



Analisis Klasterisasi Perubahan Iklim di Indonesia Menggunakan K-Means Tahun 2023

Muklas Rivai*, Rosni, Nasrullah Bacthiar

Program Studi Sains Aktuaria, Fakultas Sains, Institut Teknologi Sumatera

muklas.rivai@at.itera.ac.id

Abstract

Global climate change has a significant impact on weather patterns in Indonesia, affecting key sectors such as agriculture, fisheries, health, and the economy. To analyze complex climate data, the K-Means clustering method is used to identify similar weather patterns, which are beneficial for disaster mitigation, resource planning, and more effective climate management. The data used includes wind speed, humidity, and rainfall from 34 provinces in Indonesia in 2023. Based on the silhouette method analysis, it was found that the optimal number of clusters is two. The first cluster includes 15 provinces such as North Sumatra, Jambi, West Papua, and others. The characteristics of the first cluster are low wind speed, high humidity, and high rainfall. These conditions are influenced by tropical geographic location, stable monsoon wind patterns, topography (valleys, forests, or mountains), and proximity to water vapor sources. The second cluster consists of 19 provinces such as Aceh, East Java, West Nusa Tenggara, and others. The characteristics of the second cluster are high wind speed, low humidity, and low rainfall. Geographical factors such as open areas, dry monsoon winds, and distance from major moisture sources are considered. The results of this study are expected to aid disaster planning and mitigation related to weather patterns in Indonesia.

Keywords: clustering; weather patterns; k-means

Abstrak

Perubahan iklim global memberikan dampak signifikan terhadap pola cuaca di Indonesia, memengaruhi sektor penting seperti pertanian, perikanan, kesehatan, dan ekonomi. Untuk menganalisis data iklim yang kompleks, metode klasterisasi *K-Means* digunakan guna mengidentifikasi pola cuaca serupa, yang bermanfaat bagi mitigasi bencana, perencanaan sumber daya, dan pengelolaan iklim secara lebih efektif. Data yang digunakan mencakup kecepatan angin, Kelembapan, dan curah hujan dari 34 provinsi di Indonesia pada tahun 2023. Berdasarkan analisis metode silhouette, ditemukan bahwa jumlah klaster optimal adalah dua. Klaster pertama meliputi 15 provinsi seperti Sumatera Utara, Jambi, Papua Barat dan lainnya. Karakteristik kluster pertama yaitu kecepatan angin rendah, Kelembapan tinggi, dan curah hujan tinggi. Kondisi ini dipengaruhi oleh lokasi geografis tropis, pola angin muson stabil, topografi (lembah, hutan, atau pegunungan), serta kedekatan dengan sumber uap air. Klaster kedua Terdiri dari 19 provinsi seperti Aceh, Jawa Timur, Nusa Tenggara Barat dan lainnya. Karakteristik kluster kedua yaitu memiliki kecepatan angin tinggi, Kelembapan rendah, dan curah hujan rendah. Faktor geografis seperti wilayah terbuka, angin muson kering, dan jarak dari sumber Kelembapan utama. Hasil penelitian ini diharapkan dapat membantu perencanaan dan mitigasi bencana terkait pola cuaca di Indonesia.

Kata Kunci: klasterisasi; pola cuaca; *k-means*

1. PENDAHULUAN

Perubahan iklim global membawa dampak signifikan terhadap pola cuaca di berbagai belahan dunia, termasuk Indonesia. Sebagai negara kepulauan yang terletak di garis khatulistiwa, Indonesia memiliki keragaman iklim yang sangat kompleks dan bervariasi. Variasi ini tidak hanya mempengaruhi kehidupan sehari-hari masyarakat, tetapi juga berdampak pada sektor pertanian, perikanan, kesehatan, dan ekonomi secara keseluruhan. Memahami pola cuaca yang ada sangat penting untuk mengambil langkah-langkah yang tepat dalam perencanaan dan pengelolaan sumber daya. Perubahan pola curah hujan bahkan terbukti dapat mempengaruhi aktivitas ekonomi dan produksi di berbagai wilayah (Kotz et al., 2022). Selain itu, fenomena cuaca ekstrem seperti siklon tropis juga dapat memicu perubahan kondisi atmosfer yang berdampak pada intensitas hujan dan kondisi cuaca regional (Suhardi et al., 2020).

Cuaca adalah kondisi yang terjadi pada waktu tertentu di area tertentu yang relatif kecil dan dalam durasi yang singkat (Iskandar, 2019). Kondisi cuaca di suatu wilayah dapat berubah-ubah tergantung pada berbagai faktor yang memengaruhinya, seperti kecepatan angin, kelembapan, dan curah hujan. Perubahan karakteristik permukaan lahan dan penggunaan lahan juga dapat mempengaruhi suhu permukaan serta karakteristik curah hujan di suatu wilayah (Suharyanto et al., 2023). Selain itu, variabilitas curah hujan diketahui memiliki hubungan erat dengan perubahan suhu permukaan laut yang terjadi di berbagai wilayah dunia (Mulsandi et al., 2024). Data iklim harian yang mencakup variabel kecepatan angin, kelembapan, dan curah hujan merupakan informasi yang sangat penting untuk menganalisis kondisi cuaca. Badan Meteorologi, Klimatologi, dan Geofisika menyediakan berbagai data meteorologi dan klimatologi yang dapat dimanfaatkan dalam penelitian terkait cuaca dan iklim (BMKG, 2026). Data tersebut banyak dimanfaatkan dalam berbagai penelitian untuk menganalisis variabilitas iklim dan pola cuaca di Indonesia (Nur'utami & Hayasaka, 2022). Namun, dengan volume data yang besar serta kompleksitas hubungan antarvariabel yang tinggi, analisis konvensional sering kali tidak cukup untuk mengidentifikasi pola yang relevan dalam

Salah satu algoritma klasterisasi yang paling populer adalah metode K-Means karena memiliki proses komputasi yang relatif sederhana serta mampu mengelompokkan data dalam jumlah besar secara efisien (Kim, 2020; Wani, 2024). Meskipun berbagai penelitian telah menerapkan metode klasterisasi pada data iklim dan cuaca, penelitian yang secara khusus mengelompokkan wilayah Indonesia berdasarkan kombinasi variabel meteorologis tahun 2023, seperti curah hujan, kelembapan udara dan kecepatan angin masih sangat terbatas. Penerapan metode K-Means pada data iklim Indonesia tahun 2023 guna memperoleh pola pengelompokan wilayah yang lebih representatif berdasarkan karakteristik meteorologisnya (Han et al., 2022).

Metode K-Means telah banyak digunakan dalam berbagai bidang penelitian. Penelitian sebelumnya menunjukkan bahwa algoritma K-Means dapat digunakan untuk mengelompokkan data kesehatan, seperti distribusi imunisasi campak pada balita di Indonesia (Sari et al., 2018). Selain itu, metode ini juga digunakan dalam pengelompokan data teknologi, misalnya dalam mengelompokkan berbagai jenis merek smartphone berdasarkan karakteristik tertentu (Almayda & Saepudin, 2021). Adapun penelitian serupa yang menggunakan analisis klasterisasi dalam mengelompokkan wilayah berdasarkan indikator tertentu, diantaranya penelitian yang dilakukan oleh Nurfathullah dan Purnamasari (2024) mengenai pengelompokan provinsi di Indonesia

berdasarkan jumlah pengangguran terbuka menggunakan metode K-Means yang menghasilkan tiga klaster dengan karakteristik tingkat pengangguran yang berbeda di setiap kelompok provinsi. Selain itu, penelitian lain yang dilakukan oleh Rosni et al. (2026) menggunakan algoritma K-Means menunjukkan bahwa jumlah klaster optimal dengan metode Elbow adalah dua klaster yang menggambarkan perbedaan tingkat jumlah penyandang disabilitas antar wilayah.

Hasil penelitian tersebut menunjukkan bahwa metode K-Means memiliki kemampuan yang baik dalam mengidentifikasi kelompok data yang memiliki karakteristik serupa sehingga dapat membantu proses analisis dan pengambilan keputusan. Meskipun berbagai penelitian telah dilakukan untuk menganalisis variabilitas iklim dan pola curah hujan di berbagai wilayah, sebagian besar penelitian tersebut masih berfokus pada analisis statistik atau hubungan antara variabel iklim dengan fenomena atmosfer global seperti ENSO dan perubahan suhu permukaan laut. Di sisi lain, penelitian yang memanfaatkan teknik data mining, khususnya metode K-Means, untuk mengidentifikasi pola cuaca berdasarkan kombinasi beberapa variabel meteorologis seperti suhu, kelembapan, dan curah hujan masih relatif terbatas, terutama pada data iklim harian di Indonesia. Oleh karena itu, diperlukan pendekatan analisis yang mampu mengelompokkan data iklim secara lebih sistematis sehingga pola cuaca yang terbentuk dapat diidentifikasi dengan lebih jelas.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengeksplorasi penggunaan metode klasterisasi K-Means dalam mengelompokkan data iklim harian yang mencakup variabel suhu, kelembapan, dan curah hujan di Indonesia serta menganalisis pola cuaca yang terbentuk dari hasil pengelompokan tersebut. Diharapkan, temuan dari penelitian ini dapat memberikan kontribusi dalam meningkatkan pemahaman mengenai karakteristik pola cuaca di Indonesia serta dapat digunakan sebagai dasar dalam mendukung proses perencanaan dan pengambilan kebijakan yang berkaitan dengan mitigasi bencana, perencanaan sektor pertanian, dan pengelolaan sumber daya air.

2. METODE PENELITIAN

2.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik yang kemudian akan diolah di R-Studio. Data yang digunakan adalah data iklim dari 34 provinsi di Indonesia pada tahun 2023. Adapun observasi pada penelitian ini yaitu 34 provinsi yang mencakup Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Variabel independen yang digunakan sebanyak 3 variabel yaitu kecepatan angin (m/s) (X_1), Kelembapan (X_2), dan jumlah curah hujan (mm) (X_3). Variabel-variabel tersebut akan digunakan sebagai faktor-

faktor dalam analisis klasterisasi terhadap 34 Provinsi yang ada di Indonesia. Data pola cuaca di Indonesia pada 34 provinsi disajikan pada Tabel 1.

Tabel 1. Pola Cuaca di Indonesia

Provinsi	(X_1)	(X_2)	(X_3)
Aceh	2,71	80	1575
Sumatera Utara	2,38	86,9	975,9
Sumatera Barat	2,88	84	3548
Riau	2,97	80,5	2048,3
Jambi	2,39	82,1	1694,9
Sumatera Selatan	3,32	79,5	1947,2
Bengkulu	2,12	83,2	2668,9
Lampung	1,95	78,9	1628,1
Kepulauan Bangka Belitung	4,05	79,9	1534,7
Kepulauan Riau	3,16	84,1	2250,9
DKI Jakarta	1,54	74	2169,5
Jawa Barat	2,14	74,4	2199,3
Jawa Tengah	2,83	70	1620,7
DI Yogyakarta	0,07	82,8	2045,5
Jawa Timur	3,88	75,2	2024,7
Banten	0,99	79,3	1310,1
Bali	3,27	79,1	1133,8
Nusa Tenggara Barat	3,27	81,4	1147,9
Nusa Tenggara Timur	4,04	75,6	1406
Kalimantan Barat	1,75	85,7	2757,7
Kalimantan Tengah	2,17	80,5	2748,4
Kalimantan Selatan	1,92	81,2	2509,6
Kalimantan Timur	1,96	79,7	2069,4
Kalimantan Utara	2,2	83,7	2311,5
Sulawesi Utara	2,92	75,6	1807
Sulawesi Tengah	2,32	72,6	460,9
Sulawesi Selatan	2,89	75,2	3382
Sulawesi Tenggara	1,27	83,1	1589,6
Gorontalo	1,95	77,5	870,6
Sulawesi Barat	1,93	77,2	1167,9
Maluku	2,35	83,6	1987,2
Maluku Utara	2,6	78,3	913,4
Papua Barat	1,54	83,6	2844,6
Papua	2,6	75,5	1265,9

2.2 Prosedur Penelitian

Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mendeskripsikan karakteristik variabel independen dengan statistika deskriptif. Statistika deskriptif mencakup nilai terendah, rata-rata, dan nilai tertinggi pada

data. Langkah ini bertujuan untuk memberikan gambaran umum mengenai data dalam penelitian.

2. Melakukan standarisasi pada data. Standarisasi data bertujuan untuk mengubah skala data yang berbeda agar data dapat dibandingkan secara langsung.
3. Menguji multikolinearitas pada data dengan menggunakan nilai VIF. Uji multikolinearitas dilakukan dengan tujuan memeriksa apakah variabel independen yang digunakan saling berkorelasi, dengan dilakukan pengukuran menggunakan nilai VIF (*Variance Inflation Factor*)
4. Menentukan jumlah klaster optimal. Jumlah klaster dapat ditentukan dengan menggunakan metode *silhouette*, dengan membandingkan objek pada masing-masing klaster. Nilai berkisar antara -1 dan +1, dimana semakin tinggi nilai akan menunjukkan klaster yang lebih baik.

Silhouette Coefficient adalah gabungan dari dua metode untuk menilai kekuatan dan kualitas klaster. Metode kohesi menilai kedekatan hubungan antar objek dalam sebuah klaster, sementara metode separasi menilai sejauh mana sebuah Klaster terpisah dari klaster lainnya. Tahap perhitungan *Silhouette Coefficient*:

$$S(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \quad (2.12)$$

dengan penjabaran sebagai berikut :

$$a(x) = \frac{1}{|C(x)| - 1} \sum_{C(y) \in C(x), y \neq x} d(x, y) \quad (2.13)$$

$$b(x) = \min_{C_k \neq C(x)} \left(\frac{1}{|C_k|} \sum_{C(z) \in C_k} d(x, z) \right) \quad (2.14)$$

Keterangan :

- $S(x)$: Koefisien *silhouette*
 $a(x)$: Jarak rata-rata data ke- x dengan data lain pada *cluster* yang sama
 $b(x)$: Jarak rata-rata data ke- x dengan data lain pada *cluster* yang berbeda
 $C(x)$: *Cluster* tempat data x berada (*Cluster* ke- x)

Rentang nilai *silhouette* berada pada nilai -1 hingga 1. Nilai *silhouette* negatif mengindikasikan adanya overlapping pada struktur cluster yang dihasilkan. Nilai *silhouette* positif mengindikasikan data sudah berada pada cluster yang tepat, sehingga hasil clustering dapat dikatakan baik. Nilai *silhouette* yang muncul juga mengindikasikan kekuatan struktur clustering yang dihasilkan (Kaufman dan Rousseeuw, 1990). Semakin besar nilai *silhouette*, maka semakin baik pula hasil clustering yang akan didapat.

5. Pengelompokan dengan metode K-Means, yakni pengelompokan data dalam k klaster berdasarkan jarak antartitik data. Pengelompokan data ini akan

membentuk klaster-klaster yang memiliki kesamaan karakteristik. K-means adalah algoritma yang digunakan untuk pengelompokan data secara partisi, yang membagi data ke dalam kelompok yang berbeda. Algoritma ini bekerja dengan cara meminimalkan jarak antara setiap data dan pusat klaster (*centroid*) yang ditentukan. Proses ini melibatkan beberapa langkah, di mana data dikelompokkan berdasarkan kedekatannya dengan *centroid* yang dihitung ulang setelah setiap iterasi. Sebagai contoh, ketika diberikan sekelompok objek, K-means akan membagi objek tersebut menjadi beberapa subkelompok. Pembagian ini dilakukan berdasarkan kesamaan dan jarak setiap titik data terhadap rata-rata *centroid* yang ada. Proses ini memungkinkan pengelompokan data yang lebih terstruktur, sehingga pola dan hubungan dalam data dapat lebih mudah diidentifikasi. Dengan cara ini, K-means membantu dalam menyederhanakan analisis data dan memberikan wawasan yang lebih mendalam tentang karakteristik yang ada dalam setiap subkelompok. Berikut merupakan Langkah-langkah dari algoritma K-Means (Yanto & Handayani, 2024; Mujahidin & Hasanah, 2025):

- 1) Menentukan banyak k sebagai jumlah klasterisasi yang di ambil untuk penelitian
- 2) Menentukan pusat klaster awal (*centroid*) secara acak. Penentuan pada *centroid* awal dilakukan secara acak dari objek yang tersedia sebanyak k.
- 3) Gunakan rumus jarak Euclidian, atau jarak Euclidian, untuk menghitung jarak setiap objek ke *centroid* masing-masing klaster. Ini akan menghasilkan jarak yang paling dekat dari setiap data ke *centroid*. Persamaan jarak Euclidian adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - C_i)^2}$$

Keterangan:

$d(x, y)$: jarak antara titik data x_i dengan *centroid* C_i pada data ke- i .

x_i : nilai data pada objek ke- i yang akan dihitung jaraknya.

C_i : nilai pusat klaster (*centroid*) pada data ke- i .

- 4) Mengelompokkan setiap data kedalam klaster anantara data dengan jarak terdekat.
- 5) Menentukan pusat klaster baru dalam data. Dilakukan dengan menghitung rata-rata nilai data yang termasuk dalam kluster yang bersesuaian dengan menggunakan rumus sebagai berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i$$

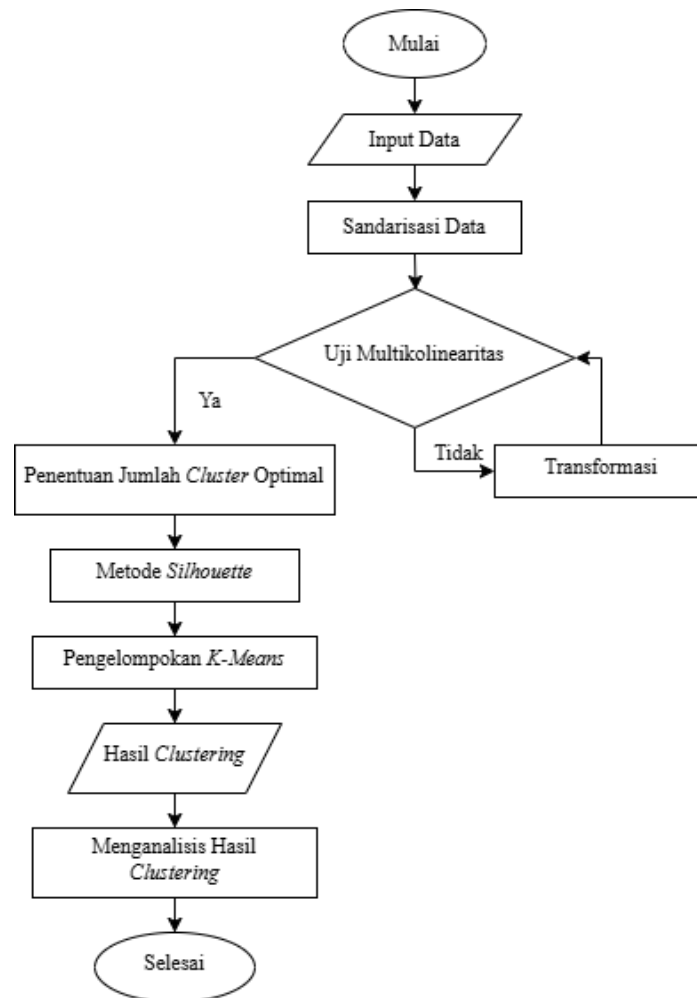
Keterangan:

C_k : pusat klaster (*centroid*) untuk klaster ke- k .

n_k : jumlah data dalam klaster k .

d_i : nilai jarak dari setiap data dalam klaster ke- k .

- 6) Melakukan perulangan dari langkah 3 hingga 5 sampai tiap klaster tidak ada yang berubah.
 - 7) Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat klaster akan digunakan sebagai parameter untuk menentukan klasifikasi data.
6. Menganalisis dan menentukan hasil klasterisasi. Setelah dilakukan pengelompokan klaster, maka dapat dilakukan perbandingan untuk melihat perbedaan antar klaster, sehingga didapat kesimpulan ataupun hasil dari tujuan penelitian dilakukan. Secara lebih jelas alur penelitian disajikan pada Gambar 1.



Gambar 1. Flowchart

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Penelitian ini menganalisis data iklim dari 34 provinsi di Indonesia pada tahun 2023. Data yang digunakan mencakup tiga variabel, yaitu kecepatan angin, kelembapan, dan

jumlah curah hujan. Berikut disajikan ringkasan statistika deskriptif dari data penelitian tersebut.

Tabel 2. Analisis Deskriptif Setiap Variabel

Variabel	Nilai Terendah	Rata-rata	Nilai Tertinggi
Kecepatan angin (m/s)	0,070	2,421	4,050
Kelembapan (%)	70,00	79,53	86,90
Jumlah curah hujan (mm)	460,9	1871,0	3548

Berdasarkan Tabel 2, diperoleh bahwa kecepatan angin memiliki rata-rata sebesar 2,421 meter per detik (m/s), dengan nilai terendah tercatat di DI Yogyakarta, yaitu 0,070 m/s, dan nilai tertinggi di Kepulauan Bangka Belitung sebesar 4,050 m/s. Rentang kecepatan angin sebesar 3,980 m/s ini menunjukkan adanya variasi antar wilayah yang kemungkinan besar dipengaruhi oleh kondisi geografis dan topografi setempat. Variabel kecepatan angin memiliki rata-rata yang lebih rendah dibandingkan kelembapan udara dan curah hujan.

Sementara itu, Kelembapan udara rata-rata berada pada angka 79,53%. Kelembapan terendah tercatat sebesar 70,00% di Jawa Tengah, sedangkan yang tertinggi mencapai 86,90% di Sumatera Utara. Variabel Kelembapan udara memiliki nilai rata-rata yang cukup tinggi, kemungkinan besar disebabkan oleh iklim tropis Indonesia yang memiliki kelembapan udara tinggi sepanjang tahun. Namun, nilai terendah di Jawa Tengah mungkin dipengaruhi oleh faktor-faktor seperti musim kering atau pola angin musiman yang mengurangi Kelembapan.

Di sisi lain, rata-rata jumlah curah hujan yang tercatat adalah 1.871 mm per tahun. Curah hujan terendah mencapai 460,9 mm di Sulawesi Tengah, sedangkan curah hujan tertinggi mencapai 3.548 mm di Sumatera Barat. Rata-rata curah hujan yang tinggi di Sumatera Barat kemungkinan besar disebabkan oleh pengaruh pegunungan barisan yang membentang di sepanjang wilayahnya. Ketika angin membawa uap air dari laut dan bertiup ke arah pegunungan, udara dipaksa naik ke ketinggian yang lebih tinggi. Saat udara naik, suhunya menjadi lebih dingin, sehingga uap air di dalamnya mengembun dan membentuk awan. Proses ini menghasilkan hujan yang lebih sering dan intens di daerah pegunungan. Sebaliknya, rendahnya curah hujan di Sulawesi Tengah dapat dikaitkan dengan kondisi topografi yang lebih datar, sehingga angin yang membawa uap air tidak menghasilkan banyak hujan. Selain itu, fenomena cuaca seperti El Niño juga bisa memperpanjang musim kering dan mengurangi curah hujan di wilayah tersebut.

3.2 Standarisasi Data

Standarisasi bertujuan untuk mengubah skala variabel menjadi skala yang sama agar tidak ada variabel yang memiliki pengaruh lebih besar hanya karena skala nilainya lebih besar. Standarisasi mengubah nilai variabel sehingga memiliki rata-rata 0 dan standar deviasi 1. Hal ini penting terutama jika variabel dalam dataset memiliki unit atau skala yang berbeda, misalnya kecepatan angin (m/s), kelembapan (%), dan curah hujan (mm). Perbedaan ukuran data yang besar dapat menyebabkan analisis kluster tidak valid.

Standarisasi sebagaimana yang umum digunakan adalah dengan bentuk *z-score* sebagai berikut:

$$Z = \frac{x - \mu}{\sigma}$$

Tabel 3. Hasil standarisasi *z-score*

Provinsi	Z Score Kecepatan angin	Z Score Kelembapan	Z Score Curah Hujan
Aceh	0,34	0,12	-0,41
Sumatera Utara	-0,05	1,84	-1,25
Sumatera Barat	0,54	1,12	2,34
Riau	0,65	0,24	0,25
Jambi	-0,04	0,64	-0,25
⋮	⋮	⋮	⋮
Sulawesi Barat	-0,58	-0,58	-0,98
Maluku	-0,08	1,02	0,16
Maluku Utara	0,21	-0,31	-1,34
Papua Barat	-1,04	1,02	1,36
Papua	0,21	-1,01	-0,85

3.3 Uji Multikolinearitas

Multikolinearitas adalah kondisi ketika variabel bebas memiliki korelasi kuat, yang dapat mengganggu keakuratan model dan mengurangi kemampuan model untuk membedakan antar variabel. Multikolineritas dapat diketahui dengan menghitung nilai *Variance Inflation Factor* (VIF) dari masing masing variabel independent. Nilai $VIF < 10$ artinya tidak terjadi multikolineritas, dan sebaliknya jika nilai $VIF > 10$ artinya terjadi multikolineritas. Rumus menghitung nilai *Variance Inflation Factor* (VIF) sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Keterangan:

R_j^2 : Nilai koefisien determinasi dari variabel independen X_j dengan X lainnya.

Tabel 4. digunakan untuk menginterpretasikan nilai VIF dalam mendeteksi adanya multikolinearitas pada variabel.

Tabel 4. Nilai VIF untuk setiap variabel

Variabel	VIF	Keterangan
Kecepatan Angin	1.06	Tidak terjadi multikolinearitas
Kelembapan	1.16	Tidak terjadi multikolinearitas
Jumlah Curah Hujan	1.10	Tidak terjadi multikolinearitas

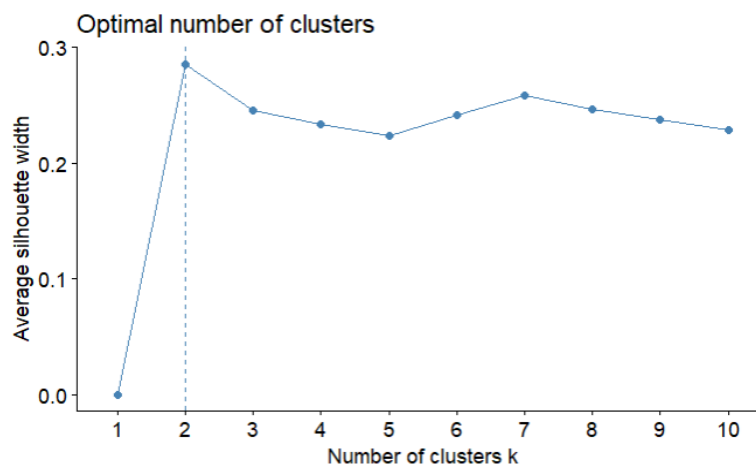
Berdasarkan hasil analisis VIF untuk variabel kecepatan angin, kelembapan, dan jumlah curah hujan, semua nilai $VIF < 10$. Dengan demikian, keputusan yang diambil adalah

tidak menolak H_0 , yang berarti tidak terdapat gejala multikolinearitas. Hal ini menunjukkan bahwa korelasi antar variabel rendah, sehingga model dapat digunakan.

3.4 Analisis Klasterisasi K-Means

Penentuan jumlah kluster dalam analisis kluster non-hierarki dapat dilakukan menggunakan berbagai metode, seperti *silhouette*, *elbow*, dan *gap-statistics*. Dalam penelitian ini, metode *silhouette* digunakan karena memberikan kejelasan visual dan hasil yang akurat dalam menentukan jumlah kluster yang optimal. Menentukan jumlah kluster yang optimal sangat penting sebagai acuan sebelum melakukan pengelompokan data. Proses ini dilakukan dengan memanfaatkan metode *silhouette* menggunakan *software R*.

Metode *silhouette* adalah pendekatan yang menggunakan nilai rata-rata untuk menilai kualitas kluster yang terbentuk. Semakin tinggi nilai rata-rata *silhouette*, semakin baik kualitas kluster tersebut.



Gambar 2. Nilai *silhouette*

Berdasarkan grafik *silhouette width* di atas, terlihat bahwa nilai rata-rata *silhouette* mencapai puncaknya pada $k = 2$, menunjukkan bahwa jumlah kluster optimal adalah $k = 2$ kluster. Hal ini mengindikasikan bahwa pengelompokan dengan dua kluster memberikan hasil yang paling jelas dan terdefinisi dengan baik. Setelah $k = 2$, nilai *silhouette width* menurun secara bertahap, yang menandakan bahwa penambahan jumlah kluster mengurangi kualitas pengelompokan. Meskipun terdapat sedikit fluktuasi pada nilai *silhouette width* untuk jumlah kluster yang lebih besar, secara keseluruhan terlihat bahwa kejelasan dan konsistensi pengelompokan menurun setelah jumlah kluster optimal tercapai. Dengan nilai *silhouette coefficient* tertinggi sebesar 0.2858 juga dicapai pada $k = 2$, memperkuat analisis bahwa dua kluster memberikan pengelompokan yang paling optimal dengan kejelasan dan definisi yang baik antar kluster.

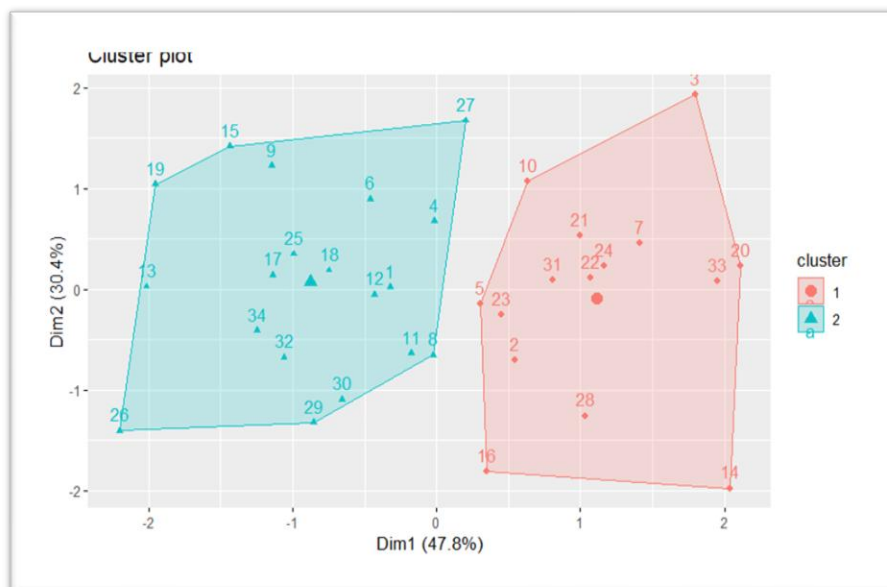
Metode ini memberikan hasil yang lebih kuat dan mendukung keputusan untuk menetapkan jumlah kluster yang paling optimal. Berdasarkan analisis, jumlah kluster

yang optimal dalam penelitian ini adalah 2, sesuai dengan hasil tertinggi pada grafik *silhouette*.

Tabel 5. Hasil Jarak Minimum Setiap Klaster

No.	Provinsi	Klaster	Min. Jarak
1	Aceh	2	0.7917170
2	Sumatera Utara	1	2.0702283
3	Sumatera Barat	1	2.1735233
4	Riau	2	1.1261174
5	Jambi	1	0.9248754
⋮	⋮	⋮	⋮
29	Gorontalo	2	1.4302473
30	Sulawesi Barat	2	1.1863961
31	Maluku	1	0.6049052
32	Maluku Utara	2	1.0437251
33	Papua Barat	1	1.0071951
34	Papua	2	0.6180342

Setelah mendapatkan hasil klaster dilakukan visualisasi hasil klasterisasi yang digunakan untuk melihat persebaran klaster. Berdasarkan gambar, diperoleh bahwa plot yang berwarna merah menunjukkan provinsi yang termasuk ke dalam klaster 1, plot berwarna biru menunjukkan provinsi yang termasuk ke dalam klaster 2.



Gambar 3. Hasil Klasterisasi Perubahan Iklim

Dimensi 1 (Dim1) menjelaskan sebesar 47,8% dari keseluruhan variasi data, yang berarti sebagian besar informasi dalam dataset diwakili oleh Dimensi 1. Dimensi 2 (Dim2) hanya menjelaskan sebesar 30,4% dari variasi data, sehingga kontribusi informasi dari Dimensi 2 lebih kecil dibandingkan dengan Dimensi 1. Setelah dibentuk dalam bentuk plot dilakukan pengelompokan data klaster dalam bentuk tabel untuk melihat provinsi mana saja yang terbentuk ke dalam masing-masing klaster. Hasil klaster disajikan pada tabel 6.

Tabel 6. Hasil Klasterisasi *K-Means*

Klaster	Provinsi	Jumlah Anggota
1	Sumatera Utara, Sumatra Barat, Jambi, Bengkulu, Kepulauan Riau, DI Yogyakarta, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tenggara, Maluku, Papua Barat	15
2	Aceh, Riau, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, Gorontalo, Maluku Utara, Papua	19

Dari hasil klaster tersebut dapat diperoleh nilai rata-rata setiap variabel yang disajikan pada Tabel 7.

Tabel 7. Karakteristik klaster pada setiap variabel

Klaster	Kecepatan Angin	Kelembapan	Curah Hujan
1	-0.562482	0.8434573	0.488414
2	0.444064	-0.6658873	-0.38559

Hasil klaster yang terbentuk merupakan interpretasi dari karakteristik setiap *Klaster* pada Tabel 8.

Tabel 8. Kategori klaster pada setiap variabel

Klaster	Kecepatan Angin	Kelembapan	Curah Hujan
1	Rendah	Tinggi	Tinggi
2	Tinggi	Rendah	Rendah

Berdasarkan Tabel 8, Klaster 1 dapat dikategorikan berdasarkan beberapa faktor geografis dan klimatologis. Karakteristik tersebut dapat dijelaskan sebagai berikut.

1. Kecepatan angin yang relatif rendah pada klaster 1 dipengaruhi oleh kondisi geografis wilayah yang mencakup Kalimantan, sebagian Sumatera (bagian utara dan barat), serta Papua. Wilayah-wilayah tersebut memiliki lokasi geografis tropis, pola angin muson yang relatif stabil, topografi yang beragam, serta kondisi tekanan atmosfer yang cenderung stabil. Faktor-faktor tersebut menyebabkan kecepatan angin berada pada tingkat sedang hingga rendah, kecuali pada kondisi tertentu seperti musim muson atau gangguan cuaca ekstrem di wilayah perairan terbuka. Selain itu, kecepatan angin umumnya lebih rendah di daerah lembah atau kawasan yang berada di bawah pegunungan karena aliran angin terhambat oleh kondisi topografi di sekitarnya. Dengan demikian, wilayah yang memiliki kecepatan angin rendah di Indonesia umumnya merupakan daerah yang terlindungi oleh pegunungan, lembah, atau hutan lebat, berada jauh dari jalur utama angin muson, serta memiliki tutupan vegetasi yang tinggi sehingga meningkatkan gesekan permukaan dan memperlambat

- pergerakan angin. Karakteristik tersebut banyak ditemukan di Kalimantan, Papua bagian pedalaman, dan Sumatera bagian tengah.
2. Wilayah dengan tingkat kelembapan udara tertinggi di Indonesia umumnya berada pada kawasan hutan hujan tropis dan wilayah pesisir, seperti Kalimantan dan Sumatera. Hutan hujan tropis menghasilkan uap air dalam jumlah besar melalui proses evaporasi dan transpirasi. Selain itu, kedekatan dengan laut menyebabkan pasokan uap air dari proses penguapan permukaan laut menjadi lebih tinggi. Angin laut yang bergerak menuju daratan turut membawa massa udara lembap, terutama pada musim penghujan. Oleh karena itu, tingginya kelembapan udara pada Klaster 1 dipengaruhi oleh kombinasi curah hujan yang tinggi, penguapan dari perairan laut, serta keberadaan vegetasi tropis yang mendukung terbentuknya kondisi atmosfer yang lembap secara alami.
 3. Wilayah Indonesia yang memiliki curah hujan tinggi umumnya berada pada kawasan yang dipengaruhi oleh pola angin muson, kondisi topografi, dan letak geografis yang mendukung terjadinya presipitasi dalam jumlah besar. Wilayah tersebut meliputi Sumatera bagian barat, Kalimantan, Papua, Sulawesi, dan Maluku. Pada Sumatera bagian barat, angin muson barat membawa massa udara lembap dari Samudra Hindia yang kemudian mengalami pengangkatan akibat keberadaan Pegunungan Bukit Barisan. Proses ini dikenal sebagai hujan orografis, yaitu hujan yang terjadi ketika udara lembap dipaksa naik oleh pegunungan, kemudian mengalami pendinginan dan kondensasi sehingga menghasilkan curah hujan yang tinggi. Di Kalimantan, keberadaan hutan hujan tropis yang luas meningkatkan proses evapotranspirasi sehingga menambah kandungan uap air di atmosfer dan mendorong terbentuknya presipitasi. Sementara itu, di Papua, massa udara lembap yang berasal dari Samudra Pasifik bertemu dengan pegunungan tinggi sehingga memicu terjadinya hujan orografis dengan intensitas yang tinggi. Dengan demikian, tingginya curah hujan pada wilayah-wilayah tersebut dipengaruhi oleh kombinasi faktor angin muson, topografi pegunungan, serta kedekatan dengan sumber uap air. Kondisi geografis dan klimatologis tersebut menjadikan Sumatera bagian barat, Kalimantan, Papua, Sulawesi, dan Maluku sebagai wilayah dengan curah hujan tinggi di Indonesia.

Sehingga, diperlukannya strategi untuk memanfaatkan berbagai sumber daya dan kondisi alam untuk berbagai tujuan yaitu dengan fokus pada pengelolaan hasil panen tropis, seperti karet, kopi, kakao, dan padi sawah, yang membutuhkan Kelembapan tinggi dan curah hujan melimpah. Investasi infrastruktur seperti saluran irigasi atau penyimpanan air tidak menjadi prioritas utama, tetapi pengendalian penyakit tanaman menjadi tantangan utama.

Di sisi lain, klaster 2 memiliki karakteristik iklim yang cenderung kering atau semi-arid. Di Indonesia, wilayah dengan karakteristik ini biasanya terletak di daerah yang dipengaruhi oleh musim kemarau panjang atau berada di bawah efek bayangan hujan. Karena memiliki beberapa alasan seperti:

1. Wilayah Indonesia yang memiliki kecepatan angin tinggi biasanya terkait dengan kondisi geografis, topografi, dan pola cuaca tertentu. Seperti Nusa Tenggara Timur (NTT) dan Nusa Tenggara Barat (NTB) karena letaknya yang dekat dengan perairan terbuka seperti Laut Timor dan Samudra Hindia membuat wilayah ini lebih terpapar angin muson. Kurangnya penghalang geografis (pegunungan atau hutan lebat) memungkinkan angin bertiup dengan cepat. Angin muson timur (musim kemarau)

- biasanya membawa kecepatan angin lebih tinggi di wilayah ini. Kesimpulan klaster 2 memiliki kecepatan angin yang tinggi dikarenakan wilayah seperti Nusa Tenggara, pantai barat Sumatera, dan Laut Maluku cenderung memiliki kecepatan angin tinggi karena kombinasi faktor geografis (wilayah terbuka), angin muson, dan tekanan atmosfer. Fenomena global seperti badai tropis juga dapat meningkatkan kecepatan angin di beberapa daerah.
2. Memiliki Kelembapan rendah seperti pada wilayah Dataran Tinggi Jawa Timur dan Nusantara Tenggara Timur memiliki Kelembapan yang lebih rendah karena terletak jauh dari sumber Kelembapan utama dan terpengaruh oleh angin darat yang kering. Secara keseluruhan, wilayah Nusa Tenggara, termasuk Flores dan Sumba, juga memiliki Kelembapan yang lebih rendah. Ini karena letaknya yang cukup jauh dari sumber Kelembapan utama, serta pengaruh angin musim yang cenderung lebih kering. Faktor utama yang menyebabkan kelembapan rendah di wilayah-wilayah ini adalah rendahnya curah hujan dan pengaruh angin monsun yang membawa udara kering.
 3. Memiliki jumlah curah hujan yang rendah dimana daerah ini sama seperti wilayah yang mengalami Kelembapan rendah karena curah hujan yang rendah di wilayah-wilayah ini terutama disebabkan oleh pengaruh angin muson kering dan letak geografis yang jauh dari sumber Kelembapan.

Sehingga, diperlukannya strategi untuk memanfaatkan kondisi ini secara optimal yaitu dengan fokus pada pengelolaan tanaman tahan kekeringan, Menanam tanaman seperti gandum, jagung, barley, zaitun, dan anggur yang tahan terhadap kondisi kering. Investasi infrastruktur seperti saluran irigasi atau penyimpanan air hujan untuk memaksimalkan penggunaan air.

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan yaitu analisis klasterisasi terhadap data iklim di 34 Provinsi di Indonesia pada tahun 2023, didapatkan kesimpulan:

1. Melalui analisis menggunakan metode *Silhouette*, diperoleh nilai *Silhouette index* sebesar 0.2858. Maka didapatkan hasil bahwa jumlah klaster optimal adalah 2 klaster. Hal ini mengindikasikan bahwa pengelompokan dengan dua klaster memberikan hasil yang paling jelas dan terdefinisi dengan baik.
2. Hasil analisis yang telah dilakukan didapatkan hasil dua klasterisasi, berdasarkan iklim cuaca di 34 Provinsi yang ada di Indonesia. Klaster 1 terdiri atas 15 Provinsi dengan karakteristik kecepatan angin yang rendah, kelembapan tinggi, dan curah hujan yang tinggi. Kondisi ini dipengaruhi oleh angin muson, topografi seperti lembah, hutan, atau pegunungan, serta vegetasi tropis yang secara alami mendukung tingginya Kelembapan udara. Sedangkan klaster 2 terdiri atas 19 Provinsi dengan karakteristik kecepatan angin yang tinggi, kelembapan rendah, dan curah hujan yang rendah. Hal ini disebabkan oleh faktor geografis seperti wilayah terbuka yang jauh dari sumber Kelembapan, seperti laut, sehingga angin muson yang melewati wilayah ini cenderung menjadi kering.

6. REKOMENDASI

Penelitian lebih lanjut masih diperlukan dalam analisis klasterisasi untuk mendapatkan hasil penelitian yang lebih spesifik dan optimal. Metode ini dapat dibandingkan dengan

klasterisasi data menggunakan algoritma lain seperti, Klasterisasi *Agglomerative Hierarchical*, *K-Medoids*, dan lain-lain. Juga dapat menggunakan metode lain seperti *Dunn index*, *Jaccard coefficient*, dan lainnya agar mendapatkan hasil dan jumlah klaster yang lebih optimal. Selain itu keterbatasan data atau faktor-faktor lain juga berpengaruh terhadap hasil analisis yang lebih akurat.

7. REFERENSI

- Almadya, A., & Saepudin, S. (2021). Penerapan data mining Clustering K-Means untuk mengelompokkan berbagai jenis merk smartphone. *Seminar Nasional Sistem Informasi dan Manajemen Informatika (SISMATIK)*, 1, 241–249.
- Badan Meteorologi, Klimatologi, dan Geofisika. (2026). *Data klimatologi Indonesia*. <https://www.bmkg.go.id>
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Iskandar, D. (2019). *Ensiklopedia seri cuaca dan iklim*. Jawa Tengah: Alprin.
- Kaufman, L., Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Kim, H. (2020). Performance analysis of Clustering K-Means algorithms for mMTC systems. In *2020 International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 30–35). IEEE. <https://doi.org/10.1109/ICTC49870.2020.9289287>
- Kotz, M., Levermann, A., & Wenz, L. (2022). The effect of rainfall changes on economic production. *Nature*, 601(7892), 223–227.
- Mujahidin, I., & Hasanah, S. H. (2025). Comparative analysis of K-means, K-medoids, and fuzzy C-means for Klasterisasi provinces in Indonesia based on rice production in 2024. *Jurnal Gaussian*, 14(2), 356–365. <https://doi.org/10.14710/j.gauss.14.2.356-365>.
- Mulsandi, A., Koesmaryono, Y., Hidayat, R., Faqih, A., & Sopaheluwakan, A. (2024). On the interannual variability of Indonesian monsoon rainfall (IMR): A literature review of the role of its external forcing. *Jurnal Meteorologi dan Geofisika*, 24(2), 115–127. <https://doi.org/10.31172/jmg.v24i2.1049>
- Nur'utami, M. N., & Hayasaka, T. (2022). Interannual variability of the Indonesian rainfall and air–sea interaction over the Indo–Pacific associated with interdecadal Pacific oscillation phases in the dry season. *Journal of the Meteorological Society of Japan. Series II*, 100(1), 77–97. <https://doi.org/10.2151/jmsj.2022-004>
- Nurfathullah, M., & Purnamasari, I. (2024). Implementasi K-Means untuk mengelompokkan provinsi di Indonesia berdasarkan indeks jumlah pengangguran terbuka. *Jurnal Mahasiswa Teknik Informatika*, 8(2), 2279–2282. <https://doi.org/10.36040/jati.v8i2.9466>
- Rosni, R., Fredella, E., & Rivai, M. (2026). Segmentasi kecamatan di Kota Bandung berdasarkan statistik penyandang disabilitas menggunakan algoritma Clustering K-Means. *Mandalika Mathematics and Education Journal*, 8(1), 227–238.
- Sari, R. W., Wanto, A., & Windarto, A. P. (2018). Implementasi RapidMiner dengan metode K-means (Studi kasus: Imunisasi campak pada balita berdasarkan provinsi). *KOMIK: Konferensi Nasional Teknologi Informasi dan Komputer*, 2(1), <https://doi.org/10.30865/komik.v2i1.930>.

- Suhardi, B., Adiputra, A., & Avrian, R. (2020). Kajian dampak cuaca ekstrem saat siklon tropis Cempaka dan Dahlia di wilayah Jawa Barat. *Jurnal Geografi, Edukasi dan Lingkungan (JGEL)*, 4(2), 61–67.
- Suharyanto, A., Maulana, A., Suprayogo, D., Devia, Y. P., & Kurniawan, S. (2023). Land surface temperature changes caused by land cover/land use properties and their impact on rainfall characteristics. *Global Journal of Environmental Science and Management*, 9(3), 353–372. <https://doi.org/10.22034/gjesm.2023.03.01>
- Wani, A. A. (2024). Comprehensive analysis of Clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Computer Science*, pp. 1-45. <https://doi.org/10.7717/peerj-cs.2286>
- Yanto, V. D., & Handayani, I. (2024). Implementation of the Clustering K-Means algorithm in determining the rate of Indramayu mango fruit. *Journal of Scientific Research, Education, and Technology (JSRET)*, 3(4), 1929–1938. <https://doi.org/10.58526/jsret.v3i4.609>.