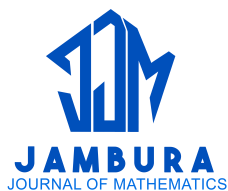


Pengelompokan Provinsi di Indonesia Menggunakan *Time Series Clustering* pada Sektor Ekspor Nonmigas

Aulia Nabila Putri, Neva Satyahadewi, dan Siti Aprizkiyandari



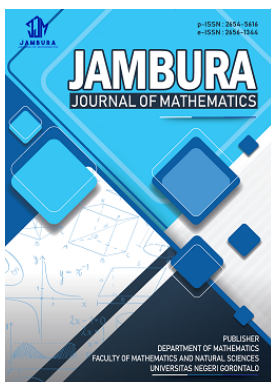
Volume 6, Issue 1, Pages 16–22, February 2024

Submit 5 September 2023, Direvisi 3 November 2023, Disetujui 12 November 2023

To Cite this Article : A. N. Putri, N. Satyahadewi, dan S. Aprizkiyandari, "Pengelompokan Provinsi di Indonesia Menggunakan *Time Series Clustering* pada Sektor Ekspor Nonmigas", *Jambura J. Math*, vol. 6, no. 1, pp. 16–22, 2024, <https://doi.org/10.37905/jjom.v6i1.21921>

© 2024 by author(s)

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS

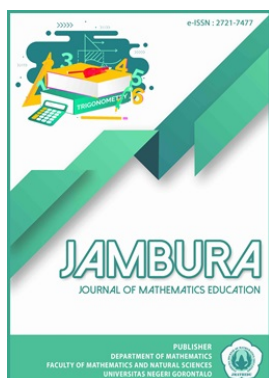


	Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
	Journal Abbreviation	:	Jambura J. Math.
	Frequency	:	Biannual (February and August)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/jjom
	Online ISSN	:	2656-1344
	Editor-in-Chief	:	Hasan S. Panigoro
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
	Google Scholar ID	:	iWLjgaUAAAAJ
	Email	:	info.jjom@ung.ac.id

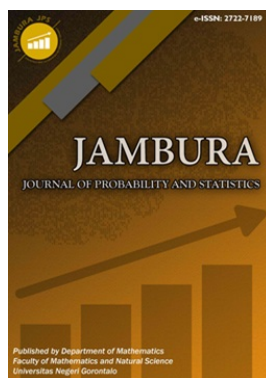
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



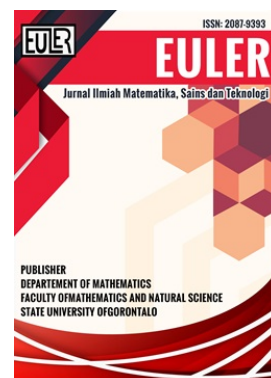
Jambura Journal of Biomathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi

Pengelompokan Provinsi di Indonesia Menggunakan *Time Series Clustering* pada Sektor Ekspor Nonmigas

Aulia Nabila Putri¹, Neva Satyahadewi^{1,*} , dan Siti Aprizkiyandari² 

¹Program Studi Statistika, Universitas Tanjungpura, Indonesia

²Program Studi Agroteknologi, Universitas Tanjungpura, Indonesia

ARTICLE HISTORY

Submit 5 September 2023
Direvisi 3 November 2023
Disetujui 12 November 2023

KATA KUNCI

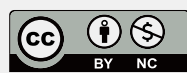
Time Series clustering
Dynamic Time Warping
Nonmigas

KEYWORDS

Time Series clustering
Dynamic Time Warping
Non-oil and Gas

ABSTRAK. Kegiatan ekspor Indonesia didominasi oleh ekspor nonmigas yang terdiri dari empat sektor, yaitu industri pengolahan, pertanian, pertambangan, dan lainnya. Pemerintah harus memperhatikan ekspor nonmigas untuk setiap provinsi karena ekspor dapat memainkan peran penting dalam pertumbuhan ekonomi suatu negara. Penelitian ini dilakukan untuk mengelompokkan provinsi di Indonesia menggunakan *time series clustering* pada sektor ekspor nonmigas berdasarkan pola data dengan memperhatikan jarak *Dynamic Time Warping (DTW)*. Sektor yang digunakan pada penelitian ini berupa sektor industri pengolahan dan sektor pertanian di 34 provinsi Indonesia pada periode 2017 – 2021. Analisis *time series clustering* menggunakan metode *average linkage* dengan jarak *DTW*, dan pemilihan jumlah cluster optimum menggunakan metode *silhouette coefficient*. Hasil analisis pada sektor industri pengolahan menghasilkan 3 cluster optimum, yaitu cluster 1 beranggotakan 1 provinsi yang memiliki ekspor industri pengolahan tinggi, cluster 2 beranggotakan 8 provinsi yang memiliki ekspor industri pengolahan sedang, dan cluster 3 beranggotakan 25 provinsi yang memiliki ekspor industri pengolahan rendah. Adapun pada sektor pertanian menghasilkan 2 cluster optimum, yaitu cluster 1 beranggotakan 5 provinsi yang memiliki ekspor industri pertanian tinggi dan cluster 2 beranggotakan 29 provinsi yang memiliki ekspor industri pertanian rendah. Hasil pengelompokan pada sektor industri pengolahan dan sektor pertanian mempunyai nilai *silhouette coefficient* sebesar 0,778 dan 0,798, sehingga dikatakan mempunyai struktur cluster yang kuat.

ABSTRACT. Indonesia's export activities are dominated by non-oil and gas exports consisting of four sectors, namely the processing industry, agriculture, mining, and others. The government must pay attention to non-oil and gas exports for each province because exports can play an essential role in a country's economic growth. This study was conducted to cluster provinces in Indonesia using *time series clustering* in the non-oil and gas export sector based on data patterns concerning *Dynamic Time Warping (DTW)* distance. The sectors used in this study are the manufacturing industry sector and the agricultural sector in 34 Indonesian provinces in the period 2017 - 2021. *Time series clustering* analysis uses the *average linkage* method with *DTW* distance and the selection of the optimum number of clusters using the *silhouette coefficient* method. The results of the analysis in the processing industry sector resulted in 3 optimum clusters, namely cluster 1 consisting of 1 province that has high processing industry exports, cluster 2 consisting of 8 provinces that have medium processing industry exports, and cluster 3 consisting of 25 provinces that have low processing industry exports. As for the agricultural sector, it produces 2 optimum clusters, namely cluster 1 consisting of 5 provinces that have high agricultural industry exports, and cluster 2 consisting of 29 provinces that have low agricultural industry exports. The clustering results in the processing industry sector and the agricultural sectors have a *silhouette coefficient* value of 0.778 and 0.798, so it is said to have a strong cluster structure.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of JJBM:* Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Ekspor merupakan kegiatan mengeluarkan barang dari wilayah suatu negara, baik secara komersial atau nonkomersial, dan barang yang diolah di luar negeri serta hasilnya dimasukkan kembali secara sah ke negara tersebut. Indonesia ialah salah satu negara yang tengah berjuang dalam meningkatkan kualitas dan kuantitas agar nilai ekspor dapat lebih laku di pasar internasional. Ekspor Indonesia didominasi oleh ekspor nonmigas yang terdiri dari empat sektor, yaitu industri pengolahan, pertanian,

pertambangan, dan lainnya [1].

Ekspor memegang peran penting untuk pertumbuhan ekonomi suatu negara. Ekspor dalam jangka panjang memiliki pengaruh untuk meningkatkan pertumbuhan ekonomi karena adanya pertumbuhan industri dalam negeri [2]. Pemerintah harus memperhatikan ekspor nonmigas untuk setiap provinsi karena ekspor nonmigas memiliki kontribusi dalam meningkatkan pertumbuhan ekonomi Indonesia. Salah satu cara yang dapat dilakukan adalah menganalisis data ekspor nonmigas tiap provinsi. Analisis yang dapat digunakan pada kasus ini ialah analisis *cluster*.

*Penulis Korespondensi.

Analisis *cluster* dilakukan guna mempermudah pemerintah agar lebih fokus dan terarah dalam melaksanakan kebijakan terhadap provinsi yang berada pada *cluster* yang sama. Analisis *cluster* (*cluster analysis*) ialah analisis statistik guna mengelompokkan objek-objek yang memiliki kesamaan menjadi satu *cluster* [3]. Secara umum analisis *cluster* diaplikasikan pada data *cross section*. Penelitian terkait metode *K-Means Clustering* menggunakan data *cross section* untuk mengelompokkan wilayah di Kabupaten Sidoarjo pada tahun 2018 berdasarkan Metode Kontrasepsi Jangka Panjang (MKJP) yang memperoleh jumlah *cluster* optimum sebanyak 3 *cluster* [4]. Analisis *cluster* juga dapat diaplikasikan pada data *time series*. Penelitian sektor ekspor nonmigas menambahkan unsur *time series* pada analisis *cluster*. Hal ini disebabkan penambahan unsur *time series* dapat memberikan informasi yang lebih kaya dibandingkan dengan analisis *cluster* yang menggunakan data *cross section* [5].

Analisis *cluster* yang diaplikasikan pada data *time series* dikenal dengan analisis *time series clustering*. Analisis *time series clustering* memiliki perbedaan dalam tata cara pengelompokan dan penggunaan algoritma dibandingkan data *cross section* [6]. Data *time series* ialah sekumpulan observasi yang terjadi secara berurutan pada periode waktu tertentu. Analisis *time series clustering* dilakukan dengan cara mengelompokkan objek-objek berdasarkan pola data dengan tetap memperhatikan pemilihan jarak serta metode pengelompokan [7]. Analisis *time series clustering* dinilai lebih efektif karena analisis tersebut menggunakan lebih dari satu periode [8]. Pengukuran jarak pada analisis *time series clustering* salah satunya menggunakan jarak *Dynamic Time Warping* (DTW). DTW merupakan sebuah algoritma untuk menghitung jarak antara dua *time series* dengan menemukan jalur pembengkokan yang optimum (*optimal warping path*) [7]. DTW digunakan untuk mengukur kesamaan antar objek penelitian dalam analisis *time series clustering*.

Penelitian terkait analisis *time series clustering* menggunakan jarak DTW telah banyak dilakukan. Hal ini dibuktikan melalui penelitian yang mengaplikasikan analisis *time series clustering* menggunakan jarak DTW untuk membangun portofolio saham yang optimum [9]. Penelitian mengenai ekspor komoditas migas dan nonmigas pada Provinsi Kalimantan Timur dengan membandingkan empat ukuran jarak, yaitu *Pearson correlation*, *Euclidean*, DTW, dan *Autocorrelation Based Distance* [10]. Hasil dari penelitian tersebut adalah penggunaan jarak terbaik menggunakan DTW dan menghasilkan jumlah *cluster* optimum sebanyak 3 *cluster*. Penelitian lainnya terkait pengelompokan kasus aktif covid-19 di Indonesia dengan mengaplikasikan analisis *time series clustering* dan memperoleh jumlah *cluster* optimum sebanyak 3 *cluster* [11].

Berdasarkan latar belakang tersebut, penelitian ini membahas analisis *time series clustering* menggunakan jarak DTW pada sektor ekspor nonmigas di Indonesia. Penelitian ini menggunakan jarak DTW karena dari beberapa rujukan menunjukkan bahwa jarak DTW menghasilkan kualitas pengelompokan yang cukup baik untuk diterapkan pada analisis *time series clustering*. Adapun sektor ekspor nonmigas yang digunakan berupa industri pengolahan dan pertanian. Hal ini dikarenakan ekspor nonmigas di Indonesia didominasi oleh sektor industri pengolahan [12]. Pada periode 2017 – 2021, sektor industri pengolahan berkontribusi sebesar 81,74% terhadap ekspor nonmigas. Selain sektor industri pengolahan, terdapat sektor lainnya yang berkontribusi

dalam meningkatkan ekspor yaitu sektor pertanian. Kementerian Pertanian menyatakan bahwa sektor pertanian memiliki potensi dalam meningkatkan ekspor di Indonesia [13]. Pada saat pandemi Covid-19 sektor pertanian terus mengalami peningkatan ekspor. Tujuan penelitian adalah mengelompokkan 34 provinsi di Indonesia pada sektor ekspor nonmigas menggunakan analisis *time series clustering* berdasarkan pola data dengan memperhatikan jarak DTW. Hasil dari penelitian ini diharapkan dapat membantu pemerintah dalam mengambil keputusan terkait kebijakan peningkatan ekspor sektor nonmigas terhadap provinsi yang berada pada *cluster* yang sama. Pemerintah dapat menggunakan informasi dari hasil penelitian untuk memahami pergerakan ekspor nonmigas pada setiap *cluster*.

2. Metode

Penelitian ini menggunakan data terbitan Badan Pusat Statistik. Data tersebut berupa data ekspor nonmigas pada sektor industri pengolahan dan pertanian di 34 provinsi Indonesia periode 2017 – 2021. Data tersebut memiliki satuan Juta US\$. Tahapan pada analisis *cluster* sebagai berikut:

1. Melakukan pengumpulan data.
2. Melakukan analisis deskriptif pada data sektor ekspor nonmigas.
3. Menghitung jarak antar objek menggunakan jarak *Dynamic Time Warping* (DTW). Jarak DTW digunakan untuk mengukur kesamaan antar objek yang nantinya diperlukan dalam proses pembentukan *cluster*.
4. Membentuk *cluster* dengan menggabungkan objek-objek berdasarkan metode *average linkage*. *Average linkage* ialah metode pengelompokan berdasarkan rata-rata jarak seluruh objek dalam satu *cluster* dengan rata-rata jarak seluruh objek pada *cluster* yang lain.
5. Menentukan jumlah *cluster* optimum berdasarkan *silhouette coefficient*.
6. Merepresentasikan karakteristik dari setiap *cluster* yang terbentuk.

2.1. Analisis Cluster

Analisis *cluster* ialah pengelompokan sejumlah objek yang mempunyai banyak kesamaan menjadi satu *cluster*, sedangkan objek yang memiliki banyak perbedaan menjadi anggota pada *cluster* yang lain [3]. Tujuan analisis *cluster* yaitu mengelompokkan objek-objek ke dalam *cluster* tertentu berdasarkan kesamaan karakteristik. Suatu *cluster* dapat dikatakan baik jika memiliki kesamaan (homogenitas) yang besar pada anggota dalam satu *cluster* dan perbedaan (heterogenitas) yang besar antar *cluster* [14]. Metode untuk analisis *cluster* mencakup metode hirarki dan tak berhirarki.

Metode hirarki merupakan suatu metode pengelompokan terhadap objek-objek yang memiliki kesamaan paling dekat menjadi satu *cluster* [14]. Analisis *cluster* dilakukan hingga membentuk pohon atau hirarki antar objek. Hasil pengelompokan metode hirarki dapat disajikan dalam bentuk *dendrogram*. Metode hirarki dilakukan apabila banyaknya jumlah *cluster* yang terbentuk tidak diketahui sebelumnya. Metode hirarki terbagi menjadi dua teknik pengelompokan, yakni *divisive* dan *agglomerative*. Salah satu metode hirarki *agglomerative* adalah metode *linkage*. Metode *linkage* terdiri dari tiga metode, yakni *single linkage*, *complete link-*

age, dan average linkage.

2.2. Metode Average Linkage

Average linkage ialah metode pengelompokan dengan menggunakan jarak rata-rata antar objek [3]. Prosedur pengelompokan metode average linkage hampir serupa dengan metode single linkage ataupun complete linkage, tetapi terdapat perbedaan dalam proses penggabungan objek berdasarkan rata-rata jarak [15]. Misalkan d_{UV} merupakan jarak dari cluster U hingga cluster V , maka jarak yang digunakan dari cluster UV hingga cluster W adalah seperti Persamaan (1).

$$d_{(UV)W} = \frac{d_{UW} + d_{VW}}{N_{(UV)}N_W} \quad (1)$$

dengan, $N_{(UV)}$ ialah jumlah objek pada cluster UV dan N_W ialah jumlah objek pada cluster W .

2.3. Jarak Dynamic Time Warping (DTW)

DTW digunakan untuk menghitung jarak dengan mencari jalur pembengkokan yang optimum (optimal warping path) diantara dua time series [7]. Jarak DTW akan memilih salah satu jalur yang memberikan jarak terkecil diantara dua time series. Misalkan terdapat dua time series, yaitu $x_t = x_1, x_2, \dots, x_n$ dan $y_t = y_1, y_2, \dots, y_n$. Penentuan jalur pembengkokan yang optimum menggunakan cost matrix C berukuran $n \times n$ dengan setiap elemen (i, j) untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, n$. n merupakan banyaknya waktu pengamatan objek dengan $t = 1, 2, \dots, n$. Perhitungan cost matrix C menggunakan persamaan berikut [16].

Pada elemen $(1, 1)$ cost matrix C dapat dihitung dengan menggunakan Persamaan (2).

$$c_{1,1} = w_{i,j} = w_{1,1} \quad (2)$$

dengan $w_{i,j} = |x_i - y_j|$

Akan tetapi, untuk elemen lainnya (i, j) dari cost matrix C dapat dihitung dengan menggunakan Persamaan (3).

$$c_{i,j} = \begin{cases} w_{i,j} + c_{i,j-1} & , i = 1 \\ w_{i,j} + c_{i-1,j} & , j = 1 \\ w_{i,j} + \min\{c_{i-1,j}, c_{i,j-1}, c_{i-1,j-1}\} & , \text{lainnya} \end{cases} \quad (3)$$

Keterangan,

$c_{i,j}$: Elemen (i, j) dari cost matrix C

$w_{i,j}$: Selisih antara x_i terhadap y_j .

Jalur pembengkokan optimum pada cost matrix C dapat diperoleh dari elemen (n, n) ke elemen $(1, 1)$ dengan mempertimbangkan nilai minimum dari $c_{i-1,j}$, $c_{i,j-1}$, $c_{i-1,j-1}$ pada Persamaan (3) [16]. Setelah menemukan jalur pembengkokan yang optimum, maka tahap selanjutnya adalah menentukan jarak DTW. Jarak DTW dapat diperoleh pada Persamaan (4).

$$d_{DTW} = c_{n,n} \quad (4)$$

Keterangan,

d_{DTW} : Jarak DTW

$c_{n,n}$: Elemen (n, n) dari cost matrix C .

2.4. Silhouette Coefficient

Validasi cluster digunakan untuk mengetahui ketepatan dan kualitas dari hasil analisis cluster [17]. Selain itu, jumlah cluster optimum dapat diperoleh melalui validasi cluster [18]. Setiap objek dapat membentuk cluster sebanyak $2 \leq K \leq (N - 1)$ dengan N merupakan jumlah objek dalam pengamatan [19]. Salah satu validasi cluster yang dapat digunakan ialah silhouette coefficient. Silhouette coefficient merupakan gabungan dari metode kohesi dan separasi untuk menentukan kebaikan objek pada suatu cluster [20]. Berikut ini perhitungan silhouette coefficient [21].

1. Menghitung rata-rata jarak objek ke- p ke semua objek yang berada dalam satu cluster yang sama menggunakan Persamaan (5).

$$a_p = \frac{1}{N_k - 1} \sum_{r=1}^{N_k-1} d(x_p, x_r), p \neq r \quad (5)$$

Keterangan,

a_p : Rata-rata jarak objek ke- p ke semua objek yang berada di dalam satu cluster

N_k : Jumlah objek pada cluster ke- k dengan $k = 1, 2, \dots, K$

$d(x_p, x_r)$: Jarak objek ke- p dengan objek ke- r dalam satu cluster.

2. Menghitung rata-rata jarak objek ke- p ke semua objek yang berada pada cluster yang berbeda menggunakan Persamaan (6), lalu diambil nilai terkecil menggunakan Persamaan (7).

$$b_p = \min\{d_p(k)\} \quad (6)$$

Jarak objek ke- p ke semua objek yang berada pada cluster yang berbeda menggunakan Persamaan (7).

$$d_p(k) = \frac{1}{N_k} \sum_{c=1}^{N_k} d(x_p, x_q) \quad (7)$$

Keterangan,

b_p : Rata-rata jarak objek ke- p ke semua objek yang berada pada cluster berbeda dengan nilai terkecil

$d_p(k)$: Jarak objek ke- p ke semua objek yang berada pada cluster ke- k yang berbeda

N_k : Jumlah objek pada cluster ke- k dengan $k = 1, 2, \dots, K$

$d(x_p, x_q)$: Jarak objek ke- p dengan objek ke- q pada cluster lainnya.

3. Menghitung nilai silhouette coefficient untuk setiap objek ke- p menggunakan Persamaan (8).

$$SC_1(p) = \frac{b_p - a_p}{\max\{a_p, b_p\}} \quad (8)$$

dengan $SC_1(p)$ merupakan nilai silhouette coefficient objek ke- p pada cluster k . Setelah mendapatkan nilai silhouette coefficient untuk setiap objek, maka tahap selanjutnya adalah menghitung rata-rata nilai silhouette coefficient dalam satu cluster menggunakan Persamaan (9).

$$SC_2(k) = \frac{1}{N_k} \sum_{i=1}^{N_k} SC_1(p) \quad (9)$$

dengan $SC_2(k)$ merupakan rata-rata nilai *silhouette coefficient* pada cluster k .

- Menghitung nilai *silhouette coefficient* dari rata-rata nilai *silhouette coefficient* setiap cluster menggunakan Persamaan (10).

$$SC = \frac{\sum_{k=1}^K (N_k \cdot SC_2(k))}{\sum_{k=1}^K N_k} \tag{10}$$

Tabel 1. Deskripsi *silhouette coefficient*

Nilai <i>Silhouette Coefficient</i>	Deskripsi
0,71 – 1,00	Struktur kuat
0,51 – 0,70	Struktur baik
0,26 – 0,50	Struktur lemah
≤ 0,25	Struktur buruk

Nilai *silhouette coefficient* menunjukkan ketepatan objek dalam suatu cluster dan kualitas dari hasil analisis cluster. *Silhouette coefficient* memiliki rentang nilai, yaitu $-1 \leq SC \leq 1$ [22]. Jika nilai *silhouette coefficient* menuju 1, maka objek-objek memiliki jarak yang sangat padat pada satu cluster dan memiliki jarak yang jauh dari cluster lainnya [23]. Validasi cluster menggunakan metode *silhouette coefficient* merujuk pada [24] yang selanjutnya disajikan pada Tabel 1.

3. Hasil dan Pembahasan

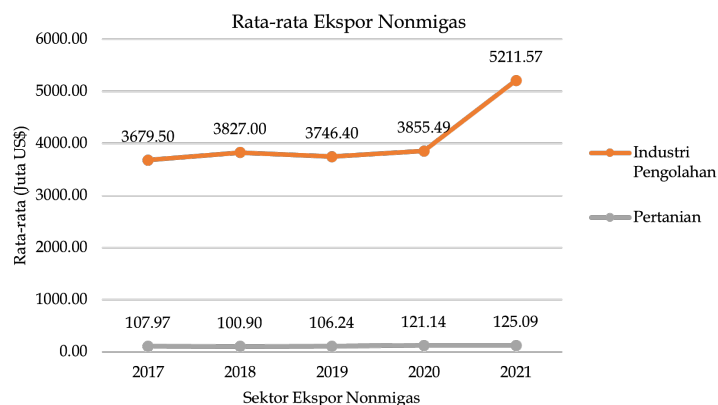
3.1. Statistik Deskriptif

Statistik deskriptif untuk sektor industri pengolahan dan sektor pertanian disajikan pada Gambar 1. Sektor industri pengolahan memperoleh pertumbuhan selama lima tahun terakhir, namun pernah mengalami penurunan pada tahun 2019 meskipun tidak signifikan. Pada saat pandemi Covid-19 kinerja sektor industri pengolahan mengalami peningkatan. Peningkatan tersebut didominasi oleh industri makanan, industri logam, industri kimia dan produk kimia, serta produk kertas [25]. Pelonggaran kebijakan *lockdown* di berbagai negara menjadi kesempatan untuk Indonesia dalam mengekspor Alat Pelindung Diri (APD). Adanya kerja sama yang baik dari pihak pemerintah pusat, pemerintah daerah, kementerian, dan pelaku industri untuk dapat melewati krisis pandemi, sehingga pada tahun 2021 rata-rata ekspor industri pengolahan di Indonesia sebesar US\$ 5211,57 juta. Industri pengolahan berkontribusi sebesar 80,78% terhadap total ekspor nonmigas pada tahun 2021 [12].

Sektor pertanian tidak mengalami perubahan yang signifikan dari tahun ke tahun, seperti terlihat pada Gambar 1. Pada saat pandemi Covid-19 ekspor pertanian tetap tumbuh konsisten. Peningkatan ekspor pertanian di tengah pandemi Covid-19 didominasi oleh komoditas buah-buahan karena adanya peningkatan permintaan dari luar negeri [26]. Rata-rata kontribusi ekspor pertanian menyumbang 2,28% dari total ekspor nonmigas pada periode 2017 – 2021 [12].

3.2. Perhitungan Jarak Dynamic Time Warping

Sebelum melakukan proses pengelompokan, terlebih dahulu dilakukan perhitungan kesamaan. Perhitungan kesamaan dapat dilakukan dengan menghitung ukuran jarak antar objek. Ukuran jarak yang digunakan pada analisis *time series clustering* adalah jarak DTW. Perhitungan jarak DTW melewati beberapa



Gambar 1. Grafik sektor ekspor nonmigas

proses, yaitu menghitung *cost matrix*, menemukan jalur pembengkokan optimum, dan menentukan jarak DTW berdasarkan Persamaan (2), (3), dan (4). Proses perhitungan jarak DTW dilakukan pada setiap objek penelitian, sehingga membentuk *matrix* jarak. Analisis *time series clustering* dilakukan pada masing-masing sektor ekspor nonmigas, yaitu sektor industri pengolahan dan sektor pertanian. Berikut ini *matrix* yang terbentuk dari perhitungan jarak DTW pada sektor industri pengolahan,

$$C_{34 \times 34} = \begin{bmatrix} 0 & \dots & 148041,8 \\ \vdots & \ddots & \vdots \\ 148041,8 & \dots & 0 \end{bmatrix},$$

sedangkan pada sektor pertanian, terbentuk *matrix* jarak DTW sebagai berikut,

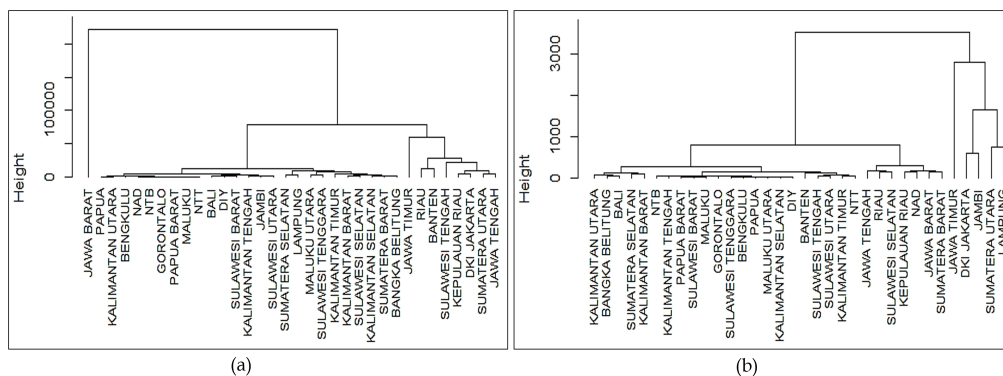
$$D_{34 \times 34} = \begin{bmatrix} 0 & \dots & 563,6 \\ \vdots & \ddots & \vdots \\ 563,6 & \dots & 0 \end{bmatrix}.$$

Setelah diperoleh jarak dari setiap objek penelitian baik sektor industri pengolahan maupun sektor pertanian, maka tahap selanjutnya adalah mengelompokkan 34 provinsi di Indonesia. Pengelompokan 34 provinsi pada penelitian ini menggunakan metode *average linkage*. Proses pengelompokan berdasarkan jarak rata-rata antar objek.

3.3. Hasil Analisis Time Series Clustering

Berdasarkan hasil pengelompokan menggunakan metode *average linkage* dengan jarak DTW, maka diperoleh dendrogram untuk setiap sektor pada Gambar 2.

Gambar 2 menunjukkan dendrogram hasil analisis *time series clustering* pada sektor industri pengolahan dan pertanian. Proses penggabungan objek dilakukan hingga semua objek membentuk suatu cluster tunggal. Setiap objek pada masing-masing sektor dapat membentuk cluster yang berjumlah 2 hingga 33 cluster. Jumlah cluster optimum yang terbentuk belum diketahui. Oleh karena itu, diperlukan validasi cluster untuk menentukan jumlah cluster optimum. Validasi cluster yang digunakan berupa metode *silhouette coefficient*. Validasi cluster perlu dilakukan baik pada sektor industri pengolahan maupun pertanian.



Gambar 2. Dendrogram pada (a) sektor industri pengolahan dan (b) sektor pertanian

3.4. Validasi Hasil Analisis Time Series Clustering

Validasi cluster dilakukan untuk mengetahui jumlah cluster optimum, ketepatan, dan kualitas dari hasil analisis cluster. Jumlah cluster optimum dapat ditentukan menggunakan metode silhouette coefficient pada cluster yang berjumlah 2 hingga 5. Hal ini dikarenakan silhouette coefficient yang dihasilkan pada cluster yang berjumlah 6 hingga 33 memiliki nilai yang semakin kecil. Nilai silhouette coefficient pada sektor industri pengolahan dan sektor pertanian ditampilkan pada Tabel 2.

Tabel 2. Nilai silhouette coefficient

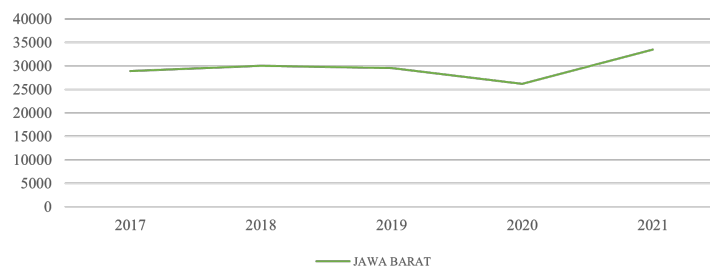
Jumlah Cluster	Sektor Industri Pengolahan	Sektor Pertanian
2	0,775	0,798
3	0,778	0,742
4	0,738	0,670
5	0,713	0,671

Jumlah cluster optimum pada sektor industri pengolahan adalah 3 cluster yang memiliki nilai silhouette coefficient sebesar 0,778. Jumlah cluster optimum pada sektor pertanian adalah 2 cluster yang memiliki nilai silhouette coefficient sebesar 0,798. Hasil pengelompokan untuk sektor industri pengolahan maupun sektor pertanian memiliki struktur cluster yang kuat. Terdapat perbedaan jumlah cluster pada sektor industri pengolahan dan sektor pertanian. Oleh karena itu, perlu diketahui karakteristik dari setiap cluster yang terbentuk.

3.5. Karakteristik Cluster pada Sektor Industri Pengolahan

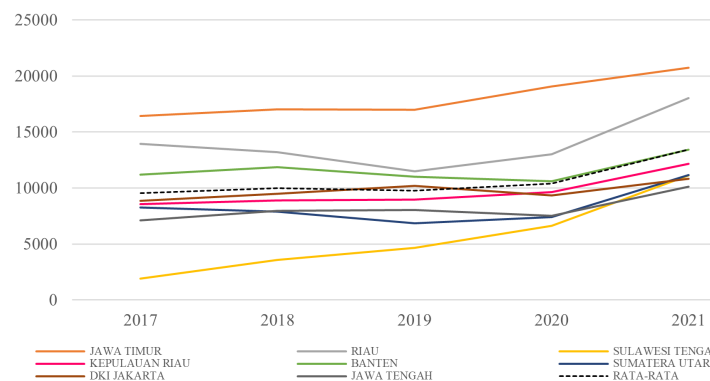
Tahap selanjutnya ialah menentukan karakteristik setiap cluster yang terbentuk. Penentuan karakteristik dapat memberikan gambaran umum mengenai provinsi yang tergabung pada masing-masing cluster. Berdasarkan perhitungan menggunakan silhouette coefficient, diperoleh jumlah cluster optimum pada sektor industri pengolahan sebanyak 3 cluster.

Cluster 1 hanya beranggotakan satu provinsi, yaitu Jawa Barat. Gambar 3 menunjukkan cluster 1 memiliki pergerakan ekspor yang landai di tahun 2019 – 2020. Namun, pada tahun 2021 mengalami peningkatan yang sangat signifikan yaitu sebesar US\$ 33495,6 Juta dibandingkan tahun sebelumnya yaitu sebesar US\$ 26218,5 Juta. Cluster 1 memiliki rata-rata ekspor sebesar US\$ 29640,40 Juta. Cluster 1 dikategorikan sebagai cluster yang memiliki rata-rata ekspor sektor industri pengolahan tertinggi. Oleh karena itu, pemerintah dapat memberikan fasilitas yang baik



Gambar 3. Cluster 1 pada sektor industri pengolahan

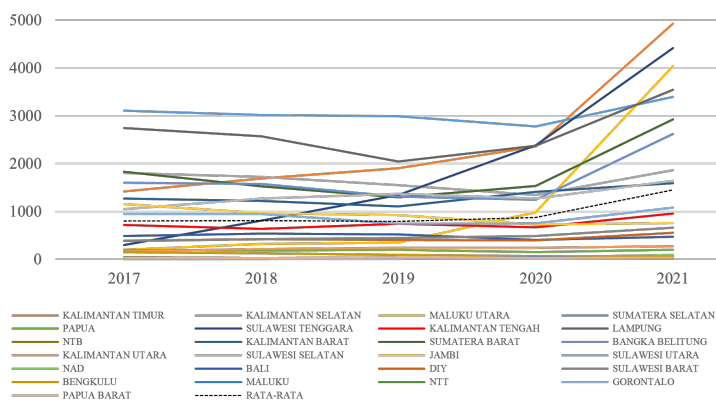
berupa modernisasi teknologi pada provinsi yang berada di cluster 1 untuk terus meningkatkan kualitas serta kuantitas ekspor.



Gambar 4. Cluster 2 pada sektor industri pengolahan

Adapun cluster 2 beranggotakan 8 provinsi, yaitu Jawa Timur, Riau, Sulawesi Tengah, Kepulauan Riau, Banten, Sumatera Utara, DKI Jakarta, dan Jawa Tengah. Gambar 4 menunjukkan provinsi yang berada di cluster 2 memiliki kecenderungan peningkatan ekspor pada tahun 2019 – 2021. Berdasarkan nilai rata-rata ekspor pada tahun 2019 sebesar US\$ 9765,84 Juta dan terus mengalami peningkatan. Pada tahun 2021 rata-rata ekspor mencapai US\$ 13441,84 Juta. Cluster 2 dikategorikan sebagai cluster dengan rata-rata ekspor sektor industri pengolahan menengah. Cluster 2 memiliki potensi yang besar dalam meningkatkan ekspor pada sektor industri pengolahan di Indonesia.

Adapun cluster 3 merupakan cluster dengan anggota terbanyak. Cluster 3 beranggotakan 25 provinsi, yaitu Kalimantan Timur, Kalimantan Selatan, Maluku Utara, Sumatera Sela-

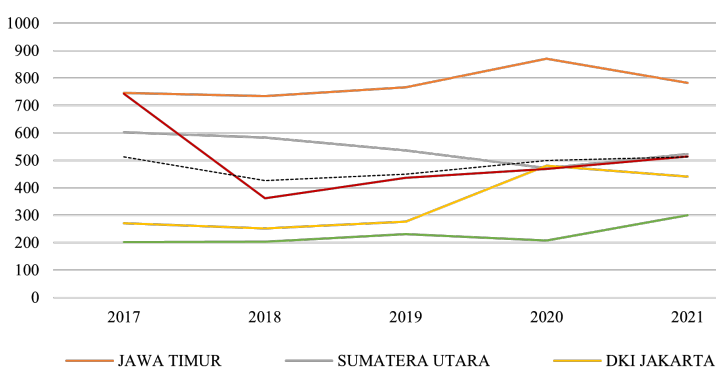


Gambar 5. Cluster 3 pada sektor industri pengolahan

tan, Papua, Sulawesi Tenggara, Kalimantan Tengah, Lampung, Nusa Tenggara Barat, Kalimantan Barat, Sumatera Barat, Bangka Belitung, Kalimantan Utara, Sulawesi Selatan, Jambi, Sulawesi Utara, Nanggroe Aceh Darussalam, Bali, Daerah Istimewa Yogyakarta, Sulawesi Barat, Bengkulu, Maluku, Nusa Tenggara Timur, Gorontalo, dan Papua Barat. Cluster 3 didominasi oleh provinsi yang berada di luar pulau Jawa dan hanya terdapat 1 provinsi yang berasal dari pulau Jawa, yaitu Daerah Istimewa Yogyakarta. Gambar 5 memberikan informasi bahwa provinsi yang berada di cluster 3 cenderung tidak mengalami perubahan yang signifikan di tahun 2017 – 2019. Akan tetapi, pada tahun 2020 – 2021 terdapat beberapa provinsi yang mengalami peningkatan ekspor yang signifikan. Berdasarkan rata-rata ekspor setiap tahunnya menunjukkan bahwa cluster 3 tidak mengalami perubahan yang signifikan di tahun 2017 – 2020. Rata-rata ekspor cluster 3 hanya sebesar US\$ 799,88 juta di tahun 2017. Namun, pada tahun 2021 terdapat lonjakan ekspor sebesar US\$ 1446,528 juta. Cluster 3 dikategorikan sebagai cluster dengan rata-rata ekspor sektor industri pengolahan terendah. Pemerintah dapat melakukan upaya peningkatan ekspor dengan memperbaiki pemerataan pembangunan yang belum maksimal di luar pulau Jawa dan menggiatkan program peningkatan kualitas SDM.

3.6. Karakteristik Cluster pada Sektor Pertanian

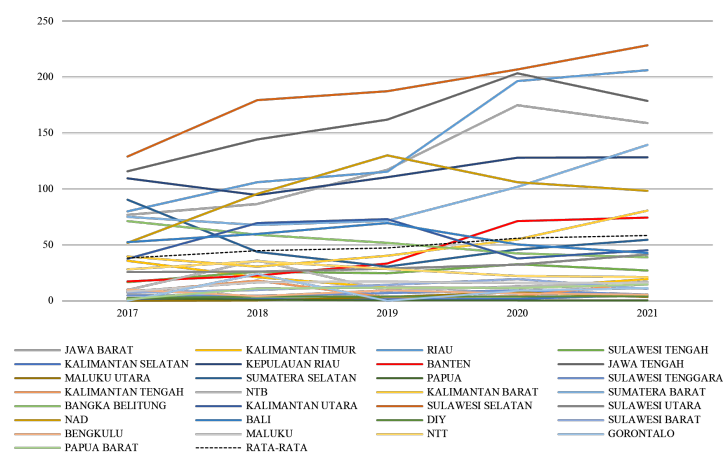
Setelah dilakukan perhitungan *silhouette coefficient*, maka tahap selanjutnya adalah menentukan karakteristik dari cluster yang terbentuk. Penentuan karakteristik dapat memberikan gambaran umum mengenai provinsi yang tergabung pada masing-masing cluster. Berdasarkan nilai *silhouette coefficient* diperoleh jumlah cluster optimum pada sektor pertanian sebanyak 2 cluster.



Gambar 6. Plot cluster 1 sektor pertanian

Cluster 1 beranggotakan 5 provinsi, yaitu Jawa Timur, Sumatera Utara, DKI Jakarta, Lampung, dan Jambi. Provinsi yang berada di cluster 1 berasal dari pulau Jawa dan Sumatera. Berdasarkan rata-rata ekspor sektor pertanian pada cluster 1 mengalami penurunan di tahun 2018 yaitu dari US\$ 512,52 juta menjadi US\$ 426,46 juta. Akan tetapi, pada tahun 2019 – 2021 selalu mengalami kenaikan. Nilai ekspor sektor pertanian pada cluster 1 lebih tinggi dibandingkan cluster 2. Oleh karena itu, cluster 1 dikategorikan sebagai cluster dengan rata-rata ekspor sektor pertanian tinggi. Untuk terus meningkatkan kualitas dan kuantitas ekspor, maka pemerintah dapat melakukan upaya dalam mengembangkan teknologi dibidang pertanian.

Gambar 6 dan Gambar 7 menunjukkan bahwa baik cluster 1 maupun cluster 2 memiliki pola data yang beragam. Adanya perbedaan pola data dengan anggota lainnya yang berada dalam satu cluster dikarenakan proses pengelompokan tidak hanya memperhatikan pola data, melainkan perlu memperhatikan penggunaan jarak DTW [11]. Jarak DTW akan mengelompokkan objek-objek dalam satu cluster yang memiliki nilai jarak terkecil.



Gambar 7. Plot cluster 2 sektor pertanian

Cluster 2 beranggotakan 29 provinsi, yaitu Jawa Barat, Kalimantan Timur, Riau, Sulawesi Tengah, Kalimantan Selatan, Kepulauan Riau, Banten, Jawa Tengah, Maluku Utara, Sumatera Selatan, Papua, Sulawesi Tenggara, Kalimantan Tengah, Nusa Tenggara Barat, Kalimantan Barat, Sumatera Barat, Bangka Belitung, Kalimantan Utara, Sulawesi Selatan, Sulawesi Utara, Nanggroe Aceh Darussalam, Bali, Daerah Istimewa Yogyakarta, Sulawesi Barat, Bengkulu, Maluku, Nusa Tenggara Timur, Gorontalo, dan Papua Barat. Berdasarkan nilai rata-rata ekspor sektor pertanian pada cluster 2 menunjukkan peningkatan setiap tahunnya. Rata-rata ekspor sektor pertanian pada tahun 2017 sebesar US\$ 38,22 juta dan terus mengalami peningkatan ekspor hingga tahun 2021 menjadi US\$ 48,87 juta. Provinsi yang berada pada cluster 2 memiliki nilai ekspor sektor pertanian yang rendah. Oleh karena itu, cluster 2 dikategorikan sebagai cluster dengan rata-rata ekspor sektor pertanian rendah. Provinsi yang berada di cluster 2 memiliki potensi dalam meningkatkan ekspor sektor pertanian. Pemerintah dapat melakukan upaya berupa peningkatan keterampilan petani di bidang pertanian.

4. Kesimpulan

Pengelompokan provinsi di Indonesia pada sektor industri pengolahan menggunakan *time series clustering* menghasilkan 3 *cluster* optimum. *Cluster* 1 hanya beranggotakan 1 provinsi memiliki pola data yang mengalami penurunan di tahun 2019 – 2020. *Cluster* 2 beranggotakan 8 provinsi memiliki pola data yang cenderung mengalami peningkatan pada tahun 2019 – 2021. *Cluster* 3 beranggotakan 25 provinsi yang memiliki pola data yang cenderung tidak mengalami perubahan yang signifikan pada tahun 2017 - 2019. Berdasarkan nilai rata-rata ekspor sektor industri pengolahan dapat diurutkan dari yang tertinggi sampai terendah, yaitu *cluster* 1, *cluster* 2, dan *cluster* 3. Pengelompokan pada sektor industri pengolahan memiliki nilai *silhouette coefficient* sebesar 0,778. Adapun pengelompokan pada sektor pertanian menghasilkan 2 *cluster* optimum. *Cluster* 1 beranggotakan 5 provinsi dan *cluster* 2 beranggotakan 29. Baik *cluster* 1 maupun *cluster* 2 memiliki pola data yang beragam dalam satu *cluster*. *Cluster* yang terbentuk dapat diurutkan dari yang tertinggi sampai terendah menurut nilai rata-rata ekspor sektor pertanian, yaitu *cluster* 1 dan *cluster* 2. Pengelompokan pada sektor pertanian mempunyai nilai *silhouette coefficient* sebesar 0,798. Hasil pengelompokan untuk sektor industri pengolahan maupun sektor pertanian mempunyai struktur *cluster* yang kuat.

Berdasarkan kesimpulan yang diperoleh dari hasil penelitian menunjukkan bahwa kualitas pengelompokan sektor ekspor nonmigas sudah baik, namun pada penelitian ini hanya menggunakan satu variabel. Saran pada penelitian selanjutnya ialah menambahkan beberapa variabel dalam analisis *time series clustering* dengan menggunakan jarak DTW. Saran untuk pihak pemerintah berdasarkan hasil penelitian pada sektor ekspor nonmigas dengan kategori tinggi dan sedang ialah mengembangkan teknologi industri berupa modernisasi mesin peralatan, dan memberikan fasilitas infrastruktur yang baik, serta menguatkan peran Industri Kecil Menengah (IKM). Adapun untuk sektor ekspor nonmigas dengan kategori rendah ialah menggiatkan program peningkatan kualitas Sumber Daya Manusia (SDM) dan pemerataan infrastruktur pada daerah luar pulau Jawa.

Referensi

- [1] Badan Pusat Statistik, *Ekspor Menurut Provinsi Asal Barang Tahun 2019*. Jakarta: BPS RI, 2020.
- [2] I. A. Syaputra and L. T. Laut, "Determinasi Ekspor Indonesia Tahun 1990-2021," *Growth: Jurnal Ilmiah Ekonomi Pembangunan*, vol. 1, no. 2, pp. 22–42, 2022.
- [3] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.
- [4] C. P. Siwi and Y. Nurfirdaus, "K-Means Cluster Analysis of Sub-Districts In Sidoarjo Based On Long-Term Contraceptive Method," *JBK*, vol. 9, no. 2, p. 161, Oct. 2020, doi: [10.20473/jbk.v9i2.2020.161-170](https://doi.org/10.20473/jbk.v9i2.2020.161-170).
- [5] D. A. N. Sirodj, I. M. Sumertajaya, and A. Kurnia, "Analisis Clustering Time Series untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Manusia Jenis Kelamin Perempuan," *Statistika*, vol. 23, no. 1, pp. 29–37, doi: [10.29313/statistika.v23i1.2181](https://doi.org/10.29313/statistika.v23i1.2181).
- [6] L. P. W. Adnyani and P. R. Sihombing, "Analisis Cluster Time Series dalam Pengelompokan Provinsi di Indonesia Berdasarkan Nilai PDRB," *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*, vol. 1, no. 1, pp. 47–54, 2021, doi: [10.46306/bay.v1i1.5](https://doi.org/10.46306/bay.v1i1.5).
- [7] A. D. Munthe, "Penerapan Clustering Time Series Untuk Menggerombolkan Provinsi di Indonesia Berdasarkan Nilai Produksi Padi," *Jurnal Litbang Sukowati : Media Penelitian dan Pengembangan*, vol. 2, no. 2, pp. 1–11, 2019, doi: [10.32630/sukowati.v2i2.61](https://doi.org/10.32630/sukowati.v2i2.61).
- [8] F. Inayah, S. Martha, and N. Imro'ah, "Pengelompokan Data Time Series Pada Distribusi Listrik Menurut Provinsi di Indonesia," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 11, no. 5, pp. 735–742, 2022, doi: [10.26418/bbimst.v11i5.58281](https://doi.org/10.26418/bbimst.v11i5.58281).
- [9] L. Gubu, D. Rosadi, and Abdurakhman, "Pembentukan Portofolio Saham Menggunakan Klastering Time Series K-Medoid dengan Ukuran Jarak Dynamic Time Warping," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 13, no. 2, pp. 35–46, 2021, doi: [10.34123/jurnalasks.v13i2.295](https://doi.org/10.34123/jurnalasks.v13i2.295).
- [10] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Pengelompokan Data Runtun Waktu menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016)," *Jurnal Ekspansional*, vol. 11, no. 1, pp. 29–37, 2020.
- [11] 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," in *Prosiding Seminar Nasional Matematika*, vol. 5, 2022, pp. 685–692.
- [12] Badan Pusat Statistik, *Analisis Komoditas Ekspor 2017 - 2021 Sektor Pertanian, Kehutanan, dan Perikanan; Industri Pengolahan; Pertambangan dan Lainnya*. BPS RI, 2022.
- [13] Kementerian Pertanian, "Peluang Ekspor Perkebunan Masih Bertahan," Dec. 21, 2020. [Online]. Available: <https://ditjenbun.pertanian.go.id/peluang-ekspor-perkebunan-masih-bertahan/>
- [14] R. Sitepu, Irmeilyana, and B. Gultom, "Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan," *Jurnal Penelitian Sains*, vol. 14, no. 3(A), pp. 11–17, 2011, doi: [10.56064/jps.v14i3.208](https://doi.org/10.56064/jps.v14i3.208).
- [15] Imasdiani, I. Purnamasari, and F. D. T. Amijaya, "Perbandingan Hasil Analisis Cluster Dengan Menggunakan Metode Average Linkage Dan Metode Ward (Studi Kasus : Kemiskinan Di Provinsi Kalimantan Timur Tahun 2018)," *Jurnal Ekspansional*, vol. 13, no. 1, pp. 9–17, 2022.
- [16] C. Cindy, C. Cynthia, V. Vito, D. Sarwinda, B. D. Handari, and G. F. Hertono, "Cluster Analysis on Dengue Incidence and Weather Data Using K-Medoids and Fuzzy C-Means Clustering Algorithms (Case Study: Spread of Dengue in the DKI Jakarta Province)," *Journal of Mathematical and Fundamental Sciences*, vol. 53, no. 3, pp. 466–486, Jan. 2022, doi: [10.5614/j.math.fund.sci.2021.53.3.9](https://doi.org/10.5614/j.math.fund.sci.2021.53.3.9).
- [17] R. Novidianto and A. T. R. Dani, "Analisis Klaster Kasus Aktif COVID-19 Menurut Provinsi di Indonesia Berdasarkan Data Deret Waktu," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, no. 2, pp. 15–24, 2020, doi: [10.34123/jurnalasks.v12i2.280](https://doi.org/10.34123/jurnalasks.v12i2.280).
- [18] T. Susilowati, D. Sugiarto, and I. Mardianto, "Uji Validasi Algoritme Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai," *Jurnal RESTI : Rekayasa Sistem dan Teknologi Informasi*, vol. 4, no. 6, pp. 1171–1178, 2020, doi: [10.29207/resti.v4i6.2492](https://doi.org/10.29207/resti.v4i6.2492).
- [19] M. Yohansa, K. A. Notodiputro, and E. Erfiani, "Dynamic Time Warping Techniques for Time Series Clustering of Covid-19 Cases in DKI Jakarta," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 13, no. 2, pp. 63–73, Nov. 2022, doi: [10.21512/comtech.v13i2.7413](https://doi.org/10.21512/comtech.v13i2.7413).
- [20] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Jurnal MATRIX*, vol. 9, no. 3, pp. 102–109, 2019, doi: [10.31940/matrix.v9i3.1662](https://doi.org/10.31940/matrix.v9i3.1662).
- [21] D. I. Yunistya, R. Goejantoro, and F. D. T. Amijaya, "The Application Of K – Harmonic Means Method In District/City Grouping (Case Study: Poverty in Kalimantan Island in 2020)," *Jurnal Matematika, Statistika, dan Komputasi*, vol. 19, no. 1, pp. 51–64, 2022, doi: [10.20956/j.v19i1.21116](https://doi.org/10.20956/j.v19i1.21116).
- [22] Nicolaus, E. Sulistianingsih, and H. Perdana, "Penentuan Jumlah Cluster Optimal pada Median Linkage dengan Indeks Validitas Silhouette," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, vol. 5, no. 2, pp. 97–102, 2016, doi: [10.26418/bbimst.v5i02.15564](https://doi.org/10.26418/bbimst.v5i02.15564).
- [23] M. Orisa and A. Faisal, "Analisis Algoritma Partitioning Around Medoid untuk Penentuan Klasterisasi," *Jurnal Teknologi Informasi dan Terapan*, vol. 8, no. 2, pp. 86–90, 2021, doi: [10.25047/jtit.v8i2.258](https://doi.org/10.25047/jtit.v8i2.258).
- [24] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons Inc, 1990.
- [25] Kementerian Perindustrian, "Alami Surplus Januari-April 2020, Ekspor Industri Pengolahan Naik 7 Persen," May 28, 2020. [Online]. Available: <https://kemenperin.go.id/artikel/21730/Alami-Surplus-Januari-April-2020,-Ekspor-Industri-Pengolahan-Naik-7-Persen>.
- [26] Kementerian Koordinator, "Pengembangan Komoditas Hortikultura untuk Peningkatan Kinerja Ekspor Nasional dan Perekonomian Daerah," Jan. 29, 2022. [Online]. Available: <https://ekon.go.id/publikasi/detail/3621/pengembangan-komoditas-hortikultura-untuk-peningkatan-kinerja-ekspor-nasional-dan-perekonomian-daerah>