E-ISSN: 3025-0994
Halaman 137-145
Volume 4 Nomor 4 Tahun 2026

Analisis Clustering Pola Penggunaan Internet Berdasarkan Wilayah di Indonesia Menggunakan Algoritma K-Means dan Elbow Method

Juwardi Wafdan[1], Abdullah[2]

[1,2] Program Studi Sistem Informasi, Universitas Islam Indragiri
e-mail: juwardiwafdan@gmail.com, abdialam@gmail.com

Abstrak

Ketimpangan akses internet antar wilayah di Indonesia masih menjadi persoalan yang belum sepenuhnya terselesaikan hingga saat ini. Kesenjangan tersebut tidak semata-mata soal ada atau tidaknya koneksi, melainkan juga menyentuh aspek kualitas jaringan, ketersediaan infrastruktur, dan seberapa dalam masyarakat di tiap daerah benar-benar memanfaatkan internet dalam kehidupan sehari-hari. Penelitian ini mengelompokkan seluruh provinsi di Indonesia berdasarkan pola dan tingkat penggunaan internet menggunakan algoritma K-Means Clustering dengan Elbow Method untuk menentukan jumlah klaster optimal. Data bersumber dari Badan Pusat Statistik (BPS) dan portal Satu Data Indonesia (data.go.id), mencakup lima indikator: persentase penduduk pengguna internet, persentase rumah tangga berakses internet, kepemilikan telepon seluler, kepadatan jaringan broadband tetap, dan rata-rata kecepatan koneksi internet per provinsi. Hasil analisis membentuk tiga klaster wilayah dengan profil digital yang berbeda secara signifikan. Evaluasi menggunakan Silhouette Score menghasilkan nilai rata-rata 0,612, mengindikasikan struktur pengelompokan yang tergolong baik. Temuan ini diharapkan menjadi rujukan bagi pengambil kebijakan dalam merancang program pemerataan infrastruktur digital yang lebih tepat sasaran.

Kata Kunci: K-Means Clustering, Elbow Method, Penggunaan Internet, Kesenjangan Digital, Indonesia

Abstract

The disparity in internet access across regions in Indonesia remains an unresolved challenge. This gap is not merely about the presence or absence of connectivity, but also encompasses network quality, infrastructure availability, and the depth of internet utilization in daily life. This study clusters all Indonesian provinces based on internet usage patterns using the K-Means Clustering algorithm, with the Elbow Method applied to determine the optimal number of clusters. Data were obtained from the Central Statistics Agency (BPS) and the Satu Data Indonesia portal (data.go.id), covering five indicators: percentage of internet users, percentage of households with internet access, mobile phone ownership, fixed broadband density, and average internet speed per province. The analysis yielded three

clusters with significantly different digital profiles. Cluster quality evaluation using the Silhouette Score produced an average of 0.612, indicating a well-formed clustering structure. These findings are expected to guide policymakers in designing more targeted digital infrastructure programs.

Keywords : *K-Means Clustering, Elbow Method, Internet Usage, Digital Divide, Indonesia*

LATAR BELAKANG

Internet bukan lagi sekadar fasilitas tambahan dalam kehidupan modern—ia telah menjelma menjadi infrastruktur dasar yang menopang hampir seluruh sektor kehidupan manusia, mulai dari pendidikan, kesehatan, perdagangan, hingga tata kelola pemerintahan. Revolusi digital yang berlangsung sejak awal abad ke-21 telah mengubah cara manusia mengakses informasi, menjalankan pekerjaan, berinteraksi sosial, bahkan cara mereka berpartisipasi dalam proses demokrasi. Pergeseran ini terjadi secara global dengan kecepatan yang belum pernah terjadi sebelumnya dalam sejarah teknologi manusia, namun dampaknya tidak dirasakan secara merata oleh semua negara, bahkan tidak merata di dalam satu negara sekalipun. Di negara-negara berkembang dengan geografi yang kompleks dan populasi yang tersebar seperti Indonesia, ketidakmerataan ini menjadi tantangan struktural yang jauh lebih dalam dari sekadar soal pembangunan menara pemancar sinyal (Cheng et al., 2022).

Di Indonesia, pertumbuhan pengguna internet memang berlangsung dengan kecepatan yang mengesankan. Dari sekitar 88 juta jiwa pada 2015, angka tersebut melonjak menjadi lebih dari 215 juta jiwa pada 2023, menjadikan Indonesia salah satu pasar digital terbesar di Asia Tenggara sekaligus negara dengan jumlah pengguna internet terbesar keempat di dunia (BPS, 2023). Ekosistem digital nasional pun berkembang pesat—platform e-commerce lokal tumbuh menjadi pemain regional, layanan transportasi berbasis aplikasi mengubah wajah mobilitas perkotaan, dan transaksi keuangan digital menembus angka yang belum pernah tercatat sebelumnya. Dari permukaan, gambaran ini tampak sangat optimistis. Namun angka nasional yang besar itu menyimpan sebuah ironi yang tidak bisa diabaikan: pertumbuhan tersebut sangat terkonsentrasi secara geografis dan tidak menyentuh seluruh lapisan wilayah secara proporsional.

Ketimpangan akses internet antara wilayah perkotaan dan pedesaan, antara Pulau Jawa dan kawasan timur Indonesia, sudah lama menjadi persoalan yang diakui namun belum sepenuhnya diselesaikan. Provinsi-provinsi di Jawa dan Bali mencatatkan tingkat penetrasi internet di atas 80 persen, dengan infrastruktur broadband yang relatif mapan dan kecepatan koneksi yang sudah memadai untuk mendukung layanan digital generasi berikutnya (BPS, 2023). Di sisi yang berlawanan, sejumlah provinsi di Papua, Maluku, dan Nusa Tenggara masih berjuang di bawah angka 50 persen dengan kecepatan koneksi yang jauh dari cukup bahkan untuk menjalankan aplikasi dasar berbasis internet. Kesenjangan ini bukan sekadar soal ada atau tidaknya sinyal—ia menyentuh dimensi yang jauh lebih dalam dan berlapis. Pertama, dimensi akses fisik yang mencakup ketersediaan jaringan dan keterjangkauan perangkat. Kedua, dimensi kualitas yang menyangkut kecepatan, stabilitas,

dan keandalan koneksi. Ketiga, dimensi pemanfaatan yang berkaitan dengan kemampuan dan kemauan masyarakat untuk menggunakan internet secara produktif dalam kehidupan nyata mereka (Kominfo, 2023).

Kondisi inilah yang dalam literatur akademis dikenal sebagai digital divide—sebuah konsep yang merujuk pada jurang pemisah antara mereka yang memiliki akses bermakna terhadap teknologi digital dan mereka yang tidak. Penelitian lintas negara telah menunjukkan secara konsisten bahwa digital divide, jika dibiarkan terus melebar, akan secara langsung memperparah ketimpangan ekonomi, memperlambat mobilitas sosial, dan menghambat pembangunan manusia di wilayah-wilayah yang tertinggal (Cheng et al., 2022; Suryana et al., 2023). Di Indonesia, relevansi dari peringatan ini sangat nyata. Suryana et al. (2023) menemukan korelasi positif yang signifikan antara tingkat penetrasi internet suatu daerah dengan pertumbuhan PDRB per kapita-nya selama periode 2017 hingga 2022. Artinya, wilayah-wilayah yang tertinggal secara digital tidak hanya rugi dalam hal akses informasi—mereka juga kehilangan peluang pertumbuhan ekonomi yang nyata dan terukur.

Persoalan ini semakin mendesak untuk ditangani mengingat arah kebijakan Indonesia yang semakin mengandalkan platform digital sebagai tulang punggung layanan publik. Program Satu Data Indonesia, sistem layanan administrasi kependudukan berbasis daring, telemedicine yang diperluas pasca pandemi COVID-19, platform pembelajaran jarak jauh, hingga program digitalisasi UMKM yang menjadi prioritas pemerintah—semuanya mensyaratkan konektivitas internet yang memadai sebagai prasyarat dasar yang tidak bisa ditawar (Kominfo, 2023). Ketika sebagian besar penduduk di suatu wilayah belum memiliki akses yang layak, program-program tersebut praktis tidak akan bisa dirasakan manfaatnya oleh mereka yang justru paling membutuhkan. Ini menciptakan paradoks kebijakan yang serius: investasi digital negara yang besar justru berpotensi hanya dinikmati oleh wilayah-wilayah yang sudah lebih maju, sementara wilayah tertinggal terus terjebak dalam siklus keterbelakangan yang sulit diputus tanpa intervensi yang tepat sasaran (Prasetyo et al., 2022).

Ironisnya, upaya untuk merancang intervensi yang tepat sasaran itu sendiri kerap terhambat oleh minimnya pemahaman yang mendalam dan berbasis data tentang pola sesungguhnya di lapangan. Laporan statistik agregat nasional memang tersedia dan mudah diakses, namun data seperti itu hanya mampu memberikan gambaran rata-rata yang secara inheren mengaburkan perbedaan nyata antar wilayah. Rata-rata nasional penetrasi internet sebesar 66 persen misalnya, tidak menceritakan apa pun tentang bagaimana 34 provinsi di Indonesia tersebar di sekitar angka tersebut—mana yang jauh di atas, mana yang jauh di bawah, dan mana yang sedang dalam proses transisi menuju kematangan digital. Lebih jauh, angka tunggal semacam itu tidak mampu memperlihatkan apakah suatu provinsi tertinggal karena masalah infrastruktur fisik, karena keterjangkauan ekonomi, atau karena faktor literasi dan adopsi teknologi yang lebih bersifat sosial-budaya (Hidayat & Permana, 2022; Nugroho & Sari, 2023).

Yang dibutuhkan adalah sebuah pendekatan analitis yang mampu memilah kondisi digital 34 provinsi secara objektif, menemukan pola pengelompokan yang mencerminkan realitas di lapangan, dan menghasilkan kategori-kategori yang dapat langsung dijadikan dasar perumusan kebijakan yang berbeda untuk konteks yang berbeda. Di sinilah teknik

clustering dalam kerangka data mining menawarkan solusi yang relevan, terukur, dan dapat diandalkan. Clustering tidak memerlukan asumsi awal tentang berapa banyak kategori yang ada atau seperti apa bentuknya—ia membiarkan struktur data itu sendiri yang memandu proses pengelompokan, sehingga hasilnya lebih jujur dan tidak bias oleh persepsi pembuat kebijakan (Han et al., 2011; Tan et al., 2019). Pendekatan semacam ini sudah terbukti efektif dalam berbagai konteks penelitian wilayah di Indonesia, termasuk untuk pengelompokan data kemiskinan (Azzahro et al., 2024), pemetaan adopsi teknologi di daerah 3T (Prasetyo et al., 2022), dan analisis ketimpangan pembangunan manusia antarkabupaten (Gunawan et al., 2023).

Dari sekian banyak algoritma clustering yang tersedia, K-Means adalah pilihan yang paling banyak digunakan dan telah teruji dalam ribuan studi selama lebih dari enam dekade sejak pertama kali diperkenalkan oleh MacQueen pada 1967. Ulasan komprehensif yang dilakukan oleh Ikotun et al. (2023) terhadap lebih dari seratus varian K-Means mengonfirmasi bahwa algoritma ini tetap relevan dan dominan dalam praktik penelitian modern, terutama karena kemampuannya memproses data multivariat dalam skala besar dengan kompleksitas komputasi yang terjangkau, sambil tetap menghasilkan klaster yang mudah diinterpretasikan bahkan oleh pemangku kepentingan yang tidak memiliki latar belakang teknis mendalam. Untuk mengatasi kelemahan utama K-Means dalam penentuan jumlah klaster, penelitian ini menggunakan Elbow Method—sebuah teknik visual yang secara empiris mengidentifikasi titik optimal berdasarkan kurva penurunan nilai Within-Cluster Sum of Squares (WCSS)—yang telah terbukti menghasilkan pengelompokan yang lebih stabil dan dapat dipertanggungjawabkan dibandingkan pemilihan jumlah klaster secara intuitif (Syakur et al., 2018).

Penelitian ini hadir untuk mengisi celah yang selama ini belum banyak disentuh: pengelompokan seluruh provinsi di Indonesia berdasarkan pola penggunaan internet secara komprehensif dan multidimensi menggunakan pendekatan machine learning berbasis data publik. Dengan memanfaatkan data yang tersedia secara terbuka dari Badan Pusat Statistik dan portal Satu Data Indonesia, penelitian ini berupaya menghasilkan pemetaan digital yang tidak hanya akurat secara statistis tetapi juga bermakna secara kebijakan. Hasil pengelompokan yang diperoleh diharapkan mampu memberikan gambaran yang lebih jujur tentang di mana sesungguhnya posisi masing-masing provinsi dalam lanskap digital Indonesia, sehingga program pemerataan infrastruktur digital yang dirancang pemerintah dapat benar-benar menyasar kelompok yang membutuhkan, bukan sekadar memperluas layanan di wilayah yang sudah siap menerimanya (Prasetyo et al., 2022; Suryana et al., 2023; Kominfo, 2023).

METODE

a. Jenis dan Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder yang sepenuhnya tersedia secara publik dan dapat diakses secara daring. Data dikumpulkan dari dua sumber utama: portal resmi Badan Pusat Statistik (BPS) di bps.go.id dan portal Satu Data Indonesia di data.go.id. Kedua sumber ini dipilih karena menyediakan data yang terstandarisasi, diperbarui secara berkala, dan telah digunakan secara luas dalam penelitian-

penelitian akademis di Indonesia (Azzahro et al., 2024; Suryana et al., 2023). Unit analisis adalah 34 provinsi di Indonesia dengan periode data tahun 2022–2023.

b. Variabel Penelitian

Terdapat lima variabel yang digunakan sebagai fitur dalam proses clustering. Pertama, persentase penduduk berusia di atas lima tahun yang menggunakan internet dalam tiga bulan terakhir. Kedua, persentase rumah tangga yang memiliki akses internet di tempat tinggal. Ketiga, persentase penduduk yang memiliki dan menggunakan telepon seluler aktif. Keempat, jumlah pelanggan layanan broadband tetap per 100 penduduk. Kelima, rata-rata kecepatan unduh koneksi internet berdasarkan data Speedtest Global Index yang dikompilasi per provinsi (BPS, 2023; Satu Data Indonesia, 2023). Kelima variabel ini merepresentasikan kondisi digital suatu wilayah dari dua dimensi: dimensi akses dan dimensi pemanfaatan (Prasetyo et al., 2022).

c. Tahapan Penelitian

Penelitian ini mengikuti kerangka Knowledge Discovery in Databases (KDD) yang terdiri atas lima tahap berurutan (Han et al., 2011). Tahap pertama adalah seleksi data, yaitu memilih variabel-variabel yang relevan dari keseluruhan dataset. Tahap kedua adalah preprocessing, mencakup penanganan nilai kosong menggunakan imputasi rata-rata dan eliminasi data duplikat. Tahap ketiga adalah transformasi data menggunakan normalisasi Z-Score agar seluruh variabel berada pada skala yang setara dan tidak ada variabel yang mendominasi proses clustering hanya karena perbedaan satuan (Tan et al., 2019). Tahap keempat adalah data mining menggunakan K-Means dengan Elbow Method untuk menentukan k optimal pada rentang k=2 hingga k=10. Tahap kelima adalah interpretasi dan evaluasi hasil clustering menggunakan Silhouette Score (Rousseeuw, 1987).

d. Implementasi dan Alat Analisis

Seluruh proses analisis diimplementasikan menggunakan bahasa pemrograman Python versi 3.10 dengan library scikit-learn untuk proses normalisasi dan clustering, pandas dan numpy untuk manipulasi data, serta matplotlib dan seaborn untuk visualisasi (Pedregosa et al., 2011). Inisialisasi centroid menggunakan metode k-means++ yang terbukti lebih stabil dibandingkan inisialisasi acak murni (Arthur & Vassilvitskii, 2007). Proses clustering dijalankan sebanyak 10 kali dengan inisialisasi berbeda untuk memastikan hasil yang diperoleh bukan produk dari inisialisasi yang kebetulan menguntungkan.

HASIL DAN PEMBAHASAN

a. Hasil Preprocessing Data

Setelah melalui tahap seleksi dan preprocessing, dataset akhir yang digunakan terdiri dari 34 baris data yang merepresentasikan 34 provinsi di Indonesia dengan 5 variabel fitur. Tidak ditemukan nilai kosong pada dataset setelah proses pembersihan data. Proses normalisasi Z-Score menghasilkan representasi data dengan mean nol dan standar deviasi satu untuk setiap fitur, sehingga tidak ada satu variabel pun yang mendominasi proses clustering semata karena perbedaan skala pengukuran (Tan et al., 2019). Secara deskriptif, variabel persentase pengguna internet memiliki rentang nilai yang paling lebar di antara kelima variabel, mencerminkan besarnya kesenjangan antarprovinsi dalam hal akses dan pemanfaatan internet (BPS, 2023).

b. Penentuan Jumlah Klaster Optimal

Analisis Elbow Method dilakukan dengan menghitung nilai WCSS untuk setiap nilai k dari 2 hingga 10. Hasil perhitungan menunjukkan penurunan nilai WCSS yang paling tajam terjadi pada transisi k=1 ke k=2 dan k=3. Setelah k=3, kurva mulai melandai dan tidak menunjukkan penurunan yang berarti, membentuk pola yang secara visual menyerupai siku (Syakur et al., 2018). Berdasarkan pengamatan ini, nilai k=3 ditetapkan sebagai jumlah klaster optimal. Keputusan ini juga dikonfirmasi oleh nilai Silhouette Score tertinggi yang dicapai pada k=3, yaitu sebesar 0,612, yang termasuk dalam kategori 'baik' menurut Rousseeuw (1987).

c. Karakteristik Klaster

Setelah proses clustering selesai, seluruh 34 provinsi dialokasikan ke dalam salah satu dari tiga klaster yang terbentuk. Tabel 1 merangkum karakteristik rata-rata setiap klaster berdasarkan lima variabel fitur yang digunakan.

Tabel 1. Karakteristik Rata-rata Setiap Klaster Hasil K-Means

Variabel	Klaster 1 (Tinggi)	Klaster 2 (Sedang)	Klaster 3 (Rendah)
Pengguna Internet (%)	82,4	67,1	48,3
RT Berakses Internet (%)	78,6	60,2	41,7
Kepemilikan HP (%)	91,3	80,5	67,9
Broadband / 100 Pddk	12,8	7,3	3,1
Kecepatan Internet (Mbps)	24,6	15,2	8,7

Klaster 1, yang dikategorikan sebagai wilayah dengan akses internet tinggi, didominasi oleh provinsi-provinsi di Pulau Jawa, Bali, dan beberapa wilayah Kalimantan yang menjadi pusat ekonomi regional. Rata-rata persentase pengguna internet di klaster ini mencapai 82,4 persen, jauh melampaui rata-rata nasional. Ketersediaan infrastruktur broadband yang relatif baik dengan nilai 12,8 pelanggan per 100 penduduk menjadi salah satu faktor pembeda utama klaster ini dari dua klaster lainnya (BPS, 2023).

Klaster 2 mencakup sebagian besar provinsi di Sumatera dan Sulawesi yang tengah mengalami pertumbuhan digital yang pesat namun belum merata. Dengan rata-rata pengguna internet sebesar 67,1 persen dan kecepatan koneksi 15,2 Mbps, wilayah-wilayah ini berada pada fase transisi menuju kematangan digital. Kondisi ini sejalan dengan temuan Nugroho dan Sari (2023) yang menyatakan bahwa segmen 'menengah' cenderung memiliki karakteristik yang paling beragam dan dinamis di antara klaster yang terbentuk.

Klaster 3, dengan tingkat akses internet terendah, sebagian besar terdiri dari provinsi-provinsi di kawasan Indonesia Timur, termasuk Papua, Maluku, dan Nusa Tenggara. Rata-rata kecepatan koneksi di klaster ini hanya 8,7 Mbps dengan kepadatan broadband 3,1 per 100 penduduk—angka yang jauh di bawah standar minimum yang

direkomendasikan untuk mendukung layanan digital berbasis cloud (Kominfo, 2023). Kondisi ini mencerminkan apa yang oleh Prasetyo et al. (2022) disebut sebagai 'jebakan infrastruktur', di mana minimnya investasi awal terus menghasilkan kondisi tertinggal yang sulit dikejar.

d. Evaluasi Kualitas Klaster

Evaluasi kualitas hasil clustering dilakukan menggunakan Silhouette Score, yang mengukur seberapa mirip setiap objek data dengan klasternya sendiri dibandingkan dengan klaster lain (Rousseeuw, 1987). Hasil evaluasi menunjukkan nilai rata-rata Silhouette Score untuk k=3 adalah 0,612. Nilai ini berada dalam kategori 'baik' karena melampaui ambang 0,50 yang ditetapkan dalam literatur (Rousseeuw, 1987). Klaster 1 memperoleh nilai Silhouette tertinggi sebesar 0,71, diikuti Klaster 3 sebesar 0,63, dan Klaster 2 sebesar 0,51. Nilai Klaster 2 yang lebih rendah mengonfirmasi bahwa kelompok wilayah 'sedang' memiliki keragaman internal yang lebih tinggi dibandingkan dua klaster lainnya (Wulandari & Kusuma, 2024).

e. Implikasi Kebijakan

Temuan penelitian ini memiliki implikasi langsung bagi pengambil kebijakan di bidang digital. Untuk Klaster 1, fokus kebijakan dapat diarahkan pada peningkatan kualitas layanan, pemanfaatan digital yang lebih produktif, serta pengembangan ekosistem digital berbasis inovasi. Untuk Klaster 2, prioritas utama adalah percepatan pemerataan infrastruktur dalam wilayah, mengingat besarnya variasi internal klaster ini. Untuk Klaster 3, intervensi yang paling mendasar adalah pembangunan infrastruktur fisik jaringan, pemberian subsidi akses, serta program literasi digital yang disesuaikan dengan konteks lokal (Kominfo, 2023; Suryana et al., 2023). Pendekatan berbasis data semacam ini sejalan dengan rekomendasi Cheng et al. (2022) bahwa kebijakan digital di negara kepulauan tidak bisa menggunakan formula seragam untuk seluruh wilayah.

SIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means Clustering dengan Elbow Method untuk mengelompokkan 34 provinsi di Indonesia berdasarkan pola penggunaan internet. Hasil analisis menghasilkan tiga klaster yang memiliki karakteristik berbeda secara signifikan: wilayah dengan akses tinggi (11 provinsi), sedang (14 provinsi), dan rendah (9 provinsi). Penentuan k=3 sebagai klaster optimal diperoleh melalui analisis Elbow Method dan dikonfirmasi oleh nilai Silhouette Score 0,612 yang termasuk kategori baik (Rousseeuw, 1987). Temuan ini mengonfirmasi bahwa kesenjangan digital di Indonesia masih nyata dan terstruktur secara geografis, di mana wilayah barat cenderung memiliki tingkat penetrasi internet yang jauh lebih tinggi dibandingkan wilayah timur (BPS, 2023; Kominfo, 2023). Hasil ini diharapkan dapat menjadi landasan empiris bagi pemerintah dalam memprioritaskan program pemerataan infrastruktur digital, khususnya untuk wilayah-wilayah dalam Klaster 3. Untuk penelitian selanjutnya, disarankan untuk menggunakan data time-series guna mengamati perubahan pola clustering dari waktu ke waktu, serta mempertimbangkan penggunaan metrik evaluasi tambahan seperti Davies-Bouldin Index (Wulandari & Kusuma, 2024) untuk memperkuat validitas hasil.

Saran

Berdasarkan hasil dan keterbatasan penelitian ini, terdapat beberapa saran yang dapat dijadikan pertimbangan untuk penelitian selanjutnya. Pertama, penelitian mendatang disarankan untuk menggunakan data time-series agar perubahan pola klaster dari waktu ke waktu dapat diamati secara dinamis, sehingga kebijakan yang dihasilkan tidak hanya menggambarkan kondisi statis melainkan juga tren perkembangan digital tiap wilayah. Kedua, penambahan variabel seperti indeks literasi digital, tingkat pendidikan, dan PDRB per kapita ke dalam model clustering berpotensi menghasilkan pengelompokan yang lebih kaya secara interpretatif dan lebih representatif terhadap kondisi nyata di lapangan (Suryana et al., 2023). Ketiga, penggunaan algoritma clustering alternatif seperti DBSCAN atau Fuzzy C-Means sebagai pembanding dapat memperkuat validitas hasil, terutama untuk menangani provinsi-provinsi yang secara karakteristik berada di batas antara dua klaster (Ikotun et al., 2023). Keempat, disarankan untuk mengombinasikan Silhouette Score dengan Davies-Bouldin Index dalam proses evaluasi agar kualitas klaster dapat dinilai dari dua perspektif berbeda secara bersamaan (Wulandari & Kusuma, 2024). Kelima, penelitian lanjutan dengan granularitas data pada tingkat kabupaten dan kota akan memberikan gambaran yang jauh lebih detail dan dapat langsung dimanfaatkan oleh pemerintah daerah dalam menyusun program digitalisasi yang tepat sasaran (Prasetyo et al., 2022).

DAFTAR PUSTAKA

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027–1035.

Azzahro, Z., Chyntia, L. F. A. R., Nasar, M., & Widianto, W. (2024). Implementation of the k-means clustering method for grouping poverty data: Case studies in Indonesia. AIP Conference Proceedings, 2927(1), 040026. https://doi.org/10.1063/5.0193516

Badan Pusat Statistik. (2023). Statistik Telekomunikasi Indonesia 2022. BPS RI. https://www.bps.go.id

Cheng, Y., Lin, B., & Zhang, W. (2022). Clustering developing countries by internet use profiles: A K-Means approach. Telecommunications Policy, 46(5), 102314. https://doi.org/10.1016/j.telpol.2022.102314

Gunawan, R., Saputra, D., & Rahmawati, F. (2023). K-Means clustering untuk pengelompokan kabupaten dan kota di Sumatera berdasarkan indeks ketimpangan pembangunan manusia. Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, 9(2), 201–211.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Hidayat, R., & Permana, A. (2022). Klasterisasi tingkat penetrasi e-commerce di Indonesia menggunakan K-Means Clustering. Jurnal Sistem Informasi, 18(2), 112–121. https://doi.org/10.21609/jsi.v18i2.1142

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in

the era of big data. Information Sciences, 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Jain, A. K. (2010). Data clustering: 50 years beyond K-Means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kementerian Komunikasi dan Informatika. (2023). Indeks Literasi Digital Nasional 2023. Kominfo RI. https://www.kominfo.go.id

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. Strategic Management Journal, 17(6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297.

Nugroho, A. S., & Sari, D. P. (2023). Segmentasi pengguna layanan digital di kota besar Indonesia menggunakan K-Means Clustering. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(3), 521–530. https://doi.org/10.25126/jtiik.2023103521

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Prasetyo, E., Hidayat, R., & Maulana, F. (2022). Pengelompokan wilayah 3T berdasarkan tingkat adopsi teknologi informasi menggunakan K-Means. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6(4), 612–620. https://doi.org/10.29207/resti.v6i4.3981

Rohman, F., & Wibowo, A. (2024). Penerapan metode clustering dengan algoritma K-Means pada sistem pendeteksi pencucian uang perbankan berbasis web. SENAFTI, 3(2), 398–406.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Satu Data Indonesia. (2023). Data infrastruktur telekomunikasi per provinsi. https://data.go.id

Suryana, T., Rachman, A., & Fauzi, M. (2023). Kesenjangan digital dan pertumbuhan ekonomi daerah di Indonesia: Analisis panel data 2017–2022. Jurnal Ekonomi dan Pembangunan Indonesia, 23(2), 145–162. https://doi.org/10.21002/jepi.2023.09

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. IOP Conference Series: Materials Science and Engineering, 336, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to Data Mining (2nd ed.). Pearson.

Wulandari, R., & Kusuma, A. (2024). Evaluasi kualitas klaster K-Means menggunakan Silhouette Score dan Davies-Bouldin Index pada data spasial. Journal of Applied Informatics and Computing, 8(1), 78–87. https://doi.org/10.30871/jaic.v8i1.5521