

# Pemetaan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Analisis *Cluster Hierarki*

Indria Wahyuni dan Sri Pingit Wulandari

Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember (ITS)

*e-mail*: sripingitwulandari@gmail.com

**Abstrak**—Kesejahteraan rakyat sangat dibutuhkan untuk masukan penyusunan kebijakan dan sebagai alat untuk melihat keadaan, memonitor, dan mengevaluasi keberhasilan pembangunan di Jawa Timur sesuai dengan kesepakatan *SDG's* untuk itu dibutuhkan informasi lebih banyak tentang kesejahteraan rakyat. Salah satunya dengan mengelompokkan Kabupaten/Kota yang kesejahteraan rakyatnya setipe dengan menggunakan analisis *cluster*. Analisis *cluster* merupakan salah satu metode statistika untuk mengelompokkan observasi ke dalam suatu kelas yang memiliki karakteristik yang sama. Berdasarkan hasil analisis *cluster hierarki*, pengelompokan Kabupaten/Kota di Jawa Timur menggunakan indikator kesejahteraan rakyat diperoleh 4 pembagian kelompok, tingkat kesejahteraan rakyat yang sangat tinggi, tinggi, sedang dan rendah. Kelompok 1 memiliki tingkat kesejahteraan rakyat sedang, kelompok 2 memiliki tingkat kesejahteraan rakyat rendah, kelompok 3 memiliki tingkat kesejahteraan rakyat tinggi dan kelompok 4 memiliki tingkat kesejahteraan rakyat sangat tinggi. Kelompok 2 memiliki tingkat kesejahteraan rakyat rendah karena memiliki kesetaraan yang kurang baik dalam hal ekonomi, kesehatan dan pendidikan.

**Kata Kunci**—Analisis Cluster, Kesejahteraan Rakyat, Metode Hierarki, Pengelompokan.

## I. PENDAHULUAN

SECARA umum kesejahteraan dapat diartikan sebagai suatu keadaan dimana segenap warga negara selalu berada dalam kondisi serba kecukupan segala kebutuhannya, baik material maupun spiritual [1]. Sejalan dengan adanya pembangunan nasional, *United Nation* atau perserikatan bangsa-bangsa mendeklarasikan kesepakatan *SDG's*. *SDG's* (*Sustainable Development Gols*) merupakan agenda pembangunan dunia yang bertujuan untuk kesejahteraan manusia secara global. Agenda tersebut merupakan program pembangunan berkelanjutan dimana terdapat 17 tujuan. Sehingga pada intinya pembangunan bertujuan untuk menciptakan lingkungan yang memungkinkan rakyat menjalankan kehidupan produktif untuk meningkatkan kesejahteraan rakyat. Terlangsir dari media Indonesia bahwa tingkat kesejahteraan masyarakat Indonesia meningkat dari posisi ke-14 menjadi jajaran lima besar pada tahun 2019.

Analisis *cluster* merupakan salah satu metode statistika dalam mengelompokkan observasi ke dalam suatu kelas yang memiliki karakteristik yang sama. Dua metode paling umum dalam analisis *cluster* adalah metode *hierarki* dan metode non *hierarki*. Penelitian ini menggunakan metode *cluster hierarki* karena jumlah *cluster* tidak ditentukan terlebih dahulu namun berdasarkan hasil analisis. Analisis *cluster hierarki* pada penelitian ini membandingkan empat metode yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward's method*. Hasil pada metode Pengelompokan *cluster hierarki*

dapat disajikan dalam sebuah diagram yang disebut *dendrogram*. *Dendrogram* digunakan untuk memudahkan dalam Pengelompokan objek-objek, karena gambar yang disajikan lebih informatif [2].

Penelitian sebelumnya terkait kesejahteraan rakyat pernah dilakukan oleh Sajidah, dengan penelitian mengenai Pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat menggunakan metode *C-means* dan *C-means clustering* yang menghasilkan kesimpulan diantaranya adalah faktor kesehatan, pendidikan dan ekonomi [3]. Selain itu penelitian yang berkaitan dengan metode analisis *cluster* adalah penelitian mengenai Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan indikator kemiskinan dengan menggunakan analisis *cluster hierarki* dengan hasil akhir terbentuk 4 kelompok dan metode terbaiknya metode *ward's method* [4]. Penelitian serupa lainnya dilakukan oleh Ghaisani mengenai analisis *cluster hierarki* untuk pengelompokan provinsi di Indonesia berdasarkan indikator demokrasi Indonesia tahun 2016 dengan hasil akhir terbentuk 3 kelompok dan metode terbaiknya *cluster centroid* [5]. Selanjutnya penelitian lainnya dilakukan oleh Fitri mengenai Pengelompokan dan pemetaan wilayah Kecamatan di Kabupaten Ponorogo berdasarkan potensi sektor pertanian menggunakan analisis *cluster hierarki* dengan hasil akhir terbentuk 3 kelompok dan metode terbaiknya adalah *average linkage* [6].

Berdasarkan uraian dari latar belakang yang dijelaskan sebelumnya, penelitian ini digunakan untuk mengetahui bagaimana hasil pengelompokan dan pemetaan indikator kesejahteraan rakyat di Jawa Timur berdasarkan Kabupaten/Kota. Sehingga tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui hasil pengelompokan dan pemetaan indikator kesejahteraan rakyat di Jawa Timur berdasarkan Kabupaten/kota.

Manfaat dari penelitian ini adalah memberikan informasi mengenai pengelompokan Kabupaten/Kota yang setipe atau sejenis untuk menjadi satu kelompok. Batasan masalah pada penelitian ini menggunakan indikator kesejahteraan rakyat dengan tiga dari 17 indikator *SGD's* yaitu ekonomi, pendidikan dan kesehatan.

## II. TINJAUAN PUSTAKA

### A. Metode Analisis Cluster Hierarki

Analisis *cluster* merupakan salah satu metode statistika dalam mengelompokkan observasi ke dalam suatu kelas yang memiliki karakteristik yang sama. Dua metode paling umum dalam analisis *cluster* adalah metode *hierarki* dan metode non *hierarki*. Penentuan metode mana yang akan dipakai

Tabel 1.  
Variabel penelitian

Variabel	Keterangan	Skala
<b>Indikator Ekonomi</b>		
$X_1$	Persentase Tingkat Pengangguran Terbuka	Rasio
$X_2$	Persentase Tingkat Partisipasi Angkatan Kerja	Rasio
$X_3$	Persentase Penduduk Miskin	Rasio
<b>Indikator Kesehatan</b>		
$X_4$	Persentase Rumah Tangga yang Memiliki Air Minum Layak	Rasio
$X_5$	Persentase Rumah Tangga Sanitasi Layak	Rasio
$X_6$	Persentase Keluhan Kesehatan	Rasio
<b>Indikator Pendidikan</b>		
$X_7$	Persentase Angka Partisipasi Sekola Berusia 5-10 Tahun	Rasio
$X_8$	Rata-rata Lama Sekolah	Rasio

Tabel 2.  
Struktur data

Kab/Kota	Variabel				
	$X_1$	$X_2$	$X_3$	...	$X_8$
1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	$X_{1,8}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	$X_{2,8}$
⋮	⋮	⋮	⋮	⋮	⋮
38	$X_{38,1}$	$X_{38,2}$	$X_{38,3}$	...	$X_{38,8}$

tergantung kepada peneliti dan konteks penelitian dengan tidak mengabaikan substansi, teori dan konsep yang berlaku. Keduanya memiliki kelebihan sendiri-sendiri. Penelitian ini menggunakan metode *cluster hierarki* karena jumlah *cluster* tidak ditentukan terlebih dahulu namun berdasarkan hasil analisis. Hasil pada metode Pengelompokan *cluster hierarki* dapat disajikan dalam sebuah diagram yang disebut *dendrogram*. *Dendrogram* digunakan untuk memudahkan dalam Pengelompokan objek-objek, karena gambar yang disajikan lebih informatif. Keuntungan metode *hierarki* adalah cepat dalam proses pengolahan sehingga menghemat waktu [2]. Sebelum melakukan *clustering*, terlebih dahulu dilakukan fungsi jarak *Euclidean* sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{1}$$

Dimana:

$d_{ij}$  = Jarak antara objek ke-i dan objek ke-j,

$i, j$  = 1, 2, ..., n

$k$  = 1, 2, ..., p

$x_{ik}$  = Nilai pengamatan objek ke-i variabel ke-k

$x_{jk}$  = Nilai pengamatan objek ke-j variabel ke-k

Berikut merupakan 4 metode dalam analisis *cluster hierarki*, diantaranya:

1) *Single Linkage*

Metode ini didasarkan pada jarak minimum. Dimulai dengan dua objek yang dipisahkan dengan jarak paling pendek maka keduanya akan ditempatkan pada *cluster* pertama, dan seterusnya. Metode ini dikenal pula dengan nama pendekatan tetangga terdekat [2]

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \tag{2}$$

2) *Complete Linkage*

Disebut juga pendekatan tetangga terjauh. Dasarnya adalah jarak maksimum. Dalam metode ini seluruh objek dalam suatu *cluster* dikaitkan satu sama lain pada suatu jarak maksimum atau dengan kesamaan minimum.

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \tag{3}$$

Tabel 3.  
Nilai *Pseudo F-Statistic*

Variabel	Jumlah Kelompok	<i>Pseudo F</i>
<i>Single Linkage</i>	2 Kelompok	1,526
	3 Kelompok	1,850
	<b>4 Kelompok</b>	<b>1,983</b>
<i>Complete Linkage</i>	2 Kelompok	27,863
	3 Kelompok	24,141
	<b>4 Kelompok</b>	<b>31,306</b>
<i>Average Linkage</i>	2 Kelompok	23,009
	<b>3 Kelompok</b>	<b>33,393</b>
	4 Kelompok	33,088
<i>Ward's Method</i>	2 Kelompok	23,009
	<b>3 Kelompok</b>	<b>33,393</b>
	4 Kelompok	33,088

Tabel 4.  
Pemilihan metode terbaik

Metode	Jumlah Kelompok	<i>ICDRate</i>
<i>Single Linkage</i>	4 Kelompok	0,851
<b><i>Complete Linkage</i></b>	<b>4 Kelompok</b>	<b>0,265</b>
<i>Average Linkage</i>	3 Kelompok	0,343
<i>Ward's Method</i>	3 Kelompok	0,343

3) *Average Linkage*

Dasarnya adalah jarak rata-rata antar observasi. Pengelompokan dimulai dari tangan atau pasangan observasi dengan jarak paling mendekati jarak rata-rata.

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ij}}{N_{(UV)} N_W} \tag{4}$$

Dimana  $d_{ik}$  adalah jarak antara objek  $i$  di dalam *cluster* ( $UV$ ) dan objek  $k$  di dalam *cluster*  $W$ , dan  $N_{(UV)}$  dan  $N_W$  adalah nomor item dari *cluster* ( $UV$ ) dan  $W$ , berturut-turut.

4) *Ward's Method*

Pada metode ini, jarak antara dua *cluster* yang terbentuk adalah *error sum of squares (ESS)* di antara dua *cluster* tersebut. Dua objek akan digabungkan jika mempunyai fungsi objektif terkecil diantara kemungkinan yang ada. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean cluster* untuk setiap pengamatan. Jika *cluster* sebanyak  $k$  maka *ESS* adalah sebagai jumlahan dari  $ESS_k$  atau  $ESS = ESS_1 + ESS_2 + \dots + ESS_k$ . Ketika semua *cluster* digabungkan dalam satu kelompok item  $N$ , maka dapat dilihat sebagai berikut [2]:

$$ESS = \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x}) \tag{5}$$

Dimana :

$x_j$  = Vektor objek ke-j

$\bar{x}$  = Rata-rata semua objek

$n$  = Banyaknya objek

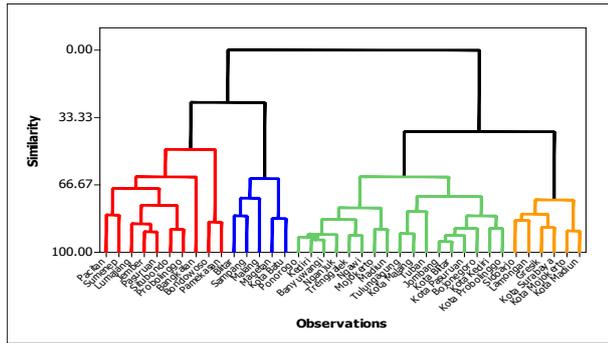
B. *Pseudo F-Statistic*

*Pseudo F-Statistic* merupakan metode statistika yang digunakan untuk mengetahui jumlah kelompok optimum. Nilai *Pseudo F* tertinggi menunjukkan bahwa kelompok tersebut memberikan hasil yang optimal, dimana keragaman dalam kelompok sangat homogen sedangkan antar kelompok heterogen [7]. Persamaan yang digunakan untuk menghitung nilai *Pseudo F-statistic* diformulasikan sebagai berikut:

$$PseudoF = \frac{\frac{R^2}{c-1}}{\frac{1-R^2}{n-c}} \tag{6}$$

Tabel 5. Pengelompokan *Complete Linkage*

Cluster	Jumlah	Kabupaten/Kota
1	10	Pacitan, Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bangkalan, Pamekasan, Sumenep, Ponorogo, Trenggalek, Tulungagung, Kediri, Banyuwangi, Mojokerto, Jombang,
2	17	Nganjuk, Madiun, Ngawi, Bojonegoro, Tuban, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo dan Kota Pasuruan.
3	5	Blitar, Malang, Magetan, Sampang dan Kota Batu.
4	6	Sidoarjo, Kota Mojokerto, Lamongan, Gresik, Kota Madiun dan Kota Surabaya.



Gambar 1. Pengelompokan *Complete Linkage*.

Dimana

$$R^2 = \frac{SST - SSW}{SST} \tag{7}$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_k)^2 \tag{8}$$

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \tag{9}$$

Keterangan:

$SST$  = Total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan

$SSW$  = Total jumlah kuadrat jarak sampel terhadap rata-rata kelompoknya

$n$  = Banyaknya sampel

$c$  = Banyaknya kelompok

$n_c$  = Banyaknya data pada kelompok  $n_c$

$p$  = Banyaknya variabel

$x_{ijk}$  = Sampel ke- $i$  pada kelompok ke- $j$  dan variabel ke- $k$

$\bar{x}_k$  = Rata-rata seluruh sampel pada variabel ke- $k$

$\bar{x}_{jk}$  = Rata-rata ke- $j$  pada variabel ke- $k$

$R^2$  digunakan untuk mengukur perbedaan antar kelompok. Nilai  $R^2$  memiliki rentang dari 0 hingga 1, apabila nilai  $R^2$  sebesar 0 menunjukkan bahwa tidak ada perbedaan di antara kelompok, sedangkan apabila nilai  $R^2$  sebesar 1 menunjukkan bahwa terdapat perbedaan di antara kelompok [8].

C. *Internal Cluster Dispersion Rate (ICDRate)*

*ICDRate (Internal Cluster Dispersion Rate)* merupakan tingkat dispersi dalam *cluster* yang digunakan untuk mengevaluasi hasil *clustering* sehingga didapatkan kriteria *cluster* terbaik. Nilai *ICDRate* didapatkan dari persentase rata-rata klasifikasi yang benar (*recovery rate*) dan tingkat dispersi *cluster* internal partisi terakhir. Nilai *ICDRate* yang semakin kecil menunjukkan bahwa *cluster* tersebut semakin baik sebab antara anggota dalam satu *cluster* memiliki perbedaan yang rendah/memiliki variasi yang kecil [9],

Tabel 6. Deskripsi kelompok baru yang terbentuk

Indikator	Variabel	Cluster			
		1	2	3	4
Ekonomi	$X_1$	Tinggi	Rendah	Sangat Rendah	Sangat Tinggi
	$X_2$	Tinggi	Rendah	Sangat Tinggi	Rendah
	$X_3$	Tinggi	Rendah	Rendah	Sangat Rendah
Kesehatan	$X_4$	Tinggi	Rendah	Sangat Tinggi	Rendah
	$X_5$	Sangat Rendah	Tinggi	Rendah	Sangat Tinggi
	$X_6$	Rendah	Tinggi	Sangat Tinggi	Sangat Rendah
Pendidikan	$X_7$	Sangat Rendah	Rendah	Tinggi	Sangat Tinggi
	$X_8$	Tinggi	Sangat Rendah	Rendah	Sangat Tinggi

Keterangan:  
 Kelompok 1: Tingkat Kesejahteraan Rakyat Sedang  
 Kelompok 2: Tingkat Kesejahteraan Rakyat Rendah  
 Kelompok 3: Tingkat Kesejahteraan Rakyat Tinggi  
 Kelompok 4: Tingkat Kesejahteraan Rakyat Sangat Tinggi



Gambar 2. Pemetaan *Complete Linkage*.

sehingga dapat diperoleh berdasarkan Persamaan (10) sebagai berikut, dimana  $SST$  sesuai dengan Persamaan (8).

$$ICDRate = 1 - \frac{SSB}{SST} = 1 - R^2 \tag{10}$$

Dimana

$$SSB = \sum_{i=1}^c \sum_{k=1}^p (\bar{x}_{ik} - \bar{x}_k)^2 \tag{11}$$

Keterangan :

$SSB$  = Total nilai jarak antar pusat *cluster*

$c$  = Banyaknya kelompok

$p$  = Banyaknya variabel

$\bar{x}_{ik}$  = Rata-rata sampel pada variabel ke- $i$  dan kelompok ke- $k$

$\bar{x}_k$  = Rata-rata seluruh sampel pada variabel ke- $k$

D. *Kesejahteraan Rakyat*

Untuk memberikan suatu gambaran mengenai kesejahteraan rakyat maka diperlukan indikator-indikator yang dapat digunakan sebagai acuan. Berikut merupakan indikator kesejahteraan rakyat:

1) *Indikator Ekonomi*

Tingkat Pengangguran Terbuka (TPT) adalah indikasi tentang penduduk kerja yang termasuk dalam kelompok pengangguran. Tingkat pengangguran terbuka diukur sebagai persentase jumlah penganggur/pencari kerja terhadap jumlah angkatan kerja. Indikator pengangguran terbuka berguna sebagai acuan pemerintah bagi pembukaan lapangan kerja

baru, dimana angka yang tinggi menunjukkan banyaknya tingkat pengangguran terbuka yang artinya semakin kecil angka persentase tingkat pengangguran terbuka maka akan semakin baik.

Tingkat Partisipasi Angkatan Kerja (TPAK) adalah perbandingan antara angkatan kerja dengan jumlah seluruh penduduk usia kerja. TPAK mengukur besarnya partisipasi angkatan kerja dalam dunia kerja. Angka persentase yang rendah menunjukkan kecilnya kesempatan kerja yang tersedia bagi penduduk usia kerja, sedangkan angka yang tinggi menunjukkan besarnya kesempatan kerja yang tersedia.

Kemiskinan yang didefinisikan sebagai permasalahan ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan. Angka persentase yang rendah menunjukkan bahwa penduduk miskin di daerah tersebut sangat rendah, artinya jika persentase rendah maka semakin baik.

### 2) Indikator Kesehatan

Rumah Tangga yang Memiliki Air Minum Layak (AML) adalah perbandingan antara jumlah rumah tangga yang memiliki akses terhadap sumber air minum layak terhadap jumlah rumah tangga., dinyatakan dalam persen (%). Selain mempertimbangkan sumber air minum, perhitungan akses terhadap air minum layak yang digunakan BPS juga mempertimbangkan sumber air yang digunakan untuk memasak/mandi/mencuci dll. Angka persentase Rumah Tangga yang Memiliki Air Minum Layak yang tinggi menunjukkan bahwa semakin tinggi persentase AML maka semakin baik.

Rumah Tangga Sanitasi Layak (SL) adalah fasilitas sanitasi yang memenuhi syarat kesehatan yaitu fasilitas tersebut digunakan oleh rumah tangga sendiri atau bersama dengan rumah tangga lain tertentu, dilengkapi dengan kloset jenis leher angsa, serta tempat pembuangan akhir atau instalasi pengolahan air limbah. Angka persentase Rumah Tangga Sanitasi Layak yang tinggi menunjukkan bahwa semakin tinggi persentase SL maka semakin baik.

Kesehatan merupakan hak setiap orang agar dapat merasa nyaman dalam kehidupan berrakyat. Indikator tingkat kesehatan dapat dilihat dari rata-rata hari sakit maupun ketersediaannya fasilitas kesehatan. Angka persentase keluhan kesehatan yang rendah menunjukkan bahwa semakin rendah persentase TTK maka semakin baik.

### 3) Indikator Pendidikan

Angka Partisipasi Sekolah (APS), yaitu angka yang menyatakan daya tampung sarana pendidikan dalam memfasilitasi penduduk berdasarkan usia tertentu untuk menikmati pendidikan. APS menunjukkan seberapa besar penduduk di suatu wilayah dapat merasakan fasilitas pendidikan. Pada penelitian ini menggunakan APS 5 tahun sampai 10 tahun. Semakin tinggi Angka Partisipasi Sekolah semakin besar jumlah penduduk yang berkesempatan mengenyam pendidikan (semakin baik). Namun demikian meningkatnya APS tidak selalu dapat diartikan bahwa pemerataan kesempatan rakyat untuk mengenyam pendidikan sudah baik.

Rata-rata Lama Sekolah (RSL) adalah jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan. RLS dapat digunakan untuk mengetahui kualitas pendidikan

rakyat dalam suatu wilayah. Penduduk tamat SD diperhitungkan lama sekolah selama 6 tahun, tamat SMP diperhitungkan lama sekolah selama 9 tahun, tamat SMA diperhitungkan lama sekolah selama 12 tahun tanpa memperhitungkan apakah pernah tinggal kelas atau tidak. Angka persentase yang tinggi menunjukkan bahwa semakin baik rata-rata lama sekolah.

## III. METODOLOGI PENELITIAN

### A. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder. Data diperoleh dari buku publikasi Badan Pusat Statistik Provinsi Jawa Timur dengan Judul "Provinsi Jawa Timur dalam Angka 2020 [3], data diambil pada hari Sabtu, 26 Desember 2020. Data yang digunakan pada masing-masing variabel sebanyak 38 Kabupaten/Kota di Jawa Timur, terdiri dari 9 Kota dan 29 Kabupaten.

### B. Variabel Penelitian

Variabel yang digunakan adalah indikator kesejahteraan rakyat di Jawa Timur yang dapat dilihat pada Tabel 1.

### C. Struktur Data

Struktur data yang digunakan pada penelitian ini dapat dilihat pada Tabel 2.

### D. Langkah-langkah Analisis

Langkah-langkah yang digunakan dalam menganalisis data penelitian ini adalah sebagai berikut:

1. Medeskripsikan kondisi pada indikator kesejahteraan rakyat tahun 2019 menggunakan analisis statistika deskriptif.
2. Melakukan analisis *cluster* untuk mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan indikator kesejahteraan rakyat dengan menggunakan *cluster hierarki*.
3. Melakukan pemilihan jumlah *cluster* yang paling optimal dengan *Pseudo Fstatistic*.
4. Membandingkan nilai *Internal Cluster Dispersion Rate (ICDRate)* antara *cluster* optimum yang terbentuk pada keempat metode *cluster hierarki*.
5. Melakukan pemetaan wilayah Kabupaten/Kota di Jawa Timur berdasarkan hasil pengelompokan

## IV. ANALISIS DAN PEMBAHASAN

Analisis yang digunakan untuk mengelompokkan Kabupaten/Kota di Jawa Timur berdasarkan Kesejahteraan Rakyat adalah *Cluster Hierarki*. Analisis *Cluster Hierarki* adalah salah satu metode pengelompokan dimana jumlah kelompok yang dibuat belum diketahui. Analisis ini menggunakan empat metode *Cluster Hierarki* yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward's Method*.

Pada penelitian ini, jumlah pengelompokan dalam metode hierarki menggunakan perkiraan jumlah kelompok 2, 3 dan 4 kelompok dan pemilihan banyaknya kelompok yang optimum dari masing-masing metode *Cluster Hierarki* dapat dilihat berdasarkan nilai *Pseudo-F* yang tertinggi dan *ICDRate* terkecil. Dihitung berdasarkan Persamaan (6) nilai *Pseudo-F* dapat dilihat pada Tabel 3.

Tabel 3 menunjukkan bahwa nilai *Pseudo-F* yang

digunakan untuk menentukan jumlah kelompok berdasarkan nilai tertinggi. Metode *Single Linkage* dan *Complete Linkage* sebanyak 4 kelompok, *Average Linkage* dan *Ward's Method* memiliki kelompok terbaik sebanyak 3 kelompok. Pengelompokan dengan metode *Cluster Hierarki* dilakukan dengan menggunakan proses *agglomerasi*.

Setelah diketahui jumlah kelompok terbaik pada masing-masing metode *Cluster Hierarki*, maka dilakukan pemilihan metode terbaik berdasarkan tingkat penyebaran internal dalam kelompok atau *internal cluster dispersion rate (Icdrate)*. Nilai *Icdrate* yang semakin kecil menunjukkan semakin baik dari hasil pengelompokan yang dilakukan oleh masing-masing metode, sebab anggota dalam satu *cluster* memiliki perbedaan yang rendah/variabel yang kecil. Hasil pemilihan terbaik dihitung berdasarkan Persamaan (10) dapat dilihat pada Tabel 4.

Tabel 4 menunjukkan hasil nilai tingkat penyebaran internal dalam kelompok untuk 4 metode *Cluster Hierarki*. Metode pengelompokan yang terbaik adalah dengan tingkat penyebaran internal dalam kelompok yang terkecil. Berdasarkan Tabel 4 dapat diketahui bahwa metode yang menghasilkan pengelompokan yang terbaik dengan nilai *Icdrate* yang terkecil sebesar 0,265 adalah metode *Complete Linkage*. Sehingga dapat disimpulkan kelompok optimum dari empat metode sebanyak 4 kelompok.

Analisis *Cluster Hierarki* dengan metode *Complete Linkage* dilakukan dengan jarak terjauh dalam kelompok. Jumlah kelompok yang optimum berdasarkan Tabel 3 adalah sebanyak 4 kelompok, dimana kelompok 1 terdiri dari 10 Kabupaten/Kota, kelompok 2 terdiri dari 17 Kabupaten/Kota, kelompok 3 terdiri dari 5 Kabupaten/Kota dan kelompok 4 terdiri dari 6 Kabupaten/Kota. Rincian anggota Kabupaten/Kota dari setiap kelompok dapat dilihat pada Tabel 5.

Tabel 5 menunjukkan anggota dari masing-masing *cluster* yang terbentuk dari metode *Complete Linkage* berdasarkan Indikator Kesejahteraan Rakyat di Jawa Timur tahun 2019. Hasil *dendrogram* berdasarkan proses *agglomerasi* dapat dilihat pada Gambar 1.

Hasil pemetaan Kabupaten/Kota di Jawa Timur berdasarkan 4 kelompok yang terbentuk dari hasil pengelompokan menggunakan analisis *cluster hierarki*, dapat dilihat pada Gambar 2.

Pemetaan pada Gambar 2 dilakukan dengan metode *Complete Linkage* berdasarkan indikator kesejahteraan rakyat di Jawa Timur dengan menggunakan jumlah kelompok sebanyak 4 *cluster*.

Karakteristik dari setiap kelompok yang sudah terbentuk didapatkan dari rata-rata setiap variabel dari kelompok yang terbentuk. Karakteristik setiap variabel akan dijelaskan pada Tabel 6 menunjukkan tentang deskripsi kelompok daerah baru yang terbentuk dari masing-masing variabel yang digunakan.

Tabel 6 dengan hasil karakteristik pada Lampiran 1 yang menjelaskan bahwa Kelompok 1 merupakan Kelompok yang memiliki kondisi kesetaraan ekonomi dan kesehatan yang baik namun memiliki kesetaraan pendidikan yang cukup baik. Kelompok 2 memiliki kondisi kesetaraan ekonomi baik namun memiliki kesetaraan kesehatan dan pendidikan yang cukup baik. Kelompok 3 memiliki kondisi kesetaraan ekonomi yang cukup baik namun memiliki kesetaraan

kesehatan dan pendidikan yang baik. Kelompok 4 memiliki kondisi kesetaraan ekonomi yang baik namun memiliki kesetaraan kesehatan dan pendidikan yang sangat baik.

Dapat dilihat bahwa secara umum kelompok 2 memiliki kesetaraan yang kurang baik. Untuk meningkatkan kesetaraan dalam segi ekonomi perlu diperhatikan dalam hal partisipasi kerja, hal ini dapat dilakukan dengan membuat lapangan pekerjaan. Sedangkan untuk meningkatkan kesetaraan kesehatan yang perlu diperhatikan dalam hal rumah tangga yang memiliki air minum layak, hal ini dapat dilakukan dengan hemat air, membuang limbah industri pada tempat yang seharusnya agar air tidak tercemar. Kemudian, untuk meningkatkan kesetaraan pendidikan yang perlu diperhatikan dalam hal angka partisipasi sekolah dan rata-rata lama sekolah, hal ini dapat dilakukan dengan menetapkan wajib belajar minimum yang tinggi.

## V. KESIMPULAN

Hasil pengelompokan dari empat metode *Cluster Hierarki* yaitu 4 kelompok dengan metode terbaik *Complete Linkage*, dilihat dari nilai *Pseudo-F* tertinggi dan *ICDRate* terkecil. Berdasarkan hasil Pemetaan Kabupaten/Kota di Jawa Timur menggunakan indikator kesejahteraan rakyat diperoleh 4 pembagian kelompok yaitu tingkat kesejahteraan yang sangat tinggi, tinggi, sedang dan rendah. Kelompok 2 memiliki tingkat kesejahteraan yang rendah, di mana kesetaraan yang kurang baik dalam hal ekonomi dapat diperhatikan pada partisipasi kerja, untuk kesehatan dapat diperhatikan pada rt yang memiliki air minum layak dan untuk pendidikan yang perlu diperhatikan adalah angka partisipasi sekolah dan rata-rata lama sekolah.

Saran yang dapat diberikan kepada Pemerintah Provinsi Jawa Timur adalah hasil penelitian ini dapat digunakan sebagai informasi untuk meningkatkan kesetaraan dalam segi ekonomi yang perlu diperhatikan dalam hal partisipasi kerja terutama pada daerah Kabupaten Bangkalan. Sedangkan untuk meningkatkan kesetaraan kesehatan yang perlu diperhatikan dalam hal rumah tangga yang memiliki air minum layak pada daerah Kabupaten Gresik dan untuk meningkatkan kesetaraan pendidikan yang perlu diperhatikan dalam hal angka partisipasi sekolah pada daerah Kota Malang dan rata-rata lama sekolah pada daerah Kabupaten Sampang.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Direktorat Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan Republik Indonesia yang telah memberikan dukungan finansial melalui Beasiswa Bidik Misi tahun 2017-2021.

## DAFTAR PUSTAKA

- [1] S. Roestam, *Pembangunan Nasional untuk Kesejahteraan Rakyat*. Jakarta: Menteri Koordinator Bidang Kesejahteraan Rakyat, 1993.
- [2] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. London: Pearson education, 2014.
- [3] A. Sajidah, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Rakyat Menggunakan Metode C-Means dan Fuzzy C-Means Clustering," Departemen Statistika: Institut Teknologi Sepuluh Nopember, 2016.
- [4] Z. Abidin, "Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kemiskinan dengan Menggunakan Analisis

- Cluster Hierarki,” Departemen Statistika: Institut Teknologi Sepuluh Nopember, 2018.
- [5] S. Y. Ghaisani, N. Hikmah, A. H. Prasetyo, and E. Widodo, “Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016,” in *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) IV*. 2019.
- [6] H. Z. Fitri, “Pengelompokan Dan Pemetaan Wilayah Kecamatan di Kabupaten Ponorogo Berdasarkan Potensi Sektor Pertanian Menggunakan Analisis Klaster,” Departemen Statistika: Institut Teknologi Sepuluh Nopember, 2017.
- [7] A. R. Orpin and V. E. Kostylev, “Toward a statistically valid method of textural sea floor characterization of benthic habitats,” *Mar. Geol.*, vol. 225, no. 1–4, pp. 209–222, 2006.
- [8] S. Sharma, *Applied Multivariate Techniques*. Los Angeles: John Wiley & Sons, 1996.
- [9] S. A. Mingoti and J. O. Lima, “Comparing SOM Neural Network with fuzzy C-means, k-means and traditional hierarchical clustering algorithms,” *Eur. J. Oper. Res.*, vol. 174, no. 3, pp. 1742–1759, 2006.