

Analysis of the Grouping of Provinces in Indonesia According to the Democracy Index With the Agglomerative Hierarchical Clustering Algorithm

Mochammad Yusuf Maulana ^{a1}, Arie Wahyu Wijayanto ^{a2}

^aPoliteknik Statistika STIS
Jalan Otto Iskandardinata 64C Jakarta, Indonesia

1211810424@stis.ac.id

ariewahyu@stis.ac.id

Abstract

The purpose of this study is to clustering the provinces in Indonesia based on the IDI dimensions in 2020. This study compares four methods of agglomerative hierarchical clustering, namely the single linkage method, average linkage, complete linkage, and the ward method. The evaluation criteria used are the silhouette index, dunn index, connectivity index, and agglomerative coefficient. The best method used is the average linkage hierarchy method with two clusters. The combination of two clusters and the ward's method produces a silhouette index value of 0.34, a dunn index value of 0.28, a connectivity index value of 5.64 and an agglomerative coefficient of 0.891 which illustrates that the method is able to classify provinces in Indonesia based on the 2020 IDI. The results of this study can be used as a consideration for the government to take policies related to policies in Indonesia.

Keywords: *Data mining, Indonesian Democracy Index, clustering, ward's method.*

Abstrak

Tujuan dari penelitian ini adalah mengelompokkan provinsi-provinsi di Indonesia berdasarkan dimensi Indeks Demokrasi Indonesia (IDI) tahun 2020. Salah satu kegunaan IDI yaitu dapat digunakan sebagai pengukuran kuantitatif terhadap kebebasan rakyat Indonesia. Penelitian ini membandingkan empat metode *agglomerative hierarchical clustering*, yaitu metode *single linkage*, *average linkage*, *complete linkage*, dan *ward's method*. Kriteria evaluasi yang digunakan adalah indeks *silhouette*, indeks *dunn*, indeks *connectivity*, dan *agglomerative coefficient*. Metode terbaik yang digunakan adalah metode hierarki *average linkage* dengan jumlah kluster yang terbentuk sebanyak dua. Kombinasi jumlah kluster sebanyak dua dan metode *ward* menghasilkan nilai indeks *silhouette* sebesar 0.34, nilai indeks *dunn* sebesar 0.28, nilai indeks *connectivity* sebesar 5.64 dan *agglomerative coefficient* sebesar 0,891 yang menggambarkan bahwa metode tersebut mampu mengelompokkan provinsi-provinsi di Indonesia berdasarkan IDI tahun 2020. Hasil penelitian ini dapat digunakan sebagai pertimbangan bagi pemerintah untuk mengambil suatu kebijakan terkait demokrasi di Indonesia.

Keywords: *Data mining, Indonesian Democracy Index, clustering, ward's method.*

1. PENDAHULUAN

Sejak runtuhnya orde baru pada tahun 1998 sampai saat ini, demokrasi menjadi pilihan dalam sistem politik di Indonesia. Abraham Lincoln menyatakan bahwa demokrasi merupakan suatu pemerintahan dari rakyat, oleh rakyat, dan untuk rakyat [1]. Hal ini dapat dikatakan bahwa kekuasaan tertinggi berada di tangan rakyat dan rakyat juga memiliki kebebasan dalam menentukan pilihannya.

Mulai tahun 2009, UNDP bekerja sama dengan Bappenas, BPS, Kemendagri, dan juga pemerintah provinsi melakukan pengukuran kuantitatif terhadap kebebasan rakyat Indonesia

melalui Indeks Demokrasi Indonesia (IDI). Indeks tersebut mengukur tingkat demokrasi dari aspek kebebasan sipil, hak-hak politik, dan lembaga demokrasi [2]. IDI berguna bagi pemerintah daerah, pusat, dan provinsi sebagai evaluasi dalam penyelenggaraan demokrasi dan untuk melakukan peningkatan kualitas demokrasi [2].

Keberhasilan kebijakan pembangunan sangat dipengaruhi oleh ketepatan dalam identifikasi kelompok dan area yang menjadi target [3]. Strategi yang dapat dilakukan adalah dengan melihat kesamaan karakteristik suatu provinsi dengan provinsi lainnya menurut variabel IDI. Hal ini akan memudahkan pemerintah melakukan identifikasi provinsi-provinsi mana saja yang memerlukan prioritas dalam pemajuan demokrasi. Oleh karena itu, penting bagi pemerintah untuk mengetahui kesamaan karakteristik variabel IDI dari provinsi-provinsi di Indonesia.

Analisis *cluster* dapat digunakan untuk melihat ukuran kemiripan karakteristik dari kelompok-kelompok yang terbentuk [4]. Pembentukan kelompok tersebut didasarkan atas kesamaan karakteristiknya. Provinsi-provinsi yang memiliki karakteristik variabel pembentuk IDI yang mirip akan berada dalam satu kelompok. Melalui cara tersebut, memungkinkan pemerintah untuk lebih mudah dalam melihat karakteristik variabel-variabel pembentuk IDI dari kelompok provinsi yang terbentuk dan mengambil kebijakan yang tepat untuk memajukan demokrasi di Indonesia.

Terdapat dua macam metode analisis kluster yang dapat digunakan, yaitu *partitioning* dan *hierarchical clustering*. Proses pengelompokan pada metode *partitioning* atau metode non-hirarki terlebih dahulu menentukan jumlah *cluster* dan titik *centroid*, sedangkan metode *hierarchical* dilakukan pengelompokan secara bertahap dengan banyaknya *cluster* yang belum diketahui [5]. Penelitian terdahulu yang pernah dilakukan yaitu pengelompokan 34 provinsi Indonesia berdasarkan indeks demokrasi Indonesia tahun 2016 [6].

Berdasarkan penelitian yang dilakukan tersebut, peneliti ingin melakukan analisis kluster dengan metode *hierarchical clustering* untuk mengelompokkan provinsi di seluruh Indonesia berdasarkan indeks demokrasi Indonesia tahun 2020.

2. METODE PENELITIAN

2.1. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik (BPS). Dataset dan *source code* R yang digunakan dapat diakses melalui <https://github.com/emyeem/Clustering-IDI-2020.git>. Data yang digunakan adalah skor dimensi Indeks Demokrasi Indonesia (IDI) berdasarkan provinsi tahun 2020. Terdapat tiga dimensi yang digunakan, yaitu dimensi Aspek Kebebasan Sipil (AKS), Aspek Hak-hak Politik (AHP), dan Aspek Lembaga Demokrasi (ALD). Berikut disajikan tabel 1 mengenai deskripsi variabel yang digunakan.

Tabel 1. Deskripsi Variabel

No.	Nama Variabel	Skala Pengukuran
1	Aspek Kebebasan Sipil (AKS)	Rasio
2	Aspek Hak-hak Politik (AHP)	Rasio
3	Aspek Lembaga Demokrasi (ALD)	Rasio

2.2. Analisis Cluster

Analisis *cluster* merupakan salah satu tipe dari *unsupervised learning* yang paling umum digunakan dalam *data mining* yang mengelompokkan objek-objek pengamatan menjadi beberapa kelompok [7]. Tujuannya adalah untuk mengelompokkan data yang memiliki karakteristik yang sama, sedangkan data dengan karakteristik yang berbeda dikelompokkan di kluster yang lainnya [8].

Sebelum dilakukan analisis *cluster*, perlu dicek terlebih dahulu apakah terdapat korelasi antar variabel dalam data dengan uji Bartlett. Jika terdapat korelasi, maka perlu dilakukan analisis faktor sebelum kemudian dapat dilanjutkan analisis *cluster*. Salah satu metode untuk analisis

faktor adalah dengan Analisis Komponen Utama (AKU), yang mereduksi data menjadi dimensi yang lebih sedikit [9]. Prinsip utama dalam AKU yaitu terdapat korelasi yang tinggi antar variabel, sehingga variabel-variabel tersebut dapat direduksi [10]. Uji Bartlett dapat dilakukan dengan persamaan berikut [11]:

Hipotesis:

H_0 : $R = 1$ (matriks korelasi sama dengan matriks identitas atau terdapat korelasi)

H_1 : $R \neq 1$ (matriks korelasi tidak sama dengan matriks identitas atau tidak terdapat korelasi)

Statistik Uji:

$$Bartlett = \ln |R| \left(n - 1 - \left(\frac{2p+5}{6} \right) \right) \quad (1)$$

dengan,

$|R|$: nilai determinan dari matriks korelasi

n : banyaknya pengamatan

p : banyaknya variabel.

Kriteria pengujian:

Tolak H_0 jika $Bartlett > X_{\frac{(p+1)(p-2)}{2}(\alpha)}^2$ atau p-value $< \alpha$, dapat dikatakan bahwa variabel-variabel saling berkorelasi atau terdapat hubungan antar variabel, sehingga perlu dilakukan analisis faktor/reduksi variabel.

2.3. Hierarchical Clustering

Teknik ini akan mengelompokkan objek-objek yang mirip pada hierarki yang berdekatan dan objek-objek yang berbeda pada hierarki yang berjauhan. Metode hierarki membangun kluster bersarang dengan menggabungkan atau memisahkannya secara berurutan [12]. Metode ini merupakan suatu metode analisis *cluster* yang dilakukan secara bertahap. Pada setiap urutan tahapan pembentukan kluster, partisi baru secara optimal digabungkan atau dipisah dari partisi sebelumnya menurut beberapa kriteria kecukupan [13]. Hasil dari metode ini dapat disajikan dalam bentuk dendogram, yang merupakan representatif visual dari seluruh tahapan yang menunjukkan proses kluster terbentuk. Selain itu juga terdapat nilai koefisien jarak pada setiap tahapannya [14]. Nilai koefisien tersebut ditentukan pada setiap pasangan poin. Terdapat berbagai macam ukuran pendekatan yang dapat digunakan untuk mengukur kesamaan jarak, antara lain skor kesamaan negatif, jarak *Euclidean* dan jarak *Manhattan* [15].

Terdapat dua macam strategi yang dapat digunakan untuk melakukan *hierarchical clustering*, yaitu *agglomerative* dan *divisive*. Perbedaannya terletak pada arah pengelompokan. Metode *agglomerative* menggunakan strategi *bottom-up* yang pada awalnya setiap objek memiliki kluster sendiri-sendiri, kemudian kluster tersebut digabung menjadi kluster-kluster yang lebih besar. Sementara itu, *divisive* menggunakan strategi *top-down* yang pada awalnya seluruh objek berada pada satu kluster besar, kemudian kluster besar tersebut dipecah-pecah menjadi kluster-kluster yang lebih kecil [16]. Terdapat lima metode *agglomerative* yang dapat digunakan dalam pembentukan kluster, yaitu *single linkage*, *average linkage*, *complete linkage*, *centroid*, dan *ward's method*. Namun, pada penelitian ini hanya akan menggunakan *single linkage*, *average linkage*, *complete linkage*, dan *ward's method* saja. Pada dasarnya, perbedaan kelima metode tersebut terdapat pada jenis jarak yang digunakan untuk menghitung tingkat kemiripan data poin.

- *Single linkage*; menggunakan jarak terkecil antara satu objek dalam kluster dengan objek dalam kluster lainnya. $d(C_i, C_j) = \min\{d(x_{ip}, x_{jq})\}$.
- *Average linkage*; menggunakan jarak rata-rata antar objek-objek dalam satu kluster dengan objek-objek dalam kluster lainnya. $d(C_i, C_j) = \max\{d(x_{ip}, x_{jq})\}$.
- *Complete linkage*; menggunakan jarak terbesar antara satu objek dalam kluster dengan objek dalam kluster lainnya. $d(C_i, C_j) = \text{avg}\{d(x_{ip}, x_{jq})\}$.
- *Centroid*; menggunakan jarak antar *centroid* (titik tengah) dari setiap kluster. $d_{12} = d(\bar{X}, \bar{Y})$.
- *Ward's method*; tidak hanya berdasarkan jarak antar observasi, tetapi juga memerhatikan keragaman dari kluster tersebut dengan meminimalkan nilai *Sum of*

Squared Error. Dua kluster digabungkan ketika memberikan nilai SSE yang paling kecil/minimal.

2.4. Cluster Validation

Uji validitas kluster dilakukan untuk menghindari ketidaksesuaian hasil dan memastikan bahwa hasil dari kluster yang terbentuk mencerminkan populasi secara umum. Terdapat beberapa metode yang dapat digunakan untuk menguji validitas hasil kluster, salah satunya validasi internal yang terdiri dari indeks *connectivity*, indeks *silhouette*, dan indeks *dunn*.

Indeks *connectivity* memiliki nilai antara 0 sampai tak hingga. Semakin kecil nilai indeks *connectivity*, maka semakin baik kluster yang terbentuk. Sementara itu, indeks *silhouette* mengukur derajat kepercayaan dalam proses *clustering* pada pengamatan tertentu. Pada indeks *silhouette*, kluster yang terbentuk dikatakan baik yaitu ketika nilai indeksnya mendekati 1. Sementara itu, indeks *dunn* merupakan rasio jarak terkecil antara observasi pada kluster yang berbeda dengan jarak terbesar pada masing-masing kluster. Semakin tinggi nilai indeks *dunn*, maka semakin baik kluster yang dihasilkan [17].

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Statistik

Berikut adalah deskripsi mengenai variabel dimensi IDI berdasarkan provinsi tahun 2020.

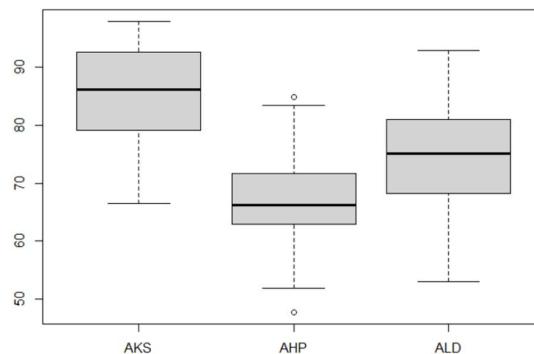
Tabel 2. Deskripsi Statistik

Variabel	Minimum	Maximum	Mean
AKS	66,59	97,93	84,75
AHP	47,78	84,95	66,67
ALD	53,09	92,97	74,79

Tabel 2 menunjukkan deskripsi statistik dari tiga variabel penelitian yang akan digunakan. Berdasarkan tabel tersebut dapat diketahui bahwa satuan variabel yang digunakan setara, sehingga tidak perlu dilakukan standarisasi.

3.2. Identifikasi Outlier

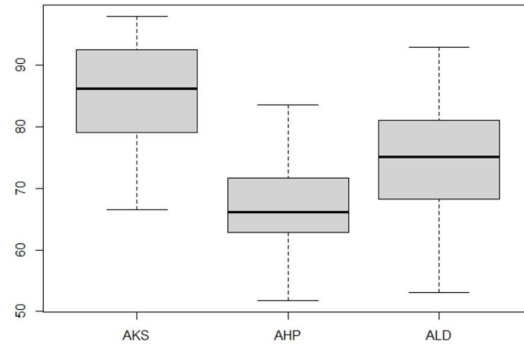
Pengecekan outlier data merupakan salah satu tahapan dari *data preprocessing* yang perlu dilakukan sebelum analisis lebih lanjut. Selain cek outlier data, pada *preprocessing* juga perlu dilakukan *cleaning data*, seperti pengecekan *missing value*, membuat duplikasi data, dan memperbaiki kesalahan pada data [18]. Berikut adalah boxplot mengenai sebaran data.



Gambar 1. Boxplot sebaran data IDI 2020.

Data dianggap sebagai outlier yaitu ketika berada di luar boxplot. Berdasarkan gambar 1, terlihat bahwa terdapat 2 outlier pada dimensi Aspek Hak Politik (AHP). Metode *winsorizing* dapat dilakukan untuk mengatasi data yang memiliki outlier dan tidak terpenuhinya asumsi [19]. Metode tersebut dapat mengubah nilai ekstrem menjadi nilai batas atas atau bawahnya [20]. Setelah

dilakukan metode *winsorizing* sebagai penanganan outlier, maka data telah bebas dari outlier. Berikut adalah boxplot mengenai sebaran data yang sudah terbebas dari outlier.



Gambar 2. Boxplot data bebas outlier.

3.3. Pengecekan Asumsi Multikolinearitas

Uji Bartlett dilakukan untuk mendeteksi gejala multikolinearitas pada data. Suatu data dikatakan memiliki gejala multikolinearitas yaitu ketika data tersebut memiliki korelasi yang kuat dengan data lainnya. Jika terdapat korelasi pada data, perlu dilakukan reduksi data untuk mengatasi gejala multikolinearitas tersebut. Berikut hasil perhitungan uji Bartlett pada data.

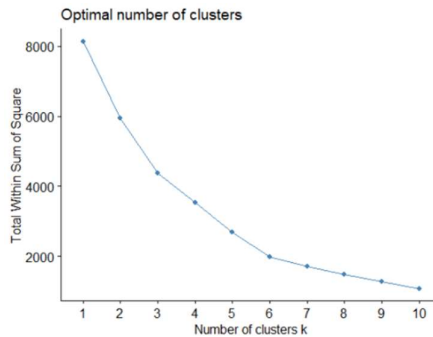
Tabel 3. Hasil Uji Bartlett.

Uji	Nilai
Bartlett	Approx. 2,331
Chi-Square	
df	3
p-value	0,51

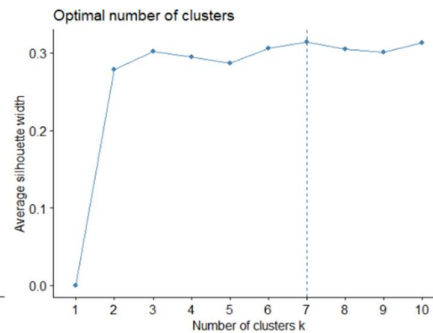
Berdasarkan tabel 3, terlihat bahwa nilai *p-value* sebesar 0,51 atau lebih besar dari tingkat signifikansi (α) 5%, sehingga keputusannya gagal tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa tidak terdapat gejala multikolinearitas pada data, sehingga tidak perlu dilakukan reduksi data.

3.4. Penentuan Jumlah Cluster Optimum

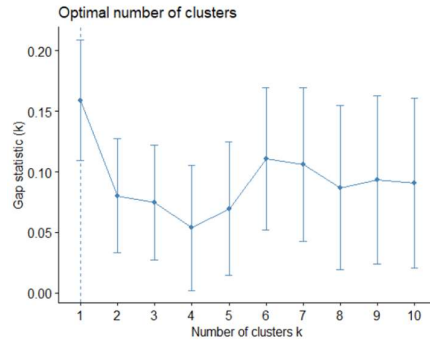
Penentuan jumlah kluster optimum dilakukan dengan metode *elbow*, metode *silhouette*, dan metode gap statistik. Berikut hasil visualisasi setiap metode penentuan jumlah kluster optimum.



Gambar 3. Elbow Method



Gambar 4. Silhouette Method



Gambar 5. Gap Statistik

Berdasarkan gambar di atas, jumlah kluster optimum pada metode *elbow* sebanyak 6 karena grafik melandai ketika kluster 6. Sementara itu, pada metode *silhouette* terlihat bahwa kluster optimum sebanyak 7, sedangkan kluster optimum pada metode gap statistik sebanyak 1. Dari ketiga metode tersebut, tidak ada metode yang menghasilkan nilai yang sama dalam menentukan banyaknya kluster optimum. Untuk mengatasi hal ini, perlu dilakukan uji validitas internal untuk menentukan berapa banyak kluster optimum yang akan digunakan. Nilai indeks validasi kluster internal dapat dilihat pada tabel 4 berikut.

Tabel 4. Nilai Indeks Validitas.

\sum kluster	2	3	4	5	6	7
<i>Connectivity</i>	5,64	12,9	22,2	23,7	26,4	28,8
<i>Dunn</i>	0,28	0,23	0,17	0,17	0,21	0,21
<i>Silhouette</i>	0,34	0,30	0,33	0,33	0,32	0,31

Tabel 5. Nilai Optimal

	Skor	\sum kluster
<i>Connectivity</i>	5,64	2
<i>Dunn</i>	0,28	2
<i>Silhouette</i>	0,34	2

Berdasarkan tabel 5, terlihat bahwa metode hierarki dengan jumlah kluster sebanyak 2 adalah yang paling optimal. Nilai indeks *connectivity* sebesar 5,64, nilai indeks *dunn* sebesar 0,28, dan nilai indeks *silhouette* sebesar 0,34.

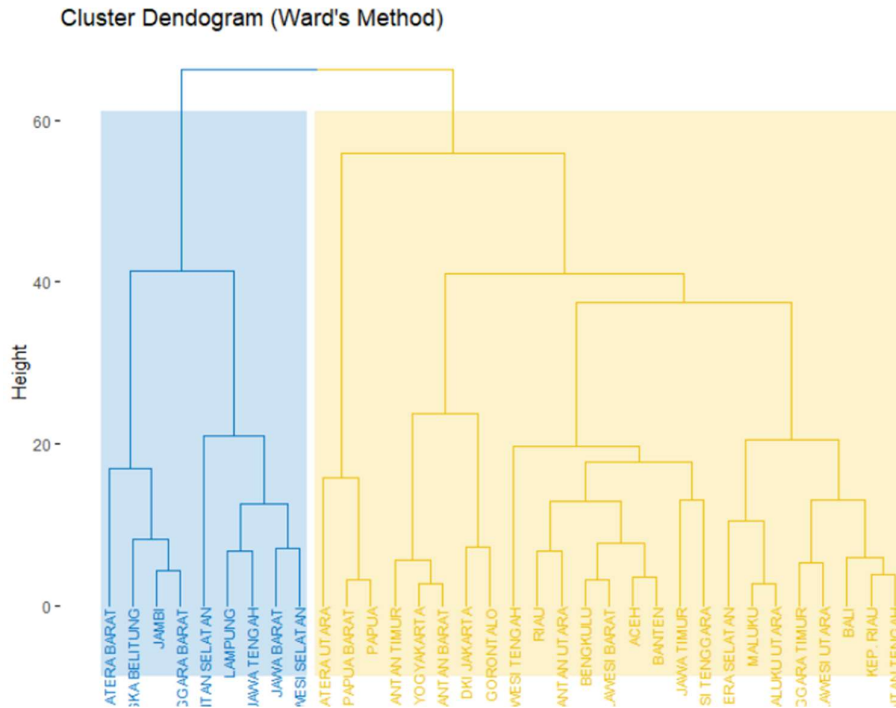
3.5. Pemilihan Metode Hierarki Terbaik

Hasil uji validitas menunjukkan bahwa metode terbaik yang dapat digunakan adalah metode hierarki dengan jumlah kluster 2. Hal ini disebabkan nilai indeks *connectivity* yang lebih kecil dan nilai indeks *silhouette* dan *dunn* yang mendekati 1. Oleh karena itu, peneliti menggunakan metode hierarki dengan jumlah kluster sebanyak 2 untuk mengelompokkan data IDI tahun 2020. Pemilihan metode hierarki terbaik didasarkan pada *agglomerative coefficient* yang nilainya paling besar [5]. Hasil perhitungan *agglomerative coefficient* dapat dilihat pada tabel 6 berikut.

Tabel 6. Hasil Perhitungan.

Metode	<i>Agglomerative Coefficient</i>
<i>Single</i>	0,588
<i>Average</i>	0,762
<i>Complete</i>	0,846
<i>Ward</i>	0,891

Berdasarkan tabel 6, didapatkan bahwa metode hierarki yang menghasilkan kluster terbaik adalah metode *ward* karena memiliki nilai *agglomerative coefficient* paling besar di antara metode hierarki yang lain. Hasil visualisasi pengelompokan provinsi berdasarkan IDI tahun 2020 dapat dilihat pada gambar 3.6 berikut.



Gambar 6. Visualisasi Dendrogram

3.6. Cluster Profiling

Tahapan ini merupakan salah satu cara untuk menginterpretasikan hasil akhir dari pembentukan kluster. Caranya adalah dengan mengidentifikasi rata-rata atau *centroid* dari setiap atribut di dalam suatu kluster dan dibandingkan dengan *centroid* pada kluster lainnya. Cluster profiling disajikan pada tabel 7 berikut.

Tabel 7. Profiling Hasil Kluster.

Kluster	AKS	AHP	ALD
1	88,995	65,580	72,772
2	72,954	69,569	80,384

Hasil clustering dari metode Ward menghasilkan dua kluster. Terdapat 25 provinsi yang dikelompokkan dalam kluster pertama dan 9 provinsi dalam kluster kedua. Interpretasi pada setiap kluster yang terbentuk adalah sebagai berikut:

- Kluster 1 terdiri dari 25 provinsi yang memiliki rata-rata lebih tinggi pada dimensi Aspek Kebebasan Sipil (AKS) dengan skor 88,995.
- Kluster 2 terdiri dari 9 provinsi yang memiliki rata-rata lebih tinggi pada dimensi Aspek Hak-hak Politik dan Aspek Lembaga Demokrasi (ALD) dengan skor 72,954.

4. KESIMPULAN

Berdasarkan penghitungan yang telah dilakukan, metode terbaik untuk mengelompokkan provinsi di Indonesia berdasarkan IDI tahun 2020 adalah metode *hierarchical clustering* dengan *ward's method*. Terdapat dua kluster yang terbentuk dari *hierarchical clustering* dengan *ward's method*, yaitu kluster 1 yang terdiri dari 25 provinsi dan kluster 2 terdiri dari 9 provinsi. Kluster 1 terdiri dari 25 provinsi yang memiliki rata-rata lebih tinggi pada dimensi Aspek Kebebasan Sipil. Sementara itu, pada kluster 2 terdiri dari 9 provinsi yang memiliki rata-rata lebih tinggi pada dimensi Aspek Hak-hak Politik dan Aspek Lembaga Demokrasi.

Berdasarkan kesimpulan yang diperoleh di atas, penulis menyarankan pada peneliti selanjutnya dapat menggunakan algoritma dan uji validitas lain untuk dibandingkan mana yang

lebih cocok. Bagi pemangku kebijakan, penelitian ini dapat digunakan sebagai pertimbangan lain dalam mengambil suatu kebijakan terkait demokrasi di Indonesia.

References

- [1] P. Nurwardani, *Pendidikan Kewarganegaraan untuk Perguruan Tinggi*. Jakarta: Direktorat Jenderal Pembelajaran dan Kemahasiswaan, 2016.
- [2] M. Raul, S. Hidayat, A. M. Gismar, S. M. Mulia, and A. Parengkuan, "MENAKAR DEMOKRASI DI INDONESIA," Jakarta, 2011.
- [3] F. Basri, *Perekonomian Indonesia Menjelang Abad XXI*. Jakarta: Erlangga, 1997.
- [4] C. Hennig, M. Meila, F. Murtagh, and R. Rocci, *Handbook of Cluster Analysis*. CRC Press, 2016.
- [5] N. Afira and A. W. Wijayanto, "Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019," *Komputika : Jurnal Sistem Komputer*, vol. 10, no. 2, pp. 101–109, Sep. 2021, doi: 10.34010/komputika.v10i2.4317.
- [6] S. Y. Ghaisani, N. Hikmah, A. H. Prasetyo, and E. Widodo, "ANALISIS CLUSTER HIRARKI UNTUK PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR DEMOKRASI INDONESIA TAHUN 2016," 2019.
- [7] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey: Pearson Education, 2007.
- [8] A. Sani, "PENERAPAN METODE K-MEANS CLUSTERING PADA PERUSAHAAN," *Jurnal Ilmiah Teknologi Informasi*, vol. 353, pp. 1–7, 2018.
- [9] N. Thamrin and A. W. Wijayanto, "Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island," *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 1, pp. 141–160, Mar. 2021, doi: 10.29244/ijsa.v5i1p141-160.
- [10] A. R. Damayanti and A. W. Wijayanto, "Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018," *EIGEN MATHEMATICS JOURNAL*, vol. 4, no. 1, pp. 8–17, Jun. 2021, doi: 10.29303/emj.v4i1.89.
- [11] S. Machfudhoh and N. Wahyuningsih, "Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur," *Jurnal Sains dan Seni POMITS*, vol. 2, no. 1, pp. 1–8, 2013.
- [12] A. M. Sikana and A. W. Wijayanto, "Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering," *Jurnal Ilmu Komputer*, vol. 14, no. 2, pp. 66–78, 2021.
- [13] S. Landau, "Cluster Analysis: Overview," 2010.
- [14] L. Ramadhani, I. Purnamasari, F. Deny, and T. Amijaya, "Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016)," *Jurnal EKSPONENSIAL*, vol. 9, no. 1, pp. 1–10, 2018.
- [15] K. Yeturu, "Machine learning algorithms, applications, and practices in data science," *Handbook of Statistics*, vol. 43, pp. 81–206, 2020, doi: 10.1016/bs.host.2020.01.002.

- [16] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data Mining dengan R*. Jakarta: IN MEDIA, 2018.
- [17] E. Irwansyah and M. Faisal, *Advanced Clustering Teori dan Aplikasi*. Yogyakarta: Deepublish, 2015.
- [18] E. Luthfi, A. Wahyu Wijayanto, and P. Statistika, "Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia," *INOVASI*, vol. 17, no. 4, pp. 761–773, 2021, [Online]. Available: <http://journal.feb.unmul.ac.id/index.php/INOVASI>
- [19] M. Destriana, N. Gusriani, and I. Irianingsih, "Klasifikasi Status Kinerja Bank yang Terdaftar di BEI dengan Pendekatan Winsorized Modified One-Step M-Estimator," *Jurnal Matematika Integratif*, vol. 14, no. 2, pp. 135–142, 2018, doi: 10.24198/jmi.v14.n2.2018.135-142.
- [20] Bookdown, "R Software Handbook." <https://bookdown.org/aschmi11/RESMHandbook/> (accessed Dec. 18, 2021).