat pasien, distribusi tenaga kesehatan, maupun fasilitas layanan. Namun, penelitian yang secara khusus mengelompokkan rumah sakit di Indonesia berdasarkan kapasitas pelayanan menggunakan indikator jumlah tempat tidur, jumlah tenaga kerja, dan total layanan masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut dengan menerapkan algoritma K-Means dan evaluasi menggunakan Silhouette Coefficient guna menghasilkan segmentasi rumah sakit berdasarkan kapasitas pelayanan secara lebih terstruktur.

3. METODE PENELITIAN



Gambar 1. Tahapan penelitian

Penelitian ini menggunakan pendekatan data *Mining* dengan metode *Clustering* untuk mengelompokkan rumah sakit di Indonesia berdasarkan kemampuan pelayanan. Proses penelitian dilakukan secara sistematis melalui beberapa tahapan, mulai dari pengumpulan masalah, pengumpulan data, prapemrosesan data, normalisasi data, penerapan algoritma K-Means, evaluasi cluster, hingga interpretasi hasil. Seluruh tahapan penerapan dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab untuk mendukung proses analisis data secara efisien dan terstruktur.

3.1. Identifikasi Masalah

Tahap identifikasi masalah dilakukan untuk menentukan permasalahan utama yang akan diselesaikan dalam penelitian ini. Permasalahan yang menjadi fokus adalah adanya variasi kapasitas pelayanan rumah sakit di Indonesia yang cukup beragam, sehingga diperlukan metode pengelompokan untuk mengetahui pola kapasitas rumah sakit berdasarkan indikator jumlah tempat tidur, jumlah tenaga kerja, dan total layanan. Pada tahap ini juga dilakukan studi pustaka untuk memperkuat dasar pemilihan metode dan menentukan variabel penelitian yang relevan.

3.2. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan memperoleh kumpulan data rumah sakit di Indonesia yang bersumber dari data terbuka. Dataset yang digunakan berisi informasi angka terkait kapasitas pelayanan rumah sakit, yaitu jumlah tempat tidur, jumlah tenaga kerja, dan total layanan. Data kemudian diimpor ke lingkungan Google Colab untuk dilakukan proses analisis dan pengolahan lebih lanjut.

3.3. Preprocessing Data

Data yang didapat masih dalam kondisi mentah sehingga perlu dilakukan preprocessing untuk meningkatkan kualitas data. Tahapan ini meliputi beberapa proses, yaitu :

a. Tahap Analisis Data Eksploratori (EDA) diterapkan guna memetakan pola sebaran data sekaligus mendeteksi keberadaan nilai pencilan (outlier) yang berpotensi memengaruhi validitas hasil analisis.
b. Pembersihan Data dilakukan dengan menghapus data yang memiliki nilai ekstrim dan tidak realistis, misalnya jumlah tempat tidur yang melebihi batas wajar.
c. Handling Missing Value dilakukan dengan menghapus data yang memiliki nilai kosong agar tidak mempengaruhi hasil analisis.
d. Feature Engineering dilakukan dengan menambahkan variabel baru berupa rasio, yaitu rasio energi kerja terhadap tempat tidur, rasio layanan terhadap tempat tidur, dan rasio layanan terhadap tenaga kerja.

Tahapan ini bertujuan untuk menghasilkan kumpulan data yang lebih representatif sehingga proses *Clustering* dapat menghasilkan kelompok yang lebih optimal.

3.4. Normalisasi Data

Usai fase prapemrosesan, data melalui tahap standarisasi dengan memanfaatkan metode StandardScaler. Prosedur ini dilakukan untuk menyeragamkan rentang nilai pada seluruh variabel, guna menghindari adanya dominasi fitur tertentu yang dapat memengaruhi akurasi perhitungan jarak dalam algoritma K-Means. Melalui penerapan normalisasi ini, setiap parameter yang dianalisis memiliki bobot yang setara, sehingga hasil pengelompokan menjadi lebih objektif dan mampu merepresentasikan karakteristik data secara seimbang.

3.5. Penerapan Algoritma K-Means

Pada tahap ini dilakukan proses *Clustering* menggunakan algoritma K-Means dengan bantuan perpustakaan Scikit-Learn. Proses K-Means meliputi penentuan jumlah cluster (k), inisialisasi centroid, penghitungan jarak antar data menggunakan Euclidean Distance, serta pembentukan cluster berdasarkan jarak data terhadap centroid. Proses iterasi dilakukan hingga centroid tidak terjadi perubahan yang signifikan sehingga menghasilkan cluster yang stabil.

3.6. Evaluasi Cluster dengan *Silhouette Coefficient*

Penentuan jumlah kelompok (K) yang paling ideal dalam penelitian ini dilakukan melalui prosedur evaluasi berdasarkan Silhouette Coefisien, yang menilai kualitas pengelompokan berdasarkan kepadatan data di dalam satu grup (intra-cluster) serta tingkat keterpisahan dengan grup lainnya (inter-cluster). Prosedur ini dijalankan secara berulang dengan memvariasikan nilai K dari 2 hingga 10 guna mengidentifikasi konfigurasi yang paling akurat dalam merepresentasikan pola distribusi data secara keseluruhan. Secara prinsip, perolehan skor yang semakin tinggi menjulang langsung dengan jarak segmentasi yang lebih tajam, sehingga nilai tertinggi yang dihasilkan dalam pengujian tersebut dijadikan sebagai acuan utama untuk menetapkan jumlah klaster final pada algoritma K-Means.

Dalam tahap interpretasi, skor yang mendekati angka 1 mencerminkan bahwa setiap objek data telah berada pada kelompok yang tepat dengan batas batasan yang sangat tegas antar klaster. Sementara itu, nilai pada rentang menengah menunjukkan adanya struktur yang cukup kokoh namun tetap menyisakan kemungkinan terjadinya tumpang tindih (overlap), sedangkan skor yang rendah menunjukkan lemahnya formasi klaster atau adanya kesalahan dalam penempatan data. Melalui pendekatan dua arah ini, peneliti dapat memastikan bahwa hasil *Clustering* yang diperoleh bukan sekedar hasil acak, melainkan didasarkan pada validasi statistik yang menunjukkan seberapa baik tiap data sesuai dengan kelompoknya masing-masing [20].

3.7. Visualisasi dan Interpretasi Hasil

Untuk memahami hasil *Clustering*, dilakukan visualisasi menggunakan Principal Component Analysis (PCA) untuk mereduksi data dimensi menjadi dua dimensi. Visualisasi ini digunakan untuk melihat distribusi cluster dan pola fragmentasi antar kelompok.

Selanjutnya dilakukan interpretasi hasil dengan analisis karakteristik masing-masing cluster berdasarkan nilai rata-rata variabel kapasitas pelayanan. Berdasarkan hasil tersebut, cluster dibagi menjadi rumah sakit dengan kapasitas rendah, sedang, dan tinggi.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Preprocessing Data

Tahap praproses dilakukan untuk meningkatkan kualitas dataset sebelum proses *Clustering*. Berdasarkan data hasil eksplorasi, ditemukan adanya nilai ekstrem pada variabel jumlah tempat tidur yang tidak realistis. Oleh karena itu dilakukan proses pembersihan data dengan menghapus data yang mempunyai nilai di atas batas wajar. Selain itu, data yang mempunyai nilai nol dan nilai hilang juga dihapus agar tidak mempengaruhi hasil analisis.

Selanjutnya dilakukan feature engineering dengan menambahkan variabel rasio, yaitu rasio tenaga kerja terhadap tempat tidur, rasio layanan terhadap tempat tidur, dan rasio layanan terhadap tenaga kerja. Penambahan fitur ini bertujuan untuk menjelaskan karakteristik kapasitas pelayanan rumah sakit. Setelah itu dilakukan normalisasi data menggunakan StandardScaler untuk menyamakan skala antar variabel sehingga tidak terjadi dominasi dalam proses komputasi jarak.

Tabel 2. Statistik Deskriptif Data Rumah Sakit

	Total Tempat Tidur	Total Layanan	Total Tenaga Kerja
mean	133.38	42.93	295.66
std	123.29	30.34	393.80
min	2.00	1.00	1.00
max	1966.00	280.00	7939.00

Hasil statistik deskriptif menunjukkan bahwa data rumah sakit mempunyai variasi yang cukup tinggi pada setiap variabel yang digunakan. Rata-rata jumlah tempat tidur sebesar 133,38 dengan standar deviasi sebesar 123,29 yang menunjukkan adanya perbedaan kapasitas yang cukup signifikan antar rumah sakit. Variabel total layanan mempunyai rata-rata sebesar 42,93 dengan standar deviasi sebesar 30,34, sedangkan jumlah tenaga kerja memiliki rata-rata sebesar 295,66 dengan standar deviasi sebesar 393,80.

Nilai minimum pada setiap variabel menunjukkan keberadaan rumah sakit dengan kapasitas sangat kecil, yaitu hanya memiliki 2 tempat tidur, 1 layanan, dan 1 tenaga kerja. Sementara itu, nilai maksimal menunjukkan adanya rumah sakit dengan kapasitas sangat besar, yakni mencapai 1966 tempat tidur, 280 layanan, dan 7939 tenaga kerja. Perbedaan yang cukup besar antara nilai minimum dan maksimum ini menunjukkan bahwa data mempunyai distribusi yang tidak merata dan cenderung heterogen.

Kondisi tersebut mengindikasikan adanya perbedaan kapasitas pelayanan yang signifikan antar rumah sakit di Indonesia, sehingga diperlukan metode *Clustering* untuk mengelompokkan rumah sakit berdasarkan karakteristik kapasitas pelayanannya.

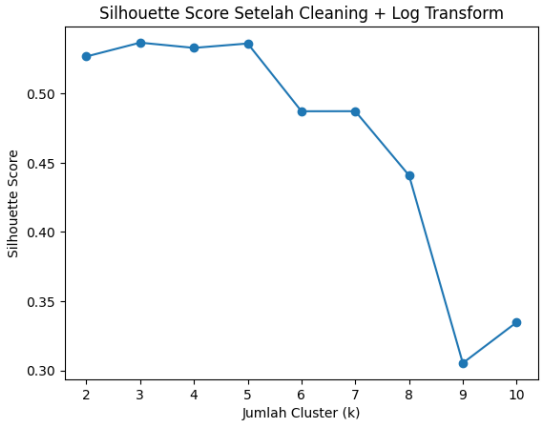
4.2. Penentuan Jumlah Cluster

Penentuan jumlah cluster dalam penelitian ini dilakukan dengan menggunakan metode Koefisien Silhouette. Metode ini digunakan untuk mengukur kualitas hasil *Clustering* berdasarkan tingkat kedekatan data dalam satu cluster (intra-cluster) dan jarak antar cluster (antar-cluster). Pengujian dilakukan dengan variasi jumlah cluster dari K=2 sampai K=10.

Hasil evaluasi Silhouette Score ditunjukkan pada Gambar 2, yang menampilkan hubungan antara jumlah cluster dengan nilai Silhouette yang dihasilkan. Berdasarkan grafik tersebut terlihat bahwa nilai Silhouette tertinggi diperoleh pada K=3 dengan nilai sebesar 0,5367. Nilai ini lebih tinggi dibandingkan dengan K=2 serta jumlah cluster lainnya, sehingga menunjukkan bahwa K=3 merupakan jumlah cluster yang paling optimal.

Selain itu, walaupun pada K=5 diperoleh nilai Silhouette yang mendekati K=3, yaitu sebesar 0,5360, namun secara umum tren grafik menunjukkan bahwa setelah K=3 nilai Silhouette cenderung mengalami penurunan. Penurunan ini menjadi lebih signifikan pada K≥6, yang menunjukkan bahwa penambahan jumlah cluster tidak meningkatkan kualitas pengelompokan, bahkan cenderung menurunkannya.

Nilai Silhouette yang berada pada rentang di atas 0,5 menunjukkan bahwa hasil *Clustering* memiliki kualitas yang cukup baik. Hal ini berarti bahwa data dalam suatu cluster memiliki tingkat kemiripan yang cukup tinggi, serta memiliki perbedaan yang cukup jelas dengan cluster lainnya. Namun demikian, masih terdapat kemungkinan terjadinya tumpang tindih antar cluster, yang menunjukkan bahwa kapasitas data pelayanan rumah sakit bersifat kontinu dan tidak memiliki batasan yang tegas antar kategori.



Gambar 2. Grafik Evaluasi Silhouette Score

Selain itu, penggunaan data yang telah melalui proses cleaning dan transformasi log ikut berkontribusi dalam meningkatkan stabilitas nilai Silhouette, sehingga menghasilkan cluster yang lebih representatif. Berdasarkan hasil tersebut maka jumlah cluster yang digunakan dalam penelitian ini adalah tiga cluster.

4.3. Hasil *Clustering* K-Means

Berdasarkan hasil adopsi algoritma K-Means dengan jumlah cluster optimal sebanyak tiga, diperoleh pengelompokan rumah sakit ke dalam tiga cluster berdasarkan kapasitas pelayanan. Hasil pengelompokan tersebut ditampilkan pada Tabel 3 yang menyajikan nilai rata-rata dari setiap variabel pada masing-masing cluster.

Berdasarkan Tabel 3 terlihat bahwa setiap cluster mempunyai karakteristik yang berbeda-beda. Cluster 1 mempunyai nilai rata-rata tertinggi pada seluruh variabel, yaitu jumlah tempat tidur sebesar 292,40, total layanan sebesar 90,40, dan jumlah tenaga kerja sebesar 785,75. Hal ini menunjukkan bahwa cluster ini merepresentasikan rumah sakit dengan kapasitas pelayanan tinggi.

Cluster 0 memiliki nilai rata-rata yang berada pada level llmenengah, yaitu jumlah tempat tidur sebesar 96,90, total layanan sebesar 32,03, dan jumlah tenaga kerja sebesar 183,40. Oleh karena itu, cluster ini dapat dikelompokkan sebagai rumah sakit dengan kapasitas pelayanan sedang.

Sementara itu, Cluster 2 memiliki nilai rata-rata terendah pada seluruh variabel, yaitu jumlah tempat tidur sebesar 57,14, jumlah layanan sebesar 23,00, dan jumlah tenaga kerja sebesar 1,14. Hal ini menunjukkan bahwa cluster ini merepresentasikan rumah sakit dengan kapasitas pelayanan rendah.

Hasil *Clustering* ini menunjukkan bahwa metode K-Means mampu mengelompokkan rumah sakit di Indonesia berdasarkan kapasitas pelayanan dengan jelas menjadi tiga kategori utama, yaitu kapasitas rendah, sedang, dan tinggi. Pengelompokkan ini memberikan gambaran yang lebih terstruktur mengenai distribusi kapasitas rumah sakit di Indonesia.

Tabel 3. Hasil *Clustering*

Cluster	Tempat Tidur	Layanan	Tenaga Kerja
0	96.90	32.03	183.40
1	292.40	90.40	785.75
2	57.14	23.00	1.14

4.4. Visualisasi Cluster

Untuk memperjelas hasil pengelompokan, dilakukan visualisasi cluster dengan menggunakan metode Principal Component Analysis (PCA). Metode ini digunakan untuk mereduksi dimensi data menjadi dua dimensi sehingga memudahkan dalam melihat distribusi dan perpecahan antar cluster.

Hasil visualisasi ditunjukkan pada Gambar 3, di mana setiap titik merepresentasikan satu rumah sakit yang telah dideklarasikan menggunakan algoritma K-Means. Warna yang berbeda menunjukkan cluster

yang berbeda, sedangkan tanda silang (centroid) menunjukkan pusat dari cluster masing-masing.

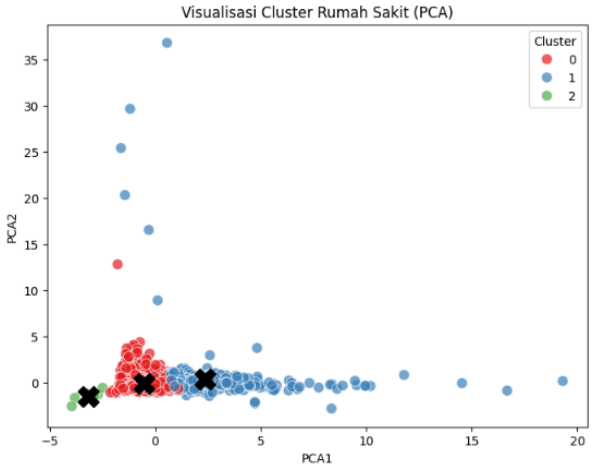
Berdasarkan Gambar 3 terlihat bahwa cluster dengan kapasitas tinggi (Cluster 1) mempunyai penyebaran data yang lebih luas dibandingkan cluster lainnya, terutama pada sumbu PCA1. Hal ini menunjukkan bahwa rumah sakit dengan kapasitas besar memiliki variasi fitur yang lebih beragam.

Cluster dengan kapasitas sedang (Cluster 0) terlihat di sekitar pusat distribusi data, yang menunjukkan bahwa sebagian besar rumah sakit memiliki karakteristik kapasitas yang relatif seragam dan berada pada tingkat menengah.

Sementara itu, cluster berkapasitas rendah (Cluster 2) terlihat berada pada area yang terpisah di sisi kiri grafik dengan penyebaran yang lebih sempit. Hal ini menunjukkan bahwa rumah sakit dengan kapasitas rendah mempunyai karakteristik yang lebih seragam dan berbeda dari cluster lainnya.

Namun begitu, masih terlihat adanya beberapa titik data yang saling berdekatan antar cluster, khususnya antara cluster berkapasitas sedang dan tinggi. Hal ini menunjukkan adanya overlap antar cluster, yang mengindikasikan bahwa batas antar kelompok tidak sepenuhnya tegas.

Kondisi tersebut menunjukkan bahwa kapasitas pelayanan rumah sakit di Indonesia bersifat kontinyu, di mana perbedaan antar rumah sakit tidak selalu terpisah secara jelas. Hal ini juga sejalan dengan hasil evaluasi Silhouette yang menunjukkan bahwa kualitas cluster termasuk cukup baik, namun tidak sepenuhnya terpisah secara sempurna.



Gambar 3. Visualisasi Cluster Rumah Sakit Menggunakan PCA

4.5. Pembahasan

Penelitian ini menerapkan algoritma K-Means untuk memetakan kapasitas pelayanan rumah sakit di Indonesia ke dalam tiga kategori, yakni tingkat rendah, sedang, dan tinggi. Melalui tahapan persiapan data yang matang dan penggunaan fitur rasio, studi ini berhasil mengidentifikasi bahwa pembagian menjadi tiga kelompok merupakan model yang paling optimal dengan nilai *Silhouette Coefficient* sebesar 0,5367. Meskipun pengelompokan ini memiliki tingkat kemiripan anggota yang kuat di setiap klasternya, ditemukan bahwa karakteristik antar data masih memiliki persinggungan tertentu yang mencerminkan kondisi nyata di lapangan.

Hasil analisis menunjukkan adanya kesenjangan yang signifikan, di mana rumah sakit pada kelompok kapasitas tinggi memiliki fasilitas dan tenaga kerja yang jauh melampaui rata-rata, sementara mayoritas rumah sakit di Indonesia berada di level menengah. Visualisasi data mengungkap bahwa batasan antar kategori tidak bersifat kaku melainkan kontinu, sehingga sering terjadi tumpang tindih antara kelompok kapasitas sedang dan tinggi. Oleh karena itu, metode *Clustering* ini lebih efektif diposisikan sebagai alat untuk memetakan distribusi dan segmentasi kapasitas layanan kesehatan nasional daripada digunakan sebagai sistem klasifikasi yang mutlak.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means untuk memetakan kapasitas pelayanan rumah sakit di Indonesia, di mana pembagian menjadi tiga kelompok utama terbukti paling optimal dengan skor *Silhouette Coefficient* sebesar 0,5367. Hasil ini menunjukkan kualitas pengelompokan yang cukup baik, dengan temuan bahwa mayoritas rumah sakit nasional saat ini berada pada kategori kapasitas menengah. Adanya tumpang tindih dalam visualisasi PCA mengindikasikan bahwa karakteristik layanan kesehatan bersifat kontinu dan tidak memiliki batas pemisah yang kaku, sehingga metode ini lebih tepat digunakan sebagai instrumen segmentasi untuk memahami pola distribusi fasilitas daripada sebagai sistem klasifikasi yang mutlak. Untuk pengembangan di masa depan, disarankan agar penelitian selanjutnya mengintegrasikan variabel yang lebih beragam serta mengeksplorasi metode *Clustering* alternatif guna menghasilkan analisis yang lebih tajam dan representatif terhadap dinamika kapasitas kesehatan di Indonesia.

DAFTAR PUSTAKA

[1] P. Lestari, N. Suarna, and W. Prihartono, “Implementasi Data Mining Clustering Dalam Mengelompokkan Penduduk Penyandang Disabilitas Menggunakan Algoritma K-Means,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 958–966, Mar. 2024.

[2] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, “The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data,” *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, Jun. 2022.

[3] M. Hafiz Ziqrullah, S. Dian Purnamasari, and I. Zuhri Yadi, “Jurnal Media Informatika [JUMIN] Clustering Data Penyakit Pasien Pada Puskesmas

Mulyaguna Menggunakan Algoritma K-Means," 2024.

[4] M. Haemmerli, T. Powell-Jackson, C. Goodman, H. Thabrany, and V. Wiseman, "Poor quality for the poor? A study of inequalities in service readiness and provider knowledge in Indonesian primary health care facilities," *Int. J. Equity Health*, vol. 20, no. 1, p. 239, Dec. 2021.

[5] F. R. Muharram *et al.*, "The Indonesia Health Workforce Quantity And Distribution," Apr. 01, 2024.

[6] Y. Mahendradhata *et al.*, "The Capacity of the Indonesian Healthcare System to Respond to COVID-19," *Front. Public Health*, vol. 9, Jul. 2021.

[7] S. S. Hilabi, S. Savina, and S. Khairunisa, "Pemanfaatan Data Analitik dalam Big Data: Studi Kasus Implementasi di Pemerintahan," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 1, Mar. 2025.

[8] Tukino, A. Hananto, R. A. Nanda, E. Novalia, E. Sediyono, and J. Sanjaya, "LSTM and Word Embedding: Classification and Prediction of Puskesmas Reviews Via Twitter," *E3S Web of Conferences*, vol. 500, p. 01018, Mar. 2024.

[9] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Comput. Sci.*, vol. 216, pp. 356–363, 2023.

[10] A. Lia Hananto *et al.*, "Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1908, no. 1, p. 012024, Jun. 2021.

[11] M. Adjie Pahlevi *et al.*, "Analisis Distribusi Tenaga Kesehatan Di Indonesia Menggunakan K-Means Clustering," 2024.

[12] A. Fatlun, S. S. Hilabi, B. Priyatna, and A. L. Hananto, "Penggunaan Algoritma K-Means Clustering CRISP-DM Dalam Mengelompokkan Drama Korea Sebagai Rekomendasi Film," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 1, Mar. 2025.

[13] A. Herlin Lutfiannisa, M. Maimunah, and P. Sukmasetya, "Clustering Data Pasien Berdasarkan Usia di Puskesmas Menerapkan Metode K-Means," *Journal of Information System Research (JOSH)*, vol. 5, no. 2, pp. 639–647, Jan. 2024.

[14] C. Diningrat, B. Priyatna, E. Novalia, and S. S. Hilabi, "Klasterisasi Kasus Kekerasan Berdasarkan Jenis Lokasi Kejadian di Jawa Barat Menggunakan Algoritma K-Means," *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 518–528, May 2025.

[15] I. Grady Favian and E. Suryani, "A Case Study of Applying Customer Segmentation in A Medical Equipment Industry," *IPTEK Journal of Proceedings Series*, vol. 0, no. 3, p. 119, Oct. 2021.

[16] F. Shepyantoni, I. Kanedi, and E. Suryana, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Pasien Rawat Inap Peserta BPJS Di Rumah Sakit Umum Daerah Kabupaten Kaur," *JURNAL MEDIA INFOTAMA*, vol. 20, no. 2, pp. 493–500, Oct. 2024.

[17] Y. G. Nengsih, K. Samosir, and B. Satria, "Segmentasi Pasien Rumah Sakit Berdasarkan Pola Kunjungan Menggunakan Algoritma K-Means Clustering untuk Optimasi Layanan Medis," *Jurnal Pendidikan Sains dan Komputer*, vol. 5, no. 01, pp. 59–69, Feb. 2025.

[18] A. K. Mulugeta, D. P. Sharma, and A. H. Mesfin, "Deep learning for medicinal plant species classification and recognition: a systematic review," *Front. Plant Sci.*, vol. 14, Jan. 2024.

[19] D. Fitriyani, M. Jajuli, and G. Garno, "IMPLEMENTASI ALGORITMA K-MEANS UNTUK KLASTERISASI DALAM PENGELOLAAN PERSEDIAAN OBAT (STUDI KASUS : APOTEK NAZA)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024.

[20] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient," 2021.