ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 13, Nomor 1, Tahun 2024, Halaman 13 - 24
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

ANALISIS PERBANDINGAN *SILHOUETTE COEFFICIENT* DAN METODE *ELBOW* PADA PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR IPM DENGAN *K-MEDOIDS*

Tias Rahmawati[1*], Yuciana Wilandari[2], Puspita Kartikasari[3]
[1,2,3]Jurusan Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail**: tiasrahmawati01507@gmail.com

DOI: 10.14710/j.gauss.13.1.13-24

Article Info:
Received: 2023-04-03
Accepted: 2023-08-08
Available Online: 2024-08-16

Keywords:
Human Development Index; KMedoids; Silhouette Coefficient; Elbow Method; Davies Bouldin Index.

Abstract: Development is a continuous process that aims to improve all aspects of life based on the prevailing societal values and predetermined life goals. The quality of life in a community is evaluated at the provincial level using the Human Development Index (HDI) based on three key indicators: economic status, health, and education. K-medoids is a method for grouping objects that may contain outliers. To determine the optimal number of clusters, the Silhouette Coefficient method is used, which assesses the degree of proximity between objects and the distance between clusters. The Elbow method is used to determine the optimal number of clusters by analyzing the percentage outcomes of different cluster quantities at a specific elbow point. The Davies Bouldin Index (DBI) is used to compare and validate the resulting clusters. The study revealed that the Silhouette Coefficient method yielded the best cluster with 3 clusters and a Silhouette Coefficient value of 0,3129, and a DBI value of 1,3184. However, using the Elbow method, the best cluster was found to be 4 clusters with an SSE value of 54,5548 and a DBI value of 1,1754. Thus, the optimal cluster configuration is 4 clusters with the Elbow method having the smallest DBI value.

1. PENDAHULUAN

Pembangunan ialah suatu proses perubahan yang berlangsung secara berkelanjutan guna meningkatkan semua bidang kehidupan dengan memperhatikan nilai-nilai yang diyakini oleh masyarakat untuk meraih tujuan hidup yang telah ditetapkan. Pembangunan yang difokuskan pada potensi, daya kreasi, inisiatif, dan kepribadian setiap masyarakat. Pandemi Covid-19 memasuki negara Indonesia dimulai awal tahun 2020 yang berdampak pada seluruh aspek bidang kehidupan salah satunya bidang ekonomi yang mengalami tekanan berat seperti IPM. Menurut Badan Pusat Statistik (2021), pada tahun 2020 IPM mengalami keterlambatan dalam laju pertumbuhan sebesar 0,03% daripada tahun sebelumnya sebesar 0,74%. Faktor yang menyebabkan terjadinya keterlambatan pada laju pertumbuhan yaitu adanya penurunan kualitas hidup yang memadai dinilai dari biaya per kapita yang disesuaikan sedangkan parameter yang lain mengalami peningkatan meskipun lambat. Namun pada tahun 2022 IPM kembali meningkat sebesar 0,86% dari tahun 2021. Menurut Badan Pusat Statistik (2022), kenaikan IPM tahun 2022 terjadi di seluruh provinsi Indonesia dengan perbandingan nilai yang tidak terlalu besar. Rata-rata peningkatan IPM dari tahun 2010 sampai dengan tahun 2022 sebesar 0,77% per tahun. Daerah di Indonesia mempunyai karakteristik dan geografis berbeda yang berpengaruh terhadap pembangunan manusianya. Selain itu, pembangunan manusia dipengaruhi oleh pertumbuhan ekonomi. Pembangunan manusia bersifat jangka panjang sedangkan pertumbuhan ekonomi tidak sepenuhnya sebanding dengan pembangunan manusia. Dalam satu dekade terakhir Indonesia mengalami pembangunan manusia yang kurang merata atau mengalami penurunan antar wilayah. Upaya untuk meningkatkan pembangunan manusia diperlukan cara yang tepat dan terkoordinasi serta mendukung program bidang satu sama

lain. Pandemi Covid-19 memberikan pengaruh bagi seluruh aspek bidang secara tidak langsung menjadi tantangan bagi Indonesia dalam menghadapi kondisi tersebut.

Menurut Supranto (2004), analisis klaster merupakan analisis yang menggabungkan objek yang memiliki karakeristik sama. Ukuran jarak digunakan untuk mengukur karakteristik, dimana semakin besar jarak yang dihasilkan menunjukkan bahwa objek semakin menjauhi pusat klaster. Dalam menganalisis metode yang digunakan pada analisis klaster salah satunya *K-Medoids* atau *Partitioning Around Medoids*. Han dan Kamber (2006) menyatakan bahwa *K-medoids* merupakan suatu teknik pengelompokan objek yang digunakan sebagai solusi mengatasi adanya pencilan. Dalam hal ini, K-medoids dianggap lebih unggul daripada metode *K-Means* yang lebih mudah dipengaruhi oleh adanya pencilan. Metode yang dipakai untuk menghitung jumlah klaster adalah menggunakan metode *Silhouette Coefficient* dan metode *Elbow*. Menurut Rousseeuw (1987) menyatakan bahwa *Silhouette Coefficient* merupakan metode yang efektif untuk menentukan jumlah klaster terbaik karena metode ini dapat memperkirakan seberapa dekat data di suatu kelompok dan selisih antar kelompok yang berlainan. Menurut Madhulata (2012), keunggulan dari metode *Elbow* bekerja dengan menghitung persentase perbandingan antara jumlah klaster yang membentuk titik siku untuk menentukan jumlah klaster terbaik.

2. TINJAUAN PUSTAKA

Menurut UNDP (*United Nations Development Programme*) 1990 dalam Badan Pusat Statistik (2021), pembangunan manusia merupakan pilihan atau keputusan pemerintah dalam skala besar yang diproses sehingga terlahirnya kebijakan dengan tujuan utama pembangunan. Perhitungan Indeks Pembangunan Manusia (IPM) didasarkan oleh empat data komponen sebagai gambaran yaitu, angka melek huruf, angka harapan hidup, pendapatan yang disesuaikan, dan rata-rata lama sekolah.

Menurut Rousseeuw dan Leroy (1996), mendeteksi pencilan berfungsi untuk mengindentifikasi sistem yang berbeda atau salah dari data lainnya sehingga berakibat fatal. Terdapat dua jenis pencilan yaitu univariat dan multivariat. Analisis klaster menggunakan pencilan multivariat sebagai alat untuk mendeteksi pencilan. Mendeteksi pencilan menggunakan kuadrat jarak Mahalanobis (d_{MD}^2) pada objek ke-i dengan pusat data menggunakan rumus sebagai berikut:

$$d_{MD}^2(i) = (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^{\boldsymbol{T}}\, \boldsymbol{\Sigma^{-1}} (\boldsymbol{x_i} - \overline{\boldsymbol{x}}),\ i = 1,2,\dots,n \qquad (1)$$

dengan $d_{MD}^2(i)$ adalah nilai kuadrat jarak mahalanobis objek ke-i dengan pusat data, $\boldsymbol{x_i}$ adalah objek pengamatan ke-i, $\overline{\boldsymbol{x}}$ adalah rata-rata variabel, dan $\boldsymbol{\Sigma^{-1}}$ adalah matriks kovarian dari variabel.

Analisis klaster adalah metode yang digunakan dalam analisis multivariat dengan tujuan sebagai metode pengelompokan objek-objek sesuai dengan karakteristik yang dimilikinya. Sehingga objek yang memiliki persamaan karakteristik atau mendekati persamaan akan dikelompokkan di dalam klaster yang sama.

Menurut Gujarati (2009 multikolinearitas dijelaskan sebagai hubungan linier lengkap atau benar-benar ada di antara satu atau beberapa variabel dalam satu set data. Salah satu metode untuk mengindetifikasi terjadinya multikolinearitas yaitu dengan mencari besarnya nilai *Varians Inflation Factor* (VIF) menggunakan Persamaan di bawah ini:

$$VIF_l = \frac{1}{(1-R_l^2)}\ ,\ l = 1, 2, \dots, \text{p} \qquad (2)$$

dengan R_l^2 adalah koefisien korelasi.

Analisis klaster merupakan pengukuran objek yang dilakukan berdasarkan adanya kemiripan antar objek satu sama lain. Sehingga dalam analisis klaster, mengukur

persamaan antar objek merupakan hal yang penting. Jarak *Euclidean (Euclidean Distance*) adalah suatu jarak yang digunakan dalam perhitungan jarak antar objek dengan cara menjumlahkan kuadrat selisih nilai dari setiap variabel objek, kemudian diambil akar kuadrat dari hasil penjumlahan tersebut. Menurut Anderberg (1973) menjelaskan bahwa terdapat formula untuk menghitung jarak *Euclidean* yang dapat digunakan, yaitu sebagai berikut:

$$d_{i,j} = \sqrt{\sum_{l=1}^{p}\left(x_{il} - x_{jl}\right)^2},\ i = 1,2,3,\ldots,n \text{ dan } j = 1,2,3,..,n \qquad (3)$$

dengan $d_{(i,j)}$ merupakan selisih antar objek i dengan objek j, x_{il} merujuk pada besarnya objek i dalam variabel ke-l, x_{jl} merujuk pada besarnya objek j dalam variabel ke-l, dan p merepresentasikan jumlah variabel.

K-Medoids adalah metode serupa dengan *K-Means* karena memiliki tujuan untuk mengurangi sensitivitas terhadap nilai ekstrim pada objek yang diperoleh dari kumpulan data, seperti yang dijelaskan oleh Vercellis (2009). Han dan Kamber (2006) menyatakan bahwa langkah-langkah dalam Partitioning Around Medoids adalah sebagai berikut:

1. Pada tahap awal klasterisasi, dilakukan inisialisasi pusat klaster sebanyak K
2. Memilih acak k medoid awal dari total n data yang tersedia.
3. Perhitungan jarak sementara antara setiap objek dengan medoid yang dipilih.
4. Selanjutnya, objek yang memiliki jarak terdekat dengan medoid ditandai dan total jaraknya dihitung menggunakan rumus jarak Euclidean.
5. Langkah ke-5 dari metode *partitioning around medoids* adalah memilih objek secara acak pada setiap klaster sebagai medoid kandidat baru, kemudian menghitung jarak antara setiap objek pada klaster dengan medoid kandidat baru tersebut.
6. Langkah ke-6 dalam metode *partitioning around medoid* dilakukan dengan menghitung total selisih jarak antara medoid awal dan medoid baru. Untuk melakukan hal ini, rumus yang digunakan adalah S = b - a, dalam hal ini variabel a mengindikasikan total jarak terpendek antara objek dengan medoid awal, sedangkan variabel b menunjukkan total jarak terpendek antara objek dengan medoid yang baru diganti. Nilai S kurang dari atau sama dengan 0, maka kembali ke langkah ke-2 dan metode berhenti ketika nilai S lebih besar dari 0.

Setelah klaster terbentuk, metode *Silhouette Coefficient* dan metode *Elbow* berfungsi sebagai penentu total klaster yang optimal. Metode *Silhouette Coefficient* berperan dalam menentukan kualitas dan kekuatan klaster yang terbentuk. Menurut Handoyo dkk (2014) perhitungan *Silhouette Coefficient* terdapat beberapa tahap yaitu:

a. Menghitung nilai rata-rata jarak antara suatu objek tertentu, misalnya objek i, dimana seluruh objek lainnya berada di klaster yang sama dengan menggunakan Persamaan di bawah ini:

$$a(i) = \frac{1}{|A|-1}\sum_{j\in A, j\neq 1} d_{i,j} \qquad (4)$$

dengan $|A|$ ialah banyaknya objek di klaster A, i, j adalah indeks dari objek, dan $d_{i,j}$ ialah selisih antar objek ke-i dengan objek ke-j.

b. Untuk menghitung rata-rata jarak objek i ke seluruh objek di klaster lain, digunakan rumus sebagai berikut:

$$d_{i,A'} = \frac{1}{|A'|}\sum_{j\in A'} d_{i,j} \qquad (5)$$

dengan $d_{i,A'}$ ialah jarak rata-rata objek i dengan seluruh objek di klaster lain dan A'adalah banyaknya objek di klaster selain A.

c. Menentukan atau memilih nilai jarak terkecil atau paling minimum, misalnya $b(i)$:

$$b(i) = \min_{A' \neq A} d_{i,A'} \quad (6)$$

d. Menghitung *Silhoutte Coefficient*

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (7)$$

$$SC = \sum_{i-1}^{n} s(i) \quad (8)$$

Silhouette Coefficient adalah suatu metrik yang menunjukkan sejauh mana data dalam sebuah kelompok memiliki kesamaan, dan dihitung secara individual untuk setiap objek dalam kelompok. Semakin mendekati angka 1 pada *Silhouette Coefficient*, maka semakin tinggi kualitas pengelompokkan pada satu kelompok tersebut. Sebaliknya, semakin mendekati nilai -1 pada *Silhouette Coefficient*, maka semakin buruk kualitas pengelompokkan pada satu klaster tersebut. Menurut Rousseeuw (1987), kriteria pengukuran *Silhouette Coefficient* sebagai berikut:

Tabel 1. Ukuran *Silhouette Coefficient*

Silhouette Coefficient	**Intepretasi yang disesuaikan**
0,7 < SC ≤ 1,0	Susunan sangat baik
0,5 < SC ≤ 0,7	Susunan baik
0,25 < SC ≤ 0,5	Susunan lemah
SC ≤ 0,25	Susunan buruk

Menurut Madhulata (2012) dalam Merliana & Santoso (2015) menyatakan bahwa metode *Elbow* adalah metode yang berfungsi untuk mendapatkan informasi dengan tujuan sebagai penentu klaster terbaik menggunakan cara menganalisis hasil persentase dari perbandingan banyaknya klaster berbentuk titik-titik siku. Penentuan klaster yang paling optimal, dapat dilakukan dengan perhitungan besarnya *Sum of Square Error* (SSE) pada setiap klaster. Rumus *Sum of Square Error* (SSE) untuk klaster ke-t (SSE_t) dan seluruh klaster (SSE_{total}) sebagai berikut (Irwanto, dkk 2012) :

$$SSE_k = \sum_{l=1}^{p} \sum_{i=1}^{m_k} |x_i - \bar{x}_{kl}|^2 \quad (9)$$

$$SSE_{total} = \sum_{k=1}^{K} SSE_k \quad (10)$$

dengan x_i adalah objek ke-i dan $\bar{x}_{kl}$ adalah central pada klaster ke-k variabel ke -l.

Menurut Dewi dan Pramita (2019) , *Davies Bouldin Index* (DBI) merupakan suatu teknik untuk mengevaluasi atau mengendalikan hasil klastering. Dalam teknik ini, klaster yang paling baik adalah klaster yang memiliki nilai DBI paling kecil. Cara mencari nilai *Davies Bouldin Index* digunakan Persamaan di bawah ini:

1. Persamaan *Sum of square within cluster* (SSW) berfungsi untuk mengetahui tingkat kohesi ke-t.

   $$SSW_k = \frac{1}{m_k} \sum_{i=1}^{m_k} d_{x_i,c_k} \quad (11)$$

   dengan m_k adalah banyaknya objek di klaster ke-k, $c_k$adalah centroid pada klaster ke-k, dan $d_{(x_i,c_k)}$adalah jarak euclidean setiap objek ke centroid.
2. *Sum of Square Between cluster* (SSB) digunakan sebagai alat ukur seberapa jauh setiap klaster berada satu sama lainnya dalam pengelompokan data.

   $$SSB_{k,t} = d_{(c_k,c_t)} \quad (12)$$

   dengan $d_{(c_k,c_t)}$ adalah jarak antar centroid klaster k dan klaster t.
3. Dilakukan perhitungan rasio ($R_{k,t}$) untuk menentukan besar kecilnya nilai antara ke-k dan klaster ke-t.

   $$R_{k,t} = \frac{SSW_k + SSW_t}{SSB_{kt}} \quad (13)$$

dengan $R_{k,t}$ adalah rasio antar klister, SSW_k adalah *Sum of Square Within Cluster k*, SSW_t adalah *Sum of Square Within Cluster* t, dan SSB_{kt} adalah separasi pada klaster *k* dan *t*.

4. Perhitungan untuk memperoleh nilai *Davies Bouldin Index* (DBI) sebagai berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^{K} \max_{k \neq t} R_{k,t} \quad (14)$$

dengan *K* adalah banyaknya klaster yang digunakan dan $R_{k,t}$ adalah rasio antara klaster *u* dan *t.*

3. METODE PENELITIAN

Informasi yang diambil sebagai dasar penelitian ini termasuk dalam jenis data sekunder berkaitan dengan Indeks Pembangunan Manusia (IPM) Tahun 2022 pada 34 provinsi di Indonesia. Data tersebut dirilis oleh Badan Pusat Statistik RI pada tahun yang sama dan disediakan oleh Badan tersebut. Variabel yang dipakai dalam penelitian yaitu indikator untuk menghitung IPM di Indonesia. Terdapat empat variabel yang digunakan, yaitu: Y_1 = yang meliputi umur harapan hidup saat lahir; Y_2 = yang meliputi harapan lama sekolah; Y_3 = yang meliputi rata-rata lama sekolah; Y_4 = yang meliputi pengeluaran perkapita disesuaikan.

Metode yang dipakai dalam analisis klaster dalam penelitian ini yaitu metode *Partitinoning Around K-medoid* atau *K-medoids* dengan menggunakan jarak *euclidean* sebagai pengukuran jaraknya serta penerapan metode *Elbow* dan *Silhouette Coefficent* sebagai pembanding kualitas klaster. Terkait dengan penelitian ini, dilakukan pengelolaan data dengan menerapkan serta mengoperasikan beberapa perangkat lunak, seperti SPSS, *Microsoft Excel*, dan *R Studio* 4.1.3. Berikut adalah tahapan-tahapan analisis yang dilakukan:

1. Memasukkan objek ke dalam *software* yang digunakan sebagai alat bantu
2. Melakukan standarisasi seluruh variabel
3. Mendeteksi objek terhadap pencilan menggunakan kuadrat jarak *mahalanobis*
4. Menguji asumsi untuk mendeteksi multikolinearitas dengan nilai *Varians Inflations Factor (VIF)*.
5. Menganalisis klaster menggunakan metode *K-Medoids*, sebagai berikut :
   1) Pada tahap awal klasterisasi, dilakukan inisialisasi pusat klaster sebanyak K
   2) Memilih acak k medoid awal dari total n data yang tersedia.
   3) Perhitungan jarak sementara antara setiap objek dengan medoid yang dipilih.
   4) Selanjutnya, objek yang memiliki jarak terdekat dengan medoid ditandai dan total jaraknya dihitung menggunakan rumus jarak Euclidean.
   5) Langkah ke-5 dari metode *partitioning around medoids* adalah memilih objek secara acak pada setiap klaster sebagai medoid kandidat baru, kemudian menghitung jarak antara setiap objek pada klaster dengan medoid kandidat baru tersebut.
   6) Langkah ke-6 dalam metode *partitioning around medoid* dilakukan dengan menghitung total selisih jarak antara medoid awal dan medoid baru. Untuk melakukan hal ini, rumus yang digunakan adalah S = b - a, dalam hal ini variabel a mengindikasikan total jarak terpendek antara objek dengan medoid awal, sedangkan variabel b menunjukkan total jarak terpendek antara objek dengan medoid yang baru diganti. Nilai S kurang dari atau sama dengan 0, maka kembali ke langkah ke-2) dan metode berhenti ketika nilai S lebih besar dari 0.
6. Melakukan validasi klaster untuk menentukan jumlah klaster optimal menggunakan metode *Elbow* dan metode *Silhouette Coefficient*

7. Melakukan evaluasi klaster dari hasil klaster yang telah didapatkan menggunakan dua langkah, sebagai berikut :
   a. Menghitung nilai *Davies Bouldin Index* pada metode *Elbow* dan *Shilhouette Coefficient.*
   b. Melakukan perbandingan dari hasil perhitungan *Davies Bouldin Index* antara metode *Elbow* dan *Silhouette Coefficient.*
8. Mengintepretasi hasil dari karakteristik daerah provinsi berdasarkan perbandingan hasil yang terbaik dari evaluasi klaster antara metode *Elbow* dan *Silhouette Coefficient*.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data indikator indeks pembangunan manusia pada seluruh provinsi di Indonesia yang disediakan oleh Badan Pusat Statistik RI Tahun 2022. Data tersebut terdiri dari empat indikator, yaitu harapan hidup saat lahir (Y_1), harapan lama sekolah (Y_2), rata-rata lama sekolah (Y_3), dan pengeluaran perkapita disesuaikan (Y_4).

Tabel 2. Data Indikator Indeks Pembangunan Manusia

Objek ke-	Provinsi	Y_1	Y_2	Y_3	Y_4
1	Aceh	70,180	14,370	9,440	9963
2	Sumatera Utara	69,610	13,310	9,710	10848
3	Sumatera Barat	69,900	14,100	9,180	11130
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
34	Papua	66,230	11,140	7,020	7146

Berdasarkan Tabel 2, dapat diketahui nilai empat variabel atau indikator indeks pembangunan manusia dari Y_1 sampai Y_4 pada 34 provinsi di Indonesia yang dimulai dari Provinsi Aceh, Provinsi Sumatera Utara, sampai Provinsi Papua. Statistika deskriptif digunakan untuk memberikan gambaran karakteristik dari setiap indikator atau variabel yang digunakan mencakup nilai minimum, standar deviasi, rata-rata (*mean*), *range*, dan maksimum

Dari informasi yang terdapat pada Tabel 3, didapatkan ciri-ciri data dari setiap variabel atau indikator yang dipakai pada penelitian ini. Variabel Y_1, yaitu provinsi dengan indikator umur harapan hidup saat lahir, memiliki angka minimal sebesar 66,230 yang terdapat di Provinsi Papua, angka rata-rata (*mean*) sebesar 70,420, dan angka maksimal sebesar 75,080 yang terdapat di Provinsi DI Yogyakarta. Variabel Y_2, yaitu persentase provinsi dengan indikator harapan lama sekolah, memiliki angka minimal sebesar 11,140 yang terdapat di Provinsi Papua, angka rata-rata (*mean*) sebesar 13,240, dan angka maksimal sebesar 15,650 yang terdapat di Provinsi DI Yogyakarta. Variabel Y_3, yaitu persentase provinsi dengan indikator rata-rata lama sekolah, memiliki angka minimal sebesar 7,020 yang terdapat di Provinsi Papua, angka rata-rata (*mean*) sebesar 8,839, dan angka maksimal sebesar 11,310 yang terdapat di Provinsi DKI Jakarta. Variabel Y_4, yaitu persentase provinsi dengan indikator pengeluaran per kapita, memiliki angka minimal sebesar 7146 yang terdapat di Provinsi Papua, angka rata-rata (*mean*) sebesar 11080, dan angka maksimal sebesar 18927 yang terdapat di Provinsi DKI Jakarta.

Tabel 3. Statistika Deskriptif Variabel Pengamatan

Variabel	Minimal	*Mean*	Maksimal	*Standar Deviasi*	*Range*

Y_1	66,230	70,420	75,080	2,415	8,850
Y_2	11,140	13,240	15,650	0,730	4,510
Y_3	7,020	8,839	11,310	0,909	4,290
Y_4	7146	11080	18927	2213	11781

Standarisasi data digunakan untuk menyetarakan skala pada variabel yang berbeda. Menghitung standarisasi menggunakan nilai *Z score* yang didapatkan dari nilai objek dikurangkan nilai rata-rata (*mean*) variabel dan membaginya dengan nilai standar deviasi. Berdasarkan nilai standarisasi empat variabel atau indikator indeks pembangunan manusia dari X_1 sampai X_4 pada 34 provinsi di Indonesia yang dimulai dari Provinsi Aceh , Provinsi Sumatera Utara, sampai Provinsi Papua.

Untuk mendeteksi pencilan dalam penelitian ini, dilakukan perhitungan jarak setiap nilai objek ke pusat data (rata-rata dari semua objek) menggunakan jarak kuadrat Mahalanobis dengan menggunakan Persamaan (1). Setelah perhitungan tersebut, ditemukan empat provinsi yang nilai jarak kuadrat Mahalanobisnya melebihi nilai kritis distribusi *chi-kuadrat* ($\chi^2_{0,95;4}$=9,487729), yaitu Provinsi DKI Jakarta dengan nilai jarak kuadrat Mahalanobis sebesar 15,9070, Provinsi D.I. Yogyakarta dengan nilai jarak kuadrat Mahalanobis sebesar 14,1669, Provinsi Maluku dengan nilai jarak kuadrat Mahalanobis sebesar 9,7599, dan nilai jarak kuadrat Mahalanobis Papua adalah 9,8560. Sehingga provinsi tersebut bersama dengan dua provinsi lainnya dianggap sebagai pencilan. Disimpulkan bahwa data indikator indeks pembangunan manusia mengandung pencilan.

Untuk menguji asumsi non-multikolinearitas dalam penelitian ini pada Tabel 4, dilakukan perhitungan besarnya *Variance Inflation Factor* (VIF) di variabel pengamatan dengan menggunakan Persamaan (2) bahwa nilai VIF pada variabel X_1 sampai X_4 tidak ada yang melebihi angka 10. Disimpulkan bahwa data indikator indeks pembangunan manusia telah memenuhi asumsi non-multikolinearitas.

Tabel 4. Nilai VIF

Variabel	Koefisien Determinasi	*VIF*
X_1	0,3888	1,569705
X_2	0,2345	1,310339
X_3	0,4782	1,896285
X_4	0,5225	1,986897

Tabel 5 menunjukkan hasil pengklasteran menggunakan metode *K-Medoids* pada *k*=2,3,4,5, dan 6. Objek medoid dan anggota setiap klaster dapat diketahui melalui proses running sintaks yang dilakukan.

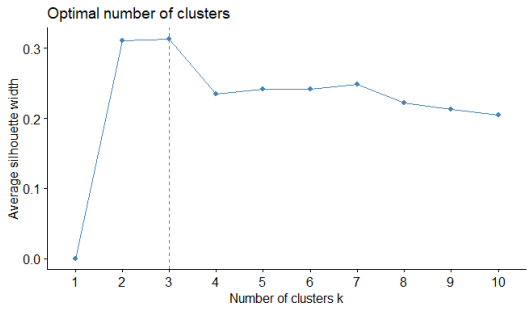
Tabel 5. Jumlah Klaster Dari Proses Klasterisasi *k*=2, 3, 4, 5, dan 6

K	Kelompok ke-	Jumlah Objek	Medoid
2	1	28	Objek ke- 5
	2	6	Objek ke- 33
3	1	23	Objek ke-5
	2	5	Objek ke-17
	3	6	Objek ke-33

4	1	10	Objek ke- 7
	2	14	Objek ke- 5
	3	5	Objek ke - 17
	4	5	Objek ke - 33
5	1	10	Objek ke-7
	2	14	Objek ke-5
	3	4	Objek ke-17
	4	1	Objek ke -14
	5	5	Objek ke -33
6	1	10	Objek ke -7
	2	14	Objek ke -5
	3	3	Objek ke -17
	4	1	Objek ke-11
	5	1	Objek ke -14
	6	5	Objek ke-33

Pada penelitian Tugas Akhir ini menggunakan dua metode untuk menentukan jumlah klaster teroptimal yaitu *Silhouette Coefficient* dan metode *Elbow.*

a. Metode *Silhoutte Coefficient*
Dari hasil grafik *output* menggunakan metode *Silhoutte Coefficient* yang tertera pada Gambar 1.



Gambar 1. Grafik *Silhoutte Coefficient*

Dari analisis grafik *Silhouette Coefficient*, ditemukan jumlah klaster optimal yang terbentuk adalah tiga klaster (*k=3*). Kesimpulan ini didasarkan pada nilai *Silhouette Coefficient* pada *k=3* lebih besar daripada nilai rata-rata pada k yang lain. Melalui penggunaan perangkat lunak *Rstudio 4.1.3 dan Microsoft Excel*. Berikut adalah nilai *Silhouette Coefficient* yang diperoleh :

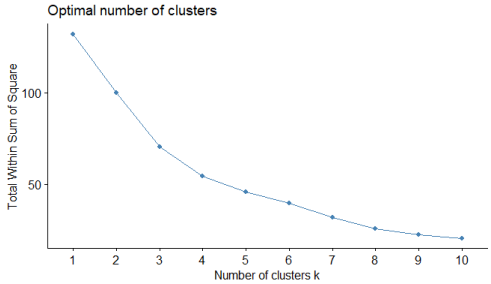
Tabel 6. Hasil Akhir Perhitungan *Silhouette Coefficient*

Total Klaster	**Nilai *Silhouette Coefficient***
2	0,3108
3	0,3129
4	0,2348
5	0,2419
6	0,2621

Dari Tabel 6. dapat diambil kesimpulan bahwa jumlah klaster paling optimal adalah *k*=3 dengan nilai *Silhoutte Coefficient* sebesar 0,3129 yang lebih besar daripada nilai *Silhouette Coefficient* jumlah klaster lainnya.

b. Metode Elbow

Diperoleh grafik output menggunakan metode *Elbow* yang tertera pada Gambar 3.



Gambar 2. Grafik Menggunakan Metode *Elbow*

Berdasarkan pada grafik metode Elbow diperoleh jumlah klaster optimal berada di *k*=4. Hal ini dikarenakan pada *k*=4 terjadinya penurunan drastis atau membentuk siku daripada lainnya. Sehingga diperoleh pehitungan SSE pada seluruh klaster sebagai berikut :

Tabel 7. Hasil Perhitungan SSE

Klaster	*Sum of Square Error*
2	119,9209
3	70,5426
4	54,5548
5	45,8322
6	35,0426

Dari output pada grafik menggunakan metode *Elbow* titik yang membentuk siku berada pada titik (*k*=4) dan hasil perhitungan *Sum Square Error* nilai yang paling optimal berada pada *k*=4 dengan hasil sebesar 54,5548 karena pada titik *k*=4 nilai SSE mengalami perubahan paling drastis sehingga dapat disimpulkan bahwa mengggunakan metode *Elbow* yaitu *k*=4. Penelitian Tugas Akhir ini menggunakan metode *Silhouette Coefficient* dan metode *Elbow* yang digunakan sebagai penentu jumlah klaster terbaik agar didapatkan jumlah dan hasil klaster yang sesuai dengan kriteria data indeks pembangunan manusia dan menggunakan metode *Davies Bouldin Index* sebagai penentu dari dua metode yang digunakan tersebut.

Tabel 8. Hasil Perhitungan *Davies Bouldin Index*

K	**Metode**	***Davies Bouldin Index***
3	*Silhoutte Coefficient*	1,3185
4	*Elbow*	1,1754

Dari hasil tersebut dapat ditarik kesimpulan bahwa 4 klaster memiliki nilai *Davies Bouldin Index* terendah dan merupakan jumlah klaster yang optimal. Dengan demikian, metode *partitioning around medoid* akan digunakan untuk melakukan pengelompokan provinsi di Indonesia berdasarkan tingkat pembangunan manusia yang terdapat pada setiap provinsi menjadi 4 klaster (k=4) dengan pengukuran jarak *Euclidean*. Hasil pengklasteran provinsi di Indonesia dengan metode *partitioning around medoids* berdasarkan data indikator indeks pembangunan manusia tahun 2022 dengan jumlah 4 klaster dan Tabel 9 memperlihatkan hasil pengukuran jarak *Euclidean* yang dihasilkan dari *running* sintaks *software Rstudio* 4.1.3.

Tabel 9. Hasil Klasterisasi Terbaik (k=4)

Klaster	Objek Medoid	Jumlah Anggota	Provinsi
1	Objek ke-7 (Provinsi Bengkulu)	10	Sulawesi Tengah, Maluku Utara, Aceh, Bengkulu, Sumatera Barat, Sulawesi Tenggara, Sulawesi Selatan, Gorontalo, Sumatera Utara, dan Maluku.
2	Objek ke- 5 (Provinsi Jambi)	14	Sumatera Selatan, Kalimantan Selatan, Riau, Banten, Jambi, Lampung, Kalimantan Tengah, Sulawesi Utara, Jawa Tengah, Kalimantan Utara, Jawa Timur, , Kalimatan Barat, Jawa Barat, dan Kep. Bangka Belitung.
3	Objek ke-17 (Provinsi Bali)	5	Bali, Kalimantan Timur, DI Yogyakarta, Kep. Riau, dan DKI Jakarta.
4	Objek ke-33 (Provinsi Papua Barat)	5	Papua, Nusa Tenggara Timur, Papua Barat, Nusa Tenggara Barat, dan Sulawesi Barat.

Tabel 10 menampilkan rata-rata nilai dari setiap variabel dalam setiap kelompok klaster yang terbentuk sebagai representasi dari hasil klasterisasi.

Tabel 10. Nilai Rata-RataVariabel di Setiap Kelompok Klaster

Variabel	Kelompok Klaster 1	Kelompok Klaster 2	Kelompok Klaster 3	Kelompok Klaster 4
Provinsi yang memiliki harapan hidup saat lahir *(Y_1)*	69,44	71,49	73,22	66,57
Provinsi yang memiliki harapan lama untuk menyelesaikan pendidikan formal (Y_2)	13,69	12,85	13,81	12,88
Provinsi yang memiliki rata-rata lama pendidikan yang dicapai (Y_3)	9,15	8,58	10,15	7,65
Provinsi yang memiliki pengeluaran per kapita disesuaikan (Y_4)	10158	11250	14892	8633

Berdasarkan Tabel 10, beberapa informasi yang diperoleh adalah sebagai berikut:

a. Klaster 1

Pada klaster pertama terdiri 10 provinsi, yaitu : Sumatera Barat, Sulawesi Tengah, Maluku, Bengkulu, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sumatera Utara, Maluku Utara, dan Aceh. Klaster 1 mempunyai dua variabel yaitu Y_2 dan Y_3 dengan rata-rata lebih rendah daripada klaster 3 dan lebih tinggi daripada klaster 2 dan klaster 4. Kemudian pada variabel Y_1 dan Y_4 lebih tinggi dari klaster 4 dan lebih rendah dari klaster 2 dan klaster 3. Dari penilaian empat indikator tersebut, provinsi yang berada pada klaster satu berada pada tingkat sedang atau baik dari segi pembangunan manusia.

b. Klaster 2

Pada klaster dua terdapat 14 provinsi, yaitu : Sumatera Selatan, Kalimantan Selatan, Riau, Banten, Jambi, Lampung, Kalimantan Tengah, Sulawesi Utara, Jawa Tengah, Kalimantan Utara, Jawa Tiimur, Kalimatan Barat, Jawa Barat, dan Kep. Bangka Belitung. Klaster 2 mempunyai variabel Y_1 dan Y_4 yang lebih rendah daripada klaster 3 dan lebih tinggi daripada klaster 1 dan klaster 4. Kemudian pada variabel Y_2 lebih rendah daripada

klaster 1, klaster 3, dan klaster 4. Pada variabel Y_3 lebih tinggi daripada klaster 4 dan lebih rendah daripada klaster 1 dan klaster 3. Dari penilaian empat indikator tersebut, provinsi yang berada pada klaster dua berada pada tingkat rendah dari segi pembangunan manusia.

c. Klaster 3

Pada klaster tiga terdapat 5 provinsi, yaitu : DI Yogyakarta, Kep. Riau, Bali, DKI Jakarta, dan Kalimantan Timur. Klaster 3 mempunyai rata-rata variabel Y_1, Y_2, Y_3, dan Y_4 lebih tinggi daripada klaster 1, klaster 2, dan klaster 4. Dari penilaian empat indikator tersebut, provinsi yang berada pada klaster tiga berada pada tingkat yang paling baik dari segi pembangunan manusia.

d. Klaster 4

Pada klaster empat terdapat 5 provinsi, yaitu : Papua Barat, Sulawesi Barat, Nusa Tenggara Barat, Papua, dan Nusa Tenggara Timur. Pada klaster 4 mempunyai rata-rata variabel Y_1, Y_3, dan Y_4 lebih rendah daripada klaster 1, klaster 2, dan klaster 4. Kemudian pada variabel Y_2 lebih tinggi dari daripada klaster 2 dan lebih rendah daripada klaster 1 dan klaster 3. Dari penilaian empat indikator tersebut, provinsi yang berada pada klaster empat berada pada tingkat paling rendah atau kurang baik dari segi pembangunan manusia.

5. KESIMPULAN

Dari hasil penelitian ini menggunakan metode Silhouette Coefficient didapatkan klaster terbaik dengan 3 klaster dan *nilai Silhouette Coefficient* yaitu 0,3129, serta nilai *Davies Bouldin Index (DBI*) yaitu 1,3185. Sementara itu, menggunakan metode Elbow didapatkan klaster terbaik dengan 4 klaster dan nilai SSE sebesar 54,5548, serta nilai Davies Bouldin Index (DBI) sebesar 1,1754. Sehingga dapat ditarik kesimpulan bahwa sesuai dengan Indeks Pembangunan Manusia, terdapat empat kelompok yang paling cocok untuk diterapkan di wilayah provinsi Indonesia, terdiri dari klaster 1 yang terdiri dari 10 provinsi pada tingkat sedang atau baik, klaster 2 yang terdiri dari 14 provinsi pada tingkat rendah, klaster 3 yang terdiri dari 5 provinsi pada tingkat tertinggi atau paling baik, dan klaster 4 yang terdiri dari 5 provinsi pada tingkat paling rendah atau kurang baik.

DAFTAR PUSTAKA

Anderberg, M. 1973. *Cluster Analysis for Application*. New York: *Academic Press.*

Dewi, D. A. I. C., & Pramita, D. A. K. 2019. Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, *9*(3), 102-109.

Gujarati, D. 2009. Dasar-Dasar Ekonometrika. Jakarta: Erlangga

Han, J., & Kamber, M. 2006. *Data Mining: Concepts and Techniques*. San Fransisco: Elsevier Inc.

Handoyo, R., Mangkudjaja, R., & Nasution, S. M. 2014. Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-means Pada Pengelompokan Dokumen. Jurnal Sifo Mikroskil, 15(2), 73-82.

Madhulatha, T.S., 2012. *An Overview On Clustering Methods*. IOSR Journal of Engineering, II(4), pp.719-725

Merliana, N. P. E., & Santoso, A. J. 2015. Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering.

Rousseeuw, P. J., & Leroy, A. M. 1996. *Robust Regression and Outlier Detection Third Edition.* New York: John Wiley & Sons, Inc.

Rousseeuw, P. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Jornal of Computational and Applied Mathematics*. Vol.20, 53–65.

Statistik, B. P. 2022. Indeks Pembangunan Manusia (IPM) Indonesia Tahun 2022. *Badan Pusat Statistik. Jakarta*.

Supranto, J. 2004. Analisis Multivariat Arti dan Interpretasi. Jakarta : PT Rineka Cipta.

Supranto, J. 2004b. Ekonometri. Jakarta: Ghalia Indonesia.

Vercellis, C. 2009. *Business Intelligence: Data Mining And Optimization For Decision Making (1st ed.).* Milan: Wiley.