



## **Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Perumahan dan Kesehatan Lingkungan Menggunakan Metode *K-Medoids***

### ***Clustering Provinces in Indonesia Based on Housing and Environmental Health Indicators Using the K-Medoids Method.***

**Alwan Fadlurohman<sup>1</sup>, Indah Manfaati Nur<sup>2</sup>**

<sup>1</sup> Universitas Muhammadiyah Semarang, Kota Semarang

<sup>2</sup> Universitas Muhammadiyah Semarang, Kota Semarang

*Corresponding author:* indahmnur@unimus.ac.id

### **Abstrak**

Indikator perumahan dan kesehatan lingkungan merupakan salah satu indikator yang sangat penting dalam upaya mewujudkan tujuan dari Tujuan Pembangunan Berkelanjutan (TPB). Kondisi perumahan dan kesehatan lingkungan di setiap provinsi di Indonesia berbeda-beda, oleh karena itu dalam melakukan prioritas peningkatan masalah perumahan dan kesehatan lingkungan juga berbeda. Tujuan atas penelitian ini guna mengklasifikasi provinsi di Indonesia atas dasar indikator perumahan dan kesehatan lingkungan untuk mengetahui tinggi rendahnya kualitas perumahan dan lingkungan di setiap provinsi. Sehingga, hasil penelitian dengan harap mampu membantu pemerintah mengoptimalkan upaya kesehatan lingkungan. Pengelompokan provinsi dilakukan dengan metode *K-Medoids* yang memiliki kelebihan *robust* terhadap data yang mengandung pencilan. Ukuran kemiripan objek dihitung dengan menggunakan metode jarak *Euclidean*. Sementara itu, pemilihan jumlah *cluster* terbaik dilakukan menggunakan indeks *silhouette* yang menghasilkan 2 cluster, dimana pada *cluster* 1 didapatkan 29 provinsi dengan nilai rata-rata indikator perumahan dan kesehatan lingkungan ( $X_2$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ , dan  $X_{10}$ ) rendah. Lalu cluster 2 didapatkan 5 provinsi dengan nilai rata-rata indikator perumahan dan kesehatan lingkungan ( $X_2$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ , dan  $X_{10}$ ) tinggi.

**Kata Kunci:** Dampak Perkotaan, Pencilan, Pengelompokkan, Sanitasi.

### ***Abstract***

Housing and environmental health indicators are very important in realizing the goals of the Sustainable Development Goals (SDGs). The condition of housing and environmental health in each province in Indonesia is different, therefore prioritizing the improvement of housing and environmental health problems is also different. This study aims to categorize provinces in Indonesia based on housing and environmental health indicators to determine the high and low quality of housing and the environment in each province. Therefore, the results of this study are to support the government in optimizing environmental health efforts. Province clustering is performed using the *K-Medoids* method which has the advantage of being *robust* to data containing outliers. The object similarity measure is calculated using the *Euclidean* distance method. Meanwhile, the selection of the best number of clusters was conducted using the *silhouette* index which resulted in 2 clusters, where in cluster 1 there were 29 provinces with low average values of housing and environmental health indicators ( $X_2$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ , dan  $X_{10}$ ) and cluster 2 obtained 5 provinces with high average values of housing and environmental health indicators ( $X_2$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ , dan  $X_{10}$ ).

**Keywords:** *Clustering, Impact of Cities, Outlier, Sanitation.*



## PENDAHULUAN

Urbanisasi yakni bagian dari proses mobilitas penduduk yang bagus didiskusikan. Hal ini dikarenakan urbanisasi di Indonesia semakin hari semakin terus terjadi. Badan Pusat Statistik (BPS) memprediksi, 56,7% penduduk Indonesia tinggal di wilayah perkotaan tahun 2020. Persentase diatas diperkirakan akan naik menjadi 66,6% di tahun 2035. Atas hal tersebut, diatas setengah populasi dunia tinggal di wilayah perkotaan. Angka ini diprediksi akan terus meningkat menjadi 6,5 miliar orang pada tahun 2050. Fenomena ini memunculkan berbagai permasalahan dan tantangan yang berkaitan dengan isu-isu berkelanjutan di sektor kesejahteraan rakyat, terutama di subektor perumahan, infrastruktur dan fasilitas pelayanan dasar.

Fenomena urbanisasi dan semakin pesatnya laju pertumbuhan jumlah penduduk akan meningkatkan kebutuhan akan perumahan dan permukiman, Oleh karena itu, pada saat ini arah kebijakan yang ditempuh pemerintah Indonesia yang ada di dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) tahun 2020-2024, yang salah satu arak kebijakannya yakni pembangunan infrastruktur dasar mencakup perumahan yang layak, akses listrik, air minum, dan sanitasi. Sejalan dengan hal ini, kesepakatan internasional pada Agenda 2030 untuk Tujuan Pembangunan Berkelanjutan (*the 2030 Agenda for Sustainable Development Goals*) juga memprioritaskan perumahan dan pembangunan infrastruktur dasar dalam capaiannya. Tujuan tersebut tercantum pada tujuan ke 6, 7, dan 11. Tujuan 6 pada Tujuan Pembangunan Berkelanjutan (TPB) fokus pada air bersih dan sanitasi layak, mengupayakan ketersediaan serta pengelolaan air bersih. Sementara tujuan fokus kepada energi bersih dan terjangkau, dengan upaya ada jaminan akses energi yang terjangkau, andal, berkelanjutan dan modern untuk semua. Sedangkan tujuan 11 fokus pada kota dan permukiman berkelanjutan, dengan usaha mewujudkan kota dan permukiman inklusif, aman, tangguh, dan berkelanjutan.

Pada tahun 2020, Badan Pusat Statistik (BPS) mengembangkan sejumlah indikator terkait perumahan dan lingkungan. Indikator ini mencakup aspek-aspek seperti kondisi fisik bangunan, fasilitas perumahan, kesehatan lingkungan, dan pengeluaran perumahan. Beberapa indikator spesifik dalam kategori ini mencakup persentase rumah tangga berdasarkan kepemilikan rumah sendiri, persentase rumah tangga berdasarkan status kepemilikan rumah kontrak/sewa, persentase rumah tangga berdasarkan jenis atap terluas yang bukan dari ijuk atau bahan lainnya, persentase rumah tangga berdasarkan jenis dinding terluas yang bukan dari bambu atau bahan lainnya, persentase rumah tangga berdasarkan jenis lantai terluas yang bukan dari tanah, persentase rumah tangga berdasarkan sumber air minum yang dianggap layak, persentase rumah tangga berdasarkan fasilitas sanitasi yang dianggap layak, persentase rumah tangga berdasarkan sumber penerangan yang berasal dari listrik, persentase rumah tangga berdasarkan sumber



penerangan dari PLN (Perusahaan Listrik Negara), dan persentase rumah tangga berdasarkan luas hunian per kapita yang kurang dari 7,2 meter persegi. Semua indikator ini membantu dalam mengevaluasi kondisi perumahan dan lingkungan di suatu wilayah, serta memberikan gambaran tentang akses dan kualitas perumahan dan infrastruktur yang tersedia bagi rumah tangga.

Pencapaian sasaran kebijakan RPJMN 2020-2024 dan juga Agenda 2030 untuk Tujuan Pembangunan Berkelanjutan perlu koordinasi perencanaan dan pelaksanaan program berbagai sektor terhubung lalu fokus target yang sama. Salah satu prasyarat keberhasilan untuk pencapaian sasaran kebijakan terlalu memihak ketepatan identifikasi target group dan target area (Basri, 2015) seperti yang dikutip oleh Ghaisani et al, 2016. Oleh karena itu, sebagai langkah awal perlu mempertimbangkan pengelompokan karakteristik provinsi di Indonesia berdasarkan Indikator Perumahan dan Kesehatan Lingkungan. Salah satu metode yang dapat digunakan untuk mengelompokkan suatu karakteristik yang sama adalah analisis *cluster*. Dengan adanya pengelompokan ini diharapkan bisa mengetahui permasalahan pada indikator perumahan di setiap provinsi di Indonesia. Contoh metode analisis *cluster* yang dapat digunakan yakni *K-Medoids*. *K-Medoids*, metode partisional clustering yang bertujuan untuk membentuk k-cluster di antara data yang paling mencirikan objek dalam kumpulan data. Dalam *K-Medoids*, medoids dari setiap cluster dianggap sebagai representasi objek yang paling "tengah" dari cluster tersebut. Algoritma *K-Medoids* berusaha untuk menemukan medoids dan mengelompokkan objek yang berdekatan dengan medoids untuk membentuk cluster baru (Wira et al., 2019). Sehingga permasalahan di atas, akan di selesaikan dengan mengelompokkan Provinsi di Indonesia yang mempunyai kesamaan karakteristik berdasarkan indikator perumahan dan kesehatan lingkungan tahun 2020 menggunakan analisis *cluster* dengan algoritma *K-Medoids*.

## **METODE**

Data yang dipakai ialah data sekunder tentang indikator perumahan dan kesehatan lingkungan tahun 2020 yang didapatkan dari Badan Pusat Statistik (BPS) dengan unit observasi yang dipakai yakni 34 provinsi di Indonesia. Variabel penelitian memakai pada Tabel 1.

Tabel 1.  
Variabel Penelitian

Variabel	Nama Variabel
X <sub>1</sub>	Persentase Rumah Tangga menurut Status Kepemilikan Rumah Milik Sendiri
X <sub>2</sub>	Persentase Rumah Tangga menurut Status Kepemilikan Rumah Kontrak/sewa
X <sub>3</sub>	Persentase Rumah Tangga menurut Atap Terlulus Bukan Ijuk/Lainnya
X <sub>4</sub>	Persentase Rumah Tangga menurut Dinding Terlulus bukan Bambu/lainnya
X <sub>5</sub>	Persentase Rumah Tangga menurut Lantai Terlulus bukan Tanah
X <sub>6</sub>	Persentase Rumah Tangga menurut Sumber Air Minum Layak
X <sub>7</sub>	Persentase Rumah Tangga menurut Sanitasi Layak
X <sub>8</sub>	Persentase Rumah Tangga menurut Sumber Penerangan dari Listrik
X <sub>9</sub>	Persentase Rumah Tangga menurut Sumber Penerangan dari Listrik PLN
X <sub>10</sub>	Persentase Rumah Tangga menurut Luas Hunian per kapita < 7.2m <sup>2</sup>

Langkah-langkah analisis yang digunakan pada penelitian ini adalah:

1. Mengumpulkan data variabel penelitian dari X<sub>1</sub> hingga X<sub>10</sub>.
2. Melakukan analisis deskriptif terhadap variabel penelitian yang digunakan.
3. Melakukan pendeteksian pencilan atau *outlier* dengan menggunakan *Boxplot*.
4. Melakukan pengujian asumsi *cluster*, yaitu uji kecukupan sampel dengan menggunakan uji *Kaiser Mayer Olkin* (KMO) dan uji multikolinearitas dengan menggunakan nilai *Varian Inflation Factor* (VIF).
5. Menerapkan algoritma *K-Medoids* dengan tahapan:
  - a. Menentukan jumlah *cluster* sebanyak *k* dengan pertimbangan teoritis dan konseptual.
  - b. Mencari jarak tiap objek terhadap *cluster* terdekat dengan menggunakan ukuran *Euclidian distance*.

$$d_{ab} = \sqrt{\sum_{k=1}^p (x_{ak} - x_{bk})^2} \quad (1)$$

dengan:

- p = banyaknya variabel  
d<sub>ab</sub> = jarak antara objek a serta objek b  
x<sub>ak</sub> = nilai objek a dari variable k  
x<sub>bk</sub> = nilai objek b dari variable k

- c. Kemudian menentukan pusat *cluster* baru secara *random* untuk dijadikan calon *non medoids* pada masing-masing objek.
  - d. Menghitung jarak antara setiap objek pada tiap-tiap *cluster* terhadap calon *non medoids*.
  - e. Menghitung antara jumlah simpangan (S) dengan jumlah jarak baru – jumlah jarak lama. Apabila  $S < 0$ , maka mengganti objek dengan data *cluster non medoids* agar terbentuk kumpulan  $k$  objek baru yang disebut *medoids*.
  - f. Selanjutnya, mengulangi cara c – e hingga *medoids* tidak berubah. Kemudian diperoleh *cluster* dan masing-masing anggota *cluster*-nya.
6. Melakukan interpretasi dari hasil *cluster* yang didapatkan dengan melihat karakteristik dari setiap *cluster*-nya.
  7. Membuat visualisasi dari hasil *clustering* menggunakan *K-Medoids*.
  8. Membuat kesimpulan.

## HASIL DAN PEMBAHASAN

Pada bagian ini, akan diuraikan menjadi beberapa bagian, yaitu analisis deskriptif, uji asumsi *cluster*, menentukan hasil *cluster*, dan visualisasi hasil *clustering* dengan metode *K-Medoids*.

### Analisis Deskriptif

Analisis deskriptif dibutuhkan guna mengamati gambaran/rangkuman umum atas data yang diperoleh melalui melihat nilai minimum, nilai rata-rata dan nilai maksimum.

Tabel 1.  
Statistik Deskriptif

Variabel	Min	Max	Mean
X <sub>1</sub>	45.04	89.20	79.06
X <sub>2</sub>	2.15	37.71	9.64
X <sub>3</sub>	73.06	99.99	97.76
X <sub>4</sub>	68.28	100.0	97.70
X <sub>5</sub>	79.35	99.84	96.99
X <sub>6</sub>	62.47	99.84	85.41
X <sub>7</sub>	40.31	96.96	79.81
X <sub>8</sub>	73.84	100.0	97.79



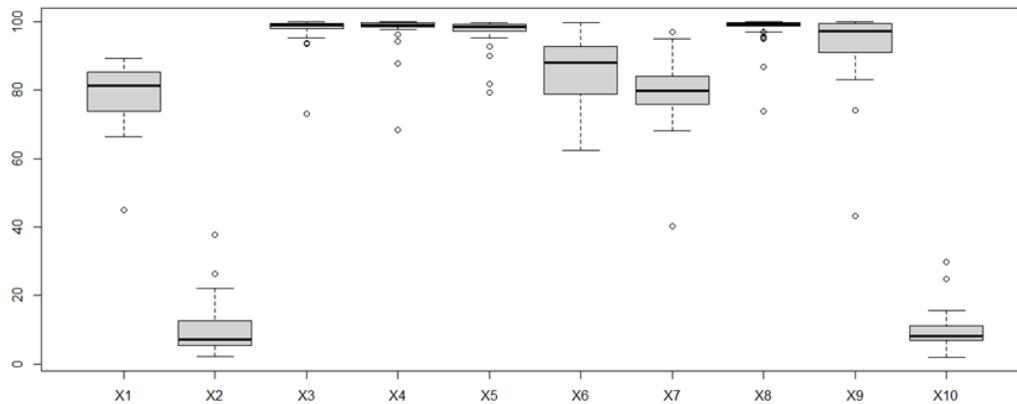
Variabel	Min	Max	Mean
X <sub>9</sub>	43.14	99.99	93.28
X <sub>10</sub>	1.99	29.92	9.59

Tabel 2 merupakan hasil statistik deskriptif. Berdasarkan Tabel 2 di atas, didapatkan bahwa pada variabel X, (persentase rumah tangga menurut status kepemilikan rumah milik sendiri) mempunyai nilai minimum sebesar 45.04, nilai maksimum sebesar 89.20 dan nilai rata-rata sebesar 79.06. Variabel X<sub>2</sub> (persentase rumah tangga menurut status kepemilikan rumah kontrak/sewa) mempunyai nilai minimum sebesar 2.15, nilai maksimum sebesar 37.71 dan nilai rata-rata sebesar 9.64. Variabel X<sub>3</sub> (persentase rumah tangga menurut atap terluas bukan ijuk/lainnya) mempunyai nilai minimum sebesar 73.06, nilai maksimum sebesar 99.99 dan nilai rata-rata sebesar 97.76. Variabel X<sub>4</sub>, (persentase rumah tangga menurut dinding terluas bukan bambu/lainnya) mempunyai nilai minimum sebesar 68.28, nilai maksimum sebesar 100.0 dan nilai rata-rata sebesar 97.70. Variabel X<sub>5</sub> (persentase rumah tangga menurut lantai terluas bukan tanah) mempunyai nilai minimum sebesar 79.35, nilai maksimum sebesar 99.98 dan nilai rata-rata sebesar 96.99. Variabel X<sub>6</sub> (persentase rumah tangga menurut sumber air minum layak) mempunyai nilai minimum sebesar 62.47, nilai maksimum sebesar 99.84 dan nilai rata-rata sebesar 85.41. Variabel X<sub>7</sub> (persentase rumah tangga menurut sanitasi layak) mempunyai nilai minimum sebesar 40.31, nilai maksimum sebesar 96.96 dan nilai rata-rata sebesar 79.81. Variabel X<sub>8</sub> (persentase rumah tangga menurut sumber penerangan dari listrik) mempunyai nilai minimum sebesar 73.84, nilai maksimum sebesar 100.0 dan nilai rata-rata sebesar 97.79. Variabel X<sub>9</sub>, (persentase rumah tangga menurut sumber penerangan dari listrik PLN) mempunyai nilai minimum sebesar 43.14, nilai maksimum sebesar 99.99 dan nilai rata-rata sebesar 93.28. Variabel X<sub>10</sub> (persentase rumah tangga menurut luas hunian per kapita < 7.2m<sup>2</sup>) mempunyai nilai minimum sebesar 1.99, nilai maksimum sebesar 29.92 dan nilai rata-rata sebesar 9.59.

### Pengecekan *Outlier*

Pada bagian ini, akan dilakukan pengecekan *outlier* terhadap data. Mendeteksi *outlier* memakai *boxplot* guna mengecek apakah ada data yang dipakai ada data *outlier*. Bila terdapat data yang atau titik yang keluar dari *boxplot* lalu data itu mempunyai data *outlier*.

Gambar 1.  
*Output Pengecekan Outlier Data.*



Dalam Gambar 1, terdapat titik-titik di luar rentang *boxplot* pada variabel X1, X2, X3, X4, X5, X6, X7, X9, dan X10, menunjukkan bahwa data yang digunakan memiliki nilai-nilai ekstrem atau *outlier*. Oleh karena itu, metode *clustering* yang lebih tepat adalah *Medoids clustering*. Dalam metode *K-Medoids*, pusat *cluster* adalah objek representatif yang disebut *medoids*. *Medoids* ini digunakan untuk menempatkan setiap objek ke dalam *cluster* yang memiliki *medoids* terdekat, sehingga metode ini cocok digunakan pada data yang mengandung *outlier*.

### Uji Asumsi

Dalam bagian ini, kami akan menguji asumsi analisis kluster, seperti Uji KMO untuk menilai apakah sampel sudah mencukupi untuk analisis kluster, serta Uji Multikolinearitas dengan menggunakan nilai VIF. Uji asumsi pertama adalah Uji KMO, yang menghasilkan nilai KMO sebesar 0,77 berdasarkan Tabel 3. Apabila nilai KMO berada dalam rentang 0,5 hingga 1, ini menunjukkan bahwa sampel mewakili populasi atau dapat dianggap sebagai sampel yang representatif. Dalam kasus ini, karena nilai KMO berada dalam rentang 0,5 hingga 1, dapat disimpulkan bahwa sampel yang digunakan dapat dianggap sebagai sampel yang mewakili populasi, dan dengan demikian, asumsi untuk sampel yang representatif terpenuhi (Ningrat et al., 2016).

Table 2.  
Hasil Uji KMO

Uji KMO
<i>Kaiser-Mayer-Olkin Measure of Sampling Adequacy = 0.516</i>

Langkah berikutnya adalah melakukan pengujian multikolinearitas dengan mengamati nilai VIF. Jika hasil VIF yang didapatkan kurang dari atau sama dengan 10, maka dapat disimpulkan bahwa tidak ada multikolinearitas, yang berarti bahwa asumsi telah terpenuhi (Mahmudan, 2020). Berikut ini adalah hipotesis yang diuji dalam pengujian multikolinearitas:

H0: Tidak ada keberadaan multikolinearitas

H1: Terdapat multikolinearitas

Tingkat signifikansi  $\alpha = 0.05$

Daerah penolakan:

H0 akan ditolak jika nilai VIF melebihi 10.

Tabel 3.  
Output Uji Multikolinieritas

Variabel	VIF
X <sub>1</sub>	10.718
X <sub>2</sub>	9.513
X <sub>3</sub>	10.815
X <sub>4</sub>	2.194
X <sub>5</sub>	6.653
X <sub>6</sub>	2.120
X <sub>7</sub>	4.372
X <sub>8</sub>	37.289
X <sub>9</sub>	18.140
X <sub>10</sub>	4.708

Berdasar Tabel 4 di atas, kita dapat menyimpulkan bahwa terdapat 6 variabel dengan nilai VIF yang kurang dari 10, yang menunjukkan bahwa tidak ada tanda-tanda multikolinieritas. Sementara itu, ada 4 variabel dengan nilai VIF yang lebih dari 10, yaitu X<sub>1</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>8</sub>, dan X<sub>9</sub>. Oleh karena itu, perlu menghapus keempat variabel tersebut. Setelah penghapusan variabel-variabel tersebut, kita dapat melihat hasil uji multikolinieritas dan uji KMO seperti yang terlihat pada Tabel 5. :

Tabel 4.  
Output Uji Multikolinieritas dan Uji KMO Setelah 4 Variabel dihilangkan

Variabel	VIF	KMO
X <sub>2</sub>	2.084	<i>Kaiser-Mayer-Olkin Measure of Adequacy = 0.516</i>
X <sub>4</sub>	1.724	
X <sub>5</sub>	2.710	
X <sub>6</sub>	1.508	
X <sub>7</sub>	3.630	

Variabel	VIF	KMO
X <sub>10</sub>	2.404	

Berdasar Tabel 5 di atas, dapat disimpulkan bahwa semua variabel memiliki nilai VIF di bawah 10. Oleh karena itu, tidak dapat menolak hipotesis nol (H<sub>0</sub>) yang menyatakan bahwa tidak ada multikolinieritas. Dengan tingkat kepercayaan 95%, data menunjukkan bahwa tidak ada dasar untuk menolak H<sub>0</sub>, yang berarti bahwa tidak ada korelasi antara keenam indikator perumahan dan lingkungan di Indonesia pada tahun 2020, berdasarkan Provinsi. Nilai KMO yang diperoleh sebesar 0.516, berada dalam rentang antara 0.5 hingga 1, yang mengindikasikan bahwa sampel ini mewakili populasi atau bersifat representatif. Dengan demikian, dapat disimpulkan bahwa semua asumsi telah terpenuhi.

### Penentuan Jumlah Cluster

Dalam studi ini, penentuan jumlah cluster atau  $k$  dilakukan menggunakan metode silhouette. Metode silhouette memanfaatkan nilai rata-rata untuk mengevaluasi kualitas cluster yang terbentuk (Anggara, et al., 2016), sehingga semakin tinggi nilai rata-ratanya, semakin baik kualitasnya. Jumlah  $s(i)$  diperhitungkan dengan menggabungkan  $a(i)$  dan  $b(i)$  (Dewi & Pramita, 2019).

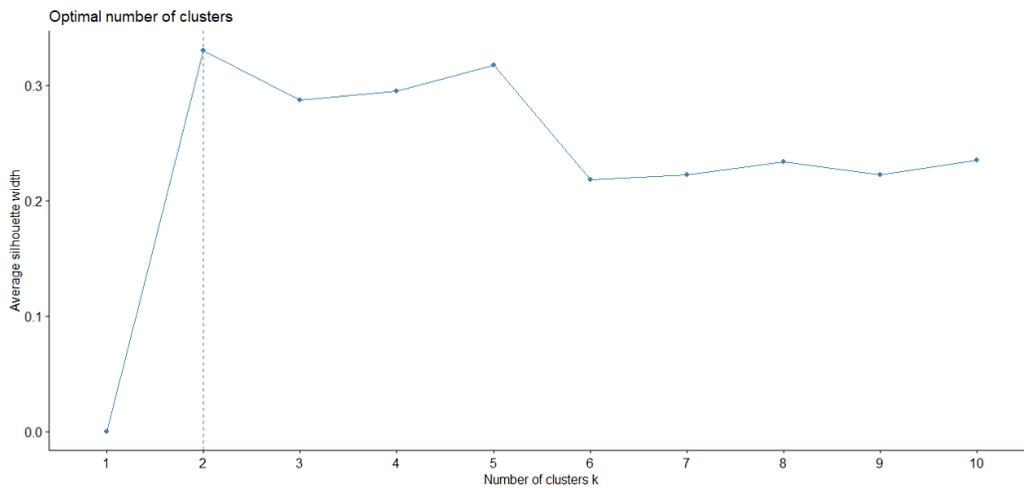
$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & \text{jika } a(i) < b(i) \\ 0 & \text{jika } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1 & \text{jika } a(i) > b(i) \end{cases} \quad (2)$$

Sehingga dapat dirumuskan sebagai berikut:

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

Berdasarkan Gambar 2, diperoleh banyaknya cluster optimum yaitu sebanyak 2 atau  $k = 2$ , hal ini dilihat dari garis putus-putus pada tampilan grafik metode silhouette, dimana garis putus-putus berada di  $k = 2$ , karena memiliki nilai Average Silhouette Width terbesar.

Gambar 2.  
Penentuan Jumlah Cluster Menggunakan Metode Silhouette



### Hasil Clustering

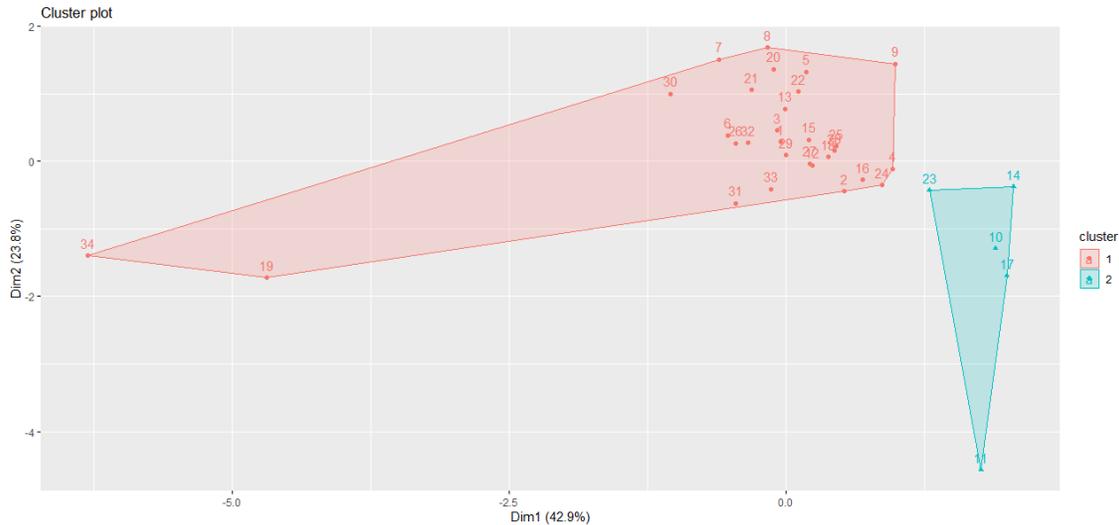
Anda dapat melihat hasil dari proses clustering dengan metode K-Medoids menggunakan Euclidean Distance sebagai metrik jarak. Dalam analisis ini, terbentuk dua kelompok (cluster), dan terdapat 34 data yang digunakan. Tabel 6 menunjukkan bahwa dari 34 data provinsi yang diamati, 29 di antaranya termasuk dalam cluster 1, sementara 5 provinsi sisanya masuk ke dalam cluster 2.

Tabel 5.  
Hasil Clustering Menggunakan Metode *K-Medoids*

Cluster	Provinsi	Jumlah
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua.	29
2	Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, dan Kalimantan Timur.	5

Berdasarkan hasil dari *clustering* pada Tabel 6, kemudian dapat dilihat menggunakan *cluster plot* seperti pada Gambar 3, dimana didapatkan bahwa *cluster 1* area warna merah dan *cluster 2* area warna hijau.

Gambar 3.  
Hasil Cluster Plot Menggunakan K-Medoids



Lalu, guna mengamati karakteristik setiap *cluster* adalah melihat nilai rata-rata dari tiap indikator *cluster*-nya. Karakteristik dari masing-masing *cluster* bisa diamati di Tabel 7.

Tabel 6.  
Karakteristik dari Masing-Masing *Cluster*

Indikator	Cluster 1	Cluster 2
X <sub>2</sub>	7.22	<b>23.7</b>
X <sub>4</sub>	97.3	<b>99.8</b>
X <sub>5</sub>	96.6	<b>99.2</b>
X <sub>6</sub>	84.0	<b>93.8</b>
X <sub>7</sub>	77.6	<b>92.7</b>
X <sub>10</sub>	9.37	<b>10.9</b>

Karakteristik setiap indikator perumahan dan kesehatan lingkungan berdasarkan Tabel 7, didapatkan bahwa *cluster 1* memiliki indikator perumahan dan kesehatan lingkungan yaitu variabel X<sub>2</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub> dan X<sub>10</sub> dengan nilai rata-rata indikator

rendah, sementara *cluster* 2 memiliki indikator perumahan dan kesehatan lingkungan yaitu variabel  $X_2$ ,  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$  dan  $X_{10}$  dengan nilai rata-rata indikator tinggi.

### Visualisasi Hasil *Clustering* Menggunakan Metode *K-Medoids*

Hasil *clustering* yang didapatkan pada Tabel 6, secara visualisasi dapat dilihat pada Gambar 4. Visualisasi dibuat untuk membuat mudah mengamati hasil *clustering* provinsi di Indonesia berdasarkan indikator perumahan dan kesehatan lingkungan pada tahun 2020.

Gambar 4.

Visualisasi Hasil *Clustering* Provinsi di Indonesia Berdasarkan Indikator Perumahan dan Kesehatan Lingkungan tahun 2020.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Cluster. The marks are labeled by Provinsi.

## KESIMPULAN

Atas dasar hasil analisis di atas, bisa ditarik simpulan dengan memakai metode *K-Medoids Clustering* diperoleh bahwa *cluster* optimal yang didapatkan dengan menggunakan metode *silhouette* adalah sebanyak 2 *cluster*. Pada *cluster* 1 terdapat 29 Provinsi, sementara pada *cluster* 2 terdapat 5 Provinsi. *Cluster* 1 memiliki karakteristik persentase rumah tangga menurut kepemilikan rumah kontrak/sewa, dinding terluas bukan bambu/lainnya, lantai terluas bukan tanah, sumber, air minum layak, sanitasi layak, luas hunian per kapita  $< 7.2m^2$  dengan nilai rata-rata indikator rendah. Sementara *cluster* 2 memiliki karakteristik persentase rumah tangga menurut kepemilikan rumah kontrak/sewa, dinding terluas bukan bambu/lainnya, lantai terluas bukan tanah, sumber, air minum layak, sanitasi layak, luas hunian per kapita  $< 7.2m^2$  dengan nilai rata-rata indikator tinggi.



Pada penelitian ini hanya fokus menggunakan jarak *Euclidean* untuk mendapatkan jarak objek, untuk penelitian selanjutnya dapat menggunakan pengukuran jarak yang lain seperti Manhattan. Selain itu, dapat juga melakukan analisis *cluster* dengan metode yang lebih *robust* terhadap *outlier* seperti *Clustering Large Application* (CLARA) dan dapat membandingkan hasilnya dengan metode *K-Medoids*.

## DAFTAR PUSTAKA

- Aurora, P., Deepali & Varshney, S. 2016. Analysis of K-Means and K-Medoids Algorithm for Big Data. *Procedia Computer Sains* Vol. 78 (Hal. 507-512).
- Basri, F.H. 1995. *Perekonomian Indonesia Menjelang Abad XII: Distorsi, Peluang dan Kendala*. Jakarta: Erlangga.
- Badan Pusat Statistika. 2020. *Indikator Perumahan dan Kesehatan Lingkungan 2020*. Jakarta: Badan Pusat Statistika.
- Ghaisani, S.Y., Hikmah, Nur., Prasetyo, A.H., & Widodo, Edy. 2016. Analisis Cluster Hirarki untuk Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016. *Prosiding-M25*. P-ISSN: 2502-6526, E-ISSN:2656-0615.
- Ningrat, F.R., Maruddani, D.A.I, & Wuryandari, T. 2016. Analisis Cluster dengan Algoritma K-Means dan Fuzzy C-Means Clustering untuk Pengelompokkan Data Obligasi Korporas. *Jurnal Gaussian* Vol. 5, No. 4 (Hal. 641-650).
- Wira, B., Budianto, A.E., & Wiguna, A.S. 2019. Impelmentasi Metode K-Medoids Clustering Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi* Vol. 1, No. 3 (Hal. 54-69).