

PERBANDINGAN KLASTER HIERARKI DAN K-MEANS DALAM MENGELOMPOKKAN TENAGA KESEHATAN BERDASARKAN RUMAH SAKIT DI KABUPATEN BOJONEGORO TAHUN 2024

Herni^{1*}, Ika Nur Laily Fitriana²

^{1,2}Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka

*Penulis korespondensi: hernipingky@gmail.com

ABSTRAK

Rumah sakit adalah institusi yang menyediakan perawatan rawat inap bagi orang sakit. Di Bojonegoro, pelayanan rumah sakit dinilai cukup baik. Namun, kekurangan tenaga perawat sering membuat keluarga pasien diminta membantu perawatan. Penelitian ini bertujuan untuk melakukan pengelompokan rumah sakit di Kabupaten Bojonegoro berdasarkan jumlah tenaga kesehatan pada tahun 2024 serta membandingkan performa dua metode analisis kluster, yaitu *Hierarchical Clustering* (Ward.D2) dan *K-Means Clustering*. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Satu Data Bojonegoro, mencakup 11 rumah sakit dengan 14 kategori profesi tenaga kesehatan. Setiap variabel memiliki skala dan rentang yang berbeda, sehingga dilakukan standarisasi menggunakan Z-score sebelum proses klusterisasi. *Hierarchical Clustering* menghasilkan tiga kluster yang berbeda, yaitu kluster rumah sakit berkapasitas SDM rendah hingga menengah, kapasitas menengah, dan kapasitas tinggi. Pemotongan dendrogram pada $k = 3$ didukung oleh struktur jarak antar kluster yang signifikan. Metode *K-Means* juga menunjukkan hasil optimal pada $k = 3$ berdasarkan *Elbow Method*. Pola pengelompokan keduanya menunjukkan konsistensi tinggi dengan mayoritas rumah sakit berada pada kluster SDM rendah hingga menengah, sementara RSUD Dr. R. Sosodoro Djatikoesoemo secara konsisten membentuk kluster tersendiri karena kapasitas tenaga kesehatan yang jauh lebih besar daripada rumah sakit lainnya. Evaluasi menggunakan *Silhouette Score* menunjukkan nilai identik antara kedua metode (0,3451), yang mengindikasikan kualitas kluster yang cukup baik dan struktur pemisahan yang stabil. Temuan penelitian ini dapat berkontribusi dalam penyusunan kebijakan pemerataan tenaga kesehatan, perencanaan kapasitas layanan, serta evaluasi kesiapan sumber daya manusia pada fasilitas kesehatan di tingkat daerah.

Kata kunci: klusterisasi, *Hierarchical Clustering*, *K-Means Clustering*, tenaga kesehatan, rumah sakit.

1 PENDAHULUAN

Rumah sakit adalah instansi kesehatan dan pusat kesehatan yang menampung dan merawat orang sakit serta menyediakan perawatan rawat inap (Hutagalung, 2022). Alokasi dana pemerintah dapat mendukung kebutuhan akan layanan kesehatan yang bermutu, sehingga rumah sakit perlu menjaga kinerja karyawan agar tujuannya tercapai dalam memberikan layanan kesehatan yang bermutu kepada masyarakat, di mana kinerja tenaga kesehatan sangat memengaruhi mutu layanan rumah sakit (Rachdiansyah & Tesmanto, 2021). Berdasarkan data yang dikumpulkan dari jumlah tenaga kesehatan di Kabupaten Bojonegoro, perawatan pasien di rumah sakit cukup baik, tetapi terdapat satu kendala. Kekurangan perawat menyebabkan tenaga kesehatan terkadang meminta bantuan keluarga pasien (Khayudin *et al.*, 2022).

Dua algoritma yang umum digunakan dalam analisis klusterisasi adalah *K-Means Clustering* dan *Hierarchical Clustering*. Algoritma K-Means dikenal memiliki kemampuan mengelola data berukuran besar dan menghasilkan pembagian kluster yang efisien, namun kelemahannya terletak pada keharusan menentukan jumlah kluster di awal serta sensitif terhadap pemilihan titik pusat kluster (Salsabila *et al.*, 2024). Sementara itu, algoritma *Hierarchical* tidak memerlukan definisi jumlah kluster di awal dan lebih intuitif dalam visualisasi hubungan antar data, tetapi memiliki keterbatasan dalam kapasitas komputasi saat diterapkan pada dataset yang besar (Februariyanti *et al.*, 2021). Oleh karena itu, studi yang membandingkan kedua algoritma ini dalam pengelolaan tenaga medis sangat diperlukan untuk menentukan pendekatan yang lebih efektif.

Beberapa penelitian sebelumnya telah membandingkan metode klusterisasi dalam konteks yang berbeda. Misalnya, penelitian oleh Nahya Nur *et al.* (2023). tentang pengelompokan daerah berisiko stunting di Indonesia menggunakan *K-Means* dan *Hierarchical Clustering* membuktikan *K-Means* lebih optimal, dengan nilai *Silhouette Coefficient* sebesar 0.48 dan *Calinski-Harabasz Index* sebesar 10.49, dibandingkan dengan *Hierarchical Clustering* yang mencatat nilai 0.47 dan 9.54. Penelitian ini menggambarkan efektivitas *K-Means* dalam mengelompokkan wilayah berdasarkan parameter tertentu. Namun, penelitian tersebut tidak secara spesifik membahas pengelompokan data tenaga medis, sehingga masih terdapat celah untuk penelitian lebih lanjut dalam penerapan kedua algoritma ini.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk membandingkan kinerja Kluster Hierarki dan *K-Means* dalam mengelompokkan tenaga kesehatan berdasarkan rumah sakit di Kabupaten Bojonegoro tahun 2024. Melalui analisis ini, diharapkan dapat diketahui algoritma yang paling efektif dalam mengelompokkan data tenaga kesehatan, serta memberikan kontribusi praktis dalam mendukung kebijakan pemerataan dan pengelolaan sumber daya manusia kesehatan di daerah.

2 METODE

2.1 Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari website Satu Data Bojonegoro yaitu data petugas kesehatan. Data observasi yang mencakup sebelas unit rumah sakit di Kabupaten Bojonegoro tercatat dalam database pada tahun 2024, yaitu RSUD Sumberrejo, RSUD Sosrogoro Djatikoesoemo, RSUD Padangan, RSM Sumberrejo, RSM Kalitidu, RSIA Fatma, RS Muna Anggita, RS Ibnu Sina, RS Bhayangkara Wahyu Tetuko, RSUD Kepohbaru dan RS Aisyiyah. Variabel-variabel penelitian yang digunakan dalam penelitian ini akan diberikan definisinya dan disajikan pada Tabel 1 sebagai berikut.

Tabel 1. Definisi Variabel Penelitian

No.	Variabel	Nama Variabel	Skala	Keterangan
1	X_1	Perawat	Rasio	Jumlah tenaga kesehatan untuk perawat
2	X_2	Penunjang	Rasio	Jumlah tenaga kesehatan untuk penunjang
3	X_3	Bidan	Rasio	Jumlah tenaga kesehatan untuk bidan
4	X_4	Dokter Spesialis	Rasio	Jumlah tenaga kesehatan untuk dokter spesialis
5	X_5	Tenaga Teknis Kefarmasian	Rasio	Jumlah tenaga kesehatan untuk tenaga teknis kefarmasian
6	X_6	Keteknisian	Rasio	Jumlah tenaga kesehatan untuk keteknisian
7	X_7	Dokter Umum	Rasio	Jumlah tenaga kesehatan untuk dokter umum

No.	Variabel	Nama Variabel	Skala	Keterangan
8	X_8	Apoteker	Rasio	Jumlah tenaga kesehatan untuk apoteker
9	X_9	Kesehatan Masyarakat	Rasio	Jumlah tenaga kesehatan untuk kesehatan masyarakat
10	X_{10}	Gizi	Rasio	Jumlah tenaga kesehatan untuk gizi
11	X_{11}	Kesehatan Lingkungan	Rasio	Jumlah tenaga kesehatan untuk kesehatan lingkungan
12	X_{12}	Keterampilan Fisik	Rasio	Jumlah tenaga kesehatan untuk keterampilan fisik
13	X_{13}	Dokter Gigi Spesialis	Rasio	Jumlah tenaga kesehatan untuk Dokter Gigi Spesialis
14	X_{14}	Dokter Gigi	Rasio	Jumlah tenaga kesehatan untuk Dokter Gigi

Adapun struktur data penelitian ditampilkan pada Tabel 2 berikut.

Tabel 2. Struktur Data Penelitian

Unit	Perawat	Penunjang	...	Dokter Gigi
RSM Kalitidu	$X_{1,1}$	$X_{1,2}$	...	$X_{1,14}$
RSM Sumberrejo	$X_{2,1}$	$X_{2,2}$	...	$X_{2,14}$
...	...	...	...	...
RS Bhayangkara	$X_{11,1}$	$X_{11,2}$		$X_{11,14}$

2.2 Standardisasi Data

Standardisasi data menggunakan metode *Z-score* penting dilakukan agar seluruh variabel berada pada skala yang sama sebelum proses klusterisasi, khususnya ketika variabel-variabel memiliki rentang nilai yang berbeda. Dengan skala yang seragam, variabel dengan nilai besar tidak akan mendominasi perhitungan jarak dan menyebabkan bias dalam hasil klaster. Sebuah studi menunjukkan bahwa ketika fitur-fitur dengan unit berbeda tidak distandarisasi sebelum penerapan *K-Means*, akurasi, presisi, recall dan *F-score* secara signifikan menurun dibandingkan ketika dilakukan standardisasi (Wongoutong, 2024). Berikut adalah persamaan yang digunakan untuk menghitung nilai standardisasi data menggunakan metode *Z-score*:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Keterangan:

Z = nilai hasil standardisasi

X = nilai asli (raw data)

μ = mean (rata-rata) variabel

σ = standar deviasi variabel

2.3 Analisis Klaster Hierarki

Hierarchical Clustering merupakan algoritma yang membangun struktur hierarki dalam proses pengelompokan data. Salah satu metode yang paling sering digunakan adalah *Agglomerative Hierarchical Clustering*, yaitu pendekatan yang menganggap setiap data sebagai

klaster terpisah pada awalnya, kemudian menggabungkannya secara bertahap hingga membentuk satu klaster besar (Satyahadewi *et al.*, 2023).

Pada proses penggabungan klaster, salah satu teknik yang umum diterapkan adalah *Ward's Linkage*, yang memiliki tujuan meminimalkan total varians dalam setiap klaster. Adapun rumus yang digunakan adalah sebagai berikut:

$$d(C_i, C_j) = \frac{|C_i| + |C_j|}{T} \|\mu_i - \mu_j\|^2 \quad (2)$$

Keterangan:

$d(C_i, C_j)$: jarak antara dua klaster

$|C_i|, |C_j|$: jumlah data dalam masing-masing klaster

T : jumlah total data

μ_i, μ_j : centroid klaster C_i dan C_j

Langkah-Langkah *Hierarchical Clustering*

1. Melakukan standarisasi data
2. Menghitung jarak antar rumah sakit menggunakan Euclidean Distance sebagai dasar pengukuran kemiripan objek.
3. Melakukan proses penggabungan klaster menggunakan metode *ward's Method*.
4. Menentukan jumlah klaster optimal ($k = 3$) berdasarkan titik pemotongan pada dendrogram.
5. Mengidentifikasi anggota klaster berdasarkan pemotongan dendrogram.
6. Melakukan profiling karakteristik tiap klaster untuk mengidentifikasi pola dan perbedaan antar kelompok.
7. Memvalidasi hasil klaster melalui visualisasi PCA guna memastikan konsistensi struktur pengelompokan.

2.4 Analisis K – Means

K-Means Clustering merupakan salah satu metode pengelompokan yang paling banyak digunakan dalam analisis data (Satrio *et al.*, 2024). Algoritma ini bekerja dengan membagi data ke dalam K klaster, di mana nilai K ditentukan terlebih dahulu sebagai jumlah kelompok yang diinginkan (Widyanti *et al.*, 2023). Proses pengelompokan diawali dengan pemilihan acak K titik pusat atau *centroid* pada ruang data. Selanjutnya, setiap data akan dialokasikan ke klaster yang memiliki *centroid* dengan jarak paling dekat (Hananto *et al.*, 2021).

Setelah data terkelompokkan, posisi centroid akan diperbarui berdasarkan nilai rata-rata dari seluruh data yang berada dalam klaster tersebut (Kastiawan *et al.*, 2024). Proses pembaruan ini diulang secara iteratif hingga tidak ada perubahan keanggotaan klaster atau hingga jumlah iterasi maksimum tercapai (Hendrastuty, 2024). Secara umum, langkah-langkah *K-Means* dapat dijabarkan sebagai berikut:

1. Melakukan standarisasi data
2. Menentukan jumlah klaster (k) yang optimal, menggunakan metode *Elbow Method* atau *Silhouette Score*.
3. Inisialisasi *centroid*, yaitu memilih titik pusat awal secara acak dari dataset.
4. Menghitung jarak antara tiap data dan centroid terdekat menggunakan *Euclidean Distance* dengan rumus:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Keterangan:

x dan y : dua titik data

x_i dan y_i : nilai pada dimensi ke- i

n : jumlah dimensi dataset

5. Mengelompokkan data ke kluster dengan centroid terdekat
6. Mengulang proses hingga posisi centroid stabil atau iterasi maksimum tercapai.

2.5 Silhouette Score

Silhouette Score digunakan untuk mengevaluasi seberapa baik suatu objek ditempatkan dalam kluster dengan membandingkan jarak intra-kluster $a(i)$ dan jarak inter-kluster terdekat $b(i)$. Konsep ini membantu menunjukkan apakah objek berada dalam kluster yang tepat atau justru lebih dekat dengan kluster lain. Menurut Astuti & Muqorobin (2021), *Silhouette Score* merupakan salah satu metrik yang efektif dalam mengukur kualitas pemisahan kluster karena mempertimbangkan dua aspek utama, yaitu kedekatan objek dengan klusternya sendiri dan jaraknya terhadap kluster terdekat. Nilai *Silhouette Score* berada pada rentang -1 hingga 1, dengan rumus sebagai berikut:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

Keterangan:

$S(i)$: skor *silhouette* untuk data ke- i

$a(i)$: rata-rata jarak antara data ke- i dan semua titik dalam klusternya

$b(i)$: rata-rata jarak antara data ke- i dan semua titik pada kluster terdekat

3 HASIL DAN PEMBAHASAN

3.1 Gambar Umum Data

Data penelitian ini terdiri dari jumlah tenaga kesehatan pada berbagai kategori profesi pada rumah sakit di Kabupaten Bojonegoro tahun 2024.

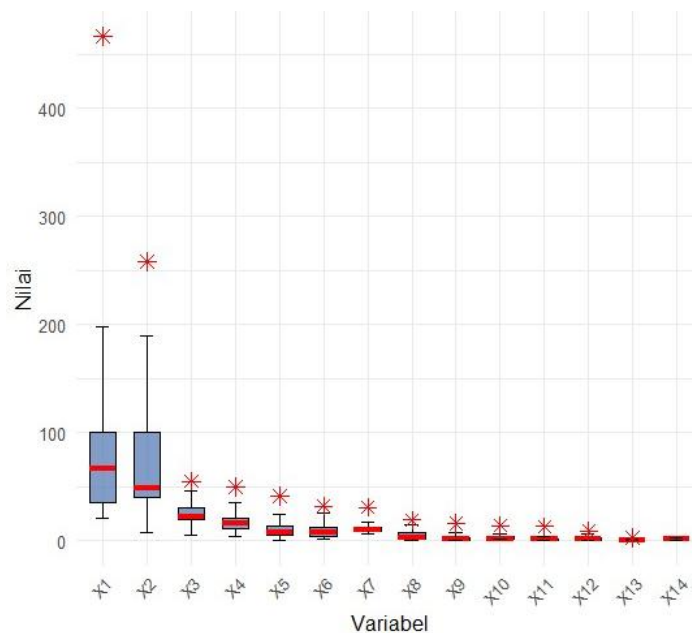
Tabel 3. Statistika Deskriptif

Variabel	Count	Mean	sd	Min	Q1	Q2	Q3	Max
X_1	11	103	127.31	21	35.5	66	100.5	467
X_2	11	80	74.35	8	40	49	100.0	258
X_3	11	26	14.96	5	20	22	30.5	55
X_4	11	19	13.49	4	11	16	20.5	50
X_5	11	11	11.57	0	5	7	13.0	41
X_6	11	10	10.09	1	4	7	12.5	32
X_7	11	13	8.40	6	9	10	12.5	31
X_8	11	5	5.37	0	2	3	7	19
X_9	11	3	4.93	0	0	2	3	16
X_{10}	11	4	3.85	1	1.5	2	3.5	14
X_{11}	11	2	3.81	0	0	1	1.5	13

Variabel	Count	Mean	sd	Min	Q1	Q2	Q3	Max
X_{12}	11	2	3.09	0	0	1	2.5	9
X_{13}	11	1	1.04	0	0	0	0.5	3
X_{14}	11	2	1.04	0	1.0	2	2.0	3

Berdasarkan Tabel 1, secara umum jumlah tenaga kesehatan antar rumah sakit menunjukkan variasi yang cukup tinggi, terlihat dari rentang nilai minimum hingga maksimum pada hampir semua kategori profesi. Variabel Perawat merupakan kategori dengan rata-rata tertinggi (103,27) dan rentang yang paling lebar (21–467), menandakan adanya perbedaan kapasitas pelayanan yang cukup mencolok antar rumah sakit. Variabel Penunjang serta Bidan juga menunjukkan sebaran yang relatif besar. Sebaliknya, kategori seperti Dokter Gigi Spesialis, Keterampilan Fisik, dan Kesehatan Lingkungan memiliki rata-rata yang rendah dan sebaran yang sempit, menggambarkan keterbatasan jumlah tenaga kesehatan pada kategori tersebut.

Informasi ini menunjukkan bahwa distribusi tenaga kesehatan antar rumah sakit tidak merata. Dengan adanya variasi skala dan rentang nilai antar variabel, proses standardisasi diperlukan sebelum analisis kluster dilakukan agar tidak terjadi dominasi variabel tertentu dalam proses pengelompokan.



Gambar 1. Boxplot Setiap Variabel Indikator Tenaga Kesehatan

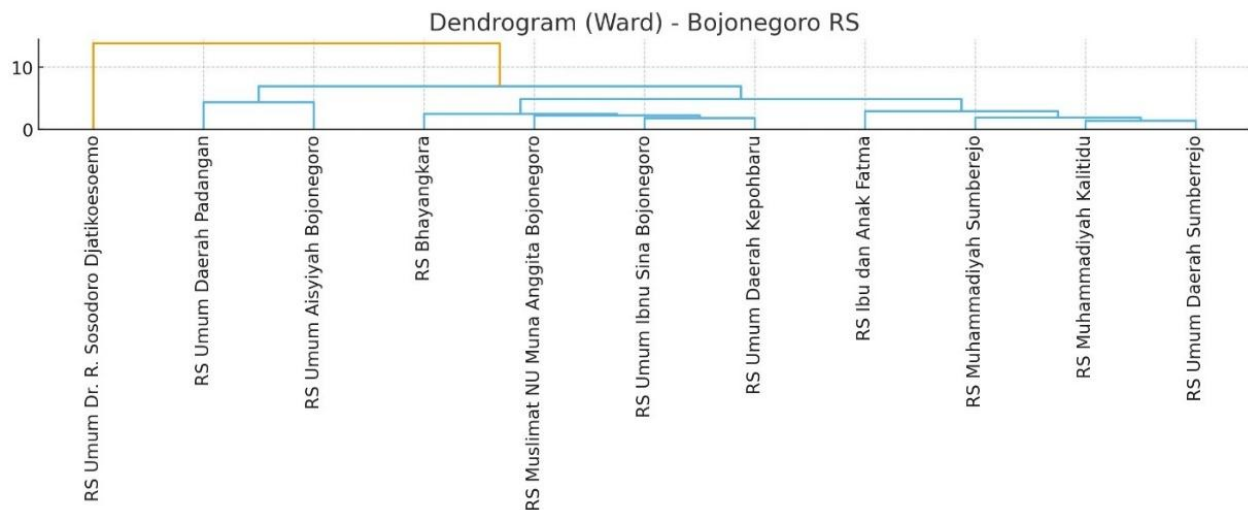
Pada *BoxPlot* diatas untuk ke-14 variabel (X_1 hingga X_{14}) menunjukkan adanya heterogenitas skala yang signifikan dalam kumpulan data. Kelompok pertama terdiri dari variabel X_1 dan X_2 yang menunjukkan variabilitas tinggi. Hal ini terlihat dari IQR yang jauh lebih lebar dibandingkan variabel lainnya, menandakan bahwa 50% data tersebar dalam rentang yang cukup luas. Khususnya untuk X_1 , rentang data dan IQR-nya sangat besar, dengan kuartil atas (Q_3) mencapai 100,5 diikuti X_2 dengan Q_3 sebesar 100. Sebaliknya, kelompok kedua yang mencakup variabel X_3 hingga X_{14} menunjukkan variabilitas rendah. Data pada kelompok ini cenderung terkompresi dengan IQR yang sempit, mengindikasikan bahwa sebagian besar nilai observasi terkonsentrasi di sekitar median dan memiliki penyebaran yang kecil. Variabel X_{13} dan X_{14} bahkan menunjukkan

tingkat dispersi paling rendah, di mana nilai kuartil bawah (Q_1) dan median (Q_2) berada sangat dekat dengan titik nol, menggambarkan distribusi data yang sangat rapat dan terkompresi.

3.2 Hasil Klaster Hierarki

3.2.1 Penentuan Jumlah Klaster

Dendrogram menunjukkan struktur hierarki pengelompokan berbasis jarak *Euclidean* dan metode penggabungan Ward.D2. Titik pemotongan pada $k = 3$ dipilih berdasarkan jarak vertikal terpanjang sebelum terjadi penggabungan cabang, yang menandakan perbedaan antar kelompok terbesar. Dengan demikian, dataset tenaga kesehatan secara alami terbagi menjadi tiga kelompok rumah sakit melalui pendekatan hierarki.



Gambar 2. Dendrogram Klaster Hierarki (Ward.D2)

3.2.2 Hasil Pengelompokan

Berdasarkan pemotongan dendrogram pada tiga klaster, diperoleh hasil pengelompokan berikut:

Tabel 4. Hasil Klaster Hierarki per Rumah Sakit

Klaster	Anggota Klaster
1	RSM Kalitidu, RSM Sumberejo, RS Ibnu Sina RSUD Sumberrejo, RSUD Kepohbaru, RS Muna Anggita, RSIA Fatma dan RS Bhayangkara
2	RSUD Padangan dan RSU Aisyiyah
3	RSUD Sosodoro Djatikoesoemo

Tabel 5. Profiling Hasil Klaster Metode *Hierarchical Clustering*

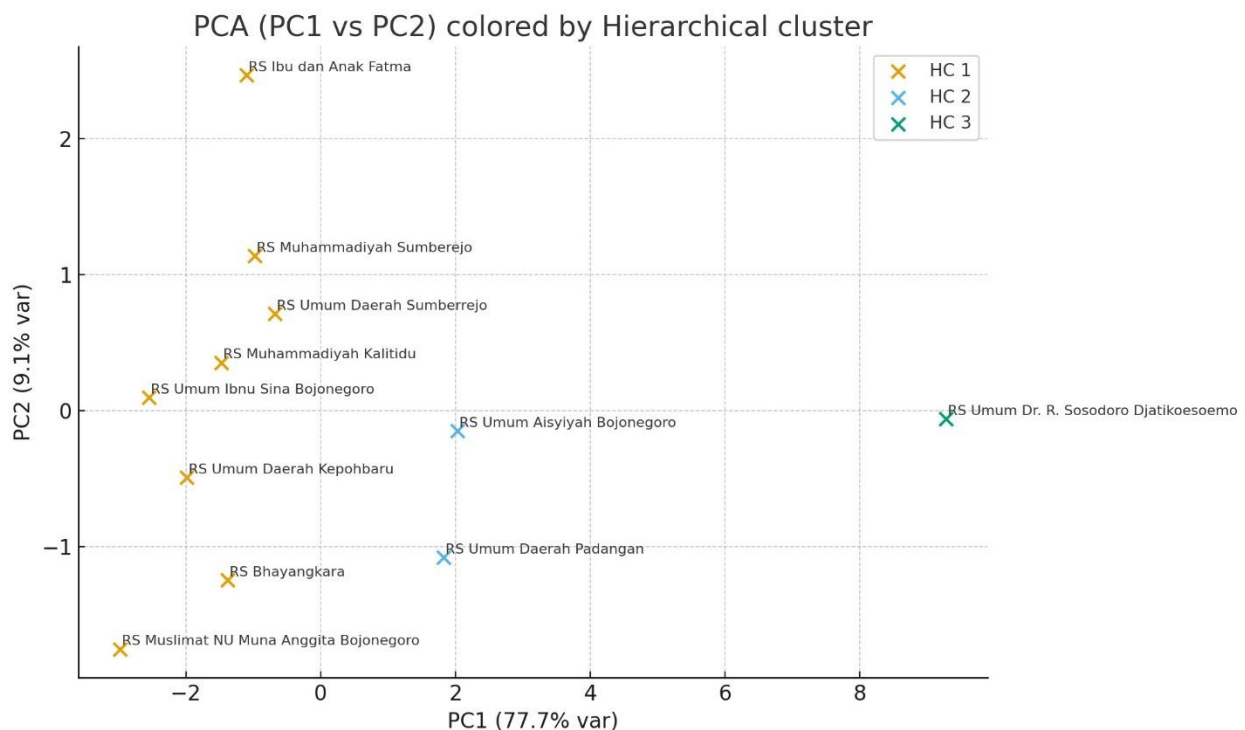
Variabel	Rata - rata		
	Klaster 1 (Rendah)	Klaster 2 (Menengah)	Klaster 3 (Tinggi)
X_1	49	139	467
X_2	43	142	258
X_3	23	23	55
X_4	13	28	50
X_5	5	19	41

X_6	6	20	32
X_7	10	19	31
X_8	3	8	19
X_9	1	5	16
X_{10}	2	5	14
X_{11}	1	2	13
X_{12}	1	5	9
X_{13}	0	1	3
X_{14}	1	2	2
Rata – rata Total	11	30	72

Klaster 3 (Tinggi) ini merupakan rumah sakit dengan rata-rata jumlah tenaga kesehatan tertinggi hampir di semua kategori. Klaster ini diisi oleh Rumah Sakit Umum Daerah (RSUD) Tipe B dan Tipe A yang berfungsi sebagai rumah sakit rujukan tertinggi di wilayah tersebut. Selain itu, RSUD Tipe B atau A yang merupakan milik Pemerintah (Daerah/Pusat) cenderung memiliki alokasi anggaran dan kewajiban standar pelayanan yang mengharuskan kelengkapan jumlah dan jenis Nakes yang tinggi, sesuai dengan profil rata-rata yang sangat superior.

3.2.3 Visualisasi PCA Berdasarkan Hasil *Hierarchical Clustering* (Ward.D2)

Untuk memahami sebaran objek dalam ruang berdimensi rendah, *Principal Component Analysis* (PCA) digunakan untuk mereduksi 14 variabel SDM menjadi dua komponen utama. Dua komponen ini menjelaskan sebagian besar variasi data (PC1 = 77,7%; PC2 = 9,1%). Visualisasi PCA berdasarkan hasil *Hierarchical Clustering* ditampilkan pada Gambar 3



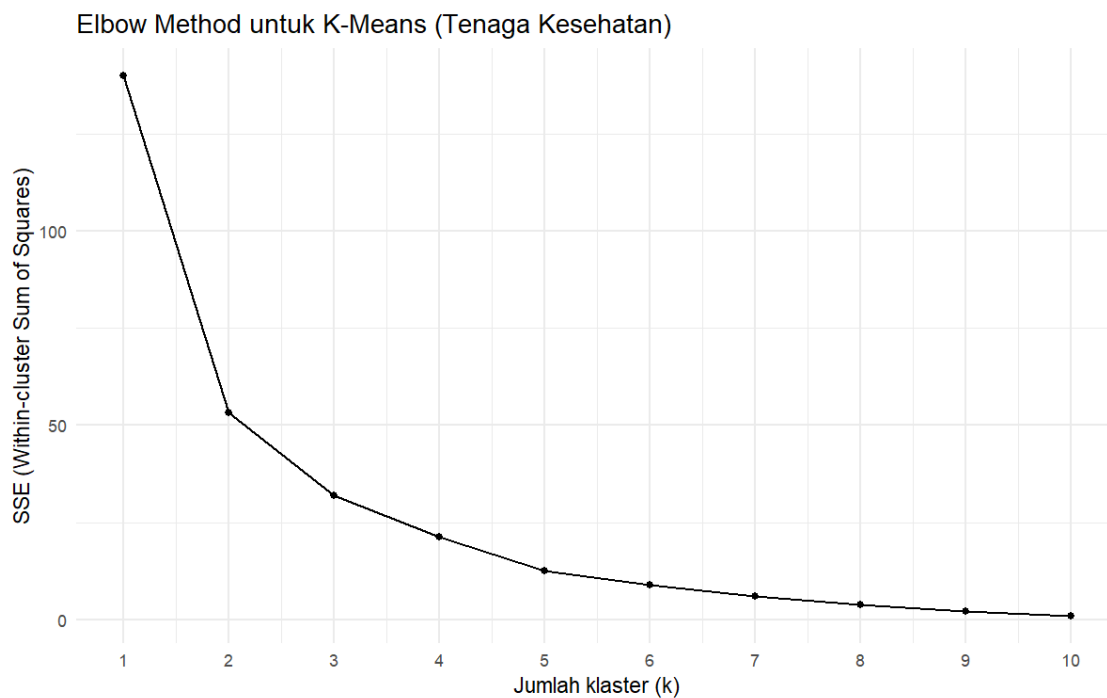
Gambar 3. Visualisasi PCA menurut metode Hierarki (Ward).

Klaster 1 berisi mayoritas rumah sakit dengan SDM kecil–menengah dan tampak bergerombol rapat di sisi kiri. Klaster 2 menempati area menengah menuju kanan atas, mencerminkan RS dengan kapasitas sedang, misalnya RSUD Padang dan RS Aisyiyah Bojonegoro. Klaster 3 kembali menempatkan RSUD Dr. R. Sosodoro Djatikoesoemo sebagai titik terpisah jelas, menegaskan bahwa rumah sakit ini memiliki profil SDM yang unik dan dominan pada sebagian besar variabel. Konsistensi antara PCA-KMeans dan PCA-Hierarki menunjukkan bahwa struktur klaster sangat stabil.

3.3 Hasil K-Means Clustering

3.3.1 Penentuan Nilai K

Penentuan jumlah klaster optimal dilakukan dengan metode elbow, yang memplot nilai *within-cluster sum of squares* (WCSS) terhadap jumlah klaster K.



Gambar 4. *Elbow Method*

Dari plot terlihat bahwa grafik menunjukkan bahwa seiring bertambahnya jumlah klaster, nilai WCSS terus menurun, namun dengan tingkat penurunan yang semakin kecil. Titik siku (*elbow point*) dalam grafik ini terlihat pada $k=3$, di mana setelah titik tersebut, penurunan WCSS mulai melambat secara signifikan. Oleh karena itu, jumlah klaster optimal yang dipilih adalah $k=3$, karena pada titik ini terjadi keseimbangan antara jumlah klaster yang cukup untuk menangkap variasi data tanpa menyebabkan pembentukan klaster yang terlalu kecil atau kurang bermakna.

3.3.2 Hasil K-Means

Berdasarkan plot pada grafik *Elbow Method* mendapatkan tiga klaster, diperoleh hasil pengelompokan sebagai berikut:

Tabel 6. Hasil Klater *K-Means* Per Rumah Sakit

Klaster	Anggota Klaster
1	RSM Kalitidu, RSM Sumberejo, RS Ibnu Sina RSUD Sumberrejo, RSUD Kepohbaru, RS Muna Anggita, RSIA Fatma dan RS Bhayangkara
2	RSUD Sosodoro Djatikoesoemo
3	RSUD Padangan dan RSU Aisyiyah

Hasil *K-Means* menunjukkan pola pengelompokan yang mirip dengan metode hierarki, meskipun terdapat perbedaan kecil pada RSUD Padangan dan RSU Aisyiyah yang dialokasikan ke Klaster 3. Klaster 1 berisi 8 rumah sakit dengan kapasitas tenaga kesehatan rendah–menengah, Klaster 2 hanya berisi RSUD Dr. R. Sosodoro Djatikoesoemo, dan Klaster 3 berisi dua rumah sakit berkapasitas SDM menengah–tinggi. Konsistensi mayoritas anggota antar dua metode klaster menunjukkan kestabilan struktur data.

3.3.3 Hasil *Centroid K-Means*

Centroid klaster dalam skala asli menggambarkan karakteristik tenaga kesehatan rata-rata di tiap klaster.

Tabel 7. *Centroid K-Means*

Variabel	Rata - rata		
	Klaster 1 (Rendah)	Klaster 2 (Tinggi)	Klaster 3 (Menengah)
X_1	49	467	139
X_2	43	258	142
X_3	23	55	23
X_4	13	50	28
X_5	5	41	19
X_6	6	32	20
X_7	10	31	19
X_8	3	19	8
X_9	1	16	5
X_{10}	2	14	5
X_{11}	1	13	2
X_{12}	1	9	5
X_{13}	0	3	1
X_{14}	1	2	2
Rata – rata Total	11	72	30

Hasil klasterisasi *K-Means* menghasilkan tiga klaster utama yang menunjukkan perbedaan kapasitas sumber daya manusia kesehatan di rumah sakit Kabupaten Bojonegoro. Centroid masing-masing klaster dalam skala asli menggambarkan karakteristik tenaga kesehatan rata-rata yang menjadi profil khas tiap kelompok.

Klaster 1 menunjukkan jumlah tenaga kesehatan yang relatif rendah hingga menengah pada seluruh kategori profesi. Nilai rata-rata perawat sebesar (49), tenaga penunjang (43), dan dokter spesialis (13) mengindikasikan keterbatasan kapasitas SDM, terutama pada tenaga medis spesialis dan tenaga kefarmasian. Kondisi ini sesuai dengan karakteristik Rumah Sakit Swasta Tipe D atau Rumah Sakit Tipe C yang baru berdiri atau dengan layanan terbatas.

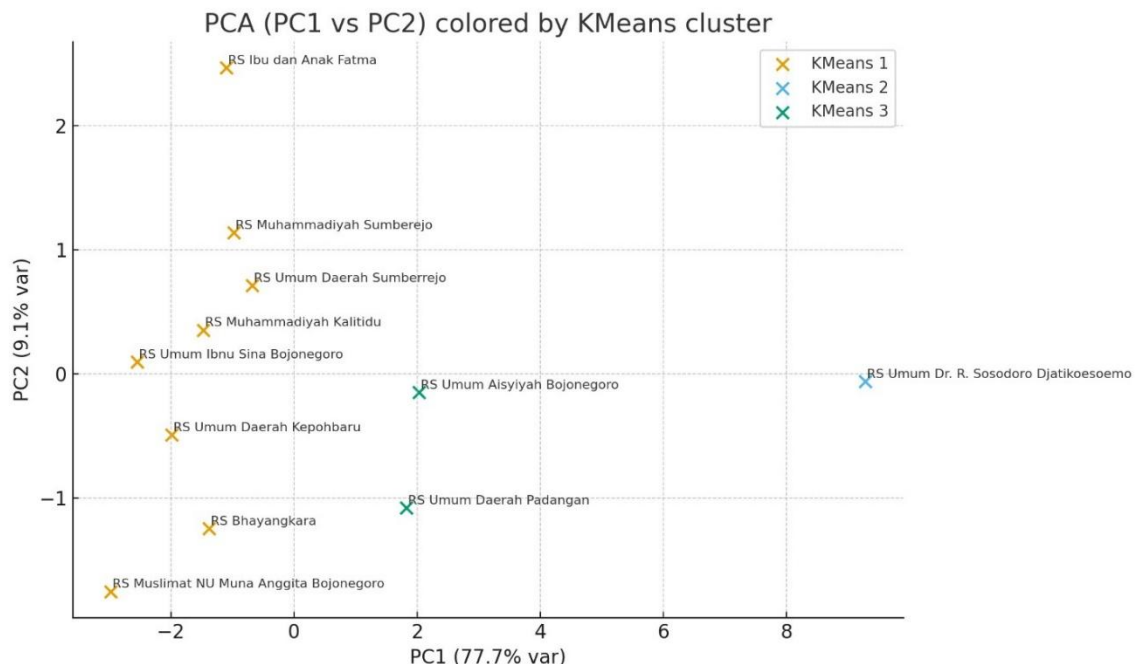
Klaster 3 menunjukkan kapasitas SDM menengah, dengan jumlah tenaga kesehatan yang lebih besar dibandingkan klaster 1 namun belum mencapai kelengkapan klaster 2. Rata-rata perawat (139), tenaga penunjang (142), dan dokter spesialis (28) menggambarkan rumah sakit yang telah memiliki layanan spesialis dasar dan struktur SDM yang lebih beragam. Kelompok ini dapat merepresentasikan Rumah Sakit Tipe C yang sudah berkembang atau beberapa Rumah Sakit Swasta Tipe B. Rumah sakit dalam klaster ini memiliki jumlah Perawat dan Penunjang yang mencukupi, tetapi jumlah Dokter Spesialisnya (28) tetap jauh di bawah Klaster 2 (50), mengindikasikan bahwa mereka memiliki layanan spesialis dasar tetapi belum mendukung semua sub-spesialisasi, menempatkannya sebagai rumah sakit kategori menengah.

Sementara itu, klaster 2 merupakan klaster dengan kapasitas SDM tertinggi, ditandai dengan nilai ekstrem pada hampir seluruh kategori tenaga kesehatan. Rata-rata perawat sebesar (467), tenaga penunjang (258), dan dokter spesialis (50) menggambarkan rumah sakit besar dengan kebutuhan layanan yang kompleks. Klaster ini didominasi oleh RSUD Dr. R. Sosodoro Djatikoesoemo yang berfungsi sebagai rumah sakit rujukan utama di Kabupaten Bojonegoro. Profil SDM yang tinggi pada klaster ini selaras dengan karakteristik rumah sakit tipe B atau A yang memiliki tuntutan pelayanan komprehensif.

Perbedaan kapasitas SDM ini juga tercermin pada rata-rata total SDM per klaster, yaitu klaster 1 sebesar (11), klaster 3 sebesar (30), dan klaster 2 mencapai (72). Nilai tersebut menegaskan adanya kesenjangan kapasitas tenaga kesehatan antar rumah sakit, sehingga hasil klaster dapat menjadi dasar perencanaan distribusi SDM yang lebih proporsional

3.3.4 Visualisasi Klaster K-Means

Untuk memahami sebaran objek dalam ruang berdimensi rendah, *Principal Component Analysis* (PCA) digunakan untuk mereduksi 14 variabel SDM menjadi dua komponen utama. Dua komponen ini menjelaskan sebagian besar variasi data ($PC1 = 77,7\%$; $PC2 = 9,1\%$). Visualisasi PCA berdasarkan hasil *K-Means* ditampilkan pada Gambar 5.



Gambar 5. Visualisasi PCA menurut *K-Means*.

Pada plot PCA, klaster 1 terletak pada area dengan skor PC1 rendah dan tampak mengelompok rapat, sesuai dengan kapasitas SDM yang seragam dan terbatas. Klaster 3 berada pada skor PC1 menengah, menggambarkan kapasitas SDM yang lebih besar dibanding klaster 1 tetapi belum mencapai klaster 2. Klaster 2 muncul sebagai titik yang jauh terpisah dari dua klaster lainnya, sejalan dengan nilai SDM yang sangat tinggi pada centroid, sehingga rumah sakit ini menjadi penentu utama arah PC1 dan berperan sebagai outlier.

3.4 Perbandingan Klaster: Hierarki vs K-Means

3.4.1 Nilai *Silhouette* Score

Silhouette score digunakan untuk menilai kualitas klaster. Nilai mendekati 1 menunjukkan klaster yang baik, sedangkan nilai mendekati 0 menunjukkan adanya tumpang tindih antar klaster.

Tabel 8. Nilai *Silhouette*

Metode	Rata-rata <i>Silhouette</i>
K-Means	0.3451
Hierarchical (Ward)	0.3451

Nilai kedua metode identik (0,3451), mengindikasikan bahwa kualitas klaster berada pada kategori cukup, dan struktur klaster relatif konsisten di antara dua metode.

4 KESIMPULAN

4.1 Kesimpulan

Berdasarkan hasil analisis klaster terhadap data sumber daya manusia (SDM) rumah sakit di Kabupaten Bojonegoro menggunakan dua pendekatan, yaitu *Hierarchical Clustering* (Ward.D2) dan *K-Means Clustering*, beberapa kesimpulan penting dapat diambil. Pertama, model klaster terbaik membagi rumah sakit menjadi tiga klaster ($k = 3$). Keputusan ini didukung oleh koherensi struktur dendrogram, kestabilan visual pada PCA, serta nilai average *silhouette* yang identik antara kedua metode (0,3451), yang menunjukkan kualitas pengelompokan yang cukup baik untuk data dengan variasi SDM yang kontras.

Kedua metode klaster menunjukkan konsistensi pola pengelompokan. *K-Means* dan *Hierarchical Clustering* (Ward.D2) menghasilkan struktur klaster yang sangat mirip, terutama dalam mengidentifikasi rumah sakit dengan karakteristik SDM yang homogen dan skala kecil–menengah sebagai satu kelompok besar. Ketiga, RSUD Dr. R. Sosodoro Djatikoesoemo membentuk klaster tersendiri karena memiliki jumlah SDM yang jauh lebih besar dibandingkan rumah sakit lain. Hal ini didukung oleh konsistensi pemisahan yang terlihat pada dendrogram, PCA, dan centroid, mempertegas karakter RSUD Sosodoro sebagai rumah sakit rujukan dengan beban layanan tinggi.

Klaster rumah sakit kecil–menengah menunjukkan homogenitas tinggi. Rumah sakit dalam klaster ini memiliki jumlah tenaga kesehatan yang relatif seragam, sehingga mudah teridentifikasi pada PCA sebagai kelompok rapat dengan variasi rendah. Kelima, penerapan PCA terbukti efektif untuk mendukung interpretasi klaster. Dua komponen utama, yaitu PC1 yang memuat 77,7% varians dan PC2 yang memuat 9,1% varians, mampu menggambarkan struktur data secara representatif. Plot PCA memperjelas hubungan antar rumah sakit dan memperkuat bukti pemisahan klaster berbasis angka centroid.

Secara keseluruhan, analisis ini memberikan gambaran komprehensif mengenai pengelompokan rumah sakit berdasarkan SDM, serta menyediakan dasar bagi pengambilan kebijakan dalam pemerataan tenaga kesehatan di daerah.

4.2 Saran

Bagi Pemerintah Daerah atau Dinas Kesehatan, hasil analisis klaster dapat dimanfaatkan sebagai dasar dalam penyusunan kebijakan distribusi tenaga kesehatan, khususnya bagi rumah sakit yang termasuk dalam kelompok dengan sumber daya manusia (SDM) rendah. Selain itu, rumah sakit yang berada dalam klaster kecil hingga menengah sebaiknya menjadi prioritas dalam pelaksanaan program peningkatan kapasitas SDM agar kesenjangan pelayanan kesehatan dapat semakin diperkecil.

Untuk pihak rumah sakit, kolaborasi antar rumah sakit dalam satu klaster sangat disarankan, terutama dalam hal perencanaan SDM, pelatihan tenaga kesehatan, serta mekanisme rujukan layanan. Rumah sakit dengan jumlah SDM yang relatif kecil juga perlu melakukan analisis lebih mendalam terhadap kebutuhan tenaga kesehatan yang bersifat kritis, dengan mempertimbangkan karakteristik klaster tempat mereka tergabung.

Bagi peneliti selanjutnya, penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan variabel lain seperti jumlah tempat tidur, tingkat BOR (Bed Occupancy Rate), cakupan layanan, maupun indikator mutu pelayanan, sehingga hasil klaster menjadi lebih multidimensional. Selain itu, penerapan metode klaster alternatif seperti DBSCAN, GMM, atau spectral clustering juga dapat dilakukan untuk membandingkan kinerja antar algoritma. Validasi eksternal, seperti cluster stability dan *cophenetic correlation*, juga direkomendasikan agar model yang dihasilkan memiliki keandalan yang lebih tinggi.

DAFTAR PUSTAKA

- A. Lia Hananto et al., "Analysis Of Drug Data Mining With Clustering Technique Using KMeans Algorithm," J Phys Conf Ser, vol. 1908, no. 1, pp. 1–8, Jun. 2021, doi: 10.1088/1742-6596/1908/1/012024.
- Astuti, D., & Muqorobin. (2021). Optimasi Metode K-Means Clustering untuk Pengelompokan Obat di Puskesmas Mertoyudan I Magelang. *Sistemasi: Jurnal Sistem Informasi*, 13(5), 2144–2160. <http://sistemasi.ftik.unisi.ac.id>
- Khayudin, B. A, dkk. (2022). *Pengalaman Perawat dalam Merawat Pasien Total Care di Ruang ICU RSUD DR R. Sosodoro Djatikoesoemo Bojonegoro*.
- Februariyanti, H., Wibowo, J. S., Santoso, D. B., & Sukur, M. (2021). Analisis Kecenderungan Informasi Menggunakan Algoritma Hierarchical Agglomerative Clustering. *Jurnal Informatika, Manajemen dan Komputer*, 13(1), 9–17.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Hutagalung, L. E. (2022). Analisa manajemen risiko Sistem Informasi Manajemen Rumah Sakit (SIMRS) pada Rumah Sakit XYZ menggunakan ISO 31000. *Jurnal TeIKA*, 12(1).
- Indra, Nahya Nur, M. Iqram, & Nurul Inayah. (2023). Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting. *Jurnal Inovtek Polbeng – Seri Informatika*, 8(2), 356–367.
- Rachdiansyah, I., & Tesmanto, J. (2021). Pengaruh Audit Manajemen Sumber Daya Manusia terhadap Kinerja Karyawan di Rumah Sakit Umum Daerah Kota Bekasi. *Aliansi: Jurnal Manajemen dan Bisnis*, 16(2), 77–84.
- Salsabila, F., Azise, N., & Ridla, M. A. (2024). Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids pada Popularitas Line Webtoon. *Prosiding Seminar Nasional Sains*

dan Teknologi Seri 02, 1(2), 3047–6569.

- Satyahadewi, N., Sinaga, S. J., & Perdana, H. (2023). Hierarchical Cluster Analysis of Districts/Cities in North Sumatra Province Based on Human Development Index Indicators Using Pseudo-F. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, 17(3), 1429–1438. <https://doi.org/10.30598/barekengvol17iss3pp1429-1438>
- Satrio, D. D., Akbar, F. A., & Al Haromainy, M. M. (2024). Pengembangan Bot Discord Sebagai Pemutar dan Rekomendasi Musik Menggunakan Metode K-Means. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 13(1), 95–106. <https://doi.org/10.35889/jutisi.v13i1.1681>
- Widyanti, T., Hilabi, S. S., Hananto, A., Tukino, & Novalia, E. (2023). Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi. *Jurnal Informasi & Teknologi (JIdT)*, 5(1), 75–82.
- Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLoS ONE*, 19(12), e0310839. <https://doi.org/10.1371/journal.pone.0310839>