Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

Pengelompokkan Pola Perubahan Cuaca Menggunakan Metode K-Medoids dan Gap Statistic

Denissya Julianthy*, Asep Id Hadiana, Edvin Ramadhan

Fakultas Sains dan Informatika, Program Studi Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia
Email: [1,*]denissyaj21@if.unjani.ac.id, [2]ahadiana@gmail.com, [3]edvin.ramadhan@gmail.com,
Email Penulis Korespondensi: denissyaj21@if.unjani.ac.id
Submitted: **29/06/2025**; Accepted: **01/09/2025**; Published**: 02/09/2025**

Abstrak−Pengelompokan pola cuaca harian merupakan proses penting untuk memahami variasi cuaca yang kompleks. Namun, metode yang sering digunakan seperti K-Means memiliki kelemahan karena sensitif terhadap outlier dan memerlukan penentuan jumlah klaster secara manual. Penelitian ini mengusulkan kombinasi metode K-Medoids dan Gap Statistic untuk menghasilkan klaster yang lebih stabil dan akurat. Data cuaca harian Semarang dari tahun 2017 hingga 2023 diproses melalui pembersihan, standarisasi dengan Standard Scaler, dan reduksi dimensi menggunakan PCA. Hasil Gap Statistik menunjukkan jumlah klaster optimal sebanyak tiga, yaitu hujan, cerah, dan berawan. Validasi klustering menghasilkan Silhouette Score sebesar 0.3793, Calinski-Harabasz Index sebesar 1490.5604, dan Davies-Bouldin Index sebesar 0.9031. Hasil tersebut menunjukkan struktur klaster cukup baik, meskipun masih terdapat ruang untuk perbaikan terutama pada pemisahan antar klaster.

Kata Kunci: Cuaca Harian; Klustering; Validasi Klaster; Gap Statistik; K-Medoids

Abstract−Clustering daily weather patterns is an important process for understanding complex weather variations. However, commonly used methods such as K-Means have weaknesses due to their sensitivity to outliers and the need for manual clustering. This study proposes a combination of the K-Medoids and Gap Statistics methods to produce more stable and accurate clusters. Semarang's daily weather data from 2017 to 2023 was processed through cleaning, standardization with Standard Scaler, and dimensionality reduction using PCA. The Gap Statistics results indicate the optimal number of clusters is three: rainy, sunny, and cloudy. The clustering evaluation yielded a Silhouette Score of 0.3793, a Calinski-Harabasz Index of 1490.5604, and a Davies-Bouldin Index of 0.9031. These results indicate a fairly good cluster structure, although there is still room for improvement, especially in the separation between clusters.

Keywords: Daily Weather; Clustering; Cluster Validation; Gap Statistics; K-Medoids

1. PENDAHULUAN

Cuaca merupakan keadaan fisik atmosfer pada suatu waktu dan tempat tertentu, yang meliputi parameter seperti suhu udara, curah hujan, kelembapan, tekanan udara, dan kecepatan angin [1]. Dalam konteks iklim tropis seperti Indonesia, dinamika cuaca sangat kompleks akibat pengaruh faktor geografis dan variabilitas iklim global seperti El Niño–Southern Oscillation (ENSO) [2]. Ketidakpastian cuaca ekstrem dapat memengaruhi berbagai sektor strategis, seperti pertanian, transportasi, dan pengelolaan sumber daya air [3]. Oleh karena itu, analisis pola perubahan cuaca berbasis data historis menjadi krusial untuk mendukung perencanaan mitigasi dan adaptasi terhadap dampak iklim [4].

Salah satu pendekatan yang digunakan dalam analisis data cuaca adalah metode clustering, yaitu pengelompokan data berdasarkan kemiripan karakteristiknya tanpa perlu label kelas [5]. Clustering sangat bermanfaat dalam mengidentifikasi pola tersembunyi pada data meteorologi multivariat, seperti suhu, kelembapan, dan curah hujan [6]. Algoritma K-Means merupakan metode yang paling umum digunakan karena kesederhanaan implementasi dan kecepatan komputasinya [7]. Namun, metode ini memiliki beberapa kelemahan penting, seperti kepekaan terhadap pencilan (outlier) dan pemilihan inisialisasi centroid yang dapat menyebabkan hasil klasterisasi tidak konsisten [8].

Sebagai alternatif dari K-Means, algoritma K-Medoids digunakan karena lebih stabil terhadap outlier dan memilih pusat klaster dari titik data aktual, sehingga menghasilkan segmentasi data yang lebih representatif [9]. Namun, salah satu tantangan dalam clustering adalah menentukan jumlah klaster yang optimal, yang dapat memengaruhi kualitas hasil analisis [6]. Untuk menjawab tantangan ini, Gap Statistic hadir sebagai metode evaluasi kuantitatif yang membandingkan nilai dispersi dalam klaster dengan data acak sebagai baseline, tanpa ketergantungan pada visualisasi subjektif seperti metode elbow [10] . Kombinasi metode K-Medoids dan Gap Statistic menawarkan pendekatan yang lebih robust, konsisten, dan dapat direproduksi dalam analisis segmentasi data cuaca [11].

Meskipun efektivitas metode K-Medoids dan Gap Statistic telah dibuktikan secara individual dalam berbagai bidang, integrasi keduanya dalam pengelompokan data cuaca harian di wilayah tropis seperti Indonesia masih sangat jarang dijumpai [12]. Sebagian besar penelitian terdahulu menggunakan metode heuristik atau pendekatan visual untuk menentukan jumlah klaster, sehingga hasil analisis menjadi subjektif dan sulit diulang secara ilmiah [13]. Studi mengenai pengelompokan data cuaca dengan evaluasi objektif jumlah klaster secara sistematis masih terbatas dan merupakan celah yang layak untuk diteliti lebih lanjut.

Penelitian ini bertujuan untuk mengelompokkan data cuaca harian berdasarkan parameter meteorologi menggunakan metode K-Medoids serta menentukan jumlah klaster optimal dengan pendekatan Gap Statistic [14]. Secara khusus, penelitian ini akan: melakukan transformasi dan normalisasi data cuaca harian (curah hujan, suhu, kelembapan, dan lama penyinaran), mengimplementasikan algoritma K-Medoids dalam proses pengelompokan, menentukan jumlah klaster optimal dengan Gap Statistic, mengevaluasi hasil klaster menggunakan metrik validasi

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

internal seperti Silhouette Score dan Davies-Bouldin Index; serta memberikan interpretasi terhadap hasil klaster sebagai dasar untuk mendukung strategi mitigasi bencana dan adaptasi perubahan iklim yang berbasis data [8].

Dalam penelitian [6]Kelemahan algoritma K-Means adalah jumlah klaster yang dibentuk oleh algoritma ini bersifat tetap, karena pusat klaster dipilih berdasarkan potongan data dalam mapper sehingga klaster yang berbeda terbentuk selama proses yang berbeda untuk set data input yang sama. Penelitian ini mengestimasi jumlah klaster yang dibentuk oleh algoritma pengelompokan berdasarkan kriteria evaluasi statistik gap. Gap statistik adalah suatu metode untuk estimasi banyaknya klaster optimal dalam satu kumpulan data. Gap statistik digunakan pada hasil pengklasteran dari beberapa metode pengklasteran, misalnya metode hirarki dan K-means.

Penelitian ini [15]karena Pulau Kalimantan, sebagai pulau terbesar di Indonesia, memiliki pola iklim yang beragam. Untuk memahami perubahan pola iklim yang semakin kompleks dalam satu dekade terakhir, diperlukan analisis yang tepat, salah satunya melalui metode K-Medoids Clustering. Metode ini dipilih karena mampu mengelompokkan data berdasarkan kesamaan atribut, lebih tahan terhadap data ekstrem, dan menghasilkan kelompok yang lebih stabil dibandingkan K-Means.

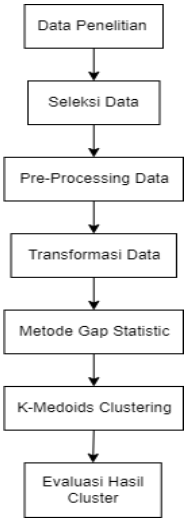
Penelitian ini [14] membahas penerapan metode Principal Component Analysis (PCA) dan Gap Statistik untuk mengatasi permasalahan multikolinieritas dan penentuan jumlah klaster optimal dalam pengelompokan data kanker payudara menggunakan algoritma K-Means. Gap Statistik diterapkan untuk menentukan jumlah klaster yang paling optimal karena K-Means tidak dapat menetapkannya secara otomatis.

Penelitian ini [11]membahas tentang pengelompokan provinsi-provinsi di Indonesia berdasarkan tingkat pengangguran tahun 2023 dengan menggunakan metode k-medoids yang dipadukan dengan principal component analysis (PCA) dan gap statistik. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi kelompok provinsi dengan karakteristik pengangguran yang serupa guna mendukung perumusan kebijakan ketenagakerjaan yang lebih tepat sasaran. Data yang digunakan mencakup sembilan variabel ketenagakerjaan dari seluruh provinsi di Indonesia. Berdasarkan analisis gap statistik, ditemukan bahwa jumlah klaster optimal adalah tiga, yaitu: klaster 1 terdiri dari 24 provinsi dengan tingkat pengangguran rendah, klaster 2 mencakup 7 provinsi dengan tingkat pengangguran sedang, dan klaster 3 terdiri dari 3 provinsi dengan tingkat pengangguran tinggi.

Berdasarkan penelitian-penelitian sebelumnya, metode k-medoids dinilai cocok untuk digunakan dalam penelitian ini yang berjudul "Pengelompokkan Pola Perubahan Cuaca Menggunakan Metode K-Medoids dengan Gap Statistic", karena metode k-medoids mampu mengelompokkan data yang memiliki banyak variasi atau nilai yang sangat berbeda secara lebih stabil dan akurat.. Metode ini dianggap lebih stabil dibandingkan metode lain karena menggunakan data asli sebagai pusat kelompok. Selain itu, gap statistik membantu menentukan jumlah kelompok yang paling pas secara otomatis. Karena itulah, kombinasi k-medoids dan gap statistik dianggap tepat untuk digunakan dalam mengelompokkan pola perubahan cuaca yang berubah-ubah dan kompleks.

2. METODOLOGI PENELITIAN

Metodologi Penelitian menjelaskan proses pada data mining yang digunakan untuk melakukan *clustering*. Data mining, atau yang juga dikenal sebagai knowledge discovery in database (KDD), merupakan proses yang mencakup pengumpulan serta pemanfaatan data historis untuk mengungkap keteraturan atau pola hubungan dalam kumpulan data berukuran besar [1]. Tujuan utama dari data mining adalah menemukan pengetahuan baru yang tersembunyi dan tidak tampak secara langsung dalam data mentah, seperti pola tersembunyi, hubungan sebab-akibat, maupun tren yang tidak terdeteksi melalui analisis konvensional [16]. Berdasarkan jurnal [12], alur metode penelitian yang dilakukan, yaitu:



Gambar 1. Metodologi Penelitian

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

Gambar 1 merupakan Tahapan dalam penelitian dimulai dari pengumpulan dan seleksi data cuaca yang relevan, dilanjutkan dengan pre-processing untuk membersihkan dan mempersiapkan data. Setelah itu, data ditransformasikan agar siap digunakan dalam proses clustering. Metode Gap Statistic digunakan untuk menentukan jumlah cluster yang optimal, kemudian dilakukan pengelompokan menggunakan algoritma K-Medoids. Tahap terakhir adalah evaluasi hasil cluster untuk memastikan kualitas dan interpretasi pengelompokan yang dihasilkan.

2.1 Data Penelitian

Tabel 1. Dataset Penelitian

Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
01/02/2017	25	30,8	26,2	86	4	5,3	6	315	4	NW
02/02/2017	23,8	29,2	25,7	87	24	0,8	5	315	3	NW
03/02/2017	23	30,2	26,1	85	17	2,3	5	315	3	NW
04/02/2017	23,2	30,8	27,4	82	23	4,2	7	315	4	NW
05/02/2017	24,2	31,2	27,3	84	11	5,7	5	315	4	NW
06/02/2017	25	31	27,7	82	0	5,7	8	315	5	NW
07/02/2017	23,8	31	27,9	78	37	5,2	9	270	6	W
08/02/2017	24,6	30,6	21,2	76	17	4,9	10	270	6	W
09/02/2017	23,2	28,4	25,8	89	22	0	8	315	1	NW
10/02/2017	23,2	28,4	25,8	89	22	0	8	315	4	W
11/02/2017	23	26,8	25,1	95	36	0	3	315	8	NW
12/02/2017	24	27,4	24,9	96	16	0,7	6	315	3	NW
13/02/2017	23,8	30,4	26,4	91	15	4,8	6	315	2	NW
14/02/2017	24,4	29,8	26,7	86	18	5,7	6	315	4	NW
15/02/2017	24,4	30,2	26,6	87	20,5	4,9	8	315	3	NW
../../….	…	…	…	…	…	…	…	…	…	…
30/12/2023	24,2	33,2	30	74	97,8	8,4	5	350	2	C
31/12/2023	25,8	33,4	30	78	0	7	4	340	2	N

Adapun contoh data yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1, yang menampilkan sebagian dari dataset cuaca harian yang akan diolah lebih lanjut pada proses clustering. Keterangan atribut: Tanggal: Menunjukkan tanggal pencatatan data cuaca harian, Tn: Temperatur minimum harian (°C), Tx: Temperatur maksimum harian (°C), Tavg: Temperatur rata-rata harian (°C), RH_avg: Kelembapan relatif rata-rata harian (%), RR: Curah hujan harian (mm), Ss: Lama penyinaran matahari dalam satuan jam, ff_x: Kecepatan angin maksimum (m/s), ddd_x: Arah angin saat kecepatan maksimum (dalam derajat), ff_avg: Kecepatan angin rata-rata harian (m/s), ddd_car: Arah angin rata-rata harian (tulisan arah angin seperti NW, W, dll).

2.3 Pre-Prosessing Data

Pre-processing data merupakan tahap awal dalam pengolahan data yang bertujuan untuk membersihkan dan mempersiapkan data agar layak digunakan dalam proses analisis lanjutan. Pada tahap ini dilakukan penghapusan data duplikat dan data yang tidak valid seperti NaN atau kode khusus seperti 8888 untuk mencegah bias pada analisis. Selain itu, dilakukan juga penyesuaian format data, misalnya konversi tipe data dari string ke numerik (float), serta pemilihan variabel yang relevan sesuai kebutuhan analisis. Proses ini bertujuan untuk menghasilkan dataset yang bersih, rapi, dan siap digunakan untuk tahap berikutnya, yaitu transformasi data dan clustering.

2.4 Data Transformasi

Transformasi data merupakan proses lanjutan setelah pre-processing yang bertujuan untuk menyelaraskan skala antar variabel agar memiliki kontribusi yang seimbang dalam proses clustering. Karena variabel data cuaca seperti suhu, kelembapan, dan curah hujan memiliki satuan dan rentang nilai yang berbeda, maka diperlukan proses normalisasi agar tidak terjadi dominasi variabel tertentu. Pada penelitian ini digunakan Standard Scaler, yaitu transformasi data dengan metode z-score yang mengubah nilai data menjadi distribusi dengan rata-rata nol dan standar deviasi satu. Penggunaan Standard Scaler dipilih karena sesuai untuk data yang berdistribusi normal maupun data dengan distribusi yang tidak diketahui, sehingga dapat menghasilkan proses clustering yang lebih akurat dan representatif terhadap pola data sebenarnya.

2.5 K-Medoids *Clustering*

K-Medoids adalah metode pengelompokan data yang membagi dataset ke dalam beberapa klaster dan lebih tahan terhadap data ekstrem dibandingkan K-Means. Metode ini menggunakan medoid, yaitu objek nyata dalam klaster, sebagai pusat klaster, sehingga lebih stabil terhadap perubahan kecil pada data [16]. Proses clustering dilakukan secara iteratif dengan memilih medoid, mengelompokkan data berdasarkan jarak terdekat, dan memperbarui medoid hingga tidak ada perubahan lagi, menandakan klaster telah terbentuk secara optimal[15]. Beberapa langkah-langkah dalam perhitungan algoritma K-Medoids sebagai berikut [2]:

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

a. Inisialisasi sejumlah pusat klaster sebanyak k.
b. Alokasikan setiap objek ke klaster terdekat menggunakan jarak Euclidean, dengan rumus:

$$d_{euc} = \sqrt{\sum_{j=1}^{p} \sum_{i=1}^{n} \left(x_{ij} - c_{kj}\right)^2} \quad (1)$$

Rumus (1) ini merupakan rumus untuk Euclidean Distance yang digunakan untuk mengukur kedekatan antara suatu data dengan pusat klaster dalam ruang multidimensi. Rumus ini menjumlahkan selisih kuadrat antara nilai pengamatan dan pusat klaster pada setiap variabel. Semakin kecil nilai jaraknya, semakin dekat posisi data terhadap pusat klaster tersebut.

c. Pilih secara acak objek dalam klaster sebagai kandidat medoid baru.
d. Hitung jarak setiap objek yang berada pada masing-masing klaster dengan kandidat medoids baru dengan menggunakan rumus Euclidian Distance diatas .
e. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika S<0, maka tukar objek dengan data klaster untuk membentuk sekumpulan k objek baru sebagai medoids.
f. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan klaster beserta anggota klaster masing-masing.

2.6 Menentukan Nilai K Optimal dengan Gap Statistik

Metode Gap Statistik digunakan untuk menentukan jumlah klaster optimal dalam analisis klaster. Metode ini membandingkan hasil *clustering* pada data asli dengan hasil *clustering* pada data acak yang memiliki distribusi yang sama [4]. Perhitungan gap statistik sebagai berikut [17]:

a. Hitung jarak data dengan pusat klaster dengan perhitungan Euclidean distance.

$$D_K = \sum_{x_i \in C_k} \sum_{x_i \in C_k} \left|\left|x_i - x_j\right|\right|^2 = 2n_k \sum_{x_i \in C_k} \left|\left|x_i - \mu_k\right|\right|^2 \quad (2)$$

Rumus (2) ini digunakan untuk menghitung total jarak antar data dalam klaster ke-k terhadap pusat klaster (medoid atau centroid) menggunakan Euclidean distance. Dalam konteks ini, D_K mewakili total jarak dalam klaster ke-k, yang dihitung sebagai dua kali jumlah data dalam klaster ($2n_k$) dikalikan dengan jumlah kuadrat dari jarak setiap titik data x_i terhadap pusat klaster μk. Nilai $\left|\left|x_i - x_j\right|\right|^2$ merupakan kuadrat dari jarak Euclidean antara titik data dan pusat klaster. Pendekatan ini memberikan ukuran seberapa kompak sebuah klaster, yaitu seberapa dekat anggota-anggotanya terhadap pusat klaster. Semakin kecil nilai DK, maka semakin padat dan baik kualitas klasternya. Rumus ini biasa digunakan dalam metode clustering seperti K-Medoids atau K-Means untuk mengevaluasi hasil pembentukan klaster.

b. Hitung kekompakan objek dalam satu klaster.

$$W_k = \sum_{k}^{K} 1 \, \frac{1}{2n_k} \, D_k \quad (3)$$

Rumus (3) ini digunakan untuk menghitung total within-cluster dispersion, yaitu ukuran kekompakan data dalam semua klaster. Nilai Wk merupakan jumlah dari semua total jarak dalam masing-masing klaster Dk, yang sebelumnya telah dihitung pada rumus (2). Untuk menghindari bias akibat perbedaan ukuran klaster, rumus ini menggunakan faktor normalisasi $\frac{1}{2n_k}$, di mana nk adalah jumlah data dalam klaster ke-k. Faktor ini memastikan bahwa klaster yang lebih besar tidak mendominasi total nilai kekompakan. Nilai Wk menggambarkan seberapa kompak data dalam setiap klaster; semakin kecil nilai Wk, semakin baik karena menunjukkan bahwa data dalam klaster tersebut lebih rapat dan tidak menyebar jauh dari pusatnya.

c. Hitung klaster optimal dengan persamaan:

$$Gap_n(k) = E_n^* \log(W_k) - \log(W_k) \quad (4)$$

Rumus (4) ini digunakan untuk menentukan jumlah klaster yang optimal menggunakan metode Gap Statistic. Nilai $Gapn(k)$ merupakan nilai selisih antara rata-rata log dari within-cluster dispersion referensi acak $E_n^* \log(W_k)$ dan log dari nilai within-cluster dispersion aktual data $\log(W_k)$. Nilai $E_n^* \log(W_k)$ ihitung berdasarkan data acak yang dibangkitkan dengan distribusi yang sama seperti data asli, sedangkan $\log(W_k)$ adalah hasil dari log within-cluster dispersion yang dihitung dari data aktual. Nilai $Gapn(k)$ yang paling besar (maksimum) menunjukkan bahwa struktur klaster yang terbentuk pada jumlah klaster k tersebut jauh lebih baik dan lebih signifikan dibandingkan dengan hasil dari data acak, sehingga jumlah klaster tersebut dianggap sebagai jumlah klaster yang optimal.

2.7 Evaluasi Hasil Klaster

2.7.1 Sillhouette Koefisien

Metode ini merupakan teknik evaluasi klaster yang menggabungkan dua pendekatan, yaitu cohessian dan separation. Cohessian diukur berdasarkan jumlah seluruh objek yang terdapat dalam sebuah klaster, sedangkan separation dihitung melalui jarak rata-rata antara setiap objek dalam sebuah klaster dengan klaster terdekatnya. Jarak antar data

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

diukur menggunakan rumus euclidean distance. Untuk memberikan informasi mengenai kualitas hasil proses *clustering*, dapat dihitung nilai silhouette untuk masing-masing klaster, bahkan untuk keseluruhan klaster. Nilai silhouette untuk seluruh data dengan jumlah klaster k didefinisikan sebagai sil(k), yang diperoleh dari rata-rata nilai silhouette semua klaster [10].

$$sil\ (c) = sil\ (k) \frac{1}{|k|} \sum_{i=1}^{k} sil\ (c_i) \tag{5}$$

Dengan hasil eveluasi [18]:

Tabel 2. Hasil Evaluasi Sillhouette Score

Nilai Koefisien Silhouette	Keterangan
0,71 - 1,00	Struktur klaster sangat kuat
0,51 - 0,70	Struktur klaster cukup baik
0,26 - 0,50	Struktur klaster lemah
≤ 0,25	Tidak terdapat struktur klaster

Pada Tabel (2) Silhouette Score merupakan metode evaluasi yang digunakan untuk mengukur kualitas hasil clustering. Nilai koefisien silhouette menunjukkan seberapa baik objek berada dalam cluster yang sesuai dan seberapa jauh objek tersebut dari cluster lain. Semakin mendekati angka 1, berarti cluster yang terbentuk semakin baik karena data benar-benar berada dalam cluster yang tepat. Sebaliknya, jika mendekati atau kurang dari nol, berarti pembentukan cluster tidak optimal. Adapun interpretasi nilai silhouette dapat dilihat pada Tabel 2, yang menjelaskan rentang nilai dan keterangannya, mulai dari struktur klaster sangat kuat (0,71–1,00), cukup baik (0,51–0,70), lemah (0,26–0,50), hingga tidak terdapat struktur klaster (≤ 0,25).

2.7.2 Davies-Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan metode yang digunakan untuk menentukan jumlah klaster yang optimal. Nilai optimal ditentukan berdasarkan tingkat kohesi dan separasi antar data atau objek. Kohesi menunjukkan seberapa dekat data atau objek dalam satu klaster terhadap pusat klaster (centroid), sedangkan separasi menggambarkan jarak antar pusat klaster. Klaster yang ideal ditandai dengan struktur yang padat dan terpisah jauh satu sama lain, yang tercermin dari nilai DBI yang rendah. Oleh karena itu, jumlah klaster yang paling sesuai adalah yang menghasilkan nilai DBI terkecil. Nilai DBI dihitung menggunakan persamaan tertentu yang mengukur rasio antara kohesi dan separasi [8]. Semakin mendekati nilai 0, nilai Davies-Bouldin Index (DBI) menunjukkan bahwa hasil *clustering* semakin baik. Nilai DBI yang lebih rendah menandakan bahwa hasil *clustering* yang diperoleh semakin optimal [9]

a. Sum of Square Within Klaster (SSW) digunakan untuk mengukur tingkat kohesi data dalam klaster ke-i.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d\ (x_j, c_i) \tag{6}$$

b. Sum of Square Between Klaster (SSB) adalah rumus yang digunakan untuk mengukur tingkat pemisahan antar klaster

$$SSB_{i,j} = d\ (c_i, c_j) \tag{7}$$

c. Setelah diperoleh nilai separasi dan kohesi, dilakukan pengukuran rasio Rij untuk menentukan tingkat perbandingan antara klaster ke-i dengan klaster ke-j.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{8}$$

d. Nilai Davies-Bouldin Index (DBI) dapat dihitung menggunakan persamaan dibawah ini:

$$DBI = \frac{1}{K} \sum_{i=1}^{k} max_{i \neq j} \left(R_{i,j}\right) \tag{9}$$

Rumus-rumus diatas digunakan untuk menghitung Davies-Bouldin Index (DBI) yang mengukur kualitas klaster. SSW dan SSB masing-masing menunjukkan tingkat kekompakan dalam klaster dan pemisahan antar klaster, sedangkan rasio $R_{i,j}$ membandingkan keduanya. Nilai DBI yang lebih rendah menunjukkan hasil klastering yang lebih baik karena klaster lebih kompak dan terpisah dengan jelas.

2.7.3 *Calinski-Harabasz Index*

Indeks Calinski-Harabasz (CH) [13] digunakan untuk mengevaluasi kualitas hasil *clustering* dengan membandingkan jumlah kuadrat antar klaster (SSB) yang menggambarkan pemisahan antar kelompok, dan jumlah kuadrat dalam klaster (SSW) yang mencerminkan kekompakan di dalam masing-masing klaster. Perbandingan ini dikalikan dengan faktor normalisasi, yaitu selisih antara jumlah data dan jumlah klaster, dibagi dengan jumlah klaster dikurangi satu. Semakin tinggi nilai indeks Calinski-Harabasz, maka semakin baik hasil *clustering*, karena menunjukkan klaster yang lebih terpisah secara jelas dan lebih homogen secara internal. Berikut perhitungan indeks validitas CH [19]:

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

$$CH = \frac{trace(SSB)}{trace(SSW)} \times \frac{N-k}{k-1} \quad (10)$$

$$SSW = \sum_{i=1}^{k} \sum_{x_i \in C_i} (x_i - \bar{x})(x_i - \bar{x})^T \quad (11)$$

$$SSB = \sum_{i=1}^{k} N_i (x_i - \bar{x_i})(x_i - \bar{x})^T \quad (12)$$

Rumus-rumus pada CHI menjelaskan Calinski-Harabasz Index (CH) yang digunakan untuk mengevaluasi hasil clustering. Nilai CH diperoleh dari rasio antara trace dari SSB (pemisahan antar klaster) dan SSW (kekompakan dalam klaster), dengan mempertimbangkan jumlah klaster k dan total data N. Semakin besar nilai CH, maka kualitas clustering dianggap semakin baik karena menunjukkan klaster yang lebih terpisah dan kompak.

3. HASIL DAN PEMBAHASAN

3.1 Pra-Prosecing

Data cuaca yang digunakan berasal dari pengamatan harian (Februari 2017 – Desember 2023). Data yang digunakan terdiri dari 2525 entri, masing-masing memiliki atribut seperti suhu (Tn, Tx, Tavg), kelembaban (RH_avg), curah hujan (RR), penyinaran matahari (ss), dan angin (ff_x, ff_avg). Pembersihan data dilakukan dengan menghapus data yang memiliki nilai kosong untuk memastikan tidak ada missing value yang dapat mempengaruhi analisis. Selain itu, proses deteksi dan penghapusan outlier dilakukan menggunakan metode Z-Score dengan batas ambang di atas 3, sehingga data yang memiliki nilai ekstrem dapat dieliminasi agar tidak mempengaruhi hasil clustering. Setelah dilakukan proses pembersihan dan penghapusan outlier, diperoleh data akhir sebanyak 2.413 baris data yang bersih dan siap digunakan untuk tahap analisis selanjutnya.

Tabel 3. Pre-Prosecing Data

Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
01/02/2017	25	30,8	26,2	86	4	5,3	6	315	4	NW
02/02/2017	23,8	29,2	25,7	87	24	0,8	5	315	3	NW
03/02/2017	23	30,2	26,1	85	17	2,3	5	315	3	NW
04/02/2017	23,2	30,8	27,4	82	23	4,2	7	315	4	NW
05/02/2017	24,2	31,2	27,3	84	11	5,7	5	315	4	NW
06/02/2017	25	31	27,7	82	0	5,7	8	315	5	NW
07/02/2017	23,8	31	27,9	78	37	5,2	9	270	6	W
08/02/2017	24,6	30,6	21,2	76	17	4,9	10	270	6	W
09/02/2017	23,2	28,4	25,8	89	22	0	8	315	1	NW
10/02/2017	23,2	28,4	25,8	89	22	0	8	315	4	W
11/02/2017	23	26,8	25,1	95	36	0	3	315	8	NW
12/02/2017	24	27,4	24,9	96	16	0,7	6	315	3	NW
13/02/2017	23,8	30,4	26,4	91	15	4,8	6	315	2	NW
14/02/2017	24,4	29,8	26,7	86	18	5,7	6	315	4	NW
15/02/2017	24,4	30,2	26,6	87	20,5	4,9	8	315	3	NW
../../….	…	…	…	…	…	…	…	…	…	…
30/12/2023	24,2	33,2	30	74	97,8	8,4	5	350	2	C
31/12/2023	25,8	33,4	30	78	0	7	4	340	2	N

Keterangan atribut Tabel 3: Tanggal: Menunjukkan tanggal pencatatan data cuaca harian, Tn: Temperatur minimum harian (°C), Tx: Temperatur maksimum harian (°C), Tavg: Temperatur rata-rata harian (°C), RH_avg: Kelembapan relatif rata-rata harian (%), RR: Curah hujan harian (mm), Ss: Lama penyinaran matahari dalam satuan jam, ff_x: Kecepatan angin maksimum (m/s), ddd_x: Arah angin saat kecepatan maksimum (dalam derajat), ff_avg: Kecepatan angin rata-rata harian (m/s), ddd_car: Arah angin rata-rata harian (tulisan arah angin seperti NW, W, dll).

3.2 Seleksi Fitur

Pada tahap ini dilakukan pemilihan data yang dianggap paling berpengaruh dalam proses pengelompokan. Dari seluruh data cuaca yang tersedia, hasil seleksi pada Tabel.4 Hasil Seleksi Fitur menunjukkan bahwa kelembapan rata-rata (RH_avg), curah hujan (RR), kecepatan angin maksimum (ff_x), dan kecepatan angin rata-rata (ff_avg) menjadi faktor utama yang digunakan. Atribut-atribut tersebut dipilih karena dianggap mewakili kondisi cuaca yang paling mempengaruhi perbedaan pola harian. Kelembapan dan curah hujan berperan penting dalam membedakan cuaca hujan dan tidak hujan, sementara kecepatan angin maksimum dan rata-rata digunakan karena angin dapat menunjukkan perbedaan kondisi atmosfer seperti adanya badai atau cuaca cerah. Dengan memilih atribut ini, proses pengelompokan menjadi lebih relevan karena faktor-faktor tersebut langsung berkaitan dengan perubahan cuaca yang signifikan.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

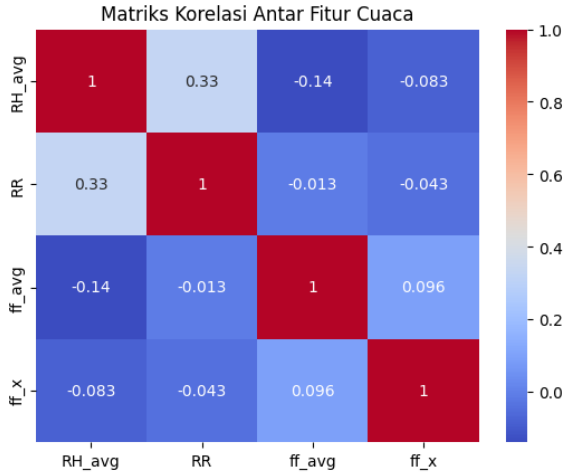
Tabel 4. Hasil Seleksi Fitur

RH_avg	RR	ff_x	ff_avg
86	4	6	4
87	24	5	3
85	17	5	3
82	23	7	4
84	11	5	4
82	0	8	5
78	37	9	6
76	17	10	6
89	22	8	1
89	22	8	4
95	36	3	8
96	16	6	3
91	15	6	2
86	18	6	4
87	20,5	8	3

3.3 Transformasi Data dan PCA

Data dinormalisasi menggunakan StandardScaler, kemudian direduksi dimensinya menjadi 2 dengan PCA untuk mempermudah proses klasterisasi dan visualisasi. Hasil dari tahapan ini yaitu:

a. PCA mengurangi dimensi dari 4 fitur menjadi 2 komponen utama
b. Variansi total yang dijelaskan oleh PCA: ± 85%



Gambar 2. Heatmap Matriks Korelasi

Gambar 2 menampilkan matriks korelasi antar fitur cuaca yang digunakan dalam penelitian ini, yaitu RH_avg (kelembaban rata-rata), RR (curah hujan), ff_avg (kecepatan angin rata-rata), dan ff_x (kecepatan angin maksimum). Nilai korelasi berkisar dari -1 hingga 1, di mana nilai mendekati 1 menunjukkan hubungan positif kuat, nilai mendekati -1 menunjukkan hubungan negatif kuat, dan nilai mendekati 0 menunjukkan tidak adanya korelasi.Dari gambar, terlihat bahwa sebagian besar korelasi antar fitur sangat lemah. Misalnya, korelasi antara RH_avg dan RR sebesar 0.33 menunjukkan hubungan positif yang lemah, artinya ketika kelembaban meningkat, curah hujan cenderung ikut meningkat, meskipun tidak selalu. Korelasi antara ff_avg dan ff_x hanya 0.096, yang menandakan bahwa kecepatan angin rata-rata dan maksimum tidak terlalu berkaitan. Korelasi antara fitur lainnya bahkan lebih rendah dan mendekati nol, seperti antara RR dan ff_avg (-0.013) atau RH_avg dan ff_avg (-0.14), yang menunjukkan bahwa fitur-fitur tersebut bekerja secara independen satu sama lain. Kesimpulannya, keempat fitur ini tidak tumpang tindih secara statistik, karena korelasinya lemah. Hal ini menunjukkan bahwa masing-masing fitur memberikan informasi yang berbeda dan saling melengkapi, sehingga sangat layak digunakan bersama dalam proses pengelompokan (clustering) cuaca.

c. Plot 2D visual menunjukkan penyebaran data yang cukup terpisah

Table 5. Hasil Transformasi Data

RH_avg	RR	ff_avg	ff_x
0,966551	-0,08501	2,183612	0,41792
1,098762	2,073654	0,997111	-0,20158

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

RH_avg	RR	ff_avg	ff_x
0,83434	1,318121	0,997111	-0,20158
0,437707	1,965721	2,183612	1,037419
0,702129	0,670522	2,183612	-0,20158
0,437707	-0,51674	3,370114	1,656918
1,363183	1,857788	-1,37589	1,656918
1,363183	1,857788	2,183612	1,656918
2,288659	1,210188	0,997111	0,41792
1,627605	1,102255	-0,18939	0,41792
0,966551	1,426055	2,183612	0,41792
1,098762	1,695888	0,997111	1,656918
1,230972	1,965721	0,997111	2,895916
0,966551	-0,51674	-0,18939	-0,82108
1,098762	-0,40881	0,997111	2,276417

Tabel 5. Hasil Transformasi Data menampilkan hasil transformasi data terhadap beberapa variabel cuaca, yaitu RH_avg (kelembaban relatif rata-rata), RR (curah hujan), ff_avg (kecepatan angin rata-rata), dan ff_x (komponen transformasi kecepatan angin). Transformasi ini dilakukan untuk menyesuaikan data dengan skala atau distribusi tertentu agar lebih siap digunakan dalam proses analisis data seperti clustering.

Table 6. Nilai Hasil PCA

PCA1	PCA2
0,776023	1,341959
-0,69695	1,969968
-0,36752	1,413501
0,64833	2,5987
0,403852	1,420003
2,162395	1,833452
-1,02146	1,843158
0,561061	3,176755
-0,67037	2,209626
-0,89675	1,428588
0,332749	2,236822
0,188833	2,52731
0,572424	3,262329
-0,66898	-0,32349
1,064571	1,541246

Berdasarkan Tabel 6. Nilai hasil PCA, komponen utama pertama (PC1) didominasi oleh fitur RH_avg dan RR, yang menunjukkan bahwa dimensi ini banyak dipengaruhi oleh faktor kelembaban dan curah hujan. Sementara itu, komponen utama kedua (PC2) didominasi oleh fitur ff_x dan ff_avg, yang mengindikasikan kontribusi besar dari faktor kecepatan angin maksimum dan rata-rata. Hasil ini menunjukkan bahwa PCA berhasil memisahkan dimensi iklim menjadi dua kelompok utama: kelembaban-curah hujan dan kecepatan angin.

3.4 Penentuan Jumlah Klaster dengan Gap Statistic

Penentuan jumlah klaster dilakukan menggunakan metode Gap Statistic untuk mengetahui berapa banyak klaster yang paling sesuai dalam mengelompokkan data. Metode ini digunakan agar hasil pengelompokan tidak terlalu sedikit atau terlalu banyak, sehingga mampu mencerminkan struktur data yang sebenarnya. Dalam proses ini, beberapa kemungkinan jumlah klaster diuji, dan hasilnya dibandingkan untuk memilih yang paling optimal, dengan hasil nilai k optimal, sebagai berikut:

Table 7. Hasil Nilai Optimal K (Gap Statistik)

Jumlah Klaster (k)	Gap Value
1	0,747073259
2	0,400317361
3	0,917266413
4	0,862788716
5	0,618000133
6	0,804453457
7	0,783295628
8	0,685447471

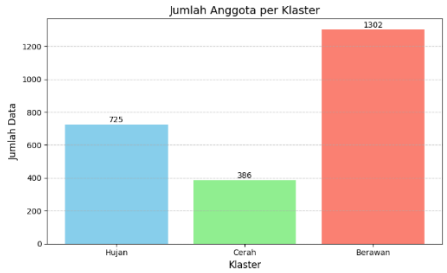
Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

Berdasarkan Tabel 7, nilai Gap Statistic tertinggi terdapat pada jumlah klaster k = 3 dengan nilai 0,9172, yang menunjukkan bahwa pembagian data menjadi tiga kelompok memberikan struktur klaster yang paling baik dibandingkan jumlah klaster lainnya. Meskipun nilai gap pada k = 4 dan k = 6 juga cukup tinggi, namun tidak melebihi nilai pada k = 3, sehingga jumlah klaster tiga dipilih sebagai yang paling optimal dalam penelitian ini karena mampu menghasilkan pemisahan data yang paling jelas dan stabil.

3.5 Proses *Clustering* dengan K-Medoids

Clustering dilakukan dengan algoritma K-Medoids berdasarkan hasil k=3. Medoid awal dipilih secara acak. Hasilnya:

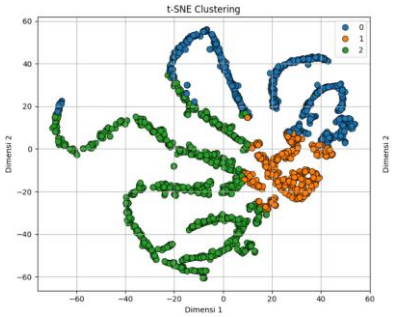
a. Distribusi Klaster:
   1. Klaster 0: 725 data
   2. Klaster 1: 386 data
   3. Klaster 2: 1302 data



Gambar 3. Distribusi Data

Grafik Gambar 3 di atas menunjukkan distribusi jumlah data pada masing-masing klaster hasil pengelompokan cuaca, yaitu hujan, cerah, dan berawan. Klaster berawan memiliki jumlah anggota terbanyak, yaitu 1.302 data, diikuti oleh klaster hujan sebanyak 725 data, dan klaster cerah sebanyak 386 data. Hal ini menunjukkan bahwa kondisi berawan merupakan pola cuaca yang paling sering terjadi dalam data yang dianalisis, sementara cuaca cerah merupakan kondisi yang paling jarang muncul dibanding dua klaster lainnya.

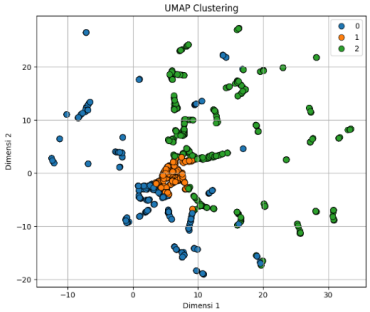
b. Visualiasi:
   Hasil divisualisasikan dengan t-SNE dan UMAP: Titik-titik dengan warna berbeda menunjukkan penyebaran klaster yang cukup baik dan tidak terlalu tumpang tindih.
   1. t-SNE *Clustering*



Gambar 4. Visualisasi t-SNE Clustering

Gambar 4 ini menunjukkan bahwa data terbagi jelas menjadi tiga klaster (label 0, 1, dan 2), dengan batas antar kelompok yang cukup terpisah dan pola sebaran yang menyerupai cabang atau struktur kompleks. t-SNE sangat baik dalam menangkap struktur lokal, sehingga cocok untuk melihat kedekatan data antar titik dalam klaster yang rapat.

   2. UMAP *Clustering*



Gambar 5. Visualisasi UMAP Clustering

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

Gambar 5 UMAP Clustering in juga menunjukkan hasil pemisahan klaster yang cukup baik. Meskipun tampilannya lebih tersebar dibanding t-SNE, UMAP mampu mempertahankan baik struktur lokal maupun global. Artinya, selain menjaga jarak antar titik dalam klaster, ia juga menggambarkan relasi antar klaster secara keseluruhan. Kedua metode memberikan gambaran visual yang membantu memahami hasil pengelompokan dari metode K-Medoids. t-SNE lebih fokus pada kedekatan lokal dan membentuk pola seperti "kelopak" atau lengan-lengan terpisah, sedangkan UMAP menunjukkan distribusi spasial klaster yang sedikit lebih longgar namun tetap mengelompok jelas. Hasil ini mendukung bahwa proses *clustering* yang dilakukan cukup berhasil memisahkan pola data secara visual.

3.6 Evaluasi

3.6.1 Evaluasi Hasil *Clustering*

Evaluasi hasil clustering dilakukan untuk menilai sejauh mana kualitas pengelompokan data yang telah dilakukan. Dalam penelitian ini, digunakan tiga metrik evaluasi yang umum digunakan, yaitu Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Masing-masing metrik memiliki fungsi yang berbeda dalam mengukur.

Tabel 8. Evaluasi Hasil *Clustering*

Metrik	Nilai	Interpretasi
Silhouette Score	0.5331	Nilai di atas 0.5 menunjukkan bahwa distribusi antar klaster cukup jelas. Artinya, data lebih dekat ke pusat klasternya dibanding ke klaster lain. Semakin mendekati 1, semakin baik pemisahan antar klaster.
Calinski-Harabasz Index	10.7583	Indeks ini mengukur rasio antara pemisahan antar klaster dan kekompakan dalam klaster. Semakin besar nilai CH, semakin baik kualitas klaster. Nilai 10.7583 menunjukkan bahwa pemisahan antar klaster cukup kuat dibanding penyebaran internal klaster.
Davies-Bouldin Index	0.8082	DBI yang lebih kecil lebih baik. Nilai mendekati 0 menandakan pemisahan antar klaster yang optimal dan kekompakan internal yang baik. Nilai 0.8082 masih tergolong bagus, meskipun jika mendekati 0.5 atau di bawahnya akan lebih ideal.

Tabel 8 menampilkan hasil evaluasi proses clustering menggunakan tiga metrik utama, yaitu Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. ecara keseluruhan, ketiga metrik ini menunjukkan bahwa hasil clustering memiliki kualitas pemisahan antar klaster yang layak dan struktur internal klaster yang cukup kompak.

3.7 Pembahasan

Dari medoid dan distribusi fitur tiap klaster:

Tabel 9. Klaster Cuaca

Klaster	Nama Cuaca	RH_avg (kelembaban)	RR(curah hujan)	ff_avg (angin rata-rata)	ff_x (angin maksimum)	Ciri utama	interpretasi
0	Hujan	Tinggi (≥ 85%)	Tinggi (≥ 20 *mm*)	Sedang - Tinggi	Sedang - Tinggi	Basah, hujan deras	Hari hujan dengan kelembaban tinggi
1	Cerah	Rendah-Sedang (≤ 82%)	Hampir 0	Rendah	Rendah	Kering, tidak hujan	Hari cerah tanpa hujan
2	Berawan	Sedang-Tinggi (82-87%)	Ringan (1-15 mm)	Sedang	Sedang	Lembab, hujan ringan	Hari berawan atau mendung ringan

Dari Tabel 10 dapat dilihat hasil pengelompokan data cuaca harian menjadi tiga klaster utama hujan, cerah, dan berawan secara langsung menjawab masalah yang diangkat dalam pendahuluan, yaitu perlunya pemahaman pola perubahan cuaca yang kompleks tanpa mengandalkan pengamatan manual. Dengan pendekatan unsupervised learning menggunakan K-Medoids dan Gap Statistic, penelitian ini menghasilkan segmentasi cuaca yang tidak hanya relevan secara statistik, tetapi juga bermakna secara klimatologis. Hal ini memungkinkan proses pengambilan keputusan berbasis data, seperti penentuan waktu tanam pertanian, perencanaan infrastruktur yang tahan cuaca ekstrem, atau sistem peringatan dini bencana berbasis pola historis cuaca.

Masing-masing klaster memiliki karakteristik yang mencerminkan kondisi nyata di Indonesia. Klaster 0 dengan kelembaban dan curah hujan tinggi diasosiasikan dengan musim hujan; sebagian besar hari pada klaster ini memiliki RR > 50 mm, selaras dengan kejadian banjir di Kalimantan Barat pada awal 2021 menurut laporan BMKG . Klaster 1 – Cerah (RR ≈ 0, kelembaban relatif rendah), klaster ini menggambarkan musim kemarau atau kondisi kering. Pola ini banyak ditemukan pada pertengahan tahun, seperti musim kemarau panjang yang terjadi di Nusa Tenggara Timur

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

(NTT), di mana curah hujan sangat minim dan kelembaban udara relatif rendah. Klaster 2 – Berawan (RR sedang, RH tinggi), klaster ini mencerminkan cuaca mendung atau hujan ringan yang biasa terjadi di masa pancaroba (peralihan musim), khususnya pada bulan April dan Oktober. Masa transisi ini sering ditandai dengan ketidakstabilan atmosfer, angin kencang, dan hujan tidak menentu [20]

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa metode K-Medoids yang dikombinasikan dengan Gap Statistic mampu mengelompokkan data cuaca harian secara efektif. Evaluasi menggunakan Silhouette Score sebesar 0,5331, Calinski-Harabasz Index sebesar 10,7583, dan Davies-Bouldin Index sebesar 0,8082 menunjukkan bahwa hasil pengelompokan memiliki struktur klaster yang cukup baik dan terpisah dengan jelas. Klaster yang terbentuk menggambarkan pola cuaca yang berbeda, yaitu cuaca hujan dengan kelembaban dan curah hujan tinggi, cuaca cerah dengan curah hujan rendah, serta cuaca berawan dengan curah hujan ringan. Hal ini menunjukkan bahwa pendekatan yang digunakan dalam penelitian ini tidak hanya unggul secara matematis, tetapi juga memberikan hasil yang bermakna secara interpretatif dalam konteks perubahan cuaca harian. Sebagai saran, penelitian selanjutnya dapat menguji stabilitas model dengan menggunakan data cuaca dari periode ekstrem seperti fenomena El Niño atau La Niña, yang memiliki pengaruh signifikan terhadap pola hujan dan suhu di Indonesia. Selain itu, disarankan untuk mengintegrasikan analisis temporal agar model tidak hanya mengenali klaster secara statis, tetapi juga dapat menangkap pergeseran pola cuaca musiman dari waktu ke waktu. Penambahan variabel atmosfer lain seperti tekanan udara atau suhu permukaan laut juga dapat meningkatkan kedalaman analisis. Penelitian lanjutan juga dapat membandingkan K-Medoids dengan algoritma lain seperti DBSCAN atau hierarchical clustering untuk mengevaluasi konsistensi dan ketepatan segmentasi pola cuaca. Hasil klaster yang diperoleh berpotensi dikembangkan menjadi bagian dari sistem peringatan dini atau pengambilan keputusan berbasis data di sektor pertanian, transportasi, dan mitigasi bencana.

REFERENCES

[1] N. T. Luchia and M. Mustakim, “Perbandingan Algoritma K-Means Dan K-Medoids Pada Pengelompokan Humidity, Temperature, Dan Voltage Di Data Center Perawang,” *Journal of Information System Research (JOSH)*, vol. 4, no. 1, pp. 184–190, Oct. 2022, doi: 10.47065/josh.v4i1.2385.

[2] F. Hardiyanti, H. S. Tambunan, and I. S. Saragih, “Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia,” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1666.

[3] H. Pohan, M. Zarlis, E. Irawan, H. Okprana, dan Y. Pranayama, “Penerapan Algoritma K-Medoids Dalam Pengelompokan Balita Stunting Di Indonesia,” JUKI: Jurnal Komputer dan Informatika, vol. 3, no. 2, pp. 97–104, Nov. 2021, doi: 10.53842/juki.v3i2.69

[4] S. Khairunnisa and M. I. Jambak, “Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2352, Oct. 2022, doi: 10.30865/mib.v6i4.4810.

[5] I. K. Khan, H. Daud, N. Zainuddin, and R. Sokkalingam, “Standardizing reference data in gap statistic for selection optimal number of cluster in K-means algorithm,” *Alexandria Engineering Journal*, vol. 118, pp. 246–260, Apr. 2025, doi: 10.1016/j.aej.2025.01.034.

[6] A. M. El-Mandouh, L. A. Abd-Elmegid, H. A. Mahmoud, and M. H. Haggag, “Optimized K-Means Clustering Model Based on Gap Statistic,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 1, pp. 183–188, 2019, doi: 10.14569/IJACSA.2019.0100124.

[7] I. K. Khan *et al.*, “Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm,” *Egyptian Informatics Journal*, vol. 27, Sep. 2024, doi: 10.1016/j.eij.2024.100504.

[8] N. Shalsadilla, S. Martha, dan H. Perdana, “Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index Dalam Pengelompokan Wilayah Kemiskinan Di Indonesia,” STATISTIKA: Journal of Theoretical Statistics and Its Applications, vol. 23, no. 1, pp. 63–72, 2023, doi: 10.29313/statistika.v23i1.1743

[9] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette Pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” Matrix: Jurnal Manajemen Teknologi dan Informatika, vol. 9, no. 3, pp. 102–109, Nov. 2019, doi: 10.31940/matrix.v9i3.1662.

[10] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat,” PENA Tek. J. Ilm. Ilmu-Ilmu Teknik, vol. 6, no. 2, pp. 48–54, Sep. 2021, doi: 10.51557/pt_jiit.v6i2.659

[11] S. Soesmono, R. Pertiwi, B. Saputri, N. Putri, and E. Widodo, “Pengelompokan Provinsi Di Indonesia Berdasarkan Tingkat Pengangguran Tahun 2023 Menggunakan K-Medoids,” Emerging Statistics And Data Science Journal, vol. 3, no. 1, pp. 498–515, Jan. 2025, doi: 10.20885/esds.vol3.iss.1.art6.

[12] Y. Diana, F. Hadi, et al., “Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang,” JOISIE Journal Of Information Systems And Informatics Engineering, vol. 7, no. 1, pp. 97–103, Jul. 2023, doi: 10.35145/joisie.v7i1.2905.

[13] M. D. Doi, A. Rusgiyono, and T. Wuryandari, “Analisis K-Medoids Dengan Validasi Indeks Pada Ipm Daerah 3t Di Indonesia,” *Jurnal Gaussian*, vol. 12, no. 2, pp. 178–188, Jul. 2023, doi: 10.14710/j.gauss.12.2.178-188.

[14] R. Afifa, M. I. Mazdadi, T. H. Saragih, F. Indriani, and M. Muliadi, “Implementasi Principal Component Analysis (PCA) Dan Gap Statistic Untuk Clustering Kanker Payudara Pada Algoritma K-Means,” Sistemasi: Jurnal Sistem Informasi, vol. 13, no. 5, pp. 1852–1864, Sep. 2024, doi: 10.32520/stmsi.v13i5.4015

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1037−1048
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7824

[15] J. Pradipta Kusuma, I. Lewenusa, and T. Handhayani, "Clustering Data Meteorologi Di Pulau Kalimantan Menggunakan Metode K-Medoids," Jurnal Eksplora Informatika, vol. 14, no. 2, pp. 129–134, 2025, doi: 10.30864/eksplora.v14i2.1131.
[16] M. P. A. Budiman and D. Winarso, "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Bulan Rawan Bencana Kabut Asap Di Kota Pekanbaru," Jurnal FASILKOM: Teknologi Informasi dan Ilmu Komputer, vol. 14, no. 1, pp. 1–8, Apr. 2024, doi: 10.25077/fasilkom.v14i1.
[17] B. Wira, A. E. Budianto, dan A. S. Wiguna, "Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang," RAINSTEK: Jurnal Terapan Sains & Teknologi, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
[18] D. Setiawan and A. Zahra, "Pengelompokan Kemiskinan di Indonesia Menggunakan Time Series Based Clustering," *Inferensi*, vol. 6, no. 1, p. 83, Mar. 2023, doi: 10.12962/j27213862.v6i1.14969.
[19] E. Luthfi and A. W. Wijayanto, "Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia," *INOVASI*, vol. 17, no. 4, pp. 761–773, Dec. 2021, doi: 10.30872/jinv.v17i4.10106.
[20] 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," *Jurnal Pendidikan Multimedia (Edsence)*, vol. 5, no. 2, pp. 91–114, Dec. 2023, doi: 10.17509/edsence.v5i2.65925.