Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

Klasterisasi Provinsi Di Indonesia Berdasarkan Indeks Masyarakat Digital Tahun 2025 Menggunakan Algoritma K-Means Clustering

Putri Wulandari*, Muhammad Irawan Pratama, Adnan Buyung Nasution

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Sumatera Utara, Indonesia
Email : [1,*]putri.wulndri@gmail.com, [2]mirawanpratama5602@gmail.com, [3]adnanbuyungnasution@uinsu.ac.id
Email Korespondensi : putri.wulndri@gmail.com
Received: **23/08/2026**, Revised: **20/09/2026**, Accepted: **24/09/2026**, Published: **25/09/2026**

Abstrak−Transformasi digital merupakan salah satu agenda prioritas pembangunan nasional Indonesia, namun kesiapan digital antarwilayah masih menunjukkan kesenjangan yang cukup lebar. Indeks Masyarakat Digital (IMD) tahun 2025 mengukur capaian digitalisasi tiap provinsi melalui empat pilar, yaitu Infrastruktur dan Ekosistem, Literasi Digital, Pemberdayaan, dan Pekerjaan. Penelitian ini bertujuan mengelompokkan 38 provinsi di Indonesia berdasarkan keempat pilar tersebut menggunakan algoritma K-Means Clustering agar pola kesenjangan digital antarwilayah dapat diidentifikasi secara objektif dan terukur. Data yang digunakan merupakan data sekunder IMD tahun 2025 yang terlebih dahulu distandarisasi menggunakan metode Z score sebelum proses klasterisasi dilakukan. Jumlah klaster optimal ditentukan melalui kombinasi metode Elbow dan Silhouette Coefficient, yang mengarahkan pemilihan pada tiga klaster dengan nilai koefisien Silhouette sebesar 0,262. Hasil klasterisasi menunjukkan provinsi terbagi menjadi tiga kelompok, yaitu klaster provinsi maju digital sebanyak 18 provinsi dengan rerata IMD 49,18, klaster provinsi berkembang sebanyak 13 provinsi dengan rerata IMD 43,16, dan klaster provinsi tertinggal sebanyak 7 provinsi dengan rerata IMD 35,11 yang didominasi wilayah kepulauan Papua. Pilar Pemberdayaan terbukti menjadi pembeda paling tajam antarklaster, sementara pilar Pekerjaan relatif seragam pada seluruh provinsi. Hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemangku kebijakan dalam merumuskan program pemerataan transformasi digital yang lebih tepat sasaran.

Kata kunci: Indeks Masyarakat Digital; K-Means Clustering; Klasterisasi Provinsi; Kesenjangan Digital; Data Mining

Abstract−Digital transformation is one of Indonesia's national development priorities, yet the level of digital readiness across regions still shows a considerable gap. The 2025 Digital Society Index (Indeks Masyarakat Digital or IMD) measures each province's digitalization achievement through four pillars, namely Infrastructure and Ecosystem, Digital Literacy, Empowerment, and Employment. This study aims to group 38 Indonesian provinces based on these four pillars using the K-Means Clustering algorithm so that the pattern of regional digital disparity can be identified objectively. The dataset used is secondary data of the 2025 IMD, which was standardized using the Z score method prior to clustering. The optimal number of clusters was determined through a combination of the Elbow method and the Silhouette Coefficient, leading to the selection of three clusters with a Silhouette Coefficient of 0.262. The clustering results show that the provinces are divided into three groups, namely a digitally advanced cluster consisting of 18 provinces with an average IMD of 49.18, a developing cluster of 13 provinces with an average IMD of 43.16, and a lagging cluster of 7 provinces with an average IMD of 35.11 that is dominated by the Papua region. The Empowerment pillar was found to be the sharpest differentiator between clusters, while the Employment pillar was relatively uniform across provinces. These findings are expected to serve as a consideration for policy makers in designing more targeted digital transformation equity programs.

Keywords: Digital Society Index; K-Means Clustering; Provincial Clustering; Digital Divide; Data Mining

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong hampir seluruh aspek kehidupan masyarakat bergeser ke ranah digital, mulai dari sektor pemerintahan, pendidikan, ekonomi, hingga interaksi sosial sehari hari. Pemerintah Indonesia melalui Kementerian Komunikasi dan Digital secara rutin mengukur tingkat kesiapan digital masyarakat melalui Indeks Masyarakat Digital Indonesia (IMDI), yang merupakan tindak lanjut dari inisiatif Presidensi G20 Indonesia dalam menyusun G20 Toolkit for Measuring Digital Literacy and Digital Skills [1]. Indeks ini disusun dari empat pilar utama, yaitu Infrastruktur dan Ekosistem, Keterampilan atau Literasi Digital, Pemberdayaan, serta Pekerjaan, yang masing masing pilar diuraikan lebih lanjut ke dalam beberapa subpilar [2].

Data IMDI tahun 2023 memperlihatkan bahwa capaian nasional meningkat dari 37,80 pada tahun 2022 menjadi 43,18 pada tahun 2023 [3],[4],[5],[6], yang menunjukkan tren perbaikan kompetensi digital masyarakat secara umum. Meskipun demikian, capaian rerata nasional tersebut menutupi kenyataan bahwa tingkat kesiapan digital antarprovinsi masih sangat bervariasi. Provinsi yang berada di kawasan perkotaan besar dan wilayah barat Indonesia pada umumnya memiliki skor infrastruktur dan literasi digital yang jauh lebih tinggi dibandingkan provinsi di kawasan timur, khususnya wilayah kepulauan Papua. Kesenjangan semacam ini apabila tidak dipetakan secara sistematis berpotensi menyebabkan kebijakan percepatan transformasi digital yang dirumuskan secara nasional menjadi kurang tepat sasaran, karena kebutuhan intervensi setiap wilayah pada dasarnya berbeda beda.

Salah satu pendekatan yang dapat digunakan untuk memetakan kesenjangan tersebut secara objektif adalah teknik klasterisasi dalam data mining. Klasterisasi memungkinkan sekumpulan provinsi dikelompokkan ke dalam beberapa kategori berdasarkan kemiripan karakteristik data tanpa memerlukan label kelas yang telah ditentukan sebelumnya. Algoritma K-Means Clustering dipilih dalam penelitian ini karena kesederhanaan komputasi, kemudahan interpretasi hasil, serta telah banyak terbukti efektif digunakan pada berbagai kajian pengelompokan wilayah di Indonesia.

Berdasarkan Penelitian Terdahulu, Kementerian Komunikasi dan Informatika (2024) melaporkan hasil pengukuran Indeks Masyarakat Digital Indonesia (IMDI) tahun 2024 yang menunjukkan kondisi perkembangan masyarakat digital di Indonesia berdasarkan indikator yang telah ditetapkan. Hasil pengukuran tersebut memberikan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

gambaran mengenai tingkat kesiapan dan kemampuan masyarakat dalam memanfaatkan teknologi digital di berbagai wilayah Indonesia [1]. Selanjutnya, Badan Pengembangan Sumber Daya Manusia Komunikasi dan Digital (2025) menjelaskan bahwa IMDI dibangun berdasarkan empat pilar utama, yaitu Infrastruktur dan Ekosistem, Keterampilan Digital, Pemberdayaan, dan Pekerjaan. Keempat pilar tersebut digunakan untuk menggambarkan berbagai aspek perkembangan masyarakat digital Indonesia secara lebih komprehensif [2]. Sementara itu, Indonesia Baik (2024) menunjukkan bahwa perkembangan masyarakat Indonesia semakin mengarah pada pemanfaatan teknologi digital dalam berbagai aktivitas kehidupan, sehingga pengukuran kondisi masyarakat digital menjadi penting untuk mengetahui perkembangan dan kesenjangan digital antarwilayah [3]. Ketiga sumber tersebut menjadi dasar dalam penelitian ini untuk memahami karakteristik dan indikator penyusun IMDI, yang selanjutnya digunakan sebagai variabel dalam proses pengelompokan 38 provinsi di Indonesia menggunakan algoritma K-Means Clustering.

Namun demikian, penerapan K-Means Clustering secara khusus pada data Indeks Masyarakat Digital tahun 2025 yang mencakup seluruh 38 provinsi di Indonesia, termasuk provinsi pemekaran baru di wilayah Papua, sejauh ini belum banyak dikaji. Sebagian besar penelitian terdahulu juga cenderung menggunakan satu metode saja dalam menentukan jumlah klaster optimal, padahal penggunaan kombinasi metode Elbow dan Silhouette Coefficient telah terbukti memberikan hasil evaluasi klaster yang lebih dapat dipertanggungjawabkan [8], [9]. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan empat pilar Indeks Masyarakat Digital tahun 2025 menggunakan algoritma K-Means Clustering, dengan penentuan jumlah klaster optimal melalui kombinasi metode Elbow dan Silhouette Coefficient, serta menginterpretasikan karakteristik setiap klaster yang terbentuk sebagai bahan pertimbangan perumusan kebijakan pemerataan digital yang lebih tepat sasaran.

Secara khusus, penelitian ini diarahkan untuk menjawab tiga permasalahan, yaitu berapa jumlah klaster optimal yang terbentuk dari data Indeks Masyarakat Digital tahun 2025, bagaimana karakteristik masing masing klaster ditinjau dari keempat pilar penyusun indeks, dan provinsi mana saja yang tergolong ke dalam kategori maju, berkembang, serta tertinggal dalam hal kesiapan masyarakat digital. Hasil penelitian ini diharapkan memberikan kontribusi baik secara akademis, yaitu memperkaya literatur penerapan K-Means Clustering pada data indeks pembangunan wilayah di Indonesia, maupun secara praktis, yaitu sebagai referensi data bagi pemerintah pusat dan daerah dalam menyusun program percepatan transformasi digital yang lebih berorientasi pada karakteristik masing masing kelompok wilayah.

2. METODE PENELITIAN

2.1 Data Mining

Data Mining adalah proses untuk menemukan pola, hubungan, informasi, atau pengetahuan yang bermanfaat dari sekumpulan data berukuran besar dengan menggunakan teknik statistik, matematika, dan algoritma komputer. Data mining digunakan untuk mengolah data menjadi informasi yang dapat membantu proses analisis dan pengambilan keputusan [10],[11],[12],[13],[14].

2.2 Tahapan Penelitian

Penelitian ini mengikuti tahapan Knowledge Discovery in Database (KDD) yang lazim digunakan pada penelitian data mining, yang terdiri atas enam tahap sebagai berikut.

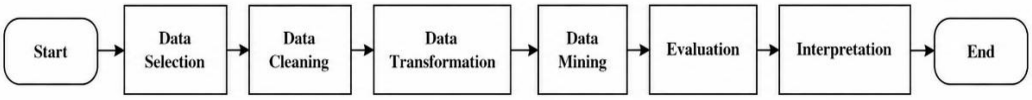
1. Seleksi data (data selection)
   Tahap awal ini dilakukan dengan memilih atribut yang benar benar relevan dengan tujuan penelitian, yaitu skor keempat pilar Indeks Masyarakat Digital, dari keseluruhan variabel yang tersedia pada berkas data mentah. Berkas data asli yang diperoleh memuat cukup banyak kolom, termasuk kode provinsi, tahun pengukuran, skor IMD secara keseluruhan, serta rincian skor pada tingkat subpilar. Kolom kode provinsi dan tahun pengukuran tidak diikutsertakan dalam proses klasterisasi karena bersifat identitas dan tidak menggambarkan karakteristik kesiapan digital. Skor subpilar juga tidak digunakan secara langsung sebagai variabel pembentuk klaster, mengingat penelitian ini difokuskan pada empat pilar utama agar hasil klasterisasi lebih mudah diinterpretasikan secara umum. Dengan demikian, keluaran dari tahap seleksi data berupa satu tabel bersih berisi nama 38 provinsi beserta empat variabel numerik yang siap diproses lebih lanjut, seperti dicontohkan pada Tabel 1.
2. Pembersihan data (data cleaning)
   Setelah atribut yang relevan ditentukan, tahap berikutnya adalah memeriksa kualitas data secara menyeluruh sebelum digunakan dalam proses klasterisasi. Pemeriksaan dilakukan untuk memastikan tidak terdapat nilai kosong atau hilang (missing value) pada setiap baris provinsi maupun kolom pilar. Selain itu, dilakukan pula pengecekan terhadap kemungkinan duplikasi baris provinsi yang dapat menyebabkan bias pada perhitungan centroid. Pemeriksaan tipe data turut dilakukan untuk memastikan seluruh nilai pada kolom pilar benar benar berformat numerik dan tidak tercampur dengan karakter non angka akibat proses ekspor data. Hasil pemeriksaan menunjukkan bahwa data pada 38 provinsi telah lengkap dan konsisten, sehingga tidak diperlukan proses imputasi maupun penghapusan baris data.
3. Transformasi data (data transformation)
   Tahap transformasi dilakukan dengan menstandardisasi seluruh variabel pilar menggunakan metode Z score, yaitu mengurangkan setiap nilai dengan rata rata kolom kemudian membaginya dengan simpangan baku kolom tersebut sebagaimana dirumuskan pada Persamaan (1). Proses ini penting dilakukan karena keempat pilar IMD memiliki rentang dan variasi nilai yang berbeda, misalnya pilar Infrastruktur dan Ekosistem yang jauh lebih lebar dibandingkan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

pilar Pekerjaan sebagaimana terlihat pada Tabel 1. Tanpa standardisasi, pilar dengan rentang nilai lebih besar akan mendominasi perhitungan jarak Euclidean pada algoritma K-Means, sehingga kontribusi pilar lain menjadi kurang bermakna dalam pembentukan klaster. Setelah proses transformasi, setiap variabel memiliki rata rata mendekati nol dan simpangan baku mendekati satu, sehingga seluruh pilar memberikan bobot yang setara pada tahap penambangan data berikutnya.

4. Penambangan data (data mining)
   Pada tahap ini, algoritma K-Means Clustering diterapkan pada data yang telah distandarisasi menggunakan bahasa pemrograman Python. Proses dimulai dengan menginisialisasi titik centroid awal secara acak, kemudian setiap provinsi dikelompokkan ke centroid terdekat berdasarkan jarak Euclidean sebagaimana dirumuskan pada Persamaan (2), dan posisi centroid diperbarui secara iteratif hingga konvergen [5]. Untuk memastikan hasil yang stabil, proses inisialisasi dan iterasi diulang sebanyak sepuluh kali percobaan (n init) pada setiap nilai k yang diuji, sehingga risiko algoritma terjebak pada solusi optimum lokal dapat diminimalkan. Pengujian dilakukan secara sistematis pada rentang jumlah klaster k mulai dari dua hingga tujuh, dan hasil pengelompokan pada setiap nilai k tersebut disimpan untuk dievaluasi pada tahap selanjutnya.
5. Evaluasi (evaluation)
   Tahap evaluasi bertujuan untuk menentukan jumlah klaster yang paling optimal di antara seluruh nilai k yang telah diuji pada tahap sebelumnya. Evaluasi dilakukan melalui kombinasi dua metode, yaitu metode Elbow yang mengamati penurunan nilai Within Cluster Sum of Squares sebagaimana dirumuskan pada Persamaan (3), serta metode Silhouette Coefficient yang dirumuskan pada Persamaan (4) untuk mengukur seberapa baik pemisahan antarklaster secara kuantitatif [10]. Kedua metode ini digunakan secara bersamaan karena metode Elbow cenderung bersifat visual dan dapat menimbulkan interpretasi yang berbeda antarpeneliti, sedangkan Silhouette Coefficient memberikan angka pasti yang lebih objektif. Hasil dari kedua metode kemudian dibandingkan dan dipertimbangkan bersama aspek interpretasi kebijakan, sebelum akhirnya ditetapkan satu nilai k sebagai jumlah klaster final yang digunakan pada tahap berikutnya.
6. Interpretasi (interpretation)
   Tahap terakhir adalah menganalisis karakteristik dari setiap klaster yang telah terbentuk berdasarkan jumlah klaster optimal yang dipilih pada tahap evaluasi. Analisis dilakukan dengan menghitung rerata skor IMD dan keempat pilar pada masing masing klaster, kemudian mengurutkan klaster tersebut dari yang memiliki rerata IMD tertinggi hingga terendah agar dapat diberi label yang mudah dipahami, misalnya maju, berkembang, dan tertinggal. Selanjutnya, daftar provinsi anggota tiap klaster diperiksa satu per satu untuk melihat apakah pengelompokan yang dihasilkan algoritma sesuai dengan kondisi geografis maupun pembangunan wilayah yang sudah umum diketahui. Hasil interpretasi ini kemudian dikaitkan dengan kondisi riil masing masing provinsi dan dijadikan dasar untuk merumuskan implikasi kebijakan pada bagian pembahasan.



Gambar 1. Tahapan Penelitian

2.3 K-Means

K-Means adalah algoritma clustering dalam data mining yang digunakan untuk mengelompokkan data ke dalam sejumlah kelompok (cluster) berdasarkan tingkat kemiripan karakteristiknya [15],[16],[17]. Algoritma ini bekerja dengan menentukan sejumlah pusat cluster (centroid), kemudian mengalokasikan setiap data ke centroid terdekat dan memperbarui posisi centroid secara berulang hingga terbentuk kelompok yang stabil, Proses pengolahan data dilakukan menggunakan bahasa pemrograman Python bersama beberapa pustaka pendukung, yaitu Pandas untuk manipulasi data, Scikit Learn untuk proses standardisasi (StandardScaler), pembentukan model K-Means (KMeans), dan perhitungan Silhouette Coefficient (silhouette score), serta Matplotlib untuk keperluan visualisasi hasil klasterisasi.
[18],[19],[20].

Empat persamaan matematis digunakan sepanjang tahapan penelitian ini, yaitu persamaan standardisasi Z score, persamaan jarak Euclidean, persamaan Within Cluster Sum of Squares, serta persamaan Silhouette Coefficient. Keempat persamaan tersebut disajikan berturut turut sebagai berikut.

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

Keterangan: Z adalah nilai hasil standardisasi, x adalah nilai asli data pada suatu pilar, μ adalah nilai rata rata pilar tersebut, dan σ adalah simpangan baku pilar tersebut.

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

Keterangan: d(x, c) adalah jarak Euclidean antara data x dan centroid c, xi adalah nilai data pada atribut ke i, ci adalah nilai centroid pada atribut ke i, dan n adalah jumlah atribut yang digunakan, dalam penelitian ini n sama dengan empat.

$$WCSS = \sum_{i=1}^{n}(x_i - c_k)^2 \quad (3)$$

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

Keterangan: WCSS adalah total jarak kuadrat setiap data terhadap centroid klasternya masing masing, xi adalah data ke i, dan ck adalah centroid klaster tempat data xi berada. Nilai WCSS inilah yang diplot terhadap berbagai jumlah klaster k pada metode Elbow.

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

Keterangan: s(i) adalah nilai Silhouette Coefficient untuk data ke i, a(i) adalah rata rata jarak data ke i terhadap seluruh data lain dalam klaster yang sama, dan b(i) adalah rata rata jarak terkecil data ke i terhadap seluruh data pada klaster tetangga terdekat. Nilai s(i) berkisar antara negatif satu hingga satu, dengan nilai di atas 0,5 menunjukkan struktur klaster kuat, nilai antara 0,25 sampai 0,5 menunjukkan struktur klaster sedang, dan nilai di bawah 0,25 menunjukkan struktur klaster lemah [6].

3. HASIL DAN PEMBAHASAN

3.1 Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis data sekunder berupa data Indeks Masyarakat Digital (IMD) tahun 2025 pada tingkat provinsi. Data mencakup 38 provinsi di Indonesia beserta skor Indeks Masyarakat Digital secara keseluruhan dan skor empat pilar penyusunnya, yaitu pilar Infrastruktur dan Ekosistem, pilar Literasi Digital, pilar Pemberdayaan, serta pilar Pekerjaan. Keempat pilar tersebut digunakan sebagai variabel dalam proses klasterisasi karena secara konseptual mewakili empat dimensi utama kesiapan digital masyarakat sebagaimana dijelaskan pada kerangka IMDI [2]. Sebagai gambaran awal mengenai bentuk data yang digunakan, cuplikan tujuh provinsi dari total 38 provinsi disajikan pada Tabel 1.

Tabel 1. Data Indeks Masyarakat Digital Tahun 2025 (Cuplikan)

Provinsi	IMD	Infrastruktur	Literasi	Pemberdayaan	Pekerjaan
DKI Jakarta	56,97	75,34	53,73	54,07	42,06
Jawa Barat	52,05	66,91	52,37	44,65	44,01
Aceh	42,64	49,60	47,69	32,77	42,13
Sumatera Utara	42,87	53,23	44,69	33,54	41,85
Sulawesi Tenggara	38,41	48,19	44,15	24,02	40,67
Papua	37,26	35,86	43,23	31,02	40,44
Papua Pegunungan	31,11	28,12	36,72	19,76	45,69

Tabel 1 memperlihatkan bahwa setiap baris merepresentasikan satu provinsi dengan lima kolom numerik, yaitu skor IMD keseluruhan dan skor empat pilar penyusunnya. Kelima kolom tersebut memiliki rentang nilai yang berbeda beda, sebagaimana terlihat pada perbandingan antara DKI Jakarta dan Papua Pegunungan, sehingga struktur data semacam ini menjadi alasan utama mengapa proses standardisasi perlu dilakukan sebelum data dimasukkan ke dalam algoritma K-Means Clustering.

3.2 Deskripsi Data

Data yang digunakan dalam penelitian ini mencakup 38 provinsi di Indonesia dengan skor Indeks Masyarakat Digital (IMD) dan empat pilar penyusunnya pada tahun 2025. Statistik deskriptif dari kelima variabel tersebut disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Data IMD Tahun 2025 (38 Provinsi)

Variabel	Minimum	Maksimum	Rerata	Simpangan Baku
IMD	31,11	56,97	44,53	5,94
Infrastruktur dan Ekosistem	28,12	75,34	53,06	10,72
Literasi Digital	36,72	55,45	49,28	4,47
Pemberdayaan	14,49	54,07	34,32	8,42
Pekerjaan	37,51	50,29	42,91	2,46

Tabel 2 memperlihatkan bahwa skor IMD nasional berkisar antara 31,11 pada nilai terendah hingga 56,97 pada nilai tertinggi, dengan rerata sebesar 44,53. Rentang skor Infrastruktur dan Ekosistem serta pilar Pemberdayaan tergolong paling lebar dibandingkan pilar lainnya, sebagaimana ditunjukkan oleh nilai simpangan baku masing masing sebesar 10,72 dan 8,42. Kondisi ini mengindikasikan bahwa kesenjangan antarprovinsi paling nyata terjadi pada aspek pemerataan infrastruktur digital dan pemanfaatan teknologi digital secara produktif. Sebaliknya, pilar Pekerjaan memiliki simpangan baku paling kecil, yaitu 2,46, yang menunjukkan bahwa skor pilar ini relatif seragam pada hampir seluruh provinsi.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

3.3 Standardisasi Data

Sebelum proses klasterisasi dilakukan, keempat variabel pilar terlebih dahulu ditransformasikan menggunakan metode Z score sehingga setiap variabel memiliki rerata nol dan simpangan baku satu. Proses ini penting dilakukan mengingat rentang nilai pilar Infrastruktur dan Ekosistem yang jauh lebih lebar dibandingkan pilar Pekerjaan, sehingga tanpa standardisasi, perhitungan jarak Euclidean pada algoritma K-Means akan didominasi oleh pilar dengan rentang nilai yang lebih besar dan menyebabkan kontribusi pilar lain menjadi kurang bermakna dalam pembentukan klaster.

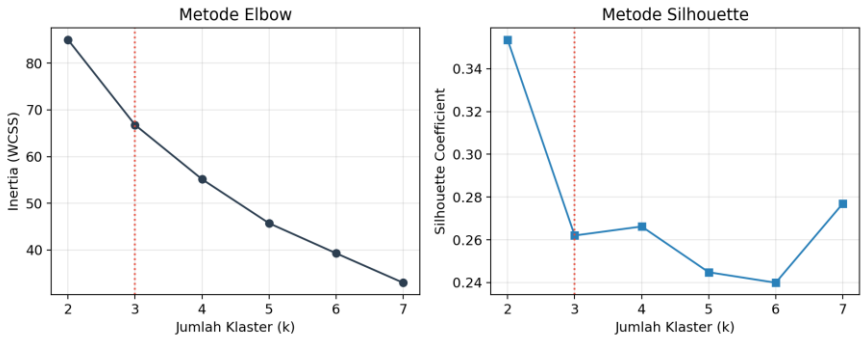
3.4 Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster optimal dilakukan dengan menguji algoritma K-Means pada rentang k sama dengan dua hingga tujuh, kemudian mengevaluasi setiap hasil menggunakan nilai Within Cluster Sum of Squares (WCSS) untuk metode Elbow dan nilai koefisien Silhouette. Hasil pengujian disajikan pada Tabel 3 dan divisualisasikan pada Gambar 1.

Tabel 3. Nilai WCSS dan Koefisien Silhouette untuk k = 2 sampai k = 7

Jumlah Klaster (k)	WCSS (Inertia)	Koefisien Silhouette	Kategori
1	85,03	0,354	Struktur sedang
2	66,79	0,262	Struktur lemah menuju sedang
3	55,14	0,266	Struktur lemah menuju sedang
4	45,71	0,245	Struktur lemah
5	39,27	0,240	Struktur lemah
6	32,96	0,277	Struktur lemah menuju sedang

Berikut Grafik Elbow pada Gambar 1 memperlihatkan penurunan nilai WCSS yang cukup tajam hingga k sama dengan tiga, kemudian penurunan mulai melandai pada k selanjutnya, sehingga titik siku secara visual berada pada k sama dengan tiga. Sementara itu, nilai koefisien Silhouette tertinggi secara absolut diperoleh pada k sama dengan dua, yaitu sebesar 0,354, namun pengelompokan pada k sama dengan dua hanya menghasilkan pembagian biner yang kurang mampu menggambarkan gradasi kesiapan digital antarprovinsi secara rinci. Nilai koefisien Silhouette pada k sama dengan tiga sebesar 0,262 relatif berdekatan dengan nilai pada k sama dengan empat sebesar 0,266, sehingga penambahan jumlah klaster tidak memberikan peningkatan kualitas pemisahan yang signifikan. Berdasarkan pertimbangan titik siku pada metode Elbow, kedekatan nilai Silhouette pada k tiga dan k empat, serta kebutuhan interpretasi kebijakan yang lazim menggunakan tiga kategori, yaitu tinggi, sedang, dan rendah [11], penelitian ini menetapkan jumlah klaster optimal sebanyak tiga klaster untuk proses klasterisasi provinsi berdasarkan Indeks Masyarakat Digital tahun 2025.



Gambar 1. *Grafik Metode Elbow (kiri) dan Grafik Metode Silhouette (kanan)*

3.5 Hasil Klasterisasi dan Karakteristik Klaster

Berdasarkan penetapan k sama dengan tiga, algoritma K-Means Clustering menghasilkan tiga kelompok provinsi dengan karakteristik yang berbeda. Rerata skor IMD dan keempat pilar pada masing masing klaster disajikan pada Tabel 4, sedangkan visualisasi sebaran klaster hasil reduksi dimensi menggunakan Principal Component Analysis (PCA) ditampilkan pada Gambar 2.

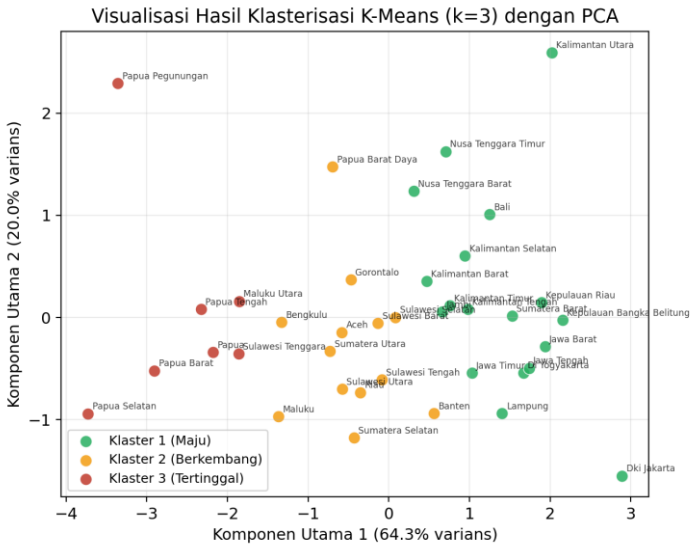
Tabel 4. Rerata Skor IMD dan Pilar pada Setiap Klaster

Klaster	Jumlah Provinsi	Rerata IMD	Infrastruktur	Literasi	Pemberdayaan	Pekerjaan
Klaster 1 (Maju)	18	49,18	60,35	52,38	40,12	44,46
Klaster 2 (Berkembang)	13	43,16	51,78	48,14	32,48	41,87
Klaster 3 (Tertinggal)	7	35,11	36,70	43,41	22,83	40,87

Berikut pada Gambar ditampilkan Visualisasi Hasil Klasterisasi K-Means (k=3) Menggunakan Reduksi Dimensi PCA

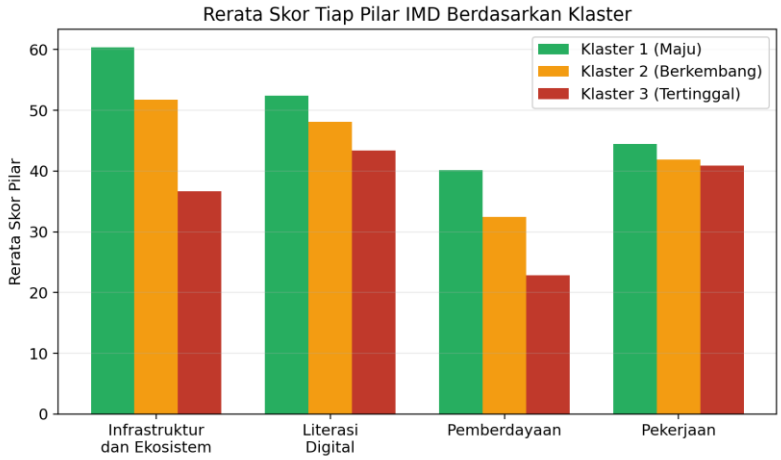
Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531



Gambar 2. Visualisasi Hasil Klasterisasi K-Means (k=3) Menggunakan Reduksi Dimensi PCA

Selanjutnya pada Gambar 3 merupakan Perbandingan Rerata Skor Tiap Pilar IMD Antarklaster



Gambar 3. Perbandingan Rerata Skor Tiap Pilar IMD Antarklaster

Klaster 1 atau kelompok provinsi maju digital beranggotakan 18 provinsi dengan rerata IMD tertinggi, yaitu 49,18. Klaster ini didominasi oleh provinsi di Pulau Jawa, Bali, Sumatera bagian tengah, serta sebagian besar wilayah Kalimantan, dengan DKI Jakarta menempati posisi tertinggi berkat skor pilar Infrastruktur dan Ekosistem yang mencapai 75,34. Daftar provinsi anggota Klaster 1 disajikan pada Tabel 5.

Tabel 5. Provinsi Anggota Klaster 1 (Maju Digital)

Provinsi	IMD	Infrastruktur	Literasi	Pemberdayaan	Pekerjaan
DKI Jakarta	56,97	75,34	53,73	54,07	42,06
Kepulauan Bangka Belitung	52,15	63,09	54,81	45,45	44,58
Jawa Barat	52,05	66,91	52,37	44,65	44,01
Jawa Tengah	51,19	65,73	53,75	41,95	43,21
DI Yogyakarta	51,13	69,73	52,93	39,37	43,12
Kepulauan Riau	50,77	57,97	55,45	44,35	44,63
Lampung	50,36	57,42	53,29	47,02	41,90
Sumatera Barat	49,70	58,00	54,15	42,32	44,08
Jawa Timur	49,17	67,22	49,87	37,95	42,73
Kalimantan Utara	48,91	55,60	54,72	37,44	50,29
Bali	48,22	63,98	50,97	34,27	46,32
Kalimantan Selatan	47,71	59,25	49,45	38,36	45,23
Kalimantan Tengah	47,31	52,41	54,55	38,79	43,62
Kalimantan Timur	47,17	56,32	50,80	38,51	43,82
Jambi	46,78	55,16	51,10	38,04	43,56

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 6
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Informatics, Electrical and Electronics Engineering
ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

Kalimantan Barat	46,22	54,09	48,49	39,03	44,26
Nusa Tenggara Barat	44,87	56,83	47,59	32,08	46,16
Nusa Tenggara Timur	44,54	51,21	54,88	28,46	46,75

Klaster 2 atau kelompok provinsi berkembang beranggotakan 13 provinsi dengan rerata IMD sebesar 43,16, yang sebagian besar berasal dari wilayah Sumatera bagian selatan, Sulawesi, serta beberapa provinsi hasil pemekaran seperti Papua Barat Daya. Daftar provinsi anggota Klaster 2 disajikan pada Tabel 6.

Tabel 6. Provinsi Anggota Klaster 2 (Berkembang)

Provinsi	IMD	Infrastruktur	Literasi	Pemberdayaan	Pekerjaan
Banten	48,01	63,17	48,16	39,62	41,51
Sumatera Selatan	45,01	54,26	44,39	40,99	40,29
Riau	44,65	57,47	45,45	35,69	41,27
Sulawesi Selatan	44,41	54,60	51,42	30,82	42,81
Sulawesi Tengah	44,01	48,65	53,03	33,76	41,11
Sulawesi Barat	43,52	49,78	51,72	31,68	42,42
Gorontalo	42,96	53,23	45,87	31,74	43,57
Sumatera Utara	42,87	53,23	44,69	33,54	41,85
Sulawesi Utara	42,73	51,56	50,04	30,23	40,68
Aceh	42,64	49,60	47,69	32,77	42,13
Bengkulu	40,65	52,91	42,24	28,57	42,09
Papua Barat Daya	40,34	42,06	48,39	28,52	45,55
Maluku	39,24	42,63	52,71	24,34	39,01

Klaster 3 atau kelompok provinsi tertinggal beranggotakan 7 provinsi dengan rerata IMD paling rendah, yaitu 35,11. Klaster ini secara mencolok didominasi oleh provinsi pemekaran di wilayah kepulauan Papua, yaitu Papua, Papua Tengah, Papua Barat, Papua Selatan, dan Papua Pegunungan, ditambah Sulawesi Tenggara dan Maluku Utara. Daftar provinsi anggota Klaster 3 disajikan pada Tabel 7.

Tabel 7. Provinsi Anggota Klaster 3 (Tertinggal)

Provinsi	IMD	Infrastruktur	Literasi	Pemberdayaan	Pekerjaan
Sulawesi Tenggara	38,41	48,19	44,15	24,02	40,67
Maluku Utara	37,43	42,48	46,35	22,76	41,61
Papua	37,26	35,86	43,23	31,02	40,44
Papua Tengah	36,02	30,14	44,13	30,27	41,13
Papua Barat	33,90	36,57	46,18	17,50	39,02
Papua Selatan	31,62	35,53	43,13	14,49	37,51
Papua Pegunungan	31,11	28,12	36,72	19,76	45,69

3.6 Pembahasan

Hasil klasterisasi menunjukkan pola kesenjangan digital yang sejalan dengan kondisi geografis dan tingkat pembangunan wilayah di Indonesia. Klaster 1 yang beranggotakan provinsi provinsi di Pulau Jawa, Bali, serta sebagian besar Kalimantan dan Sumatera bagian tengah unggul terutama pada pilar Infrastruktur dan Ekosistem dengan rerata 60,35, jauh di atas rerata Klaster 3 yang hanya 36,70. Selisih ini menegaskan bahwa pemerataan akses internet dan sarana teknologi informasi masih menjadi tantangan mendasar bagi wilayah timur Indonesia, khususnya kepulauan Papua.
Temuan yang paling menonjol dari Gambar 3 adalah pilar Pemberdayaan menjadi pembeda paling tajam antarklaster, dengan rerata berturut turut sebesar 40,12 pada Klaster 1, 32,48 pada Klaster 2, dan hanya 22,83 pada Klaster 3. Kondisi ini mengindikasikan bahwa kemampuan masyarakat dalam memanfaatkan teknologi digital secara produktif untuk kegiatan ekonomi, seperti transaksi keuangan digital dan perdagangan elektronik, masih sangat terbatas terutama di provinsi provinsi Klaster 3. Sebaliknya, pilar Pekerjaan memiliki rerata yang relatif seragam pada ketiga klaster, yaitu berkisar antara 40,87 hingga 44,46, yang menunjukkan bahwa pemanfaatan keahlian digital dalam dunia kerja secara umum masih berada pada tingkat yang belum optimal di seluruh provinsi, tidak terbatas hanya pada wilayah tertinggal saja.

Dominasi provinsi pemekaran Papua pada Klaster 3 memperkuat temuan laporan resmi IMDI yang menyebutkan bahwa peningkatan indeks nasional dari 37,80 pada tahun 2022 menjadi 43,18 pada tahun 2023 belum sepenuhnya dirasakan merata oleh seluruh wilayah . Hasil ini sejalan dengan penelitian Nurfathullah dan Purnamasari [5] yang juga menemukan pengelompokan provinsi berdasarkan indikator sosial ekonomi cenderung menempatkan sebagian provinsi di kawasan timur Indonesia pada klaster dengan capaian paling rendah. Dengan demikian, kebijakan percepatan transformasi digital yang bersifat seragam secara nasional berpotensi kurang efektif apabila tidak mempertimbangkan perbedaan karakteristik antarklaster yang telah teridentifikasi dalam penelitian ini.

Bagi provinsi pada Klaster 1, strategi kebijakan yang relevan adalah mempertahankan capaian infrastruktur yang sudah baik sambil mendorong peningkatan pilar Pemberdayaan dan Pekerjaan agar pemanfaatan teknologi digital

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

semakin berdampak pada produktivitas ekonomi. Bagi provinsi pada Klaster 2, intervensi yang diperlukan adalah penguatan literasi digital dasar dan perluasan akses layanan keuangan digital untuk mendorong pergeseran menuju kelompok yang lebih maju. Adapun bagi provinsi pada Klaster 3, khususnya wilayah kepulauan Papua, prioritas kebijakan perlu diarahkan pada pembangunan infrastruktur dasar telekomunikasi terlebih dahulu, mengingat rendahnya skor pilar Infrastruktur menjadi hambatan struktural sebelum program literasi maupun pemberdayaan digital dapat berjalan efektif.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia berdasarkan empat pilar Indeks Masyarakat Digital tahun 2025 menggunakan algoritma K-Means Clustering. Berdasarkan kombinasi metode Elbow dan Silhouette Coefficient, diperoleh tiga klaster optimal dengan nilai koefisien Silhouette sebesar 0,262. Klaster 1 atau kelompok maju digital beranggotakan 18 provinsi dengan rerata IMD 49,18 dan unggul terutama pada pilar Infrastruktur dan Ekosistem. Klaster 2 atau kelompok berkembang beranggotakan 13 provinsi dengan rerata IMD 43,16. Klaster 3 atau kelompok tertinggal beranggotakan 7 provinsi dengan rerata IMD 35,11 dan didominasi oleh provinsi pemekaran di wilayah kepulauan Papua. Pilar Pemberdayaan terbukti menjadi pembeda paling tajam antarklaster, sedangkan pilar Pekerjaan relatif seragam di seluruh provinsi, yang mengindikasikan bahwa pemanfaatan keahlian digital dalam dunia kerja masih menjadi tantangan bersama tanpa memandang tingkat kemajuan digital suatu wilayah. Hasil pengelompokan ini dapat dijadikan dasar pertimbangan bagi Kementerian Komunikasi dan Digital serta pemerintah daerah dalam merancang program transformasi digital yang lebih tepat sasaran, misalnya dengan memprioritaskan pembangunan infrastruktur telekomunikasi dasar bagi provinsi pada Klaster 3, penguatan literasi dan pemberdayaan digital bagi provinsi pada Klaster 2, serta optimalisasi pemanfaatan teknologi digital untuk produktivitas ekonomi bagi provinsi pada Klaster 1. Penelitian selanjutnya disarankan untuk melibatkan data pada tingkat subpilar yang lebih rinci, membandingkan hasil K-Means dengan algoritma clustering lain seperti Fuzzy C Means atau Hierarchical Clustering, serta menganalisis data IMD secara time series pada beberapa tahun pengukuran agar pola perkembangan maupun pergeseran klaster antarwaktu dapat diamati secara lebih mendalam.

REFERENCESS

[1] Kementerian Komunikasi dan Informatika, "Peluncuran Hasil Pengukuran Indeks Masyarakat Digital Indonesia (IMDI) Tahun 2024," Badan Pengembangan Sumber Daya Manusia Kominfo, 2024. [Online]. Available: https://bpsdm.komdigi.go.id/berita-peluncuran-hasil-pengukuran-indeks-masyarakat-digital-indonesia-imdi-tahun--46-2092

[2] Badan Pengembangan Sumber Daya Manusia Komunikasi dan Digital, "Pilar Indeks Masyarakat Digital Indonesia," Kementerian Komunikasi dan Digital Republik Indonesia, 2025. [Online]. Available: https://imdi.sdmdigital.id/tentang-kami/tentang-pilar

[3] Indonesia Baik, "Masyarakat Indonesia Semakin Digital," Kementerian Komunikasi dan Informatika, 2024. [Online]. Available: https://indonesiabaik.id/infografis/masyarakat-indonesia-semakin-digital

[4] R. A. Manik and A. Ariesta, "Data Mining untuk Klasterisasi Provinsi di Indonesia Berdasarkan Data Pokok Pendidikan Nasional," Jurnal Ticom, vol. 11, no. 3, pp. 159–164, 2023, doi: 10.70309/ticom.v11i3.115.

[5] M. Nurfathullah and I. Purnamasari, "Implementasi K-Means untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Jumlah Pengangguran Terbuka," Jurnal Mahasiswa Teknik Informatika, vol. 8, no. 2, pp. 2277–2282, 2024, doi: 10.36040/jati.v8i2.9466.

[6] M. Maesaroh, T. Nur Padilah, and J. Haerul Jaman, "Penerapan Algoritma K-Means Clustering pada Pengelompokan Daerah Penyebaran Diare di Provinsi Jawa Barat," Jurnal Mahasiswa Teknik Informatika, vol. 7, no. 4, pp. 2783–2787, 2023, doi: 10.36040/jati.v7i4.7208.

[7] S. Dewi and M. A. I. Pakereng, "Implementasi Principal Component Analysis pada K-Means untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang," Jurnal Ilmiah Penelitian dan Pembelajaran Informatika, vol. 8, no. 4, pp. 1186–1195, 2023, doi: 10.29100/jipi.v8i4.4101.

[8] M. W. Diantika, A. Rusgiyono, and B. A. Saputra, "Perbandingan Metode Optimasi Silhouette, Elbow, dan Gap Statistics dalam Menentukan Banyaknya Cluster Terbaik pada Analisis K-Means Clustering," Jurnal Gaussian, vol. 14, no. 2, pp. 335–344, 2025, doi: 10.14710/j.gauss.14.2.335-344.

[9] M. R. Syahkur, D. Hartama, and Solikhun, "Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita," Jurnal Sains dan Teknologi, vol. 13, no. 3, pp. 487–496, 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[10] M. Guntara and N. Lutfi, "Cacah Klaster pada Klasterisasi dengan Algoritma K-Means Menggunakan Silhouette Coefficient dan Elbow Method," Jurnal Teknologi Informasi, vol. 2, no. 1, pp. 43–52, 2023, doi: 10.26798/juti.v2i1.944.

[11] A. Atira and B. N. Sari, "Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokkan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan," Jurnal Ilmiah Wahana Pendidikan, vol. 9, no. 17, pp. 76–86, 2023, doi: 10.5281/zenodo.8282638.

[12] B. W. Pudjianto, Nusirwan, A. Susenna, D. Kusumasari, L. Agustina, Y. A. S. Yan, C. M. Zellatifanny, A. M. Y. Ariyanti, and F. Imran, Indeks Masyarakat Digital Indonesia 2025. Jakarta, Indonesia: Badan Pengembangan SDM Komunikasi dan Digital, 2025.

[13] M. R. Putri, G. S. Nugraha, and R. Dwiyansaputra, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering," Journal of Computer Science and Informatics Engineering (J-Cosine), vol. 7, no. 1, 2023.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

ISSN 2807-9507 (Media Online)
Vol 6, No 1, September 2026, Page: 1-9
https://djournals.com/jieee
DOI: 10.47065/jieee.v6i1.3531

[14] A. B. Saputra, U. P. Sanjaya, and I. A. Sa'ida, "Algoritma K-Means untuk Mengelompokkan Tingkat Pengangguran Terbuka (TPT) menurut Provinsi di Indonesia," Jurnal Ilmiah Informatika Global, vol. 15, no. 2, pp. 75–81, 2024, doi: 10.36982/jiig.v15i2.4359.

[15] S. J. Udayana and I. M. Widiartha, "Algoritma K-Means untuk Clustering Provinsi di Indonesia Berdasarkan Kasus Stunting," Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA), vol. 2, no. 3, pp. 573–582, 2024, doi: 10.24843/JNATIA.2024.v02.i03.p16.

[16] S. Juanita and R. D. Cahyono, "K-Means Clustering with Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews," Jurnal Teknik Informatika (JUTIF), vol. 5, no. 1, pp. 283–289, 2024, doi: 10.52436/1.jutif.2024.5.1.1349.

[17] C. Tacharri, A. Rohmani, and A. Fahmi, "Strategic Clustering of Poverty Areas in Central Java Using K-Means and Silhouette Evaluation," Sinkron: Jurnal dan Penelitian Teknik Informatika, vol. 9, no. 2, pp. 895–904, 2025, doi: 10.33395/sinkron.v9i2.14734.

[18] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient dan Metode Elbow pada Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator IPM dengan K-Medoids," Jurnal Gaussian, vol. 13, no. 1, pp. 13–24, 2024, doi: 10.14710/j.gauss.13.1.13-24.

[19] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," J. Teknoif Teknik Informatika Institut Teknologi Padang, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[20] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," IEEE Access, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.