

CLUSTERING KABUPATEN/KOTA DI PROVINSI PAPUA BERDASARKAN PRODUK DOMESTIK REGIONAL BRUTO MENURUT LAPANGAN USAHA MENGGUNAKAN SINGLE LINKAGE DAN K-MEDOIDS

Caecilia Bintang Girik Allo^{1*}

¹Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Cenderawasih

*e-mail : caecilia.bintang@fmipa.uncen.ac.id

DOI: 10.14710/j.gauss.13.4.111-120

Article Info:

Received: 2024-07-26

Accepted: 2024-10-02

Available Online: 2024-10-04

Keywords:

*Gross Regional Domestic Bruto;
Single Linkage; K-Medoids;
Silhouette Coefficient*

Abstract: Gross Regional Domestic Product (GRDP) using the production approach represents the total value added from goods and services produced by different sectors within a specific region over a defined timeframe. There are 17 business sectors used to obtain the GRDP. The growth rate of GRDP in Papua is decrease in 2023. The growth rate is only 3,44%, whereas the previous year it reached 4,11%. An analysis is needed to assist the government to enhance the GRDP in Papua. Clustering method can group districts/cities that have similar characteristics. The aims of this article is to determine the best method for clustering districts/cities in Papua using GRDP data. Single Method and K-Medoids is used in this article. Based on silhouette coefficients, Single Methods is better than K-Medoids to clustering districts/cities in Papua. Based on criteria of silhouette coefficients, number of clusters formed is three.

1. PENDAHULUAN

Berbagai visi dan misi dibentuk dalam rangka mencapai Indonesia Emas 2045. Salah satu bidang yang masuk sasaran utama dalam visi dan misi tersebut adalah bidang ekonomi. Peningkatan pendapatan per kapita dan penurunan ketimpangan antar wilayah merupakan dua sasaran utama dalam visi Indonesia Emas 2045 (Kementerian Perencanaan Pembangunan Sosial, 2024). Produk Domestik Regional Bruto (PDRB) merupakan salah satu indikator yang sangat perlu diperhatikan untuk mendukung visi tersebut.

PDRB dapat diperoleh melalui tiga pendekatan, yaitu pendekatan produksi, pengeluaran, dan pendapatan. Salah satu pendekatan PDRB yang digunakan Badan Pusat Statistika (BPS) adalah pendekatan produksi. PDRB melalui pendekatan produksi diperoleh dari 17 sektor usaha yang ada di Indonesia. PDRB biasanya disajikan dalam dua bentuk, yaitu PDRB berdasarkan harga berlaku atau PDRB berdasarkan harga konstan. PDRB berdasarkan harga berlaku digunakan untuk mengukur struktur ekonomi suatu daerah, sedangkan PDRB berdasarkan harga konstan digunakan untuk mengukur pertumbuhan ekonomi suatu daerah. Saat ini ukuran harga konstan yang digunakan adalah harga konstan 2010.

Laju pertumbuhan PDRB berdasarkan harga konstan di Provinsi Papua mengalami penurunan pada tahun 2023. Pada tahun 2022, laju pertumbuhan PDRB berdasarkan harga konstan di Provinsi Papua mencapai 4,11%. Namun, pada tahun 2023 hanya mencapai 3,44% (BPS Provinsi Papua, 2024). Apabila dilihat dari sisi nilai PDRB, kenaikan PDRB dari tahun 2022 ke 2023 tidak sebesar kenaikan nilai PDRB dari tahun 2021 ke 2022. Hal tersebut tentunya menjadi perhatian pemerintah Provinsi Papua. Diperlukan suatu analisis yang dapat membantu pemerintah dalam memperhatikan karakteristik setiap Kabupaten/Kota di Provinsi Papua agar dapat meningkatkan PDRB setiap Kabupaten/Kota.

Analisis clustering merupakan salah satu metode yang dapat digunakan untuk mengelompokkan Kabupaten/Kota yang memiliki karakteristik sama, sehingga dapat mempermudah pemerintah dalam membuat kebijakan. Secara garis besar, konsep clustering adalah setiap anggota dalam cluster memiliki karakteristik yang sama dan anggota antar cluster memiliki karakteristik yang berbeda. Berbagai macam metode clustering telah dilakukan pada berbagai macam jenis data.

Metode *K-Means* dan DBSCAN digunakan untuk clustering Kabupaten/Kota di Provinsi Papua berdasarkan PDRB (Wororomi dkk., 2023). Metode DBSCAN mampu mendeteksi Kabupaten/Kota yang menjadi outlier pada data PDRB. Sapriyanti dan Rianto (2020) memperoleh bahwa metode *K-Means* lebih ideal dibandingkan dengan metode *Single Linkage* menggunakan data *agent* pada *call center*. Metode *Single Linkage* mampu menghasilkan cluster yang lebih heterogen apabila dibandingkan dengan metode ward pada variabel – variabel yang menjadi potensi desa (Fathia dkk., 2016). Syafiyah dkk. (2020) melakukan perbandingan pada metode Hierarchical dan Non – Hierarchical menggunakan data indikator ketenagakerjaan di Jawa Barat. Hasil penelitian memperlihatkan bahwa Hierarchical Clustering menunjukkan performa lebih baik.

Metode clustering pun terus berkembang berdasarkan data. Metode *K-Means* menjadi tidak efektif apabila diaplikasikan pada data yang memiliki *outliers* (Han dan Kamber, 2006). *K-Medoids* merupakan salah satu algoritma dalam clustering yang dapat digunakan untuk data yang memiliki *outliers* (Kauffman dan Rousseeuw, 1990). Metode *K-Medoids* digunakan oleh Nahdliyah, dkk (2019) dalam mengelompokkan Kabupaten/Kota di Jawa Tengah berdasarkan jumlah kriminalitas. Ibrahim dkk (2020) juga menggunakan metode *K-Medoids* untuk mengelompokkan desa/kelurahan di Kabupaten Kutai Kartanegara berdasarkan data potensi desa.

Berdasarkan penelitian sebelumnya, penelitian ini bertujuan untuk membandingkan metode *Single Linkage* dan *K-Medoids* pada data PDRB harga konstan menurut lapangan usaha di Provinsi Papua pada tahun 2022. Pemilihan metode terbaik menggunakan *Silhouette coefficient*. Selain memperoleh metode terbaik, penelitian ini juga akan menghasilkan kelompok Kabupaten/Kota di Provinsi Papua. Peneliti berharap hasil penelitian ini dapat bermanfaat bagi pemerintah dalam penentuan kebijakan agar PDRB di Provinsi Papua terus meningkat.

2. TINJAUAN PUSTAKA

Adapun beberapa beberapa tinjauan pustaka yang digunakan dalam penelitian berupa Standarisasi Data, Asumsi Clustering, Jarak Euclidean, metode *Single Linkage*, metode *K-Medoids*, *Silhouette Coefficient*.

Sebelum melakukan analisis data, perlu diperhatikan *range* dan skala dari data. Standarisasi data dilakukan agar data berada dalam *range* yang sama. Terdapat beberapa metode standarisasi, salah satunya adalah metode *z-score* atau yang sering disebut dengan *z-score normalization*. Persamaan (1) digunakan untuk mencari *z-score*.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad (1)$$

dengan:

$i = 1, 2, \dots, n ; j = 1, 2, \dots, p$

n : banyaknya pengamatan

p : banyaknya variabel

z_{ij} : nilai standarisasi pengamatan ke – i pada variabel ke – j

x_{ij} : nilai pengamatan ke i pada variabel ke j
 \bar{x}_j : nilai rata-rata variabel ke j
 σ_{ij} : standar deviasi variabel ke j

Sebelum melakukan analisis, perlu juga dicek terkait asumsi pada analisis cluster. Menurut Hair *et al.* (2010) terdapat dua asumsi dalam analisis cluster, yaitu sampel representatif dan tidak terdapat multikolinearitas antar variabel. Sampel representatif berarti sampel yang diambil sudah mewakili populasi. Sampel representatif dapat diuji menggunakan uji *Kaiser-Mayer-Olkin* (KMO). Nilai minimum KMO adalah 0 dan nilai maksimum KMO adalah 1. Jika nilai KMO $< 0,5$ maka dapat dikatakan bahwa sampel yang diambil tidak representatif. Sebaliknya, Jika nilai KMO $\geq 0,5$ maka dapat dikatakan bahwa sampel yang diambil representatif (Hair *et al.*, 2010). Persamaan (2) dapat digunakan untuk menghitung nilai KMO.

$$KMO = \frac{\sum_{j=1}^p \sum_{k=1, k \neq j}^p r_{jk}^2}{\sum_{j=1}^p \sum_{k=1, k \neq j}^p r_{jk}^2 + \sum_{j=1}^p \sum_{k=1, k \neq j}^p \sum_{l=1}^p \rho_{j(kl)}^2} \quad (2)$$

dengan:

p : banyaknya variabel

r_{jk} : koefisien korelasi variabel ke j dan variabel ke k

$\rho_{j(kl)}$: koefisien korelasi parsial antara variabel ke j , variabel ke k , dan variabel ke l

Multikolinearitas merupakan adanya hubungan linear antar variabel. Nilai VIF dapat digunakan untuk mengecek multikolinearitas antar variabel. Jika nilai VIF < 10 , maka dapat disimpulkan tidak terjadi multikolinearitas antar variabel (Gujarati, 2004). Persamaan (3) dapat digunakan untuk mencari nilai VIF.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (3)$$

dengan R_j^2 merupakan koefisien determinasi antara variabel ke j dengan variabel lainnya yang dapat diperoleh dengan meregresikan variabel ke j dengan variabel lainnya.

Apabila terdapat multikolinearitas, maka salah satu teknik yang dapat digunakan untuk mengatasi hal tersebut adalah *Principal Component Analysis* (PCA). PCA mampu mentransformasi variabel awal menjadi variabel baru yang tidak saling berkorelasi. Walaupun menghasilkan variabel baru, PCA tidak menghilangkan informasi dari variabel awal.

Pada awal analisis cluster, banyak metode yang menggunakan jarak. Salah satu jarak yang sering digunakan apabila variabel yang dimiliki merupakan numerik adalah jarak Euclidean. Persamaan (4) dapat digunakan untuk mencari jarak antar dua pengamatan atau dua objek.

$$d_{im} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{mj})^2} \quad (4)$$

dengan:

d_{im} : jarak antara pengamatan ke i dan pengamatan ke m

x_{ij} : nilai pengamatan pengamatan ke i pada variabel ke j

x_{mj} : nilai pengamatan pengamatan ke m pada variabel ke j

Analisis cluster pada umumnya dibagi menjadi dua, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Hierarchical clustering* merupakan metode clustering yang membangun suatu hirarki cluster. Pada *hierarchical clustering* terdapat dua metode, yaitu

agglomerative dan *divisive* (Johnson & Wichern, 2007). *Agglomerative* mulai dengan menganggap masing-masing object merupakan anggota dari masing-masing cluster. Kemudian setiap cluster akan digabungkan menggunakan algoritma dari *agglomerative* hingga semua cluster tergabung menjadi satu cluster. *Divise* merupakan kebalikan dari *agglomerative*. Salah satu algoritma dalam *agglomerative* adalah Single Linkage. Pada metode hierarki, peneliti diberikan wewenang untuk menentukan banyaknya cluster yang akan dibentuk.

Berikut merupakan algoritma dari Single Linkage:

1. Membuat matriks jarak $\mathbf{D} = \{d_{im}\}$
2. Tentukan jarak terkecil dan gabungkan kedua pengamatan atau objek menjadi satu cluster. Misalkan objek yang memiliki jarak terkecil adalah U dan V , maka gabungkan U dan V menjadi satu cluster (UV).
3. Update matriks jarak tersebut dengan cara mencari jarak pada cluster yang sudah digabungkan menggunakan persamaan (5).

$$d_{UV} = \min\{d_{UW}, d_{VW}\} \quad (5)$$
4. Lakukan hingga semua objek bergabung menjadi satu cluster.

K-Medoids termasuk dalam metode non-hierarchical clustering. K-Medoids merupakan salah satu alternatif yang dapat digunakan jika didapati outlier pada data. K-Medoids juga sering dikenal sebagai Partitioning Around Medoids (PAM). Algoritma K-Medoids adalah sebagai berikut (Asmiatun *et al.*, 2020):

1. Tentukan banyaknya cluster, yaitu k .
Terdapat berbagai metode yang dapat digunakan dalam penentuan k , diantaranya Metode Elbow dan Silhouette. Pada penelitian ini akan dilakukan mulai dari $k = 2$ hingga maksimum dari k .
2. Menentukan medoids masing-masing cluster secara acak.
Penentuan medoids dilakukan secara acak berarti bisa dipilih sebarang titik yang menjadi medoid masing-masing cluster atau mengambil medoids dari pengamatan.
3. Hitung jarak setiap objek ke masing-masing medoid pada setiap cluster menggunakan persamaan (4).
4. Tentukan anggota cluster dengan memilih jarak minimum dari objek ke setiap cluster. Misalkan jarak minimum yang terpilih disimbolkan dengan d_s .
5. Hitung Total Cost (TC). Total Cost diperoleh dengan menjumlahkan jarak minimum yang terpilih pada semua cluster. Secara umum TC dapat dihitung menggunakan persamaan (6).

$$TC = \sum_{i=1}^n d_s \quad (6)$$

Total Cost pertama disimbolkan dengan TC_o .

6. Pilih secara acak kandidat medoids baru seperti pada langkah 2. Kemudian lakukan langkah 3 dan langkah 4. Total Cost terbaru disimbolkan dengan TC_n .
7. Hitung selisih total cost, yaitu S dimana $S = TC_n - TC_o$. Jika $S > 0$, maka iterasi berhenti dan anggota cluster yang digunakan adalah anggota saat menghitung TC_o . Jika $S < 0$, maka lakukan langkah 6 kembali. Perlu diingat bahwa TC_o diperoleh dari TC_n iterasi sebelumnya dan TC_n diperoleh dari penentuan medoid baru.

Silhouette Coefficient merupakan suatu nilai yang dapat digunakan untuk mengevaluasi hasil cluster. Silhouette Coefficient berkisar antara -1 hingga 1. Jika nilai Silhouette

Coefficient semakin mendekat 1, maka semakin bagus cluster yang terbentuk. Perhitungan Silhouette Coefficient dapat dilihat pada persamaan (7) (Wang dan Xu, 2019).

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (7)$$

dengan:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

$$a(i) = \frac{1}{|U| - 1} \sum_{m \in U, m \neq i} d_{im} \quad (9)$$

$$b(i) = \min_{V \neq U} \frac{1}{|V|} \sum_{z \in V} d_{iz} \quad (10)$$

dimana:

|U|: banyaknya anggota dalam cluster U

|V|: banyaknya anggota dalam cluster V

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang merupakan hasil publikasi Badan Pusat Statistik (BPS) Provinsi Papua tentang Produk Domestik Regional Bruto Menurut Lapangan Usaha. Setelah pemekaran, provinsi Papua yang semula memiliki 29 Kabupaten/Kota saat ini memiliki 9 Kabupaten/Kota.

Terdapat 17 variabel yang digunakan, yaitu Pertanian, Kehutanan, dan Perikanan (X_1); Pertambangan dan Penggalian (X_2); Industri Pengolahan (X_3); Pengadaan Listrik dan Gas (X_4); Pengadaan Air, Pengelolaan Sampah, Limbah, dan Daur Ulang (X_5); Konstruksi (X_6); Perdagangan Besar dan Eceran; Reparasi Mobil dan Sepeda Motor (X_7); Transportasi dan Pergudangan (X_8); Penyediaan Akomodasi dan Makan Minum (X_9); Informasi dan Komunikasi (X_{10}); Jasa Keuangan dan Asuransi (X_{11}); Real Estate (X_{12}); Jasa Perusahaan (X_{13}); Administrasi Pemerintahan, Pertahanan, dan Jaminan Sosial Wajib (X_{14}); Jasa Pendidikan (X_{15}); Jasa Kesehatan dan Kegiatan Sosial (X_{16}); dan Jasa Lainnya (X_{17}).

Adapun struktur data pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Struktur Data Penelitian

Kab/Kota	X_1	X_2	...	X_{17}
1	$X_{1,1}$	$X_{1,2}$...	$X_{1,17}$
2	$X_{2,1}$	$X_{2,2}$...	$X_{2,17}$
3	$X_{3,1}$	$X_{3,2}$...	$X_{3,17}$
⋮	⋮	⋮	...	⋮
9	$X_{9,1}$	$X_{9,2}$...	$X_{9,17}$

Pengolahan data menggunakan software Python. Adapun langkah-langkah pengolahan data adalah sebagai berikut:

1. Memasukkan data
2. Melakukan analisis deskriptif
3. Melakukan transformasi data menggunakan z-score menggunakan persamaan (1)
4. Pengujian multikolinearitas menggunakan persamaan (3). Jika terdapat multikolinearitas, maka digunakan PCA.

5. Melakukan clustering menggunakan algoritma Single Linkage dan K-Medoids.
6. Melakukan evaluasi menggunakan silhouette coefficient menggunakan persamaan (7)
7. Interpretasi hasil.

4. HASIL DAN PEMBAHASAN

Sebagai langkah awal dilakukan analisis deskriptif pada data agar mengetahui gambaran umum dari data. Gambar 1 menunjukkan bahwa Kabupaten Keerom, Supiori, dan Memberamo Raya tidak memiliki kontribusi sama sekali pada variabel X_5 . Pemerintah juga perlu memperhatikan tiga Kabupaten/Kota yang memiliki nilai terendah pada setiap variabelnya. Selanjutnya, dilakukan pengecekan outlier pada data.



Gambar 1. Tiga Kabupaten/Kota Dengan Nilai Terendah Setiap Variabel

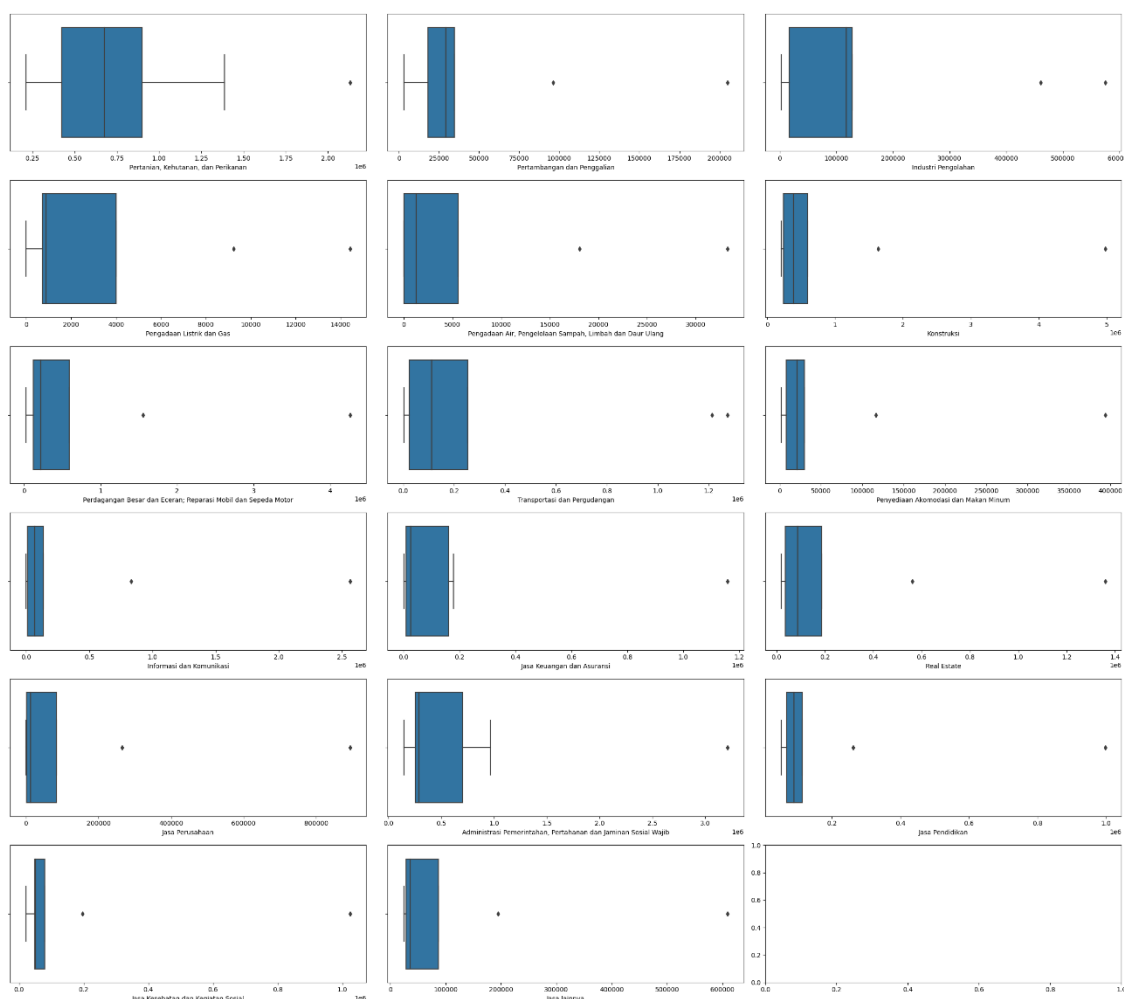
Gambar 2 menunjukkan bahwa terdapat outlier pada setiap variabel. Hal ini berarti, terdapat Kabupaten/Kota yang nilainya selalu berbeda dari yang Kabupaten/Kota lainnya. Jumlah Kabupaten/Kota yang outlier bisa satu atau dua Kabupaten/Kota. Apabila dilakukan pengecekan Kabupaten/Kota yang menjadi outlier, maka akan diperoleh Kota Jayapura dan Kabupaten Jayapura.

Selanjutnya setelah melakukan analisis deskriptif, dilakukan transformasi z -score. Setelah melakukan transformasi, dilakukan pengecekan asumsi clustering. Pada kasus ini, asumsi sampel representatif tidak dilakukan karena penelitian ini tidak menggunakan sampel. Penelitian ini menggunakan seluruh Kabupaten/kota yang berada di Provinsi Papua. Selanjutnya, dilakukan pengecekan terhadap multikolinearitas setiap variabel. Pada penelitian ini, jumlah Kabupaten/Kota hanya 9 dan jumlah variabel yang digunakan adalah 17. Hal tersebut menyebabkan terjadinya multikolinearitas. Sehingga dilakukan teknik PCA.

Tabel 2 merupakan hasil dari PCA. Jumlah komponen utama yang dipilih adalah komponen utama yang nilai eigennya lebih besar dari 1. Berdasarkan hasil pada Tabel 2, maka akan digunakan 3 Komponen Utama.

Tabel 2. Nilai Eigen, Proporsi Varian, dan Proporsi Kumulatif Komponen Utama

Komponen Utama	Nilai Eigen	Proporsi Varian	Proporsi Kumulatif
1	11,5837316	0,87701202	0,87701202
2	3,91744406	0,100303058	0,97731508
3	1,65940328	0,0179975113	0,99531259
4	0,605452003	0,00239589626	0,99770848
5	0,496604277	0,00161186803	0,99932035
6	0,245522868	0,000393996592	0,99971435
7	0,164045125	0,000175887601	0,99989024
8	0,129590613	0,000109762922	1
9	$5,30924949 \times 10^{-16}$	$1,84236145 \times 10^{-33}$	1



Gambar 2. Boxplot Setiap Variabel

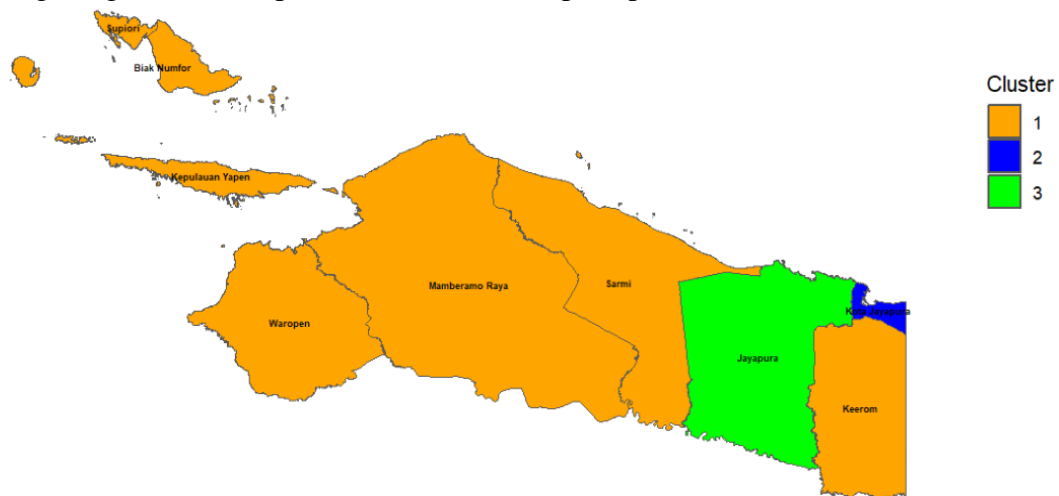
Selanjutnya dilakukan clustering menggunakan *Single Linkage* dan *K-Medoids* dengan metode evaluasi *Silhouette coefficient* (SC). Tabel 2 merupakan nilai SC untuk setiap metode. Berdasarkan hasil Tabel 3 diperoleh bahwa jumlah cluster terbaik tiap metode adalah 2. Namun metode *Single Linkage* memberikan performa lebih baik dibandingkan dengan *K-Medoids*.

Tabel 3. Nilai Silhouette Coefficient

Banyaknya Cluster	Metode Clustering	
	<i>Single Linkage</i>	<i>K-Medoids</i>

2	0,687792	0,675601
3	0,628295	0,225406
4	0,442762	0,325055
5	0,281087	0,095341
6	0,367593	0,140582
7	0,284429	-0,045437
8	0,183120	-0,019771

Perbedaan nilai SC saat $k = 2$ dan $k = 3$ tidak terlalu jauh berbeda pada Single Linkage. Berdasarkan kriteria Silhouette dapat disimpulkan bahwa nilai SC yang berada pada 0,51 sampai 0,7 berarti cluster tersebut terstruktur dengan baik (Wororomi *et al.*, 2023). Sehingga pada penelitian ini, peneliti memutuskan untuk menggunakan $k = 3$. Hasil clustering dengan $k = 3$ dapat divisualisasikan seperti pada Gambar 3.



Gambar 3. Peta Ilustrasi Hasil Cluster Terpilih

Berdasarkan Gambar 3 diperoleh bahwa cluster 3 dengan anggota Kabupaten Jayapura memiliki nilai tertinggi pada Pertanian, Kehutanan, dan Perikanan (X_1); Pertambangan dan Penggalian (X_2); serta Transportasi dan Pergudangan (X_8). Sedangkan cluster 2 yang anggotanya adalah Kota Jayapura merupakan kota yang memiliki nilai tertinggi pada setiap variabel, kecuali pada variabel Pertanian, Kehutanan, dan Perikanan (X_1); Pertambangan dan Penggalian (X_2); serta Transportasi dan Pergudangan (X_8). Berbeda dengan cluster 1 dengan anggota Kepulauan Yapen, Biak Numfor, Sarmi, Keerom, Waropen, Supriori, dan Memberamo Raya merupakan kabupaten-kabupaten dengan nilai terendah pada setiap variabel. Berdasarkan hasil tersebut dapat diperoleh bahwa cluster 1 didominasi dengan pendapatan pada 17 variabel yang rendah. Cluster 2 unggul pada variabel setiap variabel kecuali X_1 , X_2 , dan X_8 . Cluster 3 unggul pada variabel X_1 , X_2 , dan X_8 . Keunggulan pada cluster 2, yaitu Kota Jayapura dapat disebabkan karena Kota Jayapura merupakan Ibu Kota Provinsi Papua.

5. KESIMPULAN

Berdasarkan hasil clustering menggunakan *Single Linkage* dan *K-Medoids* diperoleh bahwa *Single Linkage* lebih baik dalam menghasilkan cluster pada data Produk Domestik Regional Bruto (PDRB) di Provinsi Papua. Kriteria *Silhouette Coefficient* digunakan sebagai

evaluasi metode dan diperoleh bahwa banyaknya cluster yang terbentuk adalah 3. Cluster 1 beranggotakan Kepulauan Yapen, Biak Numfor, Sarmi, Keerom, Waropen, Supriori, dan Memberamo Raya. Cluster 2 beranggotakan Kota Jayapura. Sedangkan Cluster 3 beranggotakan Kabupaten Jayapura. Pemerintah perlu memperhatikan kabupaten-kabupaten yang berada pada cluster 1 agar variabel pendukung PDRB dapat meningkat, karena kabupaten dalam cluster 1 memiliki nilai terendah pada 17 variabel.

DAFTAR PUSTAKA

- Asmiatun, S., Wakhidah, N., dan Putri, A. N. 2020. Penerapan Metode K-Medoids untuk Pengelompokan Kondisi Jalan di Kota Semarang. *Jurnal Teknik Infomatika dan Sistem Informasi* Vol. 6 No. 2 DOI: <https://doi.org/10.35957/jatisi.v6i2.193>
- Badan Pusat Statistika (BPS) Provinsi Papua. 2024. *Produk Domestik Regional Bruto Provinsi Papua Menurut Lapangan Usaha 2019 – 2023*. Jayapura: Badan Pusat Statistika (BPS) Provinsi Papua.
- Fathia, A. N., Rahmawati, R. dan Tarno. 2016. Analisis Klaster Kecamatan di Kabuapten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. *Jurnal Gaussian* Vol. 5 No. 4 DOI: 10.14710/j.gauss.5.4.801-810
- Gujarati, D. N. 2004. *Basic Econometrics* (4th Edition). Singapore: McGraw Hill Inc.
- Hair, J. F., Black, W. C., Babin, B. J., dan Anderson, R. E. 2010. *Multivariate Data Analysis* (7th Edition). New York: Pearson Prentice Hall.
- Han, J. dan Kamber, M. 2006. *Data Mining: Concept and Techniques*. San Fransisco. Morgan Kauffman Publisher.
- Ibrahim, R. N., Hayati, M. N., dan Amijaya, F. D. T. 2020. Penerapan Algoritma K-Medoids Pada Pengelompokan Wilayah Desa atau Kelurahan di Kabupaten Kutai Kartanegara (Studi Kasus: Data Hasil Pendataan Potensi Desa (PODES) Tahun 2018). *Jurnal Eksponensial* Vol. 11 No. 2 DOI: 10.30872/eksponensial.v11i2.658
- Johnson, R. A., dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis* (6th Edition). New Jersey: Prentice Hall International Inc.
- Kaufman, L., dan Rousseeuw, P. J. 1990. *Finding Groups in Data*. New York. John Willey & Sons.
- Kementerian Perencanaan Pembangunan Sosial. 2024. *Rancangan Akhir Rencana Pembangunan Jangka Panjang Nasional 2025 – 2045*. Jakarta: Kementerian Perencanaan Pembangunan Sosial.
- Nahdliyah, M. A., Widiharih, T. dan Prahutama, A. 2019. Metode K-Medoids Clustering Dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian* Vol. 8 No. 2 DOI: 10.14710/j.gauss.8.2.161-170
- Sapriyanti dan Rianto, Y. 2020. Komparasi Metode Clustering K-Means dan Single Linkage untuk Penentuan Kelompok Agent Pada Call Center. *JISAMAR: Journal of Information System, Applied, Management, Accounting and Research* Vol. 4 No. 3 <https://journal.stmikjayakarta.ac.id/index.php/jisamar/article/view/244>
- Syafiyah, U., Asrafi, I., Wicaksono, B., Puspitasari, D. P., dkk. 2022. Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020. *Seminar Nasional Official Statistics* Vol. 1 DOI: 10.34123/semnasoffstat.v2022i1.1221

- Wang, X., dan Xu, Y. 2019. An Improved Index for Clustering Validation Based On Silhouette Index and Calinski-Harabasz Index. *IOP Conference Series: Materials Science and Engineering* Vol 569 No. 5 DOI: 10.1088/1757-899X/569/5/052024
- Wororomi, J. K., Allo, C. B. G., Paranoan, N. R., Gusthvi, W. 2023. Performance of K-Means and DBSCAN Algorithm in Clustering Gross Regional Domestic Product. *JICP: Journal of International Conference Proceedings* Vol. 5 No. 5 DOI: 10.32535/jicp.v6i5.2710