

Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet Dengan Metode Centroid Linkage

Abdullah Ahmad Dzikrullah^{1, a)}

¹Jurusan Statistika FMIPA Universitas Islam Indonesia

^{a)} adzikrullah@uii.ac.id

Abstrak. Pasca pandemic saat ini, internet telah menunjukkan kekuatannya sebagai media akselerasi digital dan transformasi ekonomi. Internet berkualitas mutlak menjadi kebutuhan penting dalam menjalankan berbagai aktivitas. Pada kenyataannya, kualitas internet di Indonesia belum merata dan hanya beberapa daerah tertentu di Pulau Jawa dan Sumatera yang memiliki kualitas baik. Penelitian ini bertujuan untuk mengelompokkan 34 provinsi di Indonesia berdasarkan lima indikator kualitas jaringan internet. Metode kluster dapat memeringkatkan wilayah berdasarkan indikator kualitas internet. Metode Centroid Linkage adalah metode pengelompokan objek yang memiliki keunggulan pada penanganan data pencilan. Penelitian ini menghasilkan empat kelompok dengan kualitas internet yang rendah berada pada Provinsi Papua dan Papua Barat. Kontur dan topografi wilayah menjadi kendala pemerintah dalam proses pemerataan internet berkualitas.

Kata Kunci: analisis kluster, Centroid Linkage, pencilan.

Abstract. After the current pandemic, the internet has shown its power as a medium for digital acceleration and economic transformation. Internet quality is absolutely an essential requirement in carrying out various activities. Indonesia has unequal internet quality, and only specific areas in Java and Sumatra have good quality. This study aims to classify 34 provinces in Indonesia based on five indicators of internet network quality. The cluster method can rank regions based on internet quality indicators. The centroid Linkage method is an object grouping method that has advantages in handling outlier data. This study resulted in four groups with low internet quality located in Papua and West Papua provinces. The contours and topography of the area are obstacles for the government to distribute quality internet.

Keywords: cluster analysis, Linkage Centroid, Pencilan.

PENDAHULUAN

Di era pemulihan pandemi saat ini, internet menunjukkan keperkasaannya sebagai media akselerasi digital dan transformasi ekonomi. Kebutuhan internet berkualitas menjadi urgensi dalam menjalankan berbagai aktivitas mulai dari kebutuhan dasar dalam berkomunikasi hingga kebutuhan untuk menjalankan kegiatan ekonomi. Internet yang berkualitas mutlak menjadi prioritas utama sebagai jembatan kemajuan di berbagai bidang. Parameter kualitas internet tidak jauh dari tingkat kecepatan internet dalam mengunggah (*upload*) dan mengunduh (*download*) (G Plate, 2020).

Kecepatan internet *mobile broadband* di Indonesia menempati peringkat 120 dunia berdasarkan Data Speedtest Global Index bulan Januari 2022 lalu. Kecepatan ini cenderung menurun dan hanya berkisar pada 14,16 Mbps untuk kecepatan *download* dan rata-rata berkisar 9,50 Mbps untuk kecepatan *upload*. Kecepatan tersebut berada di bawah kecepatan akses rata-rata dunia yaitu kecepatan *download* 31,95 Mbps dan *upload* 11,32 Mbps. Sedangkan untuk kecepatan *fixed broadband*, Indonesia menduduki peringkat 115 dengan kecepatan *download* 20,60 Mbps dan *upload* 12,53 Mbps, masih jauh di bawah rata-rata kecepatan *fixed broadband* dunia yakni sebesar 74,32 Mbps untuk *download* dan *upload* 40,83 Mbps. Kecepatan internet di negeri ini

sesungguhnya tidak kalah dengan operator dunia, hanya saja kecepatan optimal hanya bisa dinikmati di wilayah ibukota provinsi dan beberapa kota besar di Indonesia (G Plate, 2020).

Berdasarkan data Kemenkominfo Tahun 2019, jaringan kabel serat optik nasional telah mencapai 342.239 km, dengan 224.453 merupakan *inland*, dan 117.786 merupakan Sistem Komunikasi Kabel Laut (SKKL). Pada capaian sinyal 4G serta sebaran serat optik, keduanya masih terkonsentrasi di Pulau Jawa dan Sumatera dengan proporsi wilayah tengah dan timur Indonesia menjadi wilayah cakupan layanan telekomunikasi paling rendah (G Plate, 2020). Hal ini sangat kontras dengan infrastruktur di ibukota provinsi khususnya kota-kota besar di Indonesia. Meskipun kualitas layanan internet di kota besar sudah menunjukkan kualitas yang baik, perlu adanya keseragaman kualitas jaringan internet agar proses pemerataan infrastruktur yang digaungkan oleh negara cepat terealisasi. Ketidakmerataan kualitas ini dipengaruhi oleh banyak faktor seperti ketersediaan fiber optik, jumlah *coverage Base Transceiver Station* (BTS) setiap daerah, kondisi wilayah, jumlah pengguna aktif internet, dan keahlian operator dalam pengelolaan jaringan (Suharni, 2021).

Berdasarkan permasalahan tersebut, peneliti ingin mengelompokkan daerah berdasarkan kualitas jaringan internet dengan menggunakan teknik statistika multivariat. Terdapat banyak metode yang dapat dipakai dalam proses pengelompokan, salah satunya yaitu analisis kluster hirarki dengan metode *Centroid Linkage*. Metode hirarki merupakan metode pengelompokan yang terstruktur dan bertahap berdasarkan kemiripan sifat antar obyek. Kemiripan sifat tersebut dapat ditentukan dari kedekatan jarak *Euclidean*. *Centroid Linkage* menitikberatkan pada jarak antar nilai tengah observasi pada variabel dalam suatu set variabel kluster. Keuntungan metode ini adalah pencilan tidak berpengaruh secara signifikan terhadap hasil kluster yang terbentuk mengingat data yang diteliti memiliki ketimpangan yang besar sehingga kemungkinan *pencilan* sangat tinggi. Dengan melakukan analisis kluster ini, diharapkan dapat memberikan gambaran kelompok wilayah yang masih memerlukan perbaikan infrastruktur komunikasi di Indonesia (Silvi, 2018).

KAJIAN TEORI

Standardisasi Data

Proses standardisasi dilakukan jika setiap variabel penelitian terdapat perbedaan satuan pengukuran yang signifikan. Perbedaan tersebut dapat mengakibatkan analisis kluster tidak valid sehingga perlu proses standardisasi dengan cara transformasi data asli agar menjadi seragam satuannya (Silvi, 2018).

$$X_{stand} = \frac{X - \text{mean}(X)}{\text{standar deviasi}(X)} \quad (1)$$

Dengan :

X_{stand}	: nilai variabel yang sudah distandarisasi
X	: variabel sebelum distandarisasi
$\text{mean}(X)$	: nilai rata-rata variabel X
$\text{standar deviasi}(X)$	: nilai standar deviasi variabel X

Jarak Euclidean

Tujuan dalam analisis kluster adalah memilah sekumpulan objek menjadi beberapa kelompok berdasarkan ukuran kemiripan dilihat dari karakteristik variabel penyusunnya. Semakin dekat jarak suatu objek terhadap objek lain, maka semakin tinggi kemiripan objek tersebut sehingga dapat dimasukkan dalam kelompok yang sama. Ukuran yang digunakan pada penelitian ini untuk menentukan besaran jarak yaitu jarak *Euclidean*. Jarak *Euclidean* adalah akar dari total kuadrat selisih nilai antar dua objek dari setiap variabel faktor (Kassambara, 2017).

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

dengan :

$d_{euc}(x, y)$: Jarak euclid antar dua objek
 x_i : objek provinsi x dengan variabel ke- i
 y_i : objek provinsi y dengan variabel ke- i
 n : banyaknya variabel faktor
 i : urutan variabel faktor dari 1 hingga n

Hasil perhitungan jarak antar objek dari 34 provinsi akan membentuk matriks yang berisi jarak ke semua objek yang akan dianalisis

$$\begin{bmatrix} d_{1,1}(x, y) & \cdots & d_{1,34}(x, y) \\ \vdots & \ddots & \vdots \\ d_{34,1}(x, y) & \cdots & d_{34,34}(x, y) \end{bmatrix}$$

Uji Asumsi

Deteksi Pencilan

Analisis statistika pada umumnya diharapkan dapat mencerminkan gambaran umum dan bersifat representatif. Oleh sebab itu, pencilan-pencilan semestinya dihilangkan dari sampel agar hasil yang diperoleh tidak terlalu ekstrim. Dalam analisis klaster, ada beberapa metode yang dapat menangani data pencilan sehingga tidak berpengaruh signifikan terhadap hasil analisis. Metode *Centroid Linkage* merupakan salah satu metode yang mentoleransi pencilan masuk ke dalam proses analisis tanpa mempengaruhi hasil yang signifikan. Deteksi pencilan dalam data dapat dilihat secara jelas menggunakan boxplot (Kassambara, 2017).

Uji Multikolinearitas

Variabel dalam analisis klaster yang digunakan semestinya tidak berkorelasi atau sebaiknya tidak ada multikolinieritas. Uji ini penting dilakukan karena korelasi antar variabel berdampak pada penurunan akurasi pengukuran kemiripan antar objek. Hal ini terjadi karena pembobotan variabel tidak homogen dan cenderung saling mempengaruhi, sementara fokus pengukuran kemiripan hanya pada setiap objek yang diteliti. Korelasi antar variabel yang tinggi mengakibatkan pengelompokan objek hanya berdasarkan beberapa variabel saja sehingga tidak semua variabel signifikan berkontribusi dalam analisis klaster. Uji asumsi ini dapat dilihat dengan nilai korelasi antar variabel yang cenderung rendah.

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} \quad (3)$$

dengan:

r_{xy} : Koefisien korelasi antara variabel X dan Y .
 x : deviasi dari mean untuk nilai variabel X
 y : deviasi dari mean untuk nilai variabel Y
 $\sum xy$: jumlah perkalian antara nilai X dan Y
 x^2 : Kuadrat dari nilai x
 y^2 : Kuadrat dari nilai y

Sedangkan uji Bartlett untuk berfungsi untuk melihat independensi variabel secara keseluruhan (Arsham, 2011).

$$B = \frac{(N-k) \ln \left(\frac{\sum_{i=1}^k (n_i-1) s_i^2}{N-k} \right) - \sum_{i=1}^k (n_i-1) \ln(s_i^2)}{1 + \frac{1}{3(k-1)} \left[\left(\sum_{i=1}^k \frac{1}{n_i-1} \right) - \frac{1}{N-k} \right]} \quad (4)$$

Dengan :

B : Koefisien Bartlett

N : jumlah semua observasi

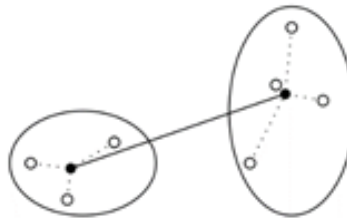
k : jumlah sampel

n_i : sampel pada observasi ke- i

s_i^2 : nilai variansi pada sampel observasi ke- i

Metode *Centroid Linkage*

Metode ini menitikberatkan pada nilai rata-rata semua obyek dalam kluster. Jarak yang digunakan dalam proses identifikasi kluster adalah kedekatan dengan titik *centroid* kluster yang terbentuk. Kluster *centroid* yang terbentuk nantinya adalah nilai tengah observasi pada variabel dalam suatu set variabel kluster. Dengan metode ini, setiap terjadi kluster baru segera terjadi perhitungan ulang *centroid* sampai terbentuk kluster yang cenderung konvergen. Keuntungan dari metode ini adalah tidak berpengaruh secara signifikan terhadap pencilaan dibandingkan dengan metode lain (Nielsen, 2016).



GAMBAR 1. Ilustrasi *Centroid Linkage*

$$[\{X_n\}, \{Y_n\}] = \left\| \left(\frac{1}{N} \sum_{n=1}^N X_n \right) - \left(\frac{1}{M} \sum_{m=1}^M Y_m \right) \right\| \quad (5)$$

Dengan :

N : jumlah semua observasi

X_n : Objek X ke- n

Y_n : Objek Y ke- n

Dendogram

Metode *Centroid Linkage* merupakan salah satu metode dalam kelompok kluster hirarki. Jenis kluster ini memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan selanjutnya. Demikian seterusnya hingga terbentuk hirarki seperti pohon atau sering disebut dendogram. Dendogram merupakan representasi visual dari proses terbentuknya kluster berdasarkan nilai koefisien jarak pada setiap langkah hingga terbentuk kluster akhir yaitu satu objek satu kluster.

Korelasi *Cophenetic*

Untuk mengetahui tingkat ketelitian metode kluster yang dipakai, peneliti dapat menggunakan koefisien korelasi *cophenetic* sebagai acuan. Nilai koefisien korelasi *cophenetic* yang tinggi mendekati 1 mengindikasikan solusi yang dihasilkan dari proses clustering sudah baik. Korelas

cophenetic merupakan alat untuk memvalidasi keamatan hasil kluster antara elemen matriks jarak Euclidean dengan elemen yang dihasilkan oleh matriks *cophenetic* (Widodo & Novita, 2018).

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{ik}^c - \bar{d}^c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{ik}^c - \bar{d}^c)^2]}} \quad (6)$$

Dengan :

r_{coph} : koefisien korelasi *cophenetic*

d_{ik} : jarak Euclidian antara objek-i dan objek-k

$\bar{d}$: rata-rata d_{ik}

d_{ik}^c : jarak *cophenetic* objek-i dan objek-k

$\bar{d}^c$: rata-rata d_{ik}^c

METODOLOGI PENELITIAN

Data Penelitian

Penelitian ini merupakan jenis analisis terapan berdasarkan kajian teori yang bertujuan untuk menyelesaikan permasalahan yang ada berdasarkan asumsi yang telah ditentukan. Objek penelitian ini adalah 34 provinsi yang ada di Indonesia dengan 5 variabel yang mempengaruhi kualitas internet. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh berdasarkan survei dan pendataan yang dilakukan oleh Badan Pusat Statistik (BPS).

TABEL 1. Variabel Penelitian

Nama Variabel	Keterangan
X_1	Luas wilayah provinsi (km ²)
X_2	Jumlah pulau dalam satu provinsi
X_3	Jumlah <i>Base Transceiver Station</i> (BTS)
X_4	proporsi individu pengguna internet (%)
X_5	Indeks keahlian mengelola jaringan

Alur Penelitian

Langkah-langkah Analisis Kluster *Centroid Linkage*

1. Membuat k kluster. Masing-masing individu atau unit observasi jadi kelompok. Kemudian dibuat matrik jarak antar objek
2. Menentukan jarak terkecil dari pasangan kluster,
3. Menggabungkan kluster u dan kluster v. kemudian update matriks jaraknya.
4. Ulangi langkah 2 & 3 sebanyak N-1 kali. Catat nilai jarak untuk setiap terjadi penggabungan kluster
5. Tentukan nilai cut off untuk menentukan kluster terbentuk. Lakukan dengan membuat dendogram, dan tentukan cut off jumlah kluster
6. Pemberian nama kluster berdasarkan profiling, yaitu melihat karakteristik kluster terbentuk secara rata-rata.

ANALISIS DAN PEMBAHASAN

Data

Data yang akan dianalisis kluster meliputi komponen objek 34 provinsi di Indonesia dan variabel sebanyak 5 komponen. Dalam analisis ini akan diteliti jumlah kelompok yang terbentuk

berdasarkan titik *centroid* yang terlihat dari dendogram nantinya. Berikut ringkasan data yang akan dianalisis:

TABEL 2. Data Penelitian

No	Provinsi	X1	X2	X3	X4	X5
1	Aceh	57956	363	1586	35.60	6.71
2	Bali	5780.06	34	557	54.08	6.26
3	Banten	9662.92	81	1085	56.25	5.85
4	Bengkulu	19919.33	9	439	40.72	6.29
5	DI Yogyakarta	3133.15	33	337	61.73	7.49
6	DKI Jakarta	664.01	113	239	73.46	6.53
7	Gorontalo	11257.07	127	293	41.78	5.78
8	Jambi	50058.16	14	727	42.68	5.8
9	Jawa Barat	35377.76	30	4476	53.94	5.57
10	Jawa Tengah	32800.69	71	4377	47.74	5.46
11	Jawa Timur	47803.49	403	4621	47.10	5.73
12	Kalimantan Barat	147307	249	922	38.38	5.31
13	Kalimantan Selatan	38744.23	158	827	50.37	5.51
14	Kalimantan Tengah	153564.5	69	542	46.73	5.65
15	Kalimantan Timur	129066.64	243	621	59.12	6.53
16	Kalimantan Utara	75467.7	196	210	54.30	6.1
17	Kep. Bangka Belitung	16424.06	507	339	45.85	5.19
18	Kep. Riau	8201.72	2025	299	65.02	6.08
19	Lampung	34623.8	172	1350	40.17	5.52
20	Maluku	46914.03	1337	527	33.89	6.89
21	Maluku Utara	31982.5	837	392	29.13	5.51
22	Nusa Tenggara Barat	18572.32	504	892	39.16	5.84
23	Nusa Tenggara Timur	48718.1	600	1031	26.29	5.65
24	Papua	319036.05	547	535	21.70	4.79
25	Papua Barat	102955.15	4514	396	43.46	6.41
26	Riau	87023.66	144	1220	44.97	6.15
27	Sulawesi Barat	16787.18	69	220	31.26	5.61
28	Sulawesi Selatan	46717.48	355	1616	43.91	6.17
29	Sulawesi Tengah	61841.29	1572	681	35.52	6.22
30	Sulawesi Tenggara	38067.7	590	612	41.92	6.43
31	Sulawesi Utara	13892.47	329	673	46.73	6.2
32	Sumatera Barat	42012.89	218	912	41.15	6.41
33	Sumatera Selatan	91592.43	24	1449	38.14	5.55
34	Sumatera Utara	72981.23	229	2567	41.38	6.24

Standardisasi Data

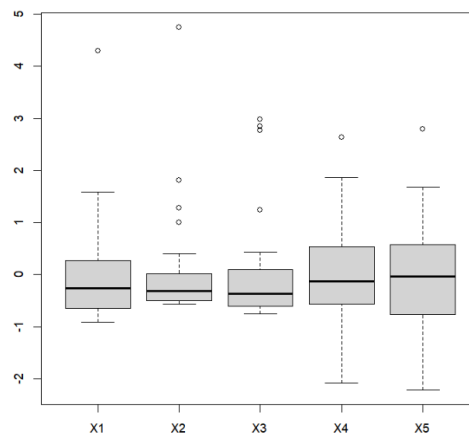
Sebelum melakukan analisis lebih lanjut terlebih dahulu dilakukan standardisasi data. Berdasarkan data awal terlihat bahwa setiap variabel memiliki satuan yang berbeda dan memiliki

selisih yang sangat tinggi apabila dibandingkan variabel satu dengan variabel yang lain. Sebagai contoh pada variabel X_1 yaitu luas wilayah sangat kontras dengan variabel X_5 yaitu indeks keahlian mengelola jaringan. Jika tidak dilakukan standardisasi, kemungkinan besar variabel X_1 cenderung akan mendominasi dalam proses pengelompokan nantinya. Proses standardisasi dilakukan menggunakan perangkat lunak R Versi 4.1.2.

Uji Asumsi

Deteksi Pencilan

Pencilan merupakan obeservasi data yang terdeteksi memiliki nilai berbeda dari sekumpulan observasi lainnya dari data tersebut. Pencilan ini termasuk observasi yang tidak biasa dan muncul pada salah satu titik ekstrim dari mayoritas data. Dampak dari outlier yang paling terasa adalah bergesernya ukuran pemusatan data khususnya nilai rata-rata. Dengan adanya pencilan, nilai rata-rata akan cenderung mendekati nilai pencilan sehingga proses analisis tidak terlalu akurat. Ukuran pemusatan data berupa median cenderung tidak terlalu berpengaruh terhadap pencilan karena median menitikberatkan pada nilai tengah dari jumlah data yang telah diurutkan. Pencilan dapat dilihat dengan jelas secara visual menggunakan boxplot sebagai berikut:



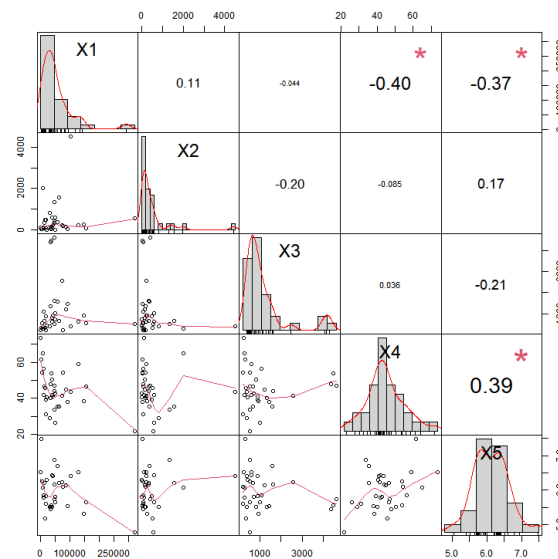
GAMBAR 2. Boxplot Setiap Variabel

Berdasarkan boxplot yang terbentuk, terdapat beberapa pencilan terlihat hampir di setiap variabel. Hal ini menunjukkan bahwa terdapat kesenjangan yang tinggi khususnya pada jumlah *Base Transceiver Station* (BTS) dan jumlah pulau yang banyak di beberapa provinsi. Dengan kondisi banyaknya pencilan tersebut dapat mengindikasikan bahwa kondisi wilayah yang berpulau-pulau menjadi pengaruh tersendiri dalam proses pembangunan infrastruktur khususnya pemasangan BTS.

Uji Multikolinearitas

Asumsi multikolinearitas dalam analisis kluster bertujuan untuk mendeteksi adanya korelasi antar kelima variabel. Korelasi yang kuat ditunjukkan dengan nilai koefisien korelasi (r) mendekati nilai 1 yang mengindikasikan adanya korelasi positif yang kuat. Sementara apabila koefisien korelasi mendekati -1, menandakan bahwa terdapat korelasi negatif yang kuat.

Berdasarkan grafik korelasi, kelima variabel menunjukkan kecenderungan memiliki koefisien korelasi yang relatif rendah. Korelasi paling tinggi ditunjukkan antara variabel X_1 (luas wilayah) dengan variabel X_4 (proporsi pengguna internet), dengan koefisien sebesar -0.40. Jika dilihat dari diagram pencar, data antar variabel cenderung menyebar dan tidak membentuk pola linear. Hal ini cukup memberikan indikasi bahwa setiap variabel saling independen dan tidak saling berpengaruh satu sama lain.



GAMBAR 3. Grafik Matriks Korelasi Antar Variabel

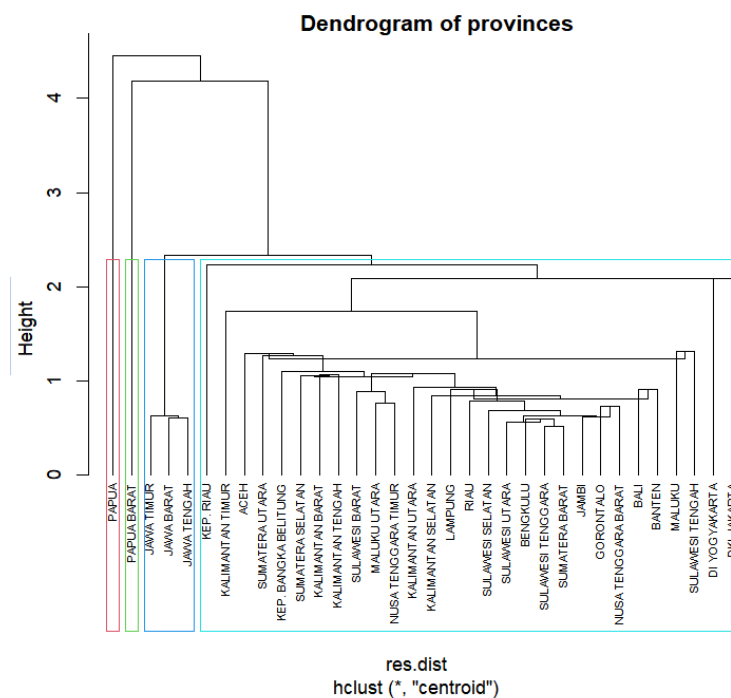
Untuk memastikan bahwa kesemua variabel tersebut homogen dan independen, akan dilakukan uji Bartlett sebagai berikut:

TABEL 3. Uji Bartlett Semua Variabel

Chi-Squared	df	p-value
17.537	10	0.0633

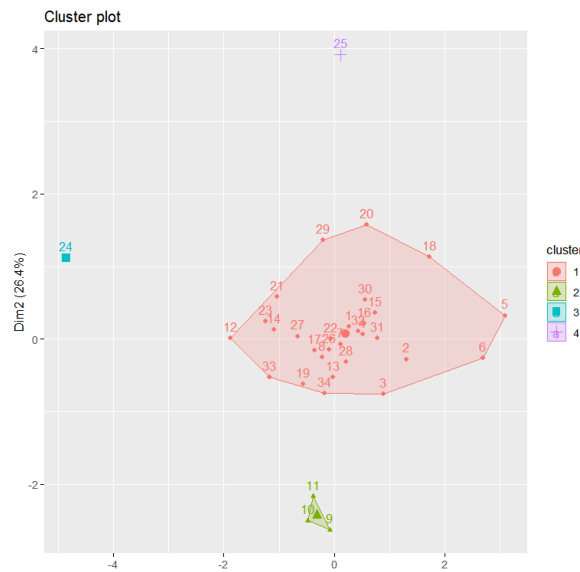
Dengan hipotesis awal adalah semua variabel tidak ada multikolinearitas dengan menggunakan tingkat kepercayaan 95%, maka dapat disimpulkan tidak menolak hipotesis awal sehingga kesemua variabel tidak terjadi multikolinearitas. Hal ini ditunjukkan dengan nilai signifikansi *p-value* lebih dari 5%.

Dendrogram dan Profilisasi Kluster



GAMBAR 7. Dendrogram 34 Provinsi

Berdasarkan hasil dendrogram yang terbentuk, proses klaster pertama adalah provinsi Papua. Klaster selanjutnya mulai terbentuk provinsi Papua Barat bersamaan dengan klaster sisa yang bercabang membentuk beberapa klaster baru yang lebih dominan dan membentuk kelompok besar. Dalam analisis klaster hirarki, penentuan jumlah klaster dapat dilihat dari proses percabangan yang terlihat mulai merata, disini peneliti menentukan jumlah klaster sebanyak 4 dikarenakan pada jarak centroid disekitar 2.2, dapat dengan mudah ditentukan klaster yang ideal. Pembagian dalam empat klaster sangat jelas dan terlihat ideal jika dilihat dalam bentuk plot pengelompokan klaster sebagai berikut:

**GAMBAR 8.** Plot Klaster 34 Provinsi

Dari hasil cluster plot yang terbentuk, pengelompokan provinsi sangat jelas terlihat mendekati pusat centroid masing-masing. Klaster 3 dan 4 terbentuk masing-masing 1 provinsi, sedangkan klaster 2 terdapat 3 provinsi, dan sisanya mayoritas berada dalam klaster 1. Provinsi DKI Jakarta masuk dalam klaster 1 yang mayoritas anggotanya berada di sekitar koordinat (0,0) yang menandakan kualitas jaringan komunikasi sudah baik. Sementara untuk klaster 2 berada pada koordinat yang cukup ideal hanya saja butuh perbaikan minor di beberapa sektor khususnya jumlah BTS yang belum merata di setiap daerah. Klaster 3 dan 4 berada pada koordinat yang kurang ideal dengan provinsi Papua perlu perbaikan dalam hal ketrampilan pengelolaan jaringan dan minimnya jumlah BTS yang berakibat jangkauan coverage BTS menjadi sangat luas sehingga kualitas jaringan sangat rendah. Sementara provinsi Papua Barat memiliki wilayah yang luas dan terdiri dari banyak pulau sehingga kualitas jaringan masih belum optimal.

TABEL 3. Klaster yang Terbentuk dari 34 Provinsi

Klaster	Provinsi
1	Kep.Riau, Kalimantan Timur, Aceh, Sumatera Utara, Kep. Bangak Belitung, Sumatera Selatan, Kalimantan Barat, Kalimantan Tengah, Sulawesi Barat, Maluku Utara, Nusa Tenggara Timur, Kalimantan Utara, Kalimantan Selatan, Lampung,Riau, Sulawesi Selatan, Sulawesi Utara, Bengkulu, Sulawesi Tenggara, Sumatera Barat, Jambi, Gorontalo, Nusa Tenggara Barat, Bali, Banten, Maluku, Sulawesi Tengah, D.I. Yogyakarta, dan DKI. Jakarta
2	Jawa Barat, Jawa Tengah, Jawa Timur
3	Papua

Dengan nilai korelasi cophenetic sebesar 0.8943226 menunjukkan bahwa klaster yang terbentuk sudah representatif dan akurat.

KESIMPULAN

Dari hasil analisis klaster yang sudah dilakukan, dapat diambil kesimpulan bahwa klaster yang terbentuk sebanyak 4 kelompok. Dengan Provinsi Papua serta Papua Barat menempati klaster yang cenderung menunjukkan kualitas jaringan internet yang rendah. Hal ini sesuai dengan pernyataan Kementerian Kominfo bahwa topografi wilayah merupakan faktor alami yang signifikan mempengaruhi kualitas jaringan. Untuk membangun infrastruktur di wilayah tersebut, perlu upaya dan investasi yang sangat besar. Pemeliharaan infrastruktur juga menjadi hal yang penting dalam proses pemerataan kualitas jaringan internet.

DAFTAR PUSTAKA

- Arsham, H. (2011). Bartlett's Test. <https://www.researchgate.net/publication/252322443>.
- BPS. (2018-2020). *Banyaknya Desa Kelurahan yang Memiliki Menara Base Transceiver Station (BTS) menurut Provinsi dan Klasifikasi Daerah*. Indonesia: <https://www.bps.go.id/indicator/2/1677/1>. diakses pada tanggal 16 Maret 2022.
- BPS. (2019). *Proporsi Individu Yang Menggunakan Internet Menurut Provinsi (Persen), 2017-2019*. Jakarta: <https://www.bps.go.id/indicator/27/1225/1/>. diakses pada tanggal 16 Maret 2022.
- BPS. (2021). *Luas Daerah dan Jumlah Pulau Menurut Provinsi, 2021*. Jakarta: https://www.bps.go.id/indikator/indikator/view_data_pub/0000/. diakses pada tanggal 16 Maret 2022.
- G Plate, J. (2020). *Rencana Strategis 2020-2024 Kementerian Komunikasi dan Informatika*. Jakarta, Indonesia: <https://web.kominfo.go.id/>.
- Kassambara, A. (2017). *Practical Guide To Cluster Analysis in R Unsupervised Machine Learning*. sthda.com.
- Nielsen, F. (2016). Hierarchical Clustering. In F. Nielsen, *Introduction to HPC with MPI for Data Science* (pp. 221-239). Springer.
- Silvi, R. (2018). Analisis Cluster dengan Data Pencilan Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika "MANTIK"*, 22-31.
- Suharni. (2021). Data Mining Analisis Cluster K-Means Pada Indeks Pembangunan Teknologi, Informasi, dan Telekomunikasi. *Journal of Technopreneurship and Information System (JTIS)*, 7-11.
- Widodo, E., & Novita, S. N. (2018). Analisis Cluster Penderita Disabilitas Mental di Provinsi D.I. Yogyakarta Tahun 2016. *KNPMP III 2018* (pp. 577-586). Surakarta: Program Studi Pendidikan Matematika FKIP UMS.