



Comparison Analysis of Clustering Methods for Clustering of Indonesian's Gender Empowerment Conditions in 2022

Aisyah 'Azizah Nur Rahmah^{a,*}, Arie Wahyu Wijayanto^b

^aPoliteknik Statistika STIS, Jl. Otto Iskandardinata Nomor 64c, Bidaracina, Jatinegara Kota Jakarta Timur, Kode Pos 13330, Indonesia. Email: 212011312@stis.ac.id

^bPoliteknik Statistika STIS, K Jl. Otto Iskandardinata Nomor 64c, Bidaracina, Jatinegara Kota Jakarta Timur, Kode Pos 13330, Indonesia. Email: ariewahyu@stis.ac.id

ABSTRACT

Gender empowerment is one of the components of gender development achievement measures that is an important agenda at the global level in realizing the Sustainable Development Goals. The Gender Empowerment Index (GEI) of Indonesia has been continuously improving since 2010, indicating an increasing involvement of women in various areas of life. However, behind this upward trend in GEI, there is still inequality at the provincial level. Therefore, there is a need to formulate development strategies, one of which is gender-based. One possible step is to categorize regions in Indonesia based on their gender empowerment characteristics so that government interventions can be targeted effectively. This research utilizes two clustering approaches, namely Hierarchical Methods and Partitioning Methods, with data consisting of three variables representing the components of GEI for 34 provinces in Indonesia in 2022. The selection of the best method and number of clusters is based on internal and stability validity, followed by the determination of the smallest within and between standard deviation ratios. From the cluster analysis results, the best method is found to be K-means with a total of 5 clusters.

Keywords : cluster, gender, hierarchical methods, partitioning methods, empowerment

Diserahkan: 01-08-2023; Diterima: 30-12-2023;

Doi: <https://doi.org/10.29303/emj.v6i2.176>

1. Pendahuluan

Sustainable Development Goals merupakan agenda pembangunan global yang telah disepakati dan menjadi komitmen seluruh negara dalam menyatukan prinsip kesejahteraan. Kesetaraan gender adalah satu dari sekian banyak isu global yang ada, sehingga kesetaraan gender termasuk dalam tujuan pembangunan berkelanjutan tersebut sebagaimana tertuang dalam tujuan kelima yakni “Mencapai kesetaraan gender dan memberdayakan semua perempuan dan anak perempuan”. Untuk mewujudkan kesetaraan gender, peran aktif perempuan dalam ekonomi

dan politik perlu menjadi perhatian karena mencerminkan kondisi pemberdayaan masyarakat berdasarkan jenis kelamin (KPPA & BPS, 2022).

Menurut Bappenas, gender adalah perbedaan terkait peran, atribut, sifat, sikap, dan perilaku yang tumbuh dan berkembang dalam masyarakat. Gender yang kerap kali menjadi objek diskriminatif sering disamaartikan dengan jenis kelamin, padahal dua hal tersebut memiliki makna yang berbeda. Jenis kelamin merupakan hal-hal yang berkaitan dengan fisik atau biologis seseorang, sedangkan gender direpresentasikan dengan perilaku yang menggambarkan kedudukan perempuan dan laki-laki dalam suatu kelompok sosial masyarakat (Judiasih, 2022).

* Corresponding author.

Alamat e-mail: 212011312@stis.ac.id

Ketimpangan gender yang masih terjadi di Indonesia menjadi bagian dari perhatian bersama dalam rangka penguatan Pengarusutamaan Gender (PUG) sebagaimana tertuang dalam Instruksi Presiden Nomor 9 Tahun 2000 tentang Pengarusutamaan Gender dalam Pembangunan Nasional dengan memerhatikan kondisi di setiap wilayahnya. Salah satu ukuran untuk melihat capaian pembangunan berbasis gender ini adalah Indeks Pemberdayaan Gender (IDG). IDG ini menekankan dalam hal partisipasi, apakah perempuan dapat memainkan peranan aktif dalam kehidupan politik dan ekonomi (BPS, 2023).

Konsep pemberdayaan memiliki perubahan dalam riwayat sejarah yang panjang. Hal ini berkaitan dengan transformasi dalam penerimaan hak perempuan dan kesetaraan yang lebih besar antara laki-laki dan perempuan (Cornwall, 2016). Namun, ukuran pemberdayaan sebagaimana yang telah dikembangkan oleh *United Nations Development Programme* (UNDP) sejak pertengahan 1990-an ini juga memiliki keterbatasan dalam aspek relasional kehidupan perempuan (Porter, 2013).

Untuk melihat gambaran kondisi pemberdayaan gender di setiap wilayahnya dapat dilakukan dengan pengelompokan wilayah untuk level provinsi berdasarkan karakteristik pemberdayaan gender yang ada. Pengelompokan tersebut salah satunya dapat didekati dengan analisis *cluster*. Terdapat dua metode *cluster* yang sering digunakan dalam penelitian, yakni *hierarchical* dan *partitioning*. Metode *hierarchical* dilakukan bertahap dengan jumlah *cluster* belum diketahui, sedangkan metode *partitioning* telah menentukan jumlah *cluster* di awal dengan pemilihan titik pusat terlebih dahulu (Afira & Wijayanto, 2021).

Dari hasil pengelompokan wilayah yang terbentuk berdasarkan model *cluster* terbaik, akan didapatkan kelompok wilayah untuk level provinsi dengan karakteristik komponen penyusun IDG yang saling berdekatan. Pada penelitian ini akan dilakukan perbandingan metode *clustering* untuk mengelompokkan provinsi di Indonesia berdasarkan kondisi pemberdayaan gender tahun 2022. Dengan demikian, dapat dilakukan perumusan kebijakan yang sesuai dengan kondisi pemberdayaan gender di wilayah tersebut mengingat setiap wilayah memiliki karakteristiknya masing-masing dan hasil dari kebijakan yang dibuat akan lebih tepat sasaran dalam upaya meningkatkan pengarusutamaan gender.

Penelitian ini dilakukan dengan merujuk pada beberapa penelitian terdahulu yang dijadikan referensi. Penelitian terdahulu yang pernah dilakukan adalah pengelompokan kabupaten/ kota di Sumatera Barat untuk melihat kondisi ketimpangan gender dengan menggunakan biplot (Diana, 2018). Selanjutnya, dengan metode *hierarchical clustering*, Nisa (2018) melakukan pengelompokan wilayah kabupaten/ kota di Jawa Timur untuk meninjau disparitas pembangunan manusia berbasis gender yang terjadi. Sukim et al., (2018) melakukan pengukuran tentang kepemimpinan perempuan di Indonesia dengan *Fuzzy c-Means Clustering* untuk level provinsi. Pengelompokan wilayah ini juga dapat dilakukan berdasarkan data gabungan *cross section*

dan *time series* sebagaimana dilakukan oleh Nurfadilah (2021) yang melihat pengelompokan wilayah kabupaten/ kota di Sulawesi Selatan berdasarkan kondisi pembangunan manusia berbasis gender menggunakan *clustering* longitudinal. Selain itu, terdapat pula penelitian yang dilakukan oleh Salsabila & Hendrawan (2021) untuk melihat kondisi pemberdayaan gender di Indonesia pascapandemi Covid-19 dengan *agglomerative hierarchical clustering* dan biplot.

2. Metodologi

2.1. Data dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari *website* Badan Pusat Statistik RI. Data tersebut adalah Indeks Pemberdayaan Gender beserta komponen-komponen pembentuknya yang akan digunakan sebagai variabel penelitian. Analisis ini menggunakan data tahun 2022 sebagai data terbaru pada saat penelitian dilakukan. Adapun jumlah unit analisis (provinsi) yang akan digunakan adalah sebanyak 34 provinsi yang tersebar di seluruh Indonesia. Berikut merupakan variabel yang digunakan dalam penelitian ini.

Tabel 1 – Variabel yang Digunakan.

Variabel	Deskripsi	Satuan
X_1	Keterlibatan perempuan dalam paerlemen	Persen
X_2	Perempuan sebagai tenaga profesional	Persen
X_3	Sumbangan pendapatan perempuan	Persen

2.2. Landasan Teori

2.2.1. Indeks Pemberdayaan Gender

Indeks Pemberdayaan Gender (IDG) merupakan salah satu indikator yang tercakup dalam pembangunan gender untuk menunjukkan apakah perempuan dapat memainkan peran aktif dalam politik dan ekonomi (BPS, 2023). IDG diukur dalam tiga dimensi, yang terdiri dari keterwakilan di parlemen, pengampilan keputusan, dan distribusi pendapatan. Penghitungan IDG didapatkan dari hasil rata-rata aritmatika dari tiga komponen pembentuknya. Masing-masing dimensi diukur dengan sebuah indikator tertentu yang dapat dilihat pada tabel berikut.

Tabel 2 – Bidang, Dimensi, dan Indikator IDG.

Bidang	Dimensi	Indikator
Politik	Keterwakilan perempuan di parlemen	Proporsi keterwakilan perempuan & laki-laki di parlemen

Bidang	Dimensi	Indikator
Manajerial	Pengambilan keputusan	Proporsi perempuan & laki-laki dari manajer, staf administrasi, pekerja profesional dan teknisi
Ekonomi	Sumbangan pendapatan	Upah buruh laki-laki & perempuan nonpertanian

2.2.2. Clustering Analysis

Data mining dapat mengekstraksi pengetahuan dari data historis untuk memprediksi hasil di masa yang akan datang. *Clustering* menjadi bagian penting dari data mining tersebut (Rani & Rohil, 2013). *Clustering analysis* atau analisis kluster merupakan teknik yang digunakan untuk mengelompokkan objek yang relatif homogen ke dalam sebuah grup atau disebut juga *cluster* (Astuti & Rezanita, 2022). *Clustering* atau disebut juga sebagai *grouping* pada dasarnya berbeda dengan klasifikasi. Dalam klasifikasi, observasi yang ada ditempatkan pada kelompok yang sudah ditentukan sebelumnya. Sedangkan dalam *clustering*, jumlah kelompok dan struktur kelompok belum diketahui sebelumnya (Johnson & Wichern, 2007). Tujuan dari pengelompokkan dengan *clustering* adalah mengelompokkan objek dengan sifat heterogen antarkelompok, tetapi memiliki kemiripan dalam satu kelompok yang relatif homogen (Talakua et al., 2017).

Proses pengelompokkan ini menggunakan konsep jarak sebagai dasarnya. Jarak menjelaskan kedekatan antarobservasi dalam melihat struktur grup yang ada (Hasanah, 2022). Ada beberapa jenis jarak yang dapat digunakan dalam *clustering analysis*, seperti jarak *Euclidean*, jarak *Manhattan*, jarak *Canberra*, jarak *Mahalanobis*, *Minkowski metric*, dan *Czekanowski coefficient*. Ukuran yang paling umum digunakan adalah jarak *Euclidean* (Johnson & Wichern, 2007). Jarak *Euclidean* mengukur jarak dua observasi yang dihubungkan secara garis lurus. Persamaan *Euclidean* tersebut mencari jarak antar objek (P, Q) di mana $P = (x_1, x_2, \dots, x_p)$ dan $Q = (y_1, y_2, \dots, y_p)$ dengan formula sebagai berikut.

$$d(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

2.2.3. Hierarchical Clustering Methods

Hierarchical clustering merupakan salah satu metode analisis dengan mengelompokkan objek-objek yang memiliki kesamaan secara bertingkat dan bertahap semacam membentuk pohon hierarki (Simanjuntak & Khaira, 2020). Metode ini merepresentasikan bagaimana mengelompokkan data menjadi "good" cluster dengan teknik yang efisien (Rencher, 2002). Dalam *hierarcichal methods* terdapat dua metode di dalamnya yakni aglomeratif dan divisif. Aglomeratif memperlakukan setiap observasi sebagai cluster tunggal yang kemudian secara berturut-turut

menggabungkan *cluster* hingga semua observasi menjadi satu *cluster* tunggal. Sebaliknya, divisif memperlakukan semua observasi dalam *cluster* tunggal dan kemudian membaginya bertahap hingga diperoleh jumlah *cluster* yang diinginkan (Rafsanjani et al., 2012).

Dalam metode aglomeratif, observasi mengalami proses penggabungan di mana observasi tersebut dari *cluster* berbeda akan masuk ke dalam kelompok *cluster* yang sama sehingga ukuran dalam *cluster* tersebut menjadi lebih besar. Secara lebih detail, aglomeratif dibagi lagi menjadi empat metode yakni *single linkage*, *complete linkage*, *average linkage*, dan *Ward's methods*. Ketiga metode tersebut didasarkan pada mekanisme pengukuran jarak antarobjek, sedangkan untuk *Ward's methods* merupakan metode yang sering digunakan pada data penelitian yang kontinu. Hasil dari proses hierarki adalah dendrogram yang secara visual merepresentasikan bagaimana *cluster* terbentuk.

2.2.4. Partitioning Clustering Methods

Partitioning clustering merupakan metode analisis untuk memngelompokkan sejumlah k observasi di mana nilai k dapat ditentukan sebelum atau pada saat prosedur *clustering* dilakukan (Johnson & Wichern, 2007). Setiap objek dalam observasi harus menjadi milik tepat suatu cluster yang ada (Swarndep & Pandya, 2016). Metode ini baik digunakan untuk data yang berukuran relatif besar. Selain itu, pengelompokkan tidak dilakukan berdasarkan matriks jarak dan kemiripan dari semua pasangan observasi yang ada sehingga lebih efisien (Rencher, 2002).

Salah satu metode yang termasuk dalam *partitioning* ini adalah K-means. Proses kerja metode ini adalah mempartisi data dalam satu atau lebih *cluster* sehingga dalam *cluster* didapatkan karakteristik sehomogen mungkin dan antar *cluster* didapatkan yang seheterogen mungkin. Algoritma pengelompokkan dengan K-means adalah sebagai berikut (Jain et al., 1999).

- 1) Tentukan jumlah *cluster*
- 2) Alokasikan data ke dalam kelompok secara acak
- 3) Hitung titik pusat (*centroid*) dari masing-masing *cluster*. Lokasi *centroid* ini diambil dari setiap *cluster* untuk rata-rata nilai dari setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah *cluster*, i menyatakan fitur ke- i dalam sebuah *cluster*, dan p menyatakan dimensi data, maka *centroid* fitur ke- i dihitung dengan formula:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (2)$$

- 4) Alokasikan masing-masing data ke *centroid* terdekat.
- 5) Ketika masih ada data yang belum konvergen atau berpindah kelompok dilihat dari perubahan nilai *centroid*-nya, maka kembali ke langkah-3.

Kekonvergenan dalam suatu proses iterasi dalam metode K-means dapat ditinjau dari nilai *squared error* yang telah mencapai minimumnya (Jain et al., 1999). Selain itu, tidak ada lagi perubahan anggota di setiap *cluster*-nya secara ekuivalen, ditunjukkan dengan tidak adanya perubahan titik pusat *cluster* (Yedla et al., 2010).

2.2.5. Validasi Cluster

Indeks validitas digunakan untuk mengevaluasi hasil *clustering* dengan tujuan mendapatkan jumlah *cluster* terbaiknya (Khairati et al., 2019). Indeks validitas ini dapat dihitung berdasarkan dua hal sebagai berikut.

- 1) *Compactness* : similaritas objek observasi pada *cluster* yang sama
- 2) *Separation* : perbedaan objek observasi pada *cluster* yang berbeda

Sedangkan, menurut Halkidi et al., (2001), indeks validitas ini dapat mencakup tiga kriteria yang terdiri dari kriteria eksternal, kriteria internal, dan kriteria relatif. Validasi *cluster* ini juga dapat dikategorikan pada tiga jenis validasi yang dilakukan dalam penelitian (Afira & Wijayanto, 2021) yakni validasi internal, validasi stabilitas, dan validasi biologis.

Dalam penelitian kali ini, akan digunakan dua jenis validasi *cluster* yakni validitas internal dan validitas stabilitas. Dalam validitas internal, terdapat tiga indeks yang terdiri dari *connectivity*, *silhouette*, dan *dunn*. Sedangkan untuk validitas stabilitas terdiri dari empat nilai yakni APN, AD, ADM, dan FOM.

2.3. Prosedur Penelitian

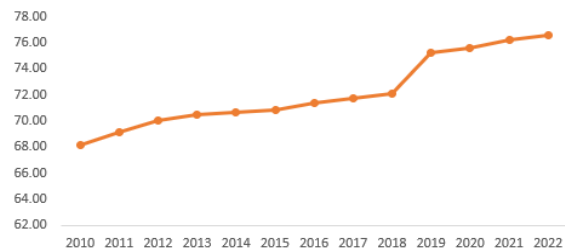
Penelitian dilakukan dengan bantuan *software* R Studio untuk melakukan rangkaian analisis data. Untuk melihat gambaran kondisi pemberdayaan gender secara umum di Indonesia tahun 2022, digunakan analisis deskriptif sederhana dengan melihat kontribusi komponen penyusun IDG yang terbesar dan terkecil pada wilayah provinsi. Pengelompokan kondisi pemberdayaan gender yang dicerminkan dengan skor IDG dilakukan dengan *clustering methods* yang selanjutnya dibandingkan antara *hierarchical methods* dan *partitioning methods* untuk mengetahui model pengelompokan mana yang terbaik. Berikut merupakan langkah-langkah yang dilakukan dalam analisis data penelitian ini.

- 1) Melakukan standarisasi data.
- 2) Melakukan pengecekan asumsi pada analisis faktor
- 3) Melakukan proses *clustering* dengan *hierarchical methods* dan *partitioning methods*.
- 4) Melakukan validasi *cluster* untuk menentukan metode yang terbaik.
- 5) Membuat visualisasi dari distribusi kelompok IDG yang terbentuk berdasarkan komponen penyusunnya.

3. Hasil dan Pembahasan

Capaian Indeks Pemberdayaan Gender (IPG) Indonesia tahun 2022 mencapai 76,59. Angka ini terus meningkat bahkan sejak lebih dari 10 tahun yang lalu dengan nilai 68,15 pada tahun 2010 dan naik cukup signifikan pada tahun 2019 dengan nilai IDG 75,24. Hal ini menunjukkan bahwa peran perempuan untuk ambil bagian dalam politik dan ekonomi semakin baik.

Indeks Pemberdayaan Gender Indonesia 2010-2022



Gambar. 1 – IDG Indonesia 2010—2022.

Namun, apabila ditinjau lebih lanjut masih terdapat ketimpangan nilai IDG pada level provinsi. Terdapat lima provinsi dengan capaian IDG di atas nilai nasional, yakni Kalimantan Tengah, Sulawesi Utara, Maluku Utara, Sulawesi Tengah, dan DI Yogyakarta. Pandemi Covid-19 yang sempat merebak di Indonesia nyatanya tidak berdampak secara terhadap indikator IDG (KPPA & BPS, 2022). Salah satu penyebabnya adalah indikator pembentuk IDG yang tidak langsung berpengaruh terhadap pandemi Covid-19 yakni keterlibatan perempuan dalam parlemen. Indikator ini didapatkan dari pemilihan umum yang dilaksanakan secara periodik setiap lima tahun sekali sedangkan dalam rentang 2020—2023 tidak mengalami fase tersebut.

Sebelum dilakukan pengelompokan wilayah pada level provinsi, perlu dilakukan pengecekan asumsi bahwa data yang digunakan cukup secara statistik dan tidak terjadi multikolinearitas (Afira & Wijayanto, 2021). Apabila asumsi ini terlanggar maka perlu dilakukan analisis faktor terlebih dahulu.

a. Uji Kecukupan Sampel

Tabel 3 – Hasil Penghitungan Kecukupan Data.

Uji	Nilai
<i>Kaiser Meyer Olkin (KMO)</i>	0,5281

Hipotesis nol dalam uji ini adalah data dinyatakan cukup dan layak untuk dianalisis. Dari hasil pengujian yang dilakukan diketahui nilai KMO lebih dari 0,5 sehingga dengan tingkat signifikansi sebesar 5% dapat dinyatakan bahwa hipotesis nol gagal ditolak yang berarti data yang digunakan dalam penelitian telah cukup dan layak untuk dianalisis lebih lanjut.

b. Uji Asumsi Nonmultikolinearitas

Tabel 4 – Pengujian Asumsi Nonmultikolinearitas.

Uji	<i>p-value</i>
<i>Bartlett Test of Sphericity</i>	0,8454

Hipotesis nol dalam uji ini adalah tidak terdapat masalah multikolinieritas. Dari hasil pengujian yang dilakukan diketahui nilai p -value dari uji *Bartlett* sebesar 0,8454 sehingga dengan tingkat signifikansi sebesar 5% dapat dinyatakan bahwa hipotesis nol gagal ditolak yang berarti tidak perlu dilakukan analisis faktor dan pengelompokkan dengan *clustering methods* dapat langsung dilakukan.

Selanjutnya, dilakukan uji validitas cluster untuk mengetahui jumlah kelompok yang optimum untuk pengelompokkan. Dalam penelitian ini, dilakukan dua metode validitas yakni internal dan stabilitas.

Tabel 5 – Nilai Indeks Validitas Internal.

Indeks	2	3	4	5
Metode : Hierarki				
<i>Connectivity</i>	9,4317	12,1218	19,2905	23,2425
<i>Dunn</i>	0,2295	0,2995	0,2201	0,3032
<i>Silhouette</i>	0,3512	0,2085	0,2675	0,3426
Metode : K-means				
<i>Connectivity</i>	10,5333	10,5595	24,9464	24,9599
<i>Dunn</i>	0,2295	0,1979	0,1706	0,2722
<i>Silhouette</i>	0,3558	0,27	0,3019	0,3467

Tabel 6 – Nilai Optimal Indeks Validitas Internal.

Indeks	Nilai	Metode	Jumlah Cluster
<i>Connectivity</i>	9,4317	Hierarki	2
<i>Dunn</i>	0,3032	Hierarki	5
<i>Silhouette</i>	0,3558	K-means	2

Tabel 7 – Nilai Indeks Validitas Stabilitas.

Indeks	2	3	4	5
Metode : Hierarki				
APN	0,0527	0,1915	0,255	0,1702
AD	2,2812	2,1256	1,992	1,7127
ADM	0,5515	0,6977	0,9483	0,792
FOM	0,944	0,9292	0,9173	0,9021
Metode : K-means				
APN	0,1187	0,3189	0,315	0,2017
AD	2,1166	2,0896	1,8318	1,6172
ADM	0,3537	0,921	0,8761	0,7154
FOM	0,9051	0,9087	0,8762	0,8623

Tabel 8 – Nilai Optimal Indeks Validitas Internal.

Indeks	Nilai	Metode	Jumlah Cluster
APN	0,0527	Hierarki	2
AD	1,6172	K-means	5
ADM	0,3537	K-means	2
FOM	0,8623	K-means	5

Komponen dalam uji validitas internal terdiri dari indeks *connectivity*, *silhouette*, dan *dunn*. Pada Tabel 5 menunjukkan bahwa dengan indeks *connectivity* metode

Hierarki dengan jumlah *cluster* 2 memiliki nilai terkecil sehingga paling optimal. Kemudian, indeks *silhouette* dan *dunn* akan bernilai optimal apabila mendekati 1 sehingga metode K-means dengan jumlah *cluster* secara berurutan 2 dan 3 menjadi nilai yang optimal.

Kemudian dilakukan uji validitas stabilitas yang terdiri dari nilai APN, AD, ADM, dan FOM. Semakin kecil nilai APN, AD, ADM, dan FOM yang diperoleh maka akan semakin optimal jumlah *cluster* yang diperoleh. Tabel 7 menunjukkan bahwa nilai terkecil untuk masing-masing komponen nilai dalam uji validitas stabilitas menghasilkan metode dan jumlah *cluster* yang berbeda.

Hasil dari uji validitas menghasilkan metode dan jumlah *cluster* yang cukup bervariasi. Namun, metode Hierarki dengan jumlah *cluster*, metode K-means dengan jumlah *cluster* 2, dan metode K-means dengan jumlah *cluster* 5 memiliki jumlah nilai optimal yang sama banyak dari indeks-indeks validitas yang digunakan. Sehingga, untuk pengelompokkan selanjutnya akan digunakan ketiga metode tersebut dan akan dilakukan evaluasi lebih lanjut di akhir untuk memilih model mana yang terbaik.

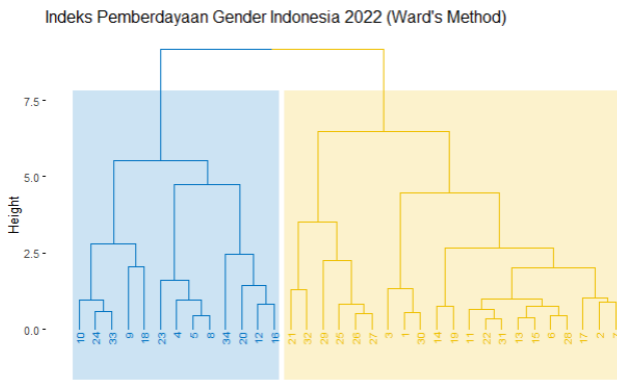
a. Hierarchical Methods

Dalam penelitian ini digunakan *hierarchical methods* dengan aglomeratif. Pengelompokkan dilakukan dengan menggunakan empat metode aglomeratif yang terdiri dari *single linkage*, *complete linkage*, *average linkage*, dan *Ward's Method*. Selanjutnya akan dipilih satu metode terbaik dengan melihat nilai *agglomerative coefficient* terbaiknya.

Tabel 9 – Agglomerative Coefficient Metode Hierarki.

Metode	Nilai
<i>Single Linkage</i>	0,5574
<i>Complete Linkage</i>	0,8569
<i>Average Linkage</i>	0,7337
<i>Ward's Methods</i>	0,8977

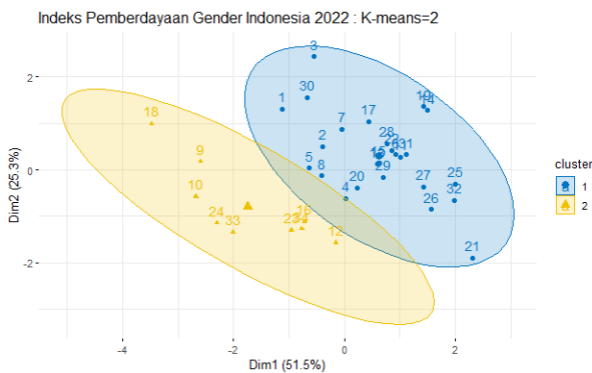
Nilai *agglomerative coefficient* merupakan ukuran struktur pengelompokkan dan struktur pengelompokkan yang paling kuat akan terpilih (Kara, 2021). Struktur pengelompokkan aglomeratif ini akan semakin kuat apabila nilai koefisiennya mendekati satu. Sehingga, berdasarkan Tabel 9 didapatkan bahwa *agglomerative hierarchy* yang menghasilkan *cluster* terbaik adalah *Ward's Methods*. Selanjutnya, pengelompokkan dengan *Hierarchical Methods* menghasilkan dendogram pengelompokkan seperti di bawah ini.



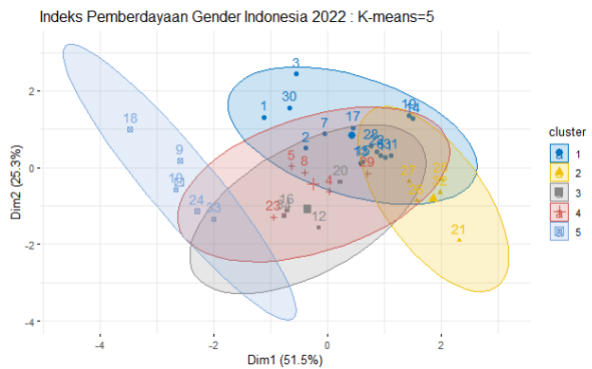
Gambar. 2 – Dendrogram Hasil Hierarki.

b. Partitioning Methods

Selanjutnya, pengelompokan dengan *Partitioning Methods* digunakan K-means yang merupakan bagian dari metode *k-clustering*. Jumlah *cluster* yang ditetapkan mengikuti hasil dari uji validitas yang dilakukan sebelumnya. Sehingga pengelompokan dengan *Partitioning Methods* menghasilkan *cluster plot* seperti di bawah ini.



Gambar. 3 – Cluster Plot K-means dengan k=2.



Gambar. 4 – Cluster Plot K-means dengan k=5.

c. Evaluasi Model dan Metode Terbaik

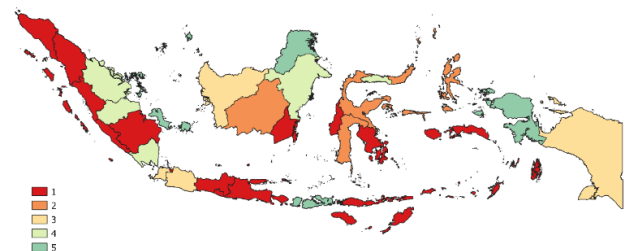
Untuk melakukan evaluasi model, digunakan nilai *average within* dan *average between* sebagaimana penelitian yang dilakukan oleh Luthfi & Wijayanto (2021). Untuk mendapatkan jarak dengan kualitas terbaik dilakukan dengan memperhatikan nilai rasio rata-rata *average within* dan *average between* yang minimum.

Tabel 10 – Nilai Rasio Simpangan Baku.

Metode	Cluster	Sw	Sb	Sw/Sb
<i>Ward's Methods</i>	2	0,1830	0,6418	0,2851
<i>K-means</i>	2	2,5383	45,6974	0,0555
<i>K-means*</i>	5	1,1735	23,0252	0,0510

Dari hasil perhitungan didapatkan bahwa nilai rasio antara *Ward's Methods* dengan K-means cukup berbeda jauh. Dengan demikian didapatkan bahwa K-means dengan jumlah cluster sebanyak 5 sebagai metode terbaik yang dipilih dengan nilai rasio simpangan terkecil. K-means memiliki keunggulan salah satunya dalam hal interpretabilitas karena adanya *centroid* dalam setiap kelompok yang lebih mudah dipahami dibandingkan dendrogram pada metode hierarkis. Selain itu, K-means baik untuk mengidentifikasi *cluster* yang berbentuk *convex-shape* karena menggunakan jarak *Euclidean* dalam penghitungan matriksnya.

Pengelompokan Provinsi di Indonesia Berdasarkan IDG 2022 dengan K-Means Clustering



Gambar. 5 – Pengelompokan dengan Metode Terbaik.

Dari pengelompokan yang diperoleh dari *Partitioning Methods* K-means di atas perlu ditinjau karakteristik yang ada dari setiap *cluster*. Kemudian didapatkan pola untuk setiap *cluster*-nya sebagai berikut.

Tabel 11 – Rata-rata Tiap Cluster.

Cluster	X_1	X_2	X_3
1	17,518	52,171	37,203
2	29,544	50,270	33,128
3	18,475	41,333	33,470
4	20,678	53,282	27,736
5	9,018	45,618	28,750

1. Kelompok wilayah 1 terdiri dari 15 provinsi yang terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, Bengkulu, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Timur, Kalimantan Selatan, Sulawesi Tenggara, Sulawesi Barat, dan Maluku. Kelompok wilayah ini memiliki karakteristik rata-rata perempuan sebagai tenaga profesional dan sumbangan pendapatan perempuan yang cukup tinggi, tetapi tingkat keterlibatan perempuan dalam parlemen yang rendah.
2. Kelompok wilayah 2 terdiri dari 5 provinsi yang terdiri dari Kalimantan Tengah, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, dan Maluku Utara. Kelompok wilayah ini memiliki karakteristik rata-rata tingkat keterlibatan perempuan dan perempuan sebagai tenaga profesional yang cukup tinggi. Namun, dari sisi sumbangan pendapatan perempuan termasuk dalam kondisi sedang.
3. Selanjutnya, terdapat 4 provinsi dalam kelompok wilayah 3 yang terdiri dari Jawa Barat, Banten, Kalimantan Barat, dan Papua. Kelompok wilayah ini memiliki karakteristik rata-rata tingkat keterlibatan perempuan dalam parlemen dan sumbangan pendapatan perempuan yang tergolong sedang, tetapi dari sisi tenaga profesional perempuannya masih rendah.
4. Terdapat 5 provinsi dalam kelompok wilayah 4 yang terdiri dari Riau, Jambi, Lampung, Kalimantan Timur, dan Gorontalo. Kelompok wilayah ini memiliki karakteristik rata-rata tenaga profesional perempuan yang tinggi dibandingkan kelompok lain dan sisi keterlibatan perempuan dalam parlemen yang sedang. Namun, dari sumbangan pendapatan perempuannya tergolong rendah.
5. Dan kelompok wilayah terakhir terdiri dari 5 wilayah yang terdiri dari Kepulauan Bangka Belitung, Kepulauan Riau, Nusa Tenggara Barat, Kalimantan Utara, dan Papua Barat. Kelompok wilayah ini memiliki karakteristik rata-rata keterlibatan perempuan dalam parlemen yang paling rendah dibanding kelompok wilayah lain, sumbangan pendapatan yang rendah, dan tenaga profesional perempuan yang tergolong sedang.

4. Kesimpulan dan Saran

1. Dengan menggunakan indikator pembentuk IDG yang terdiri dari keterlibatan perempuan dalam parlemen (X_1), perempuan sebagai tenaga profesional (X_2), dan sumbangan pendapatan perempuan (X_3), wilayah provinsi di Indonesia tahun 2022 dapat diklasifikasikan menjadi lima kelompok wilayah berdasarkan *Partitioning Methods* K-means. Klasifikasi ini merupakan metode terbaik yang dipilih berdasarkan ukuran nilai dalam validitas internal dan validitas stabilitas serta dilanjutkan dengan rasio simpangan baku yang minimum. Hasil yang diperoleh juga menunjukkan karakteristik komponen pembentuk IDG yang cukup bervariasi antar *cluster*-nya.
2. Karakteristik di setiap kelompok wilayah tersebut dapat dijadikan pertimbangan bagi pemerintah khususnya Kementerian Pemberdayaan Perempuan dan

Perlindungan Anak (KPPPA) RI dalam merumuskan kebijakan untuk mengatasi ketimpangan pembangunan gender yang masih terjadi antarprovinsi di Indonesia dalam rangka peningkatan pengarusutamaan gender.

3. Penelitian selanjutnya dapat melakukan perluasan cakupan ke level penelitian yang lebih rendah (kabupaten/ kota) dengan menggunakan variabel lain yang sekiranya memiliki pengaruh kuat dalam hal pembangunan gender.

Ucapan terima kasih

Terima kasih penulis sampaikan kepada Bapak Dr. Eng. Arie Wahyu Wijayanto, S.S.T., M.T., selaku dosen pengampu mata kuliah *Data Mining* dan *Big Data* yang telah membimbing dan memberikan masukan dalam penulisan penelitian ini. Terima kasih pula kepada seluruh pihak yang telah mendukung penulis untuk menyelesaikan segala sesuatu yang telah dimulai hingga tuntas.

DAFTAR PUSTAKA

- Afira, N., & Wijayanto, A. W. (2021). Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019. *Komputika : Jurnal Sistem Komputer*, 10(2), 101–109. <https://doi.org/10.34010/komputika.v10i2.4317>
- Astuti, C. C., & Rezanita, V. (2022). Cluster Analysis for Grouping Districts in Sidoarjo Regency Based on Education Indicators. *KnE Social Sciences*, 2022(47), 311–317. <https://doi.org/10.18502/kss.v7i10.11233>
- BPS. (2023). Indeks Pembangunan Gender.
- Cornwall, A. (2016). Women's Empowerment: What Works? *Journal of International Development*, 28(3), 342–359. <https://doi.org/https://doi.org/10.1002/jid.3210>
- Diana, R. (2018). Analisis Ketimpangan Gender Di Provinsi Sumatera Barat (Gender Inequality Analysis in West Sumatera Province). 13(1), 55–66.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Intelligent Information Systems*, 107–145. <https://doi.org/10.1023/A>
- Hasanah, I. N. (2022). Analisis Cluster Berdasarkan Dampak Ekonomi di Indonesia Akibat Pandemi Covid-19. 10(02).
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. 31, 264–323.
- Johnson, R. A., & Wincham, Dean W. (2007). Applied Multivariate Statistical Analysis. In *Applied Multivariate Statistical Analysis: Second Edition*. <https://doi.org/10.1007/978-3-540-72244-1>
- Judiasih, S. D. (2022). Implementasi Kesetaraan Gender Dalam Beberapa Aspek Kehidupan Bermasyarakat Di

- Indonesia. *ACTA DIURNAL Jurnal Ilmu Hukum Kenotariatan*, 5, 284–302.
- Kara, F. (2021). Comparison of tree diameter distributions in managed and unmanaged Kazdağı fir forests.
- 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.
- KPPA, & BPS. (2022). *Pembangunan Manusia Berbasis Gender 2022*.
- Luthfi, E., & Wijayanto, A. W. (2021). Analisis Perbandingan Metode Hierarchy, K-means, dan K-medoids Clustering dalam Pengelompokan Indeks Pembangunan Manusia. *Inovasi*, 17(4), 770–782.
- Nisa, A. H. (2018). Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Kondisi Disparitas Pembangunan Manusia Berbasis Gender. 97.
- Nurfadilah, K. (2021). Analisis Cluster Longitudinal pada Pembangunan Manusia di Sulawesi Selatan Berbasis Gender. 9(1).
- Porter, E. (2013). Rethinking Women's Empowerment. *Journal of Peacebuilding & Development*, 8(1), 1–14. <https://doi.org/https://doi.org/10.1080/15423166.2013.785657>
- Rafsanjani, M. K., Varzaneh, Z. A., Chukanlo, N. E., Rafsanjani, M. K., Varzaneh, Z. A., & Chukanlo, N. E. (2012). A survey of hierarchical clustering algorithms. 3(3), 229–240.
- Rani, Y., & Rohil, H. (2013). A study of hierarchical clustering algorithms. *International Journal of Information and Computation Technology*, 3(11), 1225–1232.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. In *Methods of Multivariate Analysis*. <https://doi.org/10.1002/0471271357>
- Salsabila, D., & Hendrawan, M. Y. (2021). Analisis Kondisi Pemberdayaan Gender di Indonesia Tahun 2020 dengan Agglomerative Hierarchical Clustering dan Biplot. 204–213.
- Sukim, Firdaus, Retnaningsih, & Utami, E. D. (2018). Mengukur Kepemimpinan Perempuan di Indonesia dengan Metode Fuzzy c-Means Clustering. 18(2), 101–112.
- Swarndep, S. J., & Pandya, S. (2016). An Overview of Partitioning Algorithms in Clustering Techniques. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 5(6), 1943–1946. <http://itd.idaho.gov/ohs/2011Data/Analysis2011.pdf>
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode K-means Untuk Pengelompokan Kabupaten/ Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. 11, 119–128.
- Yedla, M., Pathakota, S. R., & M, S. T. (2010). Enhancing K-means Clustering Algorithm with Improved Initial Center. *International Journal of Computer Science and Information Technologies*, 1(2), 121–125.