

Analisis Perbandingan K-Means dan K-Medoids dalam Pengelompokan Provinsi Berdasarkan Indeks Demokrasi Indonesia 2021

Regita Dewanti Rudianto^{1*}, Arie Wahyu Wijayanto²

^{1,2}Program Studi D-IV Statistika, Politeknik Statistika STIS

Jl. Otto Iskandardinata No.64C, Bidara Cina, Jatinegara, Kota Jakarta Timur, DKI Jakarta, Indonesia 13330

*email: 212011784@stis.ac.id

(Naskah masuk: 17 Agustus 2023; diterima untuk diterbitkan: 11 Januari 2024)

ABSTRAK – Sebagai negara demokrasi, Indonesia perlu menegakkan hak-hak politik, ekonomi, dan sosial masyarakat secara bebas serta dilindungi oleh konstitusi. Salah satu ukuran dalam menilai capaian dan kapasitas demokrasi Indonesia adalah melalui Indeks Demokrasi Indonesia (IDI). Penelitian ini bertujuan untuk melakukan pengelompokan (clustering) 34 provinsi di Indonesia berdasarkan aspek yang terdapat dalam IDI Tahun 2021. Metode clustering merupakan salah satu metode dalam data mining dan berguna dalam mengelompokkan amatan yang tidak memiliki target/kelas. Analisis clustering perlu dilakukan untuk mengetahui pengelompokan aspek IDI beserta karakteristiknya. Dalam penelitian ini, dilakukan clustering terhadap aspek-aspek dalam IDI yang mencakup Aspek Kebebasan, Aspek Kesetaraan, dan Aspek Kapasitas Lembaga Demokrasi. Metode clustering yang digunakan dalam penelitian ini yaitu K-Means dan K-Medoids. Untuk pemilihan jumlah cluster yang optimal digunakan Indeks Dunn, Indeks Silhouette, Indeks Calinski-Harabasz, dan Indeks Davies-Bouldin. Dalam menentukan model terbaik, dilakukan perbandingan menggunakan rasio antara average within (S_w) dan average between (S_b). Hasil yang diperoleh yaitu terdapat 5 cluster dalam pengelompokan IDI dengan menggunakan algoritma K-Medoids karena rasio S_w/S_b lebih kecil dibandingkan K-Means. Dengan adanya pengelompokan ini, diharapkan pemerintah dan pihak terkait dapat memanfaatkan hasil analisis untuk merumuskan kebijakan yang tepat dan menjaga kestabilan politik di Indonesia.

Kata Kunci – Indeks Demokrasi Indonesia (IDI); K-Means; K-Medoids; Indeks Dunn; Indeks Davies-Bouldin.

Comparison Analysis of K-Means and K-Medoids in Grouping Provinces Based on Indonesian Democracy Index 2021

ABSTRACT – As a democratic country, Indonesia needs to uphold the political, economic, and social rights of residents freely and protected by the constitution. One measure in assessing the achievements and capacity of Indonesian democracy is through the Indonesian Democracy Index (IDI). This study aims to cluster 34 provinces in Indonesia based on the aspects contained in the IDI Year 2021. The clustering method is one method in data mining and is practical in grouping observations that do not have a target/class. A clustering analysis was used to determine the grouping of IDI aspects and their characteristics. In this study, clustering was carried out on aspects of IDI which included Aspects of Freedom, Aspects of Equality, and Aspects of the Capacity of Democratic Institutions. The clustering methods used in this study are K-Means and K-Medoids. For the selection of the optimal number of clusters used Dunn Index, Silhouette Index, Calinski-Harabasz Index, and Davies-Bouldin Index. In determining the best model, a comparison is made using the ratio between the average within (S_w) and the average between (S_b). The results obtained are that there are 5 clusters in the IDI grouping using the K-Medoids algorithm because the ratio of S_w / S_b is smaller than K-Means. With this grouping, it is expected that the government and related parties can utilize the results of the analysis to formulate appropriate policies and maintain political stability in Indonesia.

Keywords – Indonesian Democracy Index (IDI); K-Means; K-Medoids; Dunn Index; Davies-Bouldin Index.

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang menganut sistem politik demokrasi. Demokrasi ini telah membawa perubahan politik dengan memberikan hak yang luas kepada rakyat untuk mengawasi pemerintah agar tidak terjadi penyalahgunaan kekuasaan. Dalam suatu negara demokrasi, terdapat dua karakteristik utama, yaitu dianutnya prinsip demokrasi dan adanya konstitusi yang menjadi hukum dasar kehidupan berbangsa dan bernegara [1].

Untuk mengukur sejauh mana sistem demokrasi berjalan di Indonesia, Bappenas bersama dengan UNDP menginisiasi pengukuran capaian dan kapasitas demokrasi di Indonesia. Kegiatan ini juga didukung oleh Badan Pusat Statistik (BPS) dan Badan Perencana Pembangunan Daerah (Bappeda) sehingga terbentuklah sebuah indeks yang disebut Indeks Demokrasi Indonesia (IDI). IDI adalah alat ukur kuantitatif yang digunakan untuk mengukur kemajuan dan penerapan demokrasi di Indonesia. Pengukuran indeks ini dilakukan sejak tahun 2006 dengan hasil berupa skala angka 1-100 dimana 1 adalah kinerja terendah dan 100 adalah kinerja tertinggi. Aspek yang diukur dalam IDI di antaranya Aspek Kebebasan Sipil, Aspek Hak-Hak Politik, dan Aspek Lembaga Demokrasi. Namun, sejak tahun 2021, terdapat pengukuran dengan metode baru sehingga terdapat perubahan aspek menjadi Aspek Kebebasan, Aspek Kesetaraan, dan Aspek Kapasitas Lembaga Demokrasi.

Pada tahun 2021, diketahui bahwa capaian keseluruhan aspek penyusun Indeks Demokrasi Indonesia berada dalam kategori sedang dengan angka 78,12. Nilai pada masing-masing aspek juga termasuk dalam kategori sedang, di antaranya: (1) Aspek Kebebasan sebesar 79,72; (2) Aspek Kesetaraan sebesar 78,86; dan (3) Aspek Kapasitas Lembaga Demokrasi sebesar 75,67 [2].

Dengan adanya pengukuran IDI, dapat dilihat evaluasi dan hambatan dalam sistem demokrasi di Indonesia secara keseluruhan. Salah satu contohnya adalah, pada data IDI ditemukan adanya aturan pemerintah daerah yang membatasi kebebasan sipil terutama kebebasan berkeyakinan di beberapa daerah. Selain itu, data IDI dapat digunakan untuk merencanakan pembangunan di bidang politik. Untuk itu, penting untuk diketahui bagaimana pengelompokan karakteristik IDI menurut provinsi di Indonesia. Dengan melihat pengelompokan tersebut, dapat diketahui metode yang tepat dalam merumuskan kebijakan untuk menjaga kestabilan politik di Indonesia. Salah satu cara yang dapat

dilakukan untuk mengelompokkan IDI menurut provinsi yaitu dengan metode *clustering*.

Penelitian sebelumnya mengenai *clustering* IDI telah dilakukan beberapa kali. Penelitian yang pertama berjudul "Analisis Cluster Hirarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Demokrasi Indonesia Tahun 2016". Penelitian yang dilakukan oleh Ghaisani *et al.*, pada tahun 2019 ini membandingkan lima metode *cluster* hirarki, yaitu *average linkage*, *centroid linkage*, *complete linkage*, *single linkage*, dan *ward*. Hasil yang didapatkan yaitu metode *cluster centroid* merupakan metode terbaik dalam pengelompokan IDI menurut provinsi di Indonesia [3].

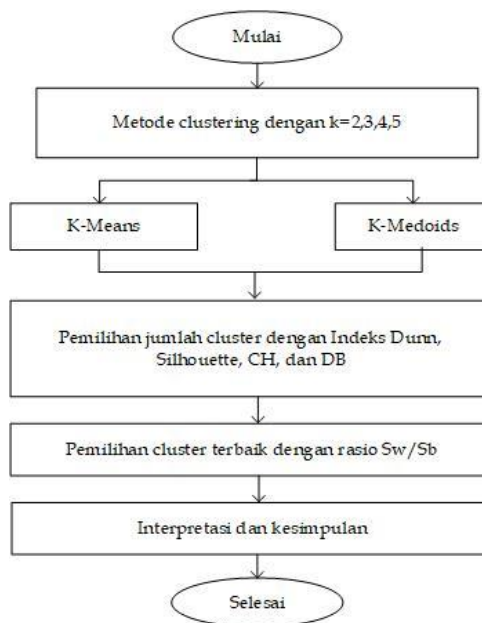
Penelitian yang kedua dilakukan oleh Nurdiana *et al.*, tahun 2022 dengan judul "Penerapan Algoritma Fuzzy C-Means dan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia". IDI yang digunakan yaitu IDI tahun 2019 dengan hasil yaitu diperoleh 3 *cluster* dimana *cluster* 1 mencakup 15 wilayah, *cluster* 2 sebanyak 12 wilayah, dan *cluster* 3 mencakup 7 wilayah [4].

Penelitian ketiga berjudul "Implementasi Rapidminer dalam Mengklasifikasikan Indeks Demokrasi" yang dilakukan oleh Nasution *et al.*, pada tahun 2022 dengan menggunakan IDI tahun 2013-2020. Hasil yang didapatkan yaitu terdapat 2 *cluster* dengan *cluster* pertama sebanyak 11 provinsi dan *cluster* kedua sebanyak 22 provinsi [5].

Berdasarkan beberapa penelitian sebelumnya, didapatkan bahwa Indeks Demokrasi yang digunakan belum menggunakan IDI yang terbaru. Selain itu, belum ada penelitian terkait IDI yang membandingkan metode *K-Means* dan *K-Medoids*. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan *cluster* aspek-aspek dalam IDI di 34 provinsi tahun 2021. Analisis ini dilakukan dengan membandingkan dua metode *clustering*, yaitu *K-Means* dan *K-Medoids*.

2. METODE DAN BAHAN

Dalam penelitian ini, digunakan metode kuantitatif dengan membandingkan dua algoritma *clustering*, yaitu *K-Means* dan *K-Medoids*. Data yang digunakan berasal dari website Badan Pusat Statistik (BPS) berupa nilai Indeks Demokrasi Indonesia tahun 2021 yang dirinci menurut 3 aspek, di antaranya Aspek Kebebasan, Aspek Kesetaraan, dan Aspek Kapasitas Lembaga Demokrasi. Jumlah observasi atau amatan sebanyak 34 provinsi di Indonesia. Tahapan penelitian ini ditampilkan dalam Gambar 1.



Gambar 1. Alur Penelitian [1]

Tahapan *Data mining*

Data mining merupakan teknik dan metode dalam mengekstrak pengetahuan dari kumpulan data yang besar dan dapat digunakan lebih lanjut untuk keperluan sains, industri, maupun penggunaan operasional dari pengetahuan tersebut. *Data mining* berkaitan dengan statistik sebagai matematika terapan yang mengekstraksi informasi yang berguna dari suatu data [6]. Selain itu, *data mining* juga menjadi teknologi yang berguna sebagai alat analisis yang kuat dan efektif untuk menginterpretasikan data yang besar dan kompleks [7]. Tahapan dalam *data mining* dapat dirinci sebagai berikut

a) *Data Selection*

Proses pemilihan data merupakan tahap paling awal dalam *data mining*. Seleksi data dilakukan dengan mencari data yang sesuai dengan variabel yang telah ditetapkan.

b) *Pre-processing/Cleaning*

Pre-processing data adalah tahapan yang mentransformasi raw data ke dalam format yang lebih mudah dan efektif untuk dilakukan proses selanjutnya [8]. Dalam tahap ini, dilakukan pengecekan *missing values* dan *noise* yang ada dalam data.

c) *Transformation*

Pada tahap transformasi data, data dilakukan *smoothing*, *attribute construction*, *aggregation*, dan *discretization*. *Smoothing* dilakukan untuk menghilangkan *noise* dari data. *Attribute construction* (konstruksi fitur) membentuk atribut baru yang diperlukan dalam *data mining*. *Aggregation* berguna untuk mengumpulkan

atribut tertentu. *Normalization* berguna untuk menormalisasi atribut agar berada pada rentang yang lebih kecil. Sementara *discretization* merupakan proses pengkategorian atribut ke dalam pengelompokan yang diinginkan [9].

d) *Data Mining*

Dalam tahap ini, dilakukan proses pencarian pola atau informasi menarik dalam data dengan teknik tertentu. *Data mining* merupakan proses pencarian dan analisis terhadap suatu data untuk mendapatkan informasi dan pengetahuan sehingga memiliki manfaat dalam pengambilan keputusan [10].

e) *Interpretation/Evaluation*

Setelah didapatkan informasi, dilakukan interpretasi sehingga mudah dimengerti oleh banyak pihak. Dalam tahap ini juga dilakukan pemeriksaan apakah hasil yang didapatkan sesuai dengan hipotesis sebelumnya [11].

K-Means Clustering

K-Means merupakan salah satu metode dalam *clustering* yang meminimalkan jumlah jarak kuadrat dari setiap titik dengan titik pusat *cluster* [12]. Pemilihan pusat *cluster* awal yang tepat sangat penting dalam pengelompokan *K-Means*. Pusat awal yang berbeda dari *cluster* dapat menghasilkan optima lokal yang berbeda dan dapat menghasilkan hasil *cluster* yang berbeda [13]. Langkah-langkah dalam melakukan *clustering* dengan algoritma *K-Means* adalah sebagai berikut [14]:

1. Menentukan jumlah *cluster* (k) yang ingin dibuat serta titik awal (*centroid*) secara acak.
2. Menghitung jarak tiap amatan terhadap *centroid*

dengan jarak Euclidian sampai ketemy titik terdekatnya. Rumus jarak Euclidian adalah sebagai berikut:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

dimana

$d(x_i, \mu_j)$ = jarak *cluster* x dengan pusat *cluster* μ pada amatan ke-i.

x_i = bobot amatan ke-i pada *cluster* yang ingin dicari jaraknya

μ_i = bobot amatan ke-i pada pusat *cluster*.

3. Menghitung nilai *centroid* yang baru dengan merata-ratakan *cluster* dengan rumus sebagai berikut:

$$C_k = \frac{1}{nk} \sum d_i \quad (2)$$

dimana

nk = jumlah amatan pada *cluster*

d_i = jumlah nilai jarak pada masing-masing *cluster*.

4. Menghitung jarak setiap objek dengan *centroid* baru,
5. Mengulangi langkah 2 dan 3 hingga konvergen. Jika tidak ada perubahan lagi, nilai rata-rata pusat *cluster* akan digunakan sebagai parameter dalam mengklasifikasikan data.

K-Medoids Clustering

K-Medoids adalah algoritma yang lebih baik daripada K-Means untuk menentukan medoids dalam kelompok yang berfungsi sebagai titik pusat kelompok. Hal ini karena pada K-Medoids, k adalah objek yang dapat meminimalkan jumlah ketidaksamaan objek data [15]. Medoid terletak pada jumlah jarak terkecil ke objek yang lain [16]. Langkah-langkah dalam melakukan algoritma K-Medoids di antaranya [17]:

1. Menentukan jumlah *cluster* sebanyak k,
2. Mengalokasikan setiap amatan ke *cluster* terdekat dengan jarak Euclidian,
3. Memilih amatan pada masing-masing *cluster* secara acak sebagai kandidat medoid baru,
4. Menghitung jarak tiap amatan pada masing-masing *cluster* dengan kandidat medoid yang baru,
5. Menghitung total simpangan (S) dengan menghitung nilai total jarak baru dikurangi jarak lama. Jika $S < 0$, tukarkan amatan dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid,
6. Mengulangi langkah 3 hingga 5 sampai konvergen sehingga didapatkan *cluster* beserta anggotanya.

Validasi Cluster

Proses validasi *cluster* digunakan untuk mencari jumlah *cluster* yang optimum [18]. Dalam menentukan jumlah *cluster* yang optimum, dapat

digunakan Indeks *Dunn*, Indeks *Silhouette*, Indeks *Calinski-Harabasz*, dan Indeks *Davies-Bouldin*.

Indeks Dunn

Indeks *Dunn* berfungsi untuk menilai validitas dari suatu *cluster* dengan asumsi bahwa antar *cluster* dipisahkan oleh suatu jarak yang besar serta diameter dalam *cluster* yang kecil. Semakin baik suatu algoritma *clustering*, nilai indeks *Dunn* nya akan semakin besar [19]. Penghitungan Indeks *Dunn* dilakukan dengan rumus berikut:

$$DI = \frac{\min d(C_i, C_j)}{\max(\text{diameter } C_k)} \quad (3)$$

Indeks Silhouette

Indeks *Silhouette* berfungsi untuk mengukur sejauh mana setiap titik data cocok dengan *clusternya* dan sejauh mana *cluster* tersebut terpisah satu sama lain. Indeks *Silhouette* digunakan untuk memvalidasi efektivitas pengelompokan berdasarkan perbedaan jarak berpasangan antar dan intra *cluster* dari dua objek data [20]. Semakin tinggi nilai indeks *Silhouette*, semakin baik validitas *clusternya*. Rumus dari Indeks *Silhouette* adalah sebagai berikut [21]:

$$SIL = \frac{1}{N} \sum_{i=1}^N S_{x_i} \quad (4)$$

$$S_{x_i} = \frac{(b_{1,i} - a_{p,i})}{\max\{a_{p,i}, b_{q,i}\}}, p \neq q \quad (5)$$

$$b_{q,i} = \min d_{q,i}; q = 1, \dots, k \quad (6)$$

$$d_{q,i} = \frac{1}{n_q} \sum_{i=1}^{n_q} d(x_i, y_j) \quad (7)$$

$$a_{p,i} = \frac{1}{n_p} \sum_{k=1}^{n_p} d(x_i, x_k) \quad (8)$$

Indeks Calinski-Harabasz

Indeks *Calinski-Harabasz* mengukur rasio dispersi antar*cluster* dengan dispersi dalam *cluster*. Semakin tinggi nilai Indeks *Calinski-Harabasz*, semakin baik validitasnya. Pemilihan indeks ini dilakukan karena ada beberapa literatur yang mendukung [22]. Rumus untuk menghitung Indeks *Calinski-Harabasz* yaitu:

$$CH = \frac{\text{trace}(SSB)}{\text{trace}(SSW)} \times \frac{N-k}{k-1} \quad (9)$$

$$SSW = \sum_{l=1}^k \sum_{x_i \in C_l} (\bar{x}_i - \bar{x}_l)(\bar{x}_i - \bar{x}_l)^T \quad (10)$$

$$SSB = \sum_{l=1}^k N_l (\bar{x}_l - \bar{x})(\bar{x}_l - \bar{x})^T \quad (11)$$

$$\bar{x}_l = \frac{1}{N} \sum_{x_i \in C_l} x_i \quad (12)$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (13)$$

Indeks Davies-Bouldin

Indeks *Davies-Bouldin* (DB) merupakan suatu pengukuran yang menghitung rata-rata nilai dari setiap titik pada himpunan data [10]. Indeks ini mewakili kesamaan rata-rata antar *cluster* dimana kesamaan merupakan metrik yang menghubungkan jarak *cluster* terhadap ukuran *cluster* [23]. Indeks ini diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Evaluasi dengan indeks ini melihat evaluasi dari internal *cluster* dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antardata hasil *cluster* [24]. Semakin rendah nilai Indeks *Davies-Bouldin*, semakin baik

validitasnya. Selain itu, Indeks *Davies-Bouldin* juga digunakan untuk menemukan jumlah sentroid *cluster* yang paling akurat [25]. Indeks *Davies-Bouldin* dapat dihitung dengan rumus sebagai berikut:

$$DB = \frac{1}{k} \sum_{p=1}^k R_p \quad (14)$$

$$R_p = \max R_{p,q}, p \neq q \quad (15)$$

$$R_{p,q} = \frac{(S_p + S_q)}{M_{pq}} \quad (16)$$

$$S_p = \frac{1}{n_p} \sum_{i=1}^{n_p} d(x_i, c_p) \quad (17)$$

$$S_q = \frac{1}{n_q} \sum_{j=1}^{n_q} d(y_j, c_q) \quad (18)$$

$$M_{pq} = d(c_p, c_q) \quad (19)$$

Evaluasi Cluster Terbaik

Untuk mengevaluasi metode yang terbaik, dapat digunakan rasio rata-rata simpangan baku antar *cluster* (*Sw*) dengan simpangan baku antar *cluster* (*Sb*). Rumus dari *Sw* dan *Sb* adalah sebagai berikut [26]:

$$S_w = \frac{1}{k} \sum_{k=1}^K S_k \quad (20)$$

$$S_b = [(K - 1)^{-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})]^{\frac{1}{2}} \quad (21)$$

3. HASIL DAN PEMBAHASAN

Analisis Deskriptif

Data yang telah didapatkan selanjutnya dianalisis dengan metode *cluster* menggunakan software R. Hasil dari analisis deskriptif berupa nilai minimal, maksimal, median, mean, kuartil 1, dan kuartil 3 dapat dilihat pada Tabel 1 berikut. Karena nilai antarvariabel tidak berbeda jauh, tidak perlu dilakukan transformasi untuk menyamakan rentang nilainya.

Tabel 1. Ringkasan Statistik [2]

Ringkasan Statistik	Aspek		
	Kebebasan	Kesetaraan	Kapasitas Lembaga Demokrasi
Minimal	59,36	62,30	56,10
Kuartil 1	76,30	74,39	63,87
Median	82,36	78,14	69,48
Mean	81,08	77,17	69,86
Kuartil 3	88,08	80,15	75,89
Maksimal	92,76	87,29	81,96

Pembentukan Model

K-Means Clustering

Metode *K-Means* diawali dengan menentukan jumlah *cluster* yang digunakan. Untuk itu, perlu adanya perbandingan antar*cluster* dengan beberapa pilihan jumlah *cluster*. Hasilnya dapat dilihat dalam Tabel 2.

Berdasarkan tabel tersebut, terlihat bahwa nilai indeks *Dunn* yang terbesar adalah saat $k=4$ dan $k=5$, Indeks *Silhouette* yang terbesar saat $k=4$, begitu pula

dengan Indeks *Calinski-Harabasz* (CH). Namun, untuk nilai Indeks *Davies-Bouldin* (DB) yang terkecil yaitu saat $k=5$. Maka dari itu, dapat disimpulkan bahwa jumlah *cluster* yang paling optimal yaitu saat $k=4$.

Tabel 2. Pemilihan Jumlah Cluster *K-Means* [3]

k	Indeks			
	<i>Dunn</i>	<i>Silhouette</i>	CH	DB
2	0,089	0,299	14,284	1,453
3	0,107	0,320	18,944	1,078
4	0,160	0,354	20,241	0,914
5	0,160	0,333	17,099	0,829

K-Medoids Clustering

Seperti yang dilakukan pada metode *K-Means*, metode *K-Medoids* juga perlu dilakukan penentuan jumlah k yang optimal. Hasilnya dapat dilihat pada Tabel 3. Perbandingan dilakukan antara jumlah *cluster* 2, 3, 4, dan 5.

Berdasarkan tabel tersebut, dapat disimpulkan bahwa jumlah *cluster* yang optimal adalah sebanyak 5 *cluster* karena nilai Indeks *Dunn*, Indeks *Silhouette*, dan Indeks *Calinski-Harabasz* (CH) terbesar di antara jumlah k lainnya. Sementara untuk Indeks *Davies-Bouldin* (DB), nilainya terkecil di antara jumlah k lainnya.

Tabel 3. Pemilihan Jumlah Cluster *K-Medoids* [4]

k	Indeks			
	<i>Dunn</i>	<i>Silhouette</i>	CH	DB
2	0,205	0,286	14,138	1,978
3	0,141	0,278	16,754	1,160
4	0,172	0,279	16,102	1,159
5	0,284	0,336	20,235	1,004

Evaluasi Model Terbaik

Selanjutnya, dilakukan perbandingan antara *K-Means* $k=4$ dengan *K-Medoids* $k=5$. Dengan menggunakan rasio antara nilai *average within* (*Sw*) dan *average between* (*Sb*), akan diperoleh rasio rata-rata simpangan baku. *Cluster* yang baik adalah saat nilai rata-rata simpangan baku dalam *cluster* kecil, sementara rata-rata simpangan baku antarkelompok besar sehingga nilai rasionya kecil [10]. Hasilnya terlihat pada tabel 4 berikut.

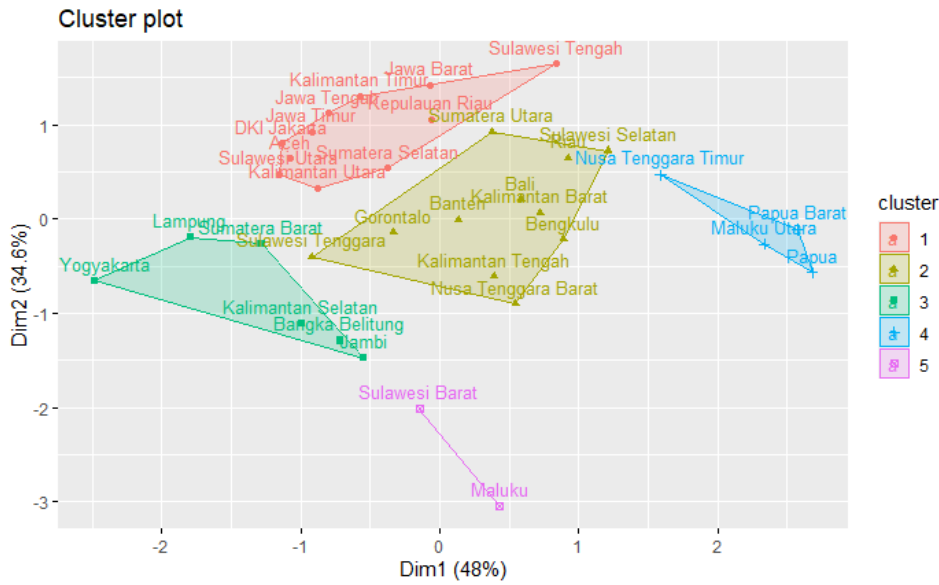
Tabel 4. Evaluasi Model [5]

Metode	Jumlah Cluster	Rasio (<i>Sw/Sb</i>)
<i>K-Means</i>	4	0,5214
<i>K-Medoids</i>	5	0,4895

Berdasarkan tabel di atas, terlihat bahwa metode *K-Medoids* memiliki rasio yang lebih kecil dibandingkan *K-Means* sehingga merupakan metode terbaik.

Selanjutnya, dilakukan pembentukan *cluster* sebanyak 5 *cluster* dengan menggunakan metode *K-Medoids*. Hasilnya terlihat pada Gambar 2 berikut.:

Visualisasi Cluster



Gambar 2. Cluster Plot Berdasarkan Provinsi [6]

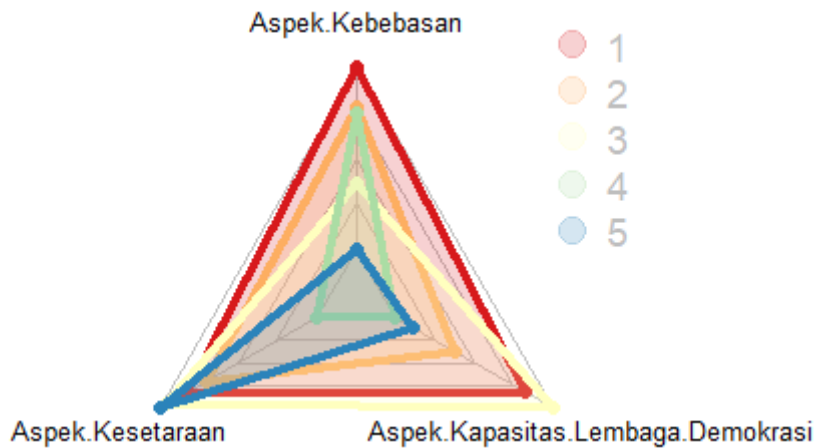
Pada gambar di atas, terlihat adanya 3 *cluster* besar, yaitu *cluster* 1, 2, dan 3. *Cluster* 1 berwarna merah, *cluster* 2 berwarna avocado, *cluster* 3 berwarna hijau, dan *cluster* 4 berwarna biru. Sementara itu,

cluster 5 yang berwarna ungu merupakan *cluster outlier* pada data ini. Berikut adalah daftar provinsi berdasarkan pengelompokan *cluster* nya.

Tabel 5. Hasil Clustering [7]

Cluster	Anggota Cluster
1	Aceh, Sumatera Selatan, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah.
2	Sumatera Utara, Riau, Bengkulu, Banten, Bali, NTB, Kalimantan Barat, Kalimantan Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo.
3	Sumatera Barat, Jambi, Lampung, Bangka Belitung, DI Yogyakarta, Kalimantan Selatan.
4	NTT, Maluku Utara, Papua Barat, Papua.
5	Sulawesi Barat, Maluku.

Selain itu, dilihat dari nilai di tiap aspeknya, setiap klaster memiliki karakteristik seperti yang ada pada Gambar 3 di bawah ini.



Gambar 3. Cluster Plot Berdasarkan Aspek [8]

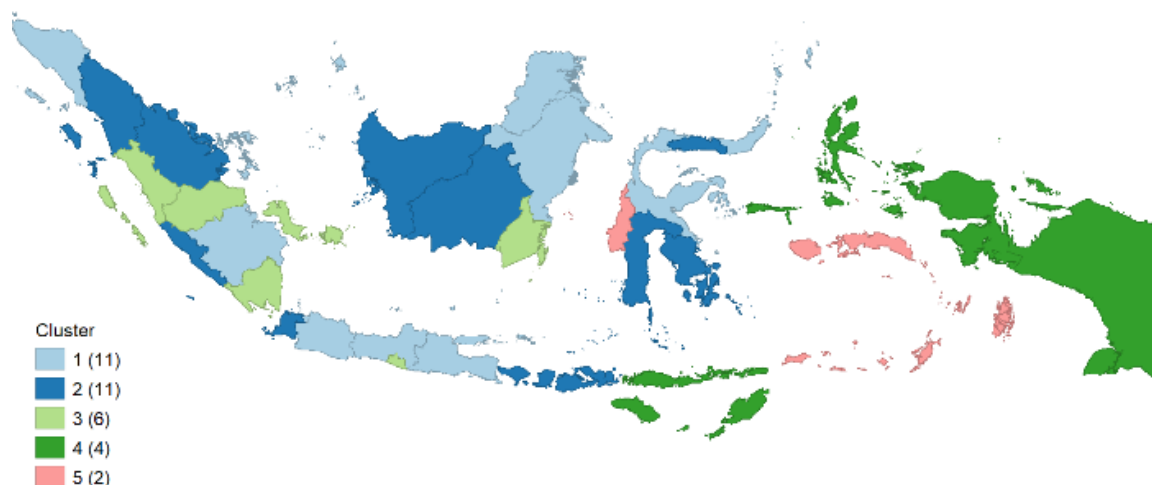
Berdasarkan gambar di atas, terlihat bahwa setiap *cluster* memiliki karakteristik masing-masing yang membedakan antara *cluster* satu dan lainnya. *Cluster* 1 memiliki nilai yang tinggi pada ketiga aspek, yaitu aspek kesetaraan, aspek kebebasan, dan aspek kapasitas lembaga demokrasi. Hal ini menunjukkan bahwa provinsi yang berada di *cluster* 1 sudah cukup baik dalam hal pelaksanaan demokrasinya. Sementara itu, *cluster* 2 memiliki nilai yang sedang di antara *cluster* lainnya di semua aspek. *Cluster* 3 memiliki nilai rata-rata aspek kapasitas lembaga demokrasi yang tertinggi di antara *cluster* lainnya, serta cukup tinggi untuk aspek kesetaraan. Namun, untuk aspek kebebasan nilainya cukup rendah. *Cluster* 4 memiliki nilai yang cukup tinggi pada aspek kebebasan, tetapi rendah untuk aspek lainnya. Sementara *cluster* 5 memiliki nilai yang tinggi pada aspek kesetaraan. Namun, untuk aspek kebebasan dan aspek kapasitas lembaga demokrasi nilainya cukup rendah.

Hasil interpretasi tersebut dapat ditunjukkan secara kuantitatif berdasarkan nilai rata-rata variabel untuk setiap *cluster* dalam Tabel 6 berikut.

Tabel 6. Karakteristik Aspek Tiap *Cluster* [9]

Cluster	Aspek		
	Kebebasan	Kesetaraan	Kapasitas Lembaga Demokrasi
1	87,67	78,55	74,78
2	82,38	76,86	66,06
3	72,10	80,10	78,31
4	81,68	68,10	58,63
5	63,30	80,66	60,77

Hasil *clustering* juga dapat ditampilkan dengan peta tematik seperti terlihat pada Gambar 3. Berdasarkan peta tersebut, terlihat bahwa *cluster* 1 tersebar di Pulau Sumatera, Jawa, Kalimantan, dan Sulawesi. Sementara *cluster* 2 berada di beberapa provinsi di Sumatera, Kalimantan, Sulawesi, Bali dan NTB. *Cluster* 3 berada di Pulau Sumatera dan Kalimantan. *Cluster* 4 dan 5 mendominasi wilayah timur Indonesia, yaitu Maluku dan Papua.



Gambar 3. Peta Tematik *Cluster* [10]

4. KESIMPULAN

Berdasarkan hasil analisis, diperoleh pengelompokan provinsi di Indonesia menurut Indeks Demokrasi Indonesia (IDI) tahun 2021. Pengelompokan yang terbaik menggunakan algoritma *K-Medoids* dengan jumlah *cluster* sebanyak 5 *cluster*. Hasil pengelompokan ini berdasarkan evaluasi rasio Sw/Sb dimana rasio untuk *K-Medoids* lebih kecil dibandingkan *K-Means*.

Cluster yang diperoleh memiliki karakteristik tertentu dimana *cluster* 1 memiliki nilai yang tinggi pada ketiga aspek, sementara *cluster* 2 memiliki nilai yang sedang di antara *cluster* lainnya di semua aspek. *Cluster* 3 memiliki nilai rata-rata aspek kapasitas lembaga demokrasi yang tertinggi di antara *cluster* lainnya. *Cluster* 4 memiliki nilai yang cukup tinggi pada aspek kebebasan, tetapi rendah untuk aspek lainnya. Selain itu, terdapat

cluster yang menjadi *outlier* dari amatan lainnya, yaitu *cluster* 5 karena hanya terdapat 2 provinsi.

Dari penelitian ini, saran yang dapat diberikan kepada pemerintah yaitu perlu dilakukan perbaikan kualitas dan penjaminan kebebasan terhadap demokrasi di setiap daerah sehingga nilai indeks demokrasi meningkat untuk semua aspek. Untuk penelitian selanjutnya, dapat dilakukan analisis *clustering* dengan menggunakan metode lain seperti *hierarchical* atau *density-based clustering*.

DAFTAR PUSTAKA

- [1] Alfauzi, R., & Effendi, O. (2020). Pembatasan Kekuasaan Berdasarkan Paham Konstitusionalisme Di Negara Demokrasi. *Politica: Jurnal Hukum Tata Negara Dan Politik Islam*, 7(2), 111-133. <https://doi.org/10.32505/politica.v7i2.1990>.
- [2] BPS. (2022). *Statistik Politik 2022*.
- [3] Ghaisani, S. Y., Hikmah, N., Prasetyo, A. H., & Widodo, E. (2019). Analisis Cluster Hirarki Untuk Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016. *Prosiding Konferensi Nasional Penelitian Matematika Dan Pembelajarannya (KNPMP) IV*, Idi.
- [4] Nurdiana, N., Nilogiri, A., & Rahman, M. (2022). Penerapan Algoritma Fuzzy C-Means dan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia Application of The Fuzzy C-Means Algorithm and Elbow Method to Grouch Provinces in Indonesia Based on The Indonesian Democ. *Jurnal Smart Teknologi*, 3(5), 2774-1702. <http://jurnal.unmuhjember.ac.id/index.php/JST>.
- [5] Nasution, M., Ritonga, A. A., & Juledi, A. P. (2022). Implementasi Rapidminer dalam Mengklasifikasikan Indeks Demokrasi. *Computer Science and Information Technology (JCoInT)*, 3(3), 99-106.
- [6] Umarani, J., Manikandan, S., Centre, D., & Nadu, T. (2020). Implementation of Data Mining Concepts in R Programming. *International Journal of Trendy Research in Engineering and Technology*. 4(1), 1-7.
- [7] Leprince, J., Miller, C., & Zeiler, W. (2021). Data mining cubes for buildings, a generic framework for multidimensional analytics of building performance data. *Energy and Buildings*, 248, 111195. <https://doi.org/10.1016/j.enbuild.2021.111195>
- [8] Ahmad, T., & Aziz, M. N. (2019). Data preprocessing and feature selection for machine learning intrusion detection systems. *ICIC Express Letters*, 13(2), 93-101. <https://doi.org/10.24507/icicel.13.02.93>.
- [9] Adjie, M. K. (2020). Penerapan Algoritma FP-Growth pada Hasil Penimbangan Kendaraan Angkutan Barang. *UIN SUSKA RIAU*.
- [10] Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review. *Faktor Exacta*, 13(1), 35. <https://doi.org/10.30998/faktorexacta.v13i1.5548>.
- [11] Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia. *Inovasi*, 17(4), 770-782. <https://journal.feb.unmul.ac.id/index.php/INOVASI/article/download/10106/1437>.
- [12] Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., Pradiftha Junfithrana, A., Imamulhak, Y., & Putra, Y. (2022). Menerapkan K-Means Clustering untuk Segmentasi Gambar Database Berwarna. *FIDELITY : Jurnal Teknik Elektro*, 2(3), 57-61. <https://doi.org/10.52005/fidelity.v2i3.116>.
- [13] Abdunnassar, A. A., & R., N. L. (2023). Performance Analysis of Kmeans with Modified Initial Centroid Selection Algorithms and Developed Kmeans9 Model. *Measurement: Sensors*, 25. <https://doi.org/https://doi.org/10.1016/j.measen.2023.100666>
- [14] Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering (Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Kasus COVID-19 Menggunakan K-Means Clustering). *Jurnal Matematika, Statiska Dan Komputasi*, 17(1), 1-13. <https://doi.org/10.20956/jmsk.v>.
- [15] Sindi, S., Ningse, W. R. O., Sihombing, I. A., R.H.Zer, F. I., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. *Jurnal Teknologi Informasi*, 4(1), 166-173. <https://doi.org/10.36294/jurti.v4i1.1296>.
- [16] Sureja, N., Chawda, B., & Vasant, A. (2022). An improved K-medoids clustering approach based on the crow search algorithm. *Journal of Computational Mathematics and Data Science*, 3(March), 100034. <https://doi.org/10.1016/j.jcmds.2022.100034>

- [17] Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 687. <https://doi.org/10.30645/senaris.v1i10.75>.
- [18] Huang, Z., Yang, C., Chen, X., Zhou, X., & Gui, W. (2022). Time series clustering method with cluster validation to identify unknown local cell conditions in the aluminum reduction cell. *Computers and Industrial Engineering*, 174(2020), 108790. <https://doi.org/10.1016/j.cie.2022.108790>
- [19] Azzahra, A., & Wijayanto, A. W. (2022). Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal. *SISTEMASI: Jurnal Sistem Informasi*, 11(2), 481-495.
- [20] Manochandar, S., Punniyamoorthy, M., & Jeyachitra, R. K. (2020). Development of new seed with modified validity measures for k-means clustering. *Computers and Industrial Engineering*, 141(July 2019), 106290. <https://doi.org/10.1016/j.cie.2020.106290>
- [21] Khairati, A. F., Adlina, A. ., Hertono, G. ., & Handari, B. . (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. *PRISMA, Prosiding Seminar Nasional Matematika*, 2, 161-170. <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906>.
- [22] Adedeji, P. A., Olatunji, O. O., Madushele, N., Akinlabi, S. A., & Adeyemo, J. A. (2022). Cluster-based wind turbine maintenance prioritization for a utility-scale wind farm. *Procedia Computer Science*, 200, 1726-1735. <https://doi.org/10.1016/j.procs.2022.01.373>
- [23] Ghaderi, H., Foreman, B., Nayebi, A., Tipirneni, S., Reddy, C. K., & Subbian, V. (2023). A self-supervised learning-based approach to clustering multivariate time-series data with missing values (SLAC-Time): An application to TBI phenotyping. *Journal of Biomedical Informatics*, 143(February), 104401. <https://doi.org/10.1016/j.jbi.2023.104401>
- [24] Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1), 95-100. <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>.
- [25] Ouertani, M. W., Manita, G., & Korbaa, O. (2022). Automatic Data Clustering Using Hybrid Chaos Game Optimization with Particle Swarm Optimization Algorithm. *Procedia Computer Science*, 207, 2677-2687. <https://doi.org/10.1016/j.procs.2022.09.326>
- [26] Millati, K., Suhaeni, C., & Susetyo, B. (2021). Penggerombolan Daerah 3T di Indonesia Berdasarkan Rasio Tenaga Kesehatan dengan Metode Penggerombolan Berhierarki dan Cluster Ensemble. *Xplore: Journal of Statistics*, 10(2), 197-213. <https://doi.org/10.29244/xplore.v10i2.744>.