

Klastering Kecepatan Internet Operator Telkomsel Berdasarkan Sebaran Site BTS (Base Transceiver Station) Menggunakan Metode DBSCAN

Arif Alifudin*, Irfan Pratama

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: ^{1,*}211210064@student.mercubuana-yogya.ac.id, ²irfanp@mercubuana-yogya.ac.id

Email Penulis Korespondensi: 211210064@student.mercubuana-yogya.ac.id

Submitted: 27/09/2024; Accepted: 30/11/2024; Published: 30/11/2024

Abstrak—Perkembangan teknologi telekomunikasi seluler kini telah mencapai tahap yang lebih maju dengan hadirnya teknologi 4G LTE. Dibandingkan dengan teknologi sebelumnya seperti 3G dan 2G, teknologi ini menawarkan kemampuan transfer data dengan kecepatan akses yang jauh lebih baik. Hal ini membuat 4G LTE menjadi tulang punggung bagi layanan komunikasi modern yang mendukung kebutuhan pengguna akan akses internet yang cepat dan stabil. Namun meskipun teknologi 4G LTE telah diterapkan secara luas, masih terdapat tantangan yang perlu diatasi untuk memastikan kualitas jaringan tetap optimal. Kualitas jaringan internet di Daerah Istimewa Yogyakarta (DIY) telah mengalami peningkatan yang signifikan dalam beberapa tahun terakhir, tetapi kendala seperti terbatasnya infrastruktur masih dirasakan terutama di area pedesaan dan wilayah terluar. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan area berdasarkan kualitas jaringan. Oleh karena itu diperlukan analisis data mining terhadap data yang ada menggunakan algoritma DBSCAN sehingga akan terbentuk cluster yang dibagi menurut kualitas jaringan. Setelah dilakukan analisis dengan menggunakan nilai epsilon = 0,5 dan nilai minpts = 5, cluster yang terbentuk yaitu sebanyak 5 cluster dengan nilai *silhouette* 0.216471397367446, yang mengindikasikan bahwa kualitas clustering tergolong rendah, yang kemungkinan disebabkan oleh distribusi data atau parameter yang kurang optimal. Meskipun demikian, hasil clustering yang diperoleh tetap memberikan wawasan yang berguna untuk analisis sebaran site dan performa jaringan.

Kata Kunci: Base Transceiver Station; Cluster; Data Mining; DBSCAN; Pre-Processing

Abstract—The development of cellular telecommunications technology has now reached a more advanced stage with the presence of 4G LTE technology. Compared to previous technologies such as 3G and 2G, this technology offers data transfer capabilities with much better access speeds. This makes 4G LTE the backbone for modern communication services that support users' needs for fast and stable internet access. However, even though 4G LTE technology has been widely implemented, there are still challenges that need to be overcome to ensure network quality remains optimal. The quality of the internet network in the Special Region of Yogyakarta (DIY) has experienced significant improvements in recent years, but obstacles such as limited infrastructure are still felt, especially in rural and outermost areas. This research aims to analyze and group areas based on network quality. Therefore, data mining analysis of existing data is needed using the DBSCAN algorithm so that clusters will be formed which are divided according to network quality. After carrying out analysis using the epsilon value = 0.5 and the minpts value = 5, the clusters formed were 5 clusters with a silhouette value of 0.216471397367446, which indicates that the quality of the clustering is relatively low, which is possibly caused by less than optimal distribution of data or parameters. Nevertheless, the clustering results obtained still provide useful insights for analyzing site distribution and network performance

Keywords: Base Transceiver Station; Cluster; Data Mining; DBSCAN; Pre-Processing

1. PENDAHULUAN

Perkembangan teknologi telekomunikasi seluler kini telah mencapai tahap yang lebih maju dengan hadirnya teknologi 4G LTE (*Long Term Evolution*). Teknologi ini menawarkan kemampuan transfer data dan kecepatan akses yang jauh lebih tinggi dibandingkan dengan teknologi sebelumnya seperti 3G dan 2G[1]. Hal ini membuat 4G LTE menjadi tulang punggung bagi layanan komunikasi modern yang mendukung kebutuhan pengguna akan akses internet yang cepat dan stabil[2]. Di dalam jaringan 4G LTE, terdapat BTS sebagai infrastruktur yang berperan untuk menjembatani komunikasi antara perangkat seluler dan jaringan operator. *Base Transceiver Station* (BTS) adalah komponen penting dalam infrastruktur jaringan seluler yang bertanggung jawab untuk mentransmisikan dan menerima sinyal radio dari dan ke perangkat seluler. Dalam konteks teknologi navigasi, BTS juga dapat dimanfaatkan untuk pengukuran posisi dan navigasi memanfaatkan sinyal yang tersedia dari jaringan seluler untuk meningkatkan akurasi sistem navigasi[3] BTS ini berperan penting dalam memastikan pengguna dapat terhubung secara nirkabel ke jaringan operator[4].

Namun, meskipun teknologi 4G LTE telah diterapkan secara luas, masih terdapat tantangan yang perlu diatasi untuk memastikan kualitas jaringan tetap optimal. Kualitas jaringan internet di Daerah Istimewa Yogyakarta (DIY), misalnya, telah mengalami peningkatan yang signifikan dalam beberapa tahun terakhir, tetapi kendala seperti terbatasnya infrastruktur masih dirasakan, terutama di area pedesaan dan wilayah terluar[5]. Hal ini menunjukkan bahwa meskipun BTS telah berfungsi sebagai tulang punggung jaringan, penambahan infrastruktur dan optimasi jaringan tetap diperlukan untuk mendukung peningkatan kualitas layanan, terutama di daerah-daerah dengan lalu lintas data tinggi.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan analisis yang dapat membantu pemetaan area dengan kualitas jaringan. Salah satu metode yang efektif untuk menangani masalah ini adalah dengan cara analisis Data Mining menggunakan algoritma DBSCAN[6]. Metode ini digunakan untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik jaringan, seperti kualitas jaringan, tingkat kecepatan akses internet dan jumlah pengguna yang tinggi di suatu area[7]. Dengan mengelompokkan data-data tersebut, operator seperti Telkomsel dapat mengidentifikasi wilayah-wilayah dengan kualitas jaringan yang bermasalah[8].

Hasil dari penerapan algoritma DBSCAN ini akan menjadi dasar bagi pengambilan keputusan dalam hal perbaikan infrastruktur, seperti penambahan BTS (*Base Transceiver Station*) atau optimasi distribusi jaringan di wilayah-wilayah dengan beban pengguna yang tinggi[9]. Selain itu, hasil analisis ini juga dapat dijadikan acuan dalam menyusun strategi operasional dan pemasaran Telkomsel. Dengan memanfaatkan informasi mengenai kelompok wilayah dengan performa jaringan rendah, operator dapat mengalokasikan sumber daya secara lebih efektif untuk meningkatkan kualitas layanan dan mempertahankan loyalitas pelanggan.

Adapun beberapa penelitian sebelumnya yang telah memberikan dukungan bagi penelitian ini. Hasil penelitian pada tahun 2020 yang dilakukan oleh Teguh Iman Hermanto menunjukkan pentingnya pemilihan parameter *minpts* dan *epsilon* dalam proses pengolahan data menggunakan algoritma *Density Based Spatial Clustering of Application With Noise*. Dengan menentukan *minpts* dan *epsilon* dapat mempengaruhi proses kerja algoritma untuk menentukan jumlah cluster dan *noise* yang terbentuk. Hasil cluster menunjukkan perbedaan kebutuhan bandwidth pada masing-masing perangkat daerah sehingga dapat dilakukan pemerataan bandwidth yang lebih optimal[7].

Kemudian penelitian pada tahun 2019 yang dilakukan oleh Disi Amalia Pramudita mengenai analisis clustering, menunjukkan bahwa provinsi dalam cluster 1 memiliki infrastruktur internet yang kurang memadai, dengan rata-rata persentase pengguna internet menurut jenjang pendidikan terendah, yaitu 40,91%. Di sisi lain, provinsi dalam cluster 2 menunjukkan infrastruktur yang baik dengan rata-rata persentase pengguna internet berdasarkan jenjang pendidikan tertinggi, mencapai 72,27%. Sementara itu, provinsi dalam cluster 3 memiliki infrastruktur yang cukup baik, dengan rata-rata persentase pengguna internet yang cukup tinggi, yaitu 57,60%. Oleh karena itu, diperlukan upaya pemerataan infrastruktur internet agar dapat memenuhi kebutuhan 10 masyarakat di seluruh Indonesia[10].

Selanjutnya penelitian pada tahun 2020 yang dilakukan oleh Suhandio Handoko, menunjukkan bahwa penerapan metode Algoritma K-Means Clustering dalam data mining telah mengidentifikasi daerah-daerah dengan tingkat penjualan produk yang berbeda, yaitu tinggi, sedang, dan rendah. Untuk daerah yang menunjukkan penjualan produk rendah, akan dilakukan strategi promosi guna meningkatkan penjualan, sedangkan untuk daerah dengan penjualan tinggi, promosi tidak akan dilakukan, karena permintaan sudah mencukupi[11].

Kemudian penelitian pada tahun 2021 yang dilakukan oleh Andi Widya Mufila Gaffar, pada penelitian ini menghasilkan tiga kluster: kluster dengan kecepatan transfer rendah (ditandai warna biru), kluster dengan kecepatan transfer normal (warna kuning), dan kluster dengan kecepatan transfer tinggi (warna hijau). Penelitian menemukan bahwa kecepatan transfer data sangat dipengaruhi oleh variabel inbound dan outbound, dengan kecepatan terendah terjadi pada saat aktivitas kampus paling sibuk[12].

Selanjutnya penelitian pada tahun 2023 yang dilakukan oleh Made Mita Wikantari, menunjukkan bahwa penerapan metode DBSCAN Clustering dalam menentukan jenis konten dan motivasi pengguna suatu topik terbukti optimal. Hal ini terlihat dari *Silhouette Score* yaitu 0,29 sehingga menghasilkan total 3 cluster. Cluster yang terbentuk adalah Cluster -1, Cluster 0, dan Cluster 1. Setelah itu dilakukan analisis pada setiap cluster dengan melihat kata-kata yang terdapat pada cluster tersebut. Sehubungan dengan Cluster -1 berisi data Noise sehingga cluster tersebut dapat diabaikan karena data tersebut merupakan hasil dari data yang tidak masuk ke dalam cluster[13].

Kemudian penelitian pada tahun 2023 yang dilakukan oleh Rizki Mubarak, menunjukkan bahwa penerapan algoritma K-Means Clustering memberikan wawasan baru mengenai penggunaan bandwidth di Pemerintah Kabupaten Padang Pariaman. Dengan informasi ini, pihak administrasi jaringan di Dinas Komunikasi dan Informatika Kabupaten Padang Pariaman dapat mengoptimalkan alokasi bandwidth sesuai dengan tingkat penggunaan setiap Organisasi Perangkat Daerah (OPD) yang ada di wilayah tersebut. Dari data penggunaan bandwidth selama tahun 2022, pengelompokan menunjukkan bahwa terdapat enam OPD dengan tingkat pemakaian tinggi, 14 OPD dengan tingkat pemakaian sedang, dan 12 OPD dengan tingkat pemakaian rendah[14].

Selanjutnya penelitian pada tahun 2023 yang dilakukan oleh Rizal Ahmad Fauzi, menunjukkan bahwa setiap metode clustering memiliki kelebihan dan kelemahan masing-masing dalam konteks analisis data komputer. K-means cocok untuk data dengan cluster yang berbentuk bulat dan ukuran yang seimbang, Hierarchical Clustering efektif untuk data dengan struktur hierarkis atau cluster yang kompleks, sedangkan DBSCAN baik digunakan untuk data dengan noise atau data penciran serta cluster yang beragam. Pemilihan metode clustering harus didasarkan pada karakteristik data, tujuan analisis, dan ketersediaan sumber daya komputasi[15].

Dan penelitian pada tahun 2024 yang dilakukan oleh Gilang Mahardika, hasil penelitian ini menunjukkan bahwa DBSCAN mampu menghasilkan clustering yang efektif dalam mengelompokkan data sensor yang

kompleks, dengan nilai *silhouette score* rata-rata sebesar 0.7 dan kemurnian sebesar 0.85. Kemampuan DBSCAN dalam menangani data dengan distribusi yang tidak merata, bentuk cluster yang tidak teratur, serta noise dan outlier, telah terbukti melalui evaluasi kinerja yang dilakukan. Hasil clustering yang dihasilkan juga relatif bersih dari data yang tidak relevan, menunjukkan keunggulan algoritma DBSCAN dalam mendeteksi dan mengelompokkan pola dalam data sensor IoT[16].

Penerapan metode ini diharapkan dapat mengoptimalkan perbaikan jaringan internet di DIY, khususnya dalam memastikan bahwa kualitas layanan dapat merata di semua wilayah, serta mendukung strategi pengembangan bisnis Telkomsel di masa mendatang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk menghasilkan cluster dari proses analisis data mining menggunakan algoritma DBSCAN memerlukan beberapa tahapan. Adapun tahapan yang dilakukan bertujuan untuk menganalisis dan mengelompokkan area berdasarkan kualitas jaringan. Tahapan tersebut terdiri dari identifikasi masalah, menganalisa masalah, studi literatur, mengumpulkan data, melakukan pre-processing serta transformasi data, menguji hasil dan kemudian kesimpulan[17]. Adapun diagram alir dari jalannya penelitian ini dapat dilihat pada gambar berikut:



Gambar 1. Tahapan Penelitian

Proses dimulai dengan mengidentifikasi masalah untuk memahami ruang lingkup dan konteks tantangan yang dihadapi kemudian dilanjutkan dengan menganalisa masalah secara mendalam guna menemukan penyebab utama dan dampaknya. Tahap berikutnya adalah mempelajari literatur, yaitu meninjau penelitian atau referensi sebelumnya yang relevan untuk memperoleh wawasan dan metode yang mendukung. Setelah itu, dilakukan pengumpulan data, baik dari sumber primer maupun sekunder, yang kemudian melalui tahap pre-processing dan transformasi data untuk membersihkan, mempersiapkan, dan menyesuaikannya agar siap digunakan dalam analisis. Selanjutnya, hasil model atau analisis diketahui melalui pengujian, memastikan validitas dan efektivitas pendekatan yang digunakan. Akhirnya, proses diakhiri dengan menyimpulkan hasil, di mana temuan dan rekomendasi disampaikan sebagai solusi untuk menyelesaikan masalah yang telah diidentifikasi.

2.2 Dataset

Data yang digunakan pada penelitian ini adalah data mengenai sebaran site dengan yang meliputi Site ID, Nilai RSRP, Unduh, Unggah dengan total data berjumlah 154 baris data. Tabel berikut adalah gambaran umum dari bentuk data yang akan digunakan pada penelitian kali[18].

Table 1. Sample Data

SITE ID	LATITUDE	LONGITUDE	NILAI RSRP	UNDUH	UNGGAH	KECEPATAN INTERNET
BTL007	-7,82809	110,4388	-77	21,9	30,8	26,35
BTL057	-7,805129	110,4225	-87	8,36	3,81	6,085
KLT055	-7,6491	110,4707	-98	32	7,9	19,95
SMN001	-7,71501	110,3561	-102	9,58	11	10,29
SMN002	-7,73252	110,3719	-87	0,89	24,5	12,695
SMN006	-7,74591	110,461	-70	39,8	32,3	36,05
SMN007	-7,77156	110,4685	-89	36,4	23,2	29,8
SMN008	-7,67919	110,3403	-87	24,8	14,6	19,7

2.3 Data Pre-Processing

Pada tahap pre-processing pengolahan data akan melibatkan beberapa langkah, salah satunya adalah data *transformation* yang akan diterapkan pada penelitian ini. Langkah yang dilakukan merupakan proses penting sebelum data diolah lebih lanjut dalam proses data mining. Pada tahap ini, data mentah akan diproses supaya lebih rapi dan mudah dipahami oleh sistem[19][20].

2.3.1 Data Transformation

Pada tahap transformasi data, normalisasi data dilakukan dengan menggunakan *StandardScaler* untuk mengubah data menjadi distribusi standar (mean = 0 dan standar deviasi = 1. Tujuannya adalah supaya data lebih teratur, seragam, dan gampang dipahami oleh sistem.

2.4 Data Mining

Data Mining adalah proses menganalisis dan mengekstrak informasi atau pola yang berguna dari kumpulan data yang besar. Proses ini menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk menemukan hubungan, tren, atau wawasan yang tersembunyi dalam data yang tidak dapat terlihat secara langsung oleh manusia[21]. Data mining sering digunakan untuk mendukung pengambilan keputusan dan prediksi dalam berbagai bidang, seperti bisnis, kesehatan, pendidikan, hingga keuangan[22].

2.4.1 Clustering

Dalam jurnal oleh K. P. Sinaga dan M.-S. Yang, clustering dijelaskan sebagai metode pembelajaran tanpa pengawasan yang bertujuan untuk mengelompokkan data yang tidak berlabel ke dalam kelompok yang *homogen* berdasarkan kesamaan karakteristik tertentu[23]. Algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah metode clustering dalam data mining yang digunakan untuk mengidentifikasi kelompok atau cluster berdasarkan kepadatan data di ruang multidimensi. Metode ini sangat populer karena kemampuannya untuk menemukan cluster dengan bentuk apa pun dan mengatasi data dengan noise atau outliers yang tidak dapat dikelompokkan.

a. *Density-Based Spatial Clustering of Application with Noise* (DBSCAN)

DBSCAN adalah algoritma clustering berbasis densitas yang digunakan untuk mengelompokkan data berdasarkan kedekatan atau kepadatan sampel dalam ruang data. DBSCAN dapat mengenali cluster dengan bentuk arbitrer dan mampu menangani data yang mengandung noise (*outlier*), menjadikannya salah satu algoritma yang populer untuk analisis data yang kompleks. DBSCAN bekerja dengan mendefinisikan lingkungan sekitar sebuah titik menggunakan radius ϵ (*epsilon*), dimana lingkungan ini disebut sebagai *neighborhood*.

Neighborhood suatu titik p adalah himpunan semua titik q dalam dataset yang jaraknya dari p tidak lebih dari ϵ . jarak antar titik biasanya dihitung menggunakan *Euclidean Distance*, dengan rumus

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \quad (1)$$

Di mana x_{ia} dan x_{ja} adalah nilai atribut ke- a untuk titik i dan j , serta p adalah jumlah dimensi data. Dengan cara ini, *neighborhood* suatu titik p didefinisikan sebagai

$$N_\epsilon(p) = \{q \in D \mid \text{distance}(p, q) \leq \epsilon\} \quad (2)$$

3. HASIL DAN PEMBAHASAN

Dataset yang telah terkumpul melalui proses sampling ini menggunakan data dari berbagai lokasi site dengan informasi variabel seperti nilai unduh (download), unggah (upload), dan sinyal RSRP. Dataset ini terdiri dari total 154 data, yang selanjutnya akan masuk ke tahap pre-processing untuk membersihkan data. Proses pre-processing ini bertujuan untuk memastikan data yang digunakan dalam algoritma clustering dapat menghasilkan

cluster yang lebih optimal. Hasil dari proses pre-processing divisualisasikan dalam bentuk gambar menggunakan library seperti *Scatter Plot* untuk memberikan gambaran mengenai distribusi cluster dalam dataset tersebut dan akan divisualisasikan dalam bentuk map menggunakan pustaka *Folium*. Hasil dari tahap pre-processing dapat dilihat pada Tabel 2 berikut.

Tabel 2. Hasil Data Transformation

UNDUH_SCALED	UNGGAH_SCALED	NILAI_RSRP_SCALED
-0.182545	-0.957755	-1.304630
-1.350693	0.416025	-0.966597
-1.775474	-0.843273	-0.713073
-0.182545	-1.358441	0.470042
1.622775	0.873951	1.061600

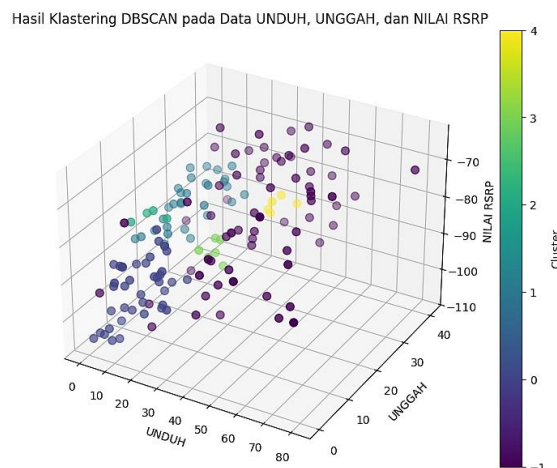
Berdasarkan Tabel 2, hasil Data Transformation di atas, proses pre-processing data dilakukan dengan metode normalisasi atau standarisasi pada tiga variabel utama, yaitu UNDUH, UNGGAH, dan NILAI RSRP. Variabel-variabel ini diubah menjadi skala yang terstandar, ditunjukkan dengan kolom UNDUH_SCALED, UNGGAH_SCALED, dan NILAI_RSRP_SCALED. Hasil transformasi menunjukkan nilai-nilai yang telah dinormalisasi dengan distribusi rata-rata mendekati nol, mencakup angka negatif dan positif. Contohnya, salah satu nilai pada variabel UNDUH_SCALED memiliki nilai -1.775474, sedangkan pada variabel UNGGAH_SCALED terdapat nilai 0.873951. Proses ini bertujuan untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam analisis lebih lanjut, seperti pada algoritma clustering yang sensitif terhadap skala data.

3.1 Implementasi DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering yang bekerja dengan membentuk area berdasarkan kepadatan data yang saling terhubung. Keunggulan utama algoritma ini dibandingkan metode lain adalah kemampuannya untuk mendeteksi outlier atau noise, yaitu data yang berada jauh dari kumpulan data lainnya. Berbeda dengan K-Means atau K-Medoids, DBSCAN tidak memerlukan jumlah cluster yang harus ditentukan sebelumnya, sehingga algoritma ini mampu mengenali pola cluster dengan bentuk yang tidak beraturan.

Dua parameter utama yang digunakan dalam DBSCAN adalah *MinPts* dan *Epsilon* (ϵ). Parameter *MinPts* menentukan jumlah data minimum yang diperlukan untuk memulai sebuah cluster, sedangkan *Epsilon* mengatur jarak maksimum antar data agar dianggap sebagai tetangga. DBSCAN menggunakan metode *Euclidean Distance* untuk menghitung jarak antar data guna mengelompokkan data berdasarkan pola kepadatannya. Pada proses inialisasi model, parameter *MinPts* diatur sebesar 5 dan *Epsilon* sebesar 0,5. Setelah parameter ditetapkan, dilakukan proses *fit-predict* untuk melakukan clustering data. Algoritma ini kemudian memberikan label kepada setiap data, di mana data dengan label -1 dianggap sebagai *noise* dan tidak dihitung sebagai bagian dari cluster.

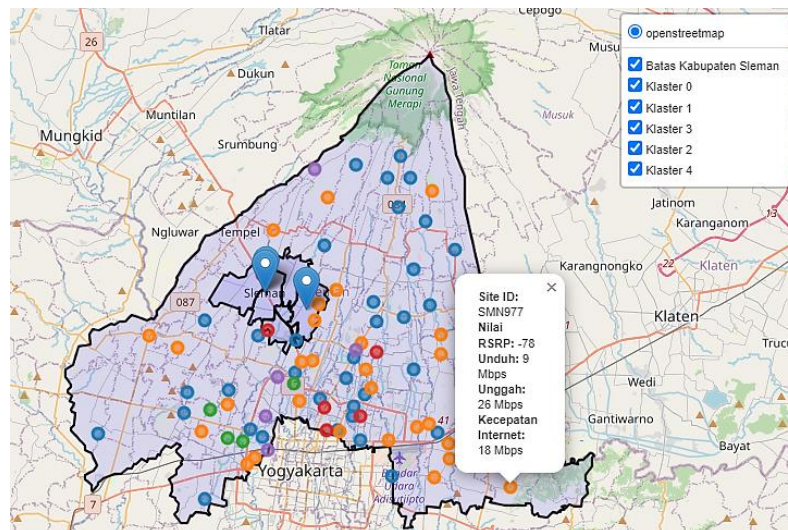
Setelah clustering selesai, dilakukan perhitungan jumlah cluster yang dihasilkan dengan mengabaikan *noise*. Hasilnya, algoritma DBSCAN berhasil mengidentifikasi 5 cluster. Evaluasi kualitas cluster dilakukan menggunakan metrik *Silhouette Score*, yang menunjukkan nilai sebesar 0.216471397367446, mengindikasikan bahwa kualitas clustering tergolong rendah, yang kemungkinan disebabkan oleh distribusi data atau parameter yang kurang optimal. Hasil clustering ini kemudian divisualisasikan menggunakan dua pendekatan utama, yaitu *Scatter Plot* dan Map, untuk memberikan gambaran yang lebih jelas tentang pola cluster yang terbentuk. Untuk hasil berdasarkan *Scatter Plot* dapat dilihat pada Gambar 2 berikut.



Gambar 2. Visualisasi Cluster Menggunakan Scatter Plot

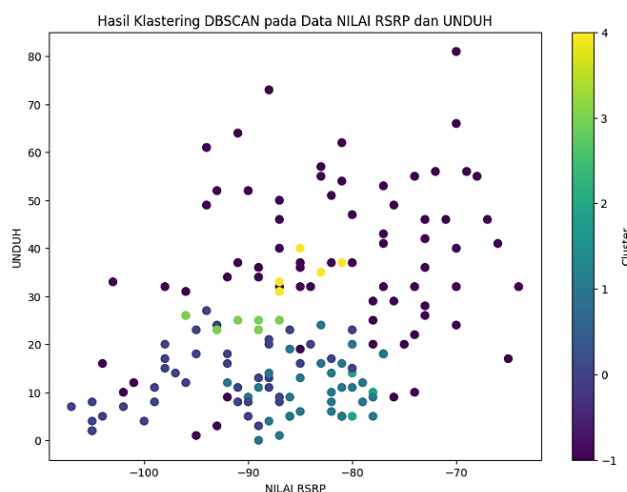
Untuk *Scatter Plot*, data divisualisasikan dalam grafik 3D dengan tiga variabel utama: UNDUH, UNGGAH, dan NILAI RSRP. Sumbu X, Y, dan Z masing-masing mewakili ketiga variabel tersebut, sementara setiap titik pada grafik merepresentasikan satu data. Warna pada titik menunjukkan cluster yang berbeda, di mana warna yang lebih terang merefleksikan NILAI RSRP yang lebih tinggi. Visualisasi ini juga mengungkapkan keberadaan beberapa titik yang tampak terisolasi atau dianggap sebagai noise, yang diberi warna khusus berbeda dari cluster lainnya.

Sementara itu, visualisasi dalam bentuk Map memberikan representasi spasial hasil clustering data site di Kabupaten Sleman, Yogyakarta. Setiap titik pada peta mewakili lokasi site, dengan warna berbeda untuk setiap cluster. Peta ini dilengkapi dengan fitur interaktif yang memungkinkan pengguna untuk menampilkan atau menyembunyikan batas wilayah Kabupaten Sleman serta cluster-cluster yang berbeda. Selain itu, saat salah satu titik pada peta diklik, akan muncul pop-up informasi detail seperti Site ID, NILAI RSRP, UNDUH, UNGGAH, dan Kecepatan Internet. Kombinasi visualisasi ini memberikan wawasan mendalam tentang hasil clustering, baik dari pola distribusi variabel maupun konteks geografisnya, mempermudah interpretasi pola cluster berdasarkan lokasi. Hasil visualisasi cluster ke bentuk map dapat dilihat pada Gambar 3 berikut.



Gambar 3. Visualisasi Hasil Cluster ke bentuk Map

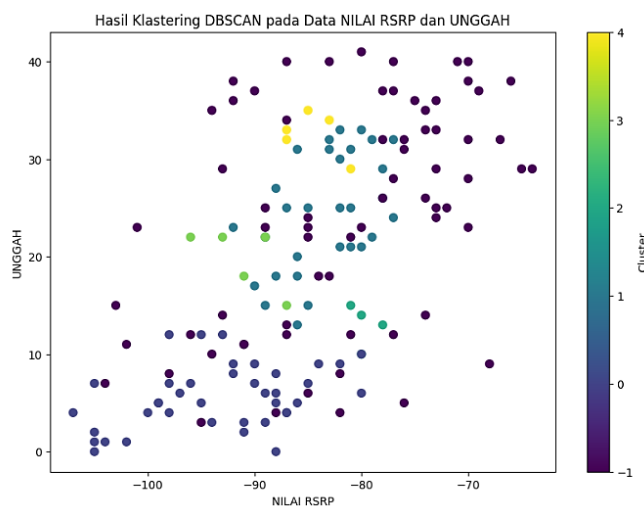
Secara detail, hasil klasterisasi menggunakan algoritma DBSCAN dapat ditampilkan untuk berbagai kombinasi variabel, yaitu antara NILAI RSRP dan UNDUH, NILAI RSRP dan UNGGAH, serta UNDUH dan UNGGAH. Visualisasi ini akan memberikan gambaran tentang pola distribusi data pada setiap pasangan variabel, menunjukkan cluster yang terbentuk, dan mengidentifikasi data yang dianggap sebagai *noise*. Analisis ini bertujuan untuk memahami hubungan antar variabel serta membantu dalam pengambilan keputusan berdasarkan pola yang terdeteksi.



Gambar 4. Hasil Visualisasi Scatter Plot Nilai RSRP dan UNDUH

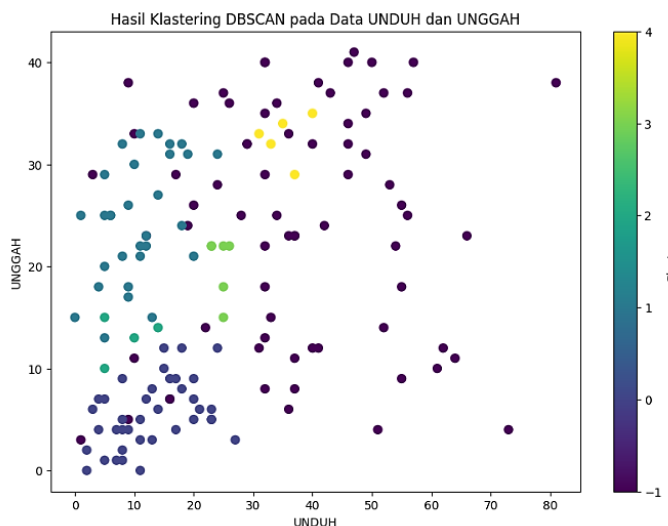
Gambar 4 diatas menunjukkan hasil visualisasi yang memperlihatkan pola hubungan antara NILAI RSRP (kekuatan sinyal) dan kecepatan UNDUH. Sebagian besar data terkelompok ke dalam beberapa cluster utama dengan distribusi yang menunjukkan adanya hubungan antara kekuatan sinyal dan kecepatan unduh. Namun,

terdapat beberapa data yang tidak sesuai dengan pola densitas sehingga dikategorikan sebagai *noise*. Analisis ini menunjukkan kemampuan DBSCAN dalam memisahkan kelompok data padat dan mendeteksi outlier, memberikan informasi berharga mengenai pengaruh kualitas sinyal terhadap kecepatan unduh.



Gambar 5. Hasil Visualisasi Scatter Plot data NILAI RSRP dan UNGGAH

Dari hasil visualisasi pada gambar 5, terlihat pola hubungan antara NILAI RSRP (kekuatan sinyal) dan kecepatan UNGGAH, di mana sebagian besar data berhasil dikelompokkan ke dalam beberapa cluster utama, sementara sebagian kecil data terdeteksi sebagai *noise*. Visualisasi ini dapat memberikan wawasan yang berguna untuk menganalisis pengaruh kekuatan sinyal terhadap kecepatan unggah.



Gambar 6. Visualisasi Scatter Plot data UNDUH dan UNGGAH

Gambar 6 di atas menunjukkan pola hubungan antara kecepatan UNDUH dan kecepatan UNGGAH. Sebagian besar data terkelompok ke dalam cluster utama, yang menunjukkan adanya keterkaitan antara kedua variabel, sementara sebagian kecil data diklasifikasikan sebagai *noise* karena posisinya yang tidak sejalan dengan kepadatan cluster lainnya. Visualisasi ini memberikan wawasan mengenai pola distribusi data pada variabel unduh dan unggah, serta membantu mengidentifikasi anomali yang dapat memengaruhi analisis lebih lanjut.

3.2 Pembahasan

Pada penelitian ini, digunakan analisis Data Mining dengan menerapkan algoritma DBSCAN untuk clustering menggunakan bahasa Python dan kode editor Visual Studio Code. Setelah dataset berhasil dikumpulkan, dilakukan tahap pre-processing data, termasuk normalisasi data, untuk memastikan skala variabel seragam. Selanjutnya, library DBSCAN dari Scikit-Learn diimpor, dan dataset yang sudah dinormalisasi diolah dengan menginisialisasi model DBSCAN menggunakan parameter MinPts = 5 dan Epsilon = 0.5. Berdasarkan pengujian terhadap 154 data, algoritma DBSCAN berhasil menghasilkan 5 cluster dengan kriteria masing-masing yaitu cluster 1 = 0 dengan rentang Nilai RSRP berkisar antara -107 hingga -80 dBm, variasi nilai UNDUH antara 3

hingga 24 Mbps, UNGGAH berada dalam rentang nilai 2 hingga 18 Mbps. Cluster 2 = 1 dengan rentang Nilai RSRP berkisar antara -87 hingga -85 dBm, variasi nilai UNDUH cukup tinggi, yaitu antara 20 hingga 30 Mbps mengindikasikan performa unduhan yang baik, nilai UNGGAH berada dalam rentang 10 hingga 15 Mbps juga menunjukkan performa unggahan yang baik. Cluster 3 = 2 dengan rentang Nilai RSRP berkisar antara -81 hingga -78 dBm, yang menunjukkan sinyal yang lebih kuat dibandingkan Cluster 0 dan 1, variasi nilai UNDUH bervariasi antara 5 hingga 14 Mbps, UNGGAH berada dalam rentang nilai 10 hingga 15 Mbps, menunjukkan performa unggahan yang cukup baik dan stabil. Cluster 4 = 3 dengan rentang Nilai RSRP berkisar antara -96 hingga -87 dBm, variasi nilai UNDUH antara 23 hingga 25 Mbps, nilai UNGGAH berada dalam rentang 15 hingga 22 Mbps. Dan Cluster 5 = 4 dengan rentang Nilai RSRP berkisar antara -87 hingga -81 dBm, variasi nilai UNDUH antara 31 hingga 40 Mbps, menunjukkan performa unduhan yang baik dan stabil, dengan kecepatan yang cukup tinggi pada seluruh data dalam cluster ini, dan UNGGAH berada dalam rentang nilai 29 hingga 35 Mbps, yang mencerminkan performa unggah yang cukup stabil dan baik. Secara keseluruhan, Cluster 5 memiliki jaringan yang stabil dengan performa unduhan dan unggahan yang baik, mencerminkan kualitas layanan yang optimal dalam hal kecepatan internet.

4. KESIMPULAN

Hasil analisis menunjukkan bahwa algoritma DBSCAN sangat cocok untuk clustering berbasis kepadatan, terutama pada dataset dengan pola distribusi yang tidak beraturan dan keberadaan outlier. Dibandingkan dengan penelitian terdahulu, DBSCAN terbukti mampu memberikan hasil clustering yang relevan dan bermanfaat dalam konteks data kepadatan seperti sebaran site. Dalam penelitian ini, DBSCAN berhasil mengelompokkan data menjadi 5 cluster dengan nilai *silhouette* 0.216471397367446, yang mengindikasikan bahwa kualitas clustering tergolong rendah, yang kemungkinan disebabkan oleh distribusi data atau parameter yang kurang optimal. Meskipun demikian, hasil clustering yang diperoleh tetap memberikan wawasan yang berguna untuk analisis sebaran site dan performa jaringan. Dengan hasil ini, penulis berharap bahwa Telkomsel dapat menggunakan informasi yang diperoleh dari analisis ini untuk mengidentifikasi wilayah dengan kepadatan site yang tinggi atau rendah. Hal ini dapat membantu perusahaan dalam pengambilan keputusan strategis, seperti optimasi jaringan dan peningkatan kualitas layanan bagi pelanggan di Kabupaten Sleman. Untuk peneliti yang ingin melanjutkan penelitian ini, disarankan untuk membandingkan hasil DBSCAN dengan algoritma clustering lain seperti K-Means atau HDBSCAN untuk melihat apakah ada perbedaan signifikan dalam hasil clustering. Selain itu, peneliti juga disarankan untuk menambahkan variabel tambahan, seperti jumlah pengguna aktif atau rata-rata throughput, guna memperkaya analisis dan memberikan pemahaman yang lebih dalam tentang faktor-faktor yang memengaruhi kualitas jaringan. Terakhir, perluasan cakupan wilayah ke area lain dengan karakteristik geografis yang berbeda juga dapat membantu menghasilkan temuan yang lebih komprehensif dan meningkatkan validitas hasil analisis ini.

REFERENCES

- [1] B. Hardiyanto, "Analisis Quality of Service (QoS) Jaringan 4G LTE Melalui Drive Test di BBPLK Bekasi Menggunakan Aplikasi Netmonitor Cell Signal Logging," *J. Eng. Edu*, vol. 6, no. 2, pp. 1–8, 2020.
- [2] C. I. Lendo and D. Chandra, "Analisis Kualitas Jaringan Internet Mahasiswa Dan Niat Perkuliahan Berbasis Daring Mahasiswa Universitas Kristen Satya Wacana," *J. TEKINKOM*, vol. 6, no. 2, 2023, doi: 10.37600/tekinkom.v6i2.1012.
- [3] Z. (Zak) M. Kassas, "Navigation with Cellular Signals of Opportunity," *Position, Navig. Timing Technol. 21st Century*, vol. 2, pp. 1171–1223, 2020, doi: 10.1002/9781119458555.ch38.
- [4] A. Yuda, "Komparasi Algoritma K-Means dan Hierarchical Clustering Untuk Mengetahui Data Customer Dalam Layanan Internet," *Inform. Sains Teknol.*, vol. 2, no. 2, pp. 9–14, 2024, doi: 10.34005/insit.v2i2.4119.
- [5] A. Y. Kodoatie and H. Sama, "Studi Analisis Kecepatan Internet Telkomsel Di Kota Batam: Studi Geografis," *Vol. 1 Nomor 1 Ed. Agustus 2020 Conference Business, Soc. Sci. Innov. Technol.*, vol. 1, pp. 498–507, 2020, [Online]. Available: <http://journal.uib.ac.id/index.php/cbssit>
- [6] K. E. Setiawan and A. Kurniawan, "Pengelompokan Rumah Sakit di Jakarta Menggunakan Model DBSCAN, Gaussian Mixture, dan Hierarchical Clustering," *J. Inform. Terpadu*, vol. 9, no. 2, pp. 149–156, 2023, doi: 10.54914/jit.v9i2.995.
- [7] T. I. Hermanto and Y. Muhyidin, "Analisis Data Sebaran Bandwidth Menggunakan Algoritma Dbscan Untuk Menentukan Tingkat Kebutuhan Bandwidth Di Kabupaten Purwakarta," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 5, no. 2, pp. 130–137, 2020, doi: 10.36341/rabit.v5i2.1388.
- [8] Telkomsel, "Telkomsel perluas jaringan 4G di wilayah DIY," 2023. <https://www.telkomsel.com/jaringan-perluasan-diy>
- [9] A. Kristianto, "Implementasi DBSCAN dalam Clustering Data Minat Mahasiswa Setelah Pandemi Covid19," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 426–431, 2022, doi: 10.24002/konstelasi.v2i2.5638.

- [10] D. A. Pramudita and Bagus Sumargo, "Pengelompokan Pengguna Internet dengan Metode K-Means Clustering," *J. Stat. dan Apl.*, vol. 3, no. 1, pp. 1–12, 2019, doi: 10.21009/jsa.03101.
- [11] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [12] A. W. M. Gaffar, "Pengklasifikasian Kecepatan Transfer Data Pada Jaringan Backbone Menggunakan K-Means," *Bul. Sist. Inf. dan Teknol. Islam*, vol. 2, no. 2, pp. 126–130, 2021, doi: 10.33096/busiti.v2i2.982.
- [13] M. M. Wikantari, Y. Sibaroni, and A. F. Ihsan, "Clustering Content Types and User Motivation Using DBSCAN on Twitter," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 741–748, 2023, doi: 10.47065/josyc.v4i4.3750.
- [14] R. Mubarak, S. Defit, and G. W. Nurcahyo, "Algoritma K-Means Clustering Penggunaan Bandwidth Internet (Studi Kasus di Pemerintah Daerah Kabupaten Padang Pariaman)," *Explor. J. Sist. Inf. dan Telemat.*, vol. 14, no. 1, p. 37, 2023, doi: 10.36448/jsit.v14i1.3037.
- [15] R. Ahmad and F. Informatika, "Analisis Perbandingan Antara Metode Clustering Untuk Pengelompokan Data Dalam Analisis Data Komputer," *Teknologipintar.org*, vol. 3, no. 11, pp. 2023–2024, 2023.
- [16] G. Mahardika, "Clustering Data Sensor IOT Dengan Algoritma DBSCAN," *J. Dunia Data*, vol. 1, no. 5, p. 16, 2024.
- [17] M. A. Zakariah, V. Afriani, and M. Zakariah, *METODOLOGI PENELITIAN KUALITATIF, KUANTITATIF, ACTION RESEARCH, RESEARCH AND DEVELOPMENT (R n D)*. Kolaka: Yayasan Pondok Pesantren Al Mawaddah Warrahmah Kolaka, 2020.
- [18] Kabar Harian, "Pengertian Dataset dan Jenis-jenisnya," *Kumparan*, 2021. <https://kumparan.com/kabar-harian/pengertian-dataset-dan-jenis-jenisnya-1wtM6xNlqpQ> (accessed Nov. 30, 2024).
- [19] Suripto, Rr nurul rahmanita, and ajeng sekar kirana, "Teknik pre-processing dan classification dalam data science," *binus*, 2022. <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science> (accessed Nov. 30, 2024).
- [20] I. M. Hamdani¹ et al., "Edukasi dan Pelatihan Data Science dan Data Preprocessing," *Juni*, vol. 2, no. 1, pp. 19–26, 2024, doi: 10.58227/intisari.v2i1.125.
- [21] muhammad arhami and muhammad nasir, *Data Mining - Algoritma dan Implementasi*. Andi, 2020.
- [22] Carudin et al., *Buku Ajar Data Mining*, Cetakan Pe. Kota Jambi: PT. Sonpedia Publishing Indonesia, 2024.
- [23] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.