

Analisis Longitudinal dan Klasterisasi K-Means Disparitas Gender dalam Akses Internet di Indonesia

Dimas Dwi Randa¹, Revi Gusriwa²

¹Sains Data, Fakultas Sains, Insitut Teknologi Sumatera

²Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

¹dimas.randa@sd.itera.ac.id, ²revirvg31@gmail.com

Abstract

The rapid development of information technology has made the internet an essential infrastructure supporting social, economic, and educational activities. However, disparities in internet access between males and females remain evident across several regions of Indonesia. This study aims to analyze gender-based disparities in internet access at the provincial level and to classify provinces based on their access characteristics. The data used consist of 38 provinces in Indonesia, represented by the percentage of internet users by gender in 2023. The K-Means clustering algorithm is employed to group provinces according to similarities in internet access levels between males and females. The optimal number of clusters is determined using the Elbow Method, resulting in three clusters ($k = 3$), while clustering performance is evaluated using the Silhouette Score with a value of 0.59. The results show that Indonesian provinces can be grouped into several clusters with distinct characteristics of internet access. Provinces with higher internet penetration tend to exhibit smaller gender gaps, whereas provinces with lower access levels show relatively larger disparities. These findings provide insights into gender-based digital inequality in Indonesia and support policies promoting equitable internet access across regions.

Keywords: gender disparity, internet access, k-means clustering, silhouette score, indonesia

Abstrak

Perkembangan teknologi informasi telah menjadikan internet sebagai sarana penting dalam mendukung aktivitas sosial, ekonomi, dan pendidikan masyarakat. Namun, kesenjangan akses internet antara laki-laki dan perempuan masih menjadi permasalahan di beberapa wilayah Indonesia. Penelitian ini bertujuan untuk menganalisis disparitas gender dalam akses internet pada tingkat provinsi serta mengelompokkan provinsi berdasarkan karakteristik akses tersebut. Data yang digunakan mencakup 38 provinsi di Indonesia berupa persentase penduduk yang mengakses internet menurut jenis kelamin pada tahun 2023. Metode yang digunakan adalah algoritma K-Means untuk mengelompokkan provinsi berdasarkan tingkat akses internet laki-laki dan perempuan. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dan menghasilkan tiga cluster ($k = 3$), sedangkan kualitas hasil clustering dievaluasi menggunakan Silhouette Score sebesar 0.59. Hasil penelitian menunjukkan bahwa provinsi di Indonesia dapat dikelompokkan ke dalam beberapa cluster dengan karakteristik tingkat akses internet yang berbeda. Provinsi dengan tingkat penetrasi internet yang lebih tinggi cenderung memiliki kesenjangan gender yang lebih kecil, sedangkan provinsi dengan tingkat akses yang lebih rendah menunjukkan kesenjangan yang lebih besar. Temuan ini memberikan gambaran mengenai pola kesenjangan digital berbasis gender di Indonesia serta dapat menjadi dasar dalam perumusan kebijakan untuk pemerataan akses internet.

Kata kunci: disparitas gender, akses internet, k-means clustering, silhouette score, indonesia

© 2026 Author

Creative Commons Attribution 4.0 International License



1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi dalam beberapa dekade terakhir telah membawa perubahan signifikan dalam berbagai aspek kehidupan masyarakat. Internet menjadi salah satu infrastruktur digital utama yang mendukung aktivitas sosial, ekonomi, pendidikan, dan pemerintahan. Akses internet memungkinkan masyarakat memperoleh informasi secara cepat, meningkatkan produktivitas, serta memperluas peluang ekonomi dan inovasi. Oleh karena itu, pemerataan akses internet menjadi indikator penting dalam pembangunan berbasis teknologi dan transformasi digital suatu negara [1].

Di Indonesia, penggunaan internet terus meningkat seiring dengan berkembangnya infrastruktur telekomunikasi dan meningkatnya kebutuhan masyarakat terhadap layanan digital. Namun, peningkatan tersebut belum sepenuhnya diikuti oleh pemerataan akses di seluruh wilayah maupun pada seluruh kelompok masyarakat. Salah satu bentuk ketimpangan yang masih menjadi perhatian adalah kesenjangan akses internet berdasarkan gender. Perbedaan akses antara laki-laki dan perempuan dipengaruhi oleh berbagai faktor, seperti kondisi sosial ekonomi, tingkat pendidikan, budaya, serta ketersediaan infrastruktur teknologi [2].

Beberapa penelitian sebelumnya menunjukkan bahwa kesenjangan digital tidak hanya terjadi antarwilayah, tetapi juga antar kelompok sosial, termasuk gender [3]. Di Indonesia, studi terkait kesenjangan digital dan akses internet telah banyak dilakukan dengan berbagai pendekatan analisis data, termasuk metode clustering untuk mengidentifikasi pola antarwilayah [6]. Oleh karena itu, diperlukan pendekatan berbasis data mining untuk mengidentifikasi pola dan struktur data secara lebih komprehensif. Salah satu metode yang dapat digunakan adalah clustering, khususnya algoritma K-Means, yang mampu mengelompokkan data berdasarkan tingkat kemiripan karakteristik [4], [5].

Dalam konteks penelitian ini, K-Means digunakan untuk mengelompokkan provinsi berdasarkan tingkat akses internet laki-laki dan perempuan sehingga dapat diketahui karakteristik kesenjangan gender di setiap kelompok wilayah. Selain itu, metode Elbow digunakan untuk menentukan jumlah cluster optimal, sedangkan Silhouette Score digunakan untuk mengevaluasi kualitas hasil clustering [5].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis disparitas gender dalam akses internet pada tingkat provinsi di Indonesia serta mengelompokkan provinsi berdasarkan karakteristik akses internet menggunakan metode K-Means clustering. Penelitian ini diharapkan dapat memberikan gambaran pola kesenjangan digital berbasis gender serta menjadi dasar dalam

perumusan kebijakan untuk pemerataan akses internet di Indonesia.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder yang bersumber dari Badan Pusat Statistik (BPS). Data yang digunakan berupa persentase penduduk yang mengakses internet menurut jenis kelamin pada tingkat provinsi di Indonesia tahun 2023. Variabel yang dianalisis meliputi persentase pengguna internet laki-laki dan perempuan. Data mencakup seluruh provinsi di Indonesia sehingga mampu merepresentasikan kondisi akses internet secara nasional. Proses penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data, praproses data, analisis eksploratif, proses clustering menggunakan K-Means, serta evaluasi hasil clustering.

2.1 Praproses dan Normalisasi Data

Tahap praproses dilakukan untuk memastikan kualitas data sebelum dianalisis, meliputi pengecekan format data, konversi ke bentuk numerik, serta penanganan nilai yang tidak sesuai. Selanjutnya dilakukan normalisasi data menggunakan StandardScaler untuk menyamakan skala antar variabel sehingga tidak terjadi dominasi fitur tertentu dalam proses clustering. Normalisasi ini penting karena algoritma K-Means sensitif terhadap skala data.

2.2 Analisis Eksploratif Data

Analisis eksploratif dilakukan menggunakan statistik deskriptif dan visualisasi untuk melihat distribusi akses internet berdasarkan gender pada setiap provinsi. Tahap ini bertujuan untuk memahami pola awal data sebelum dilakukan clustering.

2.3 Metode K-Means Clustering

Metode K-Means digunakan untuk mengelompokkan provinsi berdasarkan tingkat kemiripan akses internet laki-laki dan perempuan. Proses clustering dilakukan dengan menentukan jumlah cluster (k), menginisialisasi centroid, menghitung jarak antar data, mengelompokkan data ke cluster terdekat, serta memperbarui centroid hingga konvergen [4]. Dalam penelitian ini, algoritma K-Means dijalankan menggunakan parameter random state = 42 untuk memastikan hasil clustering bersifat konsisten dan dapat direproduksi. Perhitungan jarak menggunakan Euclidean Distance.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

2.4 Penentuan Jumlah Cluster

Jumlah cluster optimal ditentukan menggunakan metode Elbow dengan menghitung nilai Within Cluster Sum of Squares (WCSS).

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} (x_i - c_j)^2 \quad (2)$$

2.5 Evaluasi Clustering

Evaluasi hasil clustering dilakukan menggunakan Silhouette Score

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (3)$$

2.6 Analisis Gender Gap

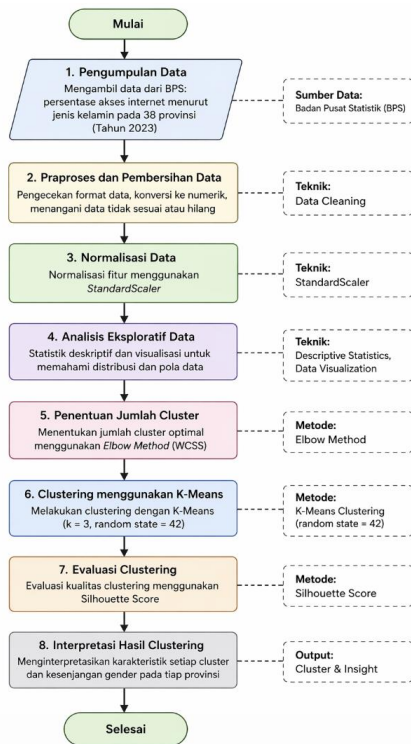
Kesenjangan gender dihitung sebagai selisih antara persentase akses internet laki-laki dan perempuan

2.7 Analisis Longitudinal (Time Series)

Analisis longitudinal digunakan untuk mengamati tren perkembangan akses internet berdasarkan gender dari waktu ke waktu. Data time series dianalisis untuk mengidentifikasi pola perubahan akses internet serta dinamika kesenjangan gender.

2.8 Workflow Penelitian

Alur penelitian dalam studi ini meliputi beberapa tahapan utama, yaitu pengumpulan data, praproses dan normalisasi data, analisis eksploratif, penentuan jumlah cluster, proses clustering menggunakan K-Means, evaluasi hasil clustering, serta interpretasi hasil. Alur tersebut dapat dilihat pada Gambar 1.



Gambar 1. Workflow penelitian

2.9 Perangkat Analisis

Pengolahan data dilakukan menggunakan Python dengan pustaka Pandas, NumPy, Scikit-learn, dan Matplotlib untuk analisis dan visualisasi data.

3. Hasil dan Pembahasan

3.1. Analisis Tren Internet Penduduk Usia 25+ Tahun

Analisis longitudinal dilakukan untuk mengamati perkembangan akses internet penduduk usia 25 tahun ke atas di Indonesia berdasarkan jenis kelamin. Hasil menunjukkan bahwa penggunaan internet mengalami peningkatan dalam beberapa tahun terakhir. Meskipun akses laki-laki masih lebih tinggi dibandingkan perempuan, kesenjangan gender cenderung menurun dari waktu ke waktu. Hal ini mengindikasikan peningkatan partisipasi perempuan dalam pemanfaatan teknologi digital. Selain itu, peningkatan akses internet pada kelompok usia dewasa menunjukkan bahwa teknologi digital semakin digunakan dalam berbagai aktivitas, seperti pekerjaan, komunikasi, dan akses informasi.

3.2. Analisis Longitudinal (Time Series)

Berdasarkan analisis time series menganalisis tren kesenjangan gender akses internet 2014–2023 dapat dilihat pada tabel 1.

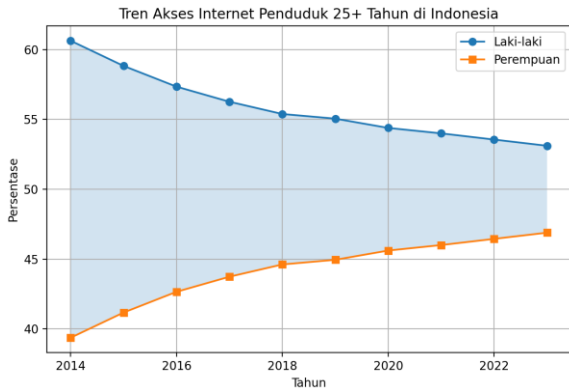
Tabel 1. Analisis tren akses internet

Tahun	Laki-laki	Perempuan	GAP
2014	60.633333	39.366667	21.266667
2015	58.831471	41.168529	17.662941
....
2020	54.390588	46.001176	8.781176
2021	53.998824	46.001176	7.997647
2022	53.556176	46.443824	7.112353
2023	53.111471	46.888529	6.222941

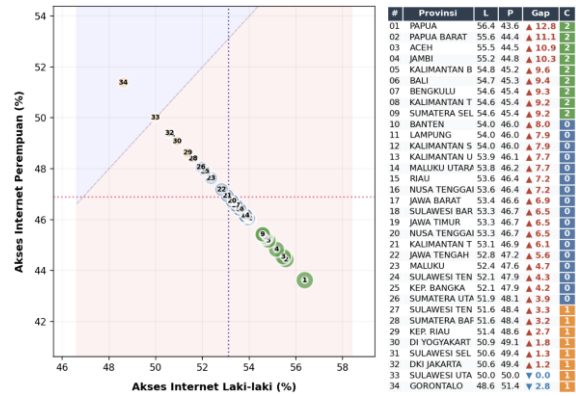
Hasil analisis tren menunjukkan bahwa akses internet laki-laki mengalami sedikit penurunan, sementara akses internet perempuan cenderung meningkat. Kondisi ini berdampak pada semakin kecilnya kesenjangan gender dalam akses internet. Secara keseluruhan, terjadi penurunan kesenjangan gender di Indonesia selama periode 2014–2023 yang mengindikasikan adanya konvergensi digital antara laki-laki dan perempuan.

3.3. Grafik Tren nasional

Analisis tren nasional dilakukan untuk melihat perkembangan akses internet berdasarkan gender di Indonesia selama periode 2014–2023. Hasil analisis menunjukkan bahwa terjadi penurunan kesenjangan akses internet antara laki-laki dan perempuan. Penurunan ini dipengaruhi oleh peningkatan akses internet pada perempuan yang berlangsung secara konsisten setiap tahun, sementara akses pada laki-laki cenderung mengalami penurunan secara bertahap. Grafiknya dapat dilihat pada gambar 2.



Gambar 2. Grafik Tren Nasional



Gambar 3. Visualisasi Clustering Provinsi

3.4. Analisis Clustering (K-Means)

Mengelompokkan provinsi berdasarkan disparitas gender akses internet, seperti pada tabel 2.

Tabel 2. Pengelompokkan Provinsi

Provinsi	L2023	P2023	gap_2023	cluster
Aceh	55.2	44.5	10.7	1
Sumatera Utara	57.1	48.0	9.1	1
Sumatera Barat	60.3	48.4	11.9	2
DKI Jakarta	78.2	73.4	4.8	0
Bali	74.1	70.5	3.6	0

Interpretasi Cluster

- a. Cluster 0 Provinsi dengan akses internet tinggi dan gender gap kecil contohnya: DKI Jakarta dan Bali
- b. Cluster 1 Provinsi dengan akses sedang dan gap sedang contohnya: Sumatera Utara dan Bali
- c. Cluster 2 Provinsi dengan akses rendah dan gap besar contohnya: Papua dan NTT

Hasil klasterisasi menunjukkan adanya perbedaan karakteristik antar provinsi dalam akses internet dan kesenjangan gender.

3.5. Visualisasi Clustering setiap provinsi

Visualisasi clustering dilakukan untuk menggambarkan distribusi provinsi berdasarkan tingkat akses internet laki-laki dan perempuan serta kesenjangan gender yang terjadi. Setiap provinsi direpresentasikan dalam bentuk titik pada bidang dua dimensi, dengan sumbu horizontal menunjukkan persentase akses internet laki-laki dan sumbu vertikal menunjukkan persentase akses internet perempuan. Pengelompokan ini bertujuan untuk memperlihatkan pola kedekatan karakteristik antarprovinsi serta perbedaan antar cluster yang terbentuk. Visualisasi Clustering dapat dilihat pada gambar 3.

3.6. Analisis Clustering Detail

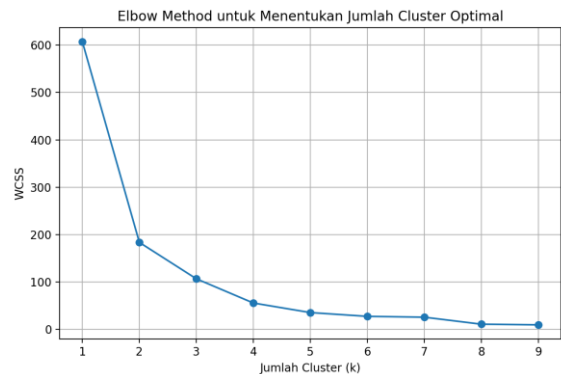
Analisis clustering detail dilakukan dengan menghitung nilai rata-rata akses internet laki-laki, perempuan, dan kesenjangan gender pada setiap cluster. Hasil perhitungan menunjukkan bahwa setiap cluster memiliki karakteristik yang berbeda. Cluster 0 memiliki rata-rata akses internet laki-laki sebesar 53,21% dan perempuan 46,79% dengan kesenjangan gender sebesar 6,41%. Cluster 1 menunjukkan tingkat akses yang relatif seimbang, dengan rata-rata laki-laki sebesar 50,67% dan perempuan 49,33%, serta kesenjangan gender yang paling kecil yaitu 1,35%. Sementara itu, cluster 2 memiliki rata-rata akses laki-laki sebesar 55,10% dan perempuan 44,90% dengan kesenjangan gender terbesar yaitu 10,20%. Hasil Analisa clustering dapat dilihat pada tabel 3.

Tabel 3. Analisis Clustering Detail

cluster	L2023	P2023	Gap 2023
0	53.205882	46.794118	6.411765
1	50.672500	49.327500	1.345000
2	55.101111	44.898889	10.202222

3.7. Penentuan Jumlah Cluster (Elbow Method)

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan menganalisis nilai Within Cluster Sum of Squares (WCSS) pada beberapa jumlah cluster. Metode ini digunakan untuk mengidentifikasi jumlah cluster yang paling sesuai berdasarkan tingkat variasi dalam setiap cluster. Gambar 4 menunjukkan Elbow Method.

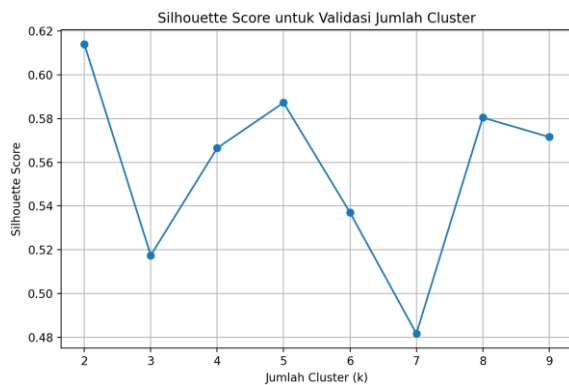


Gambar 4. Elbow Method

Berdasarkan grafik, nilai WCSS mengalami penurunan yang sangat signifikan dari k=1 ke k=2, kemudian masih menurun hingga k=3. Setelah itu, penurunannya cenderung melandai dan tidak terlalu signifikan. Pola tersebut menunjukkan terbentuknya titik siku (elbow) pada k=3, sehingga jumlah cluster optimal dalam penelitian ini ditentukan sebanyak tiga cluster.

3.8. Silhouette Score

Silhouette Score digunakan untuk mengevaluasi kualitas hasil clustering dengan mengukur tingkat kedekatan data dalam satu cluster dibandingkan dengan cluster lainnya. Semakin tinggi nilai Silhouette Score, maka semakin baik kualitas pengelompokan yang dihasilkan.

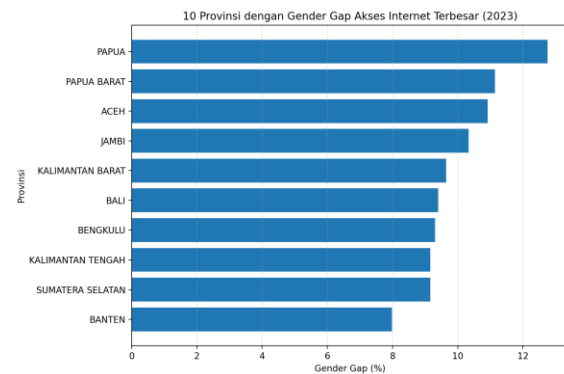


Gambar 5. Silhouette Score

Grafik pada gambar 5 menunjukkan bahwa nilai Silhouette Score tertinggi diperoleh pada k = 2, namun nilai pada k = 3 masih cukup tinggi dan stabil. Dengan mempertimbangkan hasil Elbow Method, jumlah cluster optimal yang digunakan adalah tiga.

3.9. Ranking Provinsi Berdasarkan Gender Gap

Hasil ranking menunjukkan bahwa Papua, Papua Barat, dan Aceh merupakan provinsi dengan kesenjangan gender akses internet tertinggi. Hal ini menunjukkan bahwa disparitas akses internet antar gender masih cukup signifikan di beberapa wilayah Indonesia, seperti yang terlihat pada gambar 6.

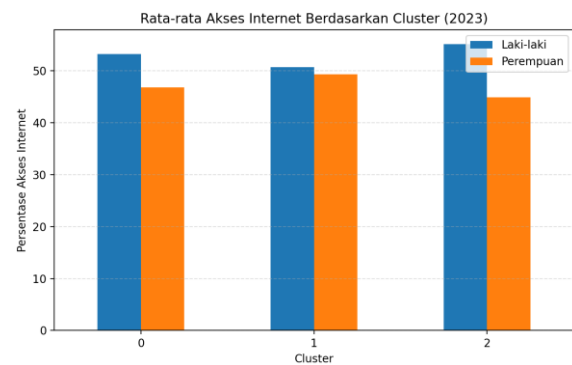


Gambar 6. Ranking Provinsi Gender Gap

3.10. Rata-rata akses internet cluster 2023

Analisis rata-rata akses internet berdasarkan cluster menunjukkan adanya perbedaan karakteristik antar

kelompok provinsi. Pada cluster 0, rata-rata akses internet laki-laki lebih tinggi dibandingkan perempuan, dengan selisih yang cukup terlihat. Cluster 1 menunjukkan kondisi yang lebih seimbang, di mana rata-rata akses internet laki-laki dan perempuan hampir sama, sehingga kesenjangan gender pada cluster ini relatif kecil. Sementara itu, cluster 2 memiliki rata-rata akses internet laki-laki yang paling tinggi, namun diikuti oleh rata-rata akses perempuan yang lebih rendah dibanding cluster lainnya. Kondisi ini menyebabkan kesenjangan gender pada cluster 2 menjadi yang paling besar. Hasil ini menunjukkan bahwa tingkat akses internet yang tinggi tidak selalu diikuti oleh kesetaraan gender, sehingga perbedaan akses antara laki-laki dan perempuan masih menjadi isu pada beberapa kelompok wilayah. Gambar 7 menunjukkan grafik rata-rata internet berdasarkan cluster.



Gambar 7. Rata-rata internet berdasarkan cluster

4. Kesimpulan

Penelitian ini berhasil menganalisis disparitas gender dalam akses internet di Indonesia serta mengelompokkan provinsi berdasarkan karakteristik akses internet menggunakan metode K-Means clustering. Hasil analisis menunjukkan bahwa terjadi penurunan kesenjangan gender dalam akses internet selama periode pengamatan, yang mengindikasikan adanya konvergensi digital antara laki-laki dan perempuan. Provinsi dengan tingkat akses internet yang tinggi cenderung memiliki kesenjangan gender yang lebih kecil, sedangkan provinsi dengan tingkat akses yang rendah menunjukkan kesenjangan yang lebih besar.

Temuan ini memiliki implikasi penting dalam perumusan kebijakan pemerataan akses internet, khususnya dalam upaya mengurangi kesenjangan digital berbasis gender di berbagai wilayah. Namun demikian, penelitian ini memiliki beberapa keterbatasan. Data yang digunakan hanya terbatas pada satu tahun pengamatan (2023) untuk analisis clustering, sehingga belum sepenuhnya menggambarkan dinamika perubahan antar waktu secara mendalam. Selain itu, variabel yang digunakan masih terbatas pada persentase akses internet berdasarkan gender tanpa

mempertimbangkan faktor lain seperti kondisi sosial ekonomi, tingkat pendidikan, dan ketersediaan infrastruktur teknologi yang juga berpengaruh terhadap kesenjangan digital. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan data dengan rentang waktu yang lebih panjang serta menambahkan variabel-variabel pendukung lainnya agar menghasilkan analisis yang lebih komprehensif dan mendalam.

Daftar Rujukan

- [1] Badan Pusat Statistik, Statistik Telekomunikasi Indonesia 2023, Jakarta: BPS, 2023.
- [2] Joshi, A., Malhotra, B., Amadi, C., Loomba, M., Misra, A., Sharma, S., dan Arora, A. 2020. Gender and the digital divide across urban slums. *Journal of Medical Internet Research*. 22(6): 1-12.
- [3] Abid, A., Kachouri, A., dan Mahjoub, M. 2021. Clustering algorithms for big data: Taxonomy and empirical analysis. *Journal of Big Data*. 8(1): 1-24.
- [4] Rousseeuw, P.J. 2020. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 20: 53-65.
- [5] Arthur, D., dan Vassilvitskii, S. 2020. K-means++: The advantages of careful seeding. *SIAM Review*. 52(1): 1-20.
- [6] Wati, S., et al. 2024. Identification of digital divide across Indonesian provinces using clustering analysis. *Statistika Journal*. 12(2): 45-60.
- [7] Prasetyo, E., dan Sari, R. 2021. Penerapan algoritma K-Means untuk pengelompokan tingkat penggunaan internet di Indonesia. *Jurnal RESTI*. 5(3): 456-463.
- [8] Rahman, A., dan Fitriani, D. 2022. Analisis clustering K-Means pada data pengguna internet di Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*. 9(2): 245-252.
- [9] Nugroho, Y., dan Putri, A. 2020. Pemetaan kesenjangan digital di Indonesia menggunakan metode K-Means clustering. *Jurnal Ilmu Komputer dan Informasi*. 13(1): 1-10.
- [10] Saputra, R., dan Widodo, A. 2021. Implementasi K-Means clustering untuk analisis disparitas akses teknologi informasi di Indonesia. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*. 10(4): 321-328.
- [11] Sari, M., dan Putra, D. 2023. Analisis disparitas akses teknologi informasi berbasis gender menggunakan data mining. *Jurnal Sistem Informasi dan Teknologi Informasi*. 11(2): 101-111