Insearch
Information System Research Journal

E-ISSN: 2775-4669
Volume 5, Nomor 2, Agustus 2025

Analisis Perbandingan Kinerja UMAP K-Means dan UMAP Agglomerative dalam Klasterisasi Indikator Perilaku Hidup Bersih dan Sehat (PHBS) Tahun 2023 di Provinsi Jawa Tengah

Friza Nur Fatmala[1], Difta Alzena Sakhi[2], Muhammad Nasrudin[3], Kartika Maulida Hindrayani[4]
Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur
Jl. Raya Rungkut Madya, Kec. Gunung Anyar, Kota Surabaya, Jawa Timur 60294, Indonesia
23083010051@student.upnjatim.ac.id[1], 23083010061@student.upnjatim.ac.id[2], nasrudin.fasilkom@upnjatim.ac.id[3], kartika.maulida.ds@upnjatim.ac.id[4]

Received: 13 Jun 2025 | Revised: 05 Aug 2025
Accepted: 11 Aug 2025 | Published: 28 Aug 2025

Abstrak

Perilaku Hidup Bersih dan Sehat (PHBS) memiliki peran penting dalam meningkatkan kualitas kesehatan masyarakat di Jawa Tengah yang masih menghadapi ketimpangan antar wilayah. Penelitian ini bertujuan untuk mengidentifikasi pola klasterisasi kabupaten/kota berdasarkan enam indikator PHBS, dengan membandingkan dua metode analisis populer: K-Means dan Agglomerative *Clustering*, yang dipadukan dengan teknik reduksi dimensi nonlinear berupa UMAP. Proses pengolahan data mencakup pengecekan nilai hilang, deteksi *outlier*, dan normalisasi menggunakan *Robust Scaler*. Hasil evaluasi dengan *One-way* MANOVA pada kedua metode menunjukkan nilai p-value < 0,05 yang mengindikasikan adanya perbedaan signifikan antar klaster. Namun, evaluasi struktural menunjukkan bahwa UMAP Agglomerative memperoleh *silhouette score* tertinggi yakni 0,63 yang sedikit melampaui UMAP K-Means yaitu 0,62. Dengan demikian, metode Agglomerative mampu membentuk dua klaster yang lebih kohesif dan terstruktur, yaitu daerah dengan pelaksanaan PHBS yang tidak merata di semua aspek (klaster 0) dan daerah dengan pelaksanaan yang relatif merata (klaster 1). Temuan ini dapat menjadi dasar dalam perumusan kebijakan kesehatan yang lebih terarah dan berbasis data di Provinsi Jawa Tengah.

Kata kunci: PHBS, Agglomerative Clustering, UMAP

Abstract

Clean and Healthy Living Behavior (PHBS) plays an important role in improving the quality of public health in Central Java, which still faces regional disparities. This study aims to identify clustering patterns of districts/cities based on six PHBS indicators by comparing two popular analysis methods: K-Means and Agglomerative Clustering, combined with a nonlinear dimension reduction technique known as UMAP. The data processing involves checking for missing values, detecting outliers, and normalizing using the Robust Scaler. The evaluation results using One-way MANOVA on both methods showed p-values < 0.05, indicating significant differences between clusters. However, structural evaluation revealed that UMAP Agglomerative achieved the highest silhouette score of 0.63, slightly surpassing UMAP K-Means at 0.62. Thus, the Agglomerative method was able to form two more cohesive and structured clusters, namely areas with uneven implementation of PHBS in all aspects (cluster 0) and areas with relatively even implementation (cluster 1). These findings can serve as a basis for formulating more targeted and data-driven health policies in Central Java Province.

Keywords: PHBS, Agglomerative Clustering, UMAP

I. PENDAHULUAN

Perilaku Hidup Bersih dan Sehat (PHBS) merupakan salah satu indikator penting dalam peningkatan derajat kesehatan masyarakat di Indonesia. Kementerian Kesehatan Republik Indonesia menetapkan PHBS sebagai pendekatan strategis di tingkat rumah tangga melalui praktik, seperti persalinan oleh tenaga kesehatan, pemberian ASI eksklusif, kebiasaan mencuci tangan pakai sabun, serta akses

terhadap air minum dan sanitasi yang layak [1]. Pemerataan capaian indikator ini sangat penting untuk mendukung target pembangunan kesehatan nasional dan Tujuan Pembangunan Berkelanjutan (TPB), khususnya tujuan ke-3 mengenai kehidupan sehat dan sejahtera [2].

Namun demikian, data dari Provinsi Jawa Tengah menunjukkan ketimpangan capaian PHBS antar kabupaten/kota. Beberapa wilayah mencatat nilai indikator yang tinggi, sedangkan lainnya tertinggal dalam aspek-aspek kunci, seperti sanitasi, air bersih, dan pelayanan kesehatan dasar [3]. Ketimpangan ini berkorelasi dengan tingginya prevalensi stunting dan risiko penyakit berbasis lingkungan. Hal ini menunjukkan bahwa pendekatan kebijakan yang seragam belum sepenuhnya efektif dan diperlukan analisis berbasis data untuk mengelompokkan wilayah berdasarkan karakteristik PHBS guna mendukung perencanaan intervensi yang lebih tepat sasaran.

Berbagai studi sebelumnya telah menerapkan metode klasterisasi, seperti K-Means dan Agglomerative Klastering dalam konteks analisis kesehatan masyarakat [4][5]. Namun, pendekatan tersebut memiliki keterbatasan dalam menangkap struktur data nonlinear, terutama pada data multivariat kompleks. *Uniform Manifold Approximation and Projection* (UMAP) sebagai metode reduksi dimensi nonlinear telah terbukti unggul dalam mempertahankan struktur lokal dan global data [6]. Selain itu, UMAP cenderung menghasilkan pemisahan klaster yang lebih jelas dengan jarak antar klaster yang proporsional serta kepadatan titik data yang rapat pada satu klaster [7]. Kombinasi UMAP dengan metode klasterisasi telah digunakan secara luas, tetapi implementasinya untuk menganalisis indikator PHBS masih jarang dilakukan, khususnya pada wilayah Jawa Tengah. Belum ada penelitian yang memanfaatkan keunggulannya dalam menghasilkan pemisahan klaster yang lebih jelas pada data multivariat. Hal ini menunjukkan adanya celah penelitian untuk mengeksplorasi pemanfaatan UMAP dalam meningkatkan ketepatan identifikasi persebaran PHBS yang kompleks di wilayah tersebut.

Penelitian ini dibatasi pada enam indikator PHBS tingkat rumah tangga yang memiliki relevansi langsung dengan kesehatan masyarakat dan tersedia secara lengkap pada publikasi resmi Badan Pusat Statistik (BPS) Provinsi Jawa Tengah tahun 2023. Pemilihan indikator ini didasarkan pada kelengkapan data dan keterbandingan antar kabupaten/kota, yang mencakup aspek kesehatan ibu dan anak, akses air minum dan sanitasi, pola makan sehat, serta perilaku merokok. Dengan mempertimbangkan batasan ruang lingkup tersebut, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator PHBS dengan membandingkan kinerja metode UMAP K-Means dan UMAP Agglomerative. Studi ini diharapkan dapat memberikan kontribusi dalam pengembangan model klasterisasi untuk data nonlinear di bidang kesehatan masyarakat serta menyediakan segmentasi spasial wilayah yang berguna bagi perumusan kebijakan berbasis bukti.

II. TINJAUAN PUSTAKA

A. *Robust Scaler*

Robust Scaler adalah metode penskalaan data yang dirancang untuk mengatasi pengaruh *outlier* dalam *preprocessing* data. Berbeda dengan standarisasi menggunakan *Standard Scaler*, yang menggunakan rata-rata dan standar deviasi, *Robust Scaler* menggunakan median dan rentang interkuartil (IQR) sehingga lebih tahan terhadap distorsi akibat nilai ekstrem [8]. Transformasi matematisnya adalah:

$$x' = \frac{x - median(x)}{(Q_3 - Q_1)} \quad (1)$$

Keterangan:
x : Nilai asli data
$median(x)$: Nilai tengah dari data
Q_3 : Kuartil atas
Q_1 : Kuartil bawah

Dengan pendekatan ini, data yang telah diskalakan akan terpusat pada median dan memiliki skala berdasarkan IQR, membuat distribusi data lebih stabil meskipun terdapat *outlier* [9]. Penggunaan *Robust Scaler* terbukti efektif dalam meningkatkan akurasi model pembelajaran mesin pada data yang tidak berdistribusi normal atau mengandung anomali, seperti pada aplikasi *credit risk modeling* dan deteksi anomali [10].

Dengan demikian, *Robust Scaler* merupakan pilihan ideal untuk menjaga stabilitas model dan validitas analisis, terutama pada data yang rawan *outlier*.

B. *Uniform Manifold Approximation and Projection* (UMAP)

UMAP adalah teknik reduksi dimensi nonlinear yang memetakan data berdimensi tinggi ke ruang berdimensi rendah dengan mempertahankan struktur lokal dan global data. UMAP membangun graf k-tetangga terdekat menggunakan bobot keterhubungan yang dihitung dengan [6].

$$w_{ij} = exp(-\frac{max(0, d(x_i, x_j) - \rho_i)}{\sigma_i}) \quad (2)$$

Keterangan:
$d(x_i, x_j)$: Jarak antara titik x_i dan x_j
ρ_i : Parameter lokal untuk titik x_i
σ_i : Parameter skala
$Fungsi\ max(0, .)$ memastikan hanya jarak lebih besar dari ρ_i yang dihitung

Kemudian, graf ini dipetakan ke ruang rendah dengan meminimalkan *cross-entropy loss* antara graf asli dan graf *embedding*:

$$C = \sum_{i \neq j} [w_{ij} log \frac{w_{ij}}{w'ij} + (1 - w_{ij}) log \frac{1 - w_{ij}}{1 - w'_{ij}}] \qquad (3)$$

Keterangan:

w_{ij} : Probabilitas kesamaan antara titik x_i dan x_j di ruang dimensi tinggi (sebelum reduksi dimensi)
w'_{ij} : Probabilitas kesamaan antara titik x_i dan x_j di ruang dimensi rendah (setelah reduksi dimensi)
$i \neq j$: Penjumlahan dilakukan untuk pasangan titik yang berbeda
log : Fungsi logaritma yang digunakan dalam perhitungan *cross-entropy loss* untuk mengukur perbedaan antara graf asli dan graf *embedding*

UMAP unggul dalam menjaga keseimbangan antara pelestarian struktur lokal dan global serta skalabilitas pada dataset besar, dengan parameter *n_neighbors* dan *min_dist* yang dapat disesuaikan. Metode ini banyak digunakan untuk visualisasi, klastering, dan analisis data di berbagai bidang, termasuk klasifikasi penyakit dan analisis teks [11] [12].

C. *Manhattan Distance*

Manhattan Distance, juga dikenal sebagai *Taxicab Distance* atau *L1 Norm*, adalah metrik jarak yang mengukur total jarak antara dua titik dengan menjumlahkan selisih absolut koordinatnya secara komponen per komponen. *Manhattan distance* antara dua item didefinisikan sebagai jumlah dari perbedaan komponen masing-masing titik [13]. Secara matematis, jaraknya dirumuskan sebagai:

$$D(x, y) = \sum_{i=1}^{n} |x_i - y_i| \qquad (4)$$

Keterangan:

x_i : Elemen ke-i dari vektor x
y_i : Elemen ke-i dari vektor y
$|x_i - y_i|$: Selisih absolut antara elemen x_i dan y_i
n : Jumlah dimensi (jumlah elemen dalam vektor)

Manhattan Distance sering diaplikasikan dalam algoritma K-Means sebagai metrik jarak alternatif yang mampu menghasilkan kualitas klastering yang kompetitif, bahkan lebih baik dibandingkan *Euclidean Distance* pada beberapa dataset. *Manhattan Distance* sering diaplikasikan dalam algoritma K-Means sebagai metrik jarak alternatif yang mampu menghasilkan kualitas klastering yang kompetitif, bahkan lebih baik dibandingkan *Euclidean Distance* pada beberapa dataset. Penelitian sebelumnya menunjukkan bahwa penggunaan *Manhattan Distance* pada data bibit padi menghasilkan klastering yang optimal, dengan nilai *silhouette score* yang mendekati atau melebihi metrik lainnya, seperti *Euclidean* dan *Cosine Similarity*, tergantung pada karakteristik data yang digunakan [14].

D. *Complete Linkage*

Complete linkage merupakan salah satu prosedur pengelompokan yang didasarkan pada jarak terjauh antara objek dalam masing-masing kelompok [15]. Jika jarak terjauh antar objek masih tergolong dekat, maka kedua klaster akan digabungkan. Namun, jika jaraknya memang terbilang jauh, klaster tersebut tidak akan digabungkan. Secara matematis, metode *complete linkage* dapat dinyatakan sebagai berikut [16].

$$D(C_i, C_j) = \max_{x_s \in C_i, x_t \in C_j} d(x_s, x_t) \qquad (5)$$

Keterangan:

$D(C_i, C_{j)}$: jarak antara klaster Ci dan klaster Cj,
$d(x_s, x_t)$: jarak antara titik x_s dan titik x_t, di mana $x_s \epsilon C_i$ dan $x_t \epsilon C_j$

Penggunaan *complete linkage* akan menghasilkan klaster yang lebih rapat karena mengurangi dampak dari data yang berjarak jauh [16]. Hal ini membuat hasil pengelompokan menjadi lebih stabil dan representatif terhadap pola utama dalam data.

E. *Euclidean Distance*

Euclidean Distance adalah metode pengukuran jarak yang digunakan untuk menentukan seberapa jauh dua titik berada dalam ruang *Euclidean*, baik itu dalam dua dimensi, tiga dimensi, maupun dimensi yang lebih tinggi [17]. Jarak ini dihitung berdasarkan akar kuadrat dari jumlah selisih kuadrat antara koordinat masing-masing titik. Adapun perhitungan matematis dari *Euclidean Distance* adalah sebagai berikut [18].

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - y_{jk})^2} \qquad (6)$$

Keterangan:

d_{ij} : Jarak klaster data pusat (i) dan data atribut (j)
n : Banyak data
x_{ij} : Data dalam klaster pusat hingga k
y_{ij} : Objek data hingga k

Dalam K-Means, *Euclidean Distance* merupakan matriks jarak yang paling umum digunakan karena cenderung memberikan hasil yang lebih optimal [19]. Kemampuannya dalam mengukur jarak geometris secara langsung membuat metode ini banyak diterapkan pada data dengan skala yang seragam dan distribusi yang relatif linier.

F. *Silhouette Score*

Silhouette score merupakan perhitungan yang menggambarkan tingkat kemiripan suatu data dengan anggota dalam klasternya, serta dihitung secara terpisah untuk setiap objek dalam kelompok tersebut [20]. Metrik ini juga mempertimbangkan sejauh mana suatu data berbeda dengan klaster lain yang paling dekat. Rumus perhitungan *silhouette score* dapat dituliskan dalam bentuk berikut [16].

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \quad (7)$$

Keterangan:

$a(i)$: Rata-rata jarak objek i ke semua objek dalam klaster sendiri

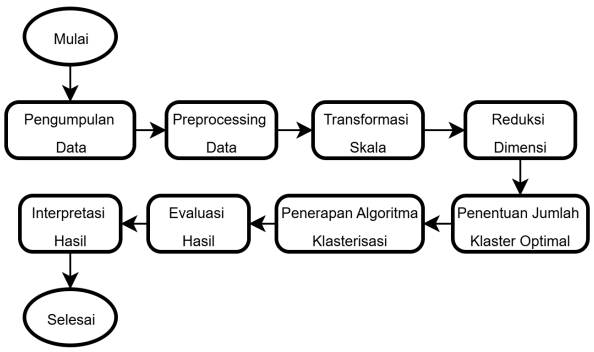
$b(i)$: Rata-rata objek i ke semua objek dalam klaster lain

Silhouette score akan menghasilkan nilai pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan kualitas yang optimal pada klaster yang terbentuk. Sebaliknya, nilai yang mendekati -1 menggambarkan kualitas pengelompokan yang buruk [20].

III. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis klaster untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator Perilaku Hidup Bersih dan Sehat (PHBS) pada tingkat rumah tangga. Tahapan penelitian disajikan pada Gambar 1.



Gambar 1. Diagram Alur Tahapan Penelitian

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Jawa Tengah tahun 2023. Unit analisis terdiri dari 35 kabupaten/kota. Enam indikator utama yang digunakan sebagai variabel input dalam analisis klaster meliputi:

1. Persentase Persalinan yang Ditolong oleh Tenaga Kesehatan [21], yang mencerminkan akses dan kualitas layanan kesehatan dalam mendukung kesehatan ibu dan anak.
2. Persentase Bayi dengan ASI Eksklusif [22], yang menunjukkan prevalensi pemberian ASI eksklusif, faktor penting dalam pertumbuhan dan perkembangan bayi.
3. Persentase Rumah Tangga dengan Akses Air Minum Layak [23], yang mencerminkan kualitas dan ketersediaan sumber air bersih bagi masyarakat.
4. Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak [24], yang menunjukkan seberapa baik akses terhadap fasilitas sanitasi yang memadai di tiap rumah tangga.
5. Rata-Rata Konsumsi Buah dan Sayur Per Kapita Seminggu [25][26], yang mencerminkan pola makan sehat dan kesejahteraan gizi masyarakat.
6. Rata-Rata Konsumsi Rokok Per Kapita Seminggu [27], yang menggambarkan prevalensi kebiasaan merokok dalam masyarakat, yang memiliki dampak langsung terhadap kesehatan.

Pemilihan indikator tersebut didasarkan pada keterkaitan langsungnya dengan perilaku dan kualitas kesehatan lingkungan rumah tangga, serta ketersediaan data yang lengkap dan terbaru pada tahun 2023.

C. *Preprocessing* Data

1. Pemeriksaan Nilai Hilang (*Missing Values*)

Pemeriksaan nilai hilang dilakukan untuk menjaga integritas data. Apabila ditemukan nilai yang hilang, penanganan dilakukan melalui dua pendekatan, yaitu penghapusan (*listwise deletion*) atau imputasi. Metode imputasi yang digunakan dapat berupa rata-rata (mean), median, atau algoritma K-Nearest Neighbors (KNN Imputation), tergantung pada distribusi dan karakteristik variabel. Pemilihan metode dilakukan dengan mempertimbangkan dampaknya terhadap struktur data.

2. Deteksi *Outlier*

Outlier diidentifikasi menggunakan metode Z-Score untuk mendeteksi penyimpangan signifikan dalam data. Observasi yang terindikasi sebagai *outlier* diidentifikasi menggunakan metode *Z-Score*, dengan ambang batas umum >|3|. Namun, *outlier* tidak langsung dieliminasi karena dalam konteks data spasial, seperti wilayah kabupaten/kota, nilai ekstrem dapat merepresentasikan karakteristik khas yang relevan secara substantif dalam pembentukan klaster.

3. Transformasi Skala Data

Untuk mengatasi perbedaan skala antar variabel dan mengurangi sensitivitas terhadap pencilan, dilakukan transformasi menggunakan *Robust Scaler*. Metode ini menggunakan median dan *interquartile range* (IQR) dalam proses penskalaan sehingga lebih stabil dibandingkan metode standardisasi konvensional, seperti *Z-Score* maupun *Min-Max Scaling*. Tujuan utama transformasi ini adalah untuk memastikan bahwa setiap variabel memiliki kontribusi proporsional dalam proses klasterisasi.

4. Reduksi Dimensi

Reduksi dimensi dilakukan menggunakan metode *Uniform Manifold Approximation and Projection* (UMAP), yang dikenal memiliki kemampuan dalam mempertahankan struktur lokal maupun global dari data berdimensi tinggi. Dibandingkan dengan metode lain, seperti PCA dan t-SNE, UMAP lebih efisien secara komputasi dan memberikan hasil visualisasi yang lebih terstruktur. Dalam penelitian ini, UMAP dikonfigurasi dengan parameter *n_neighbors* = 5, *min_dist* = 0.1, dan metrik jarak *Manhattan*, yang dipilih karena performanya yang kompetitif dalam mendeteksi pola lokal pada data spasial.

D. Implementasi Algoritma

1. Agglomerative Klastering

Penentuan jumlah klaster optimal dilakukan menggunakan metode Agglomerative Hierarchical *Clustering*, yang divisualisasikan melalui dendrogram. Dendrogram menyajikan struktur penggabungan objek secara hierarkis berdasarkan *linkage distance*. Dengan mengamati lonjakan jarak penggabungan yang paling signifikan, ditetapkan *threshold* pemotongan yang menghasilkan dua klaster utama. Pemilihan ini didasarkan pada identifikasi visual terhadap titik diskontinuitas terbesar dalam dendrogram, yang mengindikasikan pemisahan paling representatif dalam struktur data.

2. K-Means

Berdasarkan jumlah klaster optimal yang diperoleh (k=2), algoritma K-Means *Clustering* diterapkan untuk membandingkan performa metode non-hierarkis terhadap pendekatan hierarkis. Algoritma ini diinisialisasi dengan pemilihan *centroid* secara acak, kemudian dilanjutkan dengan iterasi berbasis minimisasi jarak *Euclidean* hingga mencapai konvergensi. Penggunaan jumlah klaster yang sama memastikan perbandingan yang objektif terhadap bentuk klaster dan distribusi anggotanya.

E. Evaluasi Hasil

Secara statistik, evaluasi kualitas klaster dilakukan melalui pendekatan *One-Way* Manova dengan membandingkan p-value dari beberapa statistik uji, seperti Wilks' Lambda, Pillai's Trace, Hotelling-Lawley, dan Roy's' Greatest Root terhadap nilai $\alpha = 0{,}05$. Evaluasi ini bertujuan untuk menguji apakah terdapat perbedaan karakteristik dalam kombinasi variabel indikator untuk setiap klaster yang terbentuk.

Sementara itu, evaluasi kualitas klaster secara struktural dilakukan menggunakan metrik *silhouette score*. Metrik ini mengukur tingkat kesamaan suatu objek dengan klaster tempat ia berada dibandingkan dengan klaster lain. Nilai *silhouette* yang mendekati 1 menunjukkan klaster yang kompak dan terpisah secara jelas, sedangkan nilai mendekati 0 atau negatif mengindikasikan kemungkinan tumpang tindih antarklaster.

F. Interpretasi Hasil

Interpretasi hasil klaster dilakukan dengan menganalisis nilai rata-rata dari masing-masing variabel PHBS pada setiap klaster. Pendekatan ini memungkinkan identifikasi karakteristik dominan dari tiap klaster serta membedakan profil kesehatan rumah tangga antar kelompok. Rincian statistik ini disajikan pada Tabel 7 dalam Bab Pembahasan untuk memperjelas perbedaan signifikan antarklaster.

Selanjutnya, hasil klasterisasi divisualisasikan dalam bentuk peta sebaran wilayah berdasarkan klaster di Provinsi Jawa Tengah. Visualisasi spasial ini memberikan perspektif geografis terhadap distribusi PHBS, sekaligus menjadi dasar bagi perumusan strategi intervensi kesehatan masyarakat yang lebih terarah dan berbasis data wilayah.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dataset yang digunakan mencakup enam indikator utama yang merepresentasikan Perilaku Hidup Bersih dan Sehat (PHBS) pada tingkat rumah tangga di 35 kabupaten/kota di Provinsi Jawa Tengah. Meskipun indikator PHBS secara umum terdiri dari 10 komponen, ketersediaan data dari sumber resmi membatasi penelitian ini pada enam indikator utama. Keenam indikator tersebut tetap mencerminkan aspek-aspek penting dalam pelaksanaan PHBS di masyarakat. Tabel berikut menyajikan beberapa entri dari dataset.

Tabel 1. Tampilan Dataset

Kabupaten/Kota	x_1	x_2	x_3	x_4	x_5	x_6
Cilacap	94,12	76,88	88,75	80,38	1,32	2,88
Banyumas	100,00	56,60	93,57	83,00	1,42	2,80
Purbalingga	100,00	71,09	82,82	77,19	1,34	2,92
Banjarnegara	100,00	78,06	87,56	46,09	1,40	2,59
Kebumen	98,67	73,34	87,34	93,68	1,39	2,18
…	…	…	…	…	…	…
Kota Tegal	100,00	51,51	100,00	94,12	1,23	2,29

Keterangan:

- x_1 : Persentase Persalinan yang Ditolong oleh Tenaga Kesehatan
- x_2 : Persentase Bayi dengan ASI Eksklusif
- x_3 : Persentase Rumah Tangga dengan Akses Air Minum Layak
- x_4 : Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak
- x_5 : Rata-Rata Konsumsi Buah dan Sayur Per Kapita Seminggu
- x_6 : Rata-Rata Konsumsi Rokok Per Kapita Seminggu

Dataset ini merupakan hasil penggabungan dari beberapa sumber data yang telah dihimpun menjadi satu kesatuan. Setiap variabel di dalamnya memiliki rentang nilai yang berbeda-beda sehingga diperlukan proses pengolahan lebih lanjut untuk memastikan hasil analisis yang lebih akurat.

B. Preprocessing Data

1. Pemeriksaan Nilai Hilang (*Missing Values*)

Berdasarkan hasil pemeriksaan, tidak ditemukan nilai yang hilang atau *missing value* dalam dataset. Hal ini menunjukkan bahwa informasi terkait enam indikator Perilaku Hidup Bersih dan Sehat (PHBS) untuk seluruh kabupaten/kota di Jawa Tengah telah tercatat secara lengkap. Dengan demikian, dataset dinyatakan siap untuk dianalisis tanpa memerlukan proses imputasi atau penanganan terhadap data yang tidak lengkap. Ketersediaan data yang utuh ini sangat penting dalam menjaga validitas hasil analisis dan mendukung pengambilan keputusan berbasis data yang akurat.

2. Deteksi *Outlier*

Deteksi *outlier* menggunakan metode *Z-Score* menunjukkan adanya nilai yang menyimpang pada lima variabel, yaitu persentase persalinan yang ditolong oleh tenaga kesehatan, persentase rumah tangga dengan akses air minum layak, persentase rumah tangga dengan akses sanitasi layak, rata-rata konsumsi buah dan sayur per kapita per minggu, serta rata-rata konsumsi rokok per kapita per minggu. Nilai-nilai ekstrem tersebut merepresentasikan kondisi nyata di lapangan dan dapat mencerminkan karakteristik khas dari wilayah tertentu. Oleh karena itu, mempertahankan keberadaan *outlier* merupakan langkah yang tepat untuk menjaga keutuhan dan representativitas data secara menyeluruh. Selain itu, keberadaan *outlier* yang relevan juga dapat memberikan wawasan tambahan yang tidak dapat diperoleh dari data hasil pengolahan.

3. Transformasi Skala Data

Transformasi skala data dilakukan dengan pendekatan *robust* yang berbasis median dan rentang antarkuartil (*interquartile range*/IQR). Pendekatan ini dipilih karena mampu meminimalkan pengaruh *outlier* yang masih dipertahankan dalam dataset. Berikut merupakan tampilan data setelah dilakukan transformasi skala.

Tabel 2. Data setelah Transformasi

x_1	x_2	x_3	x_4	x_5	x_6
-5,880	0,482	-1,321	-0,655	-0417	0,400
0,000	-0,352	-0,433	-0,462	0,202	0,287
0,000	0,243	-2,414	-0,889	-0,294	0,461
0,000	0,530	-1,541	-3,170	0,086	0,000
-1,330	0,336	-1,581	0,320	-0,009	-0,550
…	…	…	…	…	…
0,000	-0,563	0,752	0,352	-0,970	-0,410

Keterangan:

x_1 : Persentase Persalinan yang Ditolong oleh Tenaga Kesehatan
x_2 : Persentase Bayi dengan ASI Eksklusif
x_3 : Persentase Rumah Tangga dengan Akses Air Minum Layak
x_4 : Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak
x_5 : Rata-Rata Konsumsi Buah dan Sayur Per Kapita Seminggu
x_6 : Rata-Rata Konsumsi Rokok Per Kapita Seminggu

Hasil transformasi menunjukkan bahwa skala antarvariabel telah seragam sehingga memungkinkan algoritma klastering untuk mengelompokkan data dengan lebih adil. Langkah ini dinilai penting mengingat beberapa variabel, seperti konsumsi buah dan sayur serta rokok, memiliki rentang nilai yang berbeda secara signifikan dengan data lain.

4. Reduksi Dimensi

Proses reduksi dimensi dilakukan dengan metode UMAP (*Uniform Manifold Approximation and Projection*) guna meringkas enam variabel awal menjadi dua komponen utama pada data yang berkarakter nonlinier. Hasil transformasi dari beberapa entri pada dataset akan disajikan dalam tabel berikut.

Tabel 3. Data setelah Reduksi Dimensi

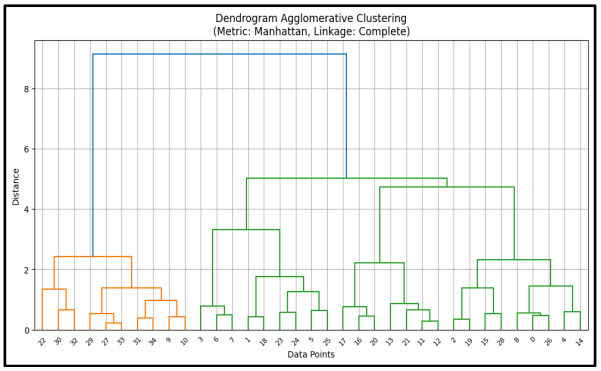
UMAP 1	UMAP 2
15,2341	10,6246
14,2414	10,4393
15,5984	9,8864
13,6699	12,0472
15,5236	11,1925
…	…
11,7192	14,5544

Nilai-nilai tersebut diperoleh setelah melakukan parameter tuning pada UMAP, yang menghasilkan kombinasi optimal berupa jumlah tetangga *(n_neighbor)* sebanyak 5, jarak minimal *(min_dist)* sebesar 0,1, dan penggunaan *Manhattan distance* sebagai ukuran jarak. Kombinasi ini memungkinkan pengelompokan data dengan karakteristik serupa secara lebih akurat sehingga struktur data nonlinier dapat direduksi dengan baik dan mendukung proses klastering yang lebih efektif.

C. Implementasi Algoritma

1. Agglomerative Klastering

Dalam Agglomerative Klastering, proses pengelompokan akan menghasilkan suatu struktur sistematis yang disebut dengan dendogram. Melalui dendogram ini, dapat ditentukan suatu *threshold* yang menjadi acuan dalam mengidentifikasi jumlah klaster paling optimal. Berikut ini merupakan visualisasi dari dendogram.



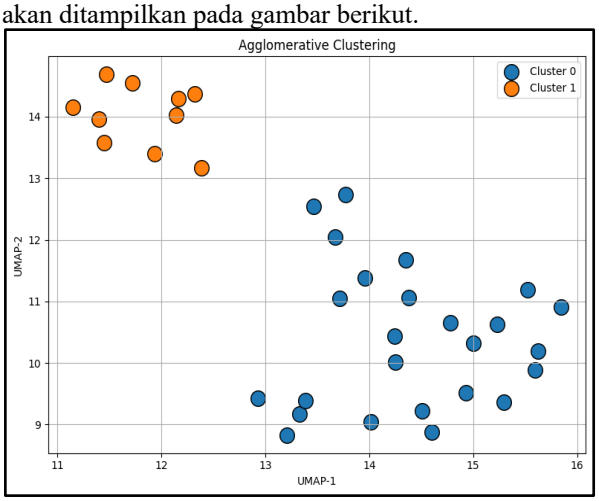
Gambar 2. Dendogram

Pada penelitian ini, penentuan *threshold* dilakukan dengan mengamati lonjakan jarak antar penggabungan klaster. Berdasarkan pengamatan visual, lonjakan jarak tertinggi terjadi pada rentang antara 5,5 hingga 8,5, yang ditandai dengan garis penggabungan berwarna biru. Oleh karena itu, *threshold* akan ditetapkan pada rentang tersebut, misalnya pada jarak ke tujuh. Penetapan *threshold* ini akan menghentikan proses penggabungan klaster sehingga diperoleh dua klaster utama sebagai hasil akhir.

Tabel 4. Jumlah Anggota Klaster Agglomerative

Klaster	Jumlah
0	25
1	10

Dari hasil klastering, terlihat bahwa jumlah anggota dari masing-masing klaster tidak seimbang. Hal ini mengindikasikan bahwa sebagian besar wilayah memiliki karakteristik penerapan PHBS yang serupa dan tergolong dalam satu kelompok besar. Sementara itu, kelompok lainnya mencakup wilayah-wilayah dengan pola yang menyimpang dari mayoritas. Secara visual, hasil klaster dengan pendekatan Agglomerative akan ditampilkan pada gambar berikut.



Gambar 3. Hasil Klaster dengan Agglomerative

Berdasarkan visualisasi di atas, tampak bahwa pendekatan Agglomerative mampu melakukan pemisahan yang cukup jelas antara dua kelompok data. Titik-titik pada setiap klaster terlihat membentuk dua kelompok tanpa adanya tumpang tindih antaranggota. Klaster pertama yang ditandai dengan warna biru cenderung tersebar di bagian kanan bawah grafik, sementara klaster kedua yang berwarna oranye lebih terkonsentrasi di bagian kiri atas. Pemisahan antarklaster yang cukup tegas ini memberikan gambaran awal bahwa struktur klaster yang terbentuk bersifat stabil dan dapat memberikan informasi penting pada tahap analisis lanjutan.

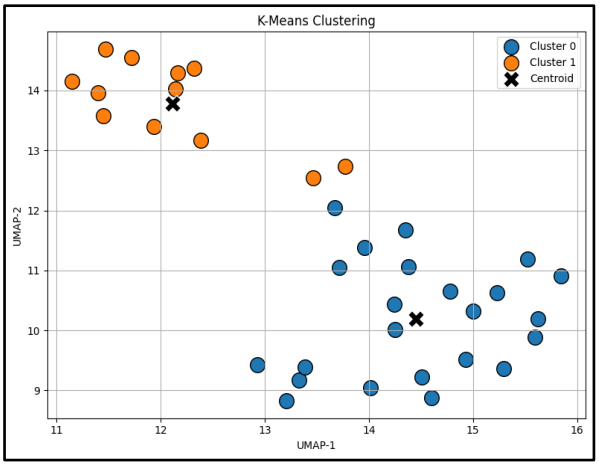
2. K-Means

Jumlah klaster yang digunakan dalam algoritma K-Means didasarkan pada hasil analisis dendrogram dari pendekatan sebelumnya, yaitu sebanyak dua klaster. Hasil pengelompokan menggunakan algoritma K-Means disajikan pada tabel berikut.

Tabel 5. Jumlah Anggota Klaster K-Means

Klaster	Jumlah
0	23
1	12

Jumlah anggota dari setiap klaster tidak berbeda secara signifikan dengan pendekatan Agglomerative. Masih terjadi ketidakseimbangan antara anggota dari masing-masing klaster. Dengan demikian, pengelompokan wilayah dari penerapan PHBS ini memang terfokus pada kumpulan data dengan pola umum serta kumpulan data yang menyimpang dari pola umum. Untuk memperjelas perbedaan antara hasil pengelompokan dengan K-Means dengan pendekatan sebelumnya, akan ditampilkan visualisasi berikut ini.



Gambar 4. Hasil Klaster dengan K-Means

Dalam visualisasi K-Means, terlihat adanya *centroid* yang menjadi dasar dalam mengidentifikasi anggota untuk masuk pada klaster tertentu. Secara

umum, pemisahan antarklaster terlihat cukup jelas, yakni data berhasil terbagi ke dalam dua klaster yang relatif terpisah secara spasial. Namun, terdapat dua anggota klaster berwarna oranye (klaster 1) yang terletak di dekat wilayah klaster berwarna biru (klaster 0). Hal ini mengindikasikan adanya sedikit tumpang tindih atau potensi ambiguitas dalam penentuan keanggotaan klaster. Dengan demikian, diperoleh indikasi awal bahwa metode Agglomerative cenderung lebih representatif dibandingkan algoritma K-Means dalam mengelompokkan dataset ini.

D. Evaluasi Hasil

Setelah tiga asumsi utama *One-Way* MANOVA, yaitu normalitas multivariat, homogenitas matriks varians-kovarians, dan independensi antar pengamatan terpenuhi, evaluasi hasil secara statistik dilakukan dengan membandingkan p-value berikut terhadap nilai $\alpha = 0{,}05$ [28].

Tabel 6. Perbandingan P-value

Statistik Uji	Agglomerative	K-Means
Wilks' Lambda	0,000002236397	0,000001749233
Pillai's Trace	0,000002236397	0,000001749233
Hotelling-Lawley Trace	0,000002236397	0,000001749233
Roy's Greatest Root	0,000002236397	0,000001749233

P-value dari keempat statistik uji menunjukkan nilai yang kurang dari 0,05 pada kedua metode klastering yang digunakan. Nilai ini mengindikasikan bahwa terdapat setidaknya satu variabel indikator yang berbeda secara signifikan pada setiap klaster. Dengan demikian, dapat disimpulkan bahwa kombinasi variabel indikator yang digunakan dalam pengelompokan berhasil membentuk dua klaster dengan karakteristik yang berbeda secara signifikan apabila didasarkan pada pendekatan statistik.

Sementara itu nilai *silhouette score* yang mewakili evaluasi hasil klastering secara struktural akan disajikan pada tabel berikut.

Tabel 6. Perbandingan *Silhouette Score*

Metode	Silhouette Score
Agglomerative	0,6320
K-Means	0,6289

Nilai yang lebih dari 0,5 ini, mengindikasikan bahwa kedua metode klastering telah berhasil mengelompokkan kabupaten/kota ke dalam klaster yang sesuai dengan karakteristik PHBS. Artinya, struktur klaster yang terbentuk relatif baik dan representatif terhadap pola yang ada dalam data. Dengan mempertimbangkan evaluasi *One-Way* Manova, nilai *silhouette score,* dan visualisasi yang telah disajikan, dapat diambil kesimpulan akhir bahwa pendekatan Agglomerative lebih optimal dalam menggambarkan pola pengelompokan data berdasarkan kemiripan karakteristik PHBS antarkabupaten/kota di Jawa Tengah.

E. Interpretasi Hasil

Hasil klastering mengungkapkan bahwa wilayah kabupaten/kota di Provinsi Jawa Tengah dapat dibagi ke dalam dua kelompok berdasarkan indikator Perilaku Hidup Bersih dan Sehat (PHBS) pada tingkat rumah tangga. Untuk memperoleh gambaran lebih jelas mengenai ciri khas masing-masing klaster, dilakukan analisis terhadap nilai rata-rata pada setiap variabel.

Tabel 7. Mean tiap Klaster

Klaster	x_1	x_2	x_3	x_4	x_5	x_6
0	99,35	64,97	92,80	83,36	1,45	2,74

Keterangan:
- x_1 : Persentase Persalinan yang Ditolong oleh Tenaga Kesehatan
- x_2 : Persentase Bayi dengan ASI Eksklusif
- x_3 : Persentase Rumah Tangga dengan Akses Air Minum Layak
- x_4 : Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak
- x_5 : Rata-Rata Konsumsi Buah dan Sayur Per Kapita Seminggu
- x_6 : Rata-Rata Konsumsi Rokok Per Kapita Seminggu

Melalui tabel tersebut dapat diketahui bahwa Klaster 0 terdiri dari kabupaten/kota dengan pelaksanaan PHBS yang relatif baik, tetapi belum merata di seluruh indikator. Beberapa indikator, seperti pemberian ASI eksklusif dan konsumsi buah serta sayur memiliki rata-rata yang lebih tinggi daripada klaster lain. Rata-rata persalinan yang ditolong oleh tenaga kesehatan juga menunjukkan hasil yang nyaris sempurna walaupun lebih rendah dari klaster lain. Di sisi lain, indikator yang berkaitan dengan lingkungan, seperti akses terhadap air minum layak dan sanitasi layak, menunjukkan capaian yang lebih rendah. Angka konsumsi rokok yang lebih tinggi juga mencerminkan adanya tantangan dalam penerapan gaya hidup sehat. Oleh karena itu, kabupaten/kota yang termasuk dalam klaster ini masih memerlukan penanganan lebih lanjut agar seluruh indikator PHBS dapat tercapai secara menyeluruh.

Sementara itu, klaster 1 menunjukkan kabupaten/kota dengan capaian indikator PHBS yang cenderung merata di berbagai aspek. Persentase akses terhadap air minum layak, sanitasi layak, dan persalinan dengan tenaga kesehatan yang lebih tinggi mencerminkan keberadaan infrastruktur dasar yang

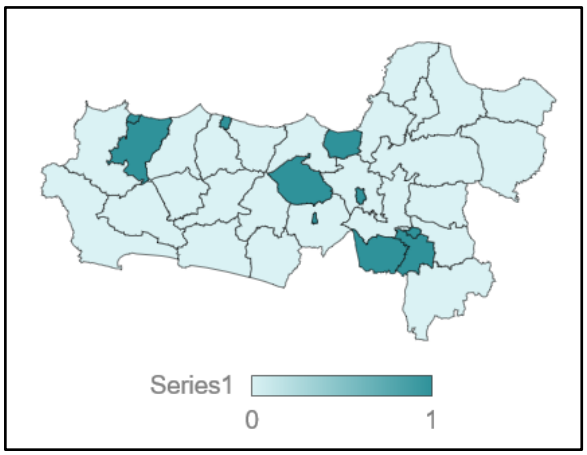
lebih baik serta pemanfaatannya secara optimal oleh masyarakat. Di samping itu, rata-rata konsumsi rokok yang relatif rendah mengindikasikan tingginya kesadaran akan pola hidup yang lebih sehat. Meskipun angka konsumsi buah dan sayur sedikit lebih rendah dari klaster 0, capaian indikator ini tetap berada pada tingkat yang cukup baik. Secara keseluruhan, klaster 1 merepresentasikan kabupaten/kota dengan implementasi PHBS yang lebih seimbang antara aspek perilaku dan ketersediaan fasilitas. Dengan demikian, diperlukan strategi pemeliharaan dan penguatan agar capaian positif ini dapat terus dipertahankan dan ditingkatkan.

Untuk mengetahui wilayah mana saja yang tergabung dalam masing-masing klaster, tabel berikut akan menyajikan daftar kabupaten/kota berdasarkan hasil pengelompokan.

Tabel 8. Persebaran Kabupaten/Kota

Klaster	Kabupaten/Kota
Klaster 0	Kab. Cilacap, Banyumas, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Purworejo, Kab. Wonosobo, Kab. Magelang, Kab. Boyolali, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Kudus, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Kendal, Kab. Batang, Kab. Pekalongan, Kab. Pemalang, Kab. Brebes.
Klaster 1	Kab. Klaten, Kab. Sukoharjo, Kab. Temanggung, Kab. Tegal, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal.

Selanjutnya, untuk mempermudah pemahaman mengenai distribusi geografis dari hasil klasterisasi, visualisasi dalam bentuk peta ditampilkan pada gambar berikut.



Gambar 5. Peta Persebaran Klaster

Dari visualisasi tersebut, terlihat bahwa kabupaten/kota dalam klaster 1 cenderung tersebar di wilayah perkotaan atau daerah dengan tingkat pembangunan yang lebih tinggi. Hasil ini mengindikasikan bahwa penerapan PHBS yang optimal umumnya terjadi di kawasan perkotaan. Kondisi tersebut dapat dikaitkan dengan tingginya tingkat kesadaran masyarakat kota terhadap pentingnya PHBS, serta ketersediaan dan kemudahan akses terhadap sarana dan prasarana pendukung yang memfasilitasi penerapan seluruh indikator PHBS.

V. KESIMPULAN

Berdasarkan hasil evaluasi, dapat dibuktikan bahwa metode Agglomerative *Clustering* menunjukkan hasil yang lebih optimal dibandingkan K-Means dalam mengelompokkan 35 kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator Perilaku Hidup Bersih dan Sehat (PHBS) tahun 2023. Hal tersebut dibuktikan dengan nilai *silhouette score* yang lebih tinggi dan penentuan anggota klaster yang lebih jelas. Pendekatan Agglomerative berhasil membagi wilayah menjadi dua klaster dengan karakteristik yang berbeda. Klaster pertama menggambarkan wilayah dengan penerapan PHBS yang belum merata akibat kurangnya akses terhadap air minum dan sanitasi layak. Sementara itu, klaster kedua memperlihatkan karakteristik wilayah dengan pelaksanaan PHBS yang lebih merata dan optimal. Hal ini ditandai dengan ketersediaan fasilitas yang memadai serta pola hidup sehat yang lebih konsisten. Hasil analisis tersebut dapat menjadi dasar dalam membentuk kebijakan untuk memperbaiki aspek yang masih kurang pada klaster pertama sekaligus mempertahankan capaian positif pada klaster kedua.

REFERENSI

[1] Kementerian Kesehatan Republik Indonesia, *Pedoman Perilaku Hidup Bersih dan Sehat (PHBS) di Tatanan Rumah Tangga*, Jakarta: Kemenkes RI, 2023, pp. 7–8. [Online]. Tersedia: https://ayosehat.kemkes.go.id/pedoman-phbs. [Diakses: 12 Juni 2025].

[2] Kementerian Perencanaan Pembangunan Nasional/Bappenas, *Lampiran I Rencana Aksi Nasional Tujuan Pembangunan Berkelanjutan (RAN TPB/SDGs) 2021–2024*, Jakarta: Bappenas, 2021, pp. 247–251. [Online]. Tersedia: https://sdgs.bappenas.go.id/website/wp-content/uploads/2023/11/Lampiran-I-RAN-SDGs-2021-2024.pdf

[3] Dinas Kesehatan Provinsi Jawa Tengah, *Profil Kesehatan Provinsi Jawa Tengah Tahun 2023*, Semarang: Dinkes Jateng, 2024. [Online]. Tersedia: https://web-api.bps.go.id/download.php?f=x8E5tIVmk1. [Diakses: 12 Juni 2025].

[4] Z. Jannah, A. S. Y. Irawan, dan E. H. Nurkifli, "Penerapan algoritma K-means dalam segmentasi daerah berdasarkan jumlah warga yang terdaftar menjadi peserta BPJS Kesehatan di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 5710–5718, 2024. [Online]. Tersedia: https://doi.org/10.36040/jati.v8i4.10019

[5] E. Ermila, G. Z. N. Fadhilah, N. W. Erdiani, dan R. A. Saputra, "Klasterisasi pola gejala depresi menggunakan Agglomerative Hierarchical Cluster Analysis berdasarkan skor depresi PHQ-9," *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 2, pp. 01–16, 2025. [Online]. Tersedia: https://journal.sinov.id/index.php/juisik/article/view/993/828

[6] L. McInnes, J. Healy, dan J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv preprint*, arXiv:1802.03426, 2018. [Online]. Tersedia: https://arxiv.org/abs/1802.03426

[7] S. Schmitz, U. Weidner, H. Hammer, dan A. Thiele, "Evaluating uniform manifold approximation and projection for dimension reduction and visualization of PolInSAR features," dalam *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXIV ISPRS Congress (2021 edition), vol. 5, no. 1, pp. 39–46, 2021. [Online]. Tersedia: 10.5194/isprs-annals-v-1-2021-39-2021.

[8] "RobustScaler — scikit-learn 1.7.0 documentation," [Online]. Tersedia: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

[9] A. Ivanov et al., "A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain," *Mathematics*, vol. 10, no. 1942, 2022. [Online]. Tersedia: https://doi.org/10.3390/math10111942

[10] *Journal of Credit Risk*, "Distributionally robust optimization approaches to credit risk management of corporate loan portfolios," 2024. [Online]. Tersedia: https://doi.org/10.21314/JCR.2024.008

[11] N. Suhandi, R. Gustriansyah, dan A. Destria, "Klasifikasi penyakit TBC menggunakan metode UMAP dan K-NN," *Bit-Tech (Binary Digital - Technology)*, vol. 7, no. 3, Apr. 2025. [Online]. Tersedia: https://doi.org/10.32877/bt.v7i3.2227

[12] I. K. T. Mertayasa dan I. D. M. B. Atmaja, "Pemodelan topik pada ulasan hotel menggunakan metode BERTopic dengan prosedur c-TF-IDF," *Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)*, vol. 1, no. 1, pp. 307-316, 2022. [Online]. Tersedia: https://doi.org/10.24843/JNATIA.2022.v01.i01.p37

[13] M. Mukhtar, M. K. M. Ali, F. Arina, A. S. Wicaksono, A. Ikhsan, W. Budiaji, S. Abdullah, D. D. A. Pertiwi, R. Zidny, Y. Oktarisa, dan R. H. Sukarna, "Hierarchical clustering algorithm-dendogram using Euclidean and Manhattan distance," *TEKNIKA: Jurnal Sains dan Teknologi*, vol. 20, no. 01, pp. 98–104, 2024. [Online]. Tersedia: http://jurnal.untirta.ac.id/index.php/ju-tek/

[14] M. S. Pangestu dan M. A. Fitriani, "Perbandingan perhitungan jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam pengelompokan Data Bibit Padi menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, pp. 141–155, 2022. [Online]. Tersedia: https://doi.org/10.30595/sainteks.v19i2.14495

[15] M. I. Afandi, *Analisis cluster hierarki dengan metode complete linkage pada kabupaten/kota di Provinsi Jawa Timur berdasarkan indikator kemiskinan*, Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia, 2020. [Online]. Tersedia: http://etheses.uin-malang.ac.id/18458/1/16610070.pdf

[16] S. D. Rahannabil, H. M. A. Ilyas, H. N. Shafira, M. A. Riam, N. N. Hastim, dan T. K. H. Siregar, "Perbandingan Agglomerative Nesting dan K-Means untuk klasterisasi ketimpangan gender berdasarkan dimensi kesehatan reproduksi," dalam *Seminar Nasional Official Statistics 2024*, pp. 459–470, 2024. [Online]. Tersedia: https://prosiding.stis.ac.id/index.php/semnasoffstat/article/download/1977/605/

[17] M. Nishom, "Perbandingan akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT),* vol. 04, no. 01, pp. 20-24, 2019. [Online]. Tersedia: https://doi.org/10.30591/jpit.v4i1.1253

[18] R. Nooraeni dan G. Nurfalah, "Kajian penerapan jarak Euclidean, Manhattan, Minkowski, dan Chebyshev pada Algoritma Clustering K-Prototype," *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, vol. 4, no. 2, pp. 72-82, 2022. [Online]. Tersedia: http://dx.doi.org/10.30872/jsakti.v4i2.9241

[19] W. W. Pribadi, A. Yunus, dan A. S. Wiguna, "Perbandingan metode k-means dengan Euclidean Distance dan Manhattan Distance pada penentuan zonasi COVID-19 di Kabupaten Malang," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 493-500, 2022. [Online]. Tersedia: https://doi.org/10.36040/jati.v6i2.4808

[20] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis perbandingan Silhouette Coefficient dan metode Elbow pada pengelompokan provinsi di Indonesia berdasarkan indikator IPM dengan K-Medoids," *Jurnal Gaussian*, vol. 13, no. 1, pp. 13-24, 2024. [Online]. Tersedia: https://doi.org/10.14710/j.gauss.13.1.13-24

[21] Badan Pusat Statistik (BPS), "Distribusi persentase wanita berumur 15-49 tahun yang pernah kawin dan melahirkan hidup dalam dua tahun terakhir menurut kabupaten/kota dan penolong persalinan di Provinsi Jawa Tengah," [Online]. Tersedia: https://jateng.bps.go.id/id/statistics-table/2/MTc4MyMy/distribusi-persentase-wanita-berumur-15-49-tahun-yang-pernah-kawin-dan-melahirkan-hidup-dalam-dua-tahun-terakhir-menurut-kabupaten-kota-dan-penolong-persalinan-di-provinsi-jawa-tengah--persen-.html. [Diakses: 12 Juni 2025].

[22] Dinas Kesehatan Provinsi Jawa Tengah, "Jumlah bayi mendapat ASI eksklusif," [Online]. Tersedia: https://data.jatengprov.go.id/dataset/d37-sdjt-jumlah-bayi-mendapat-asi-eksklusif/resource/61e48e34-59ba-43cf-a469-80efb76a56ce. [Diakses: 12 Juni 2025].

[23] Badan Pusat Statistik (BPS), "Persentase rumah tangga menurut kabupaten/kota, sumber air minum bersih, dan akses air minum layak," [Online]. Tersedia: https://jateng.bps.go.id/id/statistics-table/2/MTU2MCMy/persentase-rumah-tangga-menurut-kabupaten-kota--sumber-air-minum-bersih--dan-akses-air-minum-layak--persen-.html. [Diakses: 12 Juni 2025].

[24] Badan Pusat Statistik (BPS), "Persentase rumah tangga yang memiliki akses terhadap sanitasi layak menurut kabupaten/kota di Provinsi Jawa Tengah," [Online]. Tersedia: https://jateng.bps.go.id/id/statistics-table/3/VGtGTU5qbDFlQzl1VWxCTVNWZElXbWRhWkUwMFVUMDkjMyMzMzAw/persentase-rumah-tangga-yang-memiliki-akses-terhadap-sanitasi-layak-menurut-kabupaten-kota-di-provinsi-jawa-tengah.html?year=2023. [Diakses: 12 Juni 2025].

[25] Badan Pusat Statistik (BPS), "Rata-rata konsumsi per kapita seminggu menurut kelompok buah-buahan per kabupaten/kota," [Online]. Tersedia:

https://www.bps.go.id/id/statistics-table/2/MjEwMiMy/rata-rata-konsumsi-perkapita-seminggu-menurut-kelompok-buah-buahan-per-kabupaten-kota--satuan-komoditas-.html. [Diakses: 12 Juni 2025].

[26] Badan Pusat Statistik (BPS), "Rata-rata konsumsi per kapita seminggu menurut kelompok sayur-sayuran per kabupaten/kota," [Online]. Tersedia: https://www.bps.go.id/id/statistics-table/2/MjEwMCMy/rata-rata-konsumsi-perkapita-seminggu-menurut-kelompok-sayur-sayuran-per-kabupaten-kota--satuan-komoditas-.html. [Diakses: 12 Juni 2025].

[27] Badan Pusat Statistik (BPS), "Rata-rata konsumsi per kapita seminggu menurut kelompok rokok dan tembakau per kabupaten/kota," [Online]. Tersedia: https://www.bps.go.id/id/statistics-table/2/MjEwOCMy/rata-rata-konsumsi-perkapita-seminggu-menurut-kelompok-rokok-dan-tembakau-per-kabupaten-kota--satuan-komoditas-.html. [Diakses: 12 Juni 2025].

[28] I. R. A. Suri, Farida, dan D. Fadilah, "Pendekatan Multivariate Analysis of Variance (MANOVA) terhadap pengaruh blended learning berbasis Google classroom pada kemampuan berpikir dan minat belajar," *Sciencestatistics: Journal of Statistics, Probability, and Its Application*, vol. 2, no. 2, pp. 70-79, 2024. [Online]. Tersedia: https://doi.org/10.24127/sciencestatistics.v2i2.6119.