Komputika: Jurnal Sistem Komputer
Volume 12, Nomor 2, Oktober 2023, hlm. 211-219
Terakreditasi Peringkat 3, SK No. 148/M/KPT/2020
DOI: 10.34010/komputika.v12i2.10348

ISSN : 2252 – 9039 (Print)
ISSN : 2655 – 3198 (Online)

Analisis Cluster Kualitas Pemuda di Indonesia pada Tahun 2022 dengan *Agglomerative Hierarchical* dan *K-Means*

Jeremia Novaldi[1*], Arie Wahyu Wijayanto

[1)] Program Studi D4 Statistika, Politeknik Statistika STIS
Jl. Otto Iskandardinata No.64C Jakarta 13330
**email: 212011539@stis.ac.id*

(Naskah masuk: 8 Juli 2023; diterima untuk diterbitkan: 12 September 2023)

ABSTRAK *– Pemuda adalah generasi yang akan memegang masa depan Indonesia. Menurut BPS, seperempat penduduk Indonesia merupakan pemuda . Dengan demikian, pemerintah memerlukan gambaran mengenai kualitas pemuda saat ini untuk merumuskan kebijakan yang tepat untuk tiap daerah. Penelitian ini bertujuan untuk mengelompokan provinsi-provinsi di Indonesia menurut data kepemudaan dengan menggunakan metode hierarki aglomeratif dan K-Means. Berdasarkan nilai indeks validitas internal dan stabilitas, hierarki aglomeratif (Ward's method) dengan jumlah cluster 2 dipilih sebagai metode pengelompokan terbaik. Metode ini menghasilkan 2 cluster yang masing-masing terdiri dari 11 dan 23 provinsi. Secara umum, Cluster 1 berisi provinsi-provinsi dengan kualitas pemuda yang lebih baik, di mana nilai rata-rata RLS pemuda, persentase pemuda dengan akses internet, persentase pemuda dengan jaminan kesehatan yang lebih tinggi dari Cluster 2 meskipun memiliki TPT yang lebih tinggi. Sebaliknya, Cluster 2 memiliki nilai yang lebih tinggi pada indikator Angka Kesakitan Pemuda, persentase pemuda dengan usia kawin pertama 16 – 18 tahun, dan persentase pemudi yang melahirkan bayi dengan BBLR.*

Kata Kunci *– Clustering; Hierarki; K-Means; Pemuda; SDGs*

Clustering Analysis Using Agglomerative Hierarchical and K-Means on Youth Quality Data in Indonesia in 2022

ABSTRACT *– Youth is the generation that will hold the future of Indonesia. According to BPS, a quarter of Indonesia's population are youth. Thus, the government needs an overview of the current quality of youth to formulate appropriate policies for each region. This study aims to classify provinces in Indonesia based on youth data using agglomerative hierarchical and K-Means. According to the value of the internal validation and stability index, the agglomerative hierarchical, using Ward's method, with 2 clusters was chosen as the best clustering method. This method produces 2 clusters consisting of 11 and 23 provinces respectively. In general, Cluster 1 contains provinces with better youth quality, where the average youth schooling years, the percentage of youth with internet access, the percentage of youth with health insurance are higher than Cluster 2 despite having a higher unemployment rate. In contrast, Cluster 2 has a higher average score on the Youth Sickness Rate, the percentage of youth with first marriage age 16 – 18 years, and the percentage of young women who give birth to babies with LBW.*

Keywords *– Clustering; Hierarchical; K-Means; SDGs; Youth*

1. PENDAHULUAN

Pemuda merupakan generasi yang kelak akan mengelola suatu bangsa. Menurut Badan Pusat Statistik (BPS), pemuda merupakan penduduk yang berada di kelompok usia 16 – 30 tahun, atau penduduk yang lahir antara tahun 1992 – 2006 pada 2022. Pada tahun 2022, penduduk usia pemuda dapat digolongkan ke dalam generasi Z [1]. Dalam 10 tahun terakhir, persentase penduduk Indonesia berada di antara 23 – 24 persen, yaitu hampir seperempat dari jumlah penduduk. Provinsi dengan persentase pemuda tertinggi terletak di wilayah timur Indonesia,

yaitu Provinsi Gorontalo (26,91%), Papua (26,83%), dan Papua Barat (26,6%). Sebaliknya, persentase terendah terletak di wilayah barat Indonesia, yaitu Provinsi DI Yogyakarta (21,85%), Jawa Timur (22,21%), dan Bali (22,69%) [2].

Sejalan dengan agenda SDGs 2030, Indonesia melihat pemuda sebagai kunci percepatan pembangunan yang terdapat dalam UU No. 40 Tahun 2009 [3]. Krusialnya peran pemuda di masa yang akan datang membuat kualitas pemuda harus diperhatikan oleh pemerintah. Untuk itu, pemerintah perlu gambaran kualitas pemuda tiap daerah dengan mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakkteristik kualitas pemuda yang tersedia. Pengelompokan provinsi ini diharapkan dapat mempermudah pemerintah dalam membuat kebijakan yang dapat meningkatkan kualitas pemuda di Indonesia.

Analisis *cluster* merupakan pendekatan yang dilakukan dalam mengelompokan objek-objek berdasarkan karakteristik tertentu. Analisis *cluster* menempatkan objek-objek dengan karakteristik yang serupa ke dalam kelompok yang sama. Dengan demikian, analisis ini membagi sekumpulan objek ke dalam beberapa kelompok, di mana suatu kelompok memiliki sifat yang serupa di antara anggotanya tetapi memiliki sifat yang berbeda dengan kelompok lainnya [4].

Secara umum, metode dalam analisis *cluster* dapat dibagi ke dalam *hierarchical clustering, optimization clustering,* dan *model-based clustering* [5]. Metode hierarki melakukan pengelompokan secara bertahap dan terstruktur baik secara *agglomerative* maupun *divisive*. Metode hierarki terdiri dari *single linkage, complete linkage, average linkage,* dan *ward method*. Metode optimisasi sering juga disebut dengan metode partisi, di mana membagi data ke dalam k *cluster* dengan mencoba seluruh kemungkinan agar menemukan *cluster* yang optimum. Salah satu contoh metode partisi adalah k-means.

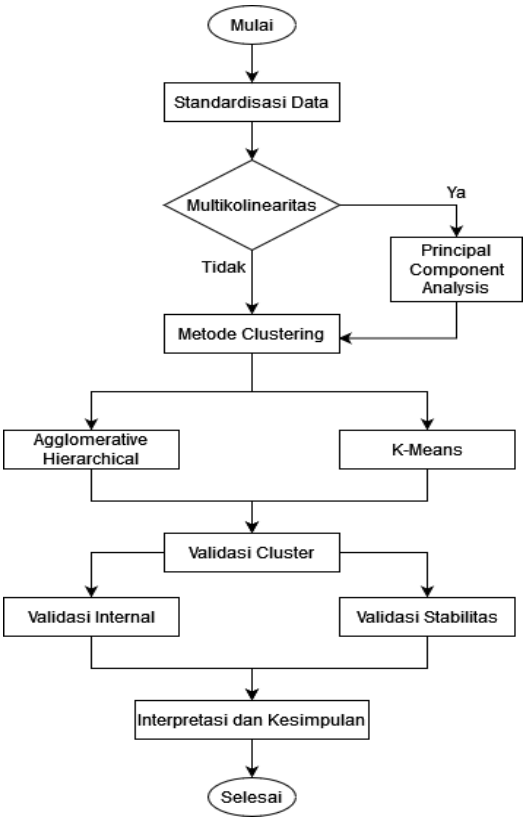
Beberapa penelitian sebelumnya yang telah dilakukan adalah pengelompokan provinsi berdasarkan data sosial ekonomi dengan k-means yang dilakukan oleh Ahmar tahun 2018 [6], pengelompokan provinsi berdasarkan ,pengelompokan provinsi berdasarkan data pelanggan air bersih dengan k-means yang dilakukan oleh Windarto tahun 2019 [7], kemiskinan dengan metode hierarki dan partisi yang dilakukan oleh Afira tahun 2021 [8], pengelompokan kabupaten/kota di Provinsi Maluku berdasarkan indikator pendidikan dengan metode ward yang dilakukan oleh Dewi tahun 2021 [9], dan pengelompokkan provinsi berdasarkan pelayanan kesehatan maternal dengan *agglomerative hierarchical* dan k-means yang dilakukan oleh Azzahra tahun 2022 [10].

Berdasarkan penelitian terdahulu, penulis ingin melakukan analisis *cluster* dengan metode hierarki secara aglomeratif dan partisi dengan k-means untuk mengelompokan 34 wilayah adminstrasi di Indonesia menggunakan dataset kualitas pemuda tahun 2022.

2. METODE DAN BAHAN

Data yang digunakan pada penelitian ini didapatkan dari publikasi Statistik Pemuda Indonesia 2022 yang dipublikasikan oleh Badan Pusat Statistik (BPS) [2]. Variabel-variabel yang digunakan untuk merepresentasikan beberapa aspek dari pemuda. Variabel rata-rata lama sekolah (RLS) pemuda dan persentase pemuda yang memiliki akses internet mewakili aspek pendidikan pemuda. Variabel angka kesakitan pemuda, persentase pemuda dengan jaminan kesehatan, persentase pemuda dengan usia kawin pertama 16-18 tahun, dan persentase pemuda Perempuan yang melahirkan bayi dengan berat badan lahir rendah (BBLR) menggambarkan keadaan kesehatan pemuda. Kondisi ekonomi pemuda digambarkan oleh variabel tingkat pengangguran terbuka (TPT) pemuda.

Gambar 1 menunjukkan tahapan-tahapan dalam penelitian ini. Tabel 1. Menyajikan variabel yang digunakan dalam penelitian ini. Penelitian ini menggunakan R-Studio untuk melakukan analisis *cluster* dan QGIS untuk menggambarkan persebaran *cluster* menggunakan peta.



Gambar 1. Diagram Alir Penelitian

Tabel 1. Daftar Variabel Penelitian

Variabel	Keterangan
X1	Rata-Rata Lama Sekolah (RLS) Pemuda
X2	Persentase Pemuda yang Memiliki Akses Internet
X3	Angka Kesakitan Pemuda
X4	Persentase Pemuda yang Memiliki Jaminan Kesehatan
X5	Tingkat Pengangguran Terbuka (TPT) Pemuda
X6	Persentase Pemuda dengan Usia Kawin Pertama 16 – 18 Tahun
X7	Persentase Pemuda Perempuan yang Melahirkan Bayi dengan Berat Badan Lahir Rendah (BBLR)

ANALISIS CLUSTER

Analisis *cluster* merupakan metode analisis yang bertujuan untuk membuat kelompok-kelompok dari sekumpulan objek yang bersifat homogen di dalam tiap kelompok dan bersifat heterogen antar kelompok**.** Metode dalam analisis *cluster* dibagi ke dalam metode hierarki dan non-hierarki [11].

Dalam menganalisis klaster terdapat dua asumsi dasar yang harus dipenuhi, yaitu kecukupan sampel dan tidak adanya multikolinearitas di antara variabel [9]. Asumsi kecukupan sampel dipenuhi apabila nilai *Kaiser-Meyer-Olkin* (KMO) lebih besar dari 0,5 [8].

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (1)$$

Keterangan :

r_{ij}^2 : korelasi antara variabel i dan j

a_{ij}^2 : korelasi parsial antara variabel i dan j

Pengecekan asumsi nonmultikolinearitas dilakukan dengan uji Bartlett dengan hipotesis awal, statistik uji, dan kriteria pengujian sebagai berikut :

Hipotesis:

$H_0 : R = 1$

$H_1 : R \neq 1$

Statistik Uji :

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

$|R|$: nilai determinan matriks korelasi

n : banyaknya pengamatan

p : banyaknya variable

Kriteria Pengujian :

Tolak H_0 jika *p-value* < α. Variabel-variabel saling berkorelasi atau terdapat multikolinearitas.

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) merupakan metode statistik yang digunakan untuk mengurangi dimensi data dengan menjaga informasi yang paling signifikan. PCA membantu mengidentifikasi pola dan hubungan yang tersembunyi dalam dataset yang kompleks dengan membentuk kombinasi linear dari peubah yang menyumbang variasi total dalam data sebanyak mungkin [12].

METODE HIERARKI

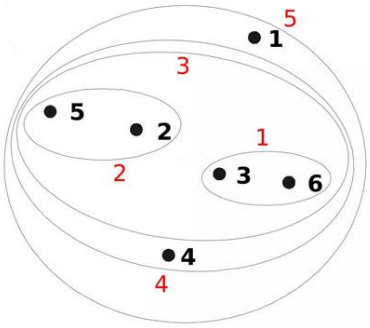
Metode hierarki merupakan metode analisis klaster di mana pengelompokan dilakukan secara teratur sehingga membentuk struktur pohon [13]. Metode ini bisa dilakukan baik secara *agglomerative (bottom up)*, yaitu menggabungkan tiap-tiap objek hingga mencapai satu kesatuan, maupun secara *divisive (top down)*, yaitu membagi objek-objek yang mulanya di dalam satu *cluster* yang selanjutnya dipartisi hingga tiap objek menjadi *cluster* tersendiri. Metode *clustering* ini Hasil dari metode ini dapat disajikan dalam *Dendogram* yang menunjukkan bagaimana *cluster* di tiap tahapannya beserta nilai koefisien jaraknya.

Nilai koefisien jarak ditentukan antara pasangan titik. Jarak antar titik dapat diukur menggunakan berbagai pendekatan seperti *Minkowski, Euclidean,* dan *Manhattan* [8].

Terdapat 4 teknik dalam melakukan analisis cluster dengan metode *agglomerative hierarchical* [14], yaitu :

1. *Single Linkage*

Jarak antar klaster ditentukan berdasarkan jarak terdekat antara objek pada suatu klaster dengan objek di klaster lainnya. Gambar 2 menunjukkan ilustrasi dari teknik *single linkage*.



Gambar 2. *Nested Clusters Single Linkage* [15]

$$d(uv)w = \min(d_{uv}d_{vw}) \quad (3)$$

Keterangan :

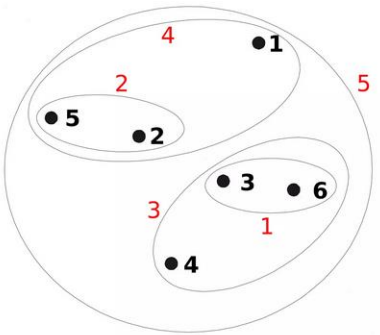
d(uv)w : jarak antara *cluster* (UV) dan *cluster* W

d_{uw} dan d_{vw} : Jarak antar objek

Penjelasan mengenai ilustrasi di atas dijelaskan sebagai berikut. Pada mulanya, Gambar 2 terdiri dari 6 *cluster* dengan masing-masing memiliki 1 objek di dalamnya. Kemudian dihitung jarak antar *cluster* menggunakan *single linkage*. Hasilnya, *cluster* 3 dan 6 digabungkan menjadi satu cluster baru karena memiliki jarak terdekat. Kemudian, jarak antara tiap cluster (36, 1, 2, 4, 5) dihitung kembali. *Cluster* 2 dan 5 memiliki jarak terdekat sehingga digabungkan menjadi *cluster* baru. Selanjutnya, jarak antara *cluster* 36, 25, 1, 4 dihitung kembali. Jarak terdekat dimiliki oleh cluster 36 dan 25 sehingga kedua *cluster* digabungkan. Tahapan ini terus berulang sampai hanya terdapat satu *cluster* saja, yaitu *cluster* 123456.

2. *Complete Linkage*

Jarak antar klaster ditentukan berdasarkan jarak terjauh antara objek pada suatu klaster dengan objek di klaster lainnya. Gambar 3 menunjukkan ilustrasi dari teknik *complete linkage*.

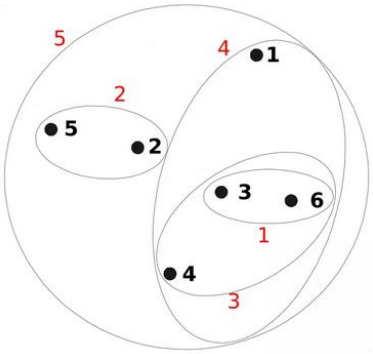


Gambar 3. *Nested Clusters Complete Linkage* [15]

$$d(uv)w = \max(d_{uv}d_{vw}) \quad (4)$$

3. *Average Linkage*

Jarak antar klaster ditentukan berdasarkan rata-rata jarak antara objek pada suatu klaster dengan klaster lainnya. Gambar 4 menunjukkan ilustrasi dari teknik *average linkage*.

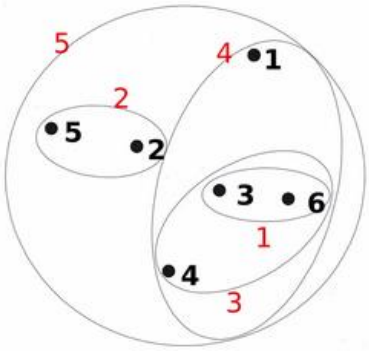


Gambar 4. *Nested Clusters Average Linkage* [15]

$$d(uv)w = \text{avg}(d_{uv}d_{vw}) \quad (5)$$

4. *Ward's Method*

Teknik ini menggabungkan *cluster* apabila total ketidaksamaan kuadrat dengan pusat cluster minimum di semua kemungkinan pilihan penggabungan [5]. Teknik ini meminimumkan jumlah kuadrat (ESS) [14]. Gambar 5 menunjukkan ilustrasi dari *ward's method*.



Gambar 5. *Nested Clusters Ward's Method* [15]

$$ESS = \sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_{ij}\right) \quad (6)$$

METODE PARTISI

Metode partisi menghasilkan suatu partisi objek, berbeda dengan struktur pengelompokan yang diperoleh dari metode hierarki [8]. Metode ini mempunyai kelebihan karena mempunyai waktu komputasi yang lebih efisien. Secara umum, metode ini dapat dibagi menjadi *k-clustering* dan *self-determining* [16]. K-Means merupakan metode *k-clustering* yang paling populer digunakan. Metode ini mengelompokan sekumpulan objek ke dalam *k-cluster* dengan prosedur yang sederhana.

Langkah-langkah dalam melakukan *clustering*

menggunakan K-Means [17] adalah sebagai berikut:

1. Menentukan jumlah klaster yang ingin dibentuk, yaitu sebanyak k.
2. Menentukan nilai *centroid* atau pusat klaster sebanyak k.
3. Menghitung jarak tiap objek terhadap tiap-tiap *centroid* menggunakan jarak *Euclidean* hingga ditemukan *centroid* terdekat dari setiap objek.
4. Mengelompokan setiap objek berdasarkan kedekatannya dengan *centroid* yang ada.
5. Mempebaharui nilai *centroid* dengan nilai rata-rata dari objek di dalam *cluster* yang terbentuk.
6. Melakukan iterasi langkah 3 sampai 5 hingga *cluster* yang terbentuk bersifat konvergen atau tidak berubah.

VALIDASI CLUSTER

Secara umum validasi *cluster* dapat dibagi menjadi validasi internal dan validasi eksternal [18]. Validasi internal mempunyai 2 kriteria, yaitu *compactness* yang mengukur seberapa mirip objek-objek yang ada di dalam *cluster* dan *separation* yang mengukur seberapa beda objek-objek antar *cluster*. Penelitian ini menggunakan indeks *connectivity, Dunn, Silhouette* dan *Davies-Bouldin* untuk memvalidasi *cluster* yang terbentuk.

Indeks *Connectivity* memiliki nilai di antara 0 sampai tak hingga di mana semakin kecil nilainya semakin bagus *cluster* yang terbentuk [8]. Indeks Dunn menghitung nilai minimum dari perbandingan antara nilai ketidaksamaan dua *cluster* dan nilai maksimum diameter *cluster* di mana semakin besar nilainya, semakin baik *cluster* yang terbentuk [19]. Indeks *Silhouette* menghitung selisih antara jarak rata-rata objek di dalam *cluster* dengan jarak minimum antar *cluster* di mana semakin mendekati positif 1 semakin baik [20].

Selain validasi Internal, penelitian ini juga menggunakan validasi stabilitas yang membandingkan hasil analisis klaster berdasarkan peniadaan variabel pada data, satu per satu. Nilai validasi stabilitas yang digunakan adalah *Average proporsion of non-overlap* (APN), *average distance* (AD), *average distance between means* (ADM), dan *figure of merit* (FOM) di mana semakin kecil nilai-nilai tersebut, semakin baik *cluster* yag dibuat [8].

3. HASIL DAN PEMBAHASAN

STATISTIK DESKRIPTIF

Tabel 2 menunjukkan bagaimana gambaran tiap variabel yang digunakan dalam menganalisis kualitas pemuda di Indonesia pada Tahun 2022. Standardisasi data diperlukan untuk menyamakan satuan pada semua variabel penelitian.

Tabel 2. Statistik Deskriptif

Variabel	Minimum	Mean	Maximum
X1	8,21	10,98	12,59
X2	32,58	89,39	98,65
X3	2,02	9,13	17,56
X4	58,65	74,80	98,80
X5	5,46	11,83	20,16
X6	10,55	19,16	28,83
X7	6,37	13,78	20,49

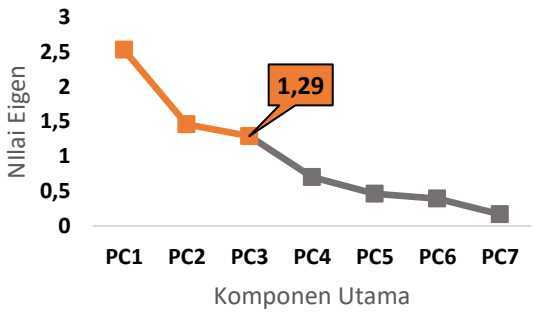
PRINCIPAL COMPONENT ANALYSIS

Berdasarkan informasi yang terdapat pada tabel 3, terdapat pelanggaran asumsi dasar untuk melakukan analisis *cluster*, yaitu adanya multikolinearitas. Pelanggaran ini ditunjukkan oleh Uji Bartlett yang menghasilkan keputusan tolak H0. Dengan demikian perlu dilakukan PCA terlebih dahulu. Sementara itu, nilai KMO yang melebihi 0,5 mengindikasikan bahwa data sudah memenuhi asumsi kecukupan sampel sehingga bisa dianalisis lebih lanjut.

Tabel 3. Uji KMO dan Bartlett

Uji		Nilai
Kaiser-Meyer-Olkin (KMO)		0,501
Bartlett	Approx. Chi-Square	68,326
	df	21
	Sig.	0,000

Menurut gambar 6, terdapat 3 komponen utama yang memiliki nilai eigen lebih dari 1 sehingga jumlah komponen utama yang optimal adalah 3. Ketiga komponen utama ini mampu menjelaskan 75,38 persen variasi dari 7 variabel sebelumnya. Tabel 4 menunjukkan *loading factor* dari ketiga komponen utama yang merupakan hasil reduksi menggunakan PCA.



Gambar 6. Diagram Nilai Eigen

Tabel 4. *Loading Factor*

Variabel	PC1	PC2	PC3
X1	-0,521	0,080	-0,071

X2	-0,474	0,402	0,211
X3	-0,015	0,750	-0,040
X4	0,008	0,067	-0,818
X5	-0,356	-0,470	0,201
X6	0,427	0,206	0,469
X7	0,440	-0,028	-0,136

VALIDASI CLUSTER

Validasi *cluster* digunakan untuk menentukan jumlah *cluster* yang paling optimal. Selain itu, validasi *cluster* juga berguna untuk mengetahui apakah metode *agglomerative hierarchy* atau K-Means yang paling sesuai dengan data pemuda di Indonesia. Tabel 5 menunjukkan beberapa nilai indeks yang digunakan dalam melakukan validasi internal. Hasil pengelompokan dapat disebut lebih baik apabila memiliki nilai Indeks *Connectivity* yang minimum serta nilai Indeks *Dunn* dan *Silhouette* yang maksimum. Tabel 6 memberikan informasi yang lebih ringkas dalam menentukan metode apa dan berapa jumlah *cluster* yang paling optimal berdasarkan tiap indeks validasi internal.

Tabel 5. Nilai Indeks Validasi Internal

Metod	Indeks	Jumlah Klaster				
		2	3	4	5	6
Hierarki	Connectivity	2,929	6,001	18,123	24,8861	26,001
	Dunn	0,763	0,393	0,182	0,229	0,228
	Silhouette	0,602	0,373	0,202	0,277	0,259
K-Means	Connectivity	2,929	16,4774	23,854	26,494	37,213
	Dunn	0,763	0,160	0,213	0,213	0,163
	Silhouette	0,602	0,290	0,293	0,302	0,229

Tabel 6. Nilai Indeks Validasi Internal Optimal

Indeks	Nilai	Metode	∑ Klaster
Connectivity	2,929	Hierarki & K-Means	2
Dunn	0,763	Hierarki & K-Means	2
Silhouette	0,602	Hierarki & K-Means	2

Tabel 7. Nilai Indeks Validasi Stabilitas

Metode	Indeks	Jumlah Klaster				
		2	3	4	5	6
Hierarki	APN	0,019	0,163	0,314	0,184	0,305
	AD	2,515	2,429	2,296	2,056	1,940
	ADM	0,117	0,551	1,144	0,985	1,044
	FOM	1,263	1,280	1,278	1,279	1,294
K-Means	APN	0,258	0,278	0,349	0,313	0,395
	AD	2,648	2,282	2,094	1,892	1,843
	ADM	0,712	0,808	1,075	0,937	1,120
	FOM	1,304	1,263	1,286	1,262	1,276

Tabel 8. Nilai Indeks Validasi Stabilitas Optimal

Indeks	Nilai	Metode	∑ Klaster
APN	0,019	Hierarki	2
AD	1,843	K-Means	6
ADM	0,117	Hierarki	2
FOM	1,263	K-Means	3

Berdasarkan tabel 6, indeks validasi internal menunjukkan bahwa baik metode hierarki maupun K-Means dengan jumlah klaster 2 merupakan metode pengelompokan yang paling optimal.

Nilai indeks validasi stabilitas dengan APN, AD, ADM, dan FOM disajikan dalam tabel 7. Menurut tabel 8, dapat disimpulkan bahwa metode hierarki dengan jumlah klaster berdasarkan nilai APN dan ADM. Sementara itu, nilai AD dan FOM menobatkan K-Means sebagai metode terbaik dengan jumlah *cluster* berturut-turut sebesar 6 dan 3.

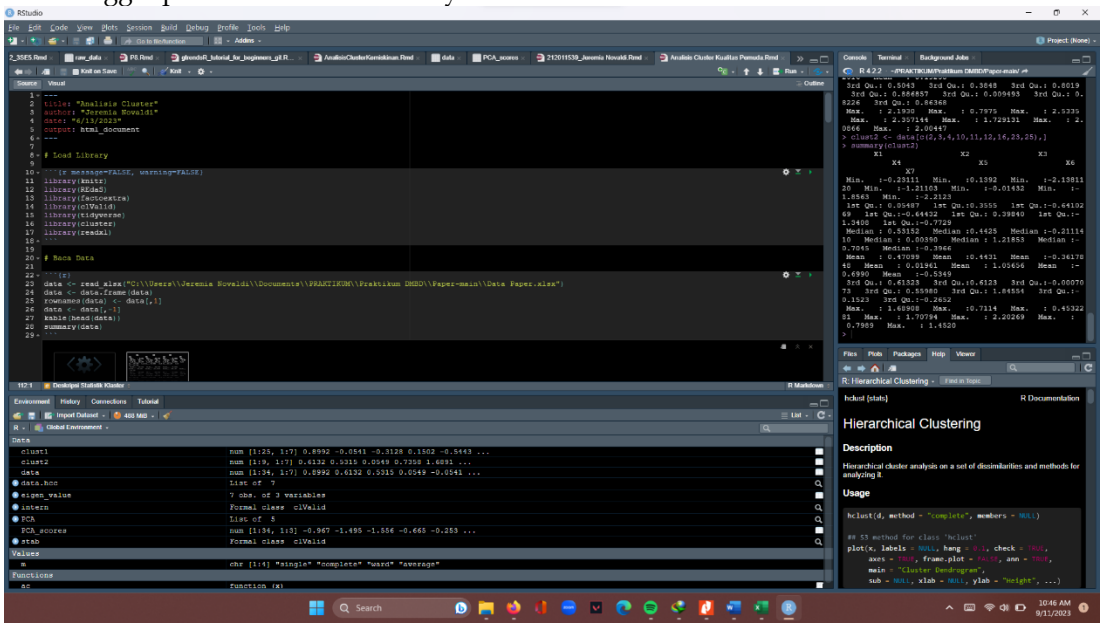
ANALISIS CLUSTER

Hasil validasi *cluster* baik dengan validasi internal maupun stabilitas menyatakan metode hierarki dengan jumlah *cluster* 2 sebagai metode pengelompokan yang terbaik. Metode *agglomerative hierarchical* dengan teknik *single linkage* menghasilkan 2 *cluster* yang berukuran 1 dan 33. Sementara itu, teknik *complete linkage* menghasilkan *cluster* berukuran 1 dan 33. Teknik *average linkage* menghasilkan *cluster* berukuran 1 dan 33. Selain itu, *Ward's* method menghasilkan *cluster* dengan anggota *cluster* sebanyak 11 dan 23.

Berdasarkan tabel 9, diperoleh *agglomerative coefficient* tertinggi pada *Ward's method* yaitu sebesar 0,858. Dengan demikian pengelompokan dengan *Ward's method* merupakan hasil *clustering* terbaik. Teknik ini membagi provinsi-provinsi ke dalam 2 *cluster* sebagai berikut :

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

Analisis cluster dilakukan menggunakan aplikasi RStudio, terlihat pada gambar proses load data dan pembacaan data untuk proses analisis.



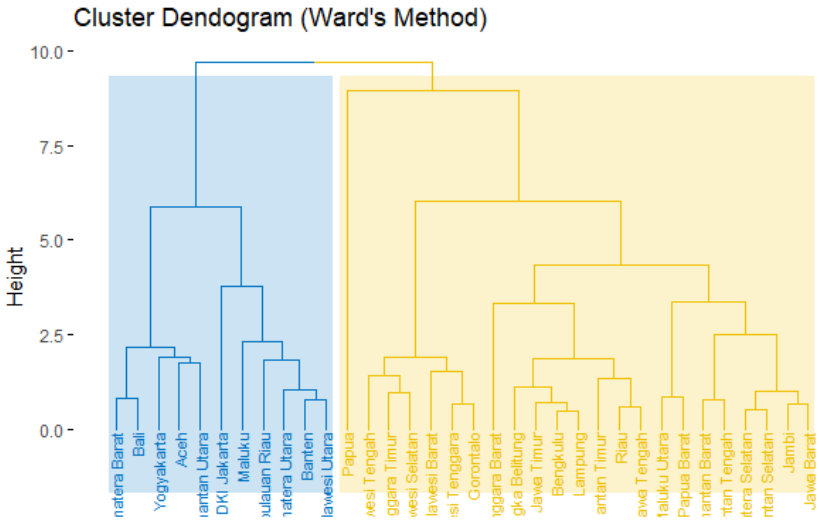
Gambar 7 Analisis Cluster

Pada table 9 terlihat nilai koefisien aglomerasi berdasarkan metode yang digunakan untuk proses clustering.

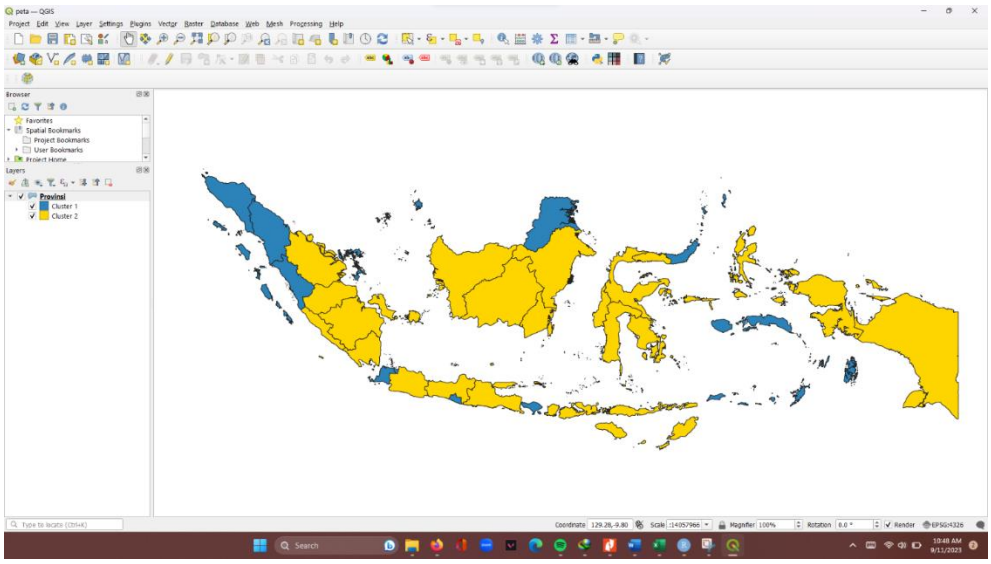
Tabel 9. Koefisien Aglomerasi

Metode	Koefisien Aglomerasi
Single Linkage	0,794
Complete Linkage	0,846
Average Linkage	0,814
Ward's Method	0,858

Selanjutnya pada gambar 8 dapat dilihat ilustrasi bagaimana clusters terbentuk menggunakan Agglomerative Hierarchical dengan Ward's Method, dimana pengelompokan dengan Ward's method merupakan hasil clustering terbaik



Gambar 8. Dendogram *Ward's Method*



Gambar 9. Peta Persebaran Cluster

. Sementara itu, Untuk melihat persebaran cluster menggunakan aplikasi QGIS. Pada Gambar 9 menunjukkan peta persebaran cluster, terlihat bahwa provinsi pada cluster 1 tersebar di seluruh Indonesia, umumnya di ujung suatu pulau.

Tabel 10 mendeskripsikan nilai dari variabel-variabel penelitian pada kedua *cluster*. Provinsi-provinsi dalam Cluster 1 memiliki rata-rata nilai RLS pemuda, persentase pemuda dengan akses internet, persentase pemuda dengan jaminan kesehatan, dan TPT pemuda yang lebih tinggi dibandingkan dengan provinsi yang ada di Cluster 2. Sebaliknya, provinsi dalam Cluster 2 memiliki rataan yang lebih tinggi dari Cluster 1 pada Angka Kesakitan Pemuda, persentase pemuda dengan usia kawin pertama 16 – 18 tahun, dan persentase pemuda perempuan yang melahirkan bayi dengan BBLR.

Tabel 10. Rata-Rata Variabel tiap Cluster

Variabel	Cluster 1	Cluster 2
X1	11,55	10,71
X2	92,73	87,80
X3	8,39	9,49
X4	79,3	72,63
X5	14,39	10,61
X6	14,22	21,52
X7	11,72	14,76

4. KESIMPULAN

Berdasarkan hasil penelitian di atas, ditemukan bahwa metode *clustering* paling baik dalam mengelompokkan provinsi-provinsi di Indonesia menggunakan data kepemudaan adalah metode *Agglomerative Hierarchical* (*Ward's method*) dengan jumlah *cluster* sebanyak 2. Secara umum,

provinsi-provinsi pada Cluster 1 memiliki kualitas pemuda yang lebih baik dibandingkan dengan Cluster 2 terlepas dari indikator TPT yang lebih tinggi. Dengan memperoleh gambaran kepemudaan pada provinsi-provinsi di Indonesia, pemerintah dapat merumuskan kebijakan yang tepat dalam meningkatkan kualitas pemuda yang merupakan penerus bangsa ini.

.

DAFTAR PUSTAKA

[1] B. Andrea, H. C. Gabriella, and J. Tímea, "Y and Z generations at workplaces," *J. Compet.*, vol. 8, no. 3, pp. 90–106, 2016, doi: 10.7441/joc.2016.03.06.

[2] Badan Pusat Statistik, *Statistik Pemuda Indonesia 2022*. Jakarta: Badan Pusat Statistik, 2022.

[3] Pemerintah Indonesia, "Undang-Undang (UU) Nomor 40 Tahun 2009 Tentang Kepemudaan." Sekretariat Negara, Jakarta, 2009, [Online]. Available: https://jdihn.go.id/files/4/2009uu040.pdf.

[4] R. Xu and D. C. Wunsch, *Clustering*. New Jersey: Wiley-IEEE Press, 2008.

[5] S. Landau and I. Chis Ster, "Cluster Analysis: Overview," *Int. Encycl. Educ. Third Ed.*, no. March 2020, pp. 72–83, 2009, doi: 10.1016/B978-0-08-044894-7.01315-4.

[6] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, "Using K-Means Clustering to Cluster Provinces in Indonesia," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018, doi: 10.1088/1742-6596/1028/1/012006.

[7] A. P. Windarto *et al.*, "Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012001.

[8] N. Afira and A. W. Wijayanto, "Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019," *Komputika J. Sist. Komput.*, vol. 10, no. 2, pp. 101–109, 2021, doi: 10.34010/komputika.v10i2.4317.

[9] D. Ls, Y. A. Lesnussa, M. W. Talakua, and M. Y. Matdoan, "Analisis Klaster untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Pendidikan dengan Menggunakan Metode Ward," *J. Stat. dan Apl.*, vol. 5, no. 1, pp. 51–60, 2021, doi: 10.21009/jsa.05105.

[10] A. Azzahra and A. W. Wijayanto, "Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokkan Provinsi Berdasarkan Pelayanan Kesehatan Maternal," *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 481–495, 2022.

[11] R. Johnson and D. . Wichern, *Applied Multivariate Statistical Analysis*, 6th editio. Pearson Prentice Hall, 2007.

[12] A. C. Rencher, *Methods of Multivariate Analysis*. 2002.

[13] Z. Nazari, D. Kang, and M. R. Asharif, "A New Hierarchical Clustering Algorithm," *Int. Conf. Intell. Informatics Biomed. Sci.*, pp. 148–152, 2015.

[14] C. Suhaeni and A. Kurnia, "Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhirarki , K-Means Cluster , dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil)," *J. Media Infotama*, vol. 14, no. 1, pp. 31–38, 2018.

[15] C. Castillo, "Hierarchical Clustering." 2015, [Online]. Available: https://www.slideshare.net/ChaToX/hierarchical-clustering-56364612.

[16] D. Zhang, K. Lee, and I. Lee, "Hierarchical Trajectory Clustering for Spatio-temporal Periodic Pattern Mining," *Expert Syst. Appl.*, 2018, doi: 10.1016/j.eswa.2017.09.040.

[17] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.

[18] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 911–916, 2010, doi: 10.1109/ICDM.2010.35.

[19] A. F. Khairati, A. . Adlina, G. . Hertono, and B. . Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906.

[20] A. Firnanda and A. W. Wijayanto, "Pengelompokan Kabupaten / Kota di Kawasan Timur Indonesia Tahun 2021 Berdasarkan Indikator Sosial Ekonomi Grouping of Regencies / Municipalities in Eastern Indonesia in 2021 Based on Socio - Economic Indicators," *Sist. J. Sist. Inf.*, vol. 12, pp. 390–403, 2023.