MATHunesa
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115

Volume 13 No 02
Tahun 2025

ANALISIS *CLUSTERING* INDEKS KEDALAMAN KEMISKINAN DI INDONESIA MENGGUNAKAN *K-MEANS*

Nanda Reza Handitia
Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya
e-mail: nandareza.21034@mhs.unesa.ac.id

A'yunin Sofro
Program Studi S1 Sains Aktuaria, FMIPA, Universitas Negeri Surabaya
e-mail: ayuninsofro@unesa.ac.id*

Abstrak

Kemiskinan merupakan salah satu isu utama dalam pembangunan berkelanjutan yang diukur menggunakan berbagai indikator, salah satunya adalah indeks kedalaman kemiskinan (P_1). Indeks tersebut memberikan gambaran yang lebih komprehensif mengenai tingkat keparahan kemiskinan dibandingkan dengan persentase penduduk miskin (P_0). Dalam penelitian ini, dilakukan analisis *cluster*ing terhadap provinsi di Indonesia berdasarkan data time series indeks kedalaman kemiskinan untuk mengidentifikasi pola perubahan yang serupa. Penelitian ini menerapkan metode *K-Means* dengan tiga pendekatan pengukuran jarak, yaitu *Dynamic Time Warping* (DTW), *Derivative Dynamic Time Warping* (DDTW), dan *Weighted Dynamic Time Warping* (WDTW). DTW digunakan untuk menangani pergeseran waktu dalam data, DDTW mempertimbangkan pola perubahan dengan menghitung turunan pertama, sedangkan WDTW memberikan bobot pada setiap titik data untuk menghasilkan hasil yang lebih stabil. Hasil *cluster*ing divalidasi menggunakan *silhouette coefficient* untuk menentukan metode yang paling optimal. Hasil penelitian menunjukkan bahwa metode *K-Means* dengan pengukuran jarak WDTW yang memiliki nilai *silhouette coefficient* sebesar 0,884 (*strong*).
Kata Kunci: Indeks Kedalaman Kemiskinan, *K-Means*, Ukuran jarak

Abstract

Poverty is one of the main issues in sustainable development that is measured using various indicators, one of which is the poverty depth index (P_1). This index provides a more comprehensive picture of the severity of poverty than the percentage of poor people (P_0). In this study, a clustering analysis of provinces in Indonesia is conducted based on time series data of the poverty depth index to identify similar patterns of change. This research applies the K-Means method with three distance measurement approaches, namely Dynamic Time Warping (DTW), Derivative Dynamic Time Warping (DDTW), and Weighted Dynamic Time Warping (WDTW). DTW is used to handle time shifts in the data, DDTW considers the pattern of change by calculating the first derivative, while WDTW assigns weights to each data point to produce more stable results. The clustering results were validated using the silhouette coefficient to determine the most optimal method. The results showed that the K-Means method with WDTW distance measurement had a silhouette coefficient value of 0.884 (strong).
Keywords: *Poverty Depth Index, K-Means, Distance measure*

PENDAHULUAN

Kemiskinan merupakan salah satu isu prioritas dalam pembangunan berkelanjutan yang tercantum dalam *Sustainable Development Goals* (SDGs) poin pertama (Utama and Sari 2023). Indikator yang umum digunakan dalam mengukur kemiskinan adalah persentase penduduk miskin (P_0), namun indikator tersebut memiliki keterbatasan karena tidak menggambarkan tingkat keparahan kemiskinan secara mendalam. Untuk mengatasi kelemahan tersebut, indeks kedalaman kemiskinan (P_1) dapat digunakan sebagai ukuran yang lebih komprehensif dalam menilai seberapa jauh kondisi ekonomi masyarakat miskin dari garis kemiskinan (Haughton and Khandker 2009). Indeks kedalaman kemiskinan merupakan ukuran rata-rata kesenjangan pengeluaran penduduk miskin terhadap garis kemiskinan (BPS, 2025). Melalui indeks tersebut juga dapat diperkirakan besaran intervensi yang diperlukan dalam program pembebasan kemiskinan dengan asumsi distribusi

bantuan yang tepat sasaran (Apriliani, Safrida, and Fajri 2023).

Analisis *clustering* adalah metode pengelompokan data ke dalam beberapa kelompok sehingga objek-objek yang memiliki kesamaan ditempatkan dalam satu *cluster*, sementara yang berbeda dimasukkan ke *cluster* lainnya (Tendean and Purba 2020). Penerapan analisis *clustering* semakin berkembang dengan penggunaan data deret waktu (*time series*). Data *time series* adalah data yang disusun berdasarkan urutan waktu dan digunakan untuk memantau perubahan suatu fenomena dari waktu ke waktu (Nababan and Alexander 2020). Pada data *time series*, perbedaan pola temporal menjadi salah satu tantangan terbesar dalam proses *clustering*. Setiap deret waktu memiliki karakteristik temporal yang berbeda-beda yang menyebabkan kesulitan dalam membandingkan dua deret waktu yang mungkin terlihat serupa tetapi memiliki pergeseran waktu yang signifikan. Oleh karena itu, aspek penting dalam analisis *clustering time series* adalah pemilihan metode pengukuran jarak yang sesuai.

Pengukuran jarak *Euclidean* merupakan metode standar untuk analisis berbasis jarak, namun metode tersebut dianggap kurang optimal untuk data yang memiliki pergeseran temporal (Middlehurst et al. 2021). Berbagai pengukuran jarak baru telah dikembangkan, seperti *Dynamic Time Warping (DTW)* yang menjadi metode paling sering digunakan dalam analisis *cluster* data *time series* karena lebih fleksibel dalam mencocokkan pola meskipun terdapat perbedaan waktu (Holder, Middlehurst, and Bagnall 2024). Namun, *DTW* memiliki kelemahan ketika ada perbedaan bentuk kurva di antara dua deret waktu. Untuk mengatasi hal tersebut, terdapat metode pengembangan dari *DTW* bernama *Derivative Dynamic Time Warping (DDTW*) yang tidak hanya memperhitungkan nilai asli dari data, tetapi juga mempertimbangkan pola perubahan dengan menghitung turunan pertama. Dengan demikian, *DDTW* lebih sensitif terhadap perubahan pola yang terjadi secara bertahap (Górecki and Łuczak 2015).

DTW dan *DDTW* mengasumsikan bahwa semua titik dalam urutan memiliki bobot yang sama. Hal tersebut menyebabkan masalah ketika satu titik pada satu urutan disesuaikan dengan banyak titik pada urutan lain. Untuk mengatasi masalah ini, terdapat pengembangan *DTW* lain yang bernama *Weighted Dynamic Time Warping (WDTW). WDTW* menambahkan bobot yang berbeda pada setiap titik data, sehingga memberikan penekanan pada fitur-fitur penting dari urutan waktu tersebut. Dengan cara tersebut, *WDTW* dapat memberikan hasil yang lebih stabil pada deret waktu dengan karakteristik yang berbeda-beda (Jeong, Jeong, and Omitaomu 2011).

Analisis *clustering* terbagi menjadi dua jenis, yaitu *clustering* hirarki dan non-hirarki (partisional). Sejumlah penelitian analisis *clustering* pada data *time series* menggunakan metode hirarki pernah dilakukan (Azizi et al. 2019; Fortuna, Nunnari, and Nunnari 2016), namun hasil penelitian menunjukkan bahwa tidak direkomendasikannya penggunaan metode hirarki karena menghasilkan terlalu banyak *cluster* dan kurang sesuai dengan pola yang terdapat pada data deret waktu. Sebaliknya, penggunaan metode non hirarki pada penelitian sejenis menghasilkan pengelompokan yang lebih baik karena kemampuannya dalam menganalisis dataset berukuran besar dan menemukan pola temporal (Alfan et al. 2017). Salah satu metode non-hirarki yang paling populer adalah *K-Means*, metode tersebut banyak digunakan dalam berbagai disiplin ilmu terutama untuk dataset berukuran besar (Holder, Middlehurst, and Bagnall 2024). Kelebihan utama dari *K-Means* adalah kesederhanaan dan efektivitas serta kecepatan komputasi (Romadhonia et al 2023).

Penelitian ini menerapkan analisis *clustering* untuk mengetahui pengelompokan provinsi di Indonesia berdasarkan data *time series* indeks kedalaman kemiskinan. Pengelompokan tersebut mampu mengidentifikasi provinsi dengan pola perubahan indeks kedalaman yang serupa, sehingga dapat dijadikan pertimbangan kebijakan pemerintah dalam mengatasi kemiskinan. Dalam penelitian digunakan tiga metode pengukuran jarak, yaitu *Dynamic Time Warping* (DTW), *Derivative Dynamic Time Warping* (DDTW), dan *Weighted Dynamic Time Warping* (WDTW). Pada ketiga jarak tersebut akan diterapkan dalam pengukuran jarak pada *K-Means*, hasil dari masing-masing metode tersebut nantinya di validasi dengan *silhouette coefficient* untuk menentukan hasil *cluster* paling optimal.

KAJIAN TEORI

INDEKS KEDALAMAN KEMISKINAN

Indeks Kedalaman Kemiskinan (*Poverty Gap Index*) menggambarkan rata-rata kesenjangan pengeluaran penduduk miskin terhadap garis kemiskinan. Indeks tersebut dapat digunakan untuk membandingkan kondisi kemiskinan antar wilayah, di mana semakin tinggi nilai P_1, semakin besar tingkat kedalaman kemiskinan di wilayah tersebut (Haughton and Khandker 2009). Dengan kata lain, nilai P_1 yang tinggi menunjukkan bahwa rata-rata pengeluaran penduduk miskin di suatu daerah semakin jauh dari garis kemiskinan, sehingga dapat mencerminkan kondisi kemiskinan yang semakin memburuk (Utama and Sari 2023).

ANALISIS *CLUSTERING*

Analisis *clustering* adalah proses mengorganisir sekumpulan data ke dalam beberapa kelompok sehingga objek-objek yang mirip akan dikelompokkan dalam satu *cluster*, sementara objek-objek yang tidak mirip akan dimasukkan ke dalam *cluster* yang berbeda (Tendean and Purba 2020). Tujuan dari analisis *cluster* adalah untuk mengorganisir data secara objektif ke dalam kelompok yang homogen di mana kemiripan antar objek dalam satu *cluster* diminimalkan dan perbedaan antar objek di *cluster* yang berbeda dimaksimalkan (Kim and Kim 2020). Analisis *clustering* dibagi menjadi dua jenis, yaitu analisis *clustering* hirarki dan non-hirarki atau partisional. Perbedaan antara analisis *cluster* hirarki dan analisis *cluster* non-hirarki terletak pada awal untuk memulai pengelompokkan (Nurfitri Imro'ah 2019). Analisis *cluster* hirarki membentuk *cluster* secara bertahap dengan menggunakan *cluster* yang telah dibentuk sebelumnya, sedangkan metode partitional menentukan semua *cluster* secara bersamaan (Madhulata 2012).

DYNAMIC TIME WARPING (DTW)

Dynamic Time Warping (DTW) distance adalah metode pengukuran jarak elastis yang paling banyak diteliti dan digunakan dalam analisis deret waktu (Ratanamahatana and Keogh 2005). Dalam prosesnya, *DTW* membentuk matriks jarak di mana setiap elemen merupakan jarak kumulatif minimum dari tiga elemen tetangga terdekat. Misalkan terdapat dua deret waktu $X = x_1, x_2, \dots, x_i, \dots, x_n$ dengan panjang n dan $Y = y_1, y_2, \dots, y_j, \dots, y_m$ dengan panjang m. Langkah pertama adalah membuat matriks jarak D ukuran $n \times m$, di mana setiap elemen (i,j) dalam matriks D merupakan selisih antara x_i dan y_j, yang didefinisikan pada Persamaan 1 sebagai berikut:

$$d_{i,j} = |x_i - y_j| \tag{1}$$

dengan $i = 1,2, \dots, n$ dan $j = 1,2, \dots, m$. Implementasi matriks D adalah sebagai berikut:

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,j} \\ d_{2,1} & d_{2,2} & \dots & d_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ d_{i,1} & d_{i,2} & \dots & d_{i,j} \end{bmatrix} \tag{2}$$

Setelah mendapatkan jarak kumulatif d_{ij}, kemudian tambahkan nilai minimum dari tiga elemen tetangga terdekat yaitu $\{d_{(i-1)(j-1),}d_{(i-1)j,}d_{i(j-1),}\}$ di mana $0 < i \leq n$ dan $0 < j \leq m$ sehingga terbentuk matriks E,. Persamaan 2.3 untuk elemen *(i,j)* pada matriks E adalah sebagai berikut:

$$E_{i,j} = d_{ij} + min\{E_{(i-1)(j-1),}E_{(i-1)j,}E_{i(j-1),}\} \tag{3}$$

Implementasi dari matriks E adalah sebagai berikut:

$$E = \begin{bmatrix} E_{1,1} & E_{1,2} & \dots & E_{1,j} \\ E_{2,1} & E_{2,2} & \dots & E_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ E_{i,1} & E_{i,2} & \dots & E_{i,j} \end{bmatrix} \tag{4}$$

Setelah terbentuk matriks E, jarak *DTW* antara dua deret waktu X dan Y dapat dihitung menggunakan Persamaan 5 berikut:

$$d_{DTW}(X,Y) = \min_{\forall w \in p} \left\{ \sum_{i,j=1}^{k} E_{i,j} \right\} \tag{5}$$

di mana
p : sekumpulan dari semua warping path
$E_{i,j}$: elemen (i,j) pada matriks E
k : panjang dari *warping path*

Warping path adalah jalur yang terdiri dari jarak minimum dari elemen $E_{1,1}$ hingga $E_{n,m}$. Syarat-syarat yang harus dipenuhi dalam menentukan *warping path* adalah sebagai berikut (Spiegel 2015):

1. *Boundary conditions*: $p_1 = (1,1)$ dan $p_k = (m, n)$. Kondisi dimana titik awal dan titik akhir dari warping path adalah dimulai dari elemen matriks (1,1) hingga elemen matriks terakhir (m, n). Hal ini bertujuan agar data yang diolah seluruhnya mulai dari awal hingga akhir.
2. *Continuity conditions*: $p_{k+1} - p_k \in \{(1,0), (0,1), (1,1)\}$. Kondisi dimana penentuan warping path harus bertahap dengan indeks i dan j dengan selisih maksimal adalah 1 pada setiap langkahnya. Hal ini bertujuan agar data yang diolah tidak ada yang terlewat.
3. *Monotonicity conditions*: $n_1 \leq n_2 \leq n_3 \leq \cdots n_k$ dan $m_1 \leq m_2 \leq m_3 \leq \cdots m_k$. Kondisi dimana penentuan *warping path* berdasarkan urutan waktu agar tidak ada data yang diolah secara berulang.

***DERIVATIVE DYNAMIC TIME WARPING* (DDTW)**

Derivative Dynamic Time Warping (DDTW) adalah pengembangan dari metode *DTW* yang diperkenalkan oleh Keogh & Pazzani (1998). Perbedaan utama *DDTW* dengan *DTW* adalah bahwa *DDTW* menerapkan proses transformasi deret waktu menjadi deret diferensial terlebih dahulu. Pada *DDTW*, deret waktu asli diubah menjadi deret diferensial berdasarkan rata-rata kemiringan antara titik data sebelum dan sesudahnya.

Misalkan deret waktu asli adalah $X = \{x_1, x_2, \ldots, x_n\}$, maka deret diferensial dari X, yaitu $X' = \{x'_2, x'_3, \ldots, x'_{n-1}\}$ di mana x'_i didefinisikan sebagai rata-rata dari kemiringan antara x_{i-1} dan x_i serta antara x_i dan x_{i+1}. Formula untuk deret diferensial dari x_i adalah sebagai berikut:

$$x'_i = \frac{(x_i - x_{i-1}) + \frac{(x_{i+1} - x_{i-1})}{2}}{2} \quad (6)$$

untuk $1 < i < n$ atau atau operasi ini dilakukan untuk semua elemen dalam deret waktu kecuali elemen pertama dan terakhir. Setelah membentuk deret diferensial, *DDTW* melanjutkan dengan perhitungan seperti pada *DTW*, yaitu menghitung jarak antara dua deret waktu yang telah didiferensialkan. Misalkan $Y = \{y_1, y_2, \ldots, y_m\}$ adalah deret waktu kedua yang akan dibandingkan, maka jarak *DDTW* antara x dan y didefinisikan sebagai Persamaan 7 berikut:

$$d_{DDTW}(x, y) = d_{DTW}(x', y') \quad (7)$$

perhitungan dilakukan dengan membangun matriks jarak D berukuran $(n-1)\times(m-1)$ untuk deret waktu x' dan y'. Setiap elemen $d(i,j)$ dalam matriks jarak dihitung sebagai berikut:

$$d(i,j) = \left(x'_{i+1}, y'_{j+1}\right)^2 \quad (8)$$

Setelah membangun matriks jarak, selanjutnya akan dibangun matriks jarak kumulatif E di mana setiap elemen $E(i,j)$ memiliki kondisi sebagai berikut:

$$E_{i,j} = d_{ij} + min\{E_{(i-1)(j-1),} E_{(i-1)j,} E_{i(j-1),}\} \quad (9)$$

dengan kondisi batas $E(1,1) = d(1,1)$ yang bermakna nilai $E(i,j)$ merupakan jarak kumulatif minimum dari elemen-elemen tetangga terdekat pada langkah sebelumnya. Setelah matriks E dibangun, jarak *DDTW* antara dua deret waktu x dan y dapat dihitung dengan persamaan berikut:

$$d_{DDTW}(x, y) = E(n-1, m-1) \quad (10)$$

di mana $E(n-1,m-1)$ adalah elemen terakhir dari matriks kumulatif, yang menyatakan jarak minimum untuk menyelaraskan kedua deret waktu.

WEIGHTED DYNAMIC TIME WARPING (WDTW)

Weighted Dynamic Time Warping (WDTW) merupakan pengembangan dari *Dynamic Time Warping* (DTW) yang pertama kali diperkenalkan oleh Jeong et al., (2011) untuk meningkatkan akurasi dalam menghitung jarak antar deret waktu. Perbedaan utama *WDTW* dengan *DTW* adalah penerapan penalti bobot yang bergantung pada jarak warping antara elemen pada deret waktu, untuk elemen yang memiliki perbedaan fase yang signifikan akan mendapatkan penalti lebih besar sehingga *warping* yang berlebihan dapat dihindari atau menghindari distorsi jarak minimum yang disebabkan oleh *outlier* (nilai pencilan) (Anantasech and Ratanamahatana 2019).

Misalkan terdapat dua deret waktu $X = x_1, x_2, \dots, x_i, \dots, x_n$ dengan panjang n dan $Y = y_1, y_2, \dots, y_j, \dots, y_m$ dengan panjang m. Langkah pertama adalah membuat matriks jarak *M* berukuran nxm yang memuat jarak kuadrat antara setiap elemen X dan Y dengan definisi sebagai berikut:

$$M_{ij} = (X_i - Y_j)^2 \quad (11)$$

Untuk mengurangi efek dari perbedaan fase, langkah selanjutnya adalah menambahkan bobot penalti pada jarak kuadrat. Bobot dihitung menggunakan fungsi logistik berdasarkan perbedaan indeks $|i-j|$ pada pada kedua deret waktu yang didefinisikan sebagai berikut:

$$w(a) = \frac{w_{max}}{1 + e^{-g(a-\frac{m}{2})}} \quad (12)$$

dimana
$w(a)$: bobot penalti
w_{max}: nilai maksimum dari bobot
g: parameter yang mengontrol kelandaian logistik
a: perbedaan indeks antara dua elemen
m: panjang deret waktu
setelah menghitung bobot penalti, langkah selanjutnya adalah membuat matriks jarak M^w berbobot dengan menerapkan bobot penalti ke matriks jarak dengan definisi sebagai berikut:

$$M^w_{i,j} = w(|i-j|)(x_i - y_j)^2 \quad (13)$$

Setelah membangun matriks jarak, selanjutnya akan dibangun matriks jarak kumulatif *E* di mana setiap elemen *E(i,j)* memiliki kondisi sebagai berikut:

$$E_{i,j} = d_{ij} + min\{E_{(i-1)(j-1),}E_{(i-1)j,}E_{i(j-1),}\} \quad (14)$$

Dengan kondisi awal $E_{(1,1)} = M^W{}_{(1,1),}$ matriks tersebut dibangun secara rekursif dengan menambahkan jarak dari matriks jarak berbobot M^w dan nilai minimum dari tetangga terdekat di *E*. Setelah matrik jarak kumulatif *E* selesai dihitung, nilai akhir jarak WDTW adalah elemen di posisi *(n,m)* pada matriks *E*, yaitu:

$$d_{wDTW}(X,Y) = E(n,m) \quad (15)$$

di mana *E(n,m)* adalah total jarak kumulatif terendah yang telah diperhitungkan berdasarkan jarak antar deret waktu dan penalti fase.

K-MEANS

K-Means adalah metode pengelompokan partisi atau non-hirarki paling terkenal dan populer, baik dalam pengelompokan standar maupun pengelompokan *time series*. Penemuan metode *K-Means* lebih dari 50 tahun lalu oleh Steinhaus (1956), kemudian dikembangkan oleh Ball & Hall (1967) serta MacQueen (1967) untuk diterapkan dalam berbagai bidang seperti psikologi, riset pemasaran, kedokteran, dan biologi. Secara umum, metode iteratif *K-Means* diterapkan sebagai berikut (Chu et al. 2012):

1. Tentukan nilai *k* atau jumlah *cluster*. Inisialisasi titik pusat (*centroid*) awal untuk setiap *cluster* secara acak.
2. Hitung jarak antara setiap objek dan *centroid*, lalu alokasikan objek ke *cluster* yang memiliki *centroid* terdekat.
3. Tentukan *centroid* baru untuk setiap *cluster* dengan menghitung rata-rata dari semua anggota *cluster* yang telah ditetapkan.
4. Ulangi langkah 2 dan 3 hingga tidak ada perubahan pada fungsi kriteria setelah satu iterasi.

GAP STATISTIC

Gap Statistic adalah metode yang diperkenalkan oleh Tibshirani et al., (2001) yang digunakan untuk menentukan jumlah *cluster* optimal. Misalkan X_{ij} adalah pengamatan pada objek ke-i dan variabel ke-j. Objek data ditetapkan pada *cluster* sejumlah k, yaitu C_1, C_2, ..., C_k dengan C_r merupakan pengamatan pada *cluster* ke-r dan n_r merupakan banyaknya anggota *cluster* r. Jika d_{ik} merupakan jarak antara objek ke-i dan objek ke-k maka didefinisikan:

$$D_r = \sum_{ik,jk} d_{ik} \quad (17)$$

di mana D_r merupakan jarak semua titik dalam *cluster* r. Kemudian, didefinisikan W_k yang merupakan jumlah kuadrat gabungan dalam k *cluster* sebagai berikut:

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \quad (18)$$

Prosedur *Gap Statistic* adalah sebagai berikut:

1. Mengelompokkan data dan memvariasikan banyaknya *cluster* mulai dari k = 1, 2, 3, ..., K yang masing-masing memberikan ukuran dispersi dalam W_k.
2. Membuat sampel bootstrap dengan cara melakukan resampling dengan pengembalikan sebanyak B kali menggunakan distribusi uniform pada daerah nilai hasil variabel pengamatan.
3. Mendapatkan nilai W_{kb}^* dari poin (2) di mana b = 1, 2, 3, ..., B dan k = 1, 2, 3, ..., K.
4. Menghitung nilai *Gap Statistic* untuk k tertentu dengan persamaan sebagai berikut:

$$Gap\ (k) = \left(\frac{1}{B}\right) \sum_{b=1}^{B} \log(W_{kb}^*) - \log(W_k) \quad (19)$$

5. Menghitung standar deviasi dengan persamaan sebagai berikut:

$$sd_k = \left[\left(\frac{1}{B}\right) \sum_{b=1}^{B} \{\log(W_{kb}^*) - I\}^2\right]^{\frac{1}{2}} \quad (20)$$

di mana $I = \left(\frac{1}{B}\right) \sum_{b=1}^{B} \log(W_{kb}^*)$

6. Menghitung nilai S_k dengan persamaan sebagai berikut:

$$s_k = sd_k \sqrt{(1 + \frac{1}{B}} \quad (21)$$

7. Menghitung nilai diffu dengan persamaan sebagai berikut:

$$Diffu = Gap\ (k) - Gap\ (k+1) - S_{k+1} \quad (22)$$

8. Mengestimasi jumlah *cluster* optimal melalui nilai k terkecil sehingga $Gap\ (k) \geq Gap\ (k+1) - S_{k+1}$ dan juga menentukan *cluster* optimal berdasarkan nilai $Diffu \geq 0$.

SILHOUETTE COEFFICIENT

Silhouette coefficient merupakan ukuran untuk menilai kualitas dan kekuatan *cluster*, khususnya untuk mengukur seberapa tepat suatu objek ditempatkan dalam *cluster* tertentu pada pengelompokan deret waktu. Metode ini mengkombinasikan dua konsep utama dari *cohesion* dan *separation,* di mana *cohesion* merujuk pada pengukuran kedekatan antara suatu objek dengan objek lain dalam satu *cluster*, sementara *separation* merujuk pada pengukuran seberapa jauh suatu objek dengan objek pada *cluster* lain (Dewi and Pramita 2019).

Proses perhitungan *Silhouette Coefficient* dimulai dengan menghitung *a(i),* yaitu rata-rata jarak objek *i* dengan objek lain di dalam *cluster* yang sama atau disebut *cohesion*. Penulisan *a(i)* dapat dituliskan dalam persamaan sebagai berikut :

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (23)$$

di mana *A* adalah *cluster*, *i* dan *j* adalah objek-objek dalam *cluster A*, dan *d(i,j)* adalah jarak antara objek *i* dan *j* ((Mario, Herry, and Nasution 2016).

Langkah berikutnya adalah menghitung *b(i),* yaitu nilai rata-rata jarak data ke-*i* terhadap *cluster* lain dan memilih nilai terkecil sebagai *separation*. Penulisan *b(i)* dapat dinyatakan dalam persamaan sebagai berikut:

$$b(i) = \min_{C \neq A} d(i, j) \quad (24)$$

di mana *C* adalah *cluster* lain yang berbeda dengan *A* sehingga dapat dinyatakan dalam persamaan lain sebagai berikut:

$$b(i) = \min_{C \neq A} \left(\frac{1}{|C|} \sum_{j \in C} d(i, j) \right) \quad (25)$$

Nilai *Silhouette Coefficient* untuk objek *i* dapat dihitung menggunakan persamaan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\ (a(i), b(i))} \quad (26)$$

Berikut kriteria pengukuran nilai *Silhouette Coefficient* (Kaufman and Rousseeuw 1990):

Tabel 1. Kriteria Koefisien Silhouette

Nilai Koefisien *Silhouette*	**Kriteria *Cluster***
0,71 - 1,00	*Strong*
0,51 - 0,70	*Good*
0,26 - 0,50	*Weak*
0,00 - 0,25	*Bad*

METODE

DATA PENELITIAN

Jenis penelitian dilakukan secara kuantitatif dengan menggunakan data yang diperoleh melalui publikasi Badan Pusat Statistik (BPS) mengenai Indeks Kedalaman Kemiskinan periode 2015 hingga 2024. Dalam penelitian ini digunakan variabel berupa indeks kedalaman kemiskinan dari 34 Provinsi di Indonesia pada rentang tersebut sehingga untuk setiap provinsi memiliki 10 data deret waktu.

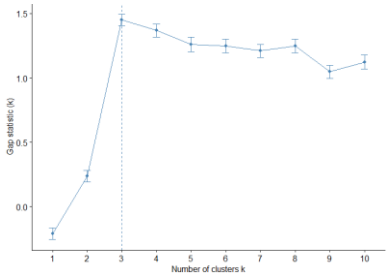
RANCANGAN PENELITIAN

Rancangan penelitian untuk mencapai tujuan dari penelitian yaitu menentukan jumlah *cluster* terbaik menggunakan *gap statistics*. Selanjutnya, dilakukan analisis *cluster*ing *K-Means* berdasarkan tiga jenis metode pengukuran jarak, yaitu *Dynamic Time Warping* (DTW), *Derivative Dynamic Time Warping* (DDTW), dan *Weighted Dynamic Time Warping* (WDTW). Hasil untuk setiap metode dilakukan validasi *cluster* menggunakan *silhouette coefficient* untuk mengetahui metode paling optimal.

HASIL DAN PEMBAHASAN

PENENTUAN JUMLAH *CLUSTER*

Langkah awal dari penelitian merupakan penentuan jumlah *cluster* terbaik yang akan digunakan dalam proses selanjutnya. Dengan menerapkan metode gap statistic didapatkan grafik sebagai berikut:



Gambar 1. Grafik *Gap Statistic*

Berdasarkan Gambar 1 ditunjukan bahwa jumlah *cluster* terbaik yang didapatkan adalah 3, hal tersebut karena ketika *cluster* sejumlah 3 diperoleh nilai *diffu* positif serta $Gap(3) \geq Gap(4)\text{-}S_4$.

HASIL ANALISIS *CLUSTERING*

Setelah mendapatkan jumlah *cluster* terbaik. Pada metode *K-Means* digunakan tiga metode pengukuran jarak untuk menentukan kedekatan antar objek dan penetapan *cluster*. Proses iterasi berhenti ketika tidak terjadi perubahan pada penetapan *cluster* setelah satu kali iterasi. Dengan demikian, didapatkan hasil analisis *cluster*ing untuk setiap jenis metode pengukuran jarak sebagai berikut:

Tabel 2. Hasil Analisis *Clustering K-Means* berdasarkan jarak *Dynamic Time Warping* (DTW)

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

Tabel 3. Hasil Analisis *Clustering K-Means* berdasarkan jarak *Derivative Dynamic Time Warping* (DDTW)

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

Tabel 4. Hasil Analisis *Clustering K-Means* berdasarkan jarak *Weighted Dynamic Time Warping* (WDTW)

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

VALIDASI HASIL *CLUSTER*

Setelah mengetahui hasil *cluster*ing dari masing-masing metode pengukuran jarak, langkah selanjutnya yaitu melakukan validasi hasil *cluster* paling optimal dari keseluruhan metode pengukuran jarak menggunakan *silhouette coefficient*. Nilai *silhouette coefficient* yang paling mendekati satu ialah nilai yang dianggap memiliki hasil *cluster* paling optimal. Perbandingan *silhouette coefficient* disajikan pada tabel berikut:

Tabel 5. Hasil *Silhouette Coefficient*

Metode	**Nilai *Silhouette Coefficient***	**Kriteria *Cluster***
K-Means dengan jarak *Dynamic Time Warping* (DTW)	0.8311712	*Strong*
K-Means dengan jarak *Derivative Dynamic Time Warping* (DDTW)	0.7031006	*Good*
K-Means dengan jarak *Weighted Dynamic Time Warping* (WDTW)	0.8848709	*Strong*

Dari hasil perbandingan *silhouette coefficient* pada Tabel 8, diketahui bahwa hasil *cluster* metode *K-Means* dengan dengan *Derivative Dynamic Time Warping* (DDTW) masuk pada kategori *good* karena memiliki nilai 0.703. Selanjutnya, hasil analisis *clustering* metode *K-Means* dengan jarak *Dynamic Time Warping* (DTW) dan *Weighted Dynamic Time Warping* (WDTW) masuk pada kategori strong karena memiliki nilai 0.831 dan 0.884. Dengan meninjau hasil nilai *silhouette coefficient* terbesar maka metode *K-Means* dengan jarak *Weighted Dynamic Time Warping* (WDTW) merupakan metode yang memiliki akurasi paling baik dalam mengelompokkan data deret waktu indeks kedalaman kemiskinan di Indonesia.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis *cluster*ing *K-Means* berdasarkan metode pengukuran jarak *Dynamic Time Warping* (DTW), *Derivative Dynamic Time Warping* (DDTW), dan *Weighted Dynamic Time Warping* (WDTW), diperoleh hasil *cluster* paling optimal menggunakan metode pengukuran jarak *Weighted Dynamic Time Warping* (WDTW) di mana masuk dalam kategori *strong*.

SARAN

Untuk penelitian selanjutnya terkait analisis *clustering* data time series dapat dilakukan pengembangan terkait pengembangan metode pengukuran jarak seperti *Weighted Derivative Dynamic Time Warping* (WDDTW).

DAFTAR PUSTAKA

Alfan, Mohammad, Dian Sukma, Aldho Riski, and Kartika Fithriasari. 2017. "Clustering Stationary and Non-Stationary Time Series Based on Autocorrelation Distance of Hierarchical and K-Means Algorithms." *International Journal of Advances in Intelligent Informatics* 3(3): 154–60.

Anantasech, Pichamon, and Chotirat Ann Ratanamahatana. 2019. "Enhanced Weighted Dynamic Time Warping for Time Series Classification." *Third International Congress on Information and Communication Technology, Advances in Intelligent Systems and Computing 797*: 655–64. doi:10.1007/978-981-13-1165-9.

Apriliani, Aini Rizqa, Safrida Safrida, and Fajri Fajri. 2023. "Pengaruh Pengangguran Terhadap Indeks Kedalaman Kemiskinan Di Provinsi Aceh." *Jurnal Ilmiah Mahasiswa Pertanian* 8(1): 73–81. doi:10.17969/jimfp.v8i1.22825.

Azizi, Elnaz, Hamed Kharrati-Shishavan, Amin Mohammadpour Shotorbani, and Behnam Mohammadi-Ivatloo. 2019. "Wind Speed

Clustering Using Linkage-Ward Method: A Case Study of Khaaf, Iran." *Gazi University Journal of Science* 32(3): 945–54. doi:10.35378/gujs.459840.

Ball, G. H., and D. J. Hall. 1967. "A Clustering Technique for Summarizing Multivariate Data." *Behavioral science* 12(2): 153–55. doi:10.1002/bs.3830120210.

Chu, Hone Jay, Churn Jung Liau, Chao Hung Lin, and Bo Song Su. 2012. "Integration of Fuzzy Cluster Analysis and Kernel Density Estimation for Tracking Typhoon Trajectories in the Taiwan Region." *Expert Systems with Applications* 39(10): 9451–57. doi:10.1016/j.eswa.2012.02.114.

Dewi, Dewa Ayu Indah Cahya, and Dewa Ayu Kadek Pramita. 2019. "Analisis Perbandingan Metode Elbow Dan Silhouette Pada Algoritma Clustering K-Medoids Dalam Pengelompokan Produksi Kerajinan Bali." *Matrix : Jurnal Manajemen Teknologi dan Informatika* 9(3): 102–9. doi:10.31940/matrix.v9i3.1662.

Fortuna, Luigi, Giuseppe Nunnari, and Silvia Nunnari. 2016. "Clustering Daily Wind Speed Time Series." : 79–89. doi:10.1007/978-3-319-38764-2_8.

Górecki, Tomasz, and Maciej Łuczak. 2015. "Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping." *Expert Systems with Applications* 42(5): 2305–12. doi:10.1016/j.eswa.2014.11.007.

Haughton, Jonathan, and Shahidur Khandker. 2009. *Handbook on Poverty and Inequality*.

Holder, Christopher, Matthew Middlehurst, and Anthony Bagnall. 2024. 66 Knowledge and Information Systems *A Review and Evaluation of Elastic Distance Functions for Time Series Clustering*. Springer London. doi:10.1007/s10115-023-01952-0.

Jeong, Young Seon, Myong K. Jeong, and Olufemi A. Omitaomu. 2011. "Weighted Dynamic Time Warping for Time Series Classification." *Pattern Recognition* 44(9): 2231–40. doi:10.1016/j.patcog.2010.09.022.

Kaufman, Leonard, and Peter Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*.

Keogh, Eamonn J, and Michael J Pazzani. 2001. "Derivative Dynamic Time Warping." *In: Proceedings of the 1st SIAM international conference on data mining*: 1–11.

Kim, Jaehwi, and Jaehee Kim. 2020. "Comparison of Time Series Clustering Methods and Application to Power Consumption Pattern Clustering." *Communications for Statistical Applications and Methods* 27(6): 589–602. doi:10.29220/CSAM.2020.27.6.589.

MacQueen, James. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1(14): 281–97. h

Madhulata, T. Soni. 2012. "An Overview of Clustering Methods." *IOSR Journal of Engineering* 2(4): 719–25. doi:10.3233/ida-2007-11602.

Mario, Anggara, Sujiani Herry, and Helfi Nasution. 2016. "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness." *Jurnal Sistem dan Teknologi Informasi* 1(1): 1–6.

Middlehurst, Matthew, James Large, Michael Flynn, Jason Lines, Aaron Bostrom, and Anthony Bagnall. 2021. "HIVE-COTE 2.0: A New Meta Ensemble for Time Series Classificatio." *ACM Digital Library* 110(Machine Language): 11–12.

Nababan, Darsono, and Eric Alexander. 2020. "Implementasi Metode Fuzzy Time Series Dengan Model Algoritma Chen Untuk Memprediksi Harga Emas." *Jurnal Teknik Informatika* 13(1): 71–78. doi:10.15408/jti.v13i1.15516.

Nurfitri Imro'ah, Indah Ayuningtias, Naomi Nessyana Debataraja,. 2019. "Analisis Cluster Non-Hirarki Dengan Metode K-Modes." *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya* 8(4): 909–16. doi:10.26418/bbimst.v8i4.36633.

Ratanamahatana, Chotirat Ann, and Eamonn Keogh. 2005. "Three Myths About Dynamic Time Warping Data Mining." *In: Proceedings of the 5th SIAM international conference on data mining*.

Romadhonia, R. W., Sofro, A. Y., Ariyanto, D., Maulana, D. A., and Prihanto, J. B. 2023. "Application of decision trees in athlete selection: A CART algorithm approach. " *Journal of Data Science*, *2023*.

Spiegel, Stephan. 2015. *Time Series Distance Measures: Segmentation, Classification, and Clustering of Temporal Data*. doi:10.14279/depositonce-4619.

Steinhaus, H. 1956. "Sur La Division Des Corps Matériels En Parties." *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques* 4: 1–6.

Tendean, Tonny, and Windania Purba. 2020. "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means." *Jurnal Sains dan Teknologi)* 1(2): 5–11.

Tibshirani, R., G. Walther, and T Hastie. 2001. "Estimating the Number of Clusters in a Data

Set Via the Gap Statistic." *Science of Computer Programming* 78(8): 1073–98. doi:10.1016/j.scico.2012.08.004.

Utama, Kharisma Pandu, and Liza Kurnia Sari. 2023. "Analisis Spasial Indeks Kedalaman Kemiskinan Tiga Provinsi Di Pulau Jawa Tahun 2021." *Seminar Nasional Official Statistics* 2023(1): 353–62. doi:10.34123/semnasoffstat.v2023i1.1640.