

Komparasi Metode Clustering Pada Provinsi di Indonesia Berdasarkan Pertumbuhan Ekonomi Tahun 2022

I Kadek Mira Merta Ningsih^{1*}, Arie Wahyu Wijayanto²

^{1,2}Program Studi Statistika Peminatan Statistika Ekonomi, D-IV Statistika, Politeknik Statistika STIS
Jl. Otto Iskandardinata No. 64C Jakarta 13330

*email: 212011657@stis.ac.id

(Naskah masuk: 19 Juli 2023; diterima untuk diterbitkan: 7 November 2023)

ABSTRAK – Pembangunan ekonomi adalah agenda sentral yang bertujuan untuk mengembangkan perekonomian suatu negara secara berkelanjutan. Mengingat bahwa ekonomi merupakan sektor yang sangat krusial maka pemerataan pertumbuhan ekonomi menjadi suatu hal penting untuk diperhatikan demi pemerataan kesejahteraan masyarakat Indonesia. Peneliti melakukan analisis terkait pengelompokan kondisi pertumbuhan ekonomi provinsi-provinsi di Indonesia tahun 2022 dengan menggunakan metode K-Means, K-Medoids, Fuzzy C-Means dan Hierarki. Data yang digunakan adalah 9 variabel pertumbuhan ekonomi pada 34 provinsi di Indonesia tahun 2022. Hasil akhir diperoleh metode Hierarki Ward dengan jumlah cluster 2 sebagai metode terbaik berdasarkan hasil validasi internal dan validasi stabilitas. Pada metode tersebut diperoleh cluster 1 berjumlah 28 provinsi, sedangkan cluster 2 berjumlah 6 provinsi. Karakteristik cluster 1 yaitu pertumbuhan ekonomi yang tinggi dilihat dari nilai faktor-faktor pembentuk IPM yang tinggi namun memiliki tingkat pengangguran terbuka yang masih tinggi, sedangkan karakteristik cluster 2 yaitu pertumbuhan ekonomi yang rendah dikarenakan nilai gini ratio dan persentase penduduk miskin yang tinggi.

Kata Kunci – Analisis Klaster; Partitioning; Hierarki; Pertumbuhan Ekonomi.

Comparison of Clustering Methods in Provinces in Indonesia Based on Economic Growth in 2022

ABSTRACT – Economic development is a central agenda that aims to develop a country's economy in sustainable manner. Given that the economy is a very crucial sector, equitable distribution of economic growth is an important thing to pay attention for the equal welfare of the Indonesian people. Researchers conducted an analysis related to the grouping of economic growth conditions of provinces in Indonesia in 2022 using the K-Means, K-Medoids, Fuzzy C-Means and Hierarchical methods. The data used are 9 variables of economic growth in 34 provinces in Indonesia in 2022. The final result was obtained by the Ward Hierarchical method with the number of cluster is 2 as the best method based on the results of internal validation and stability validation. In this method, cluster 1 is obtained totaling 28 provinces while cluster 2 totaling 6 provinces. The characteristics of cluster 1 are high economic growth seen from the value of factors forming high HDI but have high open unemployment rate, while the characteristics of cluster 2 are low economic growth due to the high gini ratio and percentage of poor people.

Keywords – Cluster Analysis; Partitioning; Hierarchical; Economic Growth.

1. PENDAHULUAN

Pembangunan ekonomi diartikan sebagai agenda sentral bagi semua negara yang bertujuan untuk mengembangkan perekonomian suatu negara secara berkelanjutan dan dianggap sebagai indikator untuk

menilai sampai seberapa jauh pembangunan suatu daerah dalam periode tertentu. Berdasarkan Badan Pusat Statistik, pertumbuhan ekonomi Indonesia tahun 2022 mencapai Rp19.588,4 triliun yang dihitung dari Produk Domestik Bruto (PDB) ADHB dengan PDB per kapita mencapai Rp71,0 juta atau US\$4.783,9. Sehingga

dapat dikatakan bahwa perekonomian Indonesia tahun 2022 tumbuh sebesar 5,31 persen lebih tinggi dibanding capaian tahun 2021 [1].

Indonesia merupakan negara berkembang dengan wilayah yang tersebar pada berbagai pulau dengan perbedaan karakteristik sehingga pemerataan ekonomi sulit tercapai. Melalui siaran pers Kementerian Koordinator Bidang Perekonomian, pemerintah tengah gencar untuk meningkatkan pemerataan ekonomi dengan menjalankan berbagai kebijakan untuk meningkatkan aktivitas ekonomi, menangani ketimpangan wilayah, serta mendorong pembangunan dan pertumbuhan ekonomi nasional untuk sekarang dan masa depan [2]. Oleh sebab itu, mengingat ekonomi merupakan sektor yang sangat krusial maka pemerataan pertumbuhan ekonomi menjadi suatu hal penting untuk diperhatikan, sebab akan berimbas pada pemerataan kesejahteraan masyarakat Indonesia.

Pertumbuhan ekonomi di suatu wilayah dapat dilihat dari berbagai faktor yaitu ketimpangan pendapatan, kemiskinan suatu wilayah, tingkat pengangguran dan kualitas sumber daya manusia yang dilihat melalui nilai Indeks Pembangunan Manusia (IPM). IPM berperan penting dalam pembangunan perekonomian modern sebab pembangunan manusia yang baik akan menjadikan faktor-faktor produksi mampu dimaksimalkan. Selain itu, pembangunan manusia yang tinggi akan menaikkan tingkat konsumsi sehingga meningkatkan pertumbuhan ekonomi [3]. BPS menyebutkan bahwa 3 dimensi pembentuk IPM adalah umur panjang dan hidup sehat, pengetahuan dan standar hidup layak [4]. Oleh sebab itu pada penelitian ini digunakan beberapa variabel yang merupakan faktor pembentuk IPM untuk mengukur kualitas sumber daya manusia Indonesia.

Penelitian Rahmadi & Parmadi (2019) menjelaskan bahwa ketimpangan yang diukur dengan *gini ratio* dan kemiskinan membawa dampak terhadap peningkatan pertumbuhan ekonomi Indonesia serta pertumbuhan ekonomi provinsi di masing-masing pulau. Dimana ketika *gini ratio* dan kemiskinan suatu wilayah tinggi maka pertumbuhan ekonomi wilayah itu rendah [5]. Sedangkan pada penelitian Ronaldo (2019) menyatakan menurut *Okun's Law*, adanya pengaruh negatif dimana peningkatan pengangguran memperlihatkan adanya ketidakmerataan [6]. Oleh sebab itu, penelitian ini ingin menganalisis terkait pengelompokan kondisi pertumbuhan ekonomi provinsi-provinsi di Indonesia pada tahun 2022. Dari hasil pengelompokan yang dilakukan akan dikaji mengenai hubungan antara pertumbuhan ekonomi dan faktor-faktor yang

memengaruhinya.

2. METODE DAN BAHAN

Data

Penelitian ini menggunakan data sekunder yang diperoleh dari *website* Badan Pusat Statistik (<https://www.bps.go.id/>). Data yang digunakan adalah data tahun 2022 dengan jumlah observasi 34 provinsi. Adapun *software* yang digunakan pada pengolahan data adalah R. Daftar variabel yang digunakan pada penelitian ini dapat dilihat di Tabel 1.

Tabel 1. Daftar variabel penelitian

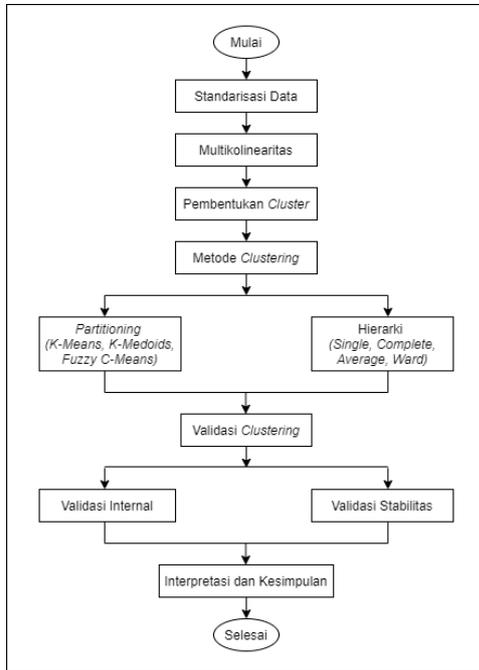
Variabel	Keterangan	Satuan
PE	Pertumbuhan Ekonomi	Persen
GINI	<i>Gini Ratio</i>	0 - 1
P0	Persentase Penduduk Miskin	Persen
TPT	Tingkat Pengangguran Terbuka	Persen
RLS	Rata-Rata Lama Sekolah	Tahun
HLS	Harapan Lama Sekolah	Tahun
PDRBP	PDRB Perkapita	Ribu Rupiah
UHH	Usia Harapan Hidup	Tahun
SL	Persentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak	Persen

Satuan data setiap variabel berbeda-beda sehingga perlu dilakukan standarisasi data terlebih dahulu pada semua variabel sebelum dilakukan analisis *cluster*. Standarisasi data yang digunakan yaitu mengurangi data dengan rata-ratanya dan membagi dengan standar deviasi [7]. Hasil standarisasi data akan menyebabkan variansi antar variabelnya tidak jauh berbeda, sehingga diperoleh pengelompokan yang tepat.

Analisis Cluster

Analisis *cluster* merupakan metode analisis yang tidak memiliki asumsi tentang jumlah kelompok atau struktur kelompok. Pengelompokan dilakukan atas dasar kesamaan atau jarak. Input yang diperlukan adalah ukuran kesamaan atau data yang persamaannya dapat dihitung. Dimana objek-objek pengamatan dibagi menjadi beberapa kelompok, objek-objek dalam satu kelompok memiliki banyak persamaan, sedangkan dengan kelompok lain memiliki banyak perbedaan

sehingga terbentuk pengelompokan secara alami [8]. Adapun kerangka pikir analisis *cluster* pada penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Kerangka Pikir Analisis Cluster

Pemeriksaan Multikolinearitas

Syarat penting yang harus terpenuhi pada analisis *cluster* adalah non-multikolinearitas. Pemeriksaan multikolinearitas dapat dilakukan dengan melihat besar kecilnya koefisien korelasi antar variabel. Jika koefisien korelasi antar variabel $\geq 0,8$ maka terdapat multikolinearitas antar variabel tersebut. Pemeriksaan lainnya dapat dilakukan dengan nilai *Variance Inflation Factors* (VIF) yang dinyatakan sebagai berikut:

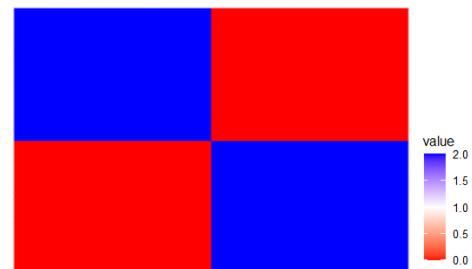
$$VIF_j = \frac{1}{1-R_j^2} \quad (1)$$

R_j^2 adalah korelasi antara variabel prediktor ke-j dengan variabel prediktor lainnya (p-1). VIF_j adalah *Variance Inflation Factors* untuk prediktor j dengan nilai dari *unity* hingga *infinity*. Nilai $VIF_j > 10$ menunjukkan adanya multikolinearitas [9].

Pemeriksaan Pembentukan Cluster

Hopkin Statistics digunakan untuk memeriksa tendensi pengelompokan suatu dataset dengan mengukur peluang suatu data berasal dari distribusi seragam yang menjadi syarat analisis *cluster*. Nilai *Hopkins Statistics* $> 0,5$ dengan k yang lebih besar

menunjukkan kecenderungan data untuk mengelompok dalam ukuran *cluster* yang lebih besar [10]. Sedangkan *Visual Assessment of cluster Tendency* (VAT Algorithm) memberikan gambaran tendensi data dengan memberikan gambaran *clustering* yang dibentuk [11]. Suatu data dapat membentuk *cluster* jika memiliki VAT Algorithm yang berbentuk seperti Gambar 2.



Gambar 2. Visual Assessment of cluster Tendency

Metode Partitioning

Konsep dasar dari metode ini adalah membagi n jumlah *cluster* ke dalam k *cluster*. *Partitioning clustering* merupakan metode pengelompokan yang bertujuan mengelompokkan objek sehingga jarak antar tiap objek ke pusat kelompok di dalam satu kelompok minimum [12]. Metode *partitioning* yang akan digunakan yaitu *K-Means*, *K-Medoids* dan *Fuzzy C-Means*.

A. K-Means

Algoritma *K-Means* mengelompokkan data dengan mencoba memisahkan sampel dalam n kelompok dengan *varians* yang sama, meminimalkan kriteria inersial atau *within-cluster sum-of-squares*. Algoritma ini membutuhkan jumlah *cluster* yang akan ditentukan. Algoritma *K-Means* membagi satu set n sampel X ke dalam C *cluster* yang saling lepas. Kriteria inersia dinyatakan dalam persamaan berikut [13]:

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2) \quad (2)$$

n : populasi, C : *cluster*, x_i : sampel ke-i, μ_j : rata-rata dari sampel di dalam *cluster*.

B. K-Medoids

Algoritma PAM (*Partitioning Around Medoid*) atau algoritma *K-Medoids* adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (*medoids*) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster*. Metode *K-Medoids* lebih cocok untuk mengelompokkan data dibandingkan metode *K-Means* [14].

C. Fuzzy C-Means

Fuzzy C-Means Clustering (FCM) menggunakan model pengelompokan *fuzzy* sehingga data dapat

menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu *cluster* ditentukan oleh derajat keanggotaannya [15]. Output yang dihasilkan dari *Fuzzy C-Means* (FCM) merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. *Fuzzy C-Means* dapat dievaluasi dengan beberapa ukuran yaitu:

1. *Partition Coefficient*

Koefisien ini digunakan untuk mengukur jumlah *overlapping* antar *cluster*. μ_{ij} adalah derajat keanggotaan titik data ke- j di dalam kelompok ke- i . Model terbaik dinyatakan dengan nilai PC maksimal. Persamaannya sebagai berikut [16]:

$$PC = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^N \mu_{ij}^2 \quad (3)$$

N : populasi, C : *cluster*, x_i : sampel ke- i , μ_{ij}^2 : keanggotaan data poin j pada *cluster* i

2. Indeks *Xie and Beni's*

Indeks ini mengukur rasio total *varians within cluster* terhadap *separation* pada *cluster*. Model terbaik adalah model yang memiliki nilai XB minimum. Indeks dirumuskan sebagai berikut [16]:

$$XB = \frac{\sum_{i=1}^C \sum_{j=1}^N (\mu_{ij})^m \left\| x_j - v_i \right\|^2}{N \min_{ij} \left\| x_j - v_i \right\|^2} \quad (4)$$

μ_{ij} : keanggotaan data poin j pada *cluster* i , x_j : sampel ke- j , v_i adalah *cluster* dari i -th *cluster*

Metode Hierarki

Metode hierarki adalah algoritma pengelompokan yang membangun *cluster* bersarang dengan menggabungkan atau memisahkan kelompok secara berurutan. Hierarki *cluster* ini direpresentasikan sebagai pohon (atau dendrogram). Dalam mengukur kesamaan jarak terdapat berbagai macam ukuran pendekatan seperti jarak *Euclidean* dan jarak *Manhattan* [17].

Agglomerative clustering mengelompokan hierarki menggunakan pendekatan dari bawah ke atas yaitu setiap pengamatan dimulai di *cluster*-nya sendiri, dan *cluster* digabungkan secara berurutan. Ada 4 metode *agglomerative* dalam pembentukan *cluster* yaitu [18]:

1. *Single Linkage*

Single linkage akan meminimalkan jarak antara pengamatan terdekat dari pasangan *cluster*.

$$d(uv)_w = \min (d_{uw}d_{vw}) \quad (5)$$

$d(uv)_w$: jarak antara *cluster* (UV) dan *cluster* W
 d_{uw} dan d_{vw} : jarak antara tetangga terdekat *cluster* U dan W, serta *cluster* V dan W.

2. *Complete Linkage*

Complete linkage akan meminimalkan jarak maksimum antara pengamatan pasangan *cluster*.

$$d(uv)_w = \max (d_{uw}d_{vw}) \quad (6)$$

3. *Average Linkage*

Average Linkage akan meminimalkan rata-rata jarak antara semua pengamatan pasangan *cluster*.

$$d(uv)_w = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)}N_w} \quad (7)$$

d_{ik} : jarak antara objek ke- i dalam *cluster* (UV) dan objek ke- k dalam *cluster* ke W

N_{uv} dan N_w : jumlah objek *cluster* (UV) dan W

4. *Ward's Linkage*

Ward's Linkage meminimalkan jumlah kuadrat (ESS) dalam semua *cluster*. Ini adalah pendekatan meminimalkan *varians* yang maknanya mirip dengan tujuan *K-Means* tetapi ditangani dengan pendekatan *agglomerative hierarchical*.

$$ESS = \sum_{j=1}^n x_j^2 - \frac{1}{n} \left(\sum_j x_j \right)^2 \quad (8)$$

n : populasi dan x_j : sampel ke- j

Validasi Internal Cluster

Validasi hasil analisis *cluster* dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Jika *cluster* tidak divalidasi, maka akan berpengaruh pada hasil analisis. Validasi *cluster* yang digunakan pada penelitian ini adalah validasi internal dan validasi stabilitas. Validasi internal yang digunakan antara lain:

1. Indeks *Connectivity*

Kepadatan hubungan mengevaluasi homogenitas dari *cluster* menggunakan *varians intra-cluster*. Hubungan ini menunjukkan posisi tetangga terdekat dari data observasi dalam sebuah *cluster*. Nilai kepadatan tersebut diukur dengan koefisien konektivitas. Nilai *connectivity* berada antara 0 s.d ∞ . *Cluster* yang terbentuk semakin baik saat nilai koefisien *connectivity* semakin rendah [19].

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_i, nn_{i(j)} \quad (9)$$

$nn_{i(j)}$: pengamatan tetangga terdekat (*nearest neighbor*) i ke j dan L

$nn_{i(j)}$: parameter yang menentukan jumlah tetangga yang berkontribusi pada *connectivity* [20].

2. Indeks *Silhouette*

Indeks *silhouette* akan mengevaluasi penempatan setiap objek dalam setiap *cluster* dengan membandingkan jarak rata-rata objek dalam satu *cluster* dan jarak antara objek dengan

cluster yang berbeda [21]. Menghitung koefisien *silhouette* yang didefinisikan sebagai rata-rata $s(i)$:

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (10)$$

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

$$b(i) = \min_{C \in \mathcal{C}} d(i, C), \text{ dan } a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

$b(i)$: nilai minimum dari jarak rata-rata objek i dengan semua objek pada cluster lain C

$a(i)$: rata-rata jarak objek ke- i dengan semua objek yang berada di dalam satu cluster A

Hasil perhitungan indeks *silhouette* berada pada range -1 s.d 1. Semakin besar nilainya akan semakin baik kualitas suatu kelompok.

3. Indeks Dunn

Indeks *dunn* memberikan skor terbaik untuk algoritma *clustering* yang menghasilkan cluster dengan kemiripan tinggi dalam suatu cluster namun kemiripan rendah dengan cluster lain. Indeks *dunn* bertujuan untuk mengidentifikasi cluster yang terpisah dengan baik. Metode ini menghitung rasio jarak antar cluster minimal dengan jarak *intra-cluster* maksimal. Sehingga semakin tinggi nilai indeks *dunn*, maka semakin optimal jumlah cluster yang dihasilkan [22].

$$D = \left\{ \frac{\min_{1 \leq i \leq k} \left(\frac{d(C_i, C_j)}{i + 1 \leq j \leq q} \right)}{\max_{1 \leq l \leq q} d(C_l)} \right\} \quad (11)$$

$d(C_i, C_j)$: ukuran kedekatan cluster i dan cluster j

$d(C_l)$: ukuran kedekatan anggota dalam cluster l

4. Indeks Davies-Bouldin

Indeks *Davies-Bouldin* merupakan salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengklasteran yang didasarkan pada nilai kohesi dan separasi. Semakin kecil nilai pada indeks, maka hasil pengklasteran semakin baik. Indeks DB dirumuskan sebagai berikut [18]:

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (12)$$

$$R_i = \max_{j=1, \dots, k, i \neq j} R_{ij}; R_{ij} = \frac{(S_i + S_j)}{d_{ij}}; d_{ij} = d(v_i, v_j);$$

v_i : nilai rata-rata pada cluster C_i ; v_j : nilai rata-rata pada cluster C_j ; $S_i = \frac{1}{|C_i|} \sum_{x \in C_i} d(x, v_i)$; $|C_i|$:

jumlah data yang termasuk cluster ke- i

Validasi Stabilitas Cluster

Validasi stabilitas membandingkan hasil analisis cluster berdasarkan penghapusan setiap kolom pada data, satu per satu. Pada validasi ini digunakan nilai *average proportion of non-overlap* (APN), *average distance* (AD), *average distance between means* (ADM), dan *figure of merit* (FOM). Semakin kecil nilai APN, AD, ADM dan FOM, semakin baik cluster yang dihasilkan [23].

3. HASIL DAN PEMBAHASAN

Deskripsi Statistik

Berikut ditampilkan deskripsi statistik mengenai variabel pertumbuhan ekonomi provinsi-provinsi di Indonesia tahun 2022 berdasarkan ukuran pemusatan dan ukuran penyebarannya.

Tabel 2. Deskripsi statistik variabel

Variabel	Minimum	Mean	Maximum
Pertumbuhan Ekonomi (PE)	-0,15	0,051	0,23
Gini Ratio (GINI)	0,255	0,343	0,459
Persentase Penduduk Miskin (P0)	4,53	10,299	26,8
Tingkat Pengangguran Terbuka (TPT)	2,34	4,966	8,31
Rata-Rata Lama Sekolah (RLS)	6,76	8,716	11,17
Harapan Lama Sekolah (HLS)	11,14	13,24	15,65
PDRB Perkapita (PDRBP)	13299	45361	182909
Usia Harapan Hidup (UHH)	65,25	70,15	75,04
Persentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak (SL)	40,81	80,97	97,12

Multikolinearitas

Sebelum memilih variabel independen yang akan digunakan, dilakukan pemeriksaan multikolinearitas. Pada hasil perhitungan koefisien korelasi *pearson*, diperoleh bahwa tidak adanya nilai $\geq 0,8$ pada pemeriksaan korelasi antar variabel independen, sehingga terbukti tidak ada multikolinearitas. Sedangkan hasil pemeriksaan multikolinearitas dengan VIF dapat dilihat dari Tabel 3.

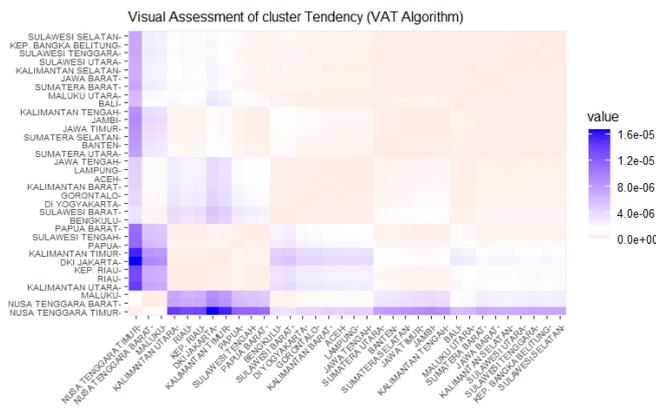
Tabel 3. Nilai VIF variabel independen

Variabel	VIF
Gini Ratio (GINI)	1,339

Persentase Penduduk Miskin (P0)	2,816
Tingkat Pengangguran Terbuka (TPT)	1,714
Rata-Rata Lama Sekolah (RLS)	3,887
Harapan Lama Sekolah (HLS)	2,368
PDRB Perkapita (PDRBP)	2,267
Usia Harapan Hidup (UHH)	2,027
Persentase Rumah Tangga yang Memiliki Akses terhadap Sanitasi Layak (SL)	2,394

Hasil yang sama diperoleh pada pemeriksaan VIF, ketika dibuat persamaan regresi dengan pertumbuhan ekonomi (PE) sebagai variabel dependen ternyata tidak ada nilai VIF variabel independen > 10 sehingga terbukti bahwa tidak ada multikolinearitas. Oleh sebab itu semua variabel independen dapat langsung dimasukkan sebagai faktor pembentuk *cluster*.

Pengecekan Pembentukan Cluster



Gambar 3. Visual Assessment of cluster Tendency pertumbuhan ekonomi provinsi-provinsi Indonesia 2022

Berdasarkan hasil pengujian *Hopkin Statistics* diperoleh nilai 0,716 > 0,5 sehingga mengindikasikan dapat dilakukannya pengelompokan pada data pertumbuhan ekonomi. Selain itu pada *Visual Assessment of cluster Tendency (VAT Algorithm)* terlihat bahwa data cocok untuk dibentuk *cluster*.

Analisis Cluster K-Means

Dilakukan uji validasi *cluster* terlebih dahulu sebelum melakukan analisis *cluster* untuk mengetahui jumlah *cluster* terbaik pada metode *K-Means*. Nilai indeks validasi internal dilihat pada Tabel 4.

Tabel 4. Validasi internal *K-Means*

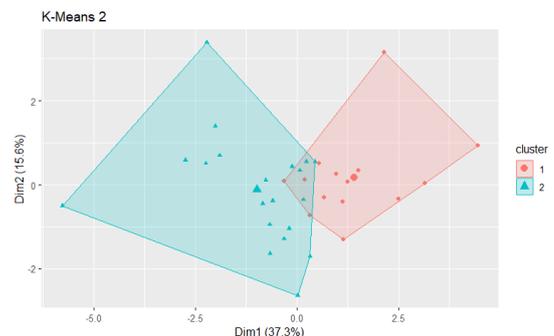
Validasi Internal	Cluster				
	2	3	4	5	6
Connectivity	6,767	16,774	18,808	18,808	29,697
Silhouette	0,302	0,319	0,342	0,397	0,337
Dunn	0,296	0,272	0,244	0,250	0,221
Davies-Bouldin	2,075	1,374	1,804	1,453	1,296

Berdasarkan Tabel 4 terlihat bahwa metode *K-Means* dengan jumlah *cluster* 2 memiliki nilai *Connectivity* lebih kecil dan nilai *Dunn* lebih besar dibandingkan jumlah *cluster* lainnya. Sedangkan nilai *Silhouette* paling tinggi pada jumlah *cluster* 5 dan nilai indeks *Davies-Bouldin* lebih rendah pada jumlah *cluster* 6. Selain itu, nilai validasi stabilitas dapat dilihat pada Tabel 5.

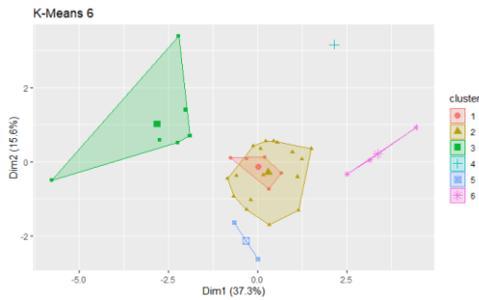
Tabel 5. Validasi stabilitas *K-Means*

Validasi Stabilitas	Cluster				
	2	3	4	5	6
APN	0,0742	0,076	0,097	0,093	0,171
AD	3,4903	3,253	2,940	2,697	2,567
ADM	0,5609	0,831	0,656	0,491	0,828
FOM	0,947	0,943	0,912	0,853	0,850

Setelah itu dilakukan uji validasi stabilitas, diperoleh nilai APN dengan metode *K-Means* jumlah *cluster* 2 lebih rendah dibandingkan jumlah *cluster* lainnya. Jika dilihat nilai AD dan FOM lebih rendah pada *K-Means* dengan 6 *cluster*. Sedangkan nilai AD lebih rendah pada *cluster* 5. Oleh sebab itu kemungkinan jumlah *cluster* yang optimum untuk metode *K-Means* adalah 2 dan 6. Persebaran data *K-Means* ditampilkan seperti Gambar 4 dan Gambar 5.



Gambar 4. Cluster Plot *K-Means* dengan 2 cluster



Gambar 5. Cluster Plot K-Means dengan 6 cluster

Pada Gambar 4 dan Gambar 5 terlihat masih adanya tumpang tindih antara cluster 1 dan cluster 2 sehingga pembentukan data menjadi 2 cluster dan 6 cluster dengan metode K-Means kurang tepat dalam mengelompokkan data.

Analisis Cluster K-Medoids

Dilakukan uji validasi cluster terlebih dahulu sebelum melakukan analisis cluster untuk mengetahui jumlah cluster terbaik pada metode K-Medoids. Nilai indeks validasi internal dapat dilihat pada Tabel 6.

Tabel 6. Validasi internal K-Medoids

Validasi Internal	Cluster				
	2	3	4	5	6
Connectivity	10,589	25,817	27,946	30,564	32,597
Silhouette	0,255	0,233	0,243	0,247	0,329
Dunn	0,271	0,147	0,152	0,164	0,180
Davies-Bouldin	1,628	2,259	1,748	1,391	1,201

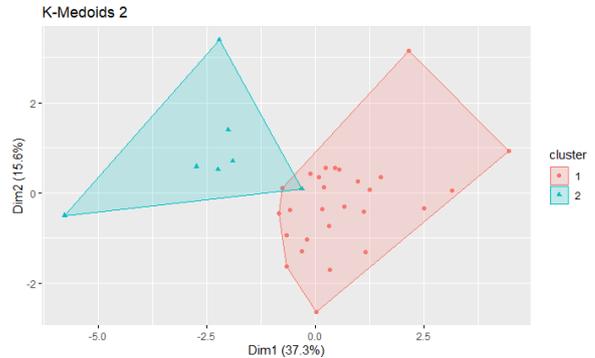
Berdasarkan Tabel 6 terlihat bahwa metode K-Medoids dengan jumlah cluster 2 memiliki nilai indeks Connectivity lebih kecil dan indeks Dunn lebih besar dibandingkan jumlah cluster lainnya. Sedangkan nilai indeks Silhouette tertinggi dan indeks Davies-Bouldin terendah ada pada K-Medoids dengan jumlah cluster 6. Sedangkan nilai validasi stabilitas dapat dilihat pada Tabel 7.

Tabel 7. Validasi stabilitas K-Medoids

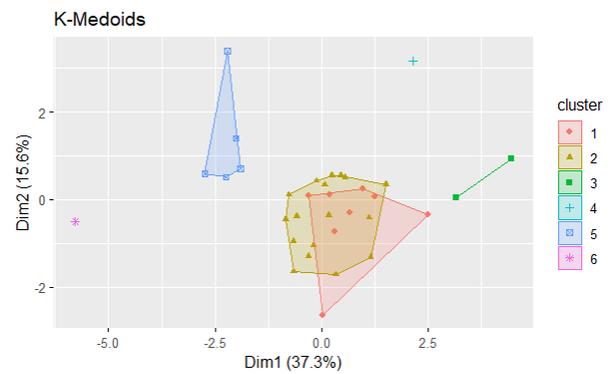
Validasi Stabilitas	Cluster				
	2	3	4	5	6
APN	0,0761	0,183	0,114	0,133	0,144
AD	3,4011	3,214	2,947	2,760	2,502
ADM	0,3083	0,673	0,618	0,654	0,586
FOM	0,9335	0,916	0,903	0,903	0,858

Setelah itu dilakukan uji validasi stabilitas, diperoleh nilai APN dan ADM terendah ada pada

metode K-Medoids dengan jumlah cluster 2. Sedangkan nilai AD dan FOM lebih rendah pada cluster 6 seperti pada Gambar 7.



Gambar 6. Cluster Plot K-Medoids dengan 2 cluster



Gambar 7. Cluster Plot K-Medoids dengan 6 cluster

Pada Gambar 6 dan Gambar 7 terlihat masih adanya tumpang tindih antara cluster 1 dan cluster 2 sehingga pembentukan data menjadi 2 cluster dan 6 cluster dengan metode K-Medoids kurang tepat dalam mengelompokkan data.

Analisis Cluster Fuzzy C-Means

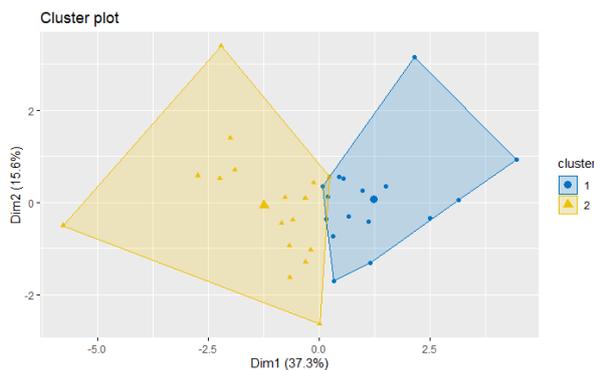
Untuk memilih analisis cluster dengan metode Fuzzy C-Means dilakukan perbandingan hasil validasi internal cluster yang ditampilkan pada Tabel 8 berikut.

Tabel 8. Perbandingan validasi internal Fuzzy C-Means

Validasi Internal	Cluster				
	2	3	4	5	6
Silhouette	0,319	0,204	0,115	0,095	0,089
Dunn	0,522	0,381	0,305	0,244	0,202
Partition Coefficient	0,522	0,381	0,305	0,244	0,202
Davies-Bouldin	2,094	2,022	2,078	2,121	2,669

<i>Xie and Beni Index</i>	4,555	6,176	$1,37 \times 10^{10}$	$9,93 \times 10^{12}$	$4,16 \times 10^{18}$
---------------------------	-------	-------	-----------------------	-----------------------	-----------------------

Berdasarkan Tabel 8 diperoleh bahwa metode *Fuzzy C-Means* yang menghasilkan *cluster* terbaik adalah *cluster* 2 karena nilai *Partition Coefficient*, *Silhouette Coefficient* dan *Dunn Index* paling tinggi, *Xie and Beni Index* lebih rendah dibandingkan *cluster* dengan jumlah lainnya, sedangkan *Davies-Bouldin Index* terendah ketiga setelah *Fuzzy C-Means* dengan 3 *cluster* dan 4 *cluster* seperti pada Gambar 8.



Gambar 8. Cluster Plot *Fuzzy C-Means* dengan 2 *cluster*

Analisis Cluster Hierarki

Dilakukan uji validasi *cluster* terlebih dahulu sebelum melakukan analisis *cluster* untuk mengetahui jumlah *cluster* terbaik pada metode hierarki. Nilai indeks validasi internal dapat dilihat pada Tabel 9.

Tabel 9. Validasi internal hierarki

Validasi Internal	Cluster				
	2	3	4	5	6
<i>Connectivity</i>	2,929	6,108	9,037	11,966	15,824
<i>Silhouette</i>	0,593	0,403	0,479	0,546	0,403
<i>Dunn</i>	0,476	0,362	0,311	0,299	0,184

Berdasarkan Tabel 9 terlihat bahwa metode hierarki dengan *cluster* 2 memiliki nilai indeks *Connectivity* paling kecil, indeks *Dunn* dan *Silhouette* lebih besar dibandingkan jumlah *cluster* lainnya. Sedangkan nilai validasi stabilitas dapat dilihat di Tabel 10.

Tabel 10. Validasi stabilitas hierarki

Validasi Stabilitas	Cluster				
	2	3	4	5	6
APN	0,006	0,032	0,044	0,053	0,058
AD	3,515	3,366	3,128	2,821	2,632
ADM	0,073	0,291	0,306	0,216	0,246
FOM	0,958	0,928	0,932	0,814	0,817

Setelah itu dilakukan uji validasi stabilitas, diperoleh nilai APN dan ADM terendah pada metode hierarki dengan jumlah *cluster* 2. Sedangkan nilai AD lebih rendah pada *cluster* 6 dan nilai FOM lebih rendah pada *cluster* 5. Oleh sebab itu jumlah *cluster* yang optimum untuk metode hierarki adalah 2. *Cluster* yang ditampilkan pada Tabel 11.

Tabel 11. Perbandingan nilai *agglomerative coefficient*

Metode	<i>Agglomerative Coefficient</i>
Single	0,615
Complete	0,779
Average	0,686
Ward	0,801

Berdasarkan Tabel 11 diperoleh bahwa metode *linkage* yang menghasilkan *cluster* terbaik yaitu metode *ward* karena nilai *agglomerative coefficient* lebih besar dibandingkan metode *linkage* lainnya.

Pemilihan Metode Cluster Terbaik

Pemilihan metode *cluster* terbaik dilakukan dengan perbandingan nilai validasi internal yang meliputi indeks *Silhouette*, *Dunn*, dan *Davies-Bouldin* yang ditampilkan pada Tabel 12.

Tabel 12. Perbandingan metode *cluster*

Metode	Validasi Internal		
	<i>Silhouette</i>	<i>Dunn</i>	<i>Davies-Bouldin</i>
<i>K-Means</i> (K=2)	0,302	0,296	2,075
<i>K-Means</i> (K=6)	0,329	0,180	1,296
<i>K-Medoids</i> (K=2)	0,255	0,271	1,628
<i>K-Medoids</i> (K=6)	0,329	0,180	1,201
<i>Fuzzy C-Means</i> (K=2)	0,319	0,522	2,094
Hierarki <i>Ward</i> (K=2)	0,593	0,476	1,443

Hasil uji validasi internal menunjukkan metode terbaik adalah Hierarki *Ward cluster* 2. Hal ini sebab pada nilai indeks *Silhouette* dibandingkan metode lain lebih mendekati 1, nilai indeks *Dunn* terbesar kedua setelah *Fuzzy C-Means* (K=2) dan nilai indeks *Davies-Bouldin* terkecil ketiga setelah metode *K-Medoids* (K=6) dan *K-Means* (K=6). Sehingga peneliti menggunakan metode Hierarki *Ward cluster* 2 sebagai metode terbaik untuk mengelompokkan pertumbuhan ekonomi 2022.

Karakteristik Hasil Cluster Metode Terbaik

Tabel 13. Deskripsi statistik setiap *cluster*

Variabel	Cluster 1	Cluster 2
Pertumbuhan Ekonomi (PE)	0,144	-0,674
Gini Ratio (GINI)	-0,177	0,826
Persentase Penduduk Miskin (P0)	-0,324	1,511
Tingkat Pengangguran Terbuka (TPT)	0,229	-1,067
Rata-Rata Lama Sekolah (RLS)	0,266	-1,241
Harapan Lama Sekolah (HLS)	0,093	-0,435
PDRB Perkapita (PDRBP)	0,105	-0,489
Usia Harapan Hidup (UHH)	0,306	-1,429
Persentase Ruta yang Memiliki Akses terhadap Sanitasi Layak (SL)	0,188	-0,876

Pada Tabel 13 dapat diketahui karakteristik dari setiap *cluster* yang terbentuk oleh metode *ward*, berdasarkan variabel pertumbuhan ekonomi yaitu :

- Kelompok 1 (28 provinsi) : rata-rata lebih tinggi pada variabel PE, TPT, RLS, HLS, PDRBP, UHH dan SL.
- Kelompok 2 (6 provinsi) : rata-rata lebih tinggi pada variabel GINI dan P0.

Berdasarkan karakteristik tersebut maka *cluster* 1 merupakan provinsi-provinsi dengan pertumbuhan ekonomi yang tinggi. Hal ini diketahui dari nilai rata-rata lama sekolah, harapan lama sekolah, PDRB perkapita, usia harapan hidup dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak yang nilainya lebih tinggi dibandingkan *cluster* 2. Dimana variabel-variabel tersebut merupakan faktor pembentuk IPM, artinya kualitas sumber daya manusia di *cluster* 1 cenderung baik, walaupun kelompok ini memiliki TPT yang lebih tinggi dibandingkan *cluster* 2. Sedangkan *cluster* 2 merupakan provinsi dengan pertumbuhan ekonomi rendah sebab nilai *gini ratio* (ukuran ketimpangan) dan persentase penduduk miskin yang lebih tinggi dibandingkan *cluster* 1.

Tingginya nilai TPT pada kelompok pertumbuhan ekonomi tinggi (*cluster* 1) disebabkan ekonomi yang baru pulih pasca pandemi Covid-19. BPS menyebutkan bahwa TPT Agustus 2022 turun sebesar 0,63 persen poin dibandingkan dengan Agustus 2021. Namun masih ada 4,15 juta orang (1,98 persen) penduduk usia kerja yang terdampak Covid-19 yang terdiri atas pengangguran (0,24 juta orang), bukan angkatan kerja (0,32 juta orang), sementara tidak bekerja (0,11 juta orang) dan penduduk bekerja yang mengalami pengurangan jam kerja (3,48 juta orang) [24].

Pemerataan pertumbuhan ekonomi antar provinsi di Indonesia pada tahun 2022 termasuk baik, karena hanya sedikit provinsi yang tergolong dalam *cluster* 2 (pertumbuhan ekonomi rendah). Selain itu, pada *cluster* 1 walaupun tergolong provinsi dengan pertumbuhan

ekonomi tinggi namun masih memiliki TPT yang tinggi, sehingga masalah lapangan pekerjaan dan penyerapan tenaga kerja masih perlu mendapatkan perhatian agar pertumbuhan ekonomi dapat ditingkatkan.

4. KESIMPULAN

Pertumbuhan ekonomi provinsi-provinsi di Indonesia tahun 2022 membentuk *cluster* dengan metode terbaik yaitu Hierarki *Ward* 2 *cluster* berdasarkan hasil validasi internal dan stabilitas dibandingkan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Hasil akhir diperoleh, *cluster* 1 berjumlah 28 provinsi, sedangkan *cluster* 2 berjumlah 6 provinsi. Karakteristik *cluster* 1 yaitu pertumbuhan ekonomi tinggi dilihat dari nilai faktor-faktor pembentuk IPM yang tinggi, namun dengan TPT yang masih tinggi, sedangkan karakteristik *cluster* 2 yaitu pertumbuhan ekonomi rendah diketahui dari nilai *gini ratio* dan persentase penduduk miskin yang tinggi. Dengan mengetahui karakteristik pertumbuhan tiap provinsi, pemerintah dapat membuat kebijakan dengan tepat dalam mengatasi masalah pertumbuhan ekonomi. Diharapkan pada penelitian berikutnya dapat menambah penggunaan metode *clustering* dan dilakukan analisis tambahan untuk menjelaskan pengaruh faktor-faktor pertumbuhan ekonomi.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Pertumbuhan Ekonomi Indonesia Triwulan IV-2022," Jakarta, 2023. [Online]. Available: <https://www.bps.go.id/pressrelease/2020/02/05/1755/ekonomi-indonesia-2019-tumbuh-5-02-persen.html>
- [2] H. Limanseto, "Memeratakan Pembangunan Ekonomi di Indonesia, Pemerintah Dorong Pengembangan PSN Prioritas di Berbagai Wilayah," Surabaya, 2022. [Online]. Available: <https://www.ekon.go.id/publikasi/detail/4660/memeratakan-pembangunan-ekonomi-di-indonesia-pemerintah-dorong-pengembangan-psn-prioritas-di-berbagai-wilayah>
- [3] A. B. Susanto and L. Rachmawati, "Pengaruh Indeks Pembangunan (IPM) dan Inflasi Terhadap Pertumbuhan Ekonomi di Kabupaten Lamongan," *J. Pendidik. Ekon.*, vol. 1, no. 3, pp. 1–18, 2013.
- [4] "Konsep Indeks Pembangunan Manusia," *Badan Pusat Statistik*. <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html> (accessed Jul. 18, 2023).

- [5] S. Rahmadi and P. Parmadi, "Pengaruh Ketimpangan Pendapatan dan Kemiskinan Terhadap Pertumbuhan Ekonomi Antar Pulau di Indonesia," *J. Paradig. Ekon.*, vol. 14, no. 2, pp. 55–66, 2019.
- [6] R. Ronaldo, "Pengaruh Inflasi dan Tingkat Pengangguran terhadap Pertumbuhan Ekonomi Makro di Indonesia," *J. Ekon.*, vol. 21, no. 2, pp. 137–153, 2019.
- [7] G. A. Dito and A. A. Nurussadad, "STA551 Pemodelan Klasifikasi - Analisis Cluster dengan R," 2021. https://rstudio-pubs-static.s3.amazonaws.com/805596_1c2dc0d4e5364833abda18b3b9deb42d.html# (accessed Nov. 03, 2023).
- [8] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Sixth. New Jersey: Pearson Prentice Hall, 2007.
- [9] C. G. Thompson, R. S. Kim, A. M. Aloe, and B. J. Becker, "Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results," *Basic Appl. Soc. Psych.*, vol. 39, no. 2, pp. 1–10, 2017.
- [10] R. F. Lanclan, "Hopkins Statistic," *ResearchGate*. https://www.researchgate.net/figure/Hopkins-statistic-of-clustering-tendency-for-k-nearest-neighbor-distances-in-chaffinches_fig5_41531093 (accessed Jul. 18, 2023).
- [11] A. Anjani Arifiyanti, F. Setiyo Darusman, and B. Widyo Trenggono, "Population Density Cluster Analysis in DKI Jakarta Province Using K-Means Algorithm," *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 772–783, 2022, [Online]. Available: <http://journal-isi.org/index.php/isi>
- [12] M. Benri, H. Metisen, and S. Latipa, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015, [Online]. Available: <https://core.ac.uk/download/pdf/287160954.pdf>
- [13] "2.3.2. K-Means," *scikit learn*. <https://scikit-learn.org/stable/modules/clustering.html#k-means> (accessed Jul. 18, 2023).
- [14] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, pp. 64–71, 2018.
- [15] E. T. Luthfi, "Fuzzy C-Means Untuk Clustering Data (Studi Kasus : Data Performance Mengajar Dosen)," *Semin. Nas. Teknol.* 2007, pp. 1–7, 2007.
- [16] A. W. Wijayanto and Takdir, "Fighting Cyber Crime in Email Spamming: An Evaluation of Fuzzy Clustering Approach to Classify Spam Messages," *Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2014*, pp. 19–24, 2014.
- [17] "2.3.6. Hierarchical Clustering," *scikit learn*. <https://scikit-learn.org/stable/modules/clustering.html#hierarchical-clustering> (accessed Jul. 18, 2023).
- [18] C. Suhaeni, A. Kurnia, and R. Ristiyanti, "Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhierarchy, 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.
- [19] N. Thamrin and A. W. Wijayanto, "Analisis Cluster dengan Menggunakan Hard Clustering dan Soft Clustering untuk Pengelompokan Tingkat Kesejahteraan Kabupaten/Kota di Pulau Jawa," *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 141–160, 2021.
- [20] N. N. Halim and E. Widodo, "Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps," *Pros. SI MaNIS (Seminar Nas. Integr. Mat. dan Nilai Islam.*, vol. 1, no. 1, pp. 188–194, 2017, [Online]. Available: <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>
- [21] M. A. Nahdliyah, T. Widiyari, and A. Prahutama, "Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018)," *J. Gaussian*, vol. 8, no. 2, pp. 161–170, 2019.
- [22] H. Malikhatin, A. Rusgiyono, and D. A. I. Maruddani, "Penerapan K-Modes Clustering dengan Validasi Dunn Index pada Pengelompokan Karakteristik Calon TKI Menggunakan R-GUI," *J. Gaussian*, vol. 10, no. 3, pp. 359–366, 2021.
- [23] 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.
- [24] Badan Pusat Statistik, "Keadaan Ketenagakerjaan Indonesia Agustus 2022," Jakarta, 2022. [Online]. Available: <https://www.bps.go.id/pressrelease/2022/11/07/1916/agustus-2022--tingkat-pengangguran-terbuka--tpt--sebesar-5-86-persen-dan-rata-rata-upah-buruh-sebesar-3-07-juta-rupiah-per-bulan.html>