

ANALISIS CLUSTER TIME SERIES DALAM PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN NILAI PDRB

Luh Putu Widya Adnyani¹, Pardomuan Robinson Sihombing²

¹Universitas Indraprasta PGRI,

²Badan Pusat Statistik, Jakarta

Email : robinson@bps.go.id

Abstract : *Cluster time series is a method of grouping time series data that can be used to group provinces in Indonesia based on the value of GRDP. Cluster analysis was performed using dynamic time warping (DTW) distances and Euclid distances. The use of these two distances is compared to determine the best clustering results, especially in the case of province grouping in Indonesia based on the value of Gross Regional Domestic Product (GRDP). The results of the study concluded that by forming six clusters, the Silhouette coefficient for the DTW method was 0.68 while the Eucliden method had a Silhouette coefficient of 0.605. Thus it can be stated that the two methods form clusters with categories of good classifications.*

Keywords: *Clustering Time Series, dynamic time warping, Eucliden*

Abstrak : *Cluster time series merupakan metode pengelompokkan data runtun waktu yang dapat digunakan untuk mengelompokkan provinsi-provinsi yang ada di Indonesia berdasarkan nilai PDRB. Analisis cluster yang dilakukan menggunakan jarak dynamic time warping (DTW) dan jarak Euclid. Penggunaan kedua jarak tersebut dibandingkan untuk mengetahui hasil penggerombolan terbaik, khususnya pada kasus pengelompokan provinsi di Indonesia berdasarkan nilai Pendapatan Domestik Regional Bruto (PDRB). Hasil penelitian menyimpulkan bahwa dengan membentuk enam cluster maka diperoleh koefisien Silhouette untuk metode DTW sebesar 0,68 sedangkan untuk metode Eucliden memiliki koefisien Silhouette sebesar 0,605. Dengan demikian dapat dinyatakan bahwa kedua metode tersebut membentuk cluster dengan kategori good classifications.*

Kata kunci: *Clustering Time Series, dynamic time warping, Eucliden*

PENDAHULUAN

Indonesia merupakan negara yang sedang melaksanakan dan mengembangkan pembangunan disegala sektor, salah satunya adalah sektor ekonomi. Pembangunan di bidang ekonomi dilaksanakan dengan tujuan agar semua masyarakat yang tersebar di berbagai provinsi di Indonesia dapat merasakan dampaknya. Pemerataan pembangunan dilaksanakan untuk mengatasi ketimpangan antar provinsi di Indonesia.

Salah satu indikator untuk melihat kondisi ekonomi suatu daerah adalah dengan melihat nilai Produk Domestik Regional Bruto (PDRB). PDRB pada dasarnya merupakan jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha dalam suatu daerah tertentu, ataupun merupakan jumlah nilai barang dan jasa akhir yang dihasilkan oleh seluruh unit ekonomi pada suatu daerah. PDRB atas dasar harga konstan merupakan jumlah seluruh nilai barang dan jasa akhir yang dihasilkan oleh unit-unit produksi dalam tahun berjalan yang dinilai atas dasar harga pada tahun dasar.

Cluster time series merupakan metode pengelompokan data runtun waktu yang dapat digunakan untuk mengelompokkan provinsi-provinsi yang ada di Indonesia berdasarkan nilai PDRB. Analisis *cluster* yang dilakukan menggunakan jarak *dynamic time warping* (DTW) dan jarak *Euclid*. Penggunaan kedua jarak tersebut dibandingkan untuk mengetahui hasil penggerombolan terbaik, khususnya pada kasus ini.

Beberapa penelitian terkait pernah dilakukan diantaranya adalah penelitian yang dilakukan oleh Aulia Sahputra (2020) dengan menggunakan jarak DTW mengelompokkan kabupaten/kota yang ada di Indonesia berdasarkan nilai inflasi. Hasil penelitian menunjukkan bahwa metode *Euclidian distance* dan algoritma ward method lebih baik dibandingkan dengan metode lainnya dan mengklasifikasikan 28 kabupaten/kota kedalam *cluster* pertama, 11 kabupaten/kota kedalam *cluster* kedua, 29 kabupaten/kota kedalam *cluster* ketiga, 13 kabupaten/kota kedalam *cluster* keempat dan 1 kabupaten/kota kedalam *cluster* kelima. Penelitian serupa juga dilakukan oleh Komang Sidhi Arta dan Edi Winarko (2016) dengan membandingkan beberapa metode jarak seperti Eros, *Euclidean distance* dan DTW dalam penggerombolan data multivariat *time series* menggunakan KNN. Hasil penelitian menunjukkan bahwa setiap metode perhitungan jarak memiliki akurasi, kepekaan dan kekhasan yang berbeda-beda untuk data set yang sama. Kombinasi data dan metode perhitungan jarak pun menunjukkan perbedaan hasil pada akurasi, kepekaan, maupun kecepatan proses perhitungan. Penelitian lainnya dilakukan oleh Munthe (2019) yang menggerombolkan provinsi di Indonesia berdasarkan nilai produksi padi dengan metode *clustering time series*. Hasil penelitian menyatakan bahwa terbentuk 3 gerombol dengan nilai koefisien silhouete sebesar 0,64 yang dikategorikan sebagai *good classification*.

Berdasarkan latar belakang di atas maka perlu dilakukan suatu penelitian dengan membandingkan jarak Euclid dan DTW dalam mengelompokkan provinsi-provinsi yang ada di Indonesia berdasarkan nilai PDRB. Dengan mengetahui *Cluster* masing-masing provinsi diharapkan dapat lebih mudah mengambil keputusan atas solusi yang berkaitan dengan pemerataan pembangunan disektor ekonomi.

METODE PENELITIAN

Data yang digunakan dalam penulisan artikel ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Data yang dimaksud adalah data Pendapatan Domestik Regional Bruto (PDRB) dari tahun 2010 hingga 2019.

Analisis dilakukan dengan membandingkan hasil penggerombolan data

berdasarkan nilai PDRB provinsi se Indonesia menggunakan jarak *Dynamic Time Warping* (DTW) dan Jarak *Euclid* menggunakan *software R*.

Analisis Cluster Time Series

Analisis Cluster merupakan teknik dalam menemukan kelompok dalam kumpulan data dengan tujuan data yang berada pada satu kelompok mempunyai kemiripan yang dekat, dan memiliki perbedaan yang jelas dengan kelompok lain (Kaufman & Rousseeuw, 1990). Analisis *cluster* deret waktu membutuhkan algoritma dan prosedur pengelompokan yang berbeda dengan data statis. Prosedur yang berbeda dilakukan karena data deret waktu merupakan sekelompok nilai pengamatan yang diperoleh pada titik waktu yang berbeda dengan selang waktu yang sama dan barisan data diasumsikan saling berhubungan satu sama lain. Selama berkembangnya pengelompokan pada data deret waktu, banyak Teknik yang dikembangkan diantaranya adalah penggunaan jarak-jarak pengelompokan yang sesuai dengan karakteristik data deret waktu (Liao, 2005)

Jarak *Euclid*

Jarak Euclidean merupakan tipe pengukuran jarak dalam analisis *cluster* untuk mengukur jarak dari obyek data ke pusat cluster. Jarak Euclid merupakan jarak geometris antar dua obyek data. Semakin dekat jarak maka semakin mirip suatu obyek data tersebut (Johnson & Wichern, 2002). Bentuk persamaan jarak *Euclid* untuk data runtun waktu u dan v dinyatakan sebagai berikut:

$$d_{uv} = \sqrt{\sum_{t=1}^n (u_t - v_t)^2}$$

(1)

dengan:

n : Panjang data u dan v

d_{uv} : Jarak Euclidean antara runtun waktu u dan v .

Dinamic Time Warping (DTW)

Jarak DTW adalah salah satu pengukuran ketakmiripan yang termasuk pendekatan tanpa model tertentu. Jarak DTW adalah minimum jarak antara dua pasang titik dengan memperhitungkan kemungkinan pergeseran titik (Berndt & Clifford, 1994), atau didefinisikan sebagai:

$$DTW(S, T) = \min_W [\sum_{k=1}^p \delta(w_k)]$$

(2)

dengan:

$S = s_1, s_2, \dots, s_i, \dots, s_n$ dan $T = t_1, t_2, \dots, t_j, \dots, t_m$ adalah suatu deret waktu,

$W = w_1, w_2, \dots, w_k$ adalah kemungkinan jalur lengkungan yang memetakan atau menjajarkan ulang anggota-anggota S dan T sehingga jarak antara keduanya minimum. Adapun jarak δ dapat berupa $\delta(i, j) = s_i - t_j$ dan w_k merujuk pada titik $(i, j)_k$ pada jalur lengkungan ke- k .

Koefisien *Silhouette*

Objek-objek dapat dimasukkan dalam sebanyak k gerombol. Batasan gerombol yang dapat dibentuk adalah $k \leq n$, dengan $k = 1$ dan $k = n$ dikecualikan. Sehingga dapat

disimpulkan bahwa $2 \leq k \leq (n - 1)$. Formula pengukuran ketepatan penggerombolan (Kaufman & Rousseeuw, 1990) dapat ditulis sebagai berikut:

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

(3)

dengan:

$a(i)$ = rata-rata ketidakmiripan setiap obyek ke-I ke semua obyek selainnya dalam anggota gerombol A.

$b(i)$ = rata-rata ketidakmiripan setiap obyek ke-I ke semua obyek selainnya dalam anggota gerombol B, jika gerombol A dianggap tidak ada.

$s(i)$ adalah koefisien *Silhouette* dengan rentang nilai $-1 \leq s(i) \leq 1$ untuk setiap objek.

Menurut Kaufman dan Rousseeuw (1990) intepretasi koefisien *Silhouette* dapat dinyatakan seperti pada tabel berikut:

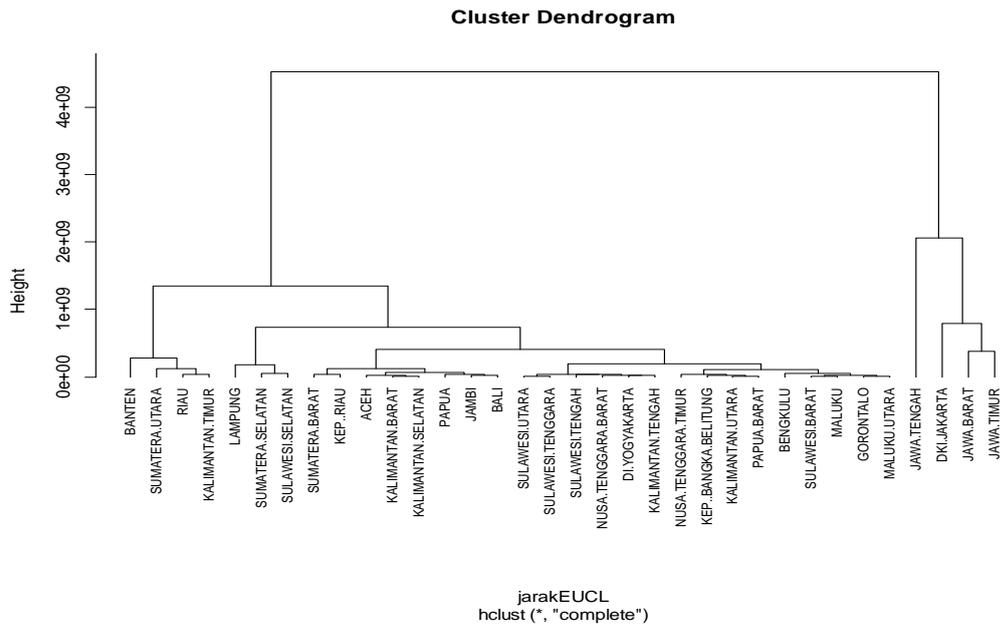
Tabel 1. Kategori Penggerombolan Berdasarkan Koefisien *Silhouette*

No	Koefisien <i>Silhouette</i>	Kategori Penggerombolan
1.	0.71-1.00	Strong Classification
2.	0.51-0.70	Good Classification
3.	0.26-0.50	Weak Classification
4.	0.00-0.25	Bad Classification

HASIL DAN PEMBAHASAN

Clustering Time Series dengan Jarak Euclid

Hasil terkait pengklasteran dengan metode Jarak Ecludid dapat dilihat pada Gambar 1. Berdasarkan Gambar 1 dapat dilihat bahwa terbentuk enam kluster dengan *cluster* pertama beranggotakan provinsi: Maluku Utara, Gorontalo, Maluku, Sulawesi Barat, Bengkulu, Papua Barat, Kalimantan Utara, Kep. Bangka Belitung, NTT, Kalteng, D.I. Yogyakarta, NTB, Sulawsi Tengah, Sulawesi Tenggara, Sulawesi Utara, Bali, Jambi, Papua, Kalimantan Selatan, Kalimantan Barat, Aceh, Kepulauan Riau dan Sumbar, sedangkan *cluster* kedua beranggotakan provinsi: Kalimantan Timur, Riau Sumatera Utara dan Banten. *Cluster* ketiga beranggotakan provinsi: Sulawesi Selatan, Sumatera Selatan, dan Lampung. *Cluster* keempat merupakan provinsi DKI Jakarta. Cluster kelima beranggotakan Jawa barat dan Jawa Timur dan cluster keenam merupakan provinsi Jawa Tengah. Untuk lebih jelasnya jumlah masing-masing cluster disajikan pada Tabel 2 di bawah.



Gambar 1. Dendrogram Penggerombolan dengan Jarak Euclid

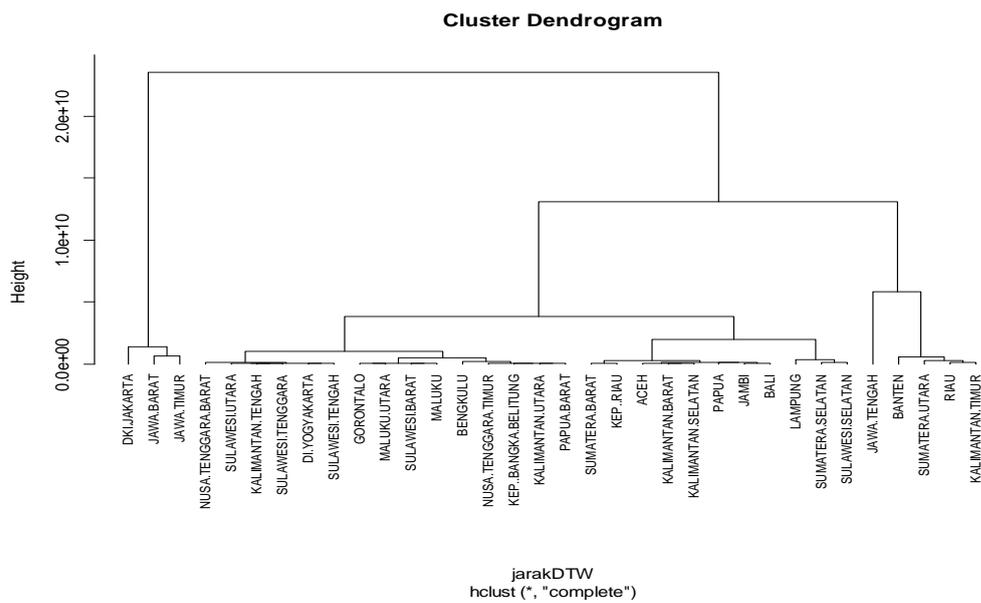
Tabel 2. Jumlah Masing-Masing Cluster (Jarak Euclid)

Kluster	Jumlah
1	23
2	4
3	3
4	1
5	2
6	1

Jika diinginkan 2 cluster besar, maka empat provinsi di Jawa membentuk cluster tersendiri yaitu DKI Jakarta, Jawa Barat, Jawa Tengah dan Jawa Timur. Keempat provinsi ini memiliki kontribusi yang terbesar terhadap total PDB Indonesia.

Selanjutnya pada Gambar 2 merupakan *dendrogram* yang dihasilkan dari analisis cluster *time series* dengan jarak *Dynamic Time Warping*. Berdasarkan Gambar 2 dapat dilihat bahwa terbentuk enam kluster dengan *cluster* pertama beranggotakan provinsi: Sumatra Barat, Kepulauan Riau, Aceh, Kalimantan Barat, Kalimantan Selatan, Papua, Jambi dan Bali; sedangkan *cluster* kedua beranggotakan provinsi: Kalimantan Timur, Riau, Sumatera Utara dan Banten. *Cluster* ketiga beranggotakan provinsi: Sulawesi Selatan, Sumatera Selatan, dan Lampung; *Cluster* keempat beranggotakan provinsi: Papua Barat, Kalimantan Utara, Kepulauan Bangka Belitung, NTT, Bengkulu, Maluku Sulawesi Barat, Maluku Utara, Gorontalo, Sulawesi Tengah, D.I. Yogyakarta, Sulawesi Tenggara, Kalimantan Tengah, Sulawesi Utara dan NTB; Cluster kelima beranggotakan DKI Jakarta, Jawa barat dan Jawa Timur dan cluster keenam merupakan provinsi Jawa Tengah. Untuk lebih

jelasan jumlah masing-masing cluster disajikan pada tabel 2 di bawah.



Gambar 2 Dendrogram Penggerombolan dengan Dynamic Time Warping (DTW)

Tabel 3. Jumlah Masing-Masing Cluster (Jarak DTW)

Kluster	Jumlah
1	8
2	4
3	3
4	15
5	3
6	1

Ukuran Ketepatan Banyaknya Gerombol

Untuk perhitungan validitas kluster akan digunakan parameter koefisien *Silhouette software*. Berdasarkan Tabel 4 dapat dilihat bahwa penggerombolan dengan jarak Euclid memiliki nilai koefisien *Silhouette* sebesar 0,605 sedangkan penggerombolan dengan jarak DTW menghasilkan koefisien *Silhouette* sebesar 0,68. Baik dengan metode jarak Euclid maupun dengan metode DTW menghasilkan klasifikasi penggerombolan dengan kategori *good classifications*. Jika dilihat dari besarnya nilai *Silhouette* maka metode DTW memberikan performa yang lebih baik.

Tabel 4. Koefisien Silhouette

Metode	Koefisien Silhoutte
Euclidian	0.605068
DTW	0.683957

KESIMPULAN

Analisis *clustering* data *time series* dengan menggunakan jarak DTW ataupun *Euclidian* menghasilkan enam *cluster* dengan jumlah anggota yang berbeda untuk masing - masing *clusternya* antara jarak Euclid dan DTW. Penggerombolan yang dibentuk menghasilkan nilai koefisien Silhouette untuk jarak Euclid dan DTW masing-masing sebesar 0,61 dan 0,68. Nilai tersebut menandakan bahwa gerombol yang dibentuk berada dalam kategori *good classification*. Jadi baik metode penggerombolan dengan jarak Euclid maupun dengan metode *dynamic time warping* dapat digunakan dalam penggerombolan untuk studi kasus penggerombolan provinsi-provinsi di Indonesia berdasarkan nilai PDRB.

DAFTAR PUSTAKA

- Artha, K. S., & Winarko, E. (2016). Perbandingan Eros, Euclidean Distance dan Dynamite Time Warping dalam Klasifikasi Data Multivariate Time Series Menggunakan KNN. *Prosiding Seminar Nasional Pendidikan Teknik Informatika*, (pp. 223-228).
- Berndt, D., & Clifford, J. (1994). Using Dynamic Time Warping to Find Patterns in Time Series. *Knowledge Discovery in Databases Workshop*, (pp. 359-370).
- Johnson, A., & Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Edisi ke-5. New York: Prentice-Hall, Inc.
- Kaufman, L., & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey (US): : John Wiley and Sons Inc.
- Liao, T. (2005). Pattern Recognition. In .. Liao, *Clustering Of Time Series Data Survey* (pp. 1857-1874).
- Munthe, A. (2019). Penerapan Clustering Time Series untuk Menggerombolkan Provinsi di Indonesia Berdasarkan Nilai Produksi Padi. *Jurnal Litbang Sukowati*, 2(2), 1-11.
- Sahputra, A. (2020). Clustering Time Series Analysis untuk Mengelompokkan Kabupaten/Kota di Indonesia Berdasarkan Nilai Inflasi. *ETD Unsyiah Universitas Syiah Kuala Banda Aceh*.

