

# CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR MENGGUNAKAN METODE K-MEANS

Gion Andrian<sup>1</sup>, Teny Handhayani<sup>2</sup>, Desi Arisandi<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: <sup>1</sup>gion.535200024@stu.untar.ac.id, <sup>3</sup>desia@fti.untar.ac.id

## Abstrak

Peran meteorologi dalam memahami pola iklim dan dampak perubahan iklim global menjadi fokus untuk mendeteksi dini perubahan iklim, terutama dampak seriusnya pada kehidupan manusia dan sektor ekonomi di kota-kota seperti Jakarta, Semarang, dan Surabaya. Studi ini difokuskan pada wilayah Indonesia timur, termasuk Papua, Maluku, dan Nusa Tenggara, dengan tujuan mengidentifikasi pola perubahan iklim menggunakan metode clustering, khususnya K-Means. Toleransi missing value sebesar 40% memiliki pengaruh besar dengan silhouette score mencapai 0.509. Penggunaan Z-Score dan penghapusan variabel arah angin maksimum juga terbukti efektif. Hasil analisis dua cluster membentuk kelompok berbeda, terutama Cluster 0 yang hanya memiliki satu kota. Perbedaan signifikan terlihat pada suhu, kelembaban, curah hujan, lama penyinaran matahari, dan kecepatan angin antar cluster, menggambarkan pola iklim yang konsisten namun keragaman kondisi meteorologi di wilayah tersebut.

**Kata kunci** – Meteorologi, Iklim, Indonesia Timur, K-means

## Abstract

*The role of meteorology in understanding climate patterns and the impacts of global climate change takes center stage in early detection of climate change, particularly its serious effects on human life and the economic sector in cities such as Jakarta, Semarang, and Surabaya. This study focuses on the eastern region of Indonesia, including Papua, Maluku, and Nusa Tenggara, with the aim of identifying climate change patterns using clustering methods, specifically K-Means. A 40% tolerance for missing values has a significant impact, with a silhouette score reaching 0.509. The use of Z-Score and the removal of the maximum wind direction variable have also proven effective. The analysis results in two clusters forming distinct groups, especially Cluster 0, which only includes one city. Significant differences are observed in temperature, humidity, rainfall, sunshine duration, and wind speed between clusters, depicting a consistent yet diverse meteorological condition pattern in the region.*

**Keywords** – Meteorology, Climate, Eastern Indonesia, K-means

## 1. PENDAHULUAN

Meteorologi merupakan ilmu yang mempelajari atmosfer, berkaitan erat dengan iklim melalui parameter seperti temperatur, kelembaban, dan curah hujan, yang menjadi dasar untuk memahami pola iklim suatu wilayah [1]. Perubahan iklim global telah menjadi sorotan utama, dengan proyeksi peningkatan suhu global sekitar 1,0°C hingga 1,5°C antara tahun 2030 dan 2052. Peningkatan ini menandakan adanya perubahan iklim yang berpotensi memberikan dampak serius

pada kehidupan manusia, terutama terhadap 68,5 juta orang, dengan kerugian ekonomi diperkirakan mencapai \$131,7 miliar. Sekitar 93% dari kerugian ekonomi tersebut disebabkan oleh badai, banjir, kebakaran hutan, dan kekeringan [2]. Pengetahuan tentang meteorologi sangat penting dalam mengidentifikasi dan memahami bagaimana faktor-faktor meteorologis berperan dalam berbagai sektor [3]. Salah satu cara untuk mengatasi perubahan iklim dan memitigasi dampaknya adalah dengan menggunakan pengetahuan meteorologi untuk analisis data. Analisis data meteorologi menjadi kunci untuk memahami karakteristik iklim suatu wilayah dan memungkinkan deteksi dini terhadap perubahan iklim.

Perubahan iklim, sebagai ancaman global, merusak keberlanjutan sektor-sektor seperti pertanian, keanekaragaman hayati, dan pariwisata. Pentingnya keterlibatan pemerintah dalam menghadapi dampaknya melalui kebijakan iklim yang canggih. Perubahan iklim menjadi prioritas serta komitmen global diperlukan untuk memastikan keberlanjutan dunia [4]. Indonesia sendiri tidak luput dari dampak perubahan iklim, terutama di kota-kota seperti Jakarta, Semarang, dan Surabaya. Kenaikan permukaan air laut, banjir, dan cuaca ekstrem menjadi masalah serius yang diperparah oleh kerusakan lingkungan, perubahan penggunaan lahan, dan urbanisasi yang tidak terkendali [5]. Oleh karena itu, perlu adanya upaya bersama baik dari tingkat nasional maupun internasional untuk mengatasi dampak perubahan iklim demi keberlanjutan lingkungan dan masyarakat.

Papua, Maluku, dan Nusa Tenggara, bagian dari Indonesia timur yang menonjol dengan keanekaragaman geografis. Meskipun terjaga dari kerusakan lingkungan namun dampak dari perubahan iklim mulai terasa serta didukung dengan adanya pemerataan ekonomi, terutama dengan mencairnya salju di Gunung Jayawijaya. Penambangan emas di Papua dan nikel di Maluku memperparah dampak perubahan iklim. Meski demikian, kondisi iklim yang masih baik memberikan peluang untuk adaptasi dan mitigasi yang lebih efektif [6, 7]. Pemanfaatan teknik *clustering*, dapat mengidentifikasi pola perubahan iklim yang khusus untuk setiap kelompok wilayah. Hal ini memungkinkan pemerintah dan para ahli untuk mengembangkan strategi adaptasi dan mitigasi yang sesuai dengan kondisi iklim setiap kelompok, termasuk Papua, Maluku, dan Nusa Tenggara. Pengelompokan ini memungkinkan penggunaan sumber daya yang lebih efisien dan pembangunan kebijakan yang lebih tepat sasaran untuk melindungi lingkungan dan masyarakat setempat.

Pemanfaatan teknik *clustering* dengan data meteorologi dengan menggunakan *K-means* juga digunakan untuk menganalisis zona iklim di Spanyol menggunakan enam teknik *clustering* (*K-means*, *K-medoids*, *SOM k-means*, *Agglomerative Hierarchical Clustering*, dan *Clustering berdasarkan Gaussian Mixture Models*) pada data meteorologi dari stasiun cuaca selama periode enam tahun menghasilkan perbedaan signifikan dalam kondisi meteorologi antara lokasi Burgos dan tiga lokasi lainnya, sementara Palma de Mallorca dan Almeria memiliki data meteorologi harian yang sangat mirip. Teknik *clustering* terbaik adalah *K-means* dalam hal beban komputasi [8]. Sistem ini bertujuan untuk menganalisis pola iklim dan perubahan iklim di kota-kota Indonesia Timur menggunakan metode *K-means*. Tujuan utama adalah menyajikan interpretasi hasil *clustering* dengan data pola iklim dan perubahannya, yang diharapkan dapat dimanfaatkan oleh para ahli di berbagai sektor. Dengan mengidentifikasi karakteristik kota-kota tersebut, diharapkan mampu memberikan dasar bagi pengembangan strategi adaptasi dan mitigasi terhadap perubahan iklim. Selain itu, diharapkan dapat memberikan panduan keputusan bagi pemangku kepentingan dalam menghadapi tantangan perubahan iklim.

## 2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berasal dari data meteorologi wilayah Indonesia Timur, diperoleh dari website resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dengan URL <https://dataonline.bmkg.go.id>. Data tersebut mencakup informasi meteorologi harian dari 33 stasiun cuaca di 6 provinsi wilayah Indonesia Timur (Maluku, Maluku Utara, Nusa Tenggara Barat, Papua, dan Papua Barat), mulai dari Januari 2010 hingga Agustus 2023 dimana setiap stasiun cuaca mewakili satu kota di wilayah tersebut. Dari total 41 stasiun yang awalnya direncanakan, hanya 33 stasiun yang dapat digunakan, sementara 8 stasiun lainnya memiliki missing value lebih dari 4 tahun dan tidak dapat digunakan dalam penelitian ini. Data yang dikumpulkan mencakup informasi tentang suhu minimum, suhu maksimum, suhu rata-rata, kelembaban, curah hujan, penyinaran matahari, kecepatan angin maksimum, arah angin kecepatan maksimum, dan kecepatan angin rata-rata. Sebanyak 163,703 data dengan *missing value* berkisar rata-rata setiap kota berkisar 3.606% hingga 26.803% dan jumlah *missing value* tertinggi untuk setiap variabel untuk masing-masing kota adalah 78.160%. Metode yang diterapkan untuk merancang sistem analisis kluster data meteorologi wilayah Indonesia Timur menggunakan metode *K-means* dipaparkan di bawah.

### 2.1 Data Pre-processing

*Preprocessing* adalah serangkaian langkah persiapan data sebelum analisis. Tujuannya adalah mengatasi kesalahan, memastikan kebersihan, dan mengoptimalkan struktur data. Tahapannya meliputi *data cleaning*, *data integration*, *data selection*, dan *data transformation* untuk memastikan kualitas dan relevansi data sebelum dilakukan analisis lebih lanjut. Pada tahap *data cleaning*, dilakukan identifikasi dan penanganan terhadap *missing value*, terutama yang disebabkan oleh kesalahan pengukuran dengan mengganti nilai yang tidak valid dengan *missing value* [9]. *Data noise* yang terjadi pada beberapa atribut pada data meteorologi yang akan digunakan adalah dengan nilai 8888 dan 9999 dihapus untuk menghindari gangguan dalam analisis, dan *data outlier* diidentifikasi serta diubah menjadi *missing value* agar tidak memengaruhi proses *clustering*. *Outlier* merupakan data yang memiliki karakteristik sangat berbeda dari data lain dalam *dataset*. *Outlier* dapat mempengaruhi statistik deskriptif dan analisis statistik, serta dapat memberikan informasi tentang kemungkinan adanya masalah dalam data atau proses yang diamati.

Data yang mengalami *missing value* dapat dilakukan tangani dengan menggunakan metode *backward fill* dan *forward fill* yang biasanya digunakan dalam mengisi *missing value data time series*. Tahap selanjutnya adalah normalisasi dalam *preprocessing* data bertujuan untuk mengubah skala data sehingga data memiliki rentang yang seragam dan dapat dibandingkan secara lebih efektif. Salah satu teknik normalisasi yang umum digunakan adalah *Min-Max scaling* dengan rentang nilai 0 hingga 1 menggunakan rumus (1) dan *z-score normalization* mengubah skor relevansi ke distribusi normal standar dengan *mean* 0 dan *standard deviation* dengan persamaan (2).

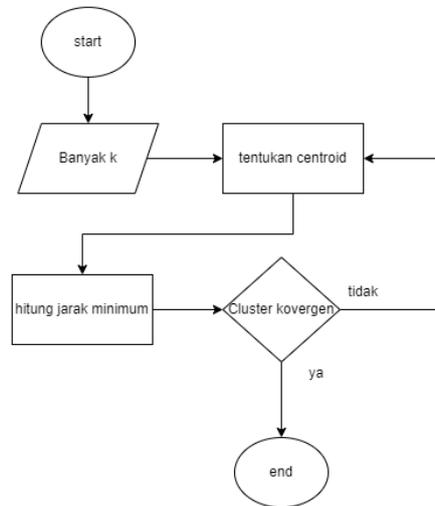
$$v = \frac{v - \min A}{\max A - \min A} \times (new_{max} - new_{min}) + new_{min} \quad (1)$$

$$v = \frac{v - a}{\sigma A} \quad (2)$$

### 2.2 K-means

Algoritma *K-means* merupakan metode *clustering* yang pertama kali diusulkan oleh Stuart Lloyd dan dikemukakan oleh Macqueen 1967 dan mulai digunakan dalam *data mining* pada tahun 1980. *K-means* adalah algoritma analisis data non-hierarki yang membagi M titik dalam N dimensi menjadi K *cluster* dengan tujuan untuk meminimalkan jumlah kuadrat dalam kluster

(*within-cluster sum of squares*). Algoritma ini mencari solusi yang bersifat "konvergen" di mana tidak ada perpindahan titik dari satu kluster ke kluster lain [12].



Gambar 1 Diagram Algoritma K-means

Algoritma *K-means* memerlukan beberapa tahap untuk mendapatkan hasil *clustering* yang konvergen untuk lebih jelasnya dapat dijelaskan pada gambar 1 dengan penjelasan dari diagram algoritma *K-means* adalah sebagai berikut [13]:

1. Menentukan nilai *k* atau jumlah *cluster* yang ingin dibentuk
2. Melakukan inisialisasi pusat *cluster* (*centroid*) biasanya dilakukan dengan cara acak (*random*) atau dengan *centroid* pertama dipilih secara acak, dan *centroid* berikutnya dipilih dengan probabilitas sebanding dengan kuadrat jaraknya terhadap *centroid* yang sudah ada (*K-means++*).
3. Melakukan perhitungan jarak terhadap *centroid* dengan menggunakan *Euclidean Distance* dengan persamaan (3)

$$Dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

$x_i$  dan  $y_i$  adalah dua titik yang ingin dihitung jaraknya pada variabel ke  $i$  dan  $n$  adalah jumlah dimensi.

4. Klasifikasikan berdasarkan hasil perhitungan jarak terkecil dari *centroid*.
5. Hitung *centroid* terbaru berdasarkan anggota *cluster* dengan persamaan (4)

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (4)$$

$v_{ij}$  adalah *centroid* kluster ke- $i$  untuk variabel ke- $j$   $N_i$  adalah jumlah data anggota kluster ke- $i$ ,  $ik$  adalah indeks dari kluster  $j$  merupakan indeks dari variabel  $X_{kj}$  nilai data ke- $k$  yang ada di dalam kluster tersebut untuk variabel ke- $j$ .

### 2.3 Silhouette score

*Silhouette score* digunakan untuk mengukur sejauh mana suatu pengelompokan (*clustering*) pada suatu kumpulan data memiliki kepadatan dan pemisahan yang baik antar kluster. Koefisien ini membantu mengevaluasi kualitas pengelompokan dengan memperhitungkan seberapa dekat titik data pada kluster tertentu dengan kluster lainnya. Penggunaan umum

Koefisien Siluet adalah untuk menilai performa pengelompokan dalam sebuah kumpulan data dan membandingkannya dengan pengelompokan lain dalam kumpulan data yang sama .

1. Perhitungan jarak rata-rata dari objek ke-I dengan semua objek yang berdalarn dalam satu klaster. Menggunakan persamaan nomor (5)

$$a(i) := \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (5)$$

Keterangan:

$A$  = banyak data pada *cluster* A

$j$  = objek lain dalam suatu *cluster* A

$d(i, j)$  = jarak objek ke- $i$  dengan  $j$

2. Perhitungan jarak rata-rata dari objek ke-I terhadap semua obek di luar klaster (berbeda klaster) dengan persamaan 6.

$$d(i, C) := \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (6)$$

Setelah melakukan perhitungan  $d(i, C)$  untuk semua klaster C ambilah nilai yang terkecil dengan persamaan 7.

$$b(i) := \min_{C \neq A} d(i, C) \quad (7)$$

3. Setelah didapat kan hasil  $a(i)$  dan  $b(i)$  makan dilakukan perhitungan nilai koefisien siluet dengan persamaan 8.

$$s(i) := \frac{b(i)-a(i)}{\max \{a(i), b(i)\}} \quad (8)$$

#### 4. HASIL DAN PEMBAHASAN

Pengujian klaster dalam penelitian ini akan menggunakan data meteorologi dari 31 kota, dengan toleransi *missing value* sebesar 45% pada 9 variabel, yang totalnya mencapai 164703 data, sebagaimana terlihat pada Tabel 1 untuk curah hujan. Variabel dengan nilai standar deviasi yang tinggi, seperti arah angin dan kecepatan maksimum, menunjukkan adanya variasi yang signifikan dalam data. Tingginya jumlah *missing value* dalam *dataset* mengharuskan pengujian mempertimbangkan beberapa skenario toleransi *missing value* untuk memilih yang terbaik.

**Tabel 1** Statistik Deskriptif

<i>Variable</i>	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Std.Deviation</i>
suhu minimum	164703	8	31	23,34758	2,558141
suhu maksimum	164703	15,1	37,7	31,22133	2,142468
suhu rata rata	164703	16,1	33,1	26,76819	2,18113
kelembaban	164703	50	100	82,87939	6,62635
curah hujan	164703	0	199,5	7,62702	15,53464
penyinaran matahari	164703	0	12	5,642554	3,236621
kecepatan angin maksimum	164703	0	40	5,179505	2,026374
arah angin kecepatan maksimum	164703	30	360	199,8971	96,75144
kecepatan angin rata-rata	164703	0	10	2,239212	1,228011

Arah angin kecepatan maksimum memiliki rentang nilai antara 0 hingga 360, oleh karena itu, pengujian ini melibatkan beberapa skenario tanpa mempertimbangkan variabel arah angin kecepatan maksimum. Pembuatan model *K-means* dilakukan melalui serangkaian eksperimen

dengan berbagai skenario, dengan tujuan mendapatkan nilai *silhouette score* tertinggi. Sebanyak 10 skenario diuji, masing-masing dicontohkan dalam Tabel 2, dan 5 skenario tambahan menggunakan data meteorologi tanpa variabel arah angin kecepatan maksimum, seperti yang tercantum pada Tabel 3.

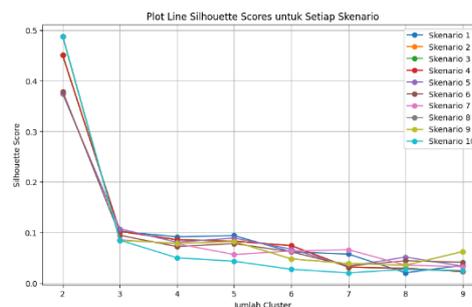
**Tabel 2** Skenario 9 variabel

skenario	Normalisasi	Random state	Algoritma	Init	Max iterasi	Toleransi Missing value
1	Z-Score	0	auto	K-Means++	300	missing value <45%
2	Z-Score	200	auto	K-Means++	1000	missing value <45%
3	Z-Score	200	lloyd	K-Means++	1000	missing value <45%
4	Z-Score	200	full	K-Means++	1000	missing value <45%
5	Min-Max	0	auto	K-Means++	300	missing value <45%
6	Min-Max	0	auto	K-Means++	300	missing value <40%
7	Z-Score	0	auto	K-Means++	300	missing value <40%
8	Z-Score	200	auto	K-Means++	1000	missing value <40%
9	Z-Score	200	full	K-Means++	1000	missing value <40%
10	Z-Score	200	full	random	0	missing value <40%

**Tabel 3** Skenario 8 variabel

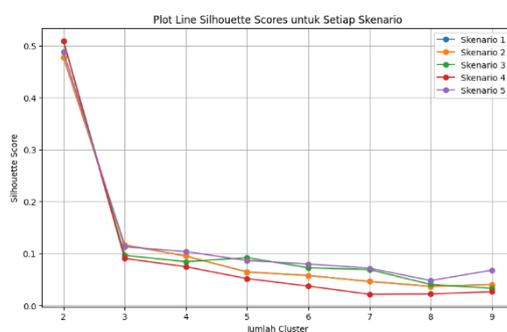
skenario	Normalisasi	Random state	Algoritma	Init	Max iterasi	Toleransi Missing value
1	Z-Score	0	auto	K-Means++	300	missing value <45%
2	Z-Score	200	full	K-Means++	1000	missing value <45%
3	Z-Score	0	auto	K-Means++	300	missing value <40%
4	Z-Score	200	full	random	1000	missing value <40%
5	Z-Score	0	full	K-Means++	300	missing value <36%

### 3.1 Pengujian K-means



**Gambar 2** Perbandingan silhouette score untuk setiap skenario 9 variabel

Hasil pada skenario 1, yang tergambar pada gambar 2, menunjukkan bahwa skenario ini mencapai hasil tertinggi dengan jumlah *cluster* 2, menghasilkan nilai *silhouette score* sebesar 0.4501. Namun, skor tersebut mengalami penurunan signifikan ketika jumlah *cluster* ditingkatkan menjadi 3, dan terus turun seiring dengan peningkatan jumlah *cluster*. Hasil pada skenario 2 hingga 4 cenderung identik dengan skenario 1, menunjukkan bahwa penggunaan algoritma penuh (full), metode *Lloyd*, dan pengaturan maksimal iterasi, serta pengaturan *state* acak (random state), tidak memiliki pengaruh signifikan. Pada skenario 5, terjadi sedikit penurunan nilai *silhouette score* menjadi 0.374, kemungkinan disebabkan oleh normalisasi menggunakan *Min-Max scaler*. Hasil tersebut juga identik dengan skenario 6, meskipun menggunakan toleransi *missing value* sebesar 40%. Pada skenario 7, *silhouette score* kembali naik pada jumlah *cluster* 2, mencapai 0.487. Skenario 8 hingga 10 menunjukkan hasil yang identik, dengan semua skenario menunjukkan bahwa nilai *silhouette* tertinggi terjadi pada jumlah *cluster* 2. Peningkatan jumlah *cluster* seringkali diikuti oleh penurunan skor *silhouette*. Konfigurasi *K-means* tidak berpengaruh secara signifikan, namun toleransi terhadap *missing value* cukup memengaruhi hasil..



**Gambar 3** Perbandingan silhouette score untuk setiap skenario 8 variabel

Dapat dilihat pada gambar 3 hasil perbandingan menggunakan data meteorologi dengan 8 variabel pada gambar 3, dapat disimpulkan bahwa skenario ke-3 dan ke-4 memberikan hasil tertinggi dengan menggunakan toleransi *missing value* sebesar 40%. Kedua skenario ini menghasilkan nilai *silhouette score* sebesar 0.509, yang merupakan nilai tertinggi di antara semua skenario yang diuji. Hasil ini konsisten dengan temuan pada 10 skenario sebelumnya dengan 9 variabel, di mana semua skenario dengan 8 variabel memiliki nilai *silhouette* tertinggi pada jumlah *cluster* 2 dan mengalami penurunan seiring dengan peningkatan jumlah *cluster*. Namun, perlu diperhatikan bahwa pada skenario ke-5, terjadi penurunan toleransi *missing value* menjadi 36%, namun hasilnya cenderung lebih menurun dibandingkan dengan toleransi *missing value* 40%. Kesimpulannya, pengaruh toleransi *missing value* pada pengujian ini sangat berpengaruh terhadap hasil, dan toleransi 40% memberikan hasil *silhouette score* terbaik sebesar 0.509.

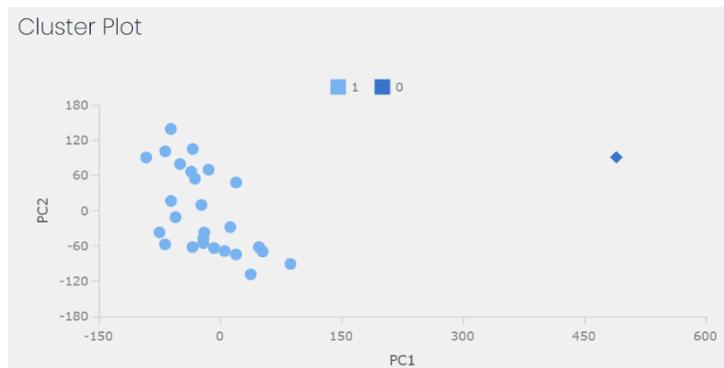
Kesimpulan dari semua skenario adalah bahwa hasil terbaik untuk *silhouette score* adalah 0.509. Ini dicapai dengan menggunakan 40% toleransi untuk *missing value* dan menerapkan standar *K-Means*. Penggunaan parameter seperti max iterasi, algoritma, *random state*, dan inti juga telah diuji, dan hasilnya tetap konsisten. Dari analisis tersebut, dapat disimpulkan bahwa jumlah optimal *cluster* yang dapat dibentuk adalah 2 *cluster*. Peningkatan jumlah *cluster* di atas 2 menghasilkan penurunan yang signifikan dalam kualitas hasil. Selain itu, perlu dicatat bahwa toleransi terhadap *missing value* memiliki pengaruh besar terhadap hasil pengujian ini.

### 3.2 Analisis Hasil Clustering

Berdasarkan hasil pengujian dengan berbagai skenario menghasilkan nilai *silhouette score* tertinggi 0.509 dengan jumlah 2 *cluster* menghasilkan anggota *cluster* yang dapat dilihat pada tabel 4. *Cluster* 0 hanya memiliki 1 kota dan *cluster* 1 memiliki 27 kota.

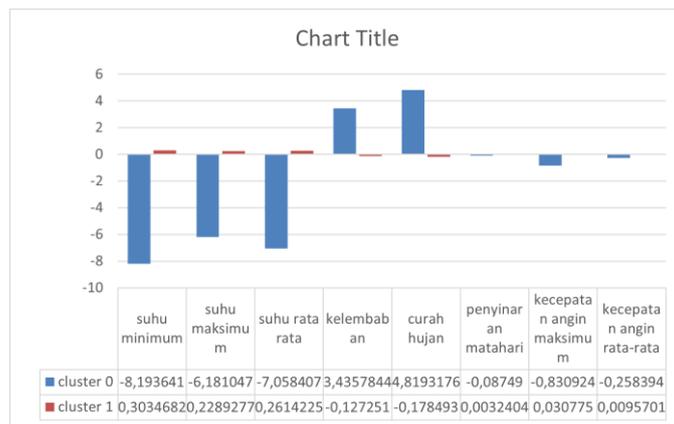
Tabel 4 Daftar kota percluster

Cluster 0	Cluster 1
Kab Manggarai	Kab Maluku Tengah, Kab Maluku Tenggara, Kab Maluku Tenggara Barat, Kab Seran Timur, Kota Ambon, Kab Halmahera Selatan, Kab Halmahera Utara, Kab Kepulauan Sula, Kota Ternate, Kota Bima, Kab Sumbawa, Kota Mataram Kab Alor, Kab Flores Timur, Kab Kupang Kab Manggarai Barat, Kab Sikka Kab Sumba Timur, Kota Kupang Kab Biak Numfor, Kab Jayapura, Kab Merauke, Kab Mimika, Kab Nabire, Kota Jayapura Kab Fak Fak, Kota Sorong



Gambar 4 Visualisasi hasil clustering

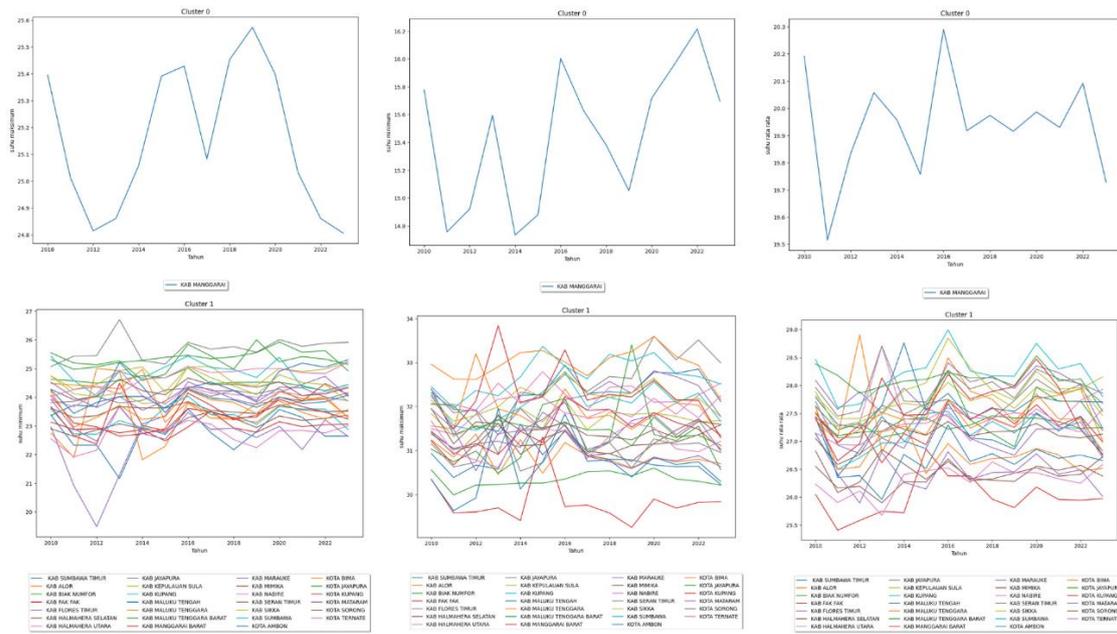
Visualisasi persebaran cluster dapat dilihat pada Gambar 4, di mana Cluster 0 terletak jauh dari Cluster 1, dan anggota Cluster 1 tampak berdekatan.



Gambar 5 Perbandingan cluster dengan rata-rata

Pada Gambar 5, perbandingan antar cluster menunjukkan bahwa Cluster 0 memiliki suhu yang jauh dibawah rata-rata, baik suhu minimum hingga rata-rata, dengan kelembaban rata-rata, curah hujan rata-rata atau rendah dibandingkan dengan Cluster 1. Selain itu, lama penyinaran matahari rata-rata, dan kecepatan angin sedikit dibawah rata-rata. Di sisi lain, Cluster 1 memiliki suhu sedikit lebih mendekati rata-rata, baik suhu minimum hingga suhu rata-rata, serta memiliki curah hujan dan kelembaban yang melebihi rata-rata. Pada Cluster 1, penyinaran matahari dan kecepatan angin juga cenderung lebih tinggi dibandingkan dengan Cluster 0.

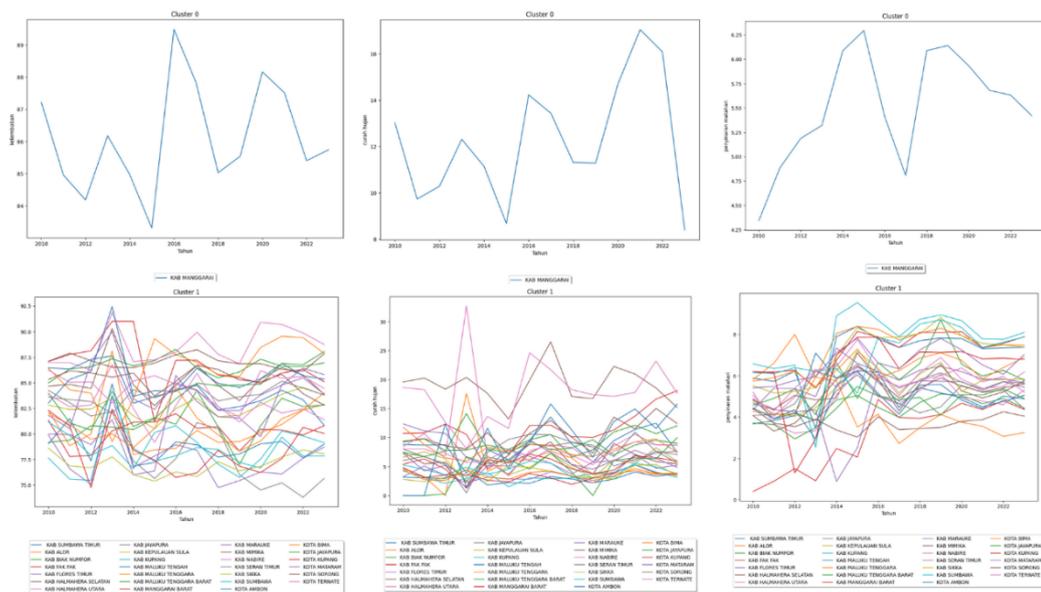
# Gion Andrian: Clustering Data Meteorologi Wilayah Indonesia Timur Menggunakan Metode K-Means



Gambar 6 Grafik tahunan suhu

Berdasarkan Gambar 6, beberapa temuan dapat diidentifikasi:

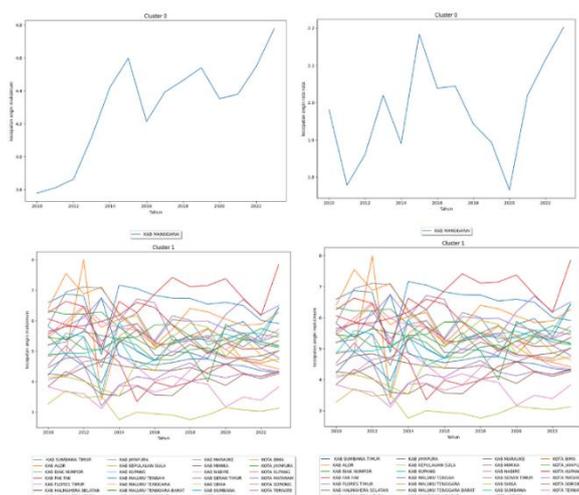
1. *Cluster 0*: Hanya terdiri dari satu anggota, yaitu Kabupaten Manggarai. Kabupaten ini menunjukkan fluktuasi suhu minimum dan maksimum yang cenderung naik setiap tahunnya. Pada suhu rata-rata, terjadi kenaikan signifikan pada tahun 2016.
2. *Cluster 2*: Memperlihatkan penurunan suhu minimum yang cukup signifikan pada Kabupaten Flores Timur pada tahun 2012. Sebaliknya, terdapat kenaikan suhu pada Kabupaten Jayapura pada tahun yang sama, serta beberapa kota lain di dalam *cluster 2*. Fluktuasi suhu minimum terjadi, meskipun tidak terlalu ekstrem, dan cenderung naik setiap tahunnya. Sementara itu, fluktuasi suhu maksimum yang ekstrem tercatat di beberapa kota pada tahun tertentu.
3. Suhu rata-rata, baik pada *Cluster 1* maupun *Cluster 2*: Menunjukkan penurunan yang sangat landai dan fluktuatif.



Gambar 7 Visualisasi hasil clustering curah hujan, kelembaban, penyinaran

Berdasarkan gambar 7 beberapa temuan dapat diidentifikasi:

1. *Cluster 0* menunjukkan tingkat kelembaban yang tinggi pada tahun 2016. Di sisi lain, *cluster 1* mengalami kelembaban tinggi secara umum pada tahun 2014, dengan hampir semua kota atau kabupaten di dalamnya terkena dampak tersebut.
2. *Cluster 0* mengalami penurunan curah hujan yang ekstrem pada tahun 2022. Sebaliknya, pada *cluster 1*, Kabupaten Halmahera Selatan mengalami kenaikan curah hujan pada tahun 2013.
3. *Cluster 0* juga mencatat penurunan curah hujan yang ekstrem.



Gambar 8 kecepatan angin

Dalam konteks kecepatan angin, dapat diamati perbedaan pada masing-masing *cluster*:

1. *Cluster 0*: Mengalami kenaikan kecepatan angin maksimum setiap tahun, dan kecepatan angin rata-rata, meskipun mengalami penurunan ekstrim pada tahun 2020.
2. *Cluster 1*: Menunjukkan fluktuasi kecepatan angin yang signifikan, baik pada kecepatan angin maksimum maupun rata-rata. Terdapat kecenderungan untuk mengalami kenaikan yang landai dari tahun ke tahun, tanpa fluktuasi yang ekstrim.

Analisis menunjukkan perbedaan mencolok antara *Cluster 0* dan *Cluster 1* dalam berbagai parameter iklim. Variasi signifikan terlihat pada suhu, kelembaban, curah hujan, lama penyinaran matahari, dan kecepatan angin antara kedua *cluster*. Terdapat fluktuasi dalam beberapa parameter meteorologi untuk masing-masing *cluster*, dan anggotanya mengalami fluktuasi yang hampir serupa untuk setiap *cluster*. Hal ini mengindikasikan adanya pola iklim yang konsisten dalam setiap *cluster*, sementara variasi antara *cluster* menunjukkan keragaman kondisi meteorologi di wilayah tersebut dan terjadinya kenaikan maupun penurunan pada setiap kota dalam anggota *cluster*.

#### 4. KESIMPULAN

Berdasarkan hasil yang didapatkan dari analisis *cluster* data meteorologi wilayah Indonesia timur dapat disimpulkan dengan poin berikut:

1. Terbaik dengan *silhouette score* 0.509, menunjukkan bahwa konfigurasi dalam *K-means* kurang signifikan, tapi toleransi *missing value* 40% berpengaruh. Penggunaan *Z-Score* lebih efektif, dan penghapusan variabel arah angin maksimum berpengaruh besar.
2. Analisis 2 *cluster* berhasil membentuk dua kelompok yang berbeda, dengan *Cluster 0* hanya memiliki 1 kota atau kabupaten. Karakteristik antar *cluster* terlihat berbeda.

3. Variasi signifikan pada suhu, kelembaban, curah hujan, lama penyinaran matahari, dan kecepatan angin antara kedua *cluster*. Fluktuasi serupa dalam parameter meteorologi di setiap *cluster* menunjukkan pola iklim konsisten, dengan variasi menunjukkan keragaman kondisi meteorologi di wilayah tersebut. Terjadi kenaikan atau penurunan pada suhu, curah hujan, kelembaban, dan lama penyinaran matahari di beberapa kota dalam anggota *cluster*, menunjukkan dinamika iklim yang beragam di setiap wilayah.

#### DAFTAR PUSTAKA

- [1] H. H. Frisinger, *History of meteorology to 1800*. Springer, 2018.
- [2] S. Fawzy, A. I. Osman, J. Doran, and D. W. Rooney, "Strategies for mitigation of climate change: a review," *Environ Chem Lett*, vol. 18, pp. 2069–2094, 2020.
- [3] Z. Chen *et al.*, "Influence of meteorological conditions on PM2. 5 concentrations across China: A review of methodology and mechanism," *Environ Int*, vol. 139, p. 105558, 2020.
- [4] K. Abbass, M. Z. Qasim, H. Song, M. Murshed, H. Mahmood, and I. Younis, "A review of the global climate change impacts, adaptation, and sustainable mitigation measures," *Environmental Science and Pollution Research*, vol. 29, no. 28, pp. 42539–42559, 2022.
- [5] E. Hermawan, "Fenomena urban heat island (UHI) pada beberapa kota besar di Indonesia sebagai salah satu dampak perubahan lingkungan global," *Jurnal Citra Widya Edukasi*, vol. 7, no. 1, pp. 33–45, 2015.
- [6] N. Sudapet, A. Sukoco, and M. I. Setiawan, "Model Integrasi Ekonomi Maritim Dan Pariwisata di Daerah Guna Peningkatan Ekonomi Indonesia Timur," *Jurnal Darussalam: Jurnal Pendidikan, Komunikasi dan Pemikiran Hukum Islam*, vol. 9, no. 1, pp. 148–160, 2017.
- [7] T. SUSENO, "ANALISIS NILAI SUMBER DAYA BIJIH BAUKSIT, NIKEL DAN EMAS PT. ANTAM TBK.," *Jurnal Teknologi Mineral dan Batubara*, vol. 6, no. 4, pp. 174–182, 2010.
- [8] Á. Arroyo, Á. Herrero, V. Tricio, and E. Corchado, "Analysis of meteorological conditions in Spain by means of clustering techniques," *Journal of Applied Logic*, vol. 24, pp. 76–89, 2017.
- [9] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*, vol. 72. Springer, 2015.
- [10] S. Walfish, "A review of statistical outlier methods," *Pharmaceutical technology*, vol. 30, no. 11, p. 82, 2006.