



Analisis Perbandingan Kinerja Algoritma K-Means dan K-Medoids dengan Reduksi Dimensi PCA pada Indikator Kesehatan dan Sosial

Aulia Rizki Firdawanti^{1,*}, Hafidlotul Fatimah Ahmad², Nur Agustiani³

¹ Sekolah Sains Data, Matematika, dan Informatika, Prodi Statistika dan Sains Data, Institut Pertanian Bogor, Bogor, Indonesia

² Sekolah Sains Data, Matematika, dan Informatika, Prodi Ilmu Komputer, Institut Pertanian Bogor, Bogor, Indonesia

³ Sekolah Sains Data, Matematika, dan Informatika, Prodi Aktuaria, Institut Pertanian Bogor, Bogor, Indonesia

Email: ^{1,*} auliafirda@apps.ipb.ac.id, ² hafidlotulftm@apps.ipb.ac.id, ³ nur_agustiani@apps.ipb.ac.id

Email Penulis Korespondensi: auliafirda@apps.ipb.ac.id

Abstrak—Kesehatan masyarakat Jawa Barat menghadapi permasalahan kompleks berupa disparitas akses layanan, gizi buruk, serta ketimpangan indikator sosial antarwilayah. Kondisi ini menuntut analisis berbasis data untuk mengidentifikasi pola kesenjangan sehingga intervensi kebijakan dapat lebih tepat sasaran. Penelitian ini bertujuan mengcluster kabupaten/kota di Jawa Barat berdasarkan indikator kesehatan dan sosial dengan pendekatan *Principal Component Analysis* (PCA) sebagai reduksi dimensi, serta algoritma K-Means dan K-Medoids untuk analisis cluster. Data 27 kabupaten/kota periode 2019–2024 dianalisis setelah standarisasi. PCA menghasilkan dua komponen utama yang menjelaskan 61,4% keragaman. Hasil *scree plot* dan *silhouette* menunjukkan jumlah cluster optimal adalah tiga. Perbandingan menunjukkan rata-rata nilai *silhouette* K-Means sebesar 0,31, sedangkan K-Medoids lebih tinggi yaitu 0,34, menandakan pemisahan yang lebih stabil dan robust terhadap outlier. Pada tahun 2024, cluster 1 berisi wilayah dengan fasilitas kesehatan memadai dan prevalensi balita kurus rendah, cluster 2 berisi daerah dengan keterbatasan puskesmas dan masalah gizi yang tinggi, sementara cluster 3 menampilkan kondisi menengah. Dengan demikian, K-Medoids lebih unggul dibandingkan K-Means karena menghasilkan pemetaan yang lebih konsisten antar tahun. Temuan ini memberi rekomendasi praktis bahwa cluster 2 memerlukan prioritas intervensi berupa peningkatan akses layanan primer dan perbaikan gizi, cluster 1 dijaga kualitas layanannya, sementara cluster 3 diarahkan pada penguatan berkelanjutan.

Kata Kunci: Clustering; Reduksi Dimensi; Indikator Kesehatan; Indikator Sosial; Jawa Barat; K-Means; K-Medoids; Principal Component Analysis

Abstract—Public health in West Java faces complex challenges, including disparities in healthcare access, malnutrition, and socio-economic inequalities across districts. These conditions require data-driven analysis to identify patterns of disparity and provide evidence-based guidance for policy intervention. This study aims to cluster districts/cities in West Java based on health and social indicators using Principal Component Analysis (PCA) for dimensionality reduction, followed by K-Means and K-Medoids algorithms for clustering. Data from 27 districts/cities during 2019–2024 were analyzed after standardization. PCA extracted two principal components explaining 61.4% of the total variance. Scree plot and silhouette results indicated three optimal clusters. Comparative analysis revealed that the average silhouette score of K-Means was 0.31, while K-Medoids achieved a higher score of 0.34, suggesting more stable and robust partitioning against outliers. In 2024, Cluster 1 consisted of regions with adequate healthcare facilities and lower prevalence of underweight children; Cluster 2 grouped regions with limited health infrastructure and higher malnutrition problems, while Cluster 3 showed intermediate conditions. Therefore, K-Medoids outperformed K-Means by producing more consistent clustering across years. These findings offer practical recommendations: Cluster 2 should be prioritized for interventions such as improving primary healthcare access and nutrition programs, Cluster 1 requires maintenance of service quality, and Cluster 3 should be targeted for gradual reinforcement.

Keywords: Clustering; Dimensionality Reduction; Health Indicators; Social Indicators; West Java; K-Means; K-Medoids; Principal Component Analysis

1. PENDAHULUAN

Kesehatan masyarakat merupakan salah satu indikator fundamental dalam menilai pembangunan suatu daerah. Derajat kesehatan tidak hanya ditentukan oleh faktor medis, tetapi juga sangat dipengaruhi oleh determinan sosial, ekonomi, lingkungan, dan budaya yang berkembang di masyarakat [1]. Faktor-faktor tersebut saling berinteraksi dan membentuk kondisi kesehatan masyarakat secara menyeluruh [2]. Tantangan ini tidak hanya mencerminkan persoalan lokal, tetapi juga berkaitan langsung dengan pencapaian Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals/SDGs), khususnya tujuan ketiga mengenai kehidupan sehat dan sejahtera (*good health and well-being*) serta tujuan kesepuluh terkait pengurangan kesenjangan. Di Indonesia, setiap provinsi memiliki tantangan kesehatan yang berbeda sesuai dengan karakteristik geografis, kepadatan penduduk, distribusi infrastruktur, serta kondisi sosial-ekonomi [3]. Jawa Barat, sebagai provinsi dengan jumlah penduduk terbesar di Indonesia, menghadapi kompleksitas permasalahan kesehatan yang sangat beragam, mulai dari keterbatasan akses fasilitas kesehatan, disparitas pelayanan antarwilayah, hingga kesenjangan indikator sosial yang berkontribusi pada kondisi kesehatan penduduk [4].

Analisis indikator kesehatan dan sosial secara simultan menjadi penting untuk memahami pola, hubungan, dan cluster wilayah. Dengan mengelompokkan wilayah berdasarkan kesamaan karakteristik, pemerintah dapat menetapkan prioritas intervensi yang lebih tepat sasaran [5]. Pendekatan semacam ini terbukti bermanfaat dalam perencanaan pembangunan dan kebijakan kesehatan [6], [7]. Metode analisis cluster seperti K-Means dan K-Medoids sering digunakan untuk mengidentifikasi kelompok homogen dalam data multidimensi. K-Means memiliki keunggulan dalam efisiensi perhitungan [8], sementara K-Medoids lebih *robust* terhadap *outlier* sehingga dapat memberikan hasil yang lebih stabil pada data sosial-kesehatan yang kompleks [9], [10], [11]. Kedua metode ini telah diaplikasikan dalam berbagai penelitian, misalnya untuk mengelompokkan wilayah berdasarkan prevalensi penyakit [12], mengevaluasi kinerja fasilitas kesehatan



[13], hingga indikator sosial-ekonomi [14],[15]. Namun, salah satu kendala utama dalam analisis multivariat adalah tingginya dimensi data, yang dapat menyebabkan redundansi dan kesulitan dalam interpretasi hasil [16]. Oleh karena itu, teknik *Principal Component Analysis* (PCA) digunakan sebagai metode reduksi dimensi untuk menyederhanakan data tanpa menghilangkan informasi penting [17], [18]. PCA mampu mengekstraksi peubah-peubah utama yang paling berpengaruh, sehingga hasil cluster menjadi lebih akurat dan mudah ditafsirkan [19],[6]. Studi serupa telah dilakukan di berbagai konteks. Misalnya, penelitian di India menggunakan PCA dan K-Means untuk menganalisis indikator kesehatan ibu dan anak, menghasilkan pemetaan wilayah prioritas dengan lebih efisien [13]. Di Indonesia, pendekatan cluster telah dimanfaatkan untuk mengevaluasi tingkat kemiskinan [12], mengukur ketahanan pangan [14], serta menilai ketersediaan layanan kesehatan [20]. Sementara itu, penelitian lain menunjukkan penerapan K-Medoids yang lebih *robust* dalam pengelompokan data kesehatan rumah sakit [15], dan bahkan digunakan untuk menilai ketahanan sistem kesehatan di Turki [21]. Studi terbaru juga menekankan pentingnya validasi hasil cluster menggunakan berbagai metrik evaluasi untuk meningkatkan reliabilitas analisis [22].

Dengan demikian, kombinasi PCA, K-Means, dan K-Medoids memberikan kerangka analisis yang kuat untuk mengidentifikasi kesenjangan serta menentukan strategi pembangunan kesehatan berbasis bukti [5]. Penelitian ini difokuskan pada Provinsi Jawa Barat dengan tujuan menganalisis cluster indikator kesehatan dan sosial masyarakat menggunakan ketiga metode tersebut. Hasil penelitian diharapkan dapat berkontribusi pada perumusan kebijakan daerah, khususnya dalam meningkatkan efisiensi alokasi sumber daya [6] serta pemerataan akses layanan kesehatan [3] [4]. Melalui pendekatan ini, pembuat kebijakan memperoleh gambaran yang lebih terstruktur mengenai kondisi kesehatan masyarakat dan faktor sosial yang memengaruhinya, sehingga strategi intervensi yang dilakukan dapat lebih efektif, adil, dan berkelanjutan [22].

2. METODOLOGI PENELITIAN

2.1 *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) merupakan teknik reduksi dimensi yang digunakan untuk mentransformasi variabel asal menjadi sejumlah komponen utama yang bersifat ortogonal. Metode ini membantu menjelaskan sebagian besar variansi data dengan jumlah komponen yang lebih sedikit dibandingkan variabel awal [17]. Dengan cara tersebut, PCA mampu mengurangi multikolinearitas sekaligus menyederhanakan analisis multivariat [18]. Selain itu, PCA juga mempermudah visualisasi hasil pengclusteran karena informasi yang relevan disajikan ringkas dan terstruktur [19].

2.2 K-Means

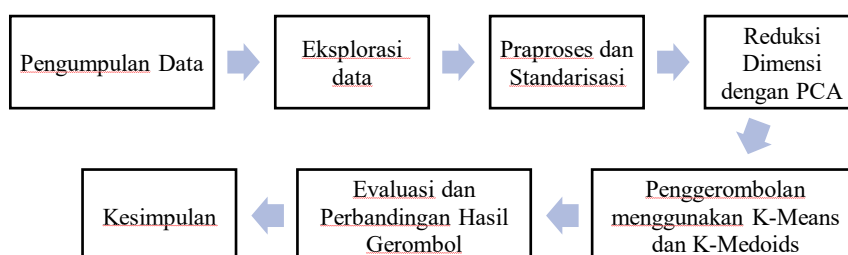
K-Means adalah algoritma pengclusteran berbasis partisi yang umum digunakan dalam analisis data multidimensi [8]. Algoritma ini bekerja dengan menentukan jumlah kluster k , menginisialisasi titik pusat kluster (centroid), kemudian mengalokasikan setiap data ke centroid terdekat menggunakan jarak Euclidean [9]. Proses tersebut dilakukan berulang hingga posisi centroid stabil. Kelebihan utama K-Means adalah kesederhanaan dan efisiensinya untuk dataset berukuran besar. Namun, algoritma ini cukup sensitif terhadap outlier dan variasi ekstrem pada data, sehingga hasilnya bisa kurang stabil dalam kondisi tertentu [14].

2.3 K-Medoids

K-Medoids merupakan varian algoritma partisi yang mirip dengan K-Means, tetapi menggunakan medoid (titik data aktual paling representatif) sebagai pusat kluster [10]. Perbedaan ini menjadikan K-Medoids lebih tahan terhadap outlier dibandingkan K-Means [11]. Dalam konteks analisis sosial dan kesehatan, K-Medoids sering dipandang lebih sesuai karena menghasilkan kluster yang konsisten dari waktu ke waktu. Penelitian sebelumnya juga menunjukkan bahwa K-Medoids dapat memberikan hasil yang lebih stabil dalam pengclusteran data kesehatan maupun sosial [15].

2.4 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis yang dirancang untuk memastikan validitas hasil serta keterandalan metodologi yang digunakan. Gambar 1 menunjukkan tahapan yang dilakukan dalam penelitian ini, adapun tahapan dimulai dari pengumpulan data, praproses data, eksplorasi data, standarisasi, reduksi dimensi menggunakan PCA, penerapan algoritma cluster (K-Means dan K-Medoids), evaluasi hasil, hingga interpretasi.



Gambar 1. Tahapan Analisis Data



2.4.1 Pengumpulan Data

Tahap pertama adalah pengumpulan data sekunder dari portal Open Data Jabar dan Badan Pusat Statistik (BPS) Jawa Barat yang relevan dengan indikator kesehatan dan sosial. Data yang digunakan meliputi peubah jumlah balita stunting, jumlah rumah tangga dengan Perilaku Hidup Bersih dan Sehat (PHBS), jumlah puskesmas, jumlah dokter spesialis gizi, jumlah balita kurang gizi, jumlah balita kurus, persentase bayi berat lahir rendah (BBLR), Indeks Pembangunan Manusia (IPM), indeks pengetahuan, persentase sanitasi layak, dan persentase penduduk miskin pada periode 2019 hingga 2024.

2.4.2. Eksplorasi Data

Tahap eksplorasi data dilakukan untuk memahami karakteristik distribusi peubah numerik sebelum memasuki analisis lebih lanjut. Proses ini penting untuk memastikan bahwa pola sebaran setiap indikator dapat dikenali sejak awal, sehingga potensi bias atau ketidakseimbangan data dapat diantisipasi [5]. Eksplorasi difokuskan pada peubah dengan satuan jumlah maupun persentase, dengan tujuan menilai keragaman nilai serta mendeteksi adanya indikasi pencilan. Selain itu, dilakukan analisis korelasi untuk melihat keterkaitan antar peubah, yang divisualisasikan menggunakan *heatmap*. Visualisasi ini membantu dalam mengidentifikasi kelompok peubah dengan hubungan yang erat, serta menjadi dasar untuk langkah reduksi dimensi pada tahap PCA berikutnya [6].

2.4.3. Praproses Data dan Standarisasi

Selanjutnya dilakukan pemeriksaan terhadap nilai hilang (*missing values*), duplikasi, dan *outlier*. Pada penelitian yang dilakukan, apabila terdapat *missing values* maka akan ditangani dengan metode imputasi, sedangkan pencilan diidentifikasi menggunakan diagram kotak garis (*boxplot*) untuk kemudian dianalisis apakah akan dihapus atau dipertahankan sesuai konteks substantif [16]. Seluruh peubah numerik distandarisasi menggunakan metode z-score untuk mencegah bias dalam proses pengclusteran [14].

2.4.4. Reduksi Dimensi dengan PCA

Setelah proses standarisasi dilakukan, tahap berikutnya adalah reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Metode ini banyak digunakan dalam analisis multivariat karena mampu merangkum data berdimensi tinggi menjadi sejumlah kecil komponen utama yang lebih ringkas [17], [23]. PCA bekerja dengan mentransformasi variabel-variabel awal ke dalam komponen baru yang saling ortogonal, sehingga mengurangi multikolinearitas sekaligus mempertahankan sebagian besar keragaman informasi dalam data [19]. Pendekatan ini telah terbukti efektif pada berbagai studi kesehatan masyarakat di Indonesia, misalnya untuk pemetaan indikator kesehatan berbasis wilayah [18], serta pada penelitian internasional terkait pola gizi dan sistem kesehatan [24].

2.4.5. Pengclusteran dengan K-Means dan K-Medoids

Tahap berikutnya adalah penerapan dua algoritma pengclusteran, yaitu K-Means dan K-Medoids. Pada K-Means, penentuan jumlah cluster (k) dilakukan terlebih dahulu menggunakan metode elbow dan *average silhouette coefficient*, yang banyak digunakan untuk mengevaluasi kualitas pemisahan cluster [20]. Setelah k ditentukan, algoritma K-Means bekerja dengan menginisialisasi *centroid* secara acak, kemudian mengalokasikan setiap titik data ke *centroid* terdekat berdasarkan jarak Euclidean. Proses ini diulang melalui pembaruan posisi *centroid* hingga mencapai kondisi konvergen [8], [9]. K-Medoids menggunakan medoid atau titik data aktual sebagai pusat cluster. Pendekatan ini membuatnya lebih robust karena tidak terlalu terpengaruh oleh pencilan atau nilai ekstrem. Berbagai penelitian menunjukkan bahwa K-Medoids mampu menghasilkan pemisahan yang lebih stabil pada data sosial-kesehatan yang kompleks [10], [11], [15].

2.4.6. Evaluasi dan Perbandingan Hasil Clustering

Evaluasi hasil cluster dilakukan dengan dua pendekatan. Pertama, evaluasi internal menggunakan *average silhouette coefficient* untuk mengukur kesesuaian (*cohesion*) dan pemisahan (*separation*) antar cluster [18], [25]. Kedua, evaluasi eksternal dilakukan dengan membandingkan komposisi anggota cluster dari kedua metode untuk mengidentifikasi konsistensi pengelompokan dan karakteristik unik masing-masing cluster [19], [24]. Evaluasi hasil pengclusteran dilakukan melalui dua pendekatan. Pertama adalah evaluasi internal, yang menggunakan *average silhouette coefficient* untuk menilai tingkat kesesuaian (*cohesion*) dan pemisahan (*separation*) antar cluster. Metode ini banyak direkomendasikan dalam studi validasi pengclusteran karena mampu memberikan ukuran numerik yang jelas terkait kualitas hasil [20]. Studi terbaru juga menekankan pentingnya penggunaan metrik validasi semacam ini untuk meningkatkan reliabilitas analisis data multidimensi [22]. Kedua adalah evaluasi eksternal, yang dilakukan dengan cara membandingkan komposisi anggota cluster hasil K-Means dan K-Medoids. Pendekatan ini membantu mengidentifikasi konsistensi keanggotaan antar algoritma sekaligus menemukan karakteristik unik pada masing-masing cluster. Evaluasi semacam ini telah diterapkan pada analisis sosial-ekonomi maupun kesehatan di beberapa studi internasional, termasuk pemetaan indikator wilayah dan ketahanan sistem kesehatan [5], [21].

2.4.7. Interpretasi Hasil

Tahap terakhir adalah interpretasi hasil analisis yang diperoleh untuk keseluruhan tahapan penelitian. Interpretasi dilakukan pada hasil K-Means, K-Medoids dan perbandingan antara keduanya.



3. HASIL DAN PEMBAHASAN

3.1 Data

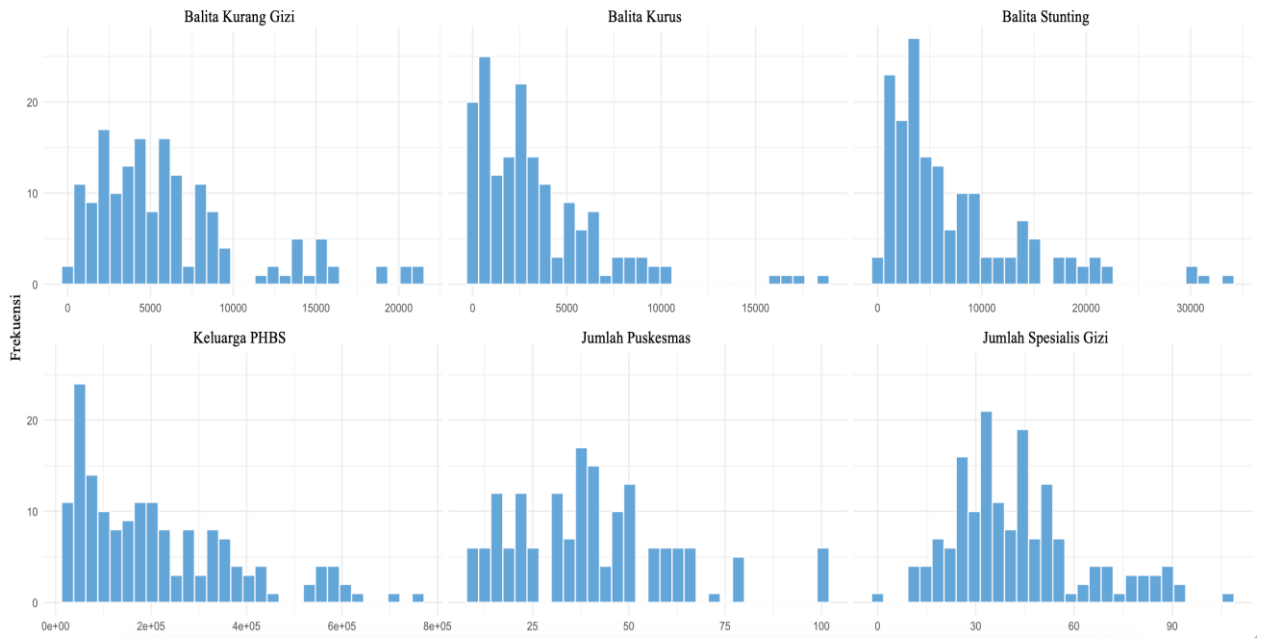
Dataset yang digunakan dalam penelitian ini terdiri dari sebelas sumber data resmi yang diperoleh dari portal Open Data Jabar dan Badan Pusat Statistik (BPS) Jawa Barat, yang mencakup indikator kesehatan dan sosial masyarakat di tingkat kabupaten/kota. Peubah kesehatan meliputi jumlah balita stunting, jumlah rumah tangga dengan perilaku hidup bersih dan sehat (PHBS), jumlah puskesmas, jumlah spesialis gizi, jumlah balita gizi buruk, jumlah balita kurus, dan persentase bayi dengan berat badan lahir rendah (BBLR). Sementara itu, peubah sosial mencakup Indeks Pembangunan Manusia (IPM), Indeks Pengetahuan, jumlah kepala keluarga dengan akses air minum layak, dan Indeks Kedalaman Kemiskinan. Tabel 1 merupakan daftar dari peubah yang dipilih berdasarkan relevansinya terhadap pengukuran kualitas kesehatan dan kesejahteraan sosial masyarakat, serta ketersediaannya dalam format terbuka yang dapat diakses publik melalui tautan resmi.

Tabel 1. Dataset yang digunakan

No	Judul Dataset	Tautan Dataset
1	Jumlah balita stunting	https://opendata.jabarprov.go.id/id/dataset/jumlah-balita-stunting-berdasarkan-kabupatenkota-di-jawa-barat
2	Jumlah Rumah Tangga dengan Perilaku Hidup Bersih dan Sehat (PHBS)	https://opendata.jabarprov.go.id/id/dataset/jumlah-rumah-tangga-dengan-perilaku-hidup-bersih-dan-sehat-phbs-berdasarkan-kabupatenkota-di-jawa-barat
3	Jumlah Puskesmas	https://opendata.jabarprov.go.id/id/dataset/jumlah-puskesmas-berdasarkan-kabupatenkota-di-jawa-barat
4	Jumlah spesialis gizi di puskesmas	https://opendata.jabarprov.go.id/id/dataset/jumlah-tenaga-kesehatan-di-puskesmas-berdasarkan-kategori-dan-jenis-kelamin-di-jawa-barat
5	Jumlah balita kurang gizi	https://opendata.jabarprov.go.id/id/dataset/jumlah-balita-berdasarkan-kategori-balita-gizi-buruk-di-jawa-barat
6	Jumlah balita kurus	https://opendata.jabarprov.go.id/id/dataset/jumlah-balita-berdasarkan-kategori-balita-gizi-buruk-di-jawa-barat
7	Persentase Bayi Berat Badan Lahir Rendah (BBLR)	https://opendata.jabarprov.go.id/id/dataset/persentase-bayi-berat-badan-lahir-rendah-bblr-berdasarkan-kabupatenkota-di-jawa-barat
8	Indeks Pembangunan Manusia (IPM)	https://jabar.bps.go.id/en/statistics-table/3/V25GaFNHaExaMnhITmIsWmRrUIJZeIjZYUc1SGR6MDkjMw==/indeks-pembangunan-manusia-menurut-kabupaten-kota-di-provinsi-jawa-barat--2022.html?year=2022
9	Indeks Pengetahuan	https://opendata.jabarprov.go.id/id/dataset/indeks-pengetahuan-berdasarkan-kabupatenkota-di-jawa-barat
10	Persentase Jumlah Kepala Keluarga dengan Akses Air Minum Berdasarkan Pelayanan Air Minum di Jawa Barat	https://opendata.jabarprov.go.id/id/dataset/jumlah-kepala-keluarga-dengan-akses-air-minum-berdasarkan-pelayanan-air-minum-di-jawa-barat
11	Indeks Kedalaman Kemiskinan	https://jabar.bps.go.id/id/statistics-table/2/ODYjMg==/indeks-kedalaman-kemiskinan-menurut-kabupaten-kota.html

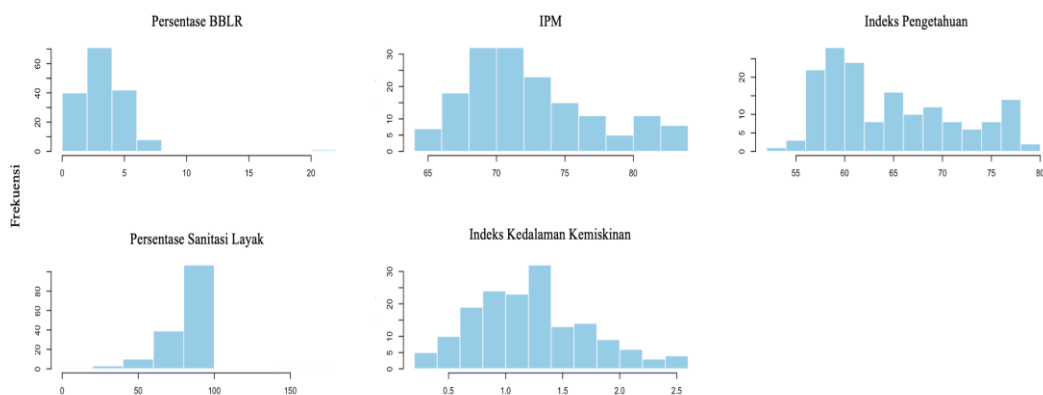
3.2 Eksplorasi Data

Eksplorasi dilakukan dengan melihat histogram setiap peubah, histogram pada peubah yang memiliki satuan jumlah seperti jumlah balita kurang gizi, jumlah balita kurus, jumlah balita stunting, jumlah rumah tangga PHBS, jumlah puskesmas dan jumlah spesialis gizi di puskesmas terdapat pada Gambar 2. Sedangkan untuk peubah persentase dan indeks seperti persentase BBLR, IPM, Indeks Pengetahuan, persentase sanitasi layak, dan indeks kedalaman kemiskinan terdapat pada Gambar 3. Hal ini dilakukan untuk melihat sebaran dari setiap peubah.



Gambar 2. Diagram distribusi peubah jumlah

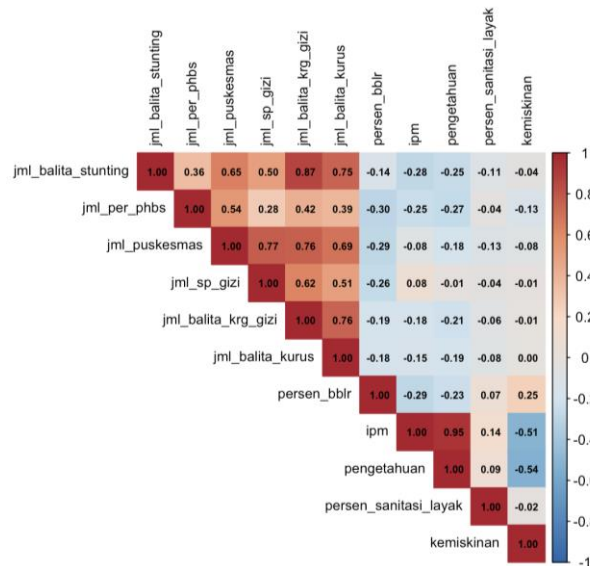
Gambar 2 menggambarkan distribusi peubah jumlah pada dataset dengan keragaman yang cukup besar antarindikator. Hal ini menunjukkan adanya pola sebaran yang berbeda di antara kabupaten/kota di Provinsi Jawa Barat. Peubah jumlah balita stunting menunjukkan distribusi yang cenderung condong ke kiri, dengan mayoritas kabupaten/kota memiliki nilai di bawah 10.000, namun terdapat beberapa wilayah dengan nilai ekstrem di atas 20.000. Pola serupa terlihat pada jumlah balita kurang gizi dan jumlah balita kurus, meskipun sebarannya lebih merata pada rentang rendah hingga menengah, dengan sejumlah kecil *outlier* pada nilai tinggi. Peubah jumlah puskesmas memiliki distribusi yang lebih terkonsentrasi di kisaran 20–50 unit per wilayah, namun terdapat beberapa daerah dengan jumlah fasilitas kesehatan yang jauh lebih tinggi, yang berpotensi mempengaruhi hasil cluster jika tidak dilakukan standarisasi. Peubah jumlah rumah tangga dengan PHBS dan jumlah spesialis gizi menunjukkan distribusi yang sangat bervariasi, dengan nilai berkisar dari sangat rendah hingga sangat tinggi, sehingga mengindikasikan adanya kesenjangan infrastruktur kesehatan antarwilayah. Perbedaan skala yang besar antarpeubah ini berimplikasi pada perlunya proses standarisasi data untuk mencegah dominasi peubah dengan rentang nilai besar terhadap hasil analisis. Selain itu, kehadiran nilai ekstrem pada beberapa peubah menegaskan pentingnya teknik reduksi dimensi seperti *Principal Component Analysis* (PCA), yang dapat memproyeksikan data ke ruang berdimensi lebih rendah dengan mempertahankan proporsi variansi terbesar.



Gambar 3. Diagram distribusi peubah persen

Gambar 3 menggambarkan distribusi peubah dengan skala persentase yang menunjukkan karakteristik yang berbeda dibandingkan peubah jumlah, baik dari segi rentang nilai maupun bentuk sebaran. Peubah persentase BBLR (bayi berat badan lahir rendah) memiliki distribusi yang condong ke kanan (*right-skewed*), dengan mayoritas kabupaten/kota berada di kisaran 2–8%, dan hanya sedikit wilayah yang memiliki angka di atas 10%. Peubah Indeks Pembangunan Manusia (IPM) relatif tersebar merata di antara nilai 65–80, menunjukkan bahwa kesenjangan capaian pembangunan manusia antarwilayah tidak terlalu ekstrem, meskipun terdapat perbedaan mencolok antara kelompok dengan IPM rendah dan tinggi. Peubah pengetahuan memiliki sebaran yang lebih lebar, mulai dari kisaran 55 hingga mendekati 80, yang mengindikasikan adanya disparitas signifikan dalam aspek literasi dan edukasi kesehatan di tingkat daerah. Pada peubah persentase sanitasi layak, sebagian besar wilayah mencatat nilai yang sangat tinggi (80–100%), namun terdapat beberapa

outlier di bawah 50%, yang berpotensi menjadi indikator keterbatasan akses sanitasi di wilayah tertentu. Sementara itu, peubah kemiskinan memiliki distribusi yang cenderung simetris dengan mayoritas daerah berada pada tingkat 1–2%, tetapi dengan beberapa wilayah ekstrem di bawah 0,5% dan di atas 2,5%. Pola distribusi yang relatif homogen pada beberapa peubah seperti IPM dan kemiskinan, serta keberadaan outlier pada sanitasi dan BBLR, menggarisbawahi pentingnya proses standarisasi sebelum dilakukan analisis multivariat. Standarisasi ini bertujuan untuk menghindari dominasi peubah dengan rentang nilai sempit maupun sangat tinggi dalam proses pengelompokan. Dengan memahami distribusi ini, langkah prapemrosesan dapat dirancang lebih tepat sehingga hasil analisis lanjutan seperti PCA dan cluster dapat merefleksikan kondisi faktual setiap kabupaten/kota secara proporsional.



Gambar 4. Heatmap Korelasi antar Peubah

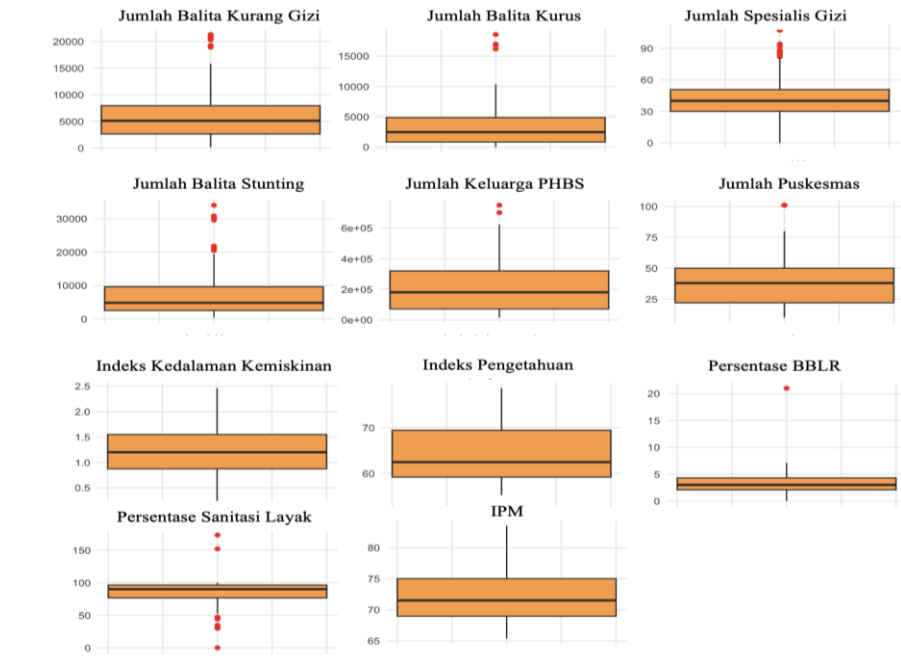
Korelasi atau hubungan antar peubah dilakukan dengan melihat visualisasi dari *heatmap* yang terbentuk. Warna pada *heatmap* yang semakin merah menunjukkan korelasi yang semakin kuat dan positif sedangkan *heatmap* yang memiliki warna yang semakin biru menunjukkan korelasi yang semakin kuat namun negatif. Berdasarkan hasil visualisasi *heatmap* korelasi Pearson pada Gambar 4, terlihat bahwa sebagian besar peubah jumlah memiliki korelasi positif yang cukup tinggi satu sama lain. Peubah jumlah balita stunting menunjukkan korelasi yang kuat dengan jumlah balita kurang gizi ($r = 0,87$) dan jumlah balita kurang gizi ($r = 0,75$), mengindikasikan adanya keterkaitan erat antara masalah stunting dan kondisi gizi buruk pada balita di tingkat daerah. Korelasi positif yang tinggi juga terlihat antara jumlah puskesmas dengan jumlah tenaga spesialis gizi ($r = 0,77$) serta jumlah balita kurang gizi ($r = 0,76$), yang mungkin merefleksikan keterpaduan distribusi fasilitas dan tenaga kesehatan dengan prevalensi kasus gizi buruk di wilayah Jawa Barat.

Pada peubah persentase, menunjukkan pola korelasi negatif terhadap indikator masalah gizi. Misalnya, persentase BBLR berkorelasi negatif dengan jumlah balita stunting ($r = -0,14$) dan jumlah balita kurang gizi ($r = -0,19$), walaupun kekuatannya relatif rendah. Indeks Pembangunan Manusia (IPM) dan indeks pengetahuan menunjukkan korelasi negatif yang cukup kuat dengan kemiskinan (masing-masing $r = -0,51$ dan $r = -0,54$). Korelasi yang sangat tinggi antara IPM dan indeks pengetahuan ($r = 0,95$). Selain itu, persentase sanitasi layak memiliki korelasi yang sangat rendah dengan hampir semua peubah lain (rentang $-0,13$ hingga $0,14$), yang mengindikasikan bahwa akses sanitasi mungkin dipengaruhi oleh faktor struktural yang berbeda dari indikator kesehatan balita atau peubah sosial ekonomi yang digunakan dalam penelitian ini. Temuan ini penting untuk proses *feature selection*, karena peubah dengan korelasi sangat tinggi (misalnya IPM dan pengetahuan) dapat dipertimbangkan untuk direduksi menggunakan metode seperti *Principal Component Analysis* (PCA) guna menghindari masalah multikolinearitas yang dapat mengganggu proses analisis cluster atau pemodelan prediktif (James *et al.*, 2021).

Secara umum, pola korelasi ini memberikan gambaran bahwa masalah gizi buruk pada balita cenderung berkaitan erat satu sama lain, sementara faktor pembangunan manusia, pendidikan, dan kemiskinan memiliki hubungan yang signifikan namun arah korelasi yang berlawanan. Informasi ini menjadi dasar penting untuk menentukan strategi transformasi data dan pemilihan peubah pada tahap analisis lanjutan.

3.3 Praproses dan Standarisasi Data

Tahap praproses dan standarisasi data yang telah dijelaskan pada bagian metodologi kemudian diaplikasikan pada dataset penelitian ini. Proses diawali dengan pengecekan terhadap data hilang (*missing values*), duplikasi, dan *outlier*, untuk memastikan bahwa data yang digunakan benar-benar bersih dan siap dianalisis. Pada penelitian ini tidak ditemukan adanya data hilang dan duplikasi sehingga imputasi untuk data hilang tidak dilakukan. Pengecekan terhadap *outlier* dilakukan dengan melihat hasil boxplot dari data penelitian yang terdapat pada Gambar 5. Setelah pengecekan outlier dilakukan, langkah berikutnya adalah melakukan standarisasi pada data.

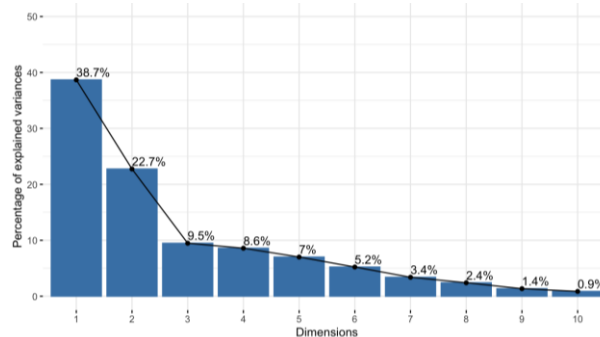


Gambar 5. *Boxplot* seluruh peubah

Hasil eksplorasi data melalui *boxplot* pada Gambar 5 menunjukkan adanya disparitas yang cukup signifikan antar kabupaten/kota pada hampir seluruh indikator kesehatan dan sosial. Indikator gizi (stunting, balita kurus, balita kurang gizi) didominasi oleh nilai rendah hingga menengah, namun dengan beberapa daerah yang menjadi outlier dengan angka kasus sangat tinggi. Hal serupa terlihat pada jumlah fasilitas kesehatan dan tenaga gizi, di mana beberapa daerah memiliki infrastruktur yang jauh lebih baik dibanding wilayah lain. Sementara itu, indikator sosial seperti IPM dan indeks pengetahuan relatif homogen dengan variasi yang tidak terlalu ekstrem, meskipun tetap terdapat daerah dengan capaian lebih tinggi dari rata-rata. Persentase sanitasi layak dan BBLR menunjukkan pola menarik, yaitu capaian rata-rata tinggi tetapi masih ada daerah yang menjadi outlier dengan kondisi sanitasi sangat rendah maupun BBLR sangat tinggi. Secara keseluruhan, pola ini menegaskan pentingnya proses standarisasi data serta penerapan PCA untuk mengurangi redundansi peubah dan memetakan perbedaan antarwilayah secara lebih komprehensif.

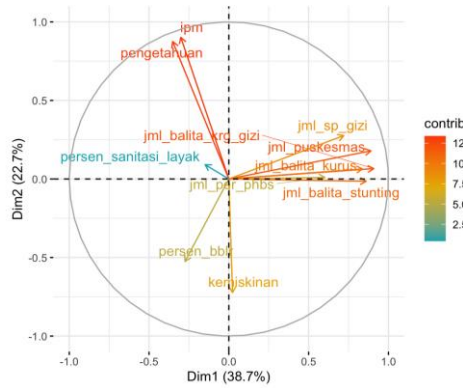
3.4 Principal Component Analysis (PCA)

Setelah data melalui tahap praproses dan standarisasi, langkah berikutnya adalah melakukan reduksi dimensi menggunakan PCA. Penerapan PCA bertujuan untuk menyederhanakan struktur data dengan tetap mempertahankan sebagian besar informasi yang terkandung, sehingga pola variasi antar wilayah dapat digambarkan secara lebih jelas.



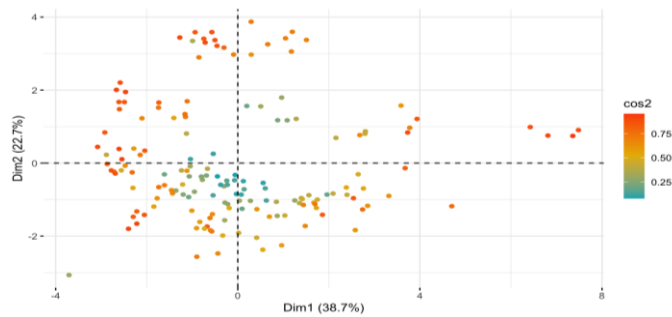
Gambar 6. *Scree Plot*

Hasil *scree plot* pada Gambar 6 menunjukkan bahwa komponen utama pertama (Dimensi 1) memiliki kontribusi sebesar 38,7% terhadap total keragaman data, sedangkan komponen utama kedua (Dimensi 2) menyumbang 22,7% varians tambahan. Dengan demikian, kedua komponen utama tersebut secara kumulatif mampu menjelaskan 61,4% varians total. Nilai ini dianggap memadai karena dalam analisis data sosial dan kesehatan, proporsi varians kumulatif di atas 60% umumnya telah memenuhi standar kelayakan untuk representasi data multivariat [17][23]. Hal ini sejalan dengan penelitian terdahulu yang menggunakan PCA untuk indikator kesehatan masyarakat, di mana ragam yang terjaga pada kisaran 60–70% sudah mampu memberikan gambaran pola yang cukup jelas tanpa menimbulkan distorsi interpretasi [18]. Setelah komponen kedua, kontribusi varians masing-masing dimensi turun secara signifikan hingga berada di bawah 10%, sehingga komponen tersebut dianggap kurang informatif untuk interpretasi awal.



Gambar 7. Biplot

Visualisasi biplot peubah pada Gambar 7 menunjukkan perbedaan kontribusi setiap peubah terhadap pembentukan dua komponen utama. Peubah jumlah balita stunting, jumlah puskesmas, jumlah spesialis gizi, jumlah balita kurus, dan jumlah balita kurang gizi memiliki korelasi positif yang tinggi dengan Dimensi 1, yang mengindikasikan bahwa dimensi ini merepresentasikan aspek ketersediaan fasilitas kesehatan serta beban permasalahan gizi. Sementara itu, peubah IPM dan indeks pengetahuan berkontribusi dominan terhadap Dimensi 2, sehingga dimensi ini lebih merefleksikan aspek kualitas sumber daya manusia. Peubah persentase sanitasi layak berada relatif dekat dengan pusat koordinat, menandakan kontribusinya terhadap kedua dimensi utama relatif rendah.

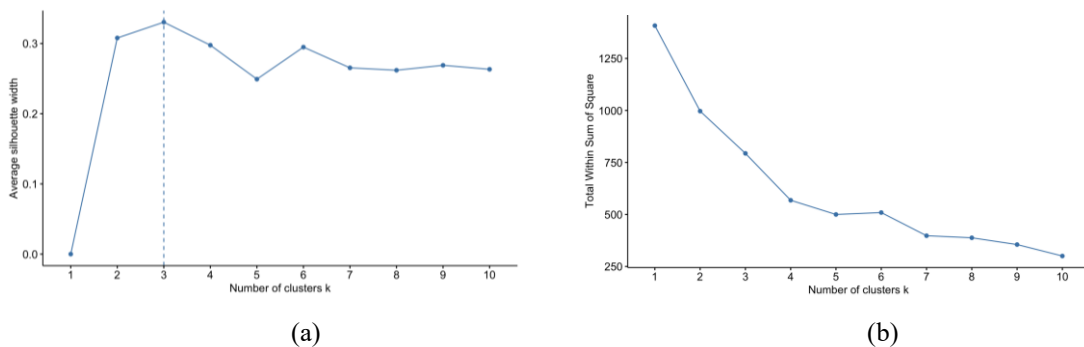


Gambar 8. Plot Individu Hasil PCA

Plot individu hasil PCA pada Gambar 8 menunjukkan distribusi kabupaten/kota di Jawa Barat dalam ruang dua dimensi yang terbentuk. Titik-titik dengan nilai \cos^2 tinggi (berwarna merah) menunjukkan representasi yang baik pada dua komponen utama ini, sedangkan titik-titik dengan nilai \cos^2 rendah (berwarna hijau kebiruan) kurang terwakili secara optimal. Daerah yang terletak pada sisi kanan plot cenderung memiliki nilai tinggi pada indikator jumlah kasus stunting, gizi buruk, balita kurus, serta jumlah fasilitas kesehatan. Sebaliknya, daerah pada sisi atas plot memiliki skor relatif tinggi pada indikator kualitas sumber daya manusia seperti IPM dan tingkat pengetahuan masyarakat. Secara umum, Dimensi 1 dapat diinterpretasikan sebagai dimensi “beban masalah gizi dan ketersediaan fasilitas kesehatan”, sedangkan Dimensi 2 dapat diinterpretasikan sebagai dimensi “kualitas sumber daya manusia”.

3.5 K-Means

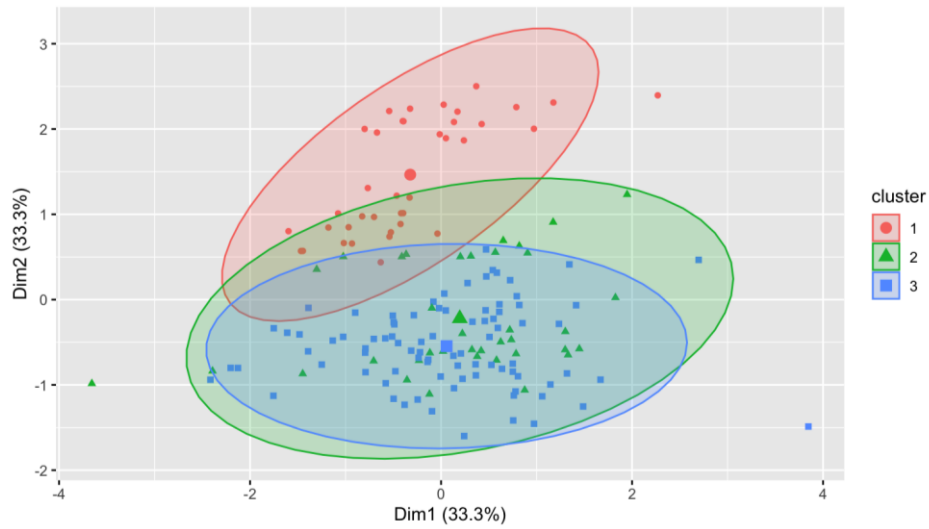
Setelah data direduksi dimensinya menggunakan PCA, tahap selanjutnya adalah penerapan algoritma K-Means untuk mengelompokkan kabupaten/kota di Jawa Barat berdasarkan indikator kesehatan dan sosial. Proses pengclustoran ini bertujuan untuk mengidentifikasi wilayah dengan karakteristik serupa sehingga dapat diperoleh gambaran pola distribusi kondisi kesehatan dan kesejahteraan masyarakat. Untuk menentukan jumlah cluster yang paling optimal, digunakan dua pendekatan evaluasi internal yaitu *average silhouette coefficient* dan metode *elbow*.



Gambar 9. Silhouette Plot (a) dan Elbow Plot (b) K-Means

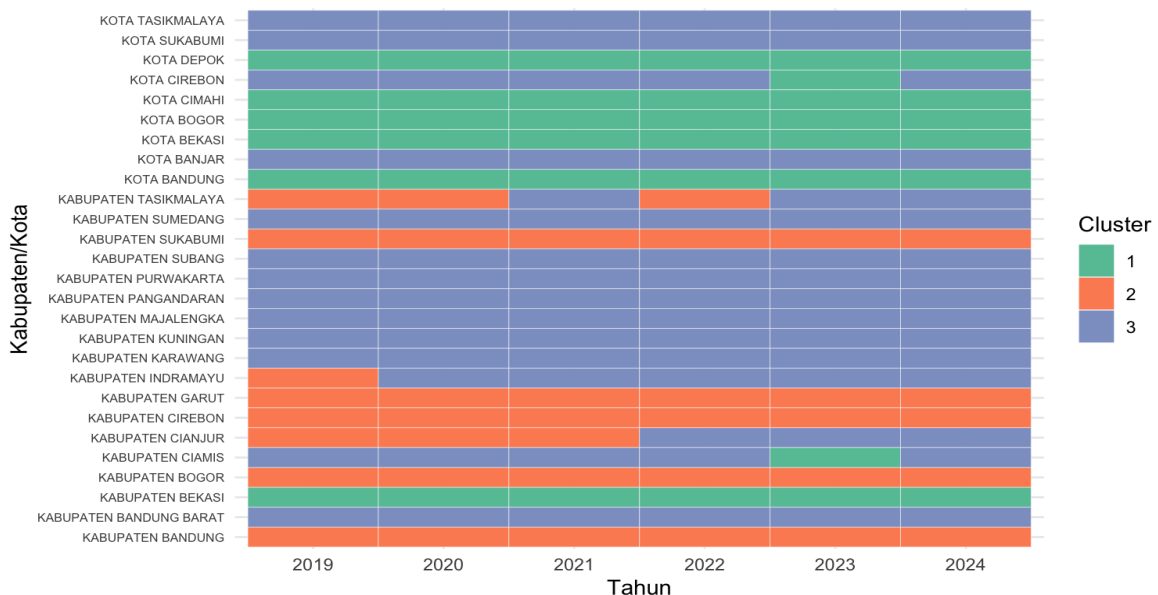


Hasil *silhouette plot* pada Gambar 9 (a) menunjukkan bahwa nilai rata-rata tertinggi diperoleh pada $k = 3$ dengan skor sebesar 0,31. Nilai ini memang tidak mencapai kategori tinggi ($>0,5$), namun masih dapat dianggap memadai dalam konteks data sosial-kesehatan yang memiliki heterogenitas tinggi antarwilayah. Skor tersebut mengindikasikan bahwa pembentukan tiga cluster mampu memberikan pemisahan yang relatif jelas antarwilayah, meskipun masih terdapat tumpang tindih pada beberapa unit observasi. Sementara itu, hasil metode elbow pada Gambar 9 (b) memperlihatkan penurunan tajam nilai *within-cluster sum of squares* (WSS) dari $k=1$ hingga $k=3$, kemudian melandai setelah titik tersebut. Pola ini menunjukkan adanya elbow yang konsisten pada $k=3$, yang menandakan bahwa penambahan jumlah cluster setelah titik tersebut tidak memberikan pengurangan WSS yang signifikan. Dengan demikian, hasil kedua metode validasi ini sama-sama mendukung pemilihan jumlah cluster optimal sebanyak tiga cluster. Hal ini menegaskan bahwa tiga cluster merupakan representasi yang paling efisien untuk mengelompokkan kabupaten/kota di Jawa Barat berdasarkan indikator kesehatan dan sosial.



Gambar 10. Plot K-Means

Gambar 10 menunjukkan hasil pengelompokan menggunakan algoritma K-Means berdasarkan dua komponen utama (PC1 dan PC2) dari analisis PCA. Terlihat bahwa data terbagi ke dalam tiga cluster yang ditandai dengan warna berbeda, yaitu cluster 1 (merah), cluster 2 (hijau), dan cluster 3 (biru). Secara umum, cluster 1 tampak terpisah lebih jelas pada sisi atas grafik, menunjukkan adanya karakteristik khusus yang membedakan anggota cluster ini dari dua cluster lainnya. Sementara itu, cluster 2 dan cluster 3 menunjukkan area yang saling tumpang tindih, menandakan adanya kemiripan karakteristik antar anggota kedua cluster tersebut. Namun demikian, masih terlihat kecenderungan penyebaran titik-titik dalam masing-masing cluster, yang mengindikasikan bahwa metode K-Means mampu menangkap pola variasi utama dalam data meskipun terdapat tumpang tindih antar cluster.

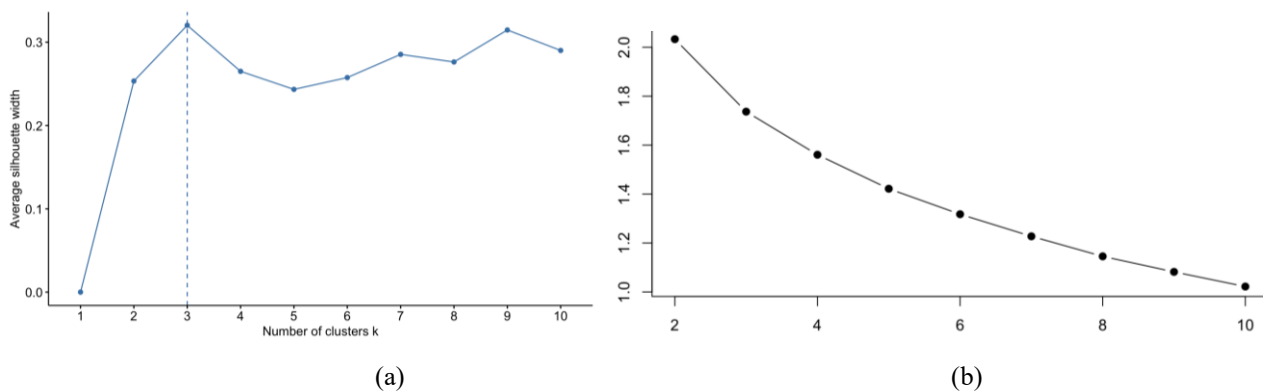


Gambar 11. Distribusi Keanggotaan Cluster menggunakan K-Means

Gambar 11 menggambarkan dinamika keanggotaan cluster kabupaten/kota di Jawa Barat selama periode 2019 hingga 2024 menggunakan K-Means. Secara umum, Cluster 3 (biru) mendominasi mayoritas wilayah pada hampir seluruh tahun pengamatan, yang menunjukkan konsistensi karakteristik sosial dan kesehatan pada kelompok ini. Cluster 1 (merah) terlihat lebih terbatas jumlahnya, dengan kecenderungan ditempati oleh kabupaten yang memiliki permasalahan kesehatan relatif lebih berat, seperti tingginya prevalensi gizi buruk atau keterbatasan fasilitas layanan dasar. Sementara itu, Cluster 2 (hijau) merepresentasikan wilayah dengan karakteristik menengah atau transisi, yang ditunjukkan oleh pergeseran keanggotaan antar tahun pada sejumlah kabupaten/kota. Perubahan posisi beberapa wilayah dari satu cluster ke cluster lain mengindikasikan adanya dinamika capaian indikator sosial-kesehatan yang cukup signifikan, baik berupa perbaikan maupun penurunan kondisi. Fenomena ini memperlihatkan bahwa meskipun terdapat kelompok wilayah yang stabil dalam jangka waktu panjang, sebagian kabupaten/kota tetap menunjukkan kerentanan terhadap perubahan kebijakan, intervensi program, maupun faktor eksternal lain seperti krisis ekonomi dan pandemi. Dengan demikian, analisis longitudinal ini menegaskan pentingnya pemantauan berkelanjutan untuk memastikan pemerataan capaian pembangunan kesehatan dan sosial di seluruh wilayah Jawa Barat.”

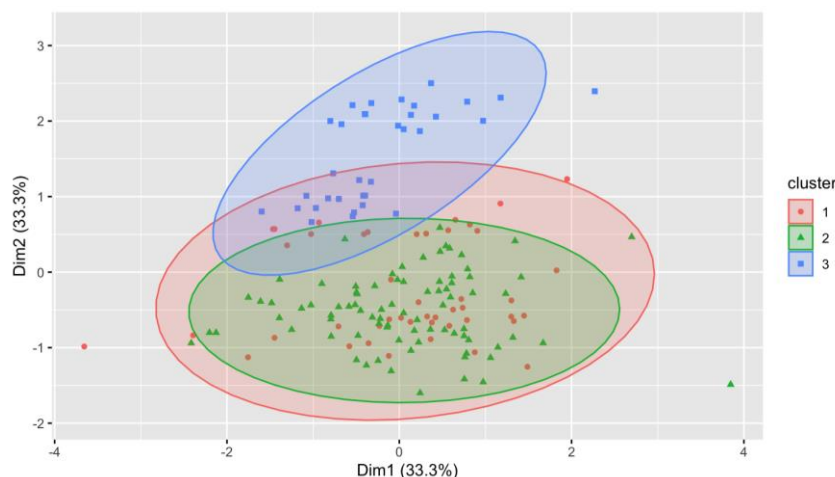
3.6 K-Medoids

Setelah penerapan algoritma K-Means, tahap berikutnya adalah analisis menggunakan algoritma K-Medoids. Metode ini dipilih karena memiliki keunggulan dalam menangani penciran (*outlier*) dengan menggunakan medoid, yaitu objek aktual yang paling representatif sebagai pusat cluster. Penerapan K-Medoids bertujuan untuk memvalidasi hasil pengclustering sebelumnya sekaligus melihat konsistensi pola yang terbentuk dalam data indikator kesehatan dan sosial di Jawa Barat.



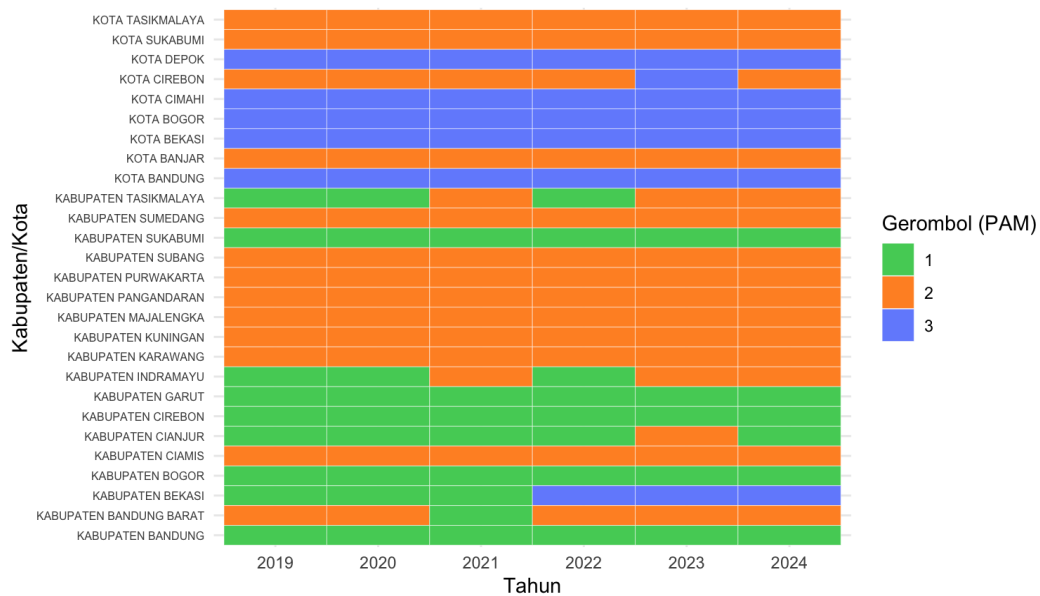
Gambar 12. Silhouette Plot (a) dan Elbow Plot (b) K-Medoids

Berdasarkan *silhouette plot* pada gambar 12 (a) nilai rata-rata tertinggi diperoleh pada $k = 3$ dengan skor sekitar 0,34, yang menunjukkan bahwa pembentukan tiga cluster memberikan kualitas pemisahan terbaik dibandingkan jumlah cluster lainnya. Meskipun nilainya tidak tergolong tinggi, skor ini cukup memadai dalam konteks data sosial dan kesehatan yang bersifat kompleks serta heterogen. Setelah $k=3$, nilai *silhouette coefficient* cenderung menurun, menandakan bahwa penambahan jumlah cluster tidak meningkatkan kualitas pengelompokan secara signifikan. Temuan ini diperkuat oleh hasil metode elbow pada *within-cluster dissimilarity* yang ditunjukkan pada gambar 12 (b), di mana terlihat penurunan tajam dari $k=2$ ke $k=3$, kemudian menurun secara lebih landai setelah titik tersebut. Pola ini mengindikasikan bahwa tiga cluster merupakan jumlah optimal untuk mewakili struktur data, dengan keseimbangan antara homogenitas internal dan pemisahan antar kelompok. Dengan demikian, hasil K-Medoids sejalan dengan temuan pada K-Means, namun lebih *robust* terhadap *outlier* karena pusat cluster ditentukan oleh objek aktual yang paling representatif dalam setiap kelompok.



Gambar 13. Plot K-Medoids

Gambar tersebut memperlihatkan hasil pengelompokan menggunakan algoritma K-Medoids (PAM) berdasarkan dua komponen utama PCA. Data terbagi menjadi tiga cluster yang ditandai dengan warna berbeda: cluster 1 (merah), cluster 2 (hijau), dan cluster 3 (biru). Tampak bahwa cluster 3 relatif lebih jelas terpisah di bagian atas grafik, menunjukkan adanya pola khusus yang konsisten dan cukup berbeda dari dua cluster lainnya. Sementara itu, cluster 1 dan cluster 2 memiliki area yang lebih tumpang tindih di bagian tengah plot, yang mengindikasikan adanya kesamaan karakteristik antar anggota kedua cluster tersebut. Keunggulan metode K-Medoids terlihat dari bentuk batas cluster yang lebih robust terhadap outlier dibandingkan K-Means, sehingga distribusi anggota tiap cluster tampak lebih stabil. Dengan demikian, hasil ini menunjukkan bahwa K-Medoids mampu menangkap pola heterogenitas data secara lebih seimbang, terutama untuk memisahkan kelompok yang memiliki karakteristik sangat berbeda seperti pada cluster 3.



Gambar 14. Distribusi Keanggotaan Cluster menggunakan K-Medoids

Gambar 14 memperlihatkan dinamika keanggotaan cluster kabupaten/kota di Jawa Barat selama periode 2019 hingga 2024 berdasarkan algoritma K-Medoids. Secara umum, Cluster 2 (jingga) mendominasi mayoritas wilayah pada sebagian besar tahun, menandakan bahwa karakteristik menengah menjadi representasi yang cukup umum di antara kabupaten/kota. Cluster 1 (hijau) cenderung stabil pada sejumlah wilayah tertentu, khususnya daerah yang memiliki kondisi sosial-ekonomi dan kesehatan relatif lebih baik, dengan konsistensi keanggotaan sepanjang periode pengamatan. Sementara itu, Cluster 3 (ungu) muncul secara lebih terbatas dan hanya pada tahun-tahun tertentu, yang mengindikasikan adanya kelompok wilayah dengan kondisi khusus atau perubahan signifikan pada indikator sosial-kesehatan. Pergeseran antar cluster yang dialami oleh beberapa kabupaten/kota dari tahun ke tahun menunjukkan adanya dinamika capaian pembangunan kesehatan dan sosial. Hal ini mencerminkan bahwa meskipun sebagian wilayah konsisten berada pada satu kategori, terdapat pula wilayah yang rentan mengalami pergeseran status akibat intervensi kebijakan, faktor lingkungan, maupun kejadian eksternal seperti pandemi. Dengan demikian, hasil pengelompokan temporal ini menggarisbawahi pentingnya pemantauan berkelanjutan untuk memastikan tercapainya pemerataan kondisi kesehatan dan sosial antarwilayah di Jawa Barat.

3.7 Perbandingan K-Means dan K-Medoids

Hasil pengelompokan dengan K-Means dan K-Medoids menunjukkan perbedaan baik dari segi pola distribusi maupun kestabilan keanggotaan cluster antar periode waktu. Pada metode K-Means, distribusi cluster cenderung tumpang tindih, terutama antara cluster 1, 2, dan 3, sehingga batas antar cluster menjadi kurang tegas. Hal ini menggambarkan bahwa K-Means lebih sensitif terhadap *outlier* dan keragaman data yang tinggi, sehingga anggota dengan karakteristik yang agak berbeda tetap dipaksakan masuk ke dalam satu cluster. Akibatnya, pola keanggotaan pada K-Means cenderung lebih bervariasi antar tahun. Beberapa kabupaten/kota, seperti Tasikmalaya, Garut, dan Bandung, terlihat berpindah cluster cukup sering sepanjang periode 2019–2024. Kondisi ini mengindikasikan bahwa K-Means menghasilkan pemisahan cluster yang lebih adaptif terhadap dinamika perubahan indikator kesehatan dan sosial, tetapi sekaligus rentan menimbulkan instabilitas dari waktu ke waktu.

Sebaliknya, hasil K-Medoids memperlihatkan pemisahan cluster yang lebih jelas dan stabil. Pada visualisasi hasil, cluster 3 muncul sebagai kelompok berbeda dengan karakteristik yang lebih homogen dibandingkan cluster lain, sedangkan cluster 1 dan 2 masih menunjukkan area yang tumpang tindih, meski dengan struktur pengelompokan yang lebih tegas dibandingkan K-Means. Distribusi keanggotaan antar tahun juga tampak lebih konsisten, di mana sebagian besar kabupaten/kota, seperti Majalengka, Kuningan, dan Indramayu, tetap berada dalam cluster yang sama selama periode 2019–2024. Meski terdapat perpindahan pada beberapa daerah, jumlahnya relatif lebih sedikit dibandingkan hasil



K-Means. Hal ini menegaskan bahwa K-Medoids lebih *robust* terhadap *outlier* maupun fluktuasi kecil pada data, sehingga memberikan hasil pengelompokan yang lebih stabil dan dapat diandalkan untuk analisis longitudinal.

4. KESIMPULAN

Penelitian ini menganalisis indikator kesehatan dan sosial masyarakat di Provinsi Jawa Barat melalui pendekatan *Principal Component Analysis* (PCA) sebagai reduksi dimensi, diikuti penerapan algoritma K-Means dan K-Medoids untuk pengelompokan wilayah. Hasil PCA menunjukkan bahwa tiga komponen utama mampu merepresentasikan sebagian besar variasi data dan menghasilkan tiga cluster optimal. Algoritma K-Means cenderung menampilkan dinamika antar tahun, dengan sejumlah kabupaten/kota mengalami perpindahan cluster akibat fluktuasi indikator kesehatan dan sosial, sedangkan K-Medoids menghasilkan struktur yang lebih konsisten dan robust sepanjang periode 2019–2024. Pada hasil tahun 2024, K-Means mengelompokkan Kota Depok, Kota Bekasi, dan Kota Cimahi ke dalam cluster dengan infrastruktur kesehatan relatif baik, sementara Kabupaten Bandung, Garut, dan Tasikmalaya masuk ke dalam cluster dengan tantangan signifikan, serta Cianjur, Subang, dan Purwakarta berada pada kategori menengah. Sebaliknya, K-Medoids memperlihatkan konsistensi dengan Majalengka, Kuningan, dan Indramayu tetap berada dalam cluster dengan kondisi lebih maju, Bandung, Garut, dan Tasikmalaya pada cluster dengan kerentanan tinggi, serta Depok, Cimahi, dan Bogor membentuk cluster tersendiri yang relatif homogen. Perbandingan ini mengindikasikan bahwa K-Means lebih sensitif dalam menangkap dinamika temporal, sedangkan K-Medoids lebih unggul dalam stabilitas jangka panjang. Dengan demikian, hasil penelitian ini memberikan dasar empiris bagi Pemerintah Provinsi Jawa Barat dalam merumuskan intervensi berbasis bukti yang disesuaikan dengan karakteristik masing-masing cluster, baik berupa pemerataan fasilitas kesehatan, penguatan layanan primer, maupun peningkatan kualitas layanan spesialis di wilayah dengan capaian lebih tinggi.

REFERENCES

- [1] U. Nations, "Transforming Our World: The 2030 Agenda for Sustainable Development." 2015. [Online]. Available: <https://sdgs.un.org/2030agenda>
- [2] U. N. D. Programme (UNDP), "Indonesia Human Development Report 2022." Jakarta, Indonesia, 2022. [Online]. Available: <https://www.undp.org/indonesia>
- [3] Badan Pusat Statistik Provinsi Jawa Barat, *Statistik Kesehatan Provinsi Jawa Barat 2023*. Bandung, Indonesia: BPS, 2023.
- [4] Kementerian Kesehatan Republik Indonesia, *Laporan Kinerja Kementerian Kesehatan 2021*. Jakarta, Indonesia: Kemenkes RI, 2021.
- [5] E. K. Korir, L. Cheruiyot, and D. K. Sang, "Comparative clustering and visualization of socioeconomic indicators using PCA, hierarchical, and K-means techniques," *Socioecon. Plann. Sci.*, vol. 93, p. 101423, 2024, doi: 10.1016/j.seps.2024.101423.
- [6] C. X. Gao, "An overview of clustering methods with guidelines for choosing among them," *Inf. Fusion*, vol. 96, pp. 35–49, 2023, doi: 10.1016/j.inffus.2023.04.005.
- [7] Y. Sheng, C. Wu, J. Wang, and L. Zhou, "Augmenting K-means clustering with qualitative data to analyze engagement with digital health technology," *J. Med. Internet Res.*, vol. 26, no. 1, p. e46287, 2024, doi: 10.2196/46287.
- [8] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 1967, pp. 281–297. [Online]. Available: <https://projecteuclid.org/euclid.bsm/1200512992>
- [9] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 1979, doi: 10.2307/2346830.
- [10] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 1990. doi: 10.1002/9780470316801.
- [11] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009, doi: 10.1016/j.eswa.2008.01.039.
- [12] H. Wibowo, R. Hidayat, and N. Safitri, "Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data sosial-ekonomi di Indonesia," *J. Teknol. Inf. Dan Ilmu Komput. JTIK*, vol. 6, no. 1, pp. 45–54, 2019.
- [13] A. Ahmed, M. Islam, and M. Rahman, "Comparative study of K-means and K-medoids algorithm for healthcare data," *Int. J. Comput. Appl.*, vol. 174, no. 21, pp. 1–5, 2021.
- [14] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
- [15] A. M. Leis, K. D. Johnson, and R. C. Riedel, "K-medoids clustering of hospital admission characteristics: Robustness against outliers," *BMC Med. Res. Methodol.*, vol. 23, no. 1, p. 56, 2023, doi: 10.1186/s12874-023-02056-0.
- [16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012. doi: 10.1016/C2009-0-61819-5.
- [17] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A*, vol. 374, no. 2065, p. 20150202, 2016, doi: 10.1098/rsta.2015.0202.
- [18] P. Nugroho and A. Kurniawan, "Pemanfaatan Principal Component Analysis untuk pemetaan indikator kesehatan di Indonesia," *J. Kesehat. Masy.*, vol. 15, no. 2, pp. 87–95, 2020.
- [19] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: Principal component analysis," *Nat. Methods*, vol. 14, no. 7, pp. 641–642, 2017, doi: 10.1038/nmeth.4346.
- [20] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.



- [21] T. Bulut, “Classification of resilience of Turkish health system to shocks using PCA and PAM clustering,” *J. Soc. Policy*, vol. 45, no. 2, pp. 125–138, 2025, doi: 10.1017/jsp.2025.2824.
- [22] C. Sheng, R. Kustra, and D. Chicco, “Comparative analysis of clustering techniques using validation metrics: A study on cognitive features from the CLSA,” *Hum. Brain Mapp.*, vol. 46, no. 5, pp. 2010–2022, 2025, doi: 10.1002/hbm.25912.
- [23] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.
- [24] A. Mageri, G. Barchitta, A. Favara, and A. Agodi, “The application of clustering on principal components for dietary patterns derivation,” *Nutrients*, vol. 15, no. 1, p. 195, 2022, doi: 10.3390/nu15010195.