

# Pengelompokan Kemiskinan di Indonesia Menggunakan *Time Series Based Clustering*

Dedi Setiawan<sup>1\*</sup>, Amalia Zahra<sup>1</sup>

<sup>1</sup>Magister Teknik Informatika, Bina Nusantara University, Jakarta, Indonesia

\*dedi.setiawan001@binus.ac.id

Received: 30 November 2022

Revised: 8 March 2023

Accepted: 15 March 2023

**ABSTRAK** – Indonesia memiliki komitmen yang besar dalam mencapai *Sustainable Development Goals* (SDGs) 2030, salah satu target dalam SDGs tersebut adalah pengentasan kemiskinan. Kemiskinan sendiri diartikan sebagai ketidakmampuan orang atau sekelompok orang dalam memenuhi kebutuhan makanan dan bukan makanan. Pada masa pandemi Indonesia mengalami peningkatan persentase kemiskinan yaitu mencapai puncaknya pada bulan September 2020 dengan angka sebesar 10,19%. Pada pencatatan terakhir bulan Maret 2022 angka tersebut sudah menurun sebesar 0,75% atau setara dengan 1,79 juta penduduk. Namun, demikian tidak semua provinsi mengalami penurunan masih ada provinsi yang mengalami peningkatan. Perlu dilakukan analisa untuk mengelompokan provinsi-provinsi di Indonesia tersebut berdasarkan persentase kemiskinan dengan tujuan untuk memberikan *treatment* yang lebih tepat. Metode analisis yang digunakan dalam penelitian adalah metode *time series based clustering* pada jarak *Dynamic Time Warping* (DTW). Algoritma penentuan kelompok yang digunakan adalah hirarki *cluster complete linkage*. Hasil analisa diperoleh bahwa penggunaan jarak DTW dapat meningkatkan nilai koefisien *silhouette* menjadi 0,75 dibandingkan menggunakan jarak *Euclidean*. Koefisien *silhouette* merupakan salah satu parameter yang digunakan untuk menentukan kebaikan hasil pengelompokan, dimana nilai 0,75 sudah dapat dikatakan hasil pengelompokan sudah sangat baik. Kelompok optimal hasil *clustering* sebanyak 3 kelompok dengan kategori kemiskinan rendah, sedang, dan tinggi. Provinsi NTT, Papua, dan Papua Barat merupakan provinsi dengan tingkat kemiskinan tertinggi, namun memiliki *progress* penurunan kemiskinan yang signifikan.

**Kata kunci**– *cluster*, DTW, hirarki, kemiskinan, dan *time series*

**ABSTRACT** – Indonesia has a strong commitment in achieving the 2030 SDGs, which one of the targets is reducing poverty. Poverty itself is defined as the inability of an individual or group to meet the basic needs of both food and non-food. During the pandemic, Indonesia experienced an increase in poverty percentage, which peaked in September 2020 with 10.19%. The latest data in March 2022, showed that the number has decreased by 0.75% or equivalent of 1.79 million people. However, the decrease does not encompass all provinces, there are several provinces that still suffer from increasing poverty. Therefore, it is necessary to group the provinces in Indonesia based on its percentage of poverty, in order to provide more appropriate treatment. The analysis method used in this research is the time series-based clustering with dtw distance. The clustering algorithm used is hierarchical cluster complete linkage. Based on the analysis result, the use of dtw distance can increase the *silhouette* coefficient value to 0.75 compared to using the Euclidean distance. The *silhouette* coefficient is one of the parameters used to determine the goodness of the clustering results, where the value of 0.75 can already be said to be very good clustering results. The optimum result of the clustering is a total of 3 groups with low, medium, and high poverty categories, where NTT, Papua, and West Papua provinces have the highest poverty rates but have significant progress in reducing poverty.

**Keywords**– *cluster*, DTW, hierarchy, poverty, and *time series*

## I. PENDAHULUAN

Salah satu target *Sustainable Development Goals* (SDGs) tahun 2030 adalah pengentasan kemiskinan, Indonesia merupakan salah satu negara yang berkomitmen atas hal tersebut. Kemiskinan sendiri diartikan sebagai ketidakmampuan seseorang dalam memenuhi kebutuhan pokok baik makanan maupun bukan makanan yang diukur dari sisi pengeluaran menurut BPS. Indonesia sendiri melalui BPS mencatat tingkat kemiskinan Maret 2022 sebesar 9,54% angka tersebut menurun sekitar 0,17% dibandingkan September 2021. Persentase kemiskinan tersebut merupakan pencapaian Indonesia terbaik selepas terjadinya Covid-19 yang melanda seluruh negara di Dunia. Indonesia mencatatkan kemiskinan pada awal pandemi, yaitu bulan Maret 2022 sebesar 9,78%. Selama kurun waktu terjadinya pandemi terjadi kenaikan kemiskinan yang signifikan dan mencapai puncak pada bulan September 2020 sebesar 10,19% dan berangsur menurun pada periode setelahnya. Kendati demikian tidak semua provinsi di Indonesia mengalami penurunan kemiskinan pasca terjadinya masa pademi, masih ada provinsi dengan persentase kemiskinan meningkat sebelum pandemi. Hal tersebut menandakan bahwa setiap provinsi di Indonesia memiliki karakteristik berbeda-beda dalam penanganan kemiskinan, sehingga peneliti tertarik dalam melakukan penelitian terkait kemiskinan ini. Salah satu metode statistika yang digunakan dalam kasus ini adalah analisis *cluster*.

Analisis *cluster* merupakan analisis statistika dalam mengelompokkan suatu objek dengan objek lain berdasarkan kedekatan karakteristik di antaranya. Telah banyak penelitian yang dilakukan menggunakan metode *cluster* ini mulai dari yang sederhana hingga kompleks [1], [2]. Penelitian terkait pengelompokkan kemiskinan kab/kota di pulau Jawa pernah dilakukan dan didapatkan hasil jumlah kelompok optimal sebanyak 2 dengan kriteria kelompok kemiskinan tinggi dan rendah [3]. Penelitian lain mengenai pengelompokan kasus kemiskinan pernah dilakukan dengan menggunakan metode *k-means clustering* menggunakan jarak *Euclidean* [4], [5]. Pengelompokan objek data yang mempertimbangkan *series* data dikenal juga sebagai *time series based clustering*. Saat ini analisis *cluster time series* telah banyak digunakan di berbagai bidang ilmu pengetahuan seperti ekonomi, psikologi, kesehatan, sosial masyarakat maupun kependudukan [6]. Penelitian *cluster time series* juga pernah dilakukan sebelumnya [7]. Hal yang membedakan dengan *cluster* pada umumnya (*cross section*) adalah metode perhitungan jarak. Salah satu metode perhitungan jarak pada *cluster time series* adalah *dynamic time warping* (DTW), di mana metode ini merupakan salah satu metode yang sering

digunakan dan menghasilkan *performance* lebih baik. Penelitian terkait penggunaan jarak DTW dalam proses *clustering* dilakukan untuk mengelompokkan kasus Covid-19 aktif di Indonesia [8].

Berdasarkan uraian di atas, peneliti tertarik untuk melakukan penelitian mengenai pengelompokan kemiskinan di Indonesia, dengan mempertimbangkan pola *series* kemiskinan di Indonesia setiap periodenya. Metode yang digunakan dalam penelitian ini adalah metode *cluster* dengan jarak DTW. Pengeleompokan pada penelitian ini menggunakan *cluster* hirarki, kelebihan dari pengelompokan hirarki adalah mudah dipahami dan diterapkan serta *output* dendrogram dari algoritma dapat digunakan untuk memahami gambaran besar serta kelompok dalam data pengamatan. Dengan menggunakan dendrogram dapat diketahui *step-step* hasil pengelompokan dan perbandingan untuk *cluster* dengan *time series* dan tidak. Penggunaan algoritma *complete linkage* pada penelitian ini dikarenakan algoritma ini biasanya menghasilkan *cluster* yang lebih rapat daripada *single-linkage* dan hasil *cluster* yang rapat ini dapat berakhir sangat berdekatan [9], sehingga diharapkan menghasilkan *cluster* yang lebih akurat. Penelitian terkait kemiskinan di Indonesia sendiri dengan menggunakan *cluster time series* masih minim sehingga diharapkan penelitian ini menjadi acuan pada penelitian-penelitian selanjutnya yang serupa. *Metric* yang digunakan dalam penentuan jumlah *cluster* optimal dan kebaikan hasil *clustering* adalah menggunakan koefisien *silhouette*. Selain itu, untuk hasil dari penelitian diharapkan dapat memberikan informasi mengenai penanganan tingkat kemiskinan setiap provinsi di Indonesia yang lebih sesuai dengan karakteristik pada masing-masing provinsi.

## II. TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai teori-teori yang berkaitan dengan penelitian ini, yang meliputi teori pengelompokan data dan mengenai kemiskinan sebagai topik dalam penelitian. Teori pengelompokan data akan dibagi menjadi dua, pengelompokan berbasis *time series* dan tidak.

### A. Analisis Cluster Hirarki

Analisis *cluster* atau kelompok adalah metode analisis yang digunakan untuk mengelompokkan beberapa objek penelitian ke dalam beberapa kelompok sesuai kedekatan karakteristik diantara objek-objek tersebut [10]. Terdapat dua tipe analisis *cluster* yaitu analisis *cluster* hirarki dan analisis *cluster* non-hirarki. Fokus penelitian ini terletak pada metode *cluster* hirarki. Analisis klaster hierarki adalah metode statistik utama untuk menemukan kelompok kasus yang homogen berdasarkan karakteristik yang diukur [11]. Pembentukan *cluster* hirarki sendiri memiliki sifat seperti struktur pohon atau hirarki yang bercabang. Tahap *cluster* dimulai dengan pengelompokan dua objek atau lebih yang mirip, yang nantinya akan dilanjutkan ke objek-objek lainnya yang terbentuk seperti pohon [12], di mana pohon *cluster* ini dapat ditampilkan dalam bentuk dendrogram. Terdapat dua metode untuk *cluster* hirarki yaitu, *divisive* dan *agglomerative*. Algoritma *agglomerative* terdiri dari beberapa algoritma, yaitu *complete linkage*, *single linkage*, *average linkage*, *ward method*, *weighted pair group method with arithmetic mean* (WPGMA), *median method*, dan *centroid method* [6]. *Complete linkage* merupakan algoritma yang digunakan dalam penelitian ini. Metode *complete linkage* menggunakan aturan jarak maksimum antar *cluster*, alasan memiliki jarak maksimum adalah objek yang memiliki sedikit kesamaan dapat dihubungkan [13]. Dalam menentukan anggota *cluster* digunakan Persamaan 1 [14].

$$d_{(i,j)k} = \max(d_{ik}, d_{jk}) \tag{1}$$

di mana  $d_{ik}$  merupakan jarak antara obyek  $i$  dan  $k$ , sedangkan  $d_{jk}$  adalah jarak antara obyek  $j$  dan  $k$ .

### B. Pengukuran Jarak Kedekatan Objek

Dalam teknik pengelompokan atau pembentukan *cluster* dilakukan dengan cara mencari dan mengelompokkan objek-objek berdasarkan kedekatannya. Teknik untuk mengukur seberapa dekat kemiripan antara suatu objek dengan objek lain menggunakan jarak terdekat. Perhitungan jarak objek pada penelitian ini dijelaskan pada poin-poin sebagai berikut:

#### 1. Jarak Euclidean

Metode jarak *euclidean* merupakan perhitungan jarak yang paling sederhana dan paling sering digunakan dalam penelitian [15]. Persamaan jarak *Euclidean* dapat dituliskan dalam Persamaan 2. Jarak objek ke- $i$  terhadap objek ke- $j$  disimbolkan dengan  $d_{ij}$  dan  $k$  adalah variabel ke-  $(1,2,3,\dots,p)$  serta disyaratkan  $i \neq j$ .

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

#### 2. Jarak DTW

Penentuan jarak kedekatan antar objek, dengan memperhatikan pola *series* data dapat menggunakan perhitungan DTW. Logika perhitungan jarak DTW menerapkan teknik pemrograman dinamis untuk menemukan semua jalur yang mungkin dan memilihnya untuk menghasilkan jarak minimum antara dua *series* data dan menampilkannya dalam bentuk matriks. Masing-masing elemen dalam matriks yang terbentuk merupakan jarak kumulatif nilai minimal dari tiga tetangga sekitarnya. Jika diberikan dua data *time series*  $Q = q_1, q_2, q_3, \dots, q_m$  dengan ukuran  $m$  dan  $C = c_1, c_2, c_3, \dots, c_n$  ber-ukuran  $n$ , sehingga terbentuk matriks  $m \times n$ . Nilai  $d_{ij}$  merupakan jarak antara  $q_i$  terhadap  $c_j$ , perhitungan nilai  $d_{ij}$  dapat dituliskan pada Persamaan 3-4.

$$d_{ij} = (q_i - c_j)^2 \tag{3}$$

$$d_{DTW}(Q,C) = \min_{\forall w \in P} \left\{ \sqrt{\sum_{k=1}^K d_{wk}^2} \right\} \tag{4}$$

$P$  adalah sekumpulan dari semua *warping path* yang mungkin, dan  $wk$  adalah elemen  $(i,j)$  ke- $k$  pada *warping path* dan  $K$  adalah panjang dari *warping path* [16].

**C. Koefisien Silhouette**

Salah satu kriteria ketepatan hasil pengelompokan adalah koefisien *silhouette*. Koefisien *silhouette* akan berada pada rentang -1 hingga 1, di mana semakin besar koefisien *silhouette* maka hasil pengelompokan semakin baik artinya antar kelompok memiliki ketidaksamaan yang tinggi dan di dalam kelompok memiliki kesamaan yang tinggi. Koefisien *silhouette* dapat dirumuskan dalam Persamaan 5, di mana  $a(i)$  adalah rata-rata jarak antar anggota dalam kelompok dan  $b(i)$  adalah nilai minimum dari rata-rata jarak dari objek  $i$  dengan objek yang berada di kelompok lainnya. Tabel 1 menampilkan kriteria koefisien *silhouette* [17].

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{5}$$

**Tabel 1** Kriteria Koefisien *Silhouette*

Nilai Koefisien <i>Silhouette</i>	Keterangan
0,71 – 1,00	Struktur yang dihasilkan kuat
0,51 – 0,70	Struktur yang dihasilkan baik
0,26 – 0,50	Struktur yang dihasilkan lemah
≤ 0,25	Tidak ada struktur substansial yang ditemukan

**D. Kemiskinan**

Pengentasan kemiskinan merupakan salah satu target SDGs Indonesia pada tahun 2023. Kemiskinan sendiri merupakan ketidakmampuan seseorang atau sekelompok orang dalam pemenuhan kebutuhan makanan dan bukan makanan yang dikurungai dari pengeluaran. Pada penelitian ini, angka kemiskinan yang dimaksud adalah persentase indikator kemiskinan P0. Indikator kemiskinan P0 merupakan persentase penduduk yang hidup di bawah garis kemiskinan nasional, menurut jenis kelamin dan kelompok umur [18]. Rumus dalam penentuan kemiskinan P0 dapat ditulis pada Persamaan 6.  $PPM$  merupakan persentase penduduk yang hidup di bawah garis kemiskinan nasional,  $JPM$  merupakan jumlah penduduk yang berada di bawah garis kemiskinan nasional pada waktu tertentu dan  $JP$  adalah jumlah penduduk pada periode waktu yang sama.

$$PPM = \frac{JPM}{JP} \times 100\% \tag{6}$$

**III. METODOLOGI PENELITIAN**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari BPS melalui *website* [www.bps.go.id](http://www.bps.go.id). Data yang dimaksud pada penelitian ini yaitu data kemiskinan P0 setiap provinsi di Indonesia dari tahun 2007 hingga Semester 1 tahun 2022. Unit observasi yang digunakan dalam penelitian sebanyak 34 provinsi, sedangkan data *series* sebanyak 26 observasi. Contoh dan struktur data yang digunakan dalam penelitian disajikan pada Tabel 2.

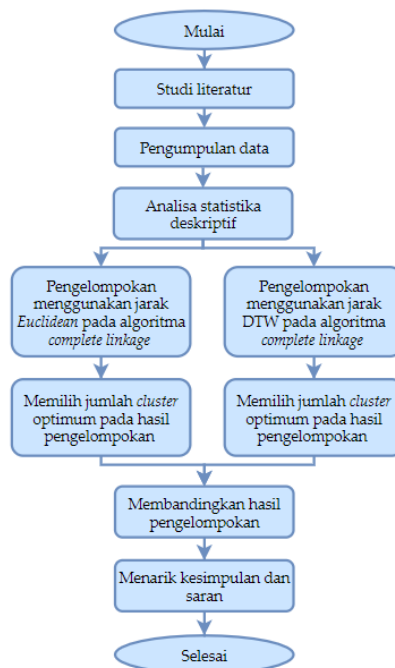
**Tabel 2** Struktur Data Penelitian

Provinsi	Persentase Penduduk Miskin (P0)					
	2007	2008	2009	...	2021 S2	2022 S1
Aceh	26,65%	23,53%	21,80%	...	15,53%	14,64%
Sumatera Utara	13,90%	12,55%	11,51%	...	8,49%	8,42%
Sumatera Barat	11,90%	10,67%	9,54%	...	6,04%	5,92%
...	...	...	...	...	...	...
Papua Barat	39,31%	35,12%	35,71%	...	21,82%	21,33%
Papua Barat	40,78%	37,08%	37,53%	...	27,38%	26,56%

\*) S1= Semester 1 dan S2= Semester 2

Langkah analisis pada penelitian dijelaskan pada tahapan-tahapan seperti yang akan dijelaskan. Tahap pertama melakukan studi literatur pada teori-teori yang digunakan dalam penelitian. Selanjutnya adalah mengumpulkan data penelitian. Tahap ketiga adalah melakukan analisis deskriptif data ini bertujuan untuk mengetahui gambaran secara umum kemiskinan di Indonesia serta mengetahui pola kemiskinan setiap provinsi. Selanjutnya adalah melakukan pengelompokan kemiskinan di setiap provinsi dengan menggunakan jarak *euclidean* sesuai Persamaan 2 dengan metode *cluster* hirarki *complete linkage* Persamaan 1. Pengelompokan pada penelitian ini dilakukan percobaan dengan jumlah

kelompok 2 hingga 6 kelompok. Pada tahapan penelitian yang sama dilakukan juga menggunakan jarak DTW sesuai Persamaan 3-4. Selanjutnya dilakukan perhitungan performa pengelompokan yang dilihat berdasarkan koefisien *silhouette* sesuai Persamaan 5. Tahapan kelima yang dilakukan adalah melakukan analisis deskriptif pada masing-masing kelompok untuk mengetahui perbedaan pada masing-masing kelompok hal ini digunakan untuk mengidentifikasi nama pada masing-masing kelompok. Tahap terakhir adalah menarik kesimpulan dan saran. Semua tahapan analisis yang telah dijelaskan tersebut dapat disederhanakan menjadi diagram alir pada Gambar 1.



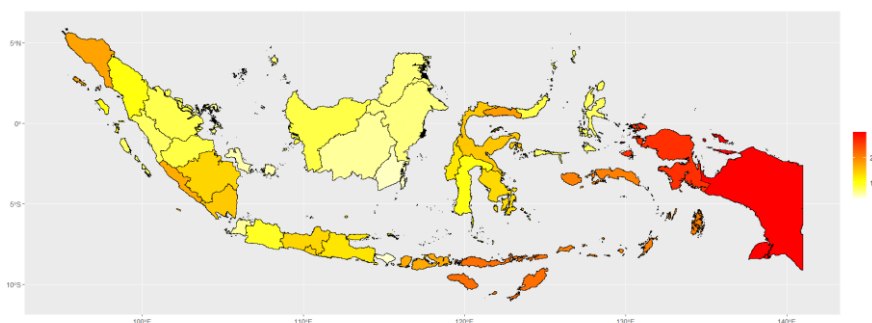
Gambar 1 Diagram Alir Penelitian

#### IV. HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai hasil penelitian, dibagi dalam beberapa sub bab yang meliputi karakteristik kemiskinan di Indonesia dan hasil pengelompokan kemiskinan.

##### A. Karakteristik Kemiskinan di Indonesia

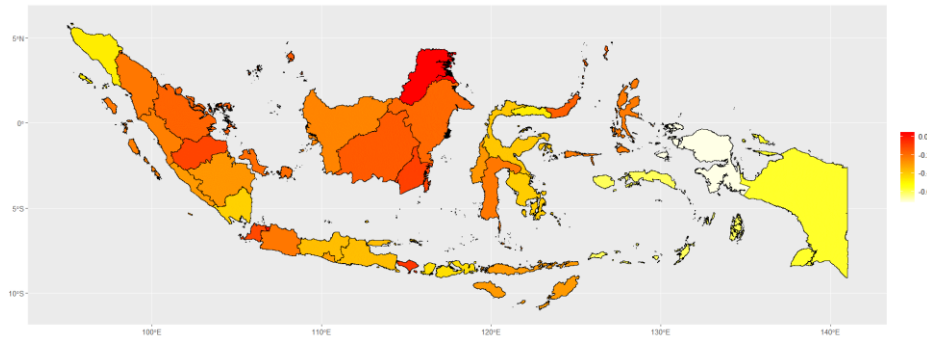
Karakteristik kemiskinan di Indonesia dapat dilihat berdasarkan statistika deskriptif berupa rata-rata. Rata-rata kemiskinan di Indonesia selama 15 tahun terakhir disajikan pada Gambar 2. Dapat dijelaskan bahwa semakin gelap gambar maka kemiskinan semakin tinggi, Indonesia Timur merupakan wilayah dengan rata-rata kemiskinan di Indonesia tinggi dibandingkan wilayah lain, di mana provinsi Papua memiliki rata-rata tertinggi selama 15 tahun terakhir dengan angka sebesar 29,96%. Kalimantan merupakan satu-satunya wilayah Indonesia yang memiliki rata-rata kemiskinan terkecil dengan rata-rata sebesar 6,52%. Provinsi dengan rata-rata kemiskinan terendah di Indonesia adalah provinsi DKI Jakarta dengan angka kemiskinan sebesar 3,94% di mana ini juga disebabkan karena semua wilayahnya merupakan perkotaan dan memiliki PDRB yang tinggi. Sebagai pembandingan rata-rata kemiskinan di Indonesia 15 tahun terakhir 11,26%, sehingga jumlah provinsi yang memiliki rata-rata diatas nasional sebanyak 16 (47.06%) provinsi dari total 34 provinsi di Indonesia.



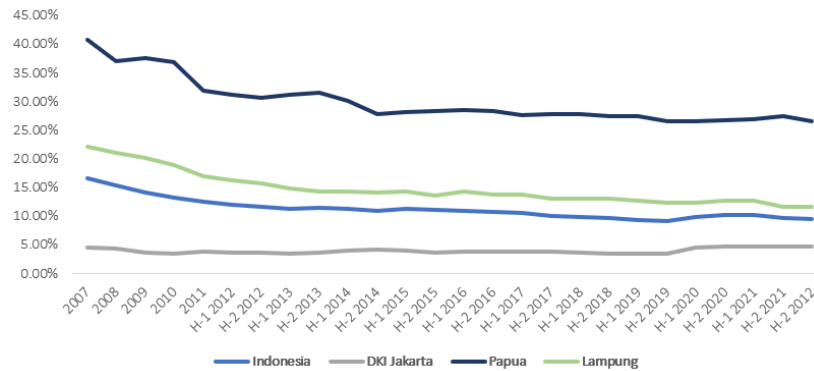
Gambar 2 Rata-rata Persentase Kemiskinan di Indonesia Tahun 2017 – Semester 1 2022

Selain melihat karakteristik kemiskinan menggunakan rata-rata, juga dilakukan analisis mengenai perubahan kemiskinan setiap periodenya. Semakin gelap warna dalam peta menandakan bahwa rata-rata perubahan kemiskinan semakin positif (meningkat), begitu juga sebaliknya (Gambar 3). Walaupun provinsi Papua merupakan provinsi yang memiliki tingkat kemiskinan tertinggi, namun Papua memiliki *progress* penurunan kemiskinan yang tertinggi

dibandingkan provinsi lain dengan rata-rata penurunan kemiskinan setiap semesternya sebesar -0,57%. Kalimantan memiliki progres penurunan tingkat kemiskinan yang paling rendah dibandingkan dengan provinsi lain di Indonesia. Gambar 4 menunjukkan perkembangan kemiskinan di Indonesia pada beberapa provinsi dan secara nasional.



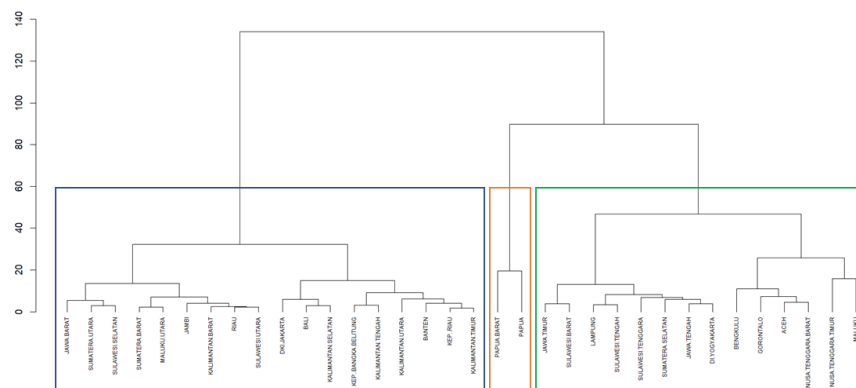
Gambar 3 Rata-rata Perubahan Persentase Kemiskinan di Indonesia Tahun 2017 – Semester 1 2022



Gambar 4 Perkembangan Persentase Kemiskinan beberapa Provinsi di Indonesia Tahun 2017 – Semester 1 2022

**B. Pengelompokan Kemiskinan di Indonesia**

Pengelompokan kemiskinan di Indonesia pada penelitian ini dilakukan dengan 2 (dua) metode, yaitu dengan jarak *euclidean* dan jarak DTW. Hasil pengelompokan kemiskinan di Indonesia dengan menggunakan jarak *euclidean* dan metode *complete linkage* pada jumlah kelompok 2 hingga 6 didapatkan hasil kelompok optimum sebanyak 3 kelompok. Penentuan jumlah kelompok optimum berdasarkan koefisien *silhouette* yang disajikan pada Tabel 5, dengan maksimum koefisien *silhouette* sebesar 0,64. Pada Gambar 5 disajikan dendrogram hasil pengelompokan dengan menggunakan jarak *euclidean*.



Gambar 5 Dendrogram Hasil Pengelompokan Menggunakan Jarak *Euclidean*

Pembagian kelompok kemiskinan terdapat 3 kelompok dengan kriteria kemiskinan rendah, sedang dan tinggi pada masing-masing provinsi di Indonesia sebagai berikut:

1. Kelompok kemiskinan rendah terdapat 18 provinsi yaitu, Sumatera Utara, Sumatera Barat, Jambi, Riau, Kep. Riau, Kep. Bangka Belitung, Banten, DKI Jakarta, Jawa Barat, Bali, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Kalimantan Timur, Sulawesi Utara, Sulawesi Selatan dan Maluku Utara.
2. Kelompok kemiskinan sedang terdapat 14 Provinsi yaitu, Aceh, Sumatera Selatan, Bengkulu, Lampung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Nusa Tenggara Timur, Nusa Tenggara Barat, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Maluku dan Gorontalo.

3. Kelompok kemiskinan tinggi terdapat 2 provinsi yaitu, Papua dan Papua Barat. Peta pengelompokan kemiskinan di Indonesia disajikan pada Gambar 6, hanya wilayah Papua yang memiliki tingkat kemiskinan dalam kelompok tinggi.



Gambar 6 Peta Hasil Pengelompokan Menggunakan Jarak Euclidean

Kriteria masing-masing kelompok disajikan dalam Tabel 3. Kelompok kemiskinan sedang memiliki rata-rata kemiskinan sekitar 2 kali dibandingkan dengan kelompok rendah serta untuk kelompok kemiskinan tinggi memiliki rata-rata sekitar 4 kali dibandingkan dengan rata-rata kelompok rendah. Namun demikian setiap semesternya penurunan kemiskinan tinggi memiliki progres yang signifikan dibandingkan kelompok lain, dengan rata-rata penurunan sebesar 0,64%.

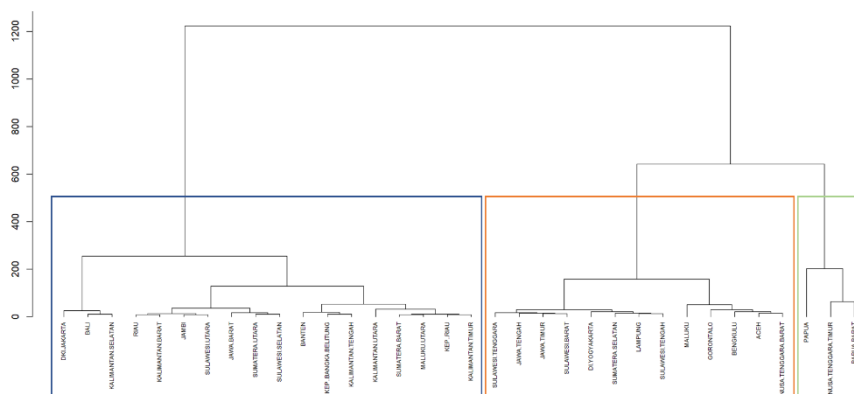
Tabel 3 Karakteristik Hasil Pengelompokan Menggunakan Jarak Euclidean

Kelompok Kemiskinan	%Rata-rata Kemiskinan 2007 - 2022	%Rata-rata Kemiskinan 2022	Trend Perubahan %Kemiskinan Setiap Semester
Rendah	07,13%	06,37%	-0,14%
Sedang	15,85%	13,27%	-0,39%
Tinggi	28,19%	23,95%	-0,64%

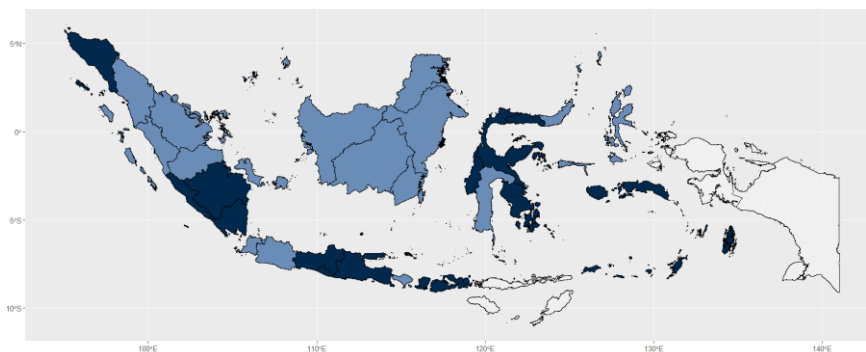
Pengelompokan kemiskinan di Indonesia juga dilakukan perhitungan dengan menggunakan jarak DTW yang sudah memperhatikan pola series data kemiskinan di Indonesia pada setiap tahun dan semester. Pengelompokan juga dilakukan dengan beberapa percobaan dengan jumlah kelompok mulai 2 hingga 6 dan hasil perhitungan koefisien *silhouette* masing-masing kelompok disajikan pada Tabel 4. Dengan menggunakan metode *complete linkage* didapatkan jumlah kelompok optimum sebanyak 3 kelompok dengan koefisien *silhouette* sebesar 0,75. Dendrogram pada pengelompokan menggunakan jarak DTW ini disajikan pada Gambar 7. Pembagian provinsi di Indonesia berdasarkan metode ini disebutkan sebagai berikut:

1. Kelompok kemiskinan rendah terdapat 18 provinsi yaitu, Sumatera Utara, Sumatera Barat, Jambi, Riau, Kep. Riau, Kep. Bangka Belitung, Banten, DKI Jakarta, Jawa Barat, Bali, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Kalimantan Timur, Sulawesi Utara, Sulawesi Selatan dan Maluku Utara.
2. Kelompok kemiskinan sedang terdapat 13 Provinsi yaitu, Aceh, Sumatera Selatan, Bengkulu, Lampung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Maluku dan Gorontalo.
3. Kelompok kemiskinan tinggi terdapat 3 provinsi yaitu, Nusa Tenggara Timur, Papua dan Papua Barat.

Peta pengelompokan kemiskinan di Indonesia disajikan pada Gambar 8. Hal yang membedakan dari hasil pengelompokan menggunakan jarak Euclidean dengan jarak DTW adalah untuk kelompok kemiskinan tinggi. Pada jarak Euclidean hanya terdapat 2 provinsi yang masuk ke dalam kelompok tinggi, yakni Papua dan Papua Barat. Namun, pada penggunaan jarak DTW, terdapat 3 provinsi dengan penambahan provinsi Nusa Tenggara Timur.



Gambar 7 Dendrogram Hasil Pengelompokan Menggunakan Jarak DTW



Gambar 8 Peta Hasil Pengelompokan Menggunakan Jarak DTW

Kriteria masing-masing kelompok disajikan dalam Tabel 4. Kelompok kemiskinan sedang memiliki rata-rata kemiskinan sekitar 2 kali dibandingkan dengan kelompok rendah serta untuk kelompok kemiskinan tinggi memiliki rata-rata sekitar 4 kali dibandingkan dengan rata-rata kelompok rendah. Namun demikian setiap semesternya penurunan kemiskinan tinggi memiliki *progress* yang signifikan dibandingkan kelompok lain, dengan rata-rata penurunan sebesar 0,53%.

Tabel 4 Karakteristik Hasil Pengelompokan Menggunakan Jarak DTW

Kelompok Kemiskinan	%Rata-rata Kemiskinan 2007 - 2022	%Rata-rata Kemiskinan 2022	Trend Perubahan %Kemiskinan Setiap Semester
Rendah	07,13%	06,37%	-0,14%
Sedang	15,40%	12,75%	-0,40%
Tinggi	26,00%	22,65%	-0,53%

### C. Perbandingan Hasil Pengelompokan

Perbandingan hasil pengelompokan kedua metode jarak yang digunakan berdasarkan koefisien *silhouette*, dengan melihat koefisien *silhouette* terbesar. Pada Tabel 5 disebutkan bahwa pada masing-masing jarak yang digunakan menghasilkan jumlah kelompok optimum sebanyak 3 kelompok. Jarak DTW menghasilkan koefisien *silhouette* yang terbesar dibandingkan dengan penggunaan jarak *Euclidean* dengan koefisien *silhouette* sebesar 0,75. Hal tersebut juga menandakan bahwa penggunaan jarak DTW ini cocok digunakan untuk kasus pengelompokan karena data yang digunakan adalah berbasis *time series*. Peningkatan nilai koefisien *silhouette* untuk jarak DTW dibandingkan dengan jarak *euclidean* sekitar 17,19%. Berdasarkan Tabel 1 mengenai kriteria *silhouette*, jarak DTW menghasilkan pengelompokan yang dikatakan kuat dibandingkan jarak *Euclidean* yang berada 1 level di bawahnya yaitu dengan kriteria hasil pengelompokan baik.

Tabel 5 Perbandingan Koefisien *Silhouette* Hasil Pengelompokan

Jumlah Kelompok	Jarak <i>Euclidean</i>	Jarak DTW
2	0,60	0,72
3	0,64	0,75
4	0,60	0,61
5	0,53	0,56
6	0,51	0,53

## V. KESIMPULAN DAN SARAN

Penelitian ini mengidentifikasi tingkat kemiskinan di Indonesia dan mengelompokan setiap provinsi di Indonesia berdasarkan karakteristiknya. Kemiskinan di Indonesia untuk beberapa tahun terakhir telah mengalami penurunan, di mana pada semester 1 tahun 2022 dengan persentase sebesar 9,54% dibandingkan tahun 2007 sebesar 16,58%. Provinsi Papua dan Papua Barat merupakan provinsi dengan tingkat kemiskinan tertinggi di Indonesia, kendati demikian tingkat penurunan kemiskinan di provinsi tersebut tergolong signifikan dibandingkan provinsi lain. DKI Jakarta dan Kalimantan Utara merupakan wilayah yang memiliki tren peningkatan kemiskinan untuk 15 tahun terakhir, di mana Kalimantan Utara memiliki nilai yang tertinggi yaitu dengan rata-rata kenaikan setiap semesternya sebesar 0,04%. Hasil pengelompokan kemiskinan di Indonesia menggunakan *time series based clustering* dengan jarak DTW didapatkan hasil yang terbaik dengan koefisien *silhouette* sebesar 0,75 dibandingkan dengan menggunakan pengelompokan menggunakan jarak *Euclidean*. Penggunaan jarak DTW pada penelitian ini dapat meningkatkan koefisien *silhouette* sebesar 17,19% dan dapat meningkatkan kategori hasil pengelompokan dengan kategori kuat. Dari hasil pengelompokan didapatkan bahwa terdapat 3 kelompok optimum dengan pengelompokan kemiskinan kategori rendah sebanyak 18

(52,94%) provinsi, sedang 13 (38,24%) provinsi dan tinggi 3 (8,82%) provinsi. Kelompok optimum penelitian ini juga sejalan dengan beberapa penelitian lain [19] [20].

Berdasarkan kesimpulan yang diperoleh, dapat dirumuskan saran sebagai pertimbangan penelitian selanjutnya adalah pengelompokan dengan menggunakan jarak DTW pada *series* data kemiskinan sudah baik, namun pada penelitian selanjutnya dapat dikombinasikan dengan menambahkan beberapa variabel. Penambahan variabel lain terkait kemiskinan dapat menggunakan metode *multivariate dynamic time warping* (m-DTW). Saran untuk pihak pemerintah berdasarkan hasil analisis ini adalah memperhatikan kelompok-kelompok dengan kemiskinan tertinggi dan kelompok yang memiliki tren peningkatan kemiskinan setiap tahunnya.

## REFERENSI

- [1] M. A. Zen, S. Wahyuningsih, A. Tri, and R. Dani, "Aplikasi Pendekatan Agglomerative Hierarchical Time Series Clustering untuk Peramalan Data Harga Minyak Goreng di Indonesia (Application of Agglomerative Hierarchical Time Series Clustering Approach for Forecasting Cooking Oil Price Data in Indonesia)," *Seminar Nasional Official Statistics 2022*, pp. 293–302, 2022.
- [2] N. Ulinnuh and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, pp. 101–108, 2022, doi: 10.30743/infotekjar.v5i1.2464.
- [3] 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.
- [4] S. Annas, B. Poerwanto, S. Sapriani, and M. F. S., "Implementation of K-Means Clustering on Poverty Indicators in Indonesia," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 257–266, Mar. 2022, doi: 10.30812/matrik.v21i2.1289.
- [5] C. A. Sugianto and T. P. O. R. Bokings, "K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province," *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 58–67, Feb. 2021, doi: 10.47709/cnahpc.v3i1.934.
- [6] A. Tri, R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, 2019, [Online]. Available: <http://ejournal.ung.ac.id/index.php/jjom-P>
- [7] X. Wang, K. Smith, and R. Hyndman, "Characteristic-Based Clustering for Time Series Data," *Data Min Knowl Discov*, vol. 13, no. 3, pp. 335–364, Sep. 2006, doi: 10.1007/s10618-005-0039-x.
- [8] M. I. Rizki, T. A. Taqqiyuddin, and J. J. Cerelia, "K-Medoids Clustering dengan Jarak Dynamic Time Warping dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif Covid-19," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, pp. 685–692, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [9] "Introduction to Hierarchical Clustering | by John Clements | Towards Data Science." <https://towardsdatascience.com/introduction-hierarchical-clustering-d3066c6b560e> (accessed Mar. 14, 2023).
- [10] A. C. Rencher, *Methods of multivariate analysis*, 2nd ed. Canada: Wiley Interscience, 2002.
- [11] Z. Zhang, F. Murtagh, S. Van Poucke, S. Lin, and P. Lan, "Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R," *Ann Transl Med*, vol. 5, no. 4, pp. 75–75, Feb. 2017, doi: 10.21037/atm.2017.02.05.
- [12] D. T. Utari and D. S. Hanun, "Hierarchical Clustering Approach for Region Analysis of Contraceptive Users," *EKSAKTA: Journal of Sciences and Data Analysis*, pp. 99–108, Sep. 2021, doi: 10.20885/eksakta.vol2.iss2.art3.
- [13] C. C. Astuti and R. S. Untari, "Applied Hierarchical Cluster Analysis with Average Linkage Algorithm," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 5, no. 1, pp. 1–7, Nov. 2017, doi: 10.18860/ca.v5i1.3862.
- [14] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. United States of America: Pearson Education, Inc.
- [15] D. Setiawan, S. A. L. Eka, R. Mubarak, Ernawati, and B. W. Otok, "Analisis Cluster pada Kabupaten di Papua Berdasarkan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia (IPM) Tahun 2013," 2017. [Online]. Available: <https://www.researchgate.net/publication/317062880>
- [16] V. Niennattrakul and C. A. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," in *Proceedings - 2007 International Conference on Multimedia and Ubiquitous Engineering, MUE 2007*, 2007, pp. 733–738. doi: 10.1109/MUE.2007.165.
- [17] A. Struyf, M. Hubert, and P. J. Rousseeuw, "Clustering in an Object-Oriented Environment."
- [18] Badan Pusat Statistik, "Badan Pusat Statistik." [https://www.bps.go.id/indikator/indikator/view\\_data/0000/data/192/sdgs\\_10/1](https://www.bps.go.id/indikator/indikator/view_data/0000/data/192/sdgs_10/1) (accessed Nov. 25, 2022).
- [19] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, "Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, pp. 557–566, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.971.
- [20] A. Astasia, "Analisis Cluster Kemiskinan dan Indeks Pembangunan Manusia di Indonesia dengan Data Outlier Menggunakan K-Medoids dan Analisis Biplot," 2021.



© 2023 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).