

Analisis Klasterisasi Kualitas Internet Seluler Menggunakan Metode K-Means dan Gaussian Mixture Model

Muhammad Aziiz Irwansyah, Allsela Meiriza*, Dinda Lestarini

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia

Email: ¹09031282227095@student.unsri.ac.id, ^{2,*}allsela@unsri.ac.id, ³dinda@unsri.ac.id

Email Penulis Korespondensi: allsela@unsri.ac.id

Submitted: 29/10/2025; Accepted: 09/12/2025; Published: 11/12/2025

Abstrak—Penelitian ini menggunakan data jaringan internet dari Ookla Open Data (Speedtest Global Performance) yang mencakup tiga variabel utama, yaitu kecepatan unduh (*download speed*), kecepatan unggah (*upload speed*), dan latensi (*latency*). Tujuan penelitian ini adalah menganalisis kondisi dan performa jaringan internet seluler di 17 kabupaten/kota Provinsi Sumatera Selatan tahun 2025 serta memberikan rekomendasi berbasis data bagi Dinas Komunikasi dan Informatika dalam upaya pemerataan dan peningkatan infrastruktur digital melalui pendekatan *Knowledge Discovery in Databases* (KDD). Metode yang diterapkan meliputi RobustScaler untuk normalisasi data, *Principal Component Analysis* (PCA) untuk reduksi dimensi, serta algoritma *K-Means* dan *Gaussian Mixture Model* (GMM) untuk klasterisasi wilayah berdasarkan kesamaan karakteristik jaringan. Hasil analisis menunjukkan bahwa kedua algoritma membentuk tiga klaster ($K=3$) dengan pola berbeda. GMM memiliki kinerja lebih stabil dibandingkan *K-Means*, dengan nilai *Silhouette* 0,426 dan *Davies–Bouldin Index* 0,284, sedangkan *K-Means* memperoleh 0,351 dan 0,688, sementara nilai *Calinski–Harabasz* GMM (9,960) yang lebih rendah dari *K-Means* (13,754) mencerminkan *trade-off* antara kepadatan klaster dan stabilitas model yang menegaskan karakter adaptif GMM terhadap variasi data. Wilayah perkotaan seperti Palembang dan Prabumulih menempati klaster berperforma tinggi, sedangkan Ogan Komering Ulu Selatan berada pada klaster berperforma rendah (18,87 Mbps; 33 ms), menunjukkan kesenjangan digital sekitar 18 Mbps antarwilayah. Temuan ini menegaskan pentingnya strategi pemerataan infrastruktur digital seperti melalui perluasan *fiber optic*, peningkatan kapasitas BTS, dan kolaborasi lintas pemangku kepentingan menuju Visi Indonesia Digital 2045.

Kata Kunci: Klasterisasi Kualitas Internet Seluler; K-Means; Gaussian Mixture Model; Ookla Open Data; Principal Component Analysis (PCA)

Abstract—This study utilizes internet network data from Ookla Open Data (Speedtest Global Performance), comprising three main variables: download speed, upload speed, and latency. The aim is to analyze the condition and performance of mobile internet networks across 17 regencies/cities in South Sumatera Province in 2025 and to provide data-driven recommendations for the Department of Communication and Informatics to promote equitable and improved digital infrastructure through a Knowledge Discovery in Databases (KDD) approach. The applied methods include RobustScaler for data normalization, Principal Component Analysis (PCA) for dimensionality reduction, and K-Means and Gaussian Mixture Model (GMM) algorithms for clustering regions based on network characteristics. The analysis shows that both algorithms form three clusters ($K=3$) with distinct patterns. GMM demonstrates higher stability than K-Means, achieving a Silhouette score of 0.426 and Davies–Bouldin Index of 0.284, compared to K-Means with 0.351 and 0.688, while the lower Calinski–Harabasz score of GMM (9.960) indicates a trade-off between cluster compactness and stability, highlighting its adaptive behavior to data variation. Urban areas such as Palembang and Prabumulih belong to the high-performance cluster, whereas Ogan Komering Ulu Selatan lies in the low-performance cluster (18.87 Mbps; 33 ms), revealing a digital gap of approximately 18 Mbps across regions. These findings emphasize the need for equitable digital infrastructure strategies through fiber-optic expansion, BTS capacity enhancement, and multi-stakeholder collaboration toward Indonesia’s Digital Vision 2045.

Keywords: Clustering of Mobile Internet Quality; K-Means; Gaussian Mixture Model; Ookla Open Data; Principal Component Analysis (PCA)

1. PENDAHULUAN

Akses internet yang merata dan berkualitas merupakan indikator penting dalam mewujudkan keberhasilan transformasi digital di Indonesia [1]. Pemerintah dengan Kementerian Komunikasi dan Informatika (Kemenkominfo) menegaskan bahwa distribusi infrastruktur digital dan perbaikan mutu jaringan adalah fokus utama dalam mendukung Visi Indonesia Digital 2045 sebagai langkah menuju Indonesia Emas 2045 [2]. Namun, kesenjangan digital antarwilayah masih menjadi tantangan utama, terutama antara kawasan perkotaan dan pedesaan di luar Pulau Jawa seperti Provinsi Sumatera Selatan yang masih mengalami ketimpangan kecepatan dan stabilitas jaringan [3]. Kondisi ini mengindikasikan perlunya analisis berbasis data yang mampu memetakan pemerataan infrastruktur dan kualitas jaringan secara objektif guna mendukung kebijakan transformasi digital yang inklusif.

Survei terbaru yang dilaksanakan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII, 2024) mengungkapkan bahwa jumlah individu yang menggunakan internet di tingkat nasional mencapai 221,56 juta orang atau sekitar 79,5 % dari keseluruhan populasi. Meskipun tingkat penetrasi internet meningkat pesat, perbedaan antara wilayah urban (69,5 %) serta rural (30,5 %) menandakan masih adanya ketimpangan signifikan [4]. Penelitian lain melaporkan bahwa kecepatan internet rata-rata di wilayah Sumatera hanya berkisar 20–30 Mbps, sedangkan di Pulau Jawa dapat melampaui 40 Mbps. Perbedaan ini disebabkan oleh persebaran infrastruktur telekomunikasi yang belum merata dan kepadatan pengguna yang tinggi [5]. Oleh karena itu, diperlukan upaya pemetaan kualitas jaringan menggunakan sumber data terbuka agar pemerintah daerah dapat memperoleh gambaran akurat terhadap kondisi digital wilayahnya.

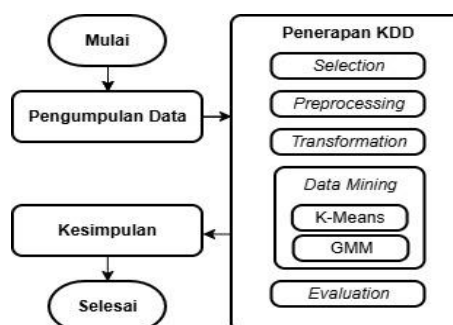
Salah satu sumber data yang dapat dimanfaatkan adalah Ookla Open Data (Speedtest Global Performance), yang menyediakan hasil pengujian kecepatan unduh (*download*), unggah (*upload*), dan latensi (*latency*) dari pengguna internet di seluruh dunia termasuk Indonesia [6]. Melalui metode Penemuan Pengetahuan dalam Basis Data (KDD), informasi itu bisa diproses dengan cara terstruktur melalui langkah-langkah pemilihan, pra-proses, perubahan, penambahan data, dan penilaian hasil [7]. Pada tahap data mining, penelitian ini menerapkan kombinasi dua metode klasterisasi yaitu *K-Means Clustering* dan *Gaussian Mixture Model (GMM)* untuk mengidentifikasi pola distribusi kualitas jaringan seluler. Wijayanti dan Kesuma (2025) menggunakan metode *K-Means* karena terbukti efisien untuk data numerik berukuran menengah hingga besar serta menghasilkan kluster yang mudah diinterpretasikan [8]. Efektivitas *K-Means* juga diperkuat oleh penelitian Fauzan, Suarna, dkk. (2025) yang menganalisis pola kecepatan dan latensi jaringan dengan nilai *Davies-Bouldin* sebesar 0,006 [9]. Sementara itu, Chyan (2022) menggunakan GMM untuk memperkuat hasil klasterisasi melalui pendekatan probabilistik yang mampu menangani variasi bentuk dan kepadatan data [10]. Penelitian Setiawan dan Kurniawan (2023) juga membuktikan bahwa model GMM dengan nilai *Davies-Bouldin* sebesar 0,645 adalah yang paling baik dan adaptif dalam mendeteksi struktur kluster kompleks dengan distribusi yang tidak linear [11].

Berbagai studi terdahulu telah mengimplementasikan pendekatan serupa pada konteks jaringan dan transformasi digital. Penelitian Alde (2025) mengaplikasikan *K-Means* untuk mengidentifikasi wilayah blank spot dan non-blank spot berdasarkan sebaran *Base Transceiver Station (BTS)* di Kabupaten Padang Pariaman, Sumatera Barat dan hasilnya efektif untuk pemetaan spasial namun terbatas pada satu wilayah dan belum mempertimbangkan performa jaringan actual [12]. Apriliana dan Widodo (2023) mengelompokkan provinsi di Indonesia berdasarkan jumlah BTS dengan *Hierarchical Clustering* [13]. Sedangkan Rahma Watin dkk (2024) menganalisis keterampilan, akses, dan fasilitas TIK antarprovinsi menggunakan kombinasi *K-Means* dan *Hierarchical Clustering* dengan nilai *Silhouette* 0,31 dan 0,37, berfokus pada indikator sosial dan infrastruktur digital tanpa meninjau performa teknis jaringan [14]. Sementara itu, studi Alizadeh, Vranken, dkk. (2020) mengusulkan pendekatan berbasis *Gaussian Mixture Model (GMM)* untuk klasifikasi dan verifikasi lalu lintas jaringan, menunjukkan akurasi hingga 97,7% hanya dengan menggunakan 9 paket awal aliran dan tetap mencapai sekitar 96,6% dengan hanya 0,5% data pelatihan, namun belum mempertimbangkan konteks spasial pemerataan jaringan antarwilayah [15].

Dari berbagai tinjauan sebelumnya, belum ditemukan kajian yang secara spesifik menyoroti pemerataan kualitas jaringan internet di Indonesia dengan pendekatan berbasis performa teknis jaringan. Sebagian besar penelitian terdahulu masih berfokus pada level provinsi, belum menjangkau skala mikro-spasial seperti kabupaten atau kota, serta cenderung menggunakan indikator sosial maupun infrastruktur digital tanpa mempertimbangkan variabel kecepatan unduh, unggah, dan latensi yang mencerminkan kinerja jaringan aktual. Selain itu, sebagian besar studi menggunakan data sebelum tahun 2025, sehingga belum mampu menggambarkan kondisi terkini perkembangan infrastruktur digital nasional. Penelitian ini berupaya mengisi kesenjangan tersebut dengan menganalisis kondisi dan performa jaringan internet seluler pada 17 kabupaten/kota di Provinsi Sumatera Selatan menggunakan data terbuka Ookla tahun 2025. Melalui penerapan kombinasi metode *K-Means* dan *Gaussian Mixture Model (GMM)* dalam kerangka *Knowledge Discovery in Databases (KDD)*, penelitian ini tidak hanya membandingkan hasil klasterisasi dari dua algoritma tersebut, tetapi juga menilai keunggulan dan keterbatasan relatif masing-masing algoritma di mana *K-Means* menawarkan efisiensi dan interpretasi yang sederhana, sedangkan GMM lebih adaptif terhadap variasi bentuk serta kepadatan data. Penelitian ini mengelompokkan wilayah berdasarkan kesamaan karakteristik jaringan serta membandingkan efektivitas kedua algoritma dalam menghasilkan pola distribusi yang representatif terhadap kondisi digital daerah. Kualitas hasil klasterisasi dievaluasi menggunakan tiga metrik internal, yaitu *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index* [16]. Dengan demikian, analisis dan evaluasi kedua metode diharapkan dapat menghasilkan pemetaan yang lebih representatif dan mendukung penyusunan kebijakan peningkatan infrastruktur digital berbasis data oleh Dinas Komunikasi dan Informatika Provinsi Sumatera Selatan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan kerangka metodologi dengan tahapan utama yang berdasar pada pendekatan *Knowledge Discovery in Databases* seperti pada Gambar 1 sebagai dasar dari proses analisis data.



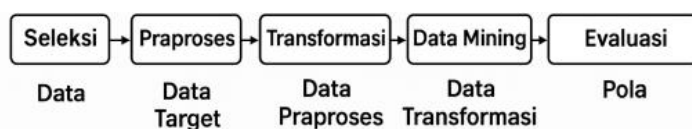
Gambar 1. Metodologi Penelitian

2.1 Tahap Pengumpulan Data

Dataset yang dipakai pada kajian ini adalah data sekunder yang didapatkan dari Ookla Open Data (Performa Global *Speedtest*) [6], yaitu data terbuka hasil pengujian kecepatan internet seluler yang dilakukan oleh pengguna *Speedtest* by Ookla di berbagai wilayah di dunia. Data tersebut diakses melalui repositori Amazon Web Services (AWS S3) dengan format *parquet* yang kemudian diolah menggunakan bahasa pemrograman Python dan pustaka *GeoPandas* untuk membatasi cakupan wilayah penelitian hanya pada Indonesia khususnya Provinsi Sumatera Selatan. Proses pengumpulan data ini menghasilkan kumpulan informasi performa jaringan yang siap digunakan sebagai dasar dalam tahapan analisis selanjutnya menggunakan pendekatan *Knowledge Discovery in Databases* (KDD).

2.2 Penerapan *Knowledge Discovery in Database*

Knowledge Discovery in Databases merupakan metode terstruktur untuk menggali wawasan baru dari data dalam kuantitas yang sangat besar. Tujuan dari tahap ini adalah mengidentifikasi pola atau informasi tersembunyi dalam data yang memiliki validitas, nilai guna, serta dapat diinterpretasikan, sehingga hasilnya dapat dimanfaatkan dalam proses pengambilan keputusan berbasis data [17].



Gambar 2. Tahap *Knowledge Discovery in Databases* (KDD)

Sebagaimana diperlihatkan pada Gambar 2, setiap tahapan KDD saling terintegrasi dan berperan dalam menghasilkan pengetahuan baru dari data mentah hingga tahap evaluasi hasil analisis. Dengan melalui tahap pemilihan, pra-pemrosesan, transformasi, penambahan data, dan evaluasi, KDD mampu membantu menemukan hubungan dan pola yang sebelumnya tidak terlihat.

2.2.1 Selection

Pada langkah pertama *selection*, data yang telah dikumpulkan dari Ookla Open Data diseleksi agar sesuai dengan fokus penelitian, yaitu menganalisis kondisi kualitas jaringan internet seluler di Provinsi Sumatera Selatan pada tahun 2025. Dari *dataset* global tersebut, penelitian ini hanya memilih data bertipe *mobile performance* dan membatasi periode analisis pada kuartal pertama hingga ketiga (Q1–Q3 2025) untuk memperoleh representasi kondisi jaringan terkini. Selanjutnya, data diproses secara spasial memakai *GeoPandas* sehingga hanya mencakup 17 kabupaten/kota yang ada di daerah Sumatera Selatan. Variabel yang dipilih mencakup kecepatan unduh, kecepatan unggah, serta latensi, yang akan menjadi dasar untuk aktivitas penambahan data pada tahap selanjutnya.

2.2.2 Preprocessing

Selanjutnya di tahap *preprocessing*, hasil seleksi dibersihkan dan disiapkan agar siap untuk lanjut dianalisis [18]. Proses ini meliputi penyaringan nilai null, standarisasi format label kuartal (Q1_2025, Q2_2025, Q3_2025), serta penghapusan duplikasi data. Setelah itu, dilakukan agregasi dengan menghitung nilai rata-rata dari setiap variabel utama, yaitu kecepatan unduh, kecepatan unggah, dan latensi berdasarkan masing-masing kabupaten/kota. Pemilihan nilai rata-rata digunakan karena dapat merepresentasikan performa umum jaringan secara keseluruhan di setiap wilayah, sementara potensi pengaruh ekstrem dari outlier dapat diminimalkan melalui proses normalisasi menggunakan *RobustScaler*, sehingga hasil agregasi tetap stabil dan proporsional [19]. Tujuannya adalah untuk memperoleh nilai representatif performa jaringan di tiap wilayah tanpa dipengaruhi oleh variasi ekstrem pada hasil pengujian individu. Tahapan ini juga memastikan konsistensi skala antar-atribut sebelum dilakukan transformasi dan analisis lebih lanjut.

2.2.3 Transformation

Tahap transformasi bertujuan untuk mengonversi data hasil preprocessing menjadi format numerik yang konsisten, terstandarisasi, dan efisien untuk diproses pada algoritma klusterisasi. Pada penelitian ini digunakan dua teknik utama, yaitu *RobustScaler* untuk normalisasi data dan *Principal Component Analysis* (PCA) untuk reduksi dimensi menjadi dua komponen utama (PC1 dan PC2).

Pada proses ini, *RobustScaler* digunakan untuk menormalisasikan skala variabel dengan mengacu pada parameter median dan *interquartile range* (IQR) sehingga mampu mengurangi pengaruh outlier pada distribusi data [20]. Sedangkan pada PCA digunakan parameter jumlah komponen ($n_components = 2$) untuk mereduksi data ke dalam dua dimensi utama. Penggunaan PCA didasarkan pada literatur terbaru yang menunjukkan bahwa reduksi dimensi mampu meningkatkan stabilitas, efisiensi, dan kualitas hasil klusterisasi dengan menghilangkan korelasi antarvariabel serta mengurangi redundansi data [21], skor positif pada PC1 atau PC2 menunjukkan karakteristik wilayah dengan performa di atas rata-rata keseluruhan, sementara skor negatif menandakan kontribusi di bawah rata-rata [22]. Dengan demikian, hasil transformasi tidak hanya mempercepat proses komputasi, tetapi juga memungkinkan visualisasi spasial yang lebih jelas terhadap pola persebaran kualitas jaringan antarwilayah.

2.2.4 Data Mining

Tahap penambangan data adalah komponen fundamental dari proses Penemuan Pengetahuan dalam Basis Data (KDD) yang bertujuan untuk mengidentifikasi pola yang tersembunyi serta relasi di antara variabel data [17]. Pada fase ini, diterapkan dua metode klusterisasi, yaitu *K-Means Clustering* dan *Gaussian Mixture Model (GMM)*. Keduanya dioperasikan menggunakan bahasa pemrograman Python di platform Google Colab.

Tabel 1. Parameter Proses Tahap Data Mining

Parameter	Nilai
Algoritma	<i>K-Means, Gaussian Mixture Model</i>
Kluster	2, 3, 4, 5
Inisialisasi (n_init)	100
Metrik Evaluasi	<i>Silhouette Score, Davies–Bouldin Index, Calinski–Harabasz Index</i>

Proses implementasi serta parameter yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1, di mana setiap algoritma diuji dengan jumlah kluster $K = 2, 3, 4,$ dan 5 secara iteratif untuk menemukan konfigurasi paling sesuai dengan karakteristik data. Nilai parameter $n_init = 100$ digunakan untuk memastikan stabilitas hasil, sedangkan tiga metrik evaluasi yaitu *Silhouette Score, Davies–Bouldin Index,* dan *Calinski–Harabasz Index* untuk menilai kualitas hasil klusterisasi yang dihasilkan oleh kedua algoritma, yaitu *K-Means* dan *Gaussian Mixture Model (GMM)*.

2.2.4.1 K-Means

Algoritma *K-Means Clustering* beroperasi dengan mengelompokkan data ke dalam sejumlah kluster yang ditentukan oleh jarak paling dekat antara data dan titik pusat kluster (*centroid*). Proses ini dimulai dengan menentukan jumlah kluster, lalu setiap data dialokasikan ke kluster dengan jarak minimum menggunakan *Euclidean Distance*. Selanjutnya, posisi *centroid* diperbarui secara bertahap melalui proses iteratif hingga mencapai kondisi stabil atau ketika perubahan jaraknya dianggap tidak signifikan [8]. Pendekatan ini bertujuan meminimalkan variansi dalam kluster dan memaksimalkan pemisahan antar kluster.

Metode *K-Means* digunakan untuk mengenali pola kesamaan performa jaringan antar kabupaten/kota berdasarkan hasil reduksi PCA dari tiga variabel utama. Algoritma ini juga efisien secara komputasi dan mampu menghasilkan pembagian data yang terstruktur serta mudah diinterpretasikan secara visual.

2.2.4.2 Gaussian Mixture Model (GMM)

Algoritma *Gaussian Mixture Model* menggunakan pendekatan probabilistik untuk membentuk kluster dengan mengasumsikan bahwa data berasal dari kombinasi beberapa distribusi *Gaussian*. Proses estimasi dilakukan melalui algoritma *Expectation–Maximization (EM)* yang memperbarui parameter mean, kovarians, dan bobot tiap distribusi hingga diperoleh model dengan kemungkinan maksimum (*maximum likelihood*) [15].

Berbeda dari pendekatan berbasis jarak, GMM menentukan probabilitas keanggotaan setiap titik terhadap masing-masing kluster, bukan penetapan pasti. Dalam penelitian ini, GMM diterapkan untuk mendeteksi variasi spasial performa jaringan internet yang tidak sepenuhnya linier. Dengan demikian, GMM menghasilkan pendekatan yang berbeda untuk pemetaan wilayah kluster.

2.2.5 Evaluation

Tahap evaluasi atau interpretasi adalah langkah terakhir dalam proses Penemuan Pengetahuan dalam Basis Data (KDD) yang bertujuan untuk menilai kualitas hasil kluster dan memberi makna secara substansial. Proses evaluasi dilakukan dengan menggunakan tiga metrik validasi internal yang sering diterapkan dalam analisis kluster, yaitu *Silhouette Score, Davies–Bouldin Index,* serta *Calinski–Harabasz Index* seperti yang ditunjukkan pada Tabel 1. Ketiga metrik ini memberikan informasi mengenai seberapa baik setiap objek terletak dalam kluster serta seberapa jelas batas antara kluster yang telah terbentuk.

- Silhouette Score (S)* menilai seberapa mirip sebuah objek dengan kluster yang sesuai dibandingkan kluster lainnya. Angka evaluasi bernilai dari -1 sampai dengan 1 , yang mana nilai yang lebih dekat ke 1 memperlihatkan bahwa objek berada pada kluster yang benar [23]
- Davies–Bouldin Index (DBI)* mengevaluasi rasio jarak antar kluster terhadap penyebaran internal kluster. Penurunan nilai DBI mengindikasikan peningkatan yang lebih baik dalam pemisahan antara kluster [24]
- Calinski–Harabasz Index (CHI)* mengevaluasi perbandingan antara variasi tiap kluster dan variasi di dalamnya. Nilai CHI yang lebih tinggi menandakan kualitas pemisahan antar kluster yang lebih baik [25]

3. HASIL DAN PEMBAHASAN

3.1 Hasil Analisis Knowledge Discovery in Database (KDD)

3.1.1 Hasil Selection

Pada tahap ini dilakukan proses selection sebagai langkah awal dalam kerangka *Knowledge Discovery in Databases* (KDD) untuk menyaring data mentah agar sesuai dengan fokus wilayah dan periode penelitian. Gambar 3 berikut menampilkan hasil penyaringan data yang telah disesuaikan dengan kriteria tersebut.

```

=== [1] Tahap SELECTION ===
Total data tahun 2025: 51 baris

```

	year	quarter	kabkota	download_mbps	upload_mbps	latency_ms	tests	kuartal
0	2025	Q1	Banyu Asin	26.8780	15.1470	24.0	2188	Q1_2025
1	2025	Q1	Empat Lawang	42.0305	14.7410	27.0	296	Q1_2025
2	2025	Q1	Lahat	24.8960	14.2430	29.0	954	Q1_2025
3	2025	Q1	Lubuklinggau	25.4310	12.7690	28.0	486	Q1_2025
4	2025	Q1	Muara Enim	21.8090	13.8005	28.0	2591	Q1_2025

Gambar 3. Hasil Ouput Tahap *Selection*

Untuk tahap pertama, dilakukan penyaringan data dari Ookla Open Data agar hanya mencakup periode tahun 2025 dan wilayah Provinsi Sumatera Selatan. Hasil seleksinya yang terdapat pada Gambar 3, menunjukkan terdapat 51 baris data yang terdiri dari 17 kabupaten/kota dan tiga kuartal (Q1–Q3 2025). Setiap baris memuat lima dari tiga variabel utama yaitu kecepatan unduh (*download speed*), kecepatan unggah (*upload speed*), dan latensi (*latency*). Data tersebut menggambarkan kondisi nyata performa jaringan pada periode terkini di Sumatera Selatan.

3.1.2 Hasil *Preprocessing*

Pada tahap kedua, data hasil seleksi pada Gambar 3 diproses agar siap digunakan dalam analisis klusterisasi. Proses *preprocessing* dilakukan dengan melakukan agregasi rata-rata dari setiap variabel utama berdasarkan masing-masing kabupaten/kota. Tujuan dari tahapan ini adalah menyederhanakan data mentah menjadi representasi yang lebih stabil dan mudah dianalisis tanpa kehilangan karakteristik utama dari setiap wilayah.

```

=== [2] Tahap PREPROCESSING ===
Data agregasi per kabupaten: 17 wilayah

```

	kabkota	download_mbps	upload_mbps	latency_ms
0	Banyu Asin	31.316500	14.773167	21.666667
1	Empat Lawang	38.344500	15.747000	28.833333
2	Lahat	27.427000	13.619500	29.833333
3	Lubuklinggau	27.377333	15.160833	29.000000
4	Muara Enim	25.814833	14.130333	26.666667
5	Musi Banyuasin	26.075000	12.243500	24.666667
6	Musi Rawas	32.773667	16.424167	29.666667
7	Musi Rawas Utara	29.940167	15.881000	29.000000
8	Ogan Ilir	25.506167	14.626667	23.333333
9	Ogan Komering Ilir	26.369500	10.833833	27.000000
10	Ogan Komering Ulu	31.054833	14.821167	24.666667
11	Ogan Komering Ulu Selatan	18.873000	8.792000	33.000000
12	Ogan Komering Ulu Timur	24.866500	13.005167	26.666667
13	Pagar Alam	24.200333	13.994667	29.500000
14	Palembang	39.488500	19.458000	22.000000
15	Penukal Abab Lematang Ilir	27.263667	14.324000	26.000000
16	Prabumulih	28.513000	16.643333	23.666667

Gambar 4. Hasil Ouput Tahap *Preprocessing*

Setelah proses agregasi dari Gambar 4, diperoleh total 17 baris data yang masing-masing merepresentasikan satu kabupaten/kota di Provinsi Sumatera Selatan. Nilai rata-rata kecepatan unduh (*download speed*) tertinggi tercatat di Kota Palembang sebesar 39,48 Mbps dengan latensi 22 ms, sedangkan nilai kecepatan unduh terendah berada di Kabupaten Ogan Komering Ulu Selatan sebesar 18,87 Mbps dengan latensi 33 ms. Hasil ini secara awal mengindikasikan adanya kesenjangan antara wilayah perkotaan dan nonperkotaan dalam hal kualitas jaringan internet.

3.1.3 Hasil *Transformation*

Selanjutnya, hasil *preprocessing* pada Gambar 4 yang telah distandarkan kemudian direduksi dimensinya menggunakan *Principal Component Analysis* (PCA) agar pola hubungan antarvariabel dapat terlihat lebih jelas dalam ruang dua dimensi. Proses transformasi menghasilkan dua komponen PC1 serta PC2, yang mewakili variasi nilai performa jaringan di setiap kabupaten/kota, sebagaimana yang ditunjukkan pada Gambar 5.

```

=== [3] Tahap TRANSFORMATION ===
✅ Data berhasil ditransformasi ke dua dimensi ut

```

	PC1	PC2	kabkota
0	0.787400	-0.936245	Banyu Asin
1	1.482948	1.085065	Empat Lawang
2	-0.592006	0.565790	Lahat
3	-0.005772	0.458352	Lubuklinggau
4	-0.397211	-0.193719	Muara Enim
5	-0.897618	-0.710722	Musi Banyuasin
6	1.019204	0.979398	Musi Rawas
7	0.543343	0.642660	Musi Rawas Utara
8	-0.045440	-0.915071	Ogan Ilir
9	-1.505829	-0.254692	Ogan Komering Ilir
10	0.580994	-0.293305	Ogan Komering Ulu
11	-3.475020	0.520734	Ogan Komering Ulu Selatan
12	-0.900371	-0.305125	Ogan Komering Ulu Timur
13	-0.814126	0.325393	Pagar Alam
14	3.351298	-0.155522	Palembang
15	-0.118509	-0.245965	Penukal Abab Lematang Ilir
16	0.986715	-0.567026	Prabumulih

Gambar 5. Hasil Ouput Tahap *Preprocessing*

Nilai PC1 mencerminkan kombinasi linier dari ketiga variabel awal yang menggambarkan arah utama variasi data, sedangkan PC2 memberikan dimensi tambahan untuk variasi sekunder yang bersifat ortogonal terhadap PC1. Sebagaimana terlihat pada Gambar 5, setiap kabupaten/kota memiliki nilai PC1 dan PC2 yang berbeda, menunjukkan adanya perbedaan karakteristik performa jaringan di masing-masing wilayah. Untuk memperjelas kontribusi masing-masing variabel terhadap komponen utama tersebut, dilakukan analisis korelasi antara variabel asli dan hasil transformasi PCA, sebagaimana disajikan pada Gambar 6.

```

=== Korelasi Variabel terhadap Komponen PCA ===

```

	PC1	PC2
download_mbps	0.919	0.198
upload_mbps	0.963	0.060
latency_ms	-0.559	0.824

Gambar 6. Korelasi Variabel Terhadap Komponen PCA

Berdasarkan Gambar 6, variabel *download_mbps* (0,919) dan *upload_mbps* (0,963) memiliki korelasi positif kuat terhadap PC1, sedangkan *latency_ms* berkorelasi negatif terhadap PC1 (-0,559) dan positif tinggi terhadap PC2 (0,824). Hal ini menunjukkan bahwa PC1 merepresentasikan dimensi utama kualitas jaringan yang berkaitan dengan kecepatan unduh dan unggah, sedangkan PC2 menggambarkan variasi stabilitas jaringan melalui perbedaan latensi antarwilayah. Nilai positif pada PC1 menandakan area dengan kecepatan tinggi dan latensi rendah, sedangkan nilai negatif menunjukkan daerah dengan performa di bawah rata-rata.

3.1.4 Hasil *Data Mining*

Di tahap ini, dua algoritma *K-Means Clustering* dan *Gaussian Mixture Model (GMM)* diterapkan untuk membentuk kelompok wilayah dengan karakteristik serupa. Kedua algoritma diuji dengan kluster (K) yang bervariasi, yaitu 2, 3, 4, dan 5, untuk menentukan konfigurasi paling optimal berdasarkan tiga metrik evaluasi internal. Hasil pengujian masing-masing algoritma ditampilkan pada Tabel 2 untuk *K-Means* dan Tabel 3 untuk *GMM* sebagai dasar analisis perbandingan performa antar-metode.

Tabel 2. *Output* Metrik Evaluasi *K-Means*

<i>K-Means</i>			
Klaster	<i>Silhouette</i>	DBI	CHI
2	0.363386	0.916888	13.095513

K-Means			
Klaster	Silhouette	DBI	CHI
3	0.351170	0.688389	13.754431
4	0.317603	0.688389	16.584041
5	0.388159	0.471603	24.044537

Hasil pengujian pada Tabel 2 menunjukkan bahwa nilai *Silhouette* tertinggi pada K-Means diperoleh pada K = 5 sebesar 0,388, sementara DBI terendah (0,472) dan CHI tertinggi (24,04) juga muncul pada konfigurasi yang sama. Hal ini menunjukkan bahwa lima klaster menghasilkan struktur pembagian yang lebih stabil dan terpisah dengan baik dibandingkan nilai K lainnya. Dengan demikian, pada model *K-Means*, peningkatan jumlah klaster cenderung meningkatkan pemisahan dan kepadatan antar-wilayah.

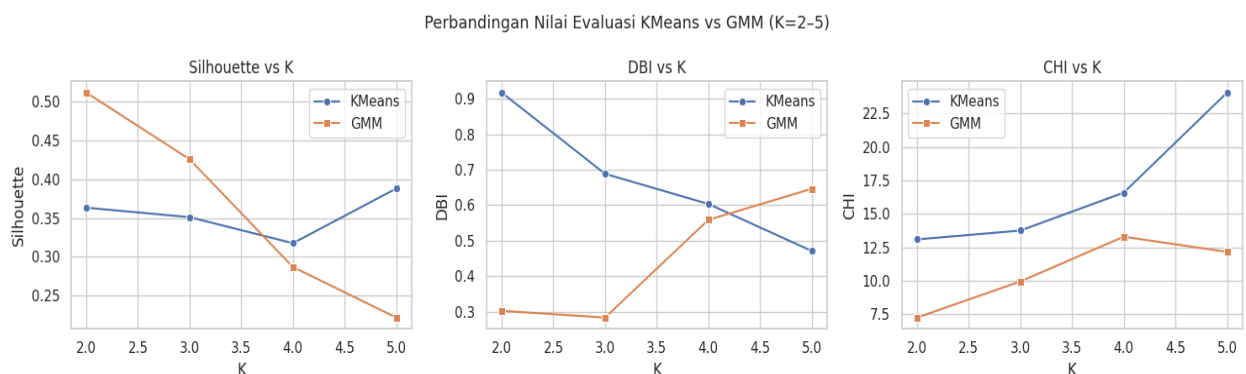
Tabel 3. Output Metrik Evaluasi GMM

GMM			
Klaster	Silhouette	DBI	CHI
2	0.511902	0.303050	7.264612
3	0.425861	0.284029	9.960720
4	0.286820	0.558914	13.296655
5	0.221626	0.646237	12.162636

Pada Tabel 3, algoritma GMM menunjukkan performa terbaik pada K = 2, dengan nilai *Silhouette* = 0,512 dan DBI = 0,303. Hasil ini menandakan bahwa pemisahan berbasis probabilistik lebih optimal ketika jumlah klaster lebih sedikit, karena GMM dapat menangkap variasi bentuk dan distribusi data dengan lebih fleksibel. Namun, nilai CHI yang relatif lebih kecil (7,26) menunjukkan bahwa kepadatan antar-klaster belum sekuat pada model *K-Means*. Secara keseluruhan, GMM menampilkan stabilitas tinggi pada jumlah klaster rendah namun dengan kompleksitas yang lebih halus dalam pemisahan data.

3.1.5 Hasil Evaluation

Berdasarkan hasil pengujian pada tahap sebelumnya, penelitian ini menetapkan K=3 sebagai konfigurasi optimal yang memberikan keseimbangan terbaik antara kepadatan, pemisahan, dan kestabilan klaster untuk kedua algoritma. Seperti yang terlihat pada Gambar 7, nilai metrik pada konfigurasi tersebut menunjukkan hasil yang konsisten, dengan *K-Means* memperoleh *Silhouette* = 0.351, DBI = 0.688, dan CHI = 13.75, sedangkan GMM mencapai *Silhouette* = 0.426, DBI = 0.284, dan CHI = 9.96.



Gambar 7. Grafik Perbandingan Nilai Evaluasi *K-Means* dan GMM

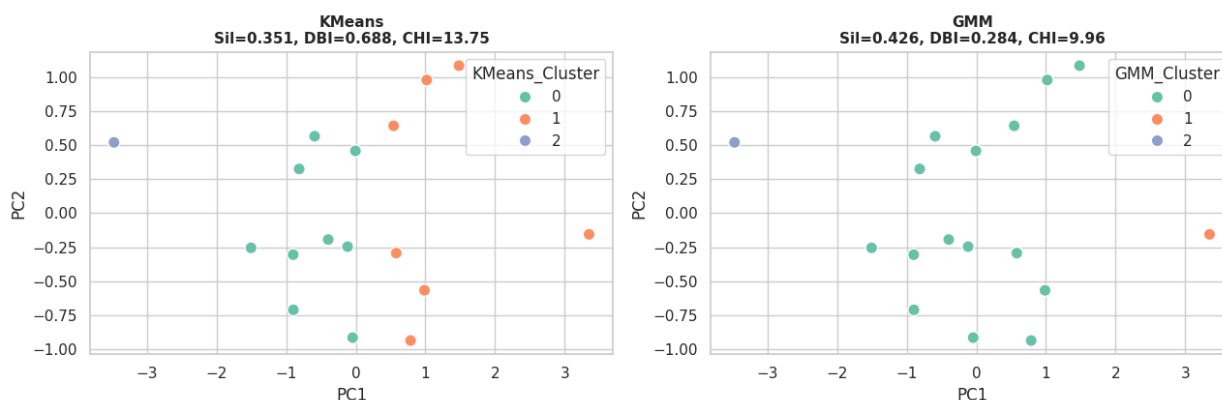
Gambar 7 juga memperkuat hasil tersebut melalui visualisasi hubungan antar-metrik terhadap variasi nilai K. Grafik pertama (*Silhouette vs K*) memperlihatkan bahwa GMM memiliki skor tertinggi pada K=2 namun menurun drastis pada K>3, sedangkan *K-Means* menunjukkan peningkatan stabil hingga K=5. Grafik kedua (*DBI vs K*) memperlihatkan pola menurun pada *K-Means* yang berarti kualitas pemisahan antar-klaster semakin baik seiring bertambahnya jumlah K. Sementara itu, grafik ketiga (*CHI vs K*) menunjukkan bahwa *K-Means* secara konsisten memperoleh nilai CHI lebih baik dibanding GMM, menegaskan kestabilan antar-klaster yang lebih baik. Berdasarkan ketiga metrik tersebut, penelitian ini menetapkan K=3 sebagai konfigurasi optimal yang seimbang untuk kedua algoritma, yang selanjutnya digunakan pada tahap evaluasi hasil klusterisasi.

● Hasil Klasterisasi GMM			● Hasil Klasterisasi K-MEANS		
Klaster		Kabupaten/Kota	Klaster		Kabupaten/Kota
0	1	Banyu Asin	0	1	Lahat
1	1	Empat Lawang	1	1	Lubuklinggau
2	1	Lahat	2	1	Muara Enim
3	1	Lubuklinggau	3	1	Musi Banyuasin
4	1	Muara Enim	4	1	Ogan Ilir
5	1	Musi Banyuasin	5	1	Ogan Komering Ilir
6	1	Musi Rawas	6	1	Ogan Komering Ulu Timur
7	1	Musi Rawas Utara	7	1	Pagar Alam
8	1	Ogan Ilir	8	1	Penukal Abab Lematang Ilir
9	1	Ogan Komering Ilir	9	2	Banyu Asin
10	1	Ogan Komering Ulu	10	2	Empat Lawang
11	1	Ogan Komering Ulu Timur	11	2	Musi Rawas
12	1	Pagar Alam	12	2	Musi Rawas Utara
13	1	Penukal Abab Lematang Ilir	13	2	Ogan Komering Ulu
14	1	Prabumulih	14	2	Palembang
15	2	Palembang	15	2	Prabumulih
16	3	Ogan Komering Ulu Selatan	16	3	Ogan Komering Ulu Selatan

Gambar 8. Hasil Klasterisasi Wilayah K-Means dan GMM

Hasil pembentukan kluster pada Gambar 8 menunjukkan bahwa algoritma *K-Means* membagi wilayah Provinsi Sumatera Selatan menjadi tiga kelompok utama. Kluster 1 terdiri atas sembilan kabupaten, yaitu Lahat, Lubuklinggau, Musi Banyuasin, hingga Penukal Abab Lematang Ilir. Kluster 2 mencakup tujuh wilayah, antara lain Banyuasin, Palembang, hingga Prabumulih. Sementara itu, Kluster 3 hanya mencakup satu wilayah, yaitu Ogan Komering Ulu Selatan. Pada algoritma *Gaussian Mixture Model* (GMM), hasil klasterisasi pada Gambar 8 juga membentuk tiga kelompok dengan pola yang hampir serupa. Kluster 1 mencakup sebagian besar wilayah, yaitu 15 kabupaten/kota di Sumatera Selatan. Kluster 2 hanya terdiri dari Kota Palembang, sedangkan Kluster 3 kembali diisi oleh Kabupaten Ogan Komering Ulu Selatan.

Perbandingan Hasil Klasterisasi KMeans vs GMM (K=3)



Gambar 9. Grafik Persebaran Pola Klasterisasi K-Means dan GMM

Gambar 9 memperlihatkan hasil klasterisasi menggunakan algoritma *K-Means* dan *Gaussian Mixture Model* (GMM) dengan jumlah kluster $K = 3$ berdasarkan hasil reduksi dua dimensi PCA (PC1 dan PC2). Kedua model berhasil membentuk tiga kelompok wilayah dengan pola yang cukup konsisten. Pada model *K-Means*, setiap kluster tampak lebih terpisah dan membentuk struktur yang relatif seimbang antara kelompok satu dengan lainnya. Sementara itu, model GMM menghasilkan pembagian kluster dengan pola yang serupa, namun distribusi titik antarwilayah terlihat sedikit lebih merata. Secara keseluruhan, kedua algoritma menunjukkan hasil pengelompokan yang sejalan, dengan perbedaan utama pada tingkat kerapatan dan sebaran antar titik yang menggambarkan variasi karakteristik jaringan antarwilayah di Provinsi Sumatera Selatan.

3.2 Pembahasan

Pada tahap ini dilakukan analisis hasil klasterisasi menggunakan dua algoritma, yaitu *K-Means* dan *Gaussian Mixture Model* (GMM), dengan konfigurasi optimal $K = 3$. Ketiga metrik evaluasi yang digunakan meliputi *Silhouette*, *Davies–Bouldin Index* (DBI), dan *Calinski–Harabasz Index* (CHI) sebagaimana ditampilkan pada Tabel 4.

Tabel 4. Hasil Metrik K=3 *K-Means* dan GMM

Metrik	K = 3	
	<i>K-Means</i>	GMM
<i>Silhouette</i>	0.351170	0.425861
DBI	0.688389	0.284029
CHI	13.754431	9.960720

Hasil pada Tabel 4 menunjukkan bahwa GMM memiliki nilai *Silhouette* lebih tinggi (0,426) dan DBI lebih rendah (0,284) dibandingkan *K-Means* yang memperoleh *Silhouette* 0,351 dan DBI 0,688. Perbedaan ini menunjukkan bahwa GMM menghasilkan kluster yang lebih seimbang dan padat, sedangkan *K-Means* cenderung membentuk batas kluster yang lebih tegas namun kurang fleksibel terhadap data yang memiliki variasi kompleks. Selain itu, nilai CHI GMM sebesar 9,96 yang lebih rendah dibanding *K-Means* (13,75) mencerminkan sifat probabilistik GMM yang cenderung membentuk kluster secara lebih halus dan stabil terhadap variasi distribusi data.

Tabel 5. Persebaran Pola Klasterisasi *K-Means* dan GMM

KMeans Cluster	Rata-rata Variabel per Kluster (KMeans):		
	Rata-rata <i>Download</i> (Mbps)	Rata-rata <i>Upload</i> (Mbps)	Rata-rata Latensi (ms)
0	26.100037	13.548722	26.962963
1	33.061595	16.249690	25.642857
2	18.873000	8.792000	33.000000

Hasil pada Tabel 5 memperlihatkan bahwa *K-Means* membentuk tiga kluster utama, di mana kluster 1 yang mencakup wilayah Palembang, Prabumulih, dan Musi Rawas menunjukkan kecepatan unduh tertinggi yaitu 33,06 Mbps dengan latensi 25,64 ms. Sebaliknya, kluster 2 yang hanya beranggotakan Kabupaten Ogan Komering Ulu Selatan memiliki kecepatan unduh terendah 18,87 Mbps dan latensi tertinggi 33 ms. Kesenjangan kecepatan unduh sekitar 14 Mbps antara wilayah performa tinggi dan rendah mencerminkan adanya perbedaan signifikan dalam pemerataan infrastruktur jaringan antarwilayah.

Tabel 6. Tabel Persebaran Pola Klasterisasi *K-Means* dan GMM

GMM Cluster	Rata-rata Variabel per Kluster (GMM):		
	Rata-rata <i>Download</i> (Mbps)	Rata-rata <i>Upload</i> (Mbps)	Rata-rata Latensi (ms)
0	28.4562	14.415222	26.677778
1	39.4885	19.458000	22.000000
2	18.8730	8.792000	33.000000

Sementara itu, hasil GMM pada Tabel 6 menunjukkan pola yang hampir serupa dengan *K-Means*, namun Palembang membentuk kluster tersendiri dengan kecepatan unduh tertinggi sebesar 39,49 Mbps dan latensi terendah 22 ms, sedangkan Ogan Komering Ulu Selatan kembali konsisten sebagai wilayah dengan performa terendah (*download* 18,87 Mbps; *latency* 33 ms) dengan kesenjangan kecepatan unduh sekitar 20 Mbps. Hal ini menunjukkan bahwa GMM lebih mampu menangkap perbedaan struktur data antarwilayah dan memisahkan area dengan karakteristik ekstrem secara lebih akurat dibanding *K-Mean*.

Perbedaan hasil antara kedua algoritma disebabkan oleh perbedaan sifat dasar keduanya. *K-Means* bersifat deterministik dan berbasis jarak *Euclidean* sehingga efektif pada data homogen dengan batas kluster tegas, sedangkan GMM menggunakan pendekatan probabilistik yang memungkinkan satu wilayah memiliki peluang keanggotaan ganda sehingga lebih adaptif terhadap distribusi data yang tidak linier. Hasil GMM yang lebih stabil menunjukkan bahwa variasi kualitas jaringan di Sumatera Selatan tidak terpisah secara kaku antarwilayah, melainkan membentuk pola yang saling tumpang tindih. Berdasarkan rata-rata kecepatan unduh, wilayah perkotaan seperti Palembang dan Prabumulih mendominasi kluster berperforma tinggi dengan kecepatan unduh 33,06 Mbps dan latensi 25,64 ms. , sedangkan Ogan Komering Ulu Selatan tertinggal dengan rata-rata selisih sekitar 18 Mbps (*download* 18,87 Mbps; *latency* 33 ms). Jika dibandingkan dengan wilayah lain, kesenjangan ini tergolong menengah karena di Pulau Sumatera rata-rata kecepatan internet berkisar 20-30 Mbps dan di Pula Jawa mencapai 40 Mbps, yang berarti pemerataan kualitas jaringan di Sumatera Selatan masih relatif lebih baik, namun belum sepenuhnya merata di seluruh kabupaten/kota.

Konsistensi rendahnya performa di Ogan Komering Ulu Selatan dapat dipengaruhi oleh kondisi geografis dan infrastruktur, seperti topografi bergelombang, kepadatan penduduk rendah, serta keterbatasan jaringan *fiber optic* dan jumlah BTS aktif. Sebaliknya, Palembang dan Prabumulih menunjukkan performa unggul karena infrastruktur digitalnya lebih padat dan terkoneksi dengan jaringan *backbone* yang kuat. Temuan ini memberikan implikasi kebijakan bagi Dinas Komunikasi dan Informatika Provinsi Sumatera Selatan untuk memprioritaskan pembangunan infrastruktur digital di wilayah berperforma rendah dengan memperluas jangkauan *fiber optic*, meningkatkan kapasitas BTS, dan memperkuat kerja sama dengan operator seluler. Sementara itu, wilayah berperforma tinggi dapat dijadikan

acuan dalam pemerataan akses jaringan di provinsi lain. Secara keseluruhan, kombinasi metode *K-Means* dan GMM dalam kerangka KDD terbukti efektif dalam memetakan variasi spasial kualitas jaringan internet seluler serta memberikan dasar bagi kebijakan peningkatan pemerataan infrastruktur menuju Visi Indonesia Digital 2045.

4. KESIMPULAN

Penelitian ini berhasil menilai kondisi dan kinerja kualitas jaringan internet seluler di 17 kabupaten/kota di Provinsi Sumatera Selatan menggunakan data Ookla Open Data tahun 2025 dengan pendekatan *Knowledge Discovery in Databases* (KDD). Hasil analisis menunjukkan adanya variasi signifikan antarwilayah, di mana Kota Palembang memiliki performa jaringan tertinggi dengan rata-rata kecepatan unduh 39,48 Mbps dan latensi 22 ms, sedangkan Kabupaten Ogan Komering Ulu Selatan menunjukkan performa terendah dengan 18,87 Mbps dan 33 ms. Rata-rata selisih sekitar 18 Mbps tersebut menggambarkan adanya kesenjangan digital yang nyata antara wilayah urban dan nonurban di provinsi ini. Proses klusterisasi menggunakan algoritma *K-Means* dan *Gaussian Mixture Model* (GMM) menghasilkan tiga kelompok wilayah ($K=3$) dengan pola pengelompokan yang serupa namun karakteristik yang berbeda. Berdasarkan hasil evaluasi metrik internal, GMM menunjukkan performa yang lebih stabil dengan nilai *Silhouette* 0,426 dan *Davies–Bouldin Index* 0,284, dibandingkan *K-Means* yang memperoleh *Silhouette* 0,351 dan DBI 0,688, menandakan bahwa GMM lebih mampu membentuk kluster yang padat dan terpisah dengan baik. Sementara itu, *Calinski–Harabasz Index* GMM sebesar 9,960 yang lebih rendah dari *K-Means* (13,754) mencerminkan adanya kompromi antara kepadatan dan pemisahan kluster, namun tetap menunjukkan representasi yang lebih realistis terhadap variasi kualitas jaringan antarwilayah. Secara keseluruhan, hasil penelitian menegaskan bahwa wilayah perkotaan seperti Palembang dan Prabumulih mendominasi kluster dengan performa tinggi, sedangkan Ogan Komering Ulu Selatan konsisten dalam kluster berperforma rendah. Kondisi ini menandakan perlunya intervensi strategis dari Dinas Komunikasi dan Informatika Provinsi Sumatera Selatan dengan langkah konkret yang dapat berupa pemerataan jaringan backbone *fiber optic*, peningkatan kapasitas *Base Transceiver Station* (BTS), dan kolaborasi aktif dengan penyedia layanan seluler untuk memperkecil kesenjangan digital antarwilayah. Penelitian ini masih terbatas pada tiga indikator teknis (*download*, *upload*, dan *latency*) serta periode analisis hingga kuartal ketiga tahun 2025, sehingga disarankan agar penelitian selanjutnya menambahkan variabel spasial seperti kepadatan BTS, jumlah pengguna aktif, serta kondisi topologi wilayah guna memberikan gambaran yang lebih komprehensif mengenai pemerataan dan kualitas layanan internet seluler di tingkat daerah.

REFERENCES

- [1] D. Sudiantini, M. P. Ayu, M. Cheimel, A. Shawirdra, M. A. Prastuti, and M. Apriyana, “Transformasi Digital : Dampak , Tantangan , Dan Peluang Untuk Pertumbuhan Ekonomi Digital,” *TRENDING J. Manaj. dan Ekon.*, vol. 1, no. 3, 2023, doi: 10.30640/trending.v1i3.1115.
- [2] Digital2024, “Launching Visi Indonesia Digital 2045: Percepatan Transformasi Digital Demi Wujudkan Indonesia Emas,” kominfo. Accessed: Oct. 20, 2025. [Online]. Available: <https://digital2045.id/launching-visi-indonesia-digital-2045-percepatan-transformasi-digital-demi-wujudkan-indonesia-emas/>
- [3] A. A. Dzirkullah, “Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet Dengan Metode Centroid Linkage,” *J. Math. Comput. Stat.*, vol. 5, no. 1, pp. 48–57, 2022, doi: 10.35580/jmathcos.v5i1.32363.
- [4] Asosiasi Penyelenggara Jasa Internet Indonesia, “APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang,” Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). Accessed: Oct. 20, 2025. [Online]. Available: <https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang>
- [5] N. Salsi Anugrah, H. Adilah, D. P. Panggabean, and R. P. Laksana, “Analisis Kecepatan Internet Seluler Di Indonesia Berdasarkan Regional,” *J. Data Anal. Information, Comput. Sci.*, vol. 2, no. 2, pp. 155–158, 2025, doi: 10.70248/jdaics.v2i2.1821.
- [6] Ookla LLC, “Speedtest by Ookla Global Fixed and Mobile Network Performance Maps,” Open Data on AWS. Accessed: Oct. 20, 2025. [Online]. Available: <https://registry.opendata.aws/speedtest-global-performance/>
- [7] O. Mar’atun Sholihah, N. Suarna, G. Dwilestari, and N. R., “Implementasi Metode K-Means Clustering Untuk Menganalisa Penerima Bantuan Di Desa Palasah,” *J. Inform. dan Teknol. Inf.*, vol. 2, no. 1, pp. 154–160, 2023, doi: 10.56854/jt.v2i1.123.
- [8] S. W. Wijayanti and A. Kesumawati, “Perbandingan Analisis K-means dan Hierarchical Clustering dalam Mengelompokkan Kecamatan di Kabupaten Grobogan Berdasarkan Jumlah Titik Kejadian Bencana Alam,” *Emerg. Stat. Data Sci. J.*, vol. 3, no. 2, pp. 597–617, 2025, doi: 10.20885/esds.vol3.iss.2.art14.
- [9] A. Fauzan, N. Suarna, I. Ali, and H. Susana, “Penerapan Algoritma K-Means Clustering Untuk Meningkatkan Model Pengelompokan Dan Kinerja Jaringan Wi-Fi Secara Optimal,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6272.
- [10] P. Chyan, “Segmentasi Kulit Manusia Dengan Ekstraksi Fitur Warna Dan Algoritma GMM-EM,” *J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 151–156, 2022, doi: 10.37792/jukanti.v5i1.468.
- [11] K. E. Setiawan and A. Kurniawan, “Model, Pengelompokan Rumah Sakit Di Jakarta Menggunakan DbSCAN, Gaussian Mixture, Dan Hierarchical Clustering,” *J. Inform. Terpadu*, vol. 9, no. 2, pp. 149–156, 2023, doi: 10.54914/jit.v9i2.995.
- [12] G. A. Alde, D. Faiza, and A. Hadi, “Analisis Jaringan Seluler Menggunakan K-Means Clustering untuk Identifikasi Daerah Blank Spot dan Non Blank Spot di Kabupaten Padang Pariaman,” *J. Pendidik. Tambusai*, vol. 9, no. 2, pp. 27236–27244, 2025, doi: 10.31004/jptam.v9i2.31229.
- [13] T. Apriliana and E. Widodo, “Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 286–296,



- 2023, doi: 10.24002/konstelasi.v3i2.7143.
- [14] R. Watin, N. P. Permatasari, A. W. Wijayanto, and W. Marsisno, "Analisis Cluster Kondisi Keterampilan, Akses dan Fasilitas Teknologi Informasi dan Komunikasi di Indonesia," *Komputika J. Sist. Komput.*, vol. 13, no. 1, pp. 83–92, 2024, doi: 10.34010/komputika.v13i1.10796.
- [15] H. Alizadeh, H. Vranken, A. Zuquete, and A. Miri, "Timely Classification and Verification of Network Traffic Using Gaussian Mixture Models," *IEEE ACCESS*, vol. 8, pp. 91287–91302, 2020, doi: 10.1109/ACCESS.2020.2992556.
- [16] R. Rahmati and A. W. Wijayanto, "Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 73–80, 2021, doi: 10.26798/jiko.v5i2.422.
- [17] J. Han, J. Pei, and H. Tong, *Data Mining Concepts and Techniques*. Katey Birtcher, 2023.
- [18] J. V. Lenda and M. A. I. Pakereng, "Analisis Tingkat Pengangguran Di Kota Palopo Menggunakan Metode K-Means," *Zo. J. Sist. Inf.*, vol. 6, no. 2, pp. 400–408, 2024, doi: 10.31849/zn.v6i2.20061.
- [19] D. Saputri, H. Mustafidah, F. Wibowo, and D. K. Hakim, "Segmentasi Provinsi di Indonesia Berdasarkan Akses Fasilitas Dasar dan Pengeluaran Rumah Tangga Menggunakan K-Means," *J. Sist. Komput. dan Inform.*, vol. 6, pp. 159–168, 2025, doi: 10.30865/json.v6i4.
- [20] V. Sharma, "A Study on Data Scaling Methods for Machine Learning," *Int. J. Glob. Acad. Sci. Res.*, vol. 1, no. 1, pp. 31–42, 2022, doi: 10.55938/ijgasr.v1i1.4.
- [21] D. Wang and Z. Ma, "Carbon meets credit : enhancing corporate ratings using principal component analysis and Gaussian mixture models," *Digit. Econ. Sustain. Dev.*, vol. 2, no. 23, 2024, doi: 10.1007/s44265-024-00048-z.
- [22] E. Elhaik, *Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated*, vol. 12, no. 1. Nature Publishing Group UK, 2022. doi: 10.1038/s41598-022-14395-4.
- [23] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids," *J. Gaussian*, vol. 13, no. 1, pp. 13–24, 2024, doi: 10.14710/j.gauss.13.1.13-24.
- [24] J. K. Wororomi *et al.*, *DATA MINING (Memahami Pola di Balik Angka)*. EUREKA MEDIA AKSARA, 2021.
- [25] I. K. Y. Triantara, M. A. Raharja, and I. B. G. Sarasvananda, "Segmentasi Pelanggan Berbasis RFMT Menggunakan K-Means dan Hierarchical Clustering," *J. Nas. Teknol. Inf. dan Apl.*, vol. 3, no. 4, pp. 917–924, 2025, doi: <https://doi.org/10.58776/jriti.v2i3.156>.