

# Perbandingan Algoritma *Partitioning* dan *Hierarchical Clustering* untuk Pengelompokan Wilayah Menurut Karakteristik Pengangguran di Pulau Jawa Tahun 2021

Delvina Nur Rahmawati<sup>1</sup>, Arie Wahyu Wijayanto<sup>2</sup>

Politeknik Statistika STIS

Jl. Otto Iskandardinata No.64C, Jakarta, Indonesia

[211910739@stis.ac.id](mailto:211910739@stis.ac.id), [ariewahyu@stis.ac.id](mailto:ariewahyu@stis.ac.id)

## **Abstract**

*Unemployment is a complex problem because it is influenced and affects various sectors of life. The Open Unemployment Rate in Indonesia in 2021 is 6.49 percent where Java Island is the island with the highest TPT compared to other islands with a total of 5,948,406 unemployed people or as much as 65.35 percent of the total number of unemployed people throughout Indonesia who are on the Java Island. Therefore, it is important to classify regions in Java Island based on the characteristics of unemployment so that the government can formulate appropriate policies to reduce unemployment. The data used includes 7 variables related to unemployment in 119 regencies/cities in Java Island in 2021. This research uses two clustering methods, namely partitioning and hierarchy to group regencies/cities in Java Island based on unemployment characteristics. The selected partitioning method is K-Means. Determining the number of clusters using internal validation and stability validation shows that the hierarchical method with the number of clusters 2 is the most optimal clusters where the Ward method is able to provide the best grouping results based on the agglomerative coefficient value.*

**Keywords:** *Unemployment, Cluster, Partitioning, Hierarchical*

## **Abstrak**

*Pengangguran merupakan masalah yang kompleks karena dipengaruhi sekaligus memengaruhi berbagai sektor kehidupan. Tingkat Pengangguran Terbuka (TPT) di Indonesia pada tahun 2021 berada pada angka 6,49 persen di mana Pulau Jawa merupakan pulau dengan TPT tertinggi dibandingkan pulau lainnya dengan jumlah pengangguran sebanyak 5.948.406 jiwa atau sebanyak 65,35 persen dari total jumlah pengangguran di seluruh Indonesia berada di Pulau Jawa. Oleh karena itu, penting untuk mengelompokkan daerah-daerah di Pulau Jawa berdasarkan karakteristik pengangguran sehingga pemerintah dapat dengan tepat merumuskan kebijakan untuk menekan angka pengangguran. Data yang digunakan meliputi 7 variabel terkait pengangguran pada 119 kabupaten/kota di Pulau Jawa tahun 2021. Penelitian ini menggunakan dua metode clustering, yaitu partitioning dan hierarki untuk mengelompokkan kabupaten/kota di Pulau Jawa berdasarkan karakteristik pengangguran. Metode partitioning yang dipilih adalah K-Means. Penentuan jumlah cluster menggunakan validasi internal dan validasi stabilitas menunjukkan bahwa metode hierarki dengan jumlah cluster 2 merupakan cluster yang paling optimal di mana metode Ward*

mampu memberikan hasil pengelompokan terbaik berdasarkan nilai *agglomerative coefficient*.

**Kata kunci:** Pengangguran, Cluster, Partitioning, Hierarki

## 1. PENDAHULUAN

Indonesia merupakan salah satu negara yang memiliki jumlah penduduk terbanyak di dunia. Berdasarkan hasil rilis Sensus Penduduk 2020, jumlah penduduk Indonesia mencapai 270,20 juta jiwa pada September 2020 [1]. Data tersebut di-update oleh Kemendagri bahwasanya hingga 30 Desember 2021 jumlah penduduk Indonesia adalah 273.879.750 jiwa [2]. Pertambahan jumlah penduduk yang kian pesat menyebabkan Indonesia terjerat pada berbagai masalah yang hingga saat ini belum mampu teratasi. Salah satu permasalahan yang masih menjadi fokus tersendiri untuk Indonesia yaitu pengangguran.

Pengangguran merupakan masalah yang kompleks karena dipengaruhi sekaligus memengaruhi berbagai sektor kehidupan. Pengangguran didefinisikan sebagai suatu keadaan di mana seseorang yang termasuk dalam angkatan kerja ingin memperoleh pekerjaan tetapi belum mendapatkan pekerjaan yang sesuai atau yang diinginkannya [3]. Pengangguran terdiri atas penduduk yang aktif mencari pekerjaan, penduduk yang sedang mempersiapkan usaha/pekerjaan baru, penduduk yang tidak mencari pekerjaan karena merasa tidak mungkin mendapat pekerjaan, dan kelompok penduduk yang tidak aktif mencari pekerjaan dengan alasan sudah mempunyai pekerjaan tetapi belum mulai bekerja [4]. Pengangguran berdampak buruk pada kesehatan keluarga, kesejahteraan masyarakat, prospek pembangunan ekonomi, kekacauan politik yang dapat berimbas pada stabilitas negara, dan bahkan dapat berefek pada bertambahnya masyarakat miskin [5]. Selain itu, tidak adanya pendapatan yang diperoleh para penganggur sementara tuntutan kebutuhan hidup yang kian tinggi dan harus terpenuhi dapat mendorong terjadinya tindak kriminal [6].

Masalah pengangguran menjadi menarik untuk ditelaah karena hampir seluruh negara di dunia masih belum mampu terbebas dari masalah tersebut. Selain itu, dampak buruk yang kompleks dari permasalahan pengangguran ini juga melatarbelakanginya dijadikan salah satu tujuan dalam *Sustainable Development Goals* (SDGs). Komitmen untuk pengangguran tercantum pada tujuan kedelapan yaitu mendukung pertumbuhan ekonomi yang inklusif dan berkelanjutan, tenaga kerja penuh dan produktif, dan pekerjaan yang layak bagi semua.

Tabel 1. Tingkat Pengangguran berdasarkan Pulau di Indonesia 2021

Pulau	Jumlah Pengangguran	Jumlah Angkatan Kerja	TPT	Distribusi Pengangguran (Persen)
Sumatera	1678731	29827775	5,628	18,44
Jawa	5948406	79799689	7,454	65,35
Bali dan Nusa Tenggara	331092	8238961	4,019	3,64
Kalimantan	469559	8466389	5,546	5,16

Sulawesi	491561	9896143	4,967	5,40
Maluku dan Papua	182703	3923618	4,656	2,01
Indonesia	9102052	140152575	6,494	100,00

Berdasarkan Tabel 1, indikator pengangguran yang biasanya dinyatakan dalam Tingkat Pengangguran Terbuka (TPT) di Indonesia pada tahun 2021 berada pada angka 6,49 persen. Pulau Jawa merupakan pulau dengan TPT tertinggi dibandingkan pulau lainnya dan satu-satunya pulau yang tingkat pengangguran terbukanya diatas angka nasional yaitu 7,45 persen dengan jumlah pengangguran sebanyak 5.948.406 jiwa di mana sebanyak 65,35 persen dari total jumlah pengangguran di seluruh Indonesia berada di Pulau Jawa. Fenomena semacam ini mengindikasikan bahwa strategi dalam penanganan pengangguran yang diterapkan belum mampu menekan angka pengangguran terutama di Pulau Jawa. Problematika yang dialami masyarakat di Pulau Jawa ini dapat menghambat kesejahteraan penduduk dan memunculkan berbagai permasalahan lainnya yang lebih kompleks.

Penanganan masalah pengangguran ini harus dilakukan dengan memperhatikan aspek-aspek yang memengaruhinya. Menurut Astuti (2017) terdapat tiga aspek yang berkaitan dengan pengangguran yaitu aspek kependudukan, ekonomi, dan pendidikan. Aspek kependudukan menunjukkan seiring bertambahnya jumlah penduduk yang tidak diikuti oleh lapangan pekerjaan yang memadai akan menimbulkan pengangguran. Selanjutnya pada aspek ekonomi, aktivitas perekonomian yang menurun akan meningkatkan jumlah pengangguran. Sedangkan aspek pendidikan menunjukkan kualitas pendidikan yang diimbangi dengan keterampilan akan mampu menurunkan pengangguran [7]. Oleh karena itu, penelitian mengenai tingkat pengangguran di Pulau Jawa dengan memperhatikan ketiga aspek yang berpengaruh tersebut menjadi sangat penting untuk dilakukan. Tujuan dari penelitian ini adalah melakukan pengelompokan kabupaten/kota di Pulau Jawa yang didasari oleh karakteristik pengangguran yang dilihat dari ketiga aspek diatas agar program-program dalam menekan angka pengangguran lebih terarah, efektif, dan tepat sasaran.

## 2. METODOLOGI PENELITIAN

### 2.1. Sumber Data

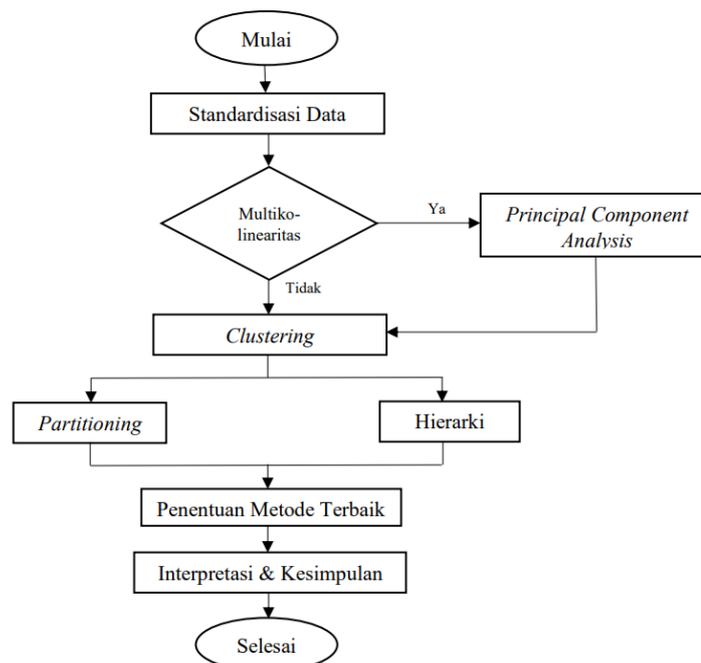
Penelitian ini menggunakan data sekunder yang diperoleh dari *website* Badan Pusat Statistik. Data yang digunakan adalah data tentang pengangguran tahun 2021 yang berjumlah 119 kabupaten/kota di Pulau Jawa. Daftar variabel yang digunakan pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Daftar Variabel Penelitian

Variabel	Keterangan
X <sub>1</sub>	Tingkat Pengangguran Terbuka
X <sub>2</sub>	Jumlah Penduduk

$X_3$	Persentase Penduduk dengan Ijazah Tertinggi SLTA sederajat
$X_4$	Upah Minimum Kab/Kota
$X_5$	Tingkat Partisipasi Angkatan Kerja
$X_6$	Persentase Penduduk Berstatus Kawin
$X_7$	Rasio Ketergantungan

*Software* yang digunakan untuk melakukan pengolahan adalah RStudio dan QGIS. *Software* RStudio digunakan untuk melakukan analisis *cluster*, sedangkan QGIS digunakan untuk membuat peta hasil *cluster*. Tahap-tahap yang dilakukan dalam menghasilkan pengelompokan wilayah dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

## 2.2. Analisis Faktor

Analisis faktor meliputi *principal component analysis* dan *common factor analysis* [8]. Analisis faktor merupakan suatu teknik interpendensi yang bertujuan untuk menjelaskan suatu set variabel berdasarkan dimensi yang lebih umum di mana suatu set variabel tersebut memiliki interkorelasi yang cukup berarti. Analisis faktor menggambarkan pola hubungan antar variabel kemudian dikelompokkan berdasarkan korelasinya sehingga membentuk variabel baru yang tidak terukur [9].

Pada dasarnya *principal component analysis* (PCA) yaitu mereduksi dimensi data multivariat dengan mencari kombinasi linier dari variabel yang menyumbang sebanyak mungkin variasi total dalam data [8]. Penentuan jumlah faktor yang akan

dibentuk tergantung pada kriteria tertentu agar diperoleh faktor yang sesuai. Kriteria pertama yang digunakan sebagai penentu jumlah faktor adalah *eigenvalue* yang artinya banyaknya faktor yang terbentuk tergantung banyaknya nilai *eigenvalue* yang lebih besar atau sama dengan satu [10]. Selain itu dalam menentukan jumlah faktor yang dibentuk juga harus memperhatikan proporsi dari keragaman total data yang dapat dijelaskan harus berkisar antara 80% - 90% [11]. Beberapa asumsi yang harus terpenuhi sebelum melakukan analisis faktor sebagai berikut [12]:

a. Kecukupan sampel

Uji *Kaizer-Meyer-Olkin* (KMO) digunakan untuk melihat kecukupan data sampel yang digunakan di mana ketika nilai KMO lebih besar dari 0,5 maka dapat dilakukan analisis faktor [12].

$H_0$  : Data layak untuk dianalisis

$H_a$  : Data tidak layak untuk dianalisis

Statistik Uji:

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

di mana,

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

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

Keputusan:

$H_0$  akan ditolak jika nilai KMO < 0,5. Sehingga dapat disimpulkan bahwa data tidak layak untuk dilakukan analisis faktor.

b. Tidak terdapat multikolinearitas

Uji *Bartlett* digunakan untuk menguji ada tidaknya multikolinearitas [12].

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

$H_a$  : Matriks korelasi tidak sama dengan matriks identitas ( $R \neq 1$ )

Statistik Uji:

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

di mana,

|R| : nilai determinan dari matriks korelasi

n : banyaknya pengamatan

p : banyaknya variabel

Keputusan:

$H_0$  akan ditolak jika nilai *pvalue* <  $\alpha$ . Sehingga dapat disimpulkan bahwa terdapat hubungan antar variabel (multikolinearitas) maka perlu dilakukan analisis faktor.

### 2.3. Analisis Cluster

Analisis *cluster* merupakan metode eksplorasi data yang digunakan untuk mendapatkan karakteristik yang tersembunyi dengan membentuk kelompok atau gerombol data tanpa adanya informasi apriori berupa label dan mekanisme

pengelompokan (*unsupervised learning*). Analisis *cluster* melakukan proses *learning* hanya berdasarkan informasi, karakteristik, dan fitur dari objek [9]. Dalam *clustering*, ukuran yang menjelaskan kemiripan (similaritas) antar objek (data) menggunakan jarak Euclidean [17].

#### 2.4. Analisis Cluster Metode Partitioning

Metode *partitioning* merupakan salah satu pendekatan yang utama pada analisis *cluster*. Metode ini lebih sederhana, mudah diterapkan, dan memiliki kompleksitas waktu yang lebih kecil dibandingkan teknik lain [18]. Metode *Partitioning* melakukan pengelompokan objek dengan cara memilah-milah objek-objek ke dalam klaster-klaster yang ada [9]. Metode *partitioning* dapat dibagi menjadi dua yaitu *k-clustering* dan *self-determining* [8]. Salah satu metode *k-clustering* adalah K-Means [13].

*K-means clustering* merupakan metode yang sederhana dan paling populer untuk pengelompokan objek ke sejumlah *k* klaster [9]. Secara umum, langkah-langkah dari algoritma K-Means sebagai berikut [14]:

- Tentukan titik pusat dan jumlah klaster (*k*).
- Tentukan titik pusat terdekat dari masing-masing data kemudian masukkan data tersebut ke klaster terdekat.
- Tentukan ulang letak titik pusat berdasarkan anggota klaster yang baru terbentuk.
- Jika belum konvergen, ulangi langkah 2 dan 3.

Kriteria konvergen adalah ketika suatu iterasi tidak ada lagi perubahan anggota pada tiap klaster dibandingkan iterasi sebelumnya, atau nilai *sum of squared error* telah mencapai nilai minimum [14].

#### 2.5. Analisis Cluster Metode Hierarki

*Hierarchical clustering* akan mengelompokkan objek-objek yang mirip pada hierarki yang berdekatan dan objek-objek yang tidak mirip pada hierarki yang berjauhan. Terdapat dua macam strategi untuk melakukan *hierarchical clustering* yaitu *Agglomerative hierarchical clustering (bottom up)* dan *Divisive Hierarchical clustering (top down)* [9]. Perbedaannya terletak pada arah pengelompokan [8].

*Agglomerative hierarchical clustering* terdiri dari beberapa metode, yaitu : [9]

- Single Linkage* : menggunakan jarak terkecil antara satu objek dalam *cluster* dengan satu objek dalam *cluster* yang lain.

$$d(C_i, C_j) = \min \{d(x_{ip}, x_{jq})\}$$

di mana,

$d(C_i, C_j)$  : jarak antara *cluster*  $C_i$  dan *cluster*  $C_j$

$d(x_{ip}, x_{jq})$  : jarak antara tetangga terdekat *cluster*  $i$  dan  $p$ , serta *cluster*  $j$  dan  $q$

- Complete Linkage* : menggunakan jarak terbesar antara satu objek dalam *cluster* dengan satu objek dalam *cluster* yang lain.

$$d(C_i, C_j) = \max \{d(x_{ip}, x_{jq})\}$$

- c. *Average Linkage* : menggunakan jarak rata-rata antar objek-objek dalam satu *cluster* dengan objek-objek dalam *cluster* yang lain.

$$d(C_i, C_j) = \text{avg} \{d(x_{ip}, x_{jq})\}$$

- d. *Ward's Method* : metode ini tidak hanya berdasarkan jarak antar observasi, tetapi juga memperhatikan keragaman dari *cluster* tersebut dengan meminimalkan nilai *Sum of Squared Error* (SSE). Dua *cluster* digabungkan ketika memberikan nilai SSE yang paling kecil/minimal [21].

$$SSE = \sum_{j=1}^N (x_j - \bar{x})'(x_j - \bar{x})$$

di mana:

$x_j$  : nilai atau data dari objek ke-j

$\bar{x}$  : rata-rata nilai dari keseluruhan objek dalam *cluster*

Hasil pengklasteran dengan metode *Hierarchical* dapat divisualisasikan dalam sebuah diagram pohon yang biasa disebut dendogram di mana banyaknya *cluster* yang terbentuk ditentukan dari dendogram yang terjadi [20].

## 2.6. Validasi Cluster

Validasi *cluster* adalah prosedur untuk mengevaluasi hasil analisis *cluster* secara kuantitatif dan objektif [15]. Terdapat tiga pendekatan untuk mengeksplorasi validitas *cluster* yaitu validasi internal, validasi stabilitas, dan validasi biologis [8]. Pada penelitian ini hanya menggunakan validasi internal dan validasi stabilitas dalam memvalidasi hasil *cluster*. Validasi internal terdiri dari beberapa indeks yaitu indeks *Connectivity*, *Silhouette*, dan indeks *Dunn* [15].

- a. Indeks *Connectivity*: digunakan untuk mengukur kepadatan guna mengidentifikasi homogenitas *cluster* yang terlihat dari nilai varians intra-*cluster* atau dikenal dengan tetangga terdekat (*nearest neighbor*) [19]. Indeks *connectivity* memiliki nilai diantara 0 sampai tak hingga. Semakin kecil nilainya berarti semakin baik *cluster* yang terbentuk [8].

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L X_{i,nni(j)}$$

di mana:

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

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

- b. Indeks *Silhouette* : dihitung sebagai derajat kepercayaan dalam proses *clustering* pada suatu pengamatan dengan *cluster* yang terbentuk dikatakan baik bila nilai indeks mendekati 1 dan kondisi sebaliknya jika nilai indeks mendekati angka -1 [15].

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

di mana:

$a(i)$  : jarak rata-rata antara  $i$  dan seluruh pengamatan lainnya pada *cluster* yang sama

$b(i)$  : jarak rata-rata antara  $i$  dengan pengamatan pada *cluster* terdekat.

- c. Indeks *Dunn* : rasio jarak terkecil antara observasi pada *cluster* yang berbeda dengan jarak terbesar pada masing-masing *cluster* data [15]. Semakin tinggi nilai indeks *Dunn*, semakin baik *cluster* yang dihasilkan [8]. *Dunn Index* merupakan fungsi validitas yang mampu memberikan penilaian yang efektif untuk pengaplikasian yang menggunakan beberapa metode *clustering* yang berbeda [16].

$$C = \frac{d_{min}}{d_{max}}$$

di mana:

$d_{min}$  : jarak terkecil antara observasi pada *cluster* yang berbeda.

$d_{max}$  : jarak terbesar pada masing-masing *cluster* data.

Validasi stabilitas digunakan untuk membandingkan hasil dari 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, menunjukkan semakin baik *cluster* yang dihasilkan [8].

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Analisis Deskriptif

Berikut merupakan deskripsi mengenai variabel-variabel yang terkait dengan pengangguran di 119 kabupaten/kota yang terletak di Pulau Jawa berdasarkan ukuran pemusatan dan ukuran penyebarannya.

Tabel 3. Deskripsi Statistik

Variabel	Rata-Rata	Maksimum	Minimum
X <sub>1</sub>	6,862	13,070	2,040
X <sub>2</sub>	1286093	5489536	28240
X <sub>3</sub>	28,39	48,13	11,71
X <sub>4</sub>	2646134	4798312	1805000
X <sub>5</sub>	68,38	80,57	56,86
X <sub>6</sub>	60,91	68,26	50,75
X <sub>7</sub>	44,96	54,73	34,33

Berdasarkan hasil analisis deskriptif yang ditunjukkan pada Tabel 3, terlihat bahwa terdapat perbedaan satuan pengukuran pada tiap variabel sehingga sebelum dilakukan *clustering* perlu dilakukan standardisasi dari ketujuh variabel tersebut terlebih dahulu.

#### 3.2. Analisis Faktor

Sebelum melakukan pengelompokan tiap kabupaten/kota di Pulau Jawa berdasarkan karakteristik penganggurannya, perlu dilakukan pengecekan asumsi terlebih dahulu meliputi uji KMO dan uji *Bartlett*. Uji KMO digunakan untuk menguji apakah data cukup secara statistik. Sedangkan uji *Bartlett* digunakan untuk menguji

mengenai ada tidaknya multikolinearitas. Jika terdapat multikolinearitas maka harus dilakukan analisis faktor terlebih dahulu.

Tabel 4. Nilai KMO dan *Bartlett Test*

Uji	Nilai
<i>Kaiser Meyer Olkin (KMO)</i>	0,7261
<i>Bartlett Test</i>	
<i>Approx. Chi-Square</i>	405,669
<i>Df</i>	21
<i>Sig.</i>	2,22e-16

Berdasarkan Tabel 4 di atas, diperoleh nilai KMO sebesar 0,7261 di mana lebih besar dari 0,5 maka keputusannya gagal tolak  $H_0$  sehingga dapat disimpulkan bahwa data cukup untuk dianalisis lebih lanjut. Kemudian pada uji *Bartlett* terlihat bahwa nilai *p-value* signifikan yang berarti bahwa variabel-variabel tersebut saling berkorelasi atau dapat dikatakan terjadi multikolinearitas sehingga perlu dilakukan *principal component analysis*.

Tabel 5. Eigenvalue

Component	Eigenvalues		
	Total	% of Variance	Cumulative %
1	3,397	48,525	48,525
2	1,567	22,386	70,911
3	0,822	11,749	82,660
4	0,459	6,557	89,217
5	0,331	4,723	93,940
6	0,244	3,479	97,419
7	0,181	2,581	100,000

Berdasarkan Tabel 5, hanya terdapat 2 faktor yang memiliki *eigenvalue* lebih dari satu, tetapi karena *persentase cumulative eigenvalue* dari kedua faktor tersebut belum mencapai 80% - 90% maka jumlah faktor yang digunakan adalah tiga. Hal ini berarti ketiga komponen secara bersama-sama dapat menjelaskan keragaman dari tujuh variabel sebesar 82,66%.

### 3.3. Penentuan Metode *Cluster* Terbaik

Setelah masalah multikolinearitas teratasi dapat dilanjutkan dengan analisis *cluster* untuk melakukan pengelompokan kabupaten/kota di Pulau Jawa berdasarkan karakteristik penganggurannya. Analisis *cluster* diawali dengan melakukan uji validitas *cluster* terlebih dahulu untuk mengetahui jumlah *cluster* terbaik beserta metode yang sesuai dengan data tersebut.

Uji validitas internal menggunakan indeks *Connectivity*, *Sillhouette*, dan *Dunn*. Semakin kecil nilai *Connectivity* semakin baik *cluster* yang didapatkan. Jumlah *cluster* terbaik ditunjukkan dengan nilai indeks *Sillhouette* dan *Dunn* yang semakin mendekati satu. Nilai indeks validitas internal dapat dilihat pada Tabel 6.

Tabel 6. Nilai Indeks Validitas Internal

	Jumlah Cluster	2	3	4	5
K-Means	<i>Connectivity</i>	20,207	21,285	39,209	44,116
	<i>Dunn</i>	0,067	0,115	0,086	0,106
	<i>Sillhoutte</i>	0,419	0,388	0,394	0,386
Hierarki	<i>Connectivity</i>	11,957	17,371	26,155	28,937
	<i>Dunn</i>	0,151	0,098	0,121	0,126
	<i>Sillhoutte</i>	0,392	0,335	0,373	0,371

Tabel 7. Nilai Optimal dari Indeks Validitas Internal

	Score	Method	Clusters
<i>Connectivity</i>	11,9567	Hierarchical	2
<i>Dunn</i>	0,1512	Hierarchical	2
<i>Sillhoutte</i>	0,4190	K-Means	2

Berdasarkan Tabel 7 dapat disimpulkan bahwa merujuk pada nilai indeks validitas internal, metode hierarki dengan jumlah *cluster* 2 yang paling optimal. Hal ini dikarenakan nilai *connectivity*-nya paling kecil dibandingkan metode dengan jumlah *cluster* lainnya.

Setelah melakukan uji validitas internal kemudian dilanjutkan uji validitas stabilitas dengan menggunakan nilai APN, AD, ADM, dan FOM. Nilai indeks validitas stabilitas dapat dilihat pada Tabel 8.

Tabel 8. Nilai Indeks Validitas Stabilitas

	Jumlah Cluster	2	3	4	5
K-Means	APN	0,1784	0,4164	0,3017	0,3669
	AD	2,5615	2,5446	2,1582	2,1080
	ADM	0,7213	1,4481	1,0001	1,1374
	FOM	1,3333	1,3182	1,2833	1,2711
Hierarki	APN	0,1340	0,3233	0,2348	0,2703
	AD	2,7397	2,7038	2,4012	2,2106
	ADM	0,6617	1,3939	1,2231	1,0559
	FOM	1,3249	1,3208	1,3184	1,2983

Tabel 9. Nilai Optimal dari Indeks Validitas Stabilitas

	Score	Method	Clusters
--	-------	--------	----------

APN	0,1340	Hierarchical	2
AD	2,1080	K-Means	5
ADM	0,6617	Hierarchical	2
FOM	1,2711	K-Means	5

Berdasarkan Tabel 9 nilai APN dan ADM menunjukkan metode hierarki dengan jumlah *cluster* 2 adalah metode terbaik. Sedangkan hanya nilai AD dan FOM yang menunjukkan metode K-Means dengan jumlah *cluster* 5 merupakan metode yang terbaik. Berdasarkan nilai indeks validitas internal dan stabilitas dapat disimpulkan bahwa untuk mengelompokkan data kabupaten/kota di Pulau Jawa berdasarkan karakteristik penganggurannya diperoleh metode hierarki dengan jumlah *cluster* 2 merupakan metode terbaik.

### 3.4. Analisis Cluster

Hasil dari uji validitas internal dan stabilitas menunjukkan bahwa metode terbaik untuk kasus ini adalah hierarki dengan jumlah *cluster* 2. Sehingga peneliti menggunakan metode hierarki dengan jumlah *cluster* 2 pada data pengangguran tahun 2021. Metode hierarki terdiri dari *single linkage*, *complete linkage*, *average linkage*, dan Ward.

Tabel 10. *Agglomerative Coefficient Metode Hierarki*

Metode	<i>Agglomerative Coefficient</i>
Single	0,7585746
Complete	0,9221624
Average	0,8568878
<b>Ward</b>	<b>0,9728054</b>

Untuk dapat menentukan metode hierarki yang menghasilkan *cluster* terbaik dapat dilakukan dengan melihat nilai dari *agglomerative coefficient*. Metode dengan nilai *agglomerative coefficient* tertinggi merupakan metode terbaik. Berdasarkan Tabel 10 diperoleh bahwa metode hierarki yang menghasilkan *cluster* terbaik yaitu metode Ward karena nilai *agglomerative coefficient* lebih besar dibandingkan metode hierarki lainnya.

Berikut hasil pengelompokan kabupaten/kota di Pulau Jawa berdasarkan karakteristik penganggurannya menggunakan metode Ward. Kelompok pertama terdiri dari 49 kabupaten/kota, yaitu:

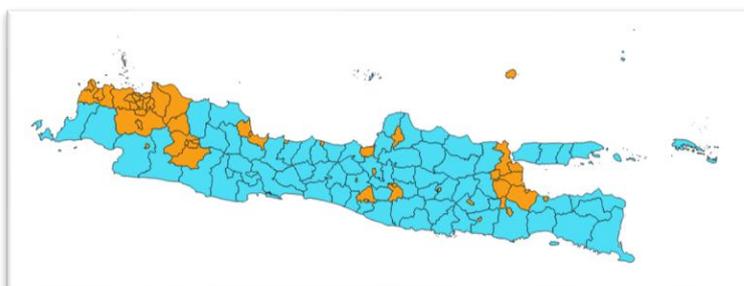
- 1) *Provinsi DKI Jakarta* : Kab. Kepulauan Seribu, Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, dan Kota Jakarta Utara.

- 2) *Provinsi Jawa Barat* : Kab. Bogor, Kab. Bandung, Kab. Cirebon, Kab. Purwakarta, Kab. Karawang, Kab. Bekasi, Kab. Bandung Barat, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, dan Kota Cimahi.
- 3) *Provinsi Jawa Tengah* : Kab. Sukoharjo, Kab. Kudus, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal.
- 4) *Provinsi Daerah Istimewa Yogyakarta* : Kab. Sleman, dan Kota Yogyakarta.
- 5) *Provinsi Jawa Timur* : Kab. Pasuruan, Kab. Sidoarjo, Kab. Mojokerto, Kab. Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, dan Kota Batu.
- 6) *Provinsi Banten* : Kab. Tangerang, Kab. Serang, Kota Tangerang, Kota Cilegon, Kota Serang, dan Kota Tangerang Selatan.

Sedangkan kelompok kedua terdiri dari 70 kabupaten/kota, yaitu:

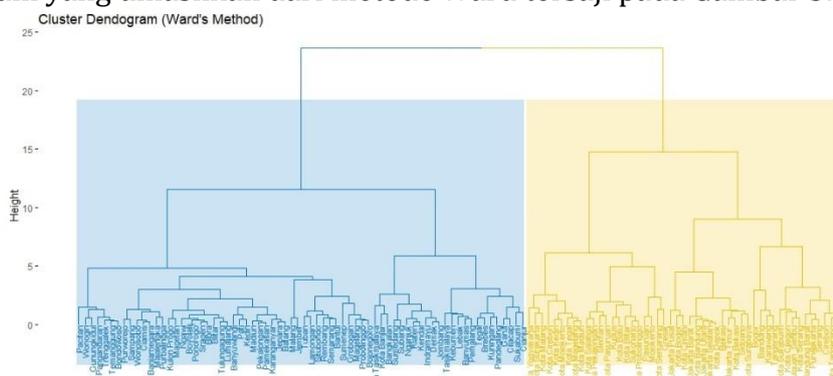
- 1) *Provinsi Jawa Barat* : Kab. Sukabumi, Kab. Cianjur, Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Kuningan, Kab. Majalengka, Kab. Sumedang, Kab. Indramayu, Kab. Subang, Kab. Pangandaran, Kota Tasikmalaya, dan Kota Banjar.
- 2) *Provinsi Jawa Tengah* : Kab. Cilacap, Kab. Banyumas, Kab. Purbalingga, Kab. Kebumen, Kab. Banjarnegara, Kab. Purworejo, Kab. Wonosobo, Kab. Magelang, Kab. Boyolali, Kab. Klaten, Kab. Wonogiri, Kab. Karanganyar, Kab. Sragen, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Pati, Kab. Jepara, Kab. Demak, Kab. Semarang, Kab. Temanggung, Kab. Kendal, Kab. Batang, Kab. Pekalongan, Kab. Pemalang, Kab. Tegal, dan Kab. Brebes.
- 3) *Provinsi Daerah Istimewa Yogyakarta* : Kab. Kulon Progo, Kab. Bantul, dan Kab. Gunungkidul.
- 4) *Provinsi Jawa Timur* : Kab. Pacitan, Kab. Ponorogo, Kab. Trenggalek, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi, Kab. Bondowoso, Kab. Situbondo, Kab. Probolinggo, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kab. Ngawi, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan, dan Kab. Sumenep.
- 5) *Provinsi Banten* : Kab. Pandeglang dan Kab. Lebak.

Penyajian kabupaten/kota di Pulau Jawa hasil *clustering* tersaji dalam bentuk peta yang ditunjukkan pada Gambar 2.



Gambar 2. Peta Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Karakteristik Pengangguran

Dendrogram yang dihasilkan dari metode Ward tersaji pada Gambar 3.



Gambar 3. Dendrogram Metode Ward

### 3.5. Karakteristik Tiap Cluster

Berikut karakteristik dari masing-masing *cluster* yang terbentuk berdasarkan variabel-variabel yang terkait dengan pengangguran di Pulau Jawa tahun 2021.

Tabel 11. Deskripsi Statistik Tiap Cluster

Variabel	Cluster 1	Cluster 2
Tingkat Pengangguran Terbuka	0,7024	0,4917
Jumlah Penduduk	0,0987	-0,0691
Persentase Penduduk dengan Ijazah Tertinggi SLTA sederajat	0,9421	-0,6595
Upah Minimum Kab/Kota	0,7832	-0,5482
Tingkat Partisipasi Angkatan Kerja	-0,5243	0,3670
Persentase Penduduk Berstatus Kawin	-0,8377	0,5864
Rasio Ketergantungan	-0,6329	0,4430

Berdasarkan Tabel 11, terlihat bahwa *cluster* pertama yang terdiri dari 49 kabupaten/kota di Pulau Jawa memiliki rata-rata yang lebih tinggi pada variabel TPT, jumlah penduduk, persentase penduduk dengan ijazah terakhir SLTA sederajat, dan UMK. Sedangkan *cluster* kedua terdiri dari 70 kabupaten/kota di Pulau Jawa memiliki rata-rata yang lebih tinggi pada variabel TPAK, persentase penduduk berstatus kawin, dan rasio ketergantungan.

## 4. SIMPULAN

Berdasarkan hasil validasi internal dan validasi stabilitas didapat metode terbaik yaitu hierarki dengan jumlah *cluster* 2 (dua). Kemudian berdasarkan hasil dari *Agglomerative Coefficient* pada metode hierarki diperoleh metode Ward yang

terbaik. Hasil akhir diperoleh untuk *cluster* pertama berjumlah 49 kabupaten/kota sedangkan *cluster* kedua berjumlah 70 kabupaten/kota. Karakteristik *cluster* pertama yaitu memiliki tingkat pengangguran yang lebih tinggi.

Dengan mengetahui karakteristik pengangguran tiap kabupaten/kota di Pulau Jawa, pemerintah terkait dapat dengan tepat dalam membuat kebijakan untuk mengatasi masalah pengangguran. Saran untuk penelitian berikutnya akan lebih baik jika menambahkan metode *clustering* dan variabel yang terkait lainnya.

#### DAFTAR PUSTAKA

- [1] BPS, *Hasil Sensus Penduduk 2020*, Jakarta: BPS, 2021.
- [2] Dirjen Kependudukan dan Pencatatan Sipil. (2022, Feb 24). 273 Juta Penduduk Indonesia Terupdate Versi Kemendagri [Online]. Available: <https://dukcapil.kemendagri.go.id/berita/baca/1032/273-juta-penduduk-indonesia-terupdate-versi-kemendagri>
- [3] S. Sukirno, *Teori Pengantar Makro Ekonomi Edisi Ketiga*, Jakarta: PT. Raja Grafindo Pesada, 2010.
- [4] BPS, *Booklet Sakernas Agustus 2021*, Jakarta: BPS, 2021.
- [5] E. Amalia, "Analisis Spasial untuk Mengidentifikasi Tingkat Pengangguran Terbuka Berdasarkan Kabupaten/Kota di Pulau Jawa Tahun 2017", Skripsi, Jakarta : Politeknik Statistika STIS, 2018.
- [6] E.N. Gunawan, " Determinan Tingkat Pengangguran Terbuka Menurut Kabupaten/Kota di Pulau Jawa Tahun 2019 menggunakan Regresi Spasial", Skripsi, Jakarta : Politeknik Statistika STIS, 2021.
- [7] W. I. Astuti, "Faktor-Faktor yang Berpengaruh terhadap Tingkat Pengangguran Terbuka di Jawa Timur dengan Pendekatan Regresi Data Panel", Skripsi, Surabaya: Institut Teknologi Sepuluh November, 2017.
- [8] N. Afira and A. W. Wijayanto, "Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki", *Jurnal Sistem Komputer*, vol. 10, no. 2, pp.101-109, 2021.
- [9] S. Pramana, et al., *Data Mining dengan R: Konsep dan Implementasi*, Jakarta: In Media, 2018.
- [10] S. Wahyuni and Y. A. Jatmiko, "Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering", *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 10, no. 1, 2018.
- [11] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, USA: Pearson Education, Inc , 2007.

- [12] S. Machfudhoh and N. Wahyuningsih, "Analisis Cluster Kabupaten/Kota Berdasarkan Pertumbuhan Ekonomi Jawa Timur," *Jurnal Sains dan Pomits*, vol. 2, no. 1, pp. 1-7, 2013.
- [13] D. Zhang, K. Lee, K., & I. Lee, "Hierarchical trajectory clustering for spatio-temporal periodic pattern mining". *Expert Systems with Applications*, no. 92, pp. 1-11, 2018.
- [14] A. K. Jain, M. N. Nurty, and P. J. Flynn, "Data Clustering: A Review", *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264-323, 1999.
- [15] N. N. Halim and E. Widodo, "Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps", in *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai-Nilai Islami)*, vol. 1, no. 1, pp. 188-194, 2017.
- [16] M. D. Simatupang and A. W. Wijayanto, "Analisis Klaster berdasarkan Tindakan Kriminalitas di Indonesia Tahun 2019", *Jurnal Statistika Industri dan Komputasi*, vol. 6, no. 1, pp. 10-19, 2021.
- [17] N. Thamrin and A. W. Wijayanto, "Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island", *Indonesian Journal of Statistics and Its Applications*, vol. 5, no. 1, pp. 141-160, 2021.
- [18] A. M. Sikana and A. W. Wijayanto, "Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering", *Jurnal Ilmu Komputer*, vol. 14, no. 2, pp. 66-78, 2021.
- [19] A. Azzahra and A. W. Wijayanto, "Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal", *SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 2, pp. 481-495, 2022.
- [20] E. Luthfi and A. W. Wijayanto, "Analisis perbandingan metode hierarchal, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia", *INOVASI*, vol. 17, no. 4, pp. 761-773, 2021.
- [21] A. R. Damayanti and A. W. Wijayanto, "Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018. *Eigen Mathematics Journal*, vol. 4, no. 1, pp. 8-17, 2021.