



Clustering Dataset Numerik Multivariat menggunakan Algoritma K-Means dengan Evaluasi Silhouette dan Davies–Bouldin Index

Mei Yanda Maharani Messakh^{1*}, Jusuf Ezra Rudolf Fransz², David Hermanus³,
Adriana Fanggalda⁴, Yulianto Triwahyuadi Polly⁵

¹⁻⁵ Program Studi Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana, Indonesia

*Penulis Korespondensi: meiyandamessakh@gmail.com

Abstract. Clustering is a data mining technique used to group data based on similarity. This study applies the K-Means clustering algorithm to analyze multivariate numerical data from the Standard Minimum Service (SPM) dataset of Nusa Tenggara Timur Province. The dataset is structured and consists of six numerical attributes representing basic service sectors, namely education, health, public works, housing and settlement, public order and community protection, and social services. The research stages include data cleaning, data standardization using the Z-score method, determination of the optimal number of clusters, and clustering using the K-Means algorithm. The quality of the clustering results is evaluated using the Silhouette Score and Davies–Bouldin Index. The results show that the optimal number of clusters is three ($k = 3$), with a Silhouette Score of 0.327 and a Davies–Bouldin Index of 0.872. The clustering process produces three data groups with different characteristics based on the average values of each sector. These results indicate that the K-Means algorithm can be used to group SPM data into meaningful clusters based on similarity patterns.

Keywords: Clustering; Data Analysis; Data Grouping; Data Mining; K-Means

Abstrak. Clustering merupakan teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan. Penelitian ini menerapkan algoritma K-Means untuk menganalisis data numerik multivariat pada dataset capaian Standar Pelayanan Minimal (SPM) Provinsi Nusa Tenggara Timur. Dataset yang digunakan bersifat terstruktur dan terdiri dari enam atribut numerik yang merepresentasikan sektor layanan dasar, yaitu pendidikan, kesehatan, pekerjaan umum, perumahan dan kawasan permukiman, ketenteraman dan ketertiban umum serta perlindungan masyarakat, dan sosial. Tahapan penelitian meliputi pembersihan data, standardisasi data menggunakan metode Z-score, penentuan jumlah kluster optimal, serta proses clustering menggunakan algoritma K-Means. Kualitas hasil clustering dievaluasi menggunakan Silhouette Score dan Davies–Bouldin Index. Hasil penelitian menunjukkan bahwa jumlah kluster optimal adalah tiga ($K = 3$), dengan nilai Silhouette Score sebesar 0,327 dan Davies–Bouldin Index sebesar 0,872. Proses clustering menghasilkan tiga kelompok data dengan karakteristik yang berbeda berdasarkan nilai rata-rata setiap sektor. Hasil ini menunjukkan bahwa algoritma K-Means dapat digunakan untuk mengelompokkan data capaian SPM ke dalam kluster yang bermakna berdasarkan pola kemiripan data.

Kata Kunci: Analisis Data; Clustering; Data Mining; K-Means; Pengelompokan Data

1. LATAR BELAKANG

Ketersediaan data numerik dalam jumlah besar pada berbagai sektor mendorong kebutuhan analisis berbasis komputasi untuk mengekstraksi pola yang tidak dapat dijelaskan melalui statistik deskriptif semata. Dalam Ilmu Komputer, kondisi ini mendorong kebutuhan *data-driven analysis* yang tidak hanya menampilkan statistik deskriptif, tetapi juga mengekstraksi pola dari data numerik multivariat agar informasi yang dihasilkan lebih mudah dipahami dan ditindaklanjuti. Salah satu pendekatan yang sering digunakan untuk eksplorasi pola pada data tanpa label adalah *unsupervised learning*, khususnya teknik pengelompokan (*clustering*), karena mampu membentuk kelompok berdasarkan kemiripan karakteristik data secara otomatis (Dewi et al., 2022; Simanjuntak et al., 2023)

Pengelompokkan banyak diterapkan pada berbagai permasalahan informatika berbasis data, misalnya pengelompokkan data bencana (Isni Rinjani et al., 2023; Murdiaty et al., 2020), pengelompokkan data pendidikan untuk identifikasi kategori siswa (Handayani, 2022; Ramadani & Fatah, 2024; Saputra & Nataliani, 2021), pengelompokkan data kesehatan dan wilayah untuk membantu pemetaan prioritas (Hutagalung & Sriani, 2024; Sembiring, 2022), hingga pengelompokkan dokumen/berita berbahasa Indonesia dengan pendekatan komputasi berbasis fitur (Simanjuntak et al., 2023). Studi-studi tersebut menunjukkan bahwa pengelompokkan relevan sebagai cara komputasional untuk merangkum dataset besar/berdimensi banyak menjadi beberapa kelompok yang lebih mudah dianalisis.

Dalam penelitian ini, data capaian Standar Pelayanan Minimal (SPM) diposisikan sebagai dataset numerik multivariat: setiap kabupaten/kota dipandang sebagai satu objek data, sedangkan nilai capaian pada beberapa sektor dipandang sebagai atribut/fitur. Dari sudut pandang informatika, karakteristik ini cocok untuk pengelompokkan karena data bersifat numerik, memiliki banyak atribut, dan tidak memiliki label kelas awal. Namun, pada praktiknya, dataset multivariat seperti ini sering masih dipakai sebatas rekap dan tabel, sehingga pola kemiripan antar-objek belum dieksplorasi secara komputasional menggunakan algoritma pengelompokkan (misalnya K-Means) dan metrik evaluasi kualitas *cluster* yang terukur (Fadlil et al., 2022; Margaretha et al., 2025).

Algoritma K-Means dipilih karena konsepnya sederhana, efisien secara komputasi, dan umum digunakan untuk data numerik berdimensi banyak. K-Means bekerja dengan membentuk pusat *cluster* (*centroid*) dan menempatkan setiap data pada cluster dengan jarak terdekat; karena berbasis jarak, prapemrosesan seperti pembersihan dan penyamaan skala nilai menjadi tahap penting agar hasil pengelompokkan tidak bias pada atribut tertentu (Fadlil, 2022; Al Rivan & Sonaru, 2022). Selain itu, pemilihan jumlah *cluster* (k) dan evaluasi kualitas hasil pengelompokkan juga krusial agar kelompok yang terbentuk benar-benar merepresentasikan struktur data, misalnya menggunakan *Silhouette* dan/atau *Davies–Bouldin Index* yang telah banyak dipakai pada studi-studi pengelompokkan di jurnal informatika Indonesia (Dewi et al., 2022; Margaretha et al., 2025).

Berdasarkan gap tersebut, penelitian ini bertujuan menerapkan algoritma K-Means untuk melakukan pengelompokkan kabupaten/kota di Provinsi Nusa Tenggara Timur berdasarkan data capaian SPM (sebagai studi kasus dataset multivariat), melalui tahapan pembersihan data, standardisasi/normalisasi, proses pengelompokkan, dan evaluasi kualitas *cluster* menggunakan metrik yang sesuai.

Luaran yang diharapkan adalah bukti penerapan alur kerja analitik berbasis Ilmu Komputer (*data preprocessing* → *clustering* → *evaluation*) pada dataset terstruktur, sehingga hasilnya dapat dibaca sebagai ringkasan pola data dan dasar eksplorasi lanjutan berbasis komputasi (Darmawan & Karmilasari, 2024; Simanjuntak et al., 2023; Ramdani et al., 2025).

2. KAJIAN TEORITIS

Data Mining dan Analisis Data

Data mining merupakan proses analisis data yang bertujuan untuk mengekstraksi pola, hubungan, dan informasi tersembunyi dari kumpulan data berukuran besar dan kompleks (Siahaan, 2022). Dalam Ilmu Komputer, *data mining* dipahami sebagai pendekatan komputasional yang mengintegrasikan teknik statistik, kecerdasan buatan, dan basis data untuk menganalisis data numerik maupun multivariat. Proses *data mining* umumnya mencakup tahap seleksi data, pembersihan data, transformasi data, penerapan algoritma analisis, serta evaluasi hasil, sebagaimana diterapkan dalam berbagai studi *clustering* berbasis *data mining* (Prasetyo et al., 2024). Pendekatan ini banyak digunakan untuk memahami struktur data secara objektif dan mendukung pengambilan keputusan berbasis data.

Clustering sebagai Metode Unsupervised Learning

Clustering adalah salah satu teknik utama dalam *data mining* yang bertujuan mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik tanpa menggunakan label kelas awal. Metode ini termasuk dalam kategori *unsupervised learning* karena proses pengelompokkan sepenuhnya bergantung pada pola yang terdapat dalam data. Dalam praktiknya, *clustering* digunakan sebagai metode eksploratif untuk mengidentifikasi struktur alami dalam dataset multivariat serta merangkul data ke dalam kelompok yang lebih mudah dianalisis (Ramdani et al., 2025).

Algoritma K-Means

Algoritma *K-Means* merupakan metode *clustering* berbasis partisi yang bekerja dengan membagi data ke dalam sejumlah *k cluster* berdasarkan jarak antara data dan pusat *cluster* (*centroid*).

Proses *K-Means* dilakukan secara iteratif melalui tahapan inialisasi *centroid*, penghitungan jarak, penentuan keanggotaan *cluster*, dan pembaruan *centroid* hingga mencapai kondisi konvergen. Beberapa penelitian di jurnal informatika Indonesia menunjukkan bahwa *K-Means* memiliki keunggulan dari sisi kesederhanaan dan efisiensi komputasi, sehingga banyak digunakan untuk pengelompokkan data numerik berdimensi banyak (Zebua, 2025).

Evaluasi Kualitas *Clustering*

Evaluasi kualitas *clustering* diperlukan untuk memastikan bahwa kelompok data yang terbentuk memiliki tingkat kohesi internal yang baik dan separasi antar *cluster* yang jelas. Salah satu metrik yang umum digunakan adalah *Silhouette Score*, yang mengukur kedekatan suatu data terhadap *cluster* tempatnya berada dibandingkan dengan *cluster* lain. Selain itu, *Davies–Bouldin Index* digunakan untuk mengukur rasio jarak intra-*cluster* dan inter-*cluster*, di mana nilai yang lebih kecil menunjukkan kualitas *clustering* yang lebih baik. Kedua metrik ini banyak digunakan dalam penelitian *clustering* berbasis data numerik di Indonesia (Sholeh & Aeni, 2023).

Sejumlah penelitian di Indonesia telah menerapkan algoritma K-Means pada berbagai jenis dataset numerik. Murdiaty et al. (2020) menerapkan teknik *clustering* pada data multivariat dan menunjukkan bahwa pendekatan ini mampu mengungkap pola data yang tidak terlihat melalui analisis deskriptif. (Ramdani et al., 2025) menggunakan K-Means untuk pengelompokan data sektoral dan menghasilkan *cluster* yang merepresentasikan perbedaan karakteristik data secara jelas. Penelitian lain oleh Zebua, (2025) serta Sholeh dan Aeni (2023) menunjukkan bahwa penggunaan metrik evaluasi *clustering* sangat penting untuk memastikan kualitas hasil pengelompokan. Hasil-hasil penelitian tersebut menjadi landasan teoritis bagi penerapan metode *clustering* pada penelitian ini.

3. METODE PENELITIAN

Penelitian ini menggunakan rancangan penelitian kuantitatif berbasis analisis data dengan pendekatan *unsupervised learning*. Pendekatan ini dipilih karena data yang dianalisis tidak memiliki label kelas awal dan tujuan penelitian adalah melakukan pengelompokan data berdasarkan kemiripan karakteristik numerik. Metode utama yang diterapkan adalah algoritma *K-Means Clustering*, yang umum digunakan dalam analisis data multivariat (Murdiaty et al., 2020; (Taryadi, 2022)). Populasi penelitian adalah seluruh data capaian Standar Pelayanan Minimal (SPM) yang tersedia di Provinsi Nusa Tenggara Timur. Sampel penelitian berupa data capaian SPM dari 22 kabupaten dan kota.

Setiap kabupaten/kota diperlakukan sebagai satu objek data, sedangkan nilai capaian pada setiap sektor layanan diperlakukan sebagai atribut numerik. Teknik pengumpulan data dilakukan melalui dokumentasi, yaitu menggunakan data sekunder yang diperoleh dari Program SKALA (Sinergi dan Kolaborasi untuk Akselerasi Layanan Dasar) Provinsi Nusa Tenggara Timur. Instrumen penelitian berupa dataset numerik yang diolah menggunakan perangkat lunak analisis data berbasis komputasi.

Tahap awal analisis dimulai dengan pembersihan data (*data cleaning*) untuk memastikan konsistensi format dan tipe data. Mengingat data capaian SPM memiliki variasi satuan dan rentang nilai antar indikator yang berbeda, dilakukan proses standardisasi data. Standardisasi dilakukan menggunakan metode *Standard Scaler (Z-score)* untuk mengubah distribusi data sehingga memiliki rata-rata 0 dan standar deviasi 1. Sebelum proses *clustering*, data numerik dilakukan standardisasi untuk menyamakan skala antar atribut agar perhitungan jarak pada algoritma *K-Means* tidak bias terhadap atribut tertentu (Al Rivan & Sonaru, 2022). Proses standardisasi dinyatakan pada Persamaan (1).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Dengan x merupakan nilai asli data, μ adalah nilai rata-rata atribut, dan σ adalah standar deviasi atribut.

Metode utama yang diterapkan dalam penelitian ini adalah algoritma *K-Means Clustering*. Algoritma ini bekerja secara iteratif untuk meminimalkan nilai *Within-Cluster Sum of Squares (WCSS)* atau fungsi objektif J , sebagaimana ditunjukkan pada Persamaan (2) (Murdiaty et al., 2020; Dewi et al., 2022).

$$J = \sum_{j=1}^K \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (2)$$

Di mana K adalah jumlah kluster, C_j adalah kluster ke- j , x_i adalah data ke- i , dan μ_j adalah *centroid* kluster ke- j .

Dalam menentukan keanggotaan kluster, jarak antara data dan titik pusat kluster (*centroid*) dihitung menggunakan jarak *Euclidean*. Penggunaan jarak *Euclidean* umum diterapkan pada algoritma *K-Means* karena konsisten dengan fungsi objektif yang meminimalkan WCSS. Perhitungan jarak *Euclidean* dinyatakan pada Persamaan (3).

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

Dengan n menyatakan jumlah atribut.

Evaluasi kualitas hasil pengelompokkan dilakukan menggunakan *Silhouette Score*. Metrik ini digunakan untuk mengukur seberapa dekat suatu titik data dengan klasternya sendiri dibandingkan dengan kluster terdekat. Nilai *Silhouette* dihitung berdasarkan Persamaan (4) (Galela, 2023)

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

Di mana $a(i)$ merupakan jarak rata-rata intra-kluster dan $b(i)$ adalah jarak rata-rata ke kluster terdekat. Nilai *Silhouette Score* berada pada rentang -1 hingga 1 .

Selain itu, kualitas kluster juga dievaluasi menggunakan *Davies–Bouldin Index (DBI)*. Indeks ini digunakan untuk mengukur rasio penyebaran data di dalam kluster terhadap jarak antar kluster. Rumus DBI ditunjukkan pada Persamaan (5) (Sembiring, 2022).

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (5)$$

Dengan S_i menyatakan dispresi kluster ke- i , S_j menyatakan dispresi kluster ke- j , dan M_{ij} menyatakan jarak antara *centroid* kluster ke- i dan kluster ke- j . Nilai DBI yang lebih kecil menunjukkan kualitas separasi kluster yang semakin baik. Model penelitian yang digunakan menggambarkan alur analisis data yang dimulai dari data input berupa dataset SPM, dilanjutkan dengan tahap pembersihan dan standardisasi data, proses pengelompokkan menggunakan algoritma *K-Means*, serta evaluasi kualitas hasil kluster menggunakan *Silhouette Score* dan *Davies–Bouldin Index*. Model ini menempatkan data sebagai objek utama analisis dan algoritma *clustering* sebagai alat komputasional untuk mengekstraksi pola kemiripan data.

4. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil analisis pengelompokkan data capaian Standar Pelayanan Minimal (SPM) di Provinsi Nusa Tenggara Timur. Analisis dilakukan melalui tahapan prapemrosesan data, penentuan jumlah kluster optimal, implementasi algoritma, serta validasi hasil terhadap data eksternal untuk memastikan keandalan model dalam mendukung pengambilan kebijakan.

Prapemrosesan Data

Dataset penelitian terdiri dari 22 kabupaten/kota dengan 6 dimensi atribut numerik capaian SPM tahun 2023. Pemahaman terhadap karakteristik data sangat penting sebelum dilakukan proses *clustering*. Tabel 1 menunjukkan ringkasan statistik deskriptif dari dataset yang digunakan.

Tabel 1. Statistik Deskriptif Dataset SPM NTT 2024.

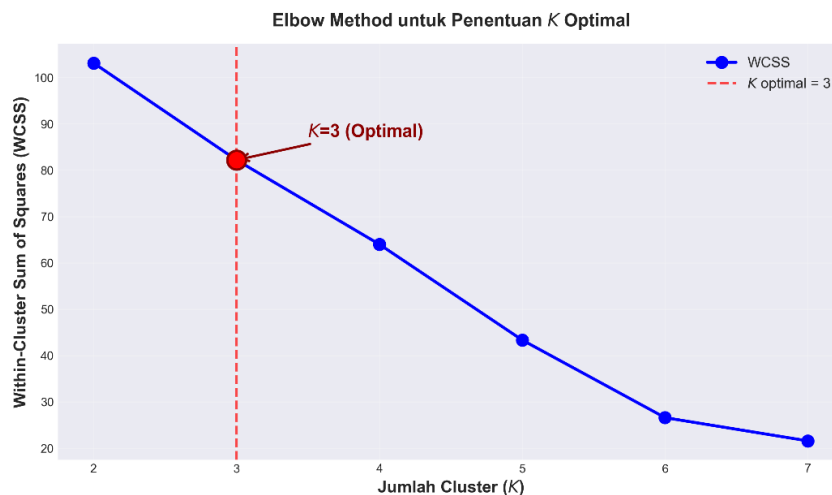
Sektor	Mean	Std	Min	Q1	Median	Q3	Max
Pendidikan	84,12	13,43	46,91	79,46	88,36	93,9	97,51
Kesehatan	82,03	7,33	65,34	79,94	81,14	83,51	100
Pekerjaan Umum (PU)	92,67	21,37	0	93,98	100	100	100
Perumahan dan Kawasan Permukiman (Perkim)	97,2	7,6	72,97	100	100	100	100
Ketentraman dan Ketertiban Umum (Trantibum)	88,34	12,9	55,88	88,1	92,05	96,67	100
Sosial	87,49	10,48	59,49	80,34	90,48	95,46	100

Berdasarkan Tabel 1, terlihat bahwa sektor Pekerjaan Umum memiliki variasi data tertinggi (standar deviasi 21,37) dengan rentang nilai yang sangat lebar (0% hingga 100%), mengindikasikan adanya ketimpangan infrastruktur yang signifikan antar wilayah. Deteksi *outlier* menggunakan metode *Interquartile Range* (IQR) mengidentifikasi adanya 18 data pencilan (13,6%), namun data tersebut tetap dipertahankan karena merepresentasikan kondisi riil ketimpangan wilayah kepulauan

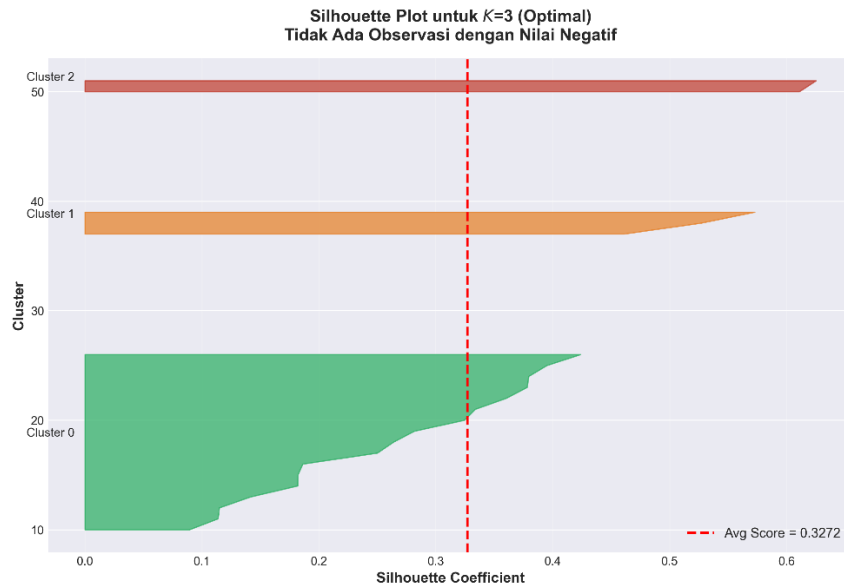
Selanjutnya, normalisasi data dilakukan menggunakan teknik *Z-Score Standardization*. Langkah ini krusial untuk mengatasi perbedaan skala antar dimensi (misalnya variasi besar pada PU dibandingkan variasi kecil pada Kesehatan), sehingga memastikan setiap atribut memiliki bobot yang setara dalam perhitungan jarak *Euclidean* pada algoritma *K-Means*.

Penentuan Jumlah Kluster Optimal

Penentuan parameter jumlah kluster (K) dilakukan melalui triangulasi tiga metrik validasi, yaitu *Elbow Method*, *Silhouette Analysis*, dan *Davies-Bouldin Index* (DBI). Gambar 1 menampilkan kurva *Elbow* dan plot *Silhouette* untuk memvisualisasikan kualitas kluster.



Gambar 1. Grafik Elbow Method menunjukkan titik optimal pada $K = 3$.



Gambar 2. Visualisasi Silhouette Plot untuk $K = 3$.

Berdasarkan Gambar 1, penurunan nilai inersia ($WCSS$) yang signifikan terjadi pada saat kluster berjumlah 3 ($K = 3$), dengan persentase penurunan sebesar 20,3% dari ($K = 2$). Tabel 2 merangkum hasil perbandingan metrik validasi untuk berbagai nilai K .

Tabel 2. Perbandingan Metrik Validasi *Cluster*.

K	$WCSS$	<i>Silhouette</i>	DBI
2,00	103,13	0,30	1,41
3,00	82,20	0,33	0,87
4,00	64,00	0,21	1,08
5,00	43,34	0,33	0,86
6,00	26,59	0,37	0,70
7,00	21,57	0,35	0,68

Konsistensi ketiga metrik mengonfirmasi bahwa ($K = 3$) adalah konfigurasi paling optimal, ditandai dengan *Silhouette Score* tertinggi (0,3272) dan nilai DBI terendah (0,8725). Visualisasi *Silhouette Plot* pada Gambar 2 juga menunjukkan tidak ada observasi dengan nilai koefisien negatif, yang mengindikasikan bahwa seluruh kabupaten telah ditempatkan pada kelompok yang tepat (*correct assignment*).

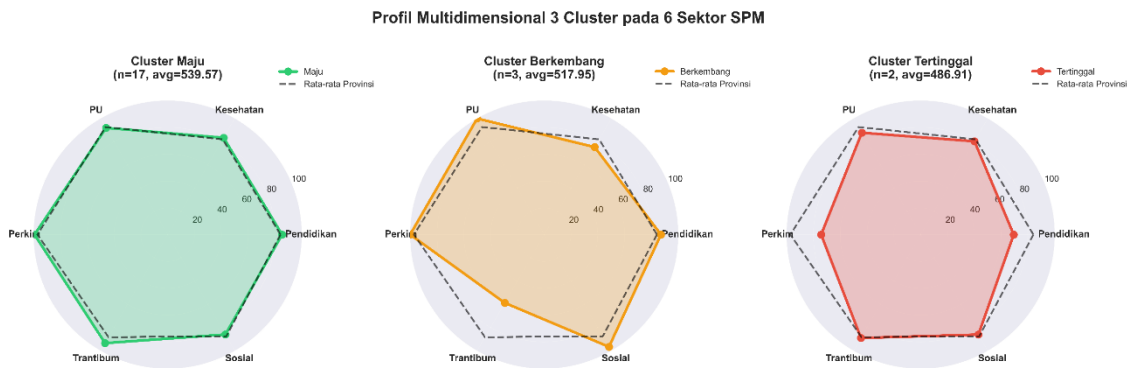
Implementasi dan Profil Kluster

Implementasi algoritma *K-Means* dengan inisialisasi *K-Means* mencapai konvergensi pada iterasi ke-2 dengan waktu komputasi 0,028 detik. Proses ini menghasilkan tiga kelompok data dengan karakteristik yang distingtif. Profil nilai rata-rata (*centroid*) untuk setiap kluster disajikan pada Tabel 3.

Tabel 3. Profil Centroid *Cluster* (Nilai Original SPM %).

Kategori	Pendidikan	Kesehatan	PU	Perkim	Trantibum	Sosial	Total Skor	Persen
Maju	85,31	83,40	91,95	99,41	93,46	86,03	539,56	89,90
Berkembang	87,12	75,36	100,00	100,00	58,81	96,66	517,95	86,30
Tertinggal	69,44	80,33	87,76	74,24	89,05	86,10	486,92	81,20
Rata-rata Provinsi	84,12	82,03	92,67	97,20	88,34	87,49	531,84	88,64

Visualisasi karakteristik multidimensional disajikan melalui *Radar Chart* pada Gambar 2.



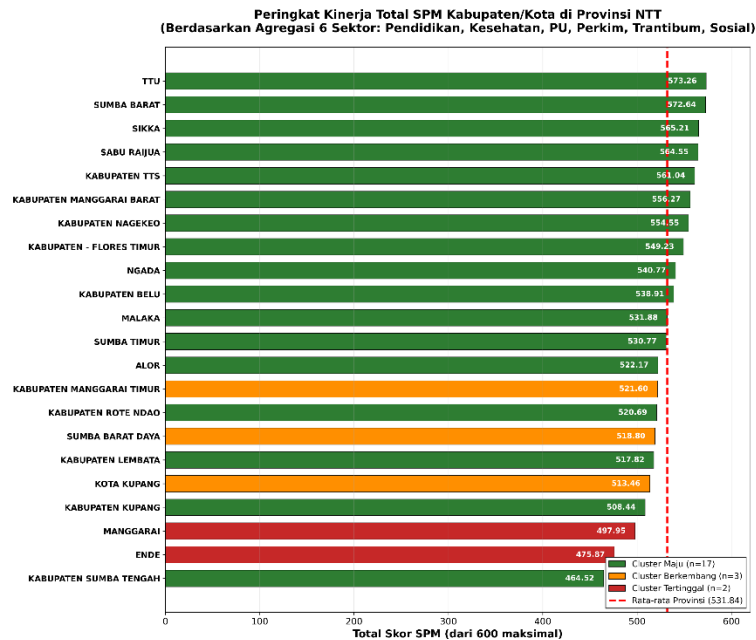
Gambar 3. Radar Chart Profil *Cluster*.

Berdasarkan Tabel 3 dan Gambar 3, terbentuk pola karakteristik sebagai berikut:

- Cluster Maju* (77,3%): Kelompok mayoritas ini menunjukkan kinerja "*Balanced Excellence*" dengan capaian konsisten tinggi di seluruh dimensi, terutama pada sektor Perkim (+2,21 poin di atas rata-rata) dan Trantibum (+5,12 poin).
- Cluster Berkembang* (13,6%): Kelompok ini menunjukkan fenomena unik "*Paradoks Infrastruktur*". Wilayah ini (Kab. Kupang, Manggarai Timur, Sabu Raijua) memiliki capaian sempurna (100%) pada infrastruktur fisik (PU dan Perkim), namun sangat rendah pada sektor Trantibum (58,81%, atau -29,53 poin di bawah rata-rata). Hal ini mengindikasikan bahwa pembangunan fisik yang masif belum tentu diikuti dengan penguatan tata kelola ketertiban umum.
- Cluster Tertinggal* (9,1%): Kelompok ini mengalami ketertinggalan multidimensi, terutama pada sektor Pendidikan (69,44%) dan Perumahan Rakyat (74,24%). Karakteristik ini dominan ditemukan pada wilayah kepulauan dengan aksesibilitas terbatas.

Validasi Hasil

Validasi dilakukan dengan membandingkan hasil *clustering* terhadap Peringkat Resmi SPM Kemendagri. Tabel 4 menunjukkan tingkat kesesuaian hasil model dengan data riil. Untuk menguji konsistensi model, dilakukan pemetaan antara total skor SPM dengan label kluster yang dihasilkan oleh algoritma. Visualisasi distribusi peringkat dan sebaran kluster disajikan pada Gambar 3.



Gambar 4. Peringkat Kinerja Total SPM Kabupaten/Kota.

Berdasarkan Gambar 4, terlihat adanya konsistensi visual (*visual consistency*) yang kuat antara total skor dengan pengelompokan algoritma.

- a) Zona Atas (Hijau): Wilayah dengan total skor di atas rata-rata provinsi (garis putus-putus merah) didominasi oleh anggota *Cluster Maju*.
- b) Zona Transisi (Oranye): Wilayah *Cluster Berkembang* tersebar di area menengah hingga menengah-bawah, yang menegaskan karakteristik uniknya (unggul di infrastruktur namun rendah di trantibum, sehingga total skornya tertekan).
- c) Zona Bawah (Merah): Wilayah dengan total skor terendah secara konsisten teridentifikasi sebagai anggota *Cluster Tertinggal*.

Tabel 4 berikut menyajikan ringkasan validasi pada sampel wilayah representatif (*top-tier* dan *bottom-tier*).

Tabel 4. Distribusi Ranking Berdasarkan Total Skor SPM.

Rank Model	Kabupaten	Jenis Cluster	Total Skor
1	TTU	Maju	573,26
2	Sumba Barat	Maju	572,64
3	Sikka	Maju	565,21
4	Sabu Raijua	Maju	564,55
5	TTS	Maju	561,04
6	Manggarai Barat	Maju	556,27
7	Nagekeo	Maju	554,55
8	Flores Timur	Maju	549,23
9	Ngada	Maju	540,77
10	Belu	Maju	538,91
11	Malaka	Maju	531,88
12	Sumba Timur	Maju	530,77
13	Alor	Maju	522,17
14	Manggarai Timur	Berkembang	521,6
15	Rote Ndao	Maju	520,69
16	Sumba Barat Daya	Berkembang	518,8
17	Lembata	Maju	517,82
18	Kota Kupang	Berkembang	513,46
19	Kabupaten Kupang	Maju	508,44
20	Manggarai	Tertinggal	497,95
21	Ende	Tertinggal	475,87
22	Sumba Tengah	Maju	464,52

Tingkat akurasi validasi yang tinggi pada peringkat teratas (*top-tier*) dan terbawah (*bottom-tier*) memperkuat reliabilitas model. Dampak kebijakan yang dapat dirumuskan antara lain: (1) *Cluster* Maju dapat dijadikan *benchmark* provinsi dan fokus pada inovasi layanan digital; (2) *Cluster* Berkembang memerlukan prioritas pada penguatan kapasitas aparatur Satpol PP (Satuan Polisi Pamong Praja) dan tata kelola keamanan; serta (3) *Cluster* Tertinggal memerlukan kebijakan afirmatif berupa Dana Alokasi Khusus (DAK) untuk percepatan infrastruktur dasar pendidikan dan perumahan.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa implementasi algoritma *K-Means Clustering* dengan teknik standardisasi *Z-Score* terbukti efektif dan valid dalam memetakan pola data multivariat capaian Standar Pelayanan Minimal (SPM) di Provinsi Nusa Tenggara Timur. Berdasarkan rangkaian pengujian parameter, ditetapkan bahwa konfigurasi optimal terbentuk pada jumlah tiga klaster ($K=3$). Hasil pengelompokan data menunjukkan dominasi *Cluster* Maju sebesar 77,3%, diikuti oleh *Cluster* Berkembang sebesar 13,6%, dan *Cluster* Tertinggal sebesar 9,1%. Secara teknis, model ini dinyatakan valid dengan nilai *Davies-Bouldin Index* sebesar 0,872 dan *Silhouette Score* 0,327 tanpa adanya kesalahan penempatan anggota.

Selain itu, visualisasi profil *centroid* berhasil mengekstraksi pengetahuan baru (*knowledge discovery*) berupa pola anomali pada *Cluster* Berkembang, yang teridentifikasi memiliki kinerja infrastruktur unggul namun mengalami penurunan signifikan pada sektor ketenteraman umum.

Berdasarkan temuan tersebut, pengembangan penelitian selanjutnya disarankan untuk menerapkan metode optimasi inialisasi titik pusat awal, seperti algoritma *K-Means++* atau *Genetic Algorithm*, guna meningkatkan stabilitas hasil. Selain itu, mengingat adanya *outlier* pada sektor Pekerjaan Umum, melakukan perbandingan performa dengan algoritma berbasis densitas seperti *DBSCAN* sangat direkomendasikan. Dari sisi implementasi praktis, hasil komputasi ini sebaiknya diintegrasikan ke dalam sistem *dashboard* monitoring berbasis web agar pemetaan karakteristik wilayah dapat dilakukan secara *real-time* dan dinamis.

DAFTAR REFERENSI

- Abid, D., Adikusuma, R. W., AL Fikri, A. M., & Hapsari, R. K. (2023). Penerapan Metode *K-Means Clustering* Untuk Analisa Penjualan Komoditas Toko Tani Indonesia. *KERNEL: Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika*, 3(2), 25–30. <https://doi.org/10.31284/j.kernel.2022.v3i2.4076>
- Al Rivan, M. E., & Sonaru, R. A. (2022). Perbandingan Metode *K-Means* dan GA *K-Means* untuk *Clustering* Dataset Heart Disease Patients. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(3), 2585–2597. <https://doi.org/10.35957/jatisi.v9i3.2799>
- Darmawan, S. A. D., & Karmilasari. (2024). Penerapan Metode & *K-Means Clustering* dan Simple Moving Average untuk Memprediksi Jenis Penyakit di Provinsi Jawa Timur. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 877–886. <https://doi.org/10.25126/jtiik.1148703>
- Dewi, F. P., Aryni, P. S., & Umaidah, Y. (2022). Implementasi Algoritma *K-Means Clustering* Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(2), 111–121. <https://doi.org/10.14421/jiska.2022.7.2.111-121>
- Fadlil, A., Riadi, I., & Mulyana, Y. (2022). Penerapan Algoritma *K-Means* pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan. *Jurnal Teknologi Informatika Dan Komputer*, 8(2), 352–366. <https://doi.org/10.37012/jtik.v8i2.1261>
- Galela, M. R. (2023, November 21). *Koefisien Silhouette untuk Menentukan Jumlah Kluster Ideal*. Kemenkeu Learning Center.
- 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>
- Hutagalung, M. I., & Sriani, S. (2024). Pengelompokan Data Penyakit THT Menggunakan Algoritma *K-Means Clustering*. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1568–1577. <https://doi.org/10.57152/malcom.v4i4.1692>

- Isni Rinjani, Saeful Anwar, & Ruli Herdiana. (2023). PENGELOMPOKAN DAERAH BENCANA ALAM MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 3(1), 35–51. <https://doi.org/10.55606/juisik.v3i1.417>
- Margaretha, A. P., Ulinnuha, N., & Intan, P. K. (2025). *Clustering* Data Kecelakaan Lalu Lintas melalui Algoritma *K-Means* dengan Seleksi Fitur Chi-Square. *INTEGER: Journal of Information Technology*, 10(2). <https://doi.org/10.31284/j.integer.0.v10i2.7529>
- Murdiaty, M., Angela, A., & Sylvia, C. (2020). Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma *K-Means*. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(3), 744. <https://doi.org/10.30865/mib.v4i3.2213>
- Prasetyo, D., Lestari, W., & Atima, V. (2024). PENERAPAN *CLUSTERING* DENGAN *K-MEANS* UNTUK PEMILIHAN MENU FAVORIT DI TETRA COFFEESHOP. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*.
- Ramadani, M. S., & Fatah, Z. (2024). ANALISIS PENGELOMPOKAN DATA NILAI SISWA UNTUK MENENTUKAN SISWA BERPRESTASI MENGGUNAKAN METODE *CLUSTERING K-MEANS*. *JISSI (Jurnal Riset Sistem Informasi)*, 103–110.
- Ramdani, R., Suarna, N., Ali, I., & Efendi, D. I. (2025). PENERAPAN ALGORITMA *K-MEANS* DALAM ANALISIS DATA KEPENDUDUKAN UNTUK OPTIMALISASI PENGELOMPOKAN DI DESA PASAWAHAN. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5639>
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode *Clustering K-Means*. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Sembiring, E. M. (2022). Penerapan *K-Means Clustering* Untuk Pengelompokan Penyebaran Demam Berdarah Dengue (DBD) Di Kabupaten Deli Serdang. *TIN: Terapan Informatika Nusantara*, 2(11), 673–677. <https://doi.org/10.47065/tin.v2i11.1503>
- Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model *Clustering* dengan Menggunakan Algoritma *K-Means*. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(1), 56. <https://doi.org/10.30998/string.v8i1.16388>
- Siahaan, M. (2022). *Data Mining* Strategi Pembangunan Infrastruktur Menggunakan Algoritma *K-Means*. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(3), 316–324. <https://doi.org/10.32736/sisfokom.v11i3.1453>
- Simanjuntak, H. T. A., Silaban, P. E. P., Manurung, J. K. S., & Sormin, V. H. (2023). Klasterisasi Berita Bahasa Indonesia Dengan Menggunakan *K-Means* Dan Word Embedding. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(3), 641–652. <https://doi.org/10.25126/jtiik.2023106468>
- Zebua, J. A. G. (2025, November 28). *ANALISIS MULTIVARIAT MENGGUNAKAN K-MEANS UNTUK MENGIDENTIFIKASI POLA PENGELOMPOKAN BIJI GANDUM*. RPub by RStudio.