

Analisis Pola Konvergensi Transpor Kelembapan Udara di Indonesia Bagian Barat Menggunakan K-Means dengan Pembobotan Statistik dan Hierarchical Shape-Based Clustering

Asri Pratiwi^{1, *)}, Tukhfatur Rizmah Azis²⁾, Anwar Fitrianto³⁾, Erfiani⁴⁾, L.M. Risman
Dwi Jumansyah⁵⁾

^{1,2,3,4,5}Program Studi Statistika dan Sains Data, IPB University, Bogor, Indonesia

*email: pratiwiasri@apps.ipb.ac.id

Abstrak

Penelitian ini menganalisis pola konvergensi transport kelembapan udara (VIMT) di wilayah Indonesia Bagian Barat menggunakan metode *K-Means* dengan pembobotan statistik dan *Hierarchical Shape-Based Clustering* berbasis *Dynamic Time Warping* (DTW). Data harian kelembapan spesifik, kecepatan angin zonal, dan meridional selama 2020–2023 digunakan untuk menghitung VIMT. Metode *clustering* digunakan untuk mengidentifikasi pola kelompok pada data transport kelembapan udara. Hasil penelitian menunjukkan bahwa konvergensi kelembapan meningkat signifikan selama periode musim hujan (November–Februari). Dengan menggunakan metode *K-Means*, diperoleh lima klaster dengan pembagian klaster yang lebih jelas dibandingkan empat klaster hasil metode *Hierarchical Clustering*. Evaluasi performa menggunakan *Silhouette* dan *Calinski-Harabasz* menunjukkan bahwa metode *K-Means* lebih unggul, dengan skor 0,37 dan 104,88 dibandingkan 0,13 dan 96,34 pada metode *Hierarchical*. Hal ini memberikan pemahaman terkait pola transport kelembapan udara yang menjadi acuan dalam prediksi pola cuaca dan iklim, sehingga mendukung upaya mitigasi dampak cuaca ekstrem di Indonesia Bagian Barat.

Kata kunci: DTW, hierarchical clustering, kelembapan udara, k-means, konvergensi, VIMT
MSC2020: 62H30

Abstract

This study analyzes the convergence patterns of Vertically Integrated Moisture Transport (VIMT) in the western region of Indonesia using the K-Means method with statistical weighting and Hierarchical Shape-Based Clustering based on Dynamic Time Warping (DTW). Daily data on specific humidity, zonal wind speed, and meridional wind speed from 2020–2023 were used to calculate VIMT. Clustering methods were utilized to identify grouping patterns in moisture transport data. The results showed that moisture convergence significantly increased during the rainy season (November–February). Using the K-Means method, five clusters with clearer separations were obtained compared to the four clusters produced by the Hierarchical Clustering method. Performance evaluation using Silhouette and Calinski-Harabasz scores indicated that the K-Means method was superior, with scores of 0.37 and 104.88 compared to 0.13 and 96.34 for the Hierarchical method. This provides an understanding of the

*) Corresponding Author

Udara...

moisture transport patterns, serving as a reference for predicting weather and climate patterns, thereby supporting efforts to mitigate the impacts of extreme weather in Western Indonesia.

Keywords: *Convergence, DTW, hierarchical clustering, k-means, moisture, VIMT*
MSC2020: 62H30

Pendahuluan

Klasifikasi tidak terawasi (*unsupervised*), yang juga dikenal sebagai *clustering* atau analisis data eksploratif, tidak terdapat data yang diberikan label sebelumnya [1]. *Clustering* bertujuan untuk membagi sekumpulan data yang terbatas dan tanpa label menjadi kelompok-kelompok struktur tersembunyi yang bersifat alami dan diskrit [2], daripada fokus pada karakterisasi yang tepat dari sampel yang tidak terlihat yang berasal dari sebaran peluang yang sama [3]. Sehingga dapat menemukan struktur atau pola tersembunyi dalam data [4]. Hierarkis memungkinkan untuk melihat bagaimana kluster berkembang secara bertahap, pada partisi langsung membagi data menjadi sejumlah kluster yang ditentukan sebelumnya, dengan tujuan mengurangi perbedaan di dalam kluster dan meningkatkan perbedaan antar kluster [5].

Menurut studi meteorologi, pemodelan atmosfer sangat bergantung pada variabel-variabel seperti kelembapan udara spesifik (*specific humidity*) yang menunjukkan jumlah uap air dalam udara, komponen angin zonal (*u*) yang merepresentasikan kecepatan angin pada arah barat-timur, dan komponen angin meridional (*v*) yang merepresentasikan kecepatan angin pada arah utara-selatan. Interaksi antara kelembapan, angin, dan hujan sangat kompleks dan memengaruhi dinamika cuaca baik secara lokal maupun global. *Specific humidity*, yang mewakili kandungan uap air, sangat penting dalam proses pembentukan awan dan hujan [6]. Komponen angin memiliki peran signifikan dalam transportasi massa udara, yang memengaruhi cuaca baik secara lokal maupun global [7]. Interaksi antara kelembapan, angin, dan hujan sangat kompleks, dengan jaringan kausal yang membantu mengidentifikasi jalur pengaruhnya [8]. Variabel-variabel ini sangat penting dalam karakterisasi kondisi atmosfer, yang krusial untuk memvalidasi model berkualitas tinggi yang digunakan dalam penelitian energi angin [9].

Namun, dengan tingginya kompleksitas data cuaca yang terus berkembang, tantangan muncul dalam hal bagaimana menyederhanakan data meteorologi yang multidimensi ini tanpa kehilangan informasi penting. Proses ini menjadi esensial dalam analisis lebih lanjut seperti *clustering* (pengelompokan) data cuaca, yang memungkinkan identifikasi pola cuaca tertentu. *Clustering* dapat digunakan untuk mengelompokkan wilayah geografis atau periode waktu berdasarkan kondisi cuaca yang mirip, yang kemudian dapat memudahkan prediksi cuaca dan mitigasi risiko bencana. Studi terbaru telah mengeksplorasi teknik clustering untuk menganalisis data meteorologi yang kompleks. Pada penelitian sebelumnya telah menerapkan *K-Means* dan *Hierarchical Agglomerative Clustering* pada data stasiun cuaca [10][11], membandingkan *K-Means* dan *K-medoids* untuk pengelompokan data cuaca di Denpasar dan mengungkap pola cuaca dominan pada interval waktu yang berbeda [12], mengevaluasi berbagai metode clustering dan klasifikasi untuk data cuaca di Delhi, menyimpulkan bahwa *K-Means clustering* paling efisien dan *Naive Bayes* memberikan hasil klasifikasi terbaik [13]. Barrera *et al*, juga menekankan pentingnya pilihan jarak dalam clustering *time series* iklim, mencatat bahwa jarak yang berbeda menonjolkan pola spasial atau temporal [14]. Studi-studi ini menunjukkan potensi teknik clustering untuk menyederhanakan data meteorologi multi-dimensi, mengidentifikasi pola cuaca, serta mendukung aplikasi di bidang pertanian, pariwisata, dan optimasi energi.

Salah satu pendekatan yang digunakan untuk menyederhanakan data meteorologi adalah dengan menghitung *Vertically Integrated Moisture Transport* (VIMT). VIMT menghitung transportasi

kelembapan secara vertikal melalui atmosfer dan diintegrasikan dengan komponen angin horizontal. Nilai VIMT dapat menggambarkan total kontribusi kelembapan dan angin terhadap pola cuaca yang terjadi [15]. Sehingga, untuk mengetahui pusat aliran massa udara yang terwakili oleh bilangan skalar, maka dapat dihitung dalam bentuk konvergensi VIMT (nilai positif) dan divergensi VIMT (nilai negatif). Informasi daerah yang memiliki konvergensi VIMT, selanjutnya disebut sebagai konvergensi transpor kelembapan udara, menunjukkan adanya peningkatan atau pemusatan kelembapan udara di wilayah tersebut, sedangkan daerah dengan divergensi VIMT berarti terdapat pengurangan kelembapan udara atau penyebaran kelembapan udara menuju wilayah lain [16].

Variabel konvergensi transpor kelembapan udara yang telah diperoleh dapat dianalisis menggunakan *clustering* untuk mengelompokkan kondisi cuaca berdasarkan karakteristik yang serupa. Pengelompokan ini dapat membantu dalam mengenali pola-pola daerah pemusatan udara basah yang penting untuk mengenali potensi wilayah dengan perkembangan cuaca ekstrem. Teknik *clustering* pada data cuaca dan iklim sudah banyak dilakukan oleh peneliti di berbagai belahan dunia, diantaranya yaitu [17] menggunakan metode *Hierarchical Clustering* untuk mengelompokkan iklim dan cuaca di Eropa. Metode *K-Means* juga banyak diterapkan dalam pengelompokan data iklim dan cuaca untuk peningkatan kualitas di berbagai bidang krusial termasuk kesehatan masyarakat, pertanian, dan perkotaan [17], [18], [19], [20].

Penelitian ini memanfaatkan metode *Hierarchical Clustering* dan *K-Means Clustering* dengan pembobotan statistik keduanya memiliki kemampuan untuk menangkap variasi pola temporal serta efisien dalam pengelompokan data berskala besar. *Hierarchical Shape-Based Clustering* dengan *Dynamic Time Warping* (DTW) memungkinkan identifikasi pola waktu yang unik, sedangkan *K-Means* yang berbasis centroid efektif dalam membentuk cluster berdasarkan sebaran statistik seperti skewness dan kurtosis. Kedua metode ini juga menawarkan fleksibilitas dalam mengeksplorasi tingkat granularitas pola dan cepat diimplementasikan. Dikarenakan data yang digunakan merupakan data *gridded time series*, maka ditambahkan teknik *shape-based* dan *Dynamic Time Warping* (DTW) pada metode *Hierarchical Clustering* dan pada metode *K-Means Clustering* ditambahkan pembobotan statistik dengan skor skewness dan kurtosis. Wilayah Indonesia bagian barat dipilih sebagai lokasi fokus penelitian karena merupakan dominasi wilayah produksi pangan (padi dan hortikultura) terbesar serta memiliki tingkat kepadatan penduduk Indonesia yang lebih tinggi dibandingkan wilayah Indonesia lainnya [21]. Analisis pola konvergensi transport kelembapan udara di wilayah Indonesia bagian barat memberikan pemahaman tentang dinamika iklim terkait pemusatan udara basah dan kering di wilayah yang krusial bagi produksi pangan dan berdampak langsung terhadap mayoritas masyarakat di Indonesia.

Hasil penelitian ini diharapkan dapat memberikan pemahaman terkait pola transport kelembapan udara di Indonesia. Selain itu, hasil penelitian ini juga dapat menjadi acuan dalam prediksi pola cuaca dan iklim di wilayah Indonesia bagian barat, serta dapat digunakan untuk mendukung upaya mitigasi dampak cuaca ekstrem.

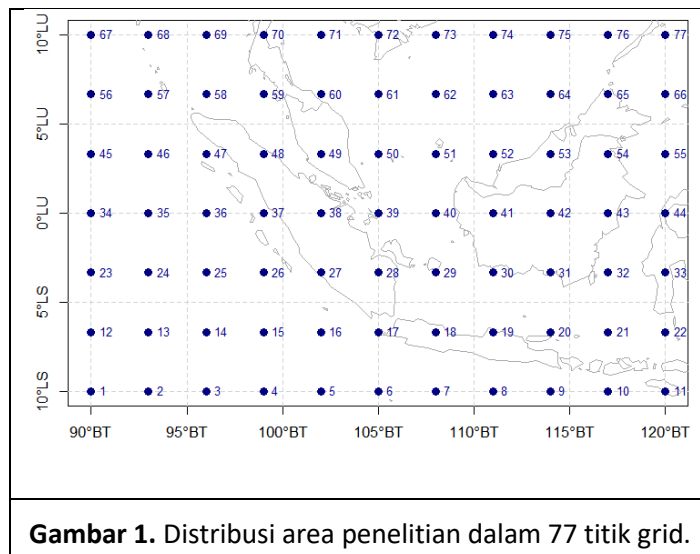
Metode

Data

Data yang digunakan dalam penelitian ini diperoleh dari situs resmi *Physical Sciences Laboratory* (PSL) - *National Oceanic and Atmospheric Administration* (NOAA) yang menyediakan data reanalisis harian dari *National Centers for Environmental Prediction* (NCEP) yang diunduh dari <https://downloads.psl.noaa.gov/Datasets/ncep.reanalysis/Dailies/pressure/> pada tanggal 12 September 2024. Data ini berupa data *gridded* dengan format *netcdf* yang memiliki resolusi spasial

Udara...

2.5°. Variabel yang diambil meliputi kecepatan angin zonal (*uwnd*), kecepatan angin meridional (*vwnd*), dan kelembapan spesifik (*shum*) pada 8 lapisan ketinggian (1000 hPa, 925 hPa, 850 hPa, 700 hPa, 600 hPa, 500 hPa, 400 hPa, dan 300 hPa) selama periode tahun 2020-2023. Data selanjutnya dipotong dengan batas koordinat 10°LU-10°LS dan 90°BT-120°BT berupa 77 grid untuk merepresentasikan wilayah Indonesia bagian barat. Data konvergensi transpor kelembaban udara harian yang dihitung dari variabel NCEP/NCAR *Reanalysis* kemudian diproses lebih lanjut untuk menghasilkan deret waktu representatif satu tahun. Proses ini dilakukan dengan merata-ratakan nilai harian selama periode empat tahun (2020-2023) menjadi satu siklus tahunan. Pendekatan ini dipilih untuk mengurangi *noise* atau fluktuasi acak, sehingga menghasilkan pola yang lebih jelas pada nilai konvergensi transpor kelembaban udara.



Tahapan Analisis

Teknik analisis data dalam penelitian ini meliputi beberapa tahapan sebagai berikut:

1. Menghitung konvergensi transpor kelembaban udara berdasarkan nilai transpor kelembaban udara harian dari data kelembapan spesifik, kecepatan angin zonal, dan meridional selama 4 tahun. Nilai-nilai tersebut kemudian dirata-ratakan untuk menghasilkan data deret waktu tahunan. Persamaan ini mencerminkan pemusatan (konvergensi) atau penyebaran (divergensi) massa udara basah pada setiap grid lokasi. Penghitungan konvergensi transpor kelembaban udara C_{MT} menggunakan persamaan integral berikut:

$$C_{MT} = -\nabla \cdot \int_{p_s}^0 qv dp \tag{1}$$

dimana:

C_{MT} = konvergensi transpor kelembaban

q = kelembapan spesifik

v = komponen angin (u dan v)

p_s = tekanan permukaan

$\nabla \cdot$ = operator divergensi

2. Menghitung *Hierarchical Shape-Based Time Series Clustering* dengan mengukur kemiripan pola deret waktu data konvergensi menggunakan *Dynamic Time Warping* (DTW) [22]. DTW memungkinkan perbandingan pola yang tidak seragam secara temporal dengan menghitung

jarak optimal antara dua deret waktu $A = (a_1, a_2, \dots, a_n)$ dan $B = (b_1, b_2, \dots, b_m)$, dengan persamaan sebagai berikut:

$$DTW(A, B) = \min\{\sum_{k=1}^K d(a_{ik}, b_{jk})\} \quad (2)$$

Dimana $d(a_{ik}, b_{jk})$ = jarak *Euclidean* antara dua titik, sedangkan K = panjang lintasan optimal.

Sehingga dapat memberikan fleksibilitas untuk menangkap variasi pola temporal.

3. Menghitung *K-Means clustering* dengan pemberian bobot berdasarkan *skewness* dan *kurtosis* [23] dari masing-masing deret waktu. *Skewness* dan *Kurtosis* digunakan untuk menghitung karakteristik sebaran statistik dari data konvergensi di setiap grid lokasi. Nilai *skewness* menunjukkan tingkat asimetri, sedangkan *kurtosis* mengukur keruncingan sebaran data. Nilai *skewness* (γ_1) dan *kurtosis* (γ_2) menggunakan persamaan sebagai berikut:

$$\gamma_1 = \frac{(n-1)(n-2)}{n} \sum_{j=1}^n \frac{(X_j - \bar{X})^3}{\sigma_x^3} \quad (3)$$

$$\gamma_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{j=1}^n \frac{(X_j - \bar{X})^4}{\sigma_x^4} - 3 \quad (4)$$

dimana:

\bar{x} = nilai rata-rata data deret waktu.

K-Means clustering kemudian dilakukan dengan menambahkan bobot w pada fitur berdasarkan *skewness* dan *kurtosis* menggunakan persamaan berikut:

$$d_{Kmeans}(A, B) = \sum_{i=1}^p w_i \cdot (a_i - b_i)^2 \quad (5)$$

Langkah ini meningkatkan sensitivitas algoritma *K-Means* terhadap variasi statistik lokal, menghasilkan pemisahan klaster yang lebih tajam dan representatif terhadap karakteristik data.

4. Menghitung validasi model menggunakan data uji. Proses evaluasi ini bertujuan untuk memastikan bahwa model yang dibangun memiliki keandalan serta mampu memberikan prediksi yang tepat dan sesuai. Evaluasi hasil clustering dilakukan dengan membandingkan hasil dari metode *Hierarchical DTW* dan *K-Means*. Kinerja *clustering* diukur dengan *Silhouette Coefficient (SC)* [24] dan *Calinski-Harabasz (CH)* [25]. *Silhouette Coefficient (SC)* didefinisikan sebagai:

$$SC = \frac{1}{n} \sum_{i=1}^n s(x_i) \quad (6)$$

dimana:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}} \quad (7)$$

Dan $SC = [-1, +1]$: $-1 = bad$, $0 = indifferent$, $1 = good$.

$a(x_i)$ = jarak rata-rata antara titik i dan anggota lain dalam klaster yang sama

$b(x_i)$ = jarak rata-rata antara titik i dan anggota klaster terdekat lainnya

Untuk DB didefinisikan sebagai berikut:

$$CH(k) = \frac{\left[\frac{B}{k-1} \right]}{\left[\frac{W}{n-k} \right]} \quad (8)$$

dimana:

B = menunjukkan ragam antar klaster

W = mewakili ragam dalam klaster

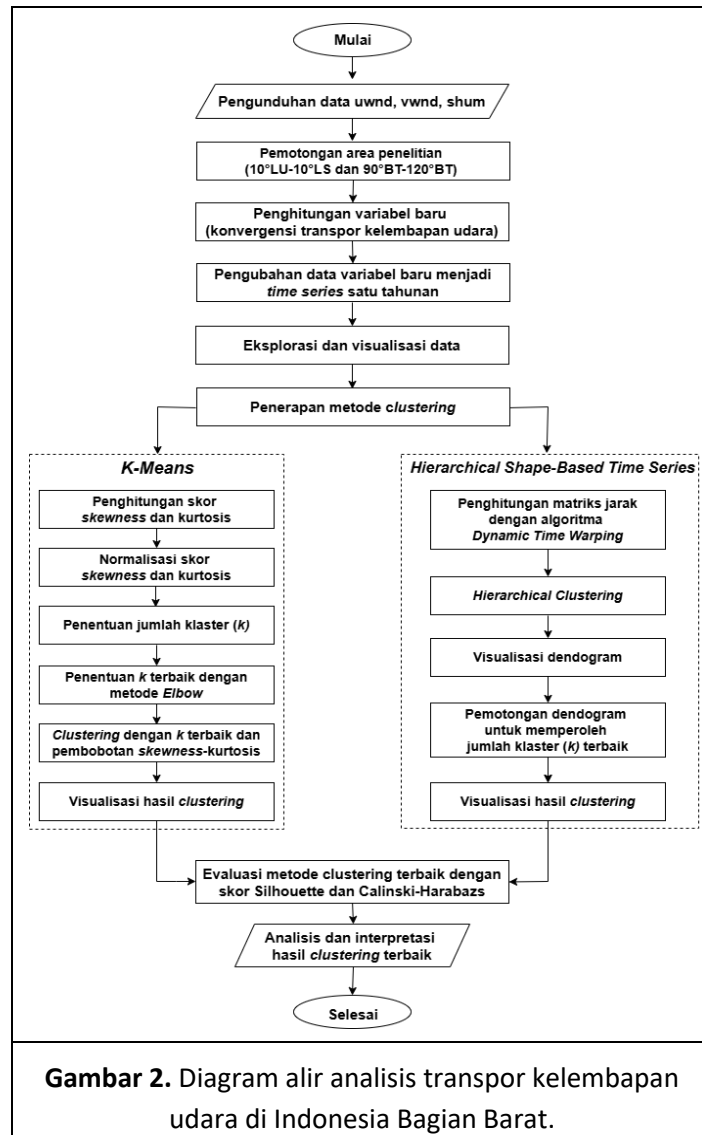
k = menunjukkan jumlah klaster

$(k - 1)$ dan $(n - k)$ = mewakili derajat kebebasan dalam klaster dan antar klaster, dan masing-masing

Semakin tinggi indeks CH yang dihitung, semakin baik hasil pengelompokan, maka nilai k akan dipilih berdasarkan indeks CH yang paling tinggi.

Udara...

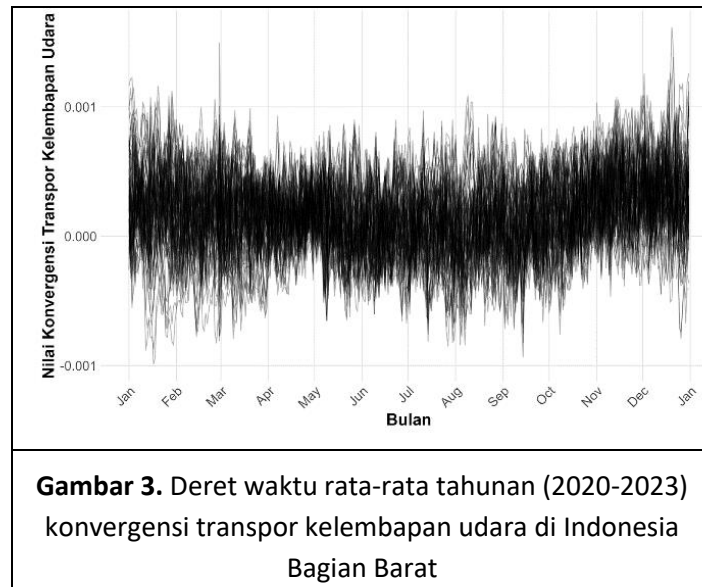
5. Hasil *clustering* dibandingkan untuk mengidentifikasi metode yang lebih efektif dalam memisahkan pola konvergensi transpor kelembaban udara berdasarkan sebaran statistik dan pola temporal.



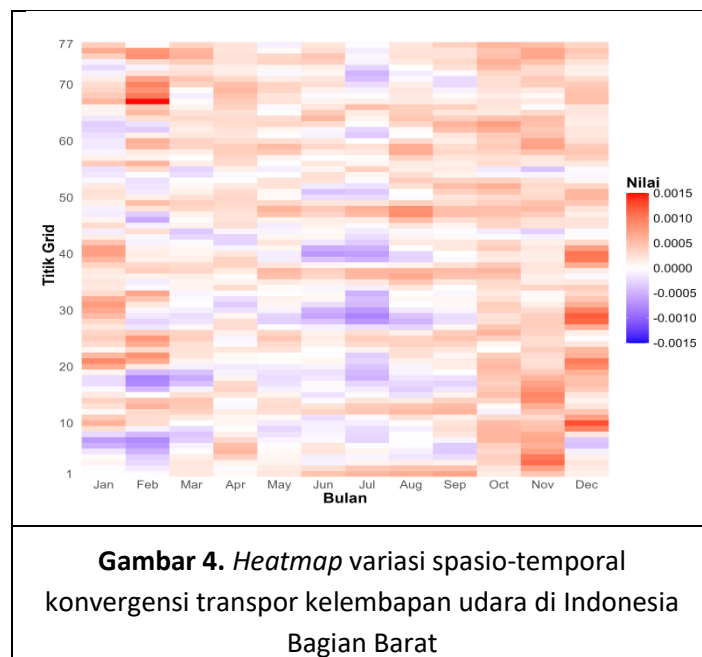
Hasil dan Diskusi

Hasil penghitungan nilai konvergensi harian dari transpor kelembaban udara di wilayah Indonesia bagian barat terlihat adanya variasi spasial dan temporal yang cukup signifikan. Nilai positif menunjukkan adanya konvergensi yang dapat diartikan sebagai sumber massa udara basah, sedangkan nilai negatif merupakan divergensi atau kondisi massa udara kering. **Gambar 2** memperlihatkan deret waktu nilai konvergensi transpor kelembaban udara bulanan yang telah dirata-ratakan selama periode 4 tahun menjadi satu siklus tahunan serta menunjukkan fluktuasi musiman dengan puncak-puncak yang terlihat jelas pada bulan-bulan tertentu. Grafik ini menunjukkan fluktuasi yang jelas sepanjang tahun, dengan amplitudo yang bervariasi. Pola musiman terlihat jelas, di mana nilai konvergensi cenderung meningkat (positif) pada bulan november hingga february terkait dengan musim hujan di wilayah Indonesia bagian barat. Sebaliknya, periode dengan nilai konvergensi yang

rendah atau negatif (divergensi) terlihat sekitar bulan Juni hingga Agustus yang berkaitan dengan berlangsungnya musim kemarau.

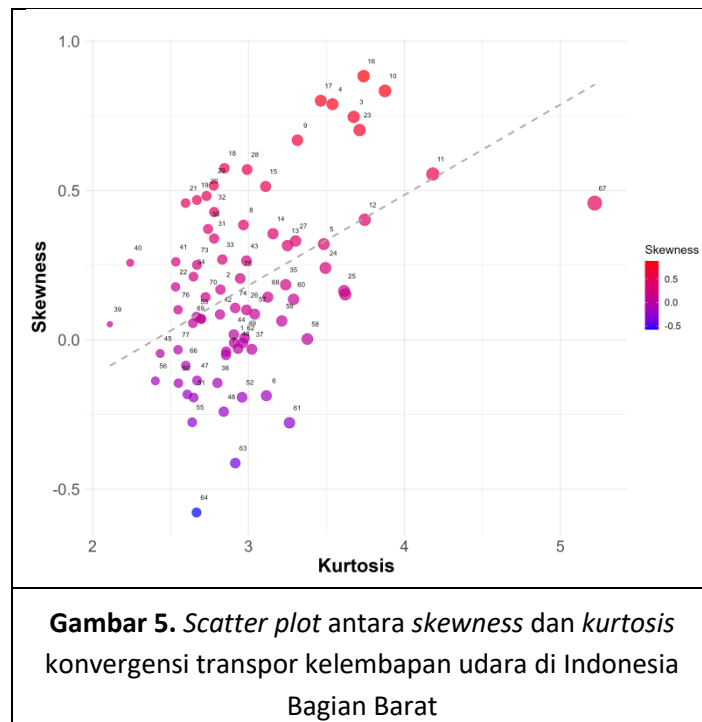


Bentuk visualisasi data secara spatio-temporal terlihat pada *heatmap* **Gambar 3**. Sumbu vertikal menunjukkan lokasi, sementara sumbu horizontal menunjukkan bulan. Warna merah menandakan nilai positif (konvergensi), sedangkan warna biru menandakan nilai negatif (divergensi). Terlihat bahwa beberapa lokasi konsisten menunjukkan konvergensi yang kuat sepanjang tahun (ditunjukkan oleh warna merah yang dominan), sementara lokasi lain menunjukkan pola yang lebih bervariasi atau bahkan cenderung mengalami divergensi. Persebaran spatio temporal ini juga menunjukkan adanya bulan-bulan basah yaitu konvergensi (nilai positif) pada bulan Oktober hingga Desember dan bulan-bulan kering yaitu bulan Juni dan Juli. Sedangkan pada Januari hingga Maret terdapat sebagian wilayah dengan kondisi basah dan sebagian kering.

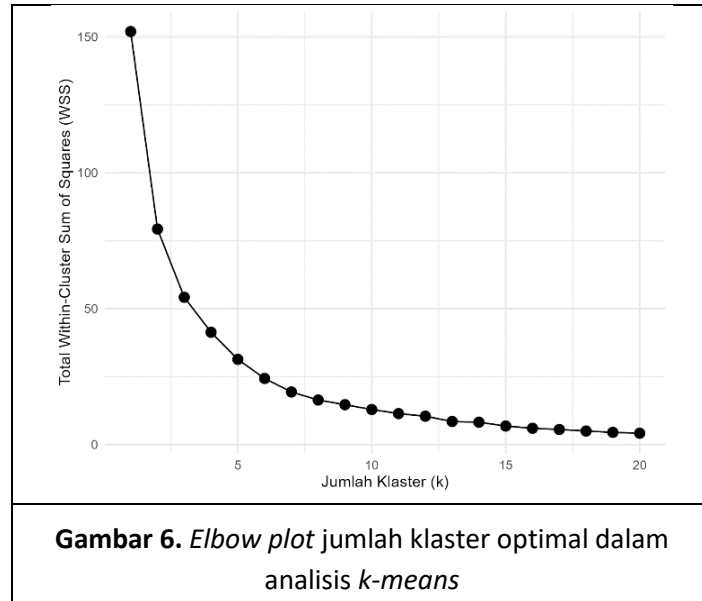


Udara...

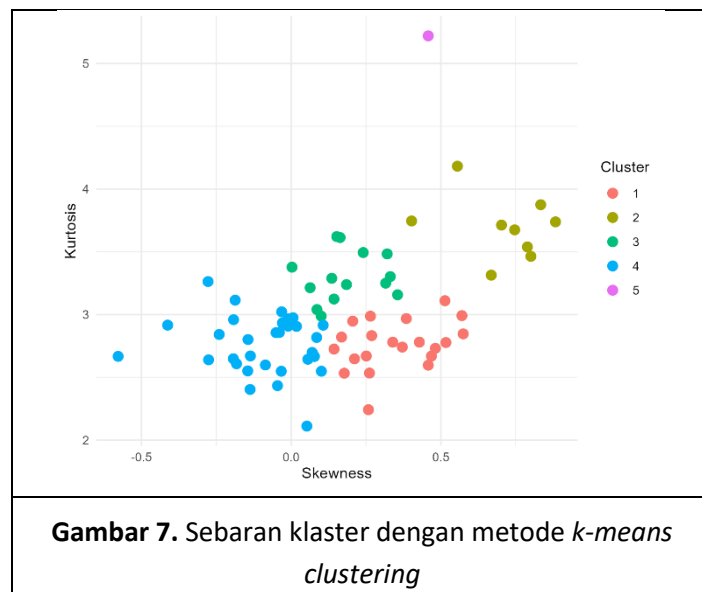
Selanjutnya dilakukan penghitungan nilai *skewness* dan *kurtosis* di 77 titik grid dengan hubungan yang ditampilkan pada scatterplot **Gambar 4**. Sumbu horizontal menunjukkan nilai *kurtosis*, sedangkan sumbu vertikal menunjukkan nilai *skewness*. Setiap titik mewakili satu lokasi. Sebagian besar lokasi menunjukkan nilai *skewness* positif dan *kurtosis* lebih besar dari 3 (leptokurtik), yang mengindikasikan sebaran data yang cenderung miring ke kanan dengan ekor yang lebih panjang atau puncak yang lebih tinggi dibandingkan sebaran normal. *Skewness* positif pada data konvergensi transpor kelembaban udara dapat diartikan bahwa nilai kelembaban udara tinggi (kondisi udara basah) lebih sering terjadi dibandingkan nilai kelembaban rendah. Beberapa lokasi menunjukkan karakteristik yang berbeda, seperti *skewness* negatif atau *kurtosis* yang lebih rendah, yang menunjukkan adanya perbedaan dalam dinamika atmosfer skala lokal.



Selanjutnya dilakukan pengelompokan dengan algoritma *K-Means clustering* berdasarkan dua fitur, yaitu *Skewness* (kemiringan) dan *Kurtosis* (keruncingan). Data juga dinormalisasi agar kedua fitur berada dalam rentang yang serupa, karena algoritma *K-Means* sensitif terhadap skala data. Dari plot *elbow* pada **Gambar 5** menunjukkan bahwa jumlah kluster optimal berada di sekitar 4-6, di mana penurunan *within-cluster sum of squares* (WSS) mulai melambat. Namun dipilih $k=8$ untuk mendapatkan variasi kluster yang lebih banyak namun tetap representatif sesuai pola kemiripan antara nilai *Skewness* dan *Kurtosis*.



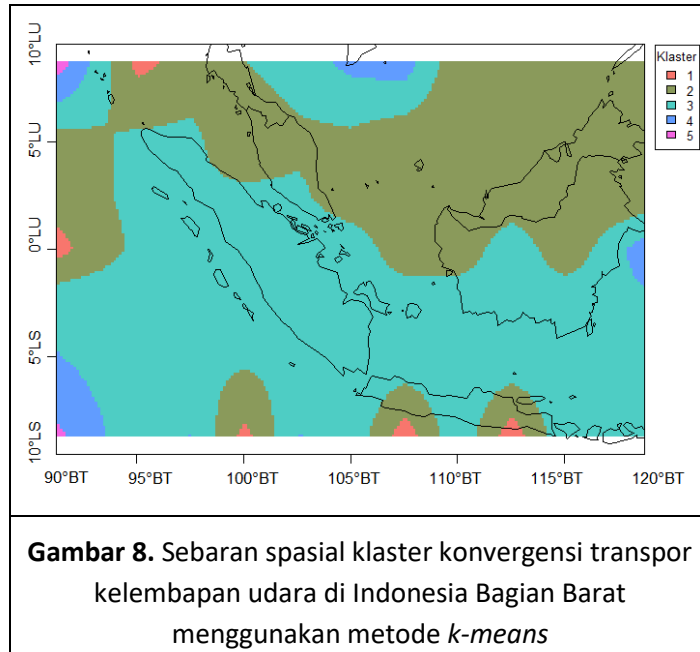
Sebaran hasil clustering metode *K-Means* dengan pembobotan nilai *kurtosis* dan *skewness* pada **Gambar 6** menunjukkan bahwa setiap kluster memiliki sebaran yang cukup baik, meskipun terdapat beberapa titik yang berada di luar rentang utama, khususnya pada kluster ke-5 yang berada cukup jauh dari kluster-kluster lainnya.



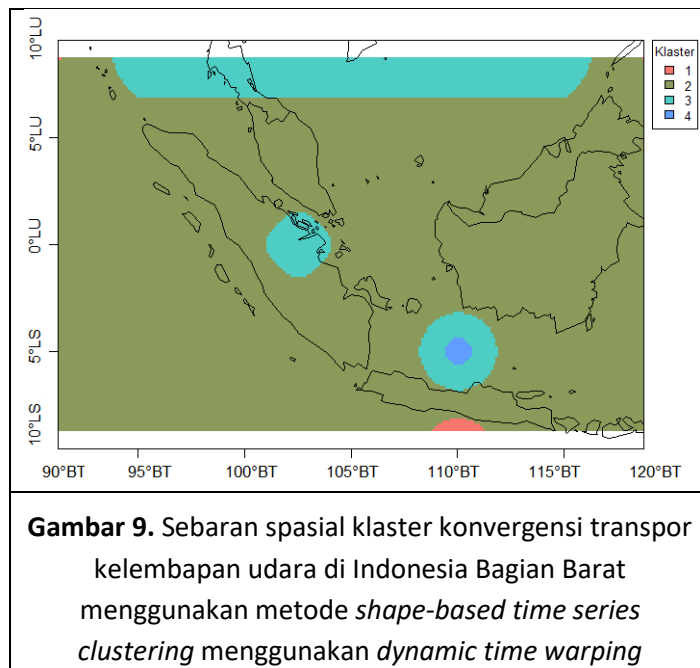
Sebaran secara spasial hasil clustering menggunakan metode *K-Means* pada **Gambar 7** menunjukkan sebaran lima kluster yang diwakili oleh warna berbeda, di mana setiap kluster menggambarkan wilayah-wilayah yang memiliki karakteristik serupa. Dominasi kluster dengan wilayah terluas terlihat pada kluster 3 dan 2. Kluster 3 (*cyan*) mendominasi sebagian besar wilayah Sumatera dan Jawa yang menunjukkan adanya kemiripan karakteristik dalam data di kedua pulau tersebut. Kluster 2 (*hijau lumut*), mendominasi sebagian besar perairan sebelah utara Indonesia dan sebagian wilayah daratan Kalimantan dan Jawa. Kluster 1 (*merah muda*) muncul di sebelah selatan Jawa dan perbatasan perairan luar Indonesia. Kluster 4 (*biru*) terlihat di sebelah timur Kalimantan dan

Udara...

perbatasan perairan luar Indonesia. Sedangkan klaster 5 (ungu) hanya di perbatasan perairan luar Indonesia.



Peta sebaran spasial klaster hasil metode Hierarchical Shape-Based Clustering dengan DTW yang mempertimbangkan kesamaan pola temporal ditunjukkan oleh **Gambar 8** dalam empat wilayah klaster. Klaster 2 (hijau lumut) mendominasi di sebagian besar wilayah Indonesia Bagian Barat yang terlihat sangat timpang dibandingkan dengan wilayah klaster lainnya. Klaster 3 (cyan) muncul di sekitar Riau dan sebelah utara Jawa Tengah. Klaster 4 (biru) membentuk area lingkaran kecil di tengah klaster 3 yang berada di sebelah utara Jawa Tengah. Klaster 1 (merah muda) membentuk area yang sangat kecil di sebelah selatan Jawa.



Kedua metode clustering dievaluasi performanya pada **Tabel 1**. Diketahui bahwa analisis dari nilai *Silhouette* dan *Calinski-Harabasz* menunjukkan kualitas pemisahan klaster yang terbentuk dalam metode clustering. Nilai *Silhouette* berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan klaster yang lebih kompak dan terpisah dengan baik, sementara nilai mendekati -1 mengindikasikan tumpang tindih antar klaster. Nilai *Silhouette* 0.3708658 untuk *K-Means* (dengan $k=5$) menunjukkan hasil klasterisasi yang lebih baik dibandingkan dengan 0.1353179 pada metode *Hierarchical Shape-Based Clustering*, yang berarti klaster *K-Means* lebih kompak dan terpisah.

Indeks *Calinski-Harabasz* tidak memiliki rentang tertentu, tetapi nilai yang lebih tinggi menunjukkan klasterisasi yang lebih baik dalam hal kompaktilitas dan pemisahan antar klaster. Nilai 104.8889 pada *K-Means* dibandingkan dengan 96.3440 pada *Hierarchical Clustering* menandakan bahwa *K-Means* memiliki struktur klaster yang lebih terpisah dan lebih jelas.

Tabel 1. Evaluasi performa metode *clustering*

Metode	<i>Silhouette</i>	<i>Calinski-Harabasz</i>
<i>K-Means Clustering</i> ($k=5$)	0.3708658	104.8889
<i>Hierarchical Shape-Based Clustering</i> dengan DTW ($k=4$)	0.1353179	96.3440

Secara umum, kedua metrik ini terutama dibandingkan secara relatif antara model klasterisasi yang diterapkan pada data yang sama untuk memilih model terbaik, sehingga *K-Means* dengan nilai yang lebih tinggi dalam kedua metrik ini terbukti lebih unggul untuk pemisahan klaster konvergensi kelembapan.

Kesimpulan

Penggunaan metode *K-Means* memiliki performa yang lebih baik dibandingkan *Hierarchical Shape-Based Clustering* dalam mengidentifikasi pola kelompok pada data transport kelembapan udara. Pada metode *K-Means* diperoleh lima klaster dengan dua klaster yang mendominasi, sedangkan pada metode *Hierarchical Shape-Based Clustering* dengan DTW menghasilkan empat klaster dengan satu klaster dominan. Berdasarkan evaluasi performa nilai *Silhouette* dan *Calinski-Harabasz*, diperoleh bahwa metode *K-Means* lebih unggul dibandingkan metode *Hierarchical Shape-Based Clustering*. Metode *K-Means* dengan pembobotan statistik (*skewness* dan *kurtosis*) mampu menghasilkan pemisahan klaster yang lebih jelas atau terpisah dengan baik.

Penelitian ini berhasil mengidentifikasi pola spasial dan temporal konvergensi transpor kelembapan udara di Indonesia Bagian Barat yang terdiri atas 77 titik grid. Pola secara spasial terlihat bahwa sebagian besar titik grid di wilayah Indonesia Bagian Barat didominasi oleh kondisi udara basah (nilai konvergensi transpor kelembapan tinggi). Hal ini ditunjukkan juga oleh hasil *clustering* metode *K-Means* yang didominasi oleh klaster 3 (*cyan*) Sumatera dan Jawa dan klaster 2 (hijau lumut) perairan utara Indonesia. Pola temporal konvergensi transpor kelembapan udara ditunjukkan oleh pola musiman kondisi udara basah dengan konvergensi kuat pada bulan November hingga Februari (musim hujan) di Indonesia Bagian Barat. Dengan kata lain, pada wilayah Sumatera, Jawa, dan perairan utara Indonesia memiliki potensi pertumbuhan awan dan hujan yang lebih besar dibandingkan wilayah lainnya, terutama pada musim hujan. Ketiga wilayah tersebut perlu mendapatkan perhatian lebih oleh pemerintah dan instansi terkait dalam hal mitigasi bencana yang terkait cuaca ekstrim.

Referensi

- [1] A. I. Károly, R. Fullér, and P. Galambos, "Unsupervised Clustering for Deep Learning: A tutorial survey," *Acta Polytech. Hungarica*, vol. 15, no. 8, 2018.
- [2] I. Nurandini and A. F. Huda, "Klastering Dokumen dengan Menambahkan Metadata Menggunakan Algoritma COATES," *KUBIK: Jurnal Publikasi Ilmiah Matematika*, vol.2, no.2, 2017. Accessed: Nov. 25, 2024. [Online]. Available: <https://journal.uinsgd.ac.id/index.php/kubik/article/view/1859>
- [3] J. Singh and D. Singh, "A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery: Taxonomy, challenges, applications and future prospects," *Adv. Eng. Informatics*, vol. 62, p. 102799, Oct. 2024, doi: 10.1016/J.AEI.2024.102799.
- [4] D. Dedy and A. Cherid, "Data Mining Pengolahan Data Calon Pekerja Migran Indonesia (PMI) dengan Penerapan Metode Klustering K-Means dan Metode Klasifikasi K-Nearest Neighbor (KNN): Studi Kasus PT. SAM," *Format J. Ilm. Tek. Inform.*, vol. 9, no. 2, pp. 166–182, Jan. 2021, doi: 10.22441/FORMAT.2020.V9.I2.008.
- [5] P. Shetty and S. Singh, "Hierarchical Clustering: A Survey," *Int. J. Appl. Res.*, vol. 7, no. 4, pp. 178–181, Apr. 2021, doi: 10.22271/ALLRESEARCH.2021.V7.I4C.8484.
- [6] B. Khouider, "Moisture and Moist Thermodynamics," *Math. Planet Earth*, pp. 23–39, 2019, doi: 10.1007/978-3-030-17775-1_2.
- [7] B. M. Smirnov, "Physical Aspects of Climate Change," *Global Energetics of the Atmosphere*, pp. 261–291, 2021, doi: 10.1007/978-3-030-90008-3_10.
- [8] A. Vázquez-Patiño, L. Campozano, D. Ballari, M. Córdova, and E. Samaniego, "Virtual Control Volume Approach to the Study of Climate Causal Flows: Identification of Humidity and Wind Pathways of Influence on Rainfall in Ecuador," *Atmos. 2020, Vol. 11, Page 848*, vol. 11, no. 8, p. 848, Aug. 2020, doi: 10.3390/ATMOS11080848.
- [9] N. Hamilton, "Atmospheric condition identification in multivariate data through a metric for total variation," *Atmos. Meas. Tech.*, vol. 13, no. 2, pp. 1019–1032, Mar. 2020, doi: 10.5194/AMT-13-1019-2020.
- [10] M. A. De Oliveira, F. A. Varone, C. W. Fraisse, R. M. Araújo, and G. G. H. Cavalheiro, "Clustering of meteorological data to improve agricultural decisions: a case study with SIMAGRO-RS," *ACM Int. Conf. Proceeding Ser.*, May 2024, doi: 10.1145/3658271.3658288.
- [11] B. N. Ruchjana, H. Khoirunnisa, I. Irianingsih, and B. Suhandi, "Perbandingan Penerapan Metode Agglomerative dengan Metode K-Means pada Data Curah Hujan di Wilayah Bogor," *KUBIK: Jurnal Publikasi Ilmiah Matematika*, vol.5 no.2, 2020. Accessed: Nov. 25, 2024. [Online]. Available: <https://journal.uinsgd.ac.id/index.php/kubik/article/view/7581>
- [12] B. Prihasto, D. Darmansyah, D. P. Yuda, F. M. Alwafi, H. N. Ekawati, and Y. P. Sari, "Comparative Analysis of K-Means and K-Medoids Clustering Methods on Weather Data of Denpasar City," *J. Pendidik. Multimed.*, vol. 5, no. 2, pp. 91–114, Dec. 2023, doi: 10.17509/EDSENCE.V5I2.65925.
- [13] R. Bhardwaj and V. Duhoon, "Study and Analysis of Time Series of Weather Data of Classification and Clustering Techniques," *Adv. Intell. Syst. Comput.*, vol. 1165, pp. 257–270, 2021, doi: 10.1007/978-981-15-5113-0_19.
- [14] Y. Barrera *et al.*, "Clustering high dimensional meteorological scenarios: Results and performance index," *Int. J. Approx. Reason.*, vol. 139, pp. 1–11, Dec. 2021, doi: 10.1016/J.IJAR.2021.08.007.
- [15] K. Chansaengkrachang, A. Luadsong, and N. Ascharyaphotha, "Vertically Integrated Moisture Flux Convergence over Southeast Asia and Its Relation to Rainfall over Thailand," *Pertanika J. Sci. Technol*, vol. 26, no. 1, pp. 235–246, 2018, Accessed: Oct. 09, 2024. [Online]. Available: <https://rbkm.kmutt.ac.th/xmlui/handle/123456789/1665>
- [16] "GRIB Parameter Database: Param 213.," European Centre for Medium-Range Weather Forecasts.(n.d.).Accessed:Sep.30,2024.[Online].Available: <https://codes.ecmwf.int/grib/param-db/213>.
- [17] G. Pernigotto, A. Walsh, A. Gasparella, and J. L. M. Hensen, "Clustering of European Climates

- and Representative Climate Identification for Building Energy Simulation Analyses”, doi: 10.26868/25222708.2019.210938.
- [18] D. Sathiaraj, X. Huang, and J. Chen, “Predicting climate types for the Continental United States using unsupervised clustering techniques,” *Environmetrics*, vol. 30, no. 4, p. e2524, Jun. 2019, doi: 10.1002/ENV.2524.
- [19] D. M. Straus, “Clustering Techniques in Climate Analysis,” *Oxford Res. Encycl. Clim. Sci.*, Feb. 2018, doi: 10.1093/ACREFORE/9780190228620.013.711.
- [20] T. Kurihana *et al.*, “Identifying Climate Patterns Using Clustering Autoencoder Techniques,” *Artif. Intell. Earth Syst.*, vol. 3, no. 3, Aug. 2024, doi: 10.1175/AIES-D-23-0035.1.
- [21] Badan Pusat Statistik, “Statistik Indonesia 2023 (No. 03200.2303),” 2023. Accessed: Sep. 30, 2024.[Online].Available: <https://www.bps.go.id/id/publication/2023/02/28/18018f9896f09f03580a614b/statistik-indonesia-2023.html>
- [22] H. Li, “On-line and dynamic time warping for time series data mining,” *Int. J. Mach. Learn. Cybern.*, vol. 6, no. 1, pp. 145–153, Feb. 2015, doi: 10.1007/S13042-014-0254-0/METRICS.
- [23] S. Demir, “Comparison of Normality Tests in Terms of Sample Sizes under Different Skewness and Kurtosis Coefficients,” *Int. J. Assess. Tools Educ.*, vol. 2022, no. 2, pp. 397–409, 2022, doi: 10.21449/ijate.1101295.
- [24] H. B. Zhou and J. T. Gao, “Automatic Method for Determining Cluster Number Based on Silhouette Coefficient,” *Adv. Mater. Res.*, vol. 951, pp. 227–230, 2014, doi: 10.4028/WWW.SCIENTIFIC.NET/AMR.951.227.
- [25] G. Zhu, X. Li, S. Zhang, X. Xu, and B. Zhang, “An improved method for k-means clustering based on internal validity indexes and inter-cluster variance,” *Int. J. Comput. Sci. Eng.*, vol. 25, no. 3, p. 253, 2022, doi: 10.1504/IJCSE.2022.123112.