

Analisis Cakupan Imunisasi Balita di Indonesia menggunakan Algoritma *K-Means* (Data Kesejahteraan Rakyat 2025)

Analysis of Immunization Coverage Among Toddlers in Indonesia using the K-Means Algorithm based on 2025 People's Welfare Data

¹Maria Angelina Luisa Makin, ²Aditia Yudhistira*

¹Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

²Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

^{1,2}Jl. Zainal Abidin Pagaralam No.9-11 Labuhan Ratu, Bandar Lampung, Indonesia

e-mail: maria_angelina_luisa_makin@teknokrat.ac.id, aditiayudhistira@teknokrat.ac.id

(received: 27 February 2026, revised: 7 March 2026, accepted: 8 March 2026)

Abstrak

Kesenjangan cakupan imunisasi balita di berbagai wilayah Indonesia menunjukkan perlunya pendekatan analisis yang mampu menggambarkan perbedaan karakteristik wilayah secara lebih mendalam. Penelitian ini bertujuan mengelompokkan cakupan imunisasi balita menggunakan algoritma *K-Means* berbasis Data Kesejahteraan Rakyat Indonesia tahun 2025 yang dipublikasikan oleh Badan Pusat Statistik. Variabel yang dianalisis meliputi riwayat pemberian imunisasi, jenis imunisasi, tempat pemberian imunisasi, serta pemberi imunisasi. Pengolahan data dilakukan menggunakan *Python* melalui platform Google Colab. Penentuan jumlah cluster menggunakan *Elbow Method* menghasilkan tiga kelompok, dengan nilai *Silhouette Score* sebesar 0,5086 yang menunjukkan kualitas pengelompokan cukup baik. Hasil penelitian menunjukkan bahwa cluster berlabel Rendah justru memiliki capaian imunisasi tertinggi (95,68%), yang mengindikasikan bahwa penamaan cluster tidak merepresentasikan tingkat capaian secara linier, melainkan mencerminkan perbedaan karakteristik operasional dan distribusi layanan imunisasi antarwilayah. Sementara itu, cluster Sedang menunjukkan capaian terendah (63,82%), dan cluster Tinggi berada pada tingkat menengah (92,23%). Analisis lebih lanjut mengindikasikan bahwa jenis imunisasi dan status pernah diberi imunisasi merupakan variabel yang paling berpengaruh dalam pembentukan cluster. Dengan tingkat kualitas *clustering* yang berada pada kategori cukup baik, metode *K-Means* dinilai mampu mengidentifikasi pola cakupan imunisasi secara memadai untuk analisis kebijakan berbasis wilayah. Temuan ini menunjukkan bahwa pendekatan *clustering* mampu mengungkap pola cakupan imunisasi yang tidak terlihat melalui analisis statistik konvensional dan dapat dimanfaatkan sebagai dasar penyusunan kebijakan imunisasi yang lebih terarah dan berbasis data.

Kata kunci: cakupan imunisasi balita, *clustering*, data kesejahteraan rakyat, *data mining*, *k-means*

Abstract

Disparities in toddler immunization coverage across different regions of Indonesia indicate the need for an analytical approach capable of capturing regional characteristics in greater depth. This study aims to cluster toddler immunization coverage using the *K-Means* algorithm based on the 2025 Indonesian People's Welfare Data published by Statistics Indonesia. The variables analyzed include immunization history, types of immunization, place of immunization, and immunization providers. Data processing was conducted using *Python* through the Google Colab platform. The determination of the optimal number of clusters using the *Elbow Method* resulted in three clusters, with a *Silhouette Score* of 0.5086, indicating a moderately good clustering quality. The results show that the cluster labeled Low surprisingly exhibits the highest immunization coverage (95.68%), suggesting that the cluster labels do not represent coverage levels in a linear manner but instead reflect differences in operational characteristics and the distribution of immunization services across regions. Meanwhile, the Medium cluster shows the lowest coverage (63.82%), and the High cluster falls at an intermediate level (92.23%). Further analysis indicates that the type of immunization and immunization history are the most influential variables in cluster formation. With clustering quality categorized as moderately

<http://sistemasi.ftik.unisi.ac.id>

good, the *K-Means* method is considered capable of adequately identifying immunization coverage patterns for region-based policy analysis. These findings demonstrate that a clustering approach can reveal immunization coverage patterns that are not captured through conventional statistical analysis and can serve as a basis for more targeted and data-driven immunization policy formulation.

Keywords: clustering, data mining, k-means, toddler immunization coverage, welfare data

1 Pendahuluan

Imunisasi dasar merupakan bagian penting dari pelayanan kesehatan yang berfungsi efektif dalam menurunkan tingkat kesakitan, kecacatan, serta angka kematian akibat penyakit-penyakit yang dapat dicegah melalui imunisasi[1]. Keberhasilan pelaksanaan imunisasi dasar di Indonesia sangat dipengaruhi oleh sinergi regulasi pemerintah, dukungan organisasi yang kuat, dan kualitas sumber daya manusia yang profesional, yang mencerminkan efektivitas sistem pelayanan kesehatan secara keseluruhan[2]. Cakupan imunisasi di Indonesia menunjukkan ketimpangan antarwilayah, yang dipengaruhi oleh beragam faktor sosial-demografis serta perbedaan karakteristik geografis dan ketersediaan layanan Kesehatan[3]. Berbagai strategi pelayanan kesehatan primer di berbagai negara diperlukan untuk mengatasi tantangan cakupan imunisasi balita, termasuk penguatan sistem primer dan kampanye kesehatan yang responsif terhadap kondisi lokal[4]. Ketimpangan cakupan imunisasi berpotensi membuka peluang timbulnya kembali penyakit menular serius yang seharusnya dapat dicegah melalui vaksinasi yang efektif dan cakupan yang memadai[5].

Data Kesejahteraan Rakyat Indonesia tahun 2025 yang dipublikasikan oleh Badan Pusat Statistik menyediakan informasi terkait riwayat pemberian imunisasi, jenis imunisasi, tempat pemberian imunisasi, pemberi imunisasi, serta alasan balita tidak mendapatkan imunisasi[6]. Studi sebelumnya juga menunjukkan bahwa cakupan imunisasi balita dapat mengalami perubahan signifikan akibat faktor sosial, seperti kondisi pandemi, yang berdampak pada akses dan pemanfaatan layanan imunisasi di berbagai daerah[7]. Meskipun data ini cukup lengkap, pemanfaatannya masih terbatas pada analisis statistik sederhana sehingga belum mampu menggambarkan pola dan karakteristik wilayah secara mendalam. Permasalahan yang muncul adalah bagaimana memanfaatkan data tersebut secara optimal sehingga dapat menghasilkan informasi yang relevan untuk mendukung pengambilan keputusan berbasis data dalam program imunisasi.

Pendekatan kuantitatif yang mengintegrasikan teknik *machine learning* untuk memahami pola cakupan imunisasi telah diterapkan dalam berbagai studi kesehatan masyarakat, menunjukkan bahwa metode seperti *clustering* mampu mengeksplorasi data imunisasi secara multidimensional[8]. Penelitian sebelumnya menunjukkan bahwa penerapan algoritma *K-Means* pada data vaksinasi mampu mengelompokkan wilayah berdasarkan pola cakupan imunisasi serta mengidentifikasi kelompok populasi dengan tingkat partisipasi vaksinasi yang berbeda[9]. Studi lain di tingkat internasional juga melaporkan bahwa metode *K-Means* efektif digunakan dalam analisis data imunisasi berbasis wilayah dan dapat mendukung perumusan kebijakan kesehatan masyarakat yang lebih terarah[10]. Pendekatan *data mining* digunakan untuk mengekstraksi informasi bermakna serta mengidentifikasi pola tersembunyi dari kumpulan data dalam jumlah besar[11]. Salah satu teknik pengelompokan yang paling banyak digunakan adalah algoritma *K-Means*, yaitu metode yang mengelompokkan data ke dalam sejumlah cluster berdasarkan tingkat kemiripan antar atribut. Dalam bidang kesehatan, algoritma ini telah digunakan untuk analisis dan pemetaan data, termasuk pada pengolahan data imunisasi, sehingga mampu mengidentifikasi kelompok wilayah dengan karakteristik cakupan yang berbeda dan mendukung penentuan prioritas intervensi secara lebih tepat. Meskipun demikian, pemanfaatan algoritma *K-Means* yang secara khusus mengintegrasikan Data Kesejahteraan Rakyat Indonesia tahun 2025 dengan variabel imunisasi yang lengkap masih relatif terbatas[12]. Selain itu, sebagian besar penelitian sebelumnya lebih berfokus pada analisis data imunisasi pada skala wilayah tertentu atau menggunakan jumlah variabel yang terbatas, sehingga gambaran pola cakupan imunisasi secara nasional dengan mempertimbangkan berbagai faktor terkait pelaksanaan imunisasi masih belum banyak dikaji secara komprehensif.

Berdasarkan kondisi tersebut, diperlukan perumusan masalah yang jelas agar analisis yang dilakukan dapat terarah. Oleh karena itu, rumusan masalah dalam penelitian ini adalah bagaimana mengelompokkan cakupan imunisasi balita di Indonesia menggunakan algoritma *K-Means* berdasarkan Data Kesejahteraan Rakyat Indonesia tahun 2025 serta bagaimana karakteristik setiap

<http://sistemasi.ftik.unisi.ac.id>

cluster yang terbentuk berdasarkan variabel riwayat pemberian imunisasi, jenis imunisasi, tempat pemberian imunisasi, pemberi imunisasi, dan alasan balita tidak memperoleh imunisasi. Perumusan masalah tersebut menjadi dasar dalam menentukan pendekatan analisis yang digunakan untuk menggali pola data imunisasi secara lebih mendalam.

Kebaruan penelitian ini terletak pada pemanfaatan Data Kesejahteraan Rakyat Indonesia tahun 2025 sebagai sumber data utama untuk analisis cakupan imunisasi balita pada tingkat nasional dengan pendekatan *data mining*. Berbeda dengan beberapa penelitian sebelumnya yang umumnya menggunakan data pada skala wilayah tertentu atau variabel yang lebih terbatas, penelitian ini mengintegrasikan beberapa variabel penting yang berkaitan dengan pelaksanaan imunisasi balita. Selain itu, penelitian ini tidak hanya melakukan proses pengelompokan menggunakan algoritma *K-Means*, tetapi juga melakukan evaluasi kualitas hasil *clustering* melalui metode *Elbow Method* dan *Silhouette Score* untuk memastikan bahwa struktur cluster yang terbentuk memiliki tingkat pemisahan yang memadai. Pendekatan ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai pola cakupan imunisasi balita di Indonesia serta mendukung analisis kebijakan kesehatan yang lebih berbasis data.

Berdasarkan rumusan masalah tersebut, penelitian ini bertujuan untuk menganalisis cakupan imunisasi balita di Indonesia menggunakan algoritma *K-Means*. Variabel yang digunakan meliputi riwayat pemberian imunisasi, jenis imunisasi, tempat pemberian imunisasi, pemberi imunisasi, dan alasan tidak diberi imunisasi. Data diolah menggunakan *Python* melalui *platform* Google Colab dan dikelompokkan ke dalam tiga cluster, yaitu cakupan tinggi, sedang, dan rendah. Hasil *clustering* divisualisasikan dan dianalisis untuk memahami karakteristik masing-masing cluster, sehingga dapat digunakan sebagai dasar rekomendasi kebijakan dalam meningkatkan pemerataan dan kualitas program imunisasi.

2 Tinjauan Literatur

Imunisasi balita merupakan salah satu indikator penting dalam evaluasi program kesehatan masyarakat karena berkaitan dengan tingkat perlindungan anak terhadap penyakit menular. Analisis cakupan imunisasi diperlukan untuk memahami variasi capaian imunisasi antarwilayah serta faktor-faktor yang memengaruhinya, sehingga hasilnya dapat digunakan sebagai dasar perencanaan dan evaluasi kebijakan Kesehatan[13].

Data mining menjadi pendekatan yang banyak digunakan dalam pengolahan data kesehatan karena mampu mengekstraksi informasi dan pola dari data berukuran besar. Salah satu teknik yang sering diterapkan adalah *clustering*, yaitu proses pengelompokan data berdasarkan tingkat kemiripan atribut. Teknik ini memungkinkan pemetaan kondisi wilayah atau kelompok sasaran secara lebih terstruktur dan objektif[14].

Algoritma *K-Means* merupakan metode *clustering* yang populer karena memiliki proses perhitungan yang sederhana dan efisien. Algoritma ini bekerja dengan menentukan pusat cluster (*centroid*) dan mengelompokkan data berdasarkan jarak terdekat ke *centroid* tersebut. Beberapa penelitian terdahulu menunjukkan bahwa *K-Means* efektif digunakan untuk mengelompokkan data imunisasi balita ke dalam kategori cakupan tinggi, sedang, dan rendah, sehingga memudahkan interpretasi dan analisis kondisi imunisasi[12], [14].

Penelitian terkait pengelompokan data imunisasi balita menggunakan algoritma *K-Means* umumnya masih dilakukan pada skala wilayah tertentu atau menggunakan variabel yang terbatas. Hal ini menyebabkan hasil *clustering* belum sepenuhnya merepresentasikan kondisi imunisasi secara menyeluruh. Oleh karena itu, diperlukan penelitian dengan cakupan data yang lebih luas serta variabel imunisasi yang lebih lengkap agar hasil pengelompokan dapat memberikan gambaran yang lebih akurat[12], [13].

Berbeda dengan penelitian sebelumnya, penelitian ini menggunakan Data Kesejahteraan Rakyat Indonesia tahun 2025 dengan cakupan nasional dan variabel imunisasi yang lebih lengkap. Hasil pengelompokan tidak hanya menggambarkan tingkat cakupan imunisasi, tetapi juga menganalisis karakteristik setiap cluster untuk mendukung perumusan rekomendasi kebijakan.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* untuk mengidentifikasi pola cakupan imunisasi balita di Indonesia berdasarkan Data Kesejahteraan Rakyat Indonesia tahun 2025 yang dipublikasikan oleh Badan Pusat Statistik. Data dianalisis dengan memanfaatkan teknik *clustering* guna mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristik imunisasi. Pendekatan ini banyak digunakan dalam penelitian kesehatan masyarakat karena mampu menyederhanakan data berskala besar menjadi kelompok yang lebih mudah dianalisis[15].

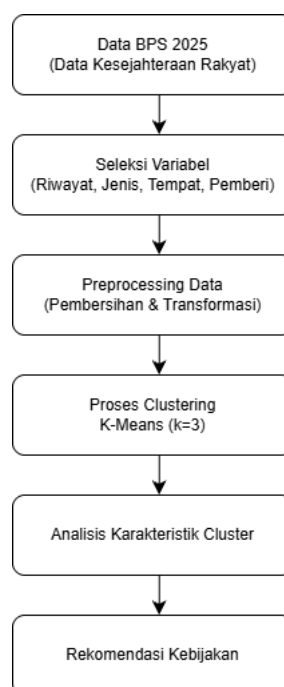
Variabel yang digunakan dalam penelitian ini meliputi riwayat pemberian imunisasi, jenis imunisasi, tempat pemberian imunisasi, pemberi imunisasi, serta alasan balita tidak memperoleh imunisasi. Pemilihan variabel tersebut bertujuan untuk menggambarkan kondisi cakupan imunisasi secara lebih komprehensif dan mendukung interpretasi hasil *clustering* secara lebih bermakna[16].

Proses pengolahan data dilakukan menggunakan bahasa pemrograman *Python* melalui platform Google Colab dengan bantuan pustaka *pandas* untuk manipulasi data, *scikit-learn* untuk proses *clustering*, serta *matplotlib* untuk visualisasi hasil. Penggunaan *Python* dalam analisis data kesehatan terbukti efektif karena bersifat fleksibel, mudah direproduksi, dan mampu menangani data dalam jumlah besar secara efisien[15].

Tahapan penelitian meliputi pengumpulan data, pembersihan data, transformasi data, proses *clustering* menggunakan algoritma *K-Means*, visualisasi hasil pengelompokan, serta analisis karakteristik setiap cluster. Hasil analisis selanjutnya digunakan sebagai dasar dalam perumusan rekomendasi kebijakan untuk meningkatkan pemerataan dan kualitas pelayanan imunisasi[17].

3.1 Alur Proses Penelitian

Alur proses penelitian ditunjukkan pada Gambar 1. Tahapan dimulai dari pengambilan *dataset* dari BPS, dilanjutkan dengan proses pembersihan dan transformasi data, penerapan algoritma *K-Means*, visualisasi hasil *clustering*, serta analisis karakteristik cluster untuk penyusunan



rekomendasi kebijakan.

Gambar 1 Alur proses penelitian

3.1.1 Pengumpulan Data dan Seleksi Variabel

Tahap awal dimulai dengan pengumpulan data primer dari sumber resmi, yakni *Statistik Kesejahteraan Rakyat 2025* yang dipublikasikan oleh Badan Pusat Statistik dan dapat diakses melalui

<http://sistemasi.ftik.unisi.ac.id>

<https://www.bps.go.id/publication/2025/10/10/b37dd4fb6b1727f8cf3c0271/statistik-kesejahteraan-rakyat-2025.html>[6]. Pada langkah ini dilakukan seleksi variabel yang relevan dengan cakupan imunisasi balita seperti status imunisasi, jenis vaksin, lokasi layanan, dan pemberi layanan. Seleksi ini penting untuk memastikan bahwa hanya atribut yang bermakna secara statistik yang akan diproses pada tahap berikutnya.

3.1.2 Preprocessing Data (Pembersihan dan Transformasi)

Tahap pembersihan data dilakukan dengan memeriksa keberadaan nilai yang hilang (*missing value*) pada setiap variabel dalam dataset. Nilai yang tidak tersedia kemudian ditangani dengan menggantinya menggunakan nilai rata-rata (*mean*) dari masing-masing variabel sehingga struktur data tetap dapat digunakan dalam proses analisis[18]. Selain itu, kolom yang tidak relevan seperti atribut yang tidak memiliki informasi penting juga dihapus agar dataset lebih terfokus pada variabel penelitian.

Setelah proses pembersihan selesai, dilakukan tahap transformasi data sebelum proses *clustering*. Pada tahap ini dilakukan normalisasi menggunakan metode *Min-Max scaling* untuk menyesuaikan rentang nilai antar variabel sehingga perbedaan skala data dapat diminimalkan sebelum algoritma *K-Means* dijalankan[19]. Normalisasi ini bertujuan agar setiap variabel memiliki kontribusi yang seimbang dalam proses perhitungan jarak antar data sehingga hasil pengelompokan menjadi lebih stabil dan mudah diinterpretasikan.

3.1.3 Proses Clustering dengan K -Means

Setelah data siap, proses *clustering* dilakukan dengan algoritma *K-Means* dengan jumlah cluster yang ditetapkan sebanyak tiga ($k = 3$). Metode *Elbow* digunakan untuk menentukan jumlah cluster optimal dengan mengamati nilai *Within-cluster sum of squares* sebagai metrik evaluasi awal. Algoritma kemudian mengelompokkan data berdasarkan jarak terdekat ke pusat cluster (*centroid*) yang diiterasi hingga *konvergen*[20].

Dalam implementasinya menggunakan pustaka *scikit-learn*, algoritma *K-Means* dijalankan dengan beberapa parameter pengaturan. Parameter *init* digunakan untuk menentukan metode inisialisasi *centroid* awal yang berfungsi sebagai titik awal proses pengelompokan. Parameter *max_iter* menunjukkan jumlah maksimum iterasi yang diizinkan selama proses pembaruan *centroid* berlangsung hingga mencapai kondisi konvergen. Selain itu, parameter *random_state* digunakan untuk mengatur nilai awal proses acak sehingga hasil pengelompokan dapat direproduksi secara konsisten ketika analisis dijalankan kembali[21].

3.1.4 Analisis Karakteristik Cluster

Hasil dari *clustering* selanjutnya dianalisis untuk memahami karakteristik setiap cluster, seperti kecenderungan cakupan imunisasi, dominasi variabel tertentu, dan perbedaan antar kelompok hasil pengelompokan. Analisis ini membantu menafsirkan cluster sebagai kategori cakupan rendah, sedang, dan tinggi, serta memberikan konteks kebijakan yang relevan[22].

3.1.5 Rekomendasi Kebijakan

Berdasarkan pola yang ditemukan, tahap akhir merupakan penyusunan rekomendasi kebijakan guna mendukung optimalisasi cakupan imunisasi berdasarkan karakteristik tiap cluster.

3.2 Variabel Penelitian

Struktur variabel penelitian yang digunakan sebagai dasar proses *clustering* ditunjukkan pada Tabel 1.

Tabel 1 Variabel penelitian

Variabel	Keterangan
Riwayat Pemberian Imunisasi	Status Pemberian Imunisasi Balita
Jenis Imunisasi	Jenis Vaksin Yang Diterima
Tempat Pemberian Imunisasi	Lokasi Pemberian Imunisasi
Pemberi Imunisasi	Tenaga atau Fasilitas Pemberian Imunisasi

Tabel 1 menjelaskan variabel yang dijadikan dasar dalam proses pengelompokan data pada penelitian ini. Variabel Riwayat Pemberian Imunisasi digunakan untuk melihat apakah balita telah menerima imunisasi sesuai ketentuan, sementara Jenis Imunisasi menunjukkan macam vaksin yang diperoleh. Variabel Tempat Pemberian Imunisasi menggambarkan lokasi pelaksanaan imunisasi, dan

Pemberi Imunisasi merujuk pada pihak atau tenaga kesehatan yang memberikan layanan tersebut. Keempat variabel tersebut dipilih karena mampu merepresentasikan kondisi dan karakteristik pelaksanaan imunisasi balita sehingga dapat dianalisis lebih lanjut melalui metode *clustering*[23].

3.3 Metode Clustering

Algoritma *K-Means* digunakan karena memiliki proses komputasi yang sederhana, efisien, serta mampu menghasilkan pengelompokan yang mudah diinterpretasikan[14]. Dalam penelitian ini, jumlah cluster ditetapkan sebanyak tiga kelompok yang merepresentasikan tingkat cakupan imunisasi tinggi, sedang, dan rendah.

Secara matematis, algoritma *K-Means* bertujuan untuk memperkecil total variasi di dalam setiap kelompok yang terbentuk. Variasi tersebut diukur melalui nilai *within-cluster sum of squares* (WCSS), yaitu jumlah kuadrat jarak antara setiap data dengan pusat clusternya. Fungsi objektif *K-Means* dapat dinyatakan sebagaimana persamaan 1 berikut :

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} |x_j - \mu_i|^2 \quad (1)$$

Keterangan :

k = jumlah cluster

C_i = himpunan data pada cluster ke- i

x_j = data ke- j

μ_i = *centroid* (titik pusat) cluster ke- i

$\|x_j - \mu_i\|^2$ = jarak *Euclidean* antara data dan *centroid*

Setelah *centroid* awal ditentukan, setiap data dihitung jaraknya terhadap seluruh *centroid* sesuai Persamaan (1), kemudian data ditempatkan pada cluster dengan jarak terdekat. Proses ini diulang hingga perubahan *centroid* sangat kecil atau tidak ada lagi pergerakan.

Perhitungan jarak antar data dan *centroid* menggunakan *Euclidean Distance* :

Setiap jarak antara titik data x dan y dihitung sesuai Persamaan (2) sebelum data ditempatkan pada cluster terdekat.

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

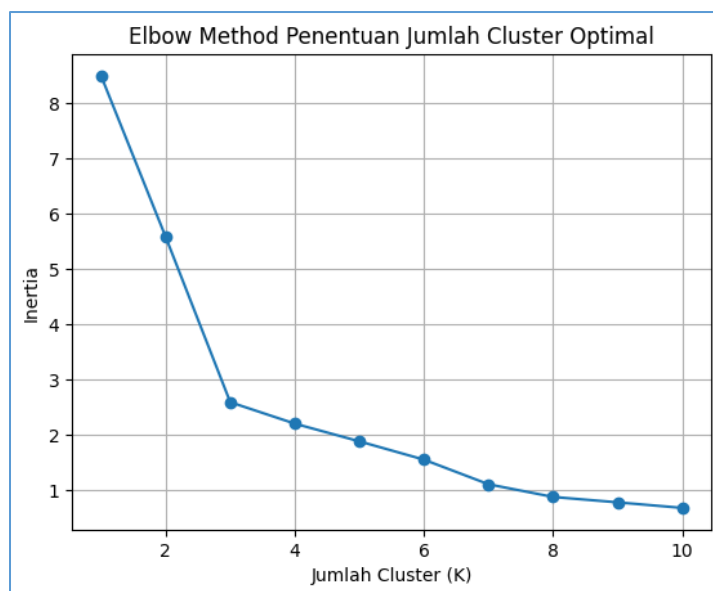
Dengan x dan y merupakan dua titik pada ruang berdimensi n .

Proses pengelompokan dilakukan secara berulang. Tahapan dimulai dari penentuan *centroid* awal, kemudian setiap data dihitung jaraknya terhadap seluruh *centroid*. Data akan ditempatkan pada cluster dengan jarak terdekat. Setelah itu, posisi *centroid* diperbarui berdasarkan rata-rata anggota cluster yang baru. Proses ini terus dilakukan hingga perubahan *centroid* sudah sangat kecil atau tidak lagi mengalami perpindahan. Hasil akhir kemudian divisualisasikan agar pola penyebaran data dan karakter tiap cluster dapat diamati dengan lebih jelas, sehingga memudahkan interpretasi dan analisis lanjutan[24].

4 Hasil dan Pembahasan

4.1 Hasil Pengelompokan Cakupan Imunisasi Balita

Penerapan algoritma *K-Means* dalam penelitian ini menghasilkan tiga kelompok cluster yang digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan karakteristik data imunisasi. Jumlah cluster ditentukan melalui *Elbow Method*, di mana pola penurunan nilai *within-cluster sum of squares* menunjukkan titik siku pada nilai $K = 3$. Dengan demikian, jumlah tersebut dinilai paling representatif dalam menggambarkan struktur pengelompokan data.



Gambar 2 Grafik *elbow method* penentuan jumlah cluster optimal

Grafik pada Gambar 2 menunjukkan bahwa nilai penurunan WCSS mulai melandai pada $K = 3$ sehingga jumlah cluster tersebut dipilih sebagai konfigurasi yang paling sesuai. Perlu ditegaskan bahwa penggunaan label *Tinggi*, *Sedang*, dan *Rendah* pada hasil *clustering* tidak secara langsung merefleksikan tingkat capaian indikator imunisasi secara linier. Label tersebut digunakan sebagai penanda kategori hasil pengelompokan berdasarkan dimensi tertentu dalam data, seperti distribusi wilayah atau karakteristik fasilitas layanan kesehatan. Penjelasan lebih lanjut mengenai perbedaan karakteristik masing-masing cluster serta keterkaitannya dengan indikator imunisasi dibahas secara mendalam pada bagian 4.3.

Berdasarkan hasil pengelompokan, mayoritas provinsi tergolong ke dalam cluster Rendah, sementara cluster Tinggi umumnya mencakup provinsi dengan karakteristik wilayah yang lebih terkonsentrasi dan akses layanan kesehatan yang relatif lebih baik. Adapun cluster Sedang terdiri dari jumlah provinsi yang lebih terbatas, yang menunjukkan posisi transisi antara kedua kelompok lainnya dengan capaian imunisasi yang belum optimal namun tidak berada pada kondisi terendah.

Tabel 2 Hasil pengelompokan provinsi berdasarkan kategori cluster (*tinggi, sedang, rendah*)

Provinsi	Kategori Cluster
Aceh	Sedang
Sumatera Utara	Rendah
Sumatera Barat	Rendah
Riau	Rendah
Jambi	Rendah
Sumatera Selatan	Rendah
Bengkulu	Rendah
Lampung	Rendah
Kepulauan Bangka Belitung	Tinggi
Kepulauan Riau	Tinggi
DKI Jakarta	Tinggi
Jawa Barat	Rendah
Jawa Tengah	Rendah
DI Yogyakarta	Tinggi
Jawa Timur	Rendah
Banten	Tinggi
Bali	Tinggi
Nusa Tenggara Barat	Rendah
Nusa Tenggara Timur	Rendah
Kalimantan Barat	Rendah

<http://sistemasi.ftik.unisi.ac.id>

Kalimantan Tengah	Rendah
Kalimantan Selatan	Rendah
Kalimantan Timur	Tinggi
Kalimantan Utara	Rendah
Sulawesi Utara	Tinggi
Sulawesi Tengah	Rendah
Sulawesi Selatan	Rendah
Sulawesi Tenggara	Rendah
Gorontalo	Rendah
Sulawesi Barat	Rendah
Maluku	Rendah
Maluku Utara	Rendah
Papua Barat	Rendah
Papua Barat Daya	Rendah
Papua	Tinggi
Papua Selatan	Rendah
Papua Tengah	Sedang
Papua Pegunungan	Sedang

Tabel 2 menyajikan hasil pengelompokan provinsi berdasarkan proses *clustering* yang telah dilakukan. Terlihat bahwa sebagian besar provinsi masuk ke dalam kategori Rendah, sedangkan kategori Tinggi dan Sedang mencakup jumlah wilayah yang lebih sedikit. Perbedaan komposisi ini menunjukkan adanya variasi karakteristik imunisasi antarprovinsi yang menjadi dasar untuk analisis lebih lanjut pada bagian berikutnya.

4.2 Evaluasi Kualitas dan Tingkat Keakuratan Cluster

Untuk menilai kualitas hasil *clustering*, penelitian ini menggunakan *Silhouette Score* sebagai ukuran tingkat keakuratan dan konsistensi pengelompokan. Berdasarkan hasil perhitungan, nilai *Silhouette Score* yang diperoleh sebesar 0,5086. Nilai ini menunjukkan bahwa struktur cluster yang terbentuk berada pada kategori cukup baik, di mana jarak antar data dalam satu cluster relatif lebih dekat dibandingkan jarak antar cluster yang berbeda.

Nilai *Silhouette Score* di atas 0,5 mengindikasikan bahwa algoritma *K-Means* mampu membedakan karakteristik masing-masing cluster secara memadai[25]. Dengan demikian, hasil pengelompokan dapat dianggap layak digunakan sebagai dasar analisis lebih lanjut, khususnya dalam mengidentifikasi perbedaan pola cakupan imunisasi antarwilayah.

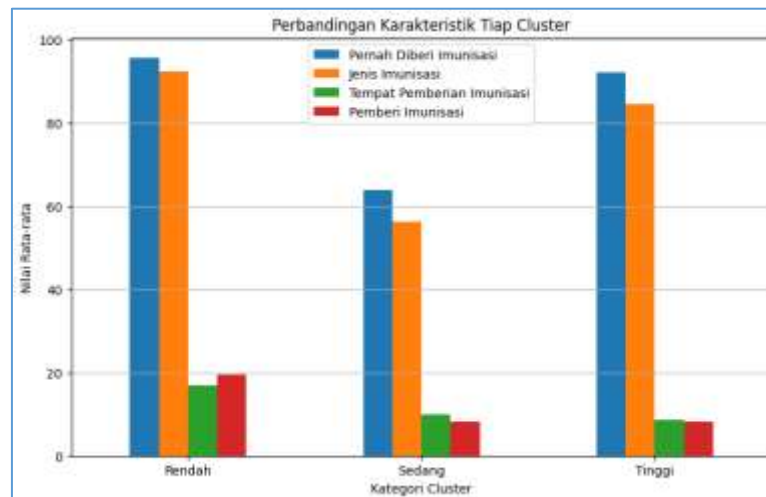
Tabel 3 Nilai evaluasi *clustering* menggunakan *silhouette score*

Metode Clustering	Jumlah Cluster (k)	Nilai <i>Silhouette Score</i>	Kategori Kualitas
<i>K-Means</i>	3	0,5086	Cukup Baik

Tabel 3 merangkum hasil evaluasi *clustering* berdasarkan nilai *Silhouette Score* yang diperoleh. Informasi tersebut memperjelas bahwa konfigurasi tiga cluster yang digunakan dalam penelitian ini telah melalui tahap pengujian kualitas sebelum dilakukan analisis lebih lanjut.

4.3 Analisis Karakteristik Setiap Cluster

Penggambaran karakteristik cluster dilakukan melalui analisis deskriptif dengan menghitung nilai rata-rata dari empat variabel yang berkaitan dengan kebijakan imunisasi, yaitu status pernah diberi imunisasi, jenis imunisasi, tempat pemberian imunisasi, serta pemberi imunisasi. Ringkasan hasil perhitungan disajikan pada Tabel 4 dan divisualisasikan melalui Gambar 3. Hasil tersebut menunjukkan bahwa pembagian cluster tidak sepenuhnya mengikuti asumsi linier antara label kategori dan capaian indikator, melainkan memperlihatkan variasi performa yang bersifat tidak seragam antar kelompok



Gambar 3 Diagram batang perbandingan karakteristik rata-rata variabel antar cluster

Tabel 4 Rata-rata nilai variabel imunisasi pada setiap cluster

Kategori Cluster	Pernah Diberi Imunisasi	Jenis Imunisasi	Tempat Pemberian Imunisasi	Pemberi Imunisasi
Rendah	95.675556	92.388889	17.005556	19.52
Sedang	63.820000	56.203333	9.936667	8.176667
Tinggi	92.232692	84.543077	8.69	8.190385

Cluster Rendah justru menunjukkan capaian paling optimal pada seluruh variabel yang dianalisis. Rata-rata balita yang pernah menerima imunisasi mencapai 95,68%, dengan nilai jenis imunisasi sebesar 92,39%. Selain itu, cluster ini didukung oleh ketersediaan sarana dan tenaga imunisasi yang relatif lebih tinggi, tercermin dari rata-rata tempat pemberian imunisasi sebesar 17,01 serta jumlah pemberi imunisasi sebesar 19,52. Kondisi tersebut mengindikasikan bahwa meskipun berlabel “rendah”, cluster ini kemungkinan merepresentasikan wilayah dengan karakteristik operasional yang lebih terkonsentrasi, sehingga proses penyelenggaraan layanan imunisasi dapat dilakukan secara lebih intensif dan efektif.

Pada sisi lain, Cluster Tinggi menempati posisi menengah dari segi capaian indikator. Nilai rata-rata pernah diberi imunisasi tercatat sebesar 92,23%, sedangkan jenis imunisasi mencapai 84,54%. Angka ini menunjukkan penurunan relatif dibandingkan Cluster Rendah, yang mengisyaratkan adanya tantangan dalam menjaga kualitas layanan seiring dengan perluasan cakupan wilayah. Hal tersebut juga tercermin pada variabel pendukung, di mana nilai rata-rata tempat dan pemberi imunisasi berada pada kisaran yang lebih rendah, masing-masing sebesar 8,69 dan 8,19. Kondisi ini mengindikasikan adanya tekanan kapasitas layanan pada cluster dengan wilayah cakupan yang lebih luas.

Sementara itu, Cluster Sedang memperlihatkan performa paling rendah pada seluruh variabel utama. Rata-rata balita yang pernah menerima imunisasi hanya mencapai 63,82%, dengan jenis imunisasi sebesar 56,20%. Keterbatasan tersebut turut diperkuat oleh rendahnya nilai fasilitas dan tenaga pendukung, yaitu masing-masing sebesar 9,94 dan 8,18. Posisi cluster ini berada di antara dua ekstrem tanpa memperoleh keunggulan signifikan dari keduanya, baik dari sisi intensitas layanan maupun skala cakupan. Dengan selisih capaian yang cukup besar dibandingkan nilai tertinggi, cluster ini menunjukkan potensi peningkatan yang masih luas dan memerlukan perhatian kebijakan yang lebih terarah.

4.4 Visualisasi Hasil Clustering

Untuk memperjelas pemisahan antar cluster, visualisasi dilakukan menggunakan pendekatan *Principal Component Analysis* (PCA) dua dimensi. Hasil visualisasi pada Gambar 4 menunjukkan bahwa masing-masing cluster cenderung membentuk kelompok yang relatif terpisah, meskipun masih

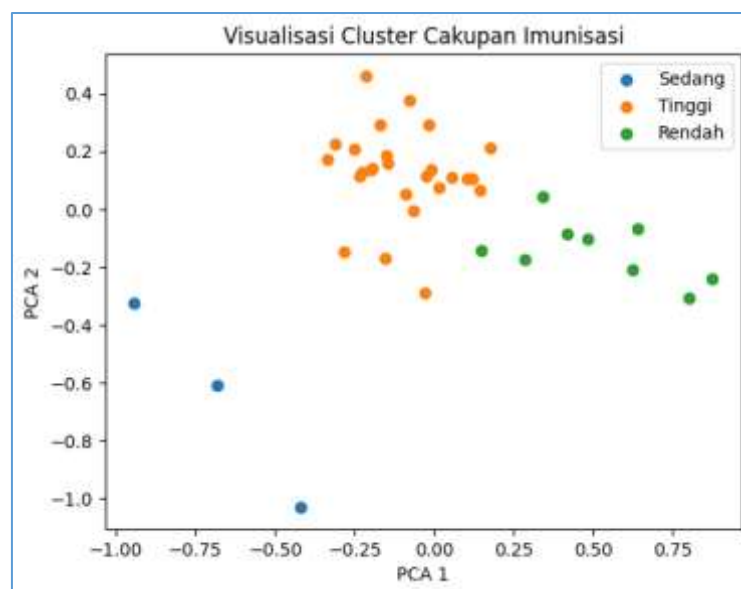
<http://sistemasi.ftik.unisi.ac.id>

terdapat beberapa titik data yang berdekatan. Hal ini sejalan dengan nilai *Silhouette Score* yang berada pada kategori cukup baik.

Nilai *Silhouette Score* tersebut mengindikasikan bahwa tingkat pemisahan antar cluster belum sepenuhnya optimal, di mana sebagian data masih memiliki kedekatan dengan cluster lain. Hal ini tercermin pada visualisasi PCA yang menunjukkan adanya beberapa titik data yang berada pada area perbatasan antar kelompok. Fenomena tersebut mencerminkan adanya kesamaan karakteristik imunitas antarprovinsi, sehingga batas pemisahan antar cluster tidak terbentuk secara tegas.

Namun demikian, tingkat pemisahan yang moderat ini tidak dapat diinterpretasikan sebagai kelemahan metode maupun kualitas data. Sebaliknya, kondisi tersebut merefleksikan karakteristik data sosial dan kesehatan masyarakat yang bersifat kompleks, di mana perbedaan antarwilayah umumnya berlangsung secara bertahap dan tidak menunjukkan pemisahan yang ekstrem. Oleh karena itu, hasil *clustering* dengan kualitas pemisahan yang cukup baik tetap relevan dalam menggambarkan variasi cakupan imunitas balita di Indonesia.

Penentuan jumlah cluster sebanyak tiga kelompok ($k = 3$) didasarkan pada hasil *Elbow Method* yang menunjukkan titik optimal pada nilai tersebut. Konfigurasi ini dipilih untuk menjaga keseimbangan antara kemampuan model dalam membedakan karakteristik wilayah dan kemudahan interpretasi hasil analisis. Meskipun penambahan jumlah cluster berpotensi meningkatkan nilai evaluasi numerik, langkah tersebut berisiko menghasilkan segmentasi yang terlalu rinci dan kurang efektif dalam mendukung analisis kebijakan. Dengan demikian, nilai *Silhouette Score* yang berada pada kategori cukup baik tidak menjadi keterbatasan utama penelitian, melainkan mencerminkan karakteristik data yang realistis dan relevan untuk analisis kebijakan kesehatan berbasis wilayah.



Gambar 4 Visualisasi hasil *clustering* menggunakan PCA

Berdasarkan Gambar 4, data cakupan imunitas terbagi menjadi tiga kelompok yang relatif terpisah. Cluster Tinggi tampak mendominasi dan berada di area tengah grafik, sedangkan Cluster Rendah dan Cluster Sedang menempati posisi yang berbeda pada sumbu komponen utama (PCA 1 dan PCA 2), menunjukkan perbedaan karakteristik antarwilayah. Pemisahan ini memperlihatkan bahwa proses *clustering* mampu membedakan tingkat cakupan imunitas secara cukup jelas.

4.5 Analisis Variabel Paling Berpengaruh

Analisis dilakukan untuk melihat variabel yang paling membedakan karakteristik antar cluster dengan menggunakan nilai *range*, yaitu selisih antara nilai rata-rata tertinggi dan terendah pada setiap variabel. Hasil perhitungan menunjukkan bahwa variabel jenis imunitas memiliki nilai *range* terbesar ($\pm 36,19$), diikuti oleh variabel pernah diberi imunitas ($\pm 31,86$). Sementara itu, variabel pemberi imunitas dan tempat pemberian imunitas memiliki nilai *range* yang lebih kecil, masing-masing

sebesar $\pm 11,34$ dan $\pm 8,32$. Rincian urutan variabel berdasarkan tingkat pengaruh dapat dilihat pada Tabel 5.

Tabel 5 Urutan variabel berdasarkan tingkat pengaruh terhadap pembentukan cluster

Nama Variabel	Nilai Range	Kategori Pengaruh
Jenis Imunisasi	36,19	Sangat Berpengaruh
Pernah Diberi Imunisasi	31,86	Berpengaruh
Pemberi Imunisasi	11,34	Cukup Berpengaruh
Tempat Pemberian Imunisasi	8,32	Kurang Berpengaruh

Besarnya nilai *range* pada variabel jenis imunisasi dan status pernah diberi imunisasi menunjukkan bahwa kedua variabel tersebut berperan dominan dalam membedakan profil imunisasi antar cluster. Hal ini mengindikasikan bahwa variasi kelengkapan jenis vaksin serta tingkat penerimaan imunisasi menjadi faktor utama dalam segmentasi cluster. Sebaliknya, variabel yang berkaitan dengan fasilitas dan tenaga imunisasi menunjukkan perbedaan yang relatif kecil antar cluster, sehingga kontribusinya bersifat pendukung dan tidak menjadi pembeda utama.

4.6 Implikasi Hasil Terhadap Kebijakan Kesehatan

Berdasarkan hasil pengelompokan dan analisis karakteristik tiap cluster, upaya peningkatan cakupan imunisasi perlu dirancang secara berbeda sesuai dengan kondisi masing-masing kelompok wilayah. Hal ini penting mengingat hasil *clustering* menunjukkan bahwa label kategori tidak selalu sejalan secara langsung dengan tingkat capaian indikator imunisasi.

Cluster Rendah, meskipun secara nomenklatur dikategorikan sebagai “rendah”, justru memperlihatkan capaian imunisasi yang paling baik. Kondisi ini mengindikasikan bahwa kebijakan pada cluster tersebut tidak berfokus pada peningkatan, melainkan pada upaya menjaga keberlanjutan kualitas layanan. Strategi yang dapat diterapkan antara lain mempertahankan standar pelayanan yang sudah berjalan, mendokumentasikan praktik pelaksanaan imunisasi yang efektif, serta memastikan tidak terjadi penurunan mutu layanan di masa mendatang. Karakteristik wilayah yang relatif lebih terkonsentrasi diduga mempermudah pengelolaan dan intensifikasi layanan imunisasi.

Sebaliknya, Cluster Sedang menunjukkan capaian terendah pada seluruh indikator utama, sehingga memerlukan perhatian kebijakan yang lebih intensif. Intervensi pada cluster ini perlu diarahkan pada perbaikan mendasar, seperti evaluasi sistem distribusi vaksin, peningkatan akses layanan imunisasi, serta penguatan kapasitas tenaga kesehatan. Selisih capaian yang cukup besar dibandingkan cluster dengan performa terbaik menunjukkan bahwa upaya perbaikan tidak dapat bersifat parsial, melainkan membutuhkan pendekatan yang lebih terstruktur dan berkelanjutan.

Sementara itu, Cluster Tinggi berada pada posisi menengah dengan capaian imunisasi yang relatif baik namun masih berada di bawah cluster dengan performa tertinggi. Kebijakan yang relevan bagi cluster ini berfokus pada upaya penyetaraan kualitas layanan, khususnya melalui peningkatan kelengkapan jenis imunisasi dan penguatan sarana serta tenaga pendukung. Dengan strategi tersebut, kesenjangan capaian antar cluster diharapkan dapat diperkecil.

Secara keseluruhan, pendekatan berbasis hasil *clustering* ini memungkinkan penentuan prioritas kebijakan yang lebih tepat sasaran. Intervensi dapat difokuskan secara maksimal pada cluster dengan performa terendah, pemeliharaan kualitas dilakukan pada cluster dengan capaian optimal, serta peningkatan bertahap diarahkan pada cluster dengan capaian menengah. Temuan ini menunjukkan bahwa algoritma *K-Means* dapat dimanfaatkan sebagai alat analisis pendukung dalam perumusan kebijakan kesehatan yang lebih terarah dan berbasis data.

5 Kesimpulan

Analisis Data Kesejahteraan Rakyat Indonesia tahun 2025 menggunakan algoritma *K-Means* menghasilkan pengelompokan 38 provinsi ke dalam tiga cluster berdasarkan karakteristik cakupan imunisasi balita. Penentuan jumlah cluster melalui *Elbow Method* serta evaluasi menggunakan *Silhouette Score* sebesar 0,5086 menunjukkan bahwa pembagian menjadi tiga kelompok telah membentuk struktur cluster yang cukup baik. Hasil pengelompokan menunjukkan bahwa klasifikasi

<http://sistemasi.ftik.unisi.ac.id>

cluster tidak selalu sejalan dengan tingkat capaian imunisasi, di mana cluster yang diberi label Rendah justru memiliki persentase balita yang pernah menerima imunisasi tertinggi sebesar 95,68%, sedangkan cluster Sedang memiliki cakupan yang relatif lebih rendah sebesar 63,82%. Analisis juga menunjukkan bahwa variabel jenis imunisasi dan riwayat pemberian imunisasi menjadi faktor yang paling berpengaruh dalam membedakan karakteristik antarcluster. Penelitian ini memberikan gambaran pola cakupan imunisasi berbasis pengelompokan data provinsi, namun masih memiliki keterbatasan pada penggunaan satu metode *clustering* serta belum mempertimbangkan faktor spasial dan sosial ekonomi secara langsung. Oleh karena itu, penelitian selanjutnya disarankan menggunakan metode *clustering* lain sebagai pembanding serta menambahkan variabel pendukung agar hasil analisis dapat memberikan pemahaman yang lebih komprehensif dalam mendukung perumusan kebijakan peningkatan cakupan imunisasi.

Referensi

- [1] F. Yusma *et al.*, “Gambaran Cakupan Imunisasi Dasar Balita pada Masa Pandemi COVID-19 Tahun 2020–2022,” *The Indonesian Journal of General Medicine*, Vol. 21, No. 1, pp. 39–64, Dec. 2025.
- [2] R. S. Putri and R. D. Wulandari, “Literature Review: Implementasi Program Imunisasi Dasar Lengkap di Indonesia Tahun 2020–2025,” *PREPOTIF : Jurnal Kesehatan Masyarakat*, Vol. 9, No. 2, pp. 6695–6702, 2025.
- [3] I. Siramaneerat and F. Agushybana, “Inequalities in Immunization Coverage in Indonesia: a Multilevel Analysis,” *Rural Remote Health*, Vol. 21, No. 3, p. 6348, Aug. 2021, DOI: 10.22605/RRH6348.
- [4] C. A. M. Freitas, A. L. Q. Vilasbôas, and N. M. de Brito Lima Prado, “Strategies and Challenges for Immunisation of Children under 5 Years in Primary Health Care in Different Countries: A Scoping Review,” *BMC Public Health*, Vol. 24, No. 1, Dec. 2025, DOI: 10.1186/s12889-025-24846-y.
- [5] World Health Organization (WHO) and Ministry of Health Indonesia, “Indonesia EPI Factsheet 2024,” Jakarta (New Delhi untuk Kantor Regional WHO SEARO), Jul. 2024.
- [6] Badan Pusat Statistik, “Statistik Kesejahteraan Rakyat 2025,” Jakarta, 2025.
- [7] M. Uzla, L. Jutomo, and P. Romeo, “Comparison of the Coverage of Basic Immunization for Children under Five before and during the COVID-19 Pandemic,” *Pancasakti Journal of Public Health Science and Research*, Vol. 3, No. 2, pp. 94–99, 2023, DOI: 10.47650/pjphsr.v3i2.573.
- [8] A. Noroozi, S. Muntaha Esha, and M. Ghari, “Predictors of Childhood Vaccination Uptake in England: An Explainable Machine Learning Analysis of Longitudinal Regional Data (2021–2024),” Thesis, University of Wolverhampton, Wolverhampton, UK, 2025.
- [9] L. Meng *et al.*, “Cluster Analysis of Adults Unvaccinated for COVID-19 based on Behavioral and Social Factors, National Immunization Survey-Adult COVID Module, United States,” *Prev. Med. (Baltim.)*, Vol. 167, Feb. 2023, DOI: 10.1016/j.ypmed.2022.107415.
- [10] M. Maleki, “Clustering Analysis of US COVID-19 Rates, Vaccine Participation, and Socioeconomic Factors,” Apr. 2024, doi: 10.3390/healthcare12151458.
- [11] A. R. Hasugian and I. Zufria, “Diving Into Public Sentiments: SVM-based Analysis of Water Service Opinions,” *Journal of Computer Networks, Architecture and High Performance Computing*, Vol. 7, No. 1, pp. 200–213, Jan. 2025, DOI: 10.47709/cnahpc.v7i1.5231.
- [12] M. N. Tiyyar and S. Samsudin, “Klasifikasi Data Sasaran Imunisasi Bayi dan Baduta pada Puskesmas berbasis Web Metode Clustering Algoritma K-Means,” *Jurnal Teknologi Sistem Informasi dan Aplikasi*, Vol. 7, No. 1, pp. 143–154, Jan. 2024, DOI: 10.32493/jtsi.v7i1.37321.
- [13] M. Usniyah Sari and I. Muhandhis, “Pengembangan Sistem Pendukung Keputusan Cakupan Vaksinasi Anak menggunakan Metode K-Means Clustering: Decision Support System Development of Childhood Vaccination Completeness using K-Means Clustering Method,” *Jurnal Sistem Informasi dan Bisnis Cerdas*, Vol. 17, No. 2, Aug. 2024.
- [14] R. R. Aria, “Implementasi Algoritma K-Means untuk Pengelompokan Data Imunisasi Balita dengan Metode CRISP-DM,” *Remik*, Vol. 9, No. 1, pp. 189–197, Jan. 2025, DOI: 10.33395/remik.v9i1.14391.

- [15] W. T. Wu et al., “Data Mining in Clinical Big Data: The Frequently used Databases, Steps, and Methodological Models,” Dec. 01, 2021, BioMed Central Ltd. DOI: 10.1186/s40779-021-00338-z.
- [16] H. Chu and A. Rammohan, “Childhood Immunization and Age-Appropriate Vaccinations in Indonesia,” *BMC Public Health*, Vol. 22, No. 1, Dec. 2022, DOI: 10.1186/s12889-022-14408-x.
- [17] S. D. Alfian, R. Abdulah, and E. Hak, “Development of a Prediction Rule for Incomplete Vaccination among Children in Indonesia,” *BMC Public Health*, Vol. 24, No. 1, Dec. 2025, DOI: 10.1186/s12889-025-23109-0.
- [18] P. Koukaras and C. Tjortjis, “Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices,” Oct. 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). DOI: 10.3390/ai6100257.
- [19] S. E. Saqila, I. P. Ferina, and A. Iskandar, “Analisis Perbandingan Kinerja Clustering Data Mining untuk Normalisasi Dataset,” *Jurnal Sistem Komputer dan Informatika (JSON)*, Vol. 5, No. 2, p. 356, Dec. 2023, DOI: 10.30865/json.v5i2.6919.
- [20] R. E. Yuliana, M. Ulya, and M. Jamhuri, “Mall Customer Segmentation using K-Means Clustering Optimized by the Elbow Method://creativecommons.org/licenses/by-sa/4.0),” *Riset Mahasiswa Matematika*, Vol. 4, No. 5, pp. 273–286, 2025, DOI: 10.18860/jrmm/v4i5.33389.
- [21] D. R. Nugraha, A. T. Zy, and A. S. Sunge, “The use of K-Means Algorithm Clustering in Grouping Life Expectancy (Case Study: Provinces in Indonesia),” *Journal of Computer Networks, Architecture and High Performance Computing*, Vol. 6, No. 3, 2024.
- [22] S. Wulandari, H. S. Husin, W. R. Sukmaningsih, S. T. Manajemen, and I. Kesehatan, “Data Mining K-Means: Clustering Health and Complaints Resident in Indonesia,” *International Journal Of Health Science*, Vol. 1, No. 3, pp. 57–63, 2021, [Online]. Available: <https://www.bps.go.id>.
- [23] E. Widyowati, Almaini, and W. I. P. E. Sari, “Faktor-Faktor yang Berhubungan dengan Kelengkapan Imunisasi Dasar pada Bayi di Puskesmas Embong Ijuk Kabupaten Kepahiang Tahun 2023,” *Journal of Midwifery (Jurnal Kebidanan)*, Vol. 11, No. 2, pp. 215–226, Oct. 2023.
- [24] M. S. Hassan, F. Abdulla, and M. Hossain, “Vaccination Coverage and Its Associated Factors among Children Under-5 in Somalia,” *BMC Public Health*, Vol. 24, No. 1, Dec. 2025, DOI: 10.1186/s12889-025-21529-6.
- [25] S. S. Fourteen, A. H. P. B. Anugrah, M. N. A. Putra, J. A. Shabrina, M. A. Winoto, and Z. Alfian, “Perbandingan Hasil Klasterisasi K-Means berdasarkan Silhouette Score dan Inertia pada 5 Dataset Berbeda.”