

Clustering Data Meteorologi di Pulau Kalimantan Menggunakan Metode K-Medoids

Jordi Pradipta Kusuma¹⁾ Teny Handhayani, PhD.²⁾ Irvan Lewenusa, M. Kom³⁾

¹⁾²⁾³⁾ Teknik Informatika, FTI, Universitas Tarumanaraga Jl. Letjen S

Parman no 1, Jakarta 11440 Indonesia

¹⁾email : jordi.535190052@stu.untar.ac.id ²⁾email : tenyh@fti.untar.ac.id ³⁾email : irvanl@fti.untar.ac.id

ABSTRACT

Meteorologi adalah cabang ilmu atmosfer yang mencakup kimia atmosfer dan fisika atmosfer, dengan fokus utama pada prakiraan cuaca. Dari percobaan menggunakan algoritma K-Medoids dan jumlah cluster 2 didapatkan nilai koefisien silhouette sebesar 0.09480563804974917, dengan parameter yang digunakan adalah temperatur minimum, maksimum, rata-rata, kelembaban rata-rata, kecepatan angin, dan kerapatan angin rata-rata. Clustering dilakukan dengan tujuan untuk mendapatkan pemahaman bagaimana pola cuaca di Pulau Kalimantan.

Kata Kunci

Clustering, Data mining, K-Medoids, Meteorologi

1. Pendahuluan

Pulau Kalimantan merupakan pulau terbesar di Indonesia yang memiliki beragam kekayaan alam dan juga lingkungan. Salah satu aspek penting dalam kelangsungan lingkungan hidup adalah pola iklim. Pola iklim mempengaruhi banyak aspek kehidupan, termasuk pertanian Masyarakat, turun naiknya ekonomi, dan kesejahteraan Masyarakat. Meteorologi merupakan salah satu cabang dari geografi yang berfokus pada mempelajari atmosfer, khususnya bagian bawah. Data – data yang terdapat Meteorologi, termasuk data iklim harian, kelembapan, arah dan kecepatan angin, curah hujan, dan juga temperatur. Hal – hal ini tentunya dapat mempengaruhi aspek kehidupan, Sebagai contoh proyek Pembangunan yang berlangsung pada kota – kota tersebut akan terhambat bahkan berhenti jika curah hujan tinggi. Unsur – unsur pada iklim juga dapat mempengaruhi pengambilan keputusan yang dilakukan dengan bermacam – macam tujuan.

Klasterisasi atau yang biasa disebut clustering adalah pengelompokan data yang dihasilkan dari

penyatuan data – data yang lebih kecil berdasarkan adanya kemiripan satu sama lain. Clustering memiliki tujuan untuk menemukan suatu objek kelompok berdasarkan perubahan perilaku dengan karakteristik yang mendasari hal tersebut berada dalam kelompok yang sama. Badan Meteorologi Klimatologi dan Geofisika (BMKG) memiliki data historis dari iklim harian yang dapat digunakan untuk dilakukan klasterisasi untuk mengetahui perubahan perilaku yang terjadi dalam satu dekade terakhir. Data ini meliputi curah hujan, temperatur, kelembapan, kecepatan angin, arah angin, dan lamanya penyinaran matahari. Clustering data ini dilakukan untuk mengetahui adanya kenaikan atau penurunan yang terjadi dalam satu dekade terakhir.

Dalam menghadapi perubahan pola iklim yang semakin kompleks, diperlukan pendekatan analisis yang tepat untuk mengelompokkan kota-kota di Pulau Kalimantan berdasarkan kesamaan pola iklim. Salah satu metode yang dapat digunakan adalah metode K-Medoids Clustering. Metode K-Medoids Clustering adalah teknik Clustering yang bertujuan untuk mengelompokkan data berdasarkan kesamaan atribut tertentu, dalam hal ini adalah pola iklim. Dengan menggunakan metode ini, akan lebih mudah untuk mengidentifikasi kelompok kota dengan pola iklim yang serupa dan memahami karakteristik iklim di kota - kota tersebut.

K-Medoids merupakan suatu algoritma yang digunakan untuk menemukan medoids didalam sebuah kelompok (cluster) yang merupakan titik pusat dari suatu kelompok (cluster). Algoritma K-Medoids lebih baik dibandingkan dengan K-Means karena pada K-Medoids kita menemukan k sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada K-Means menggunakan jumlah jarak euclidean distances untuk objek data [1].

Latar belakang tersebut merupakan alasan dilakukannya penelitian ini. Penelitian ini bertujuan untuk melakukan klasterisasi pola iklim di kota-kota di Pulau Kalimantan menggunakan

metode K-Medoids Clustering. Dengan pendekatan ini, diharapkan akan diperoleh pemahaman yang lebih mendalam tentang pola iklim di kota-kota tersebut dan hubungannya dengan faktor-faktor lingkungan dan geografis yang mempengaruhi pola iklim di wilayah tersebut.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining yang biasa disebut sebagai knowledge discovery in database (KDD) merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam himpunan data yang berukuran besar.

Teknik data mining secara garis besar dibagi dalam dua kelompok, yaitu verifikasi dan discovery. Metode verifikasi meliputi teknik-teknik statistik seperti goodness of fit, dan analisis variansi. Sedangkan, metode discovery dibagi lagi menjadi model prediktif dan model deskriptif. Model prediktif melakukan prediksi terhadap data dengan menggunakan hasil yang telah diketahui dari data yang berbeda. Model prediktif dapat dibuat berdasarkan penggunaan data historis lain. Sementara itu, model deskriptif bertujuan mengidentifikasi pola-pola atau hubungan antar data dan memberikan cara untuk mengeksplorasi karakteristik data yang diselidiki [2].

2.2 Clustering

Clustering pada suatu data adalah suatu tahapan untuk menggolongkan himpunan data yang atribut kelasnya belum dideskripsikan, secara konsep clustering adalah untuk memaksimalkan dan meminimalkan kemiripan intra antar kelas. sebagai contoh, ada suatu himpunan obyek, proses pertama dapat di klasterisasi menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompok klasifikasi tertentu. Cluster juga dapat diartikan sebagai kelompok. Maka analisa clustering pada dasarnya akan menghasilkan sejumlah cluster (kelompok) [3].

Clustering, sebagai pembelajaran unsupervised learning, ditujukan untuk menemukan pengelompokan alami dari sekumpulan pola, titik, atau objek. Dalam algoritma cluster, masalah yang signifikan adalah tidak adanya pendekatan deterministik yang mendasari pengguna untuk memutuskan metode pengelompokan mana yang paling cocok dengan sekumpulan data masukan

yang diberikan [4]. Selain itu, tantangan dalam clustering juga mencakup menentukan jumlah klaster yang optimal serta memahami kompleksitas variasi dalam data yang dimiliki.

2.3 Meteorologi

Meteorologi atau ilmu cuaca adalah cabang dari ilmu atmosfer yang mencakup kimia atmosfer dan fisika atmosfer, dengan fokus utama berada pada ilmu prakiraan cuaca. Studi dibidang ini telah dilakukan selama ribuan tahun meski kemajuan yang signifikan baru terjadi pada abad ke-18 [5]. Data yang terkandung terbagi menjadi beberapa unsur seperti suhu, angin, kelembapan, curah hujan, dan lain – lain.

Meteorologi merupakan bidang ilmu yang berfokus pada analisis serta pemahaman tentang fenomena cuaca yang terjadi di dalam lapisan atmosfer bumi. Disamping itu, meteorologi juga mendalami permasalahan yang timbul dalam konteks atmosfer.

2.4 K-Medoids

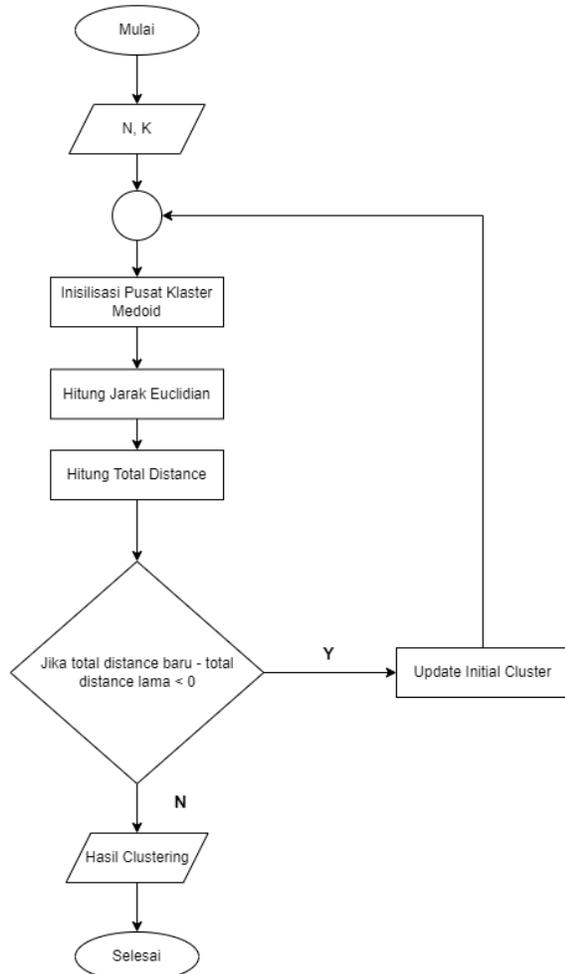
K-Medoids atau algoritma Partition Around Medoids (PAM) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma PAM termasuk metode Partitioning clustering untuk mengelompokkan sekumpulan objek menjadi cluster. Medoid adalah representasi cluster pada PAM dari sekumpulan objek yang mewakili cluster [6].

Dalam pengertian ini, medoid adalah objek dengan jarak rata-rata terkecil dari semua objek lain dalam suatu cluster. Dengan kata lain, medoid merupakan titik cluster yang paling representatif dalam hal jarak antar objek. Proses clustering data menggunakan algoritma PAM berfokus pada pengambilan sampel medoid secara optimal untuk membentuk cluster dengan jarak internal minimum antar objek di dalamnya.

Algoritma k-Means memiliki kelemahan utama yaitu sensitif terhadap outlier karena objek dengan nilai yang sangat besar dapat mengubah distribusi data. Alih-alih mengambil nilai rata-rata objek dalam sebuah klaster sebagai titik acuan, medoid dapat digunakan, yang merupakan objek yang paling terletak di pusat klaster. Dengan demikian, metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah perbedaan antara setiap objek dan titik acuannya yang sesuai [7].

Pada algoritma K-Medoids, salah satu anggota cluster digunakan sebagai perwakilan yang disebut medoid. Proses pengelompokan K-Medoids dengan cara mengorganisasikan item data ke

dalam cluster secara iteratif dan memberi nama medoid untuk setiap cluster hingga cluster medoid bertemu. Adapun alur dalam melakukan proses clustering data menggunakan algoritma K-Medoids dapat dilihat pada Gambar1 [8].



Gambar 1. Alur kerja metode K-Medoids.

1. Pada langkah awal algoritma K-Medoids, dilakukan inisialisasi dengan menetapkan jumlah pusat cluster K (Jumlah Klaster) yang akan dibentuk.
2. Melalui perhitungan Jarak Euclidean, data-data (objek) akan dikelompokkan ke dalam cluster yang memiliki jarak terdekat di antaranya. Rumus Jarak Euclidean dapat dilihat pada persamaan 1.

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} \quad (1)$$

Keterangan :

i dan j = indeks untuk dua titik atau vektor yang sedang dihitung jaraknya.

p = dimensi dari vektor, yang menunjukkan bahwa vektor tersebut memiliki elemen sebanyak p .

x_{ia} dan x_{ja} = komponen atau nilai dari vektor i dan j pada dimensi ke- a .

Pemilihan objek secara acak dari setiap cluster digunakan untuk menentukan medoid baru yang akan mewakili cluster tersebut.

3. Menggunakan medoid baru sebagai wakil, untuk menghitung jarak dari setiap objek di dalam setiap anggota cluster.
4. Simpangan total (S) dihitung dengan menjumlahkan jarak baru berdasarkan jarak antara medoid baru dengan objek dalam cluster. Jika nilai S menjadi negatif maka langkah selanjutnya adalah mengganti objek medoid dengan salah satu objek cluster sehingga terbentuk himpunan k objek baru yang akan menjadi medoid. Hal ini bertujuan untuk meningkatkan representasi cluster dengan memilih medoid yang lebih baik.
5. Ulangi langkah 3 sampai 5 hingga tidak ada perubahan pada medoid untuk mendapatkan cluster dan grup cluster yang sesuai.

Algoritma K-Medoids secara komputasi lebih rentan dibandingkan dengan K-Means karena perhitungan medoid berdasarkan frekuensi yang telah terjadi. K-Medoids memiliki karakteristik yang berpotensi penting, yang pusatnya termasuk dalam data yang diindeks sendiri [9].

2.5 Silhouette Coefficient

Silhouette Coefficient adalah sebuah metode untuk mengukur kualitas hasil clustering. Metode ini menggunakan dua ukuran, yaitu separasi dan kohesi. Separasi mengukur seberapa jauh antar cluster, sedangkan kohesi mengukur seberapa dekat antar objek dalam satu cluster [10]. Nilai Silhouette Coefficient berkisar antara -1 sampai 1. Semakin tinggi nilainya, semakin baik kualitas hasil clustering. Validasi clustering dilakukan untuk mengukur kekuatan dan kualitas hasil clustering.

Metode ini merupakan gabungan dari metode cohesion dan separation. Tahapan perhitungan Silhouette Coefficient adalah sebagai berikut [11]:

1. Hitung jarak rata-rata dari suatu objek misalkan objek ke- i dengan semua objek lain yang berada di dalam satu cluster.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2)$$

Keterangan :

i = objek yang dihitung

j = objek lain dalam satu klaster

$d(i, j)$ = adalah jarak antara objek i dengan j

2. Hitung rata-rata jarak dari objek ke- i tersebut dengan semua objek pada cluster lainnya, kemudian ambillah nilai terkecilnya.

- 3.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (3)$$

Keterangan :

$d(i, C)$ = jarak rata-rata objek i dengan semua objek pada cluster lain C dimana $A \neq C$.

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

4. Menentukan nilai minimumnya yaitu $b(i)$ yang menunjukkan perbedaan rata-rata objek i untuk kelompok yang terdekat dengan tetangganya dapat dituliskan dengan persamaan 2.5.

$$b(i) = \min_{C \neq A} d(i, C) \quad (4)$$

5. Menghitung nilai *silhouette* dengan persamaan 5.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

6. Hasil perhitungan $s(i)$ berada pada kisaran -1 hingga 1. Nilai $s(i)$ dapat diartikan sebagai:

$s(i) \approx 1$ artinya objek i terletak di kelompok yang tepat (dalam A),

$s(i) \approx 0$ artinya objek i terletak di antara 2 kelompok (A dan B),

$s(i) \approx -1$ artinya objek i terletak di kelompok yang tidak tepat (lebih dekat ke B daripada A).

7. Menghitung *Coefficient silhouette* yang didefinisikan sebagai rata-rata $s(i)$ yaitu :

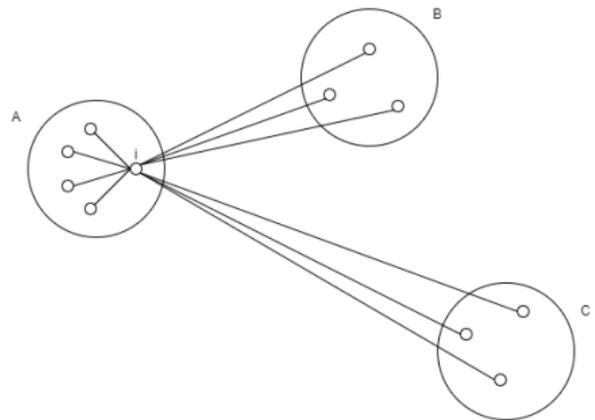
$$SC = \frac{1}{n} \sum_{i=1}^n s(i), \quad (6)$$

dengan n adalah banyak pengamatan.

Pengelompokkan optimal terjadi ketika nilai SC mencapai maksimum, yang mengindikasikan pengecilan jarak dalam kelompok ($a(i)$) sekaligus peningkatan jarak antar kelompok ($b(i)$). Besaran nilai *Coefficient silhouette* mengikuti skala yang diberikan dalam tabel berikut.

Tabel 1 Skala nilai *Coefficient Silhouette* dan Interpretasinya

| <i>Coefficient Silhouette</i> | Interpretasi |
|-------------------------------|---|
| $0.7 < SC \leq 1$ | Hubungan yang kuat (<i>strong correlation</i>) terlihat antara objek dan pembentukan kelompok. |
| $0.5 < SC \leq 0.7$ | Ada keterkaitan yang moderat (<i>moderate correlation</i>) antara objek dan pembentukan kelompok. |
| $0.25 < SC \leq 0.5$ | Terdapat hubungan yang ringan (<i>weak correlation</i>) antara objek dan pembentukan kelompok. |
| $SC \leq 0.25$ | Tidak ada keterkaitan antara objek dan pembentukan kelompok. |



Gambar 2. Ilustrasi *Silhouette* [12].

3. Hasil Percobaan

Percobaan clustering dilakukan dengan cara menerapkan metode algoritma K-Medoid kedalam data meteorologi pada 17 kota di Pulau Kalimantan. Kota – kota yang dilakukan cluster serta stasiun meteorologi nya dapat dilihat pada tabel berikut.

Tabel 2. Nama kota dan stasiun Meteorologi.

| Nama Kota | Stasiun Meteorologi |
|-------------------------|--|
| Pontianak | Stasiun Klimatologi Kalimantan Barat |
| Kab. Sambas | Stasiun Meteorologi Paloh |
| Kab. Ketapang | Stasiun Meteorologi Rahadi Oesman |
| Kab. Sintang | Stasiun Meteorologi Tebelian |
| Kab. Kapuas Hulu | Stasiun Meteorologi Pangsuma |
| Kab. Melawi | Stasiun Meteorologi Nangapinoh |
| Kab. Kubu Raya | Stasiun Meteorologi Supadio |
| Banjarmasin | Stasiun Meteorologi Syamsudin Noor |
| Banjarbaru | Stasiun Klimatologi Kalimantan Selatan |
| Palangkaraya | Stasiun Meteorologi Tjilik Riwut |
| Kab. Kotawaringin Timur | Stasiun Meteorologi H. Asan |
| Kab. Barito Selatan | Stasiun Meteorologi Sanggu |
| Balikpapan | Stasiun Meteorologi Sultan Aji |

| | |
|--------------|---|
| | Muhammad Sulaiman Sepingga |
| Kab. Berau | Stasiun Meteorologi Kalimantan |
| | Stasiun Meteorologi Aji Pangeran Tumenggung Pranoto |
| Samarinda | |
| Tarakan | Stasiun Meteorologi Juwata |
| Kab. Nunukan | Stasiun Meteorologi Nunukan |

Percobaan clustering yang dilakukan dimulai dengan tahap pengumpulan data. Data yang dikumpulkan merupakan data meteorologi dari Pulau Kalimantan. Data bersumber dari salah satu lembaga pemerintah yaitu Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), pada tahap pengumpulannya situs BMKG diakses melalui halaman resminya yaitu <https://dataonline.bmkg.go.id/home>. Karena keterbatasan dari perangkat yang digunakan dalam percobaan clustering hanya dapat menggunakan 6 parameter. Parameter tersebut meliputi temperatur minimum, temperatur maximum, temperatur rata – rata, kelembapan, kecepatan angin, kecepatan angin rata – rata. Alur kerja dari percobaan yang dilakukan dapat dilihat pada gambar berikut.



Gambar 2. Alur percobaan *clustering*

Dalam percobaan yang dilakukan metode Pra Proses untuk mengisi missing value yang digunakan adalah metode interpolasi linear. Metode interpolasi linear merupakan metode yang digunakan untuk menaksirkan nilai Tengah. Rumus dari metode ini dapat dilihat pada persamaan 7.

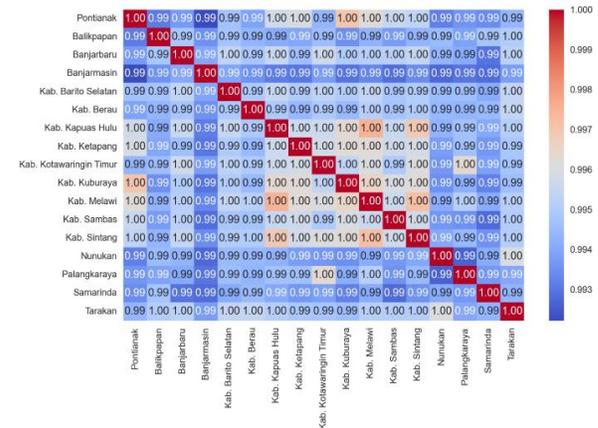
$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} (x - x_0) \quad (7)$$

Keterangan :

- $f(x)$ = data yang dicari.
- $f(x_0)$ = data pada titik awal.
- $f(x_1)$ = data pada titik akhir.
- (x_0) = nomor yang dicari.
- x = nomor yang dicari.
- x_1 = nomor titik awal.
- x_2 = nomor titik akhir.

Dalam data meteorologi, data yang digunakan memiliki format yang berbeda – beda mulai dari data temperatur, kelembapan, dan juga kecepatan angin. Oleh, karena itu data perlu dirubah menjadi satu format yang sama. Normalisasi dilakukan untuk mengubah format data yang tidak beraturan tadi menjadi format yang lebih standar dan

konsisten. Metode normalisasi yang digunakan dalam percobaan ini merupakan metode MinMaxScaller. Metode ini merupakan salah satu metode yang cocok untuk digunakan dalam clustering, Bersama dengan StandardScaler dan RobustScaler. Metode MinMaxScaller merubah format data menjadi rentang antara 0 sampai dengan 1.



Gambar 3. Korelasi *heatmap* parameter.

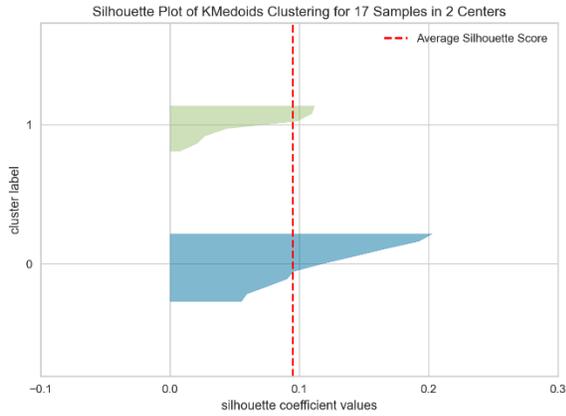
Pada Gambar 3, setelah dilakukan normalisasi data meteorologi yang digunakan memiliki korelasi yang baik satu sama lain. Hal ini dibuktikan dengan banyaknya parameter yang menyentuh angka 1 dan juga beberapa yang bernilai 0.99.

Percobaan dilakukan dengan cara mencoba tiap jumlah dari clustering mulai dari jumlah cluster 2 sampai dengan 10. Untuk menentukan jumlah cluster yang terbaik hasil silhouette coefficient perlu dilihat untuk masing – masing cluster.

```
Percobaan dengan banyaknya cluster : 2, silhouette Score: 0.09480563804974917
Percobaan dengan banyaknya cluster : 3, silhouette Score: 0.05053352776935339
Percobaan dengan banyaknya cluster : 4, silhouette Score: 0.06026879536537952
Percobaan dengan banyaknya cluster : 5, silhouette Score: 0.04731699952580342
Percobaan dengan banyaknya cluster : 6, silhouette Score: 0.046554166779054336
Percobaan dengan banyaknya cluster : 7, silhouette Score: 0.047653954491937114
Percobaan dengan banyaknya cluster : 8, silhouette Score: 0.038178123153274755
Percobaan dengan banyaknya cluster : 9, silhouette Score: 0.03754213572946353
Percobaan dengan banyaknya cluster : 10, silhouette Score: 0.04019643344851855
```

Gambar 4. Hasil percobaan *clustering*.

Dari percobaan yang dilakukan skor silhouette coefficient terbaik didapat pada percobaan dengan 2 cluster dengan skor 0.09480563804974917. Visualisasi dari hasil silhouette yang dilakukan dapat dilihat pada gambar berikut.



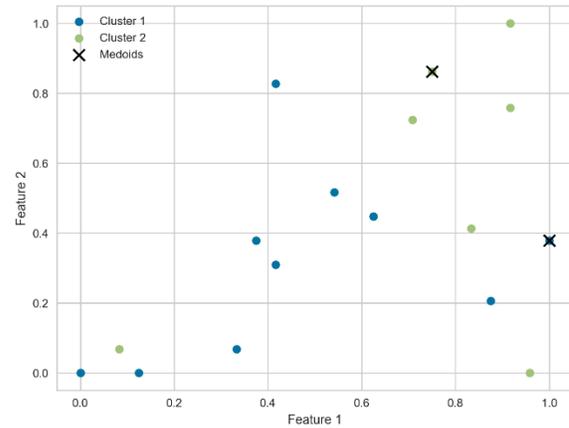
Gambar 5. Visualisasi *silhouette*.

Anggota dari masing – masing cluster dari percobaan yang dilakukan dapat dilihat pada gambar berikut.

| | ↑ Cluster |
|-------------------------|-----------|
| Pontianak | 0 |
| Banjarbaru | 0 |
| Kab. Barito Selatan | 0 |
| Kab. Berau | 0 |
| Kab. Kapuas Hulu | 0 |
| Kab. Kotawaringin Timur | 0 |
| Kab. Kuburaya | 0 |
| Kab. Melawi | 0 |
| Kab. Sambas | 0 |
| Kab. Sintang | 0 |
| Balikpapan | 1 |
| Banjarmasin | 1 |
| Kab. Ketapang | 1 |
| Nunukan | 1 |
| Palangkaraya | 1 |
| Samarinda | 1 |
| Tarakan | 1 |

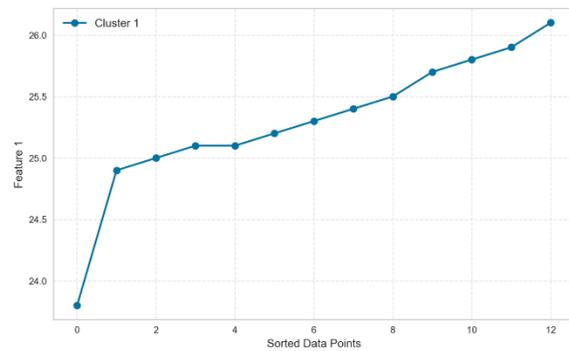
Gambar 6. Anggota *cluster* tiap kota.

Sebaran data dari percobaan clustering yang dilakukan dan juga titik mana yang dijadikan medoid dapat dilihat pada gambar berikut.



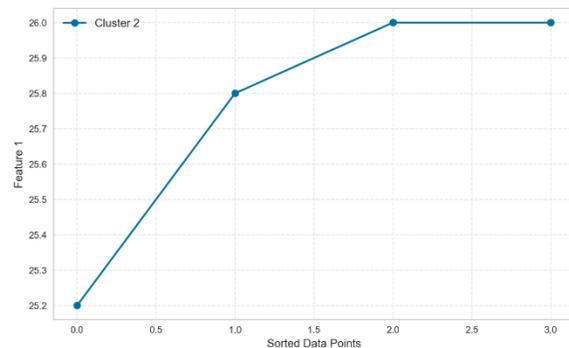
Gambar 6. Sebaran data dan medoid.

Dari percobaan terhadap semua data yang telah dilakukan, selanjutnya dilakukan percobaan cluster untuk menentukan pola dari masing – masing parameter pada tiap cluster.



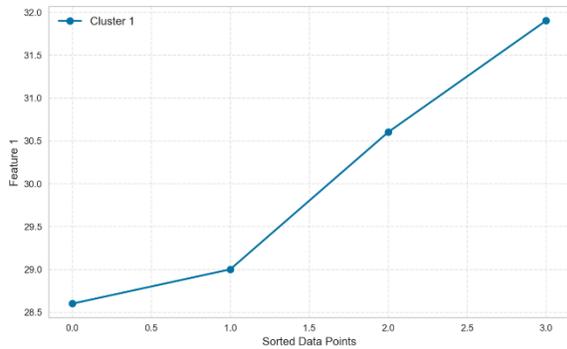
Gambar 7. Pola data temperatur minimum *cluster 1*.

Pola data dari parameter temperatur minimum cluster 2 dapat dilihat pada gambar berikut.



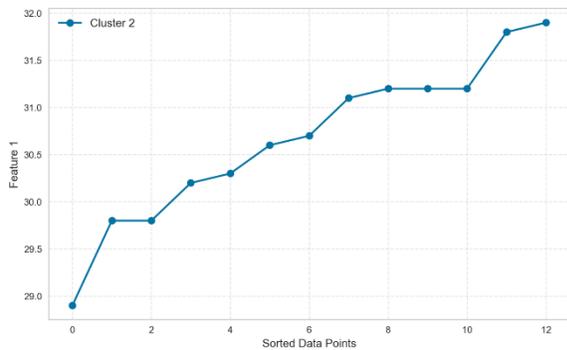
Gambar 8. Pola data temperatur minimum *cluster 2*.

Parameter selanjutnya yang dilakukan percobaan adalah temperatur maximum. Pola data dari parameter temperature maximum cluster 1 dapat dilihat pada gambar berikut.



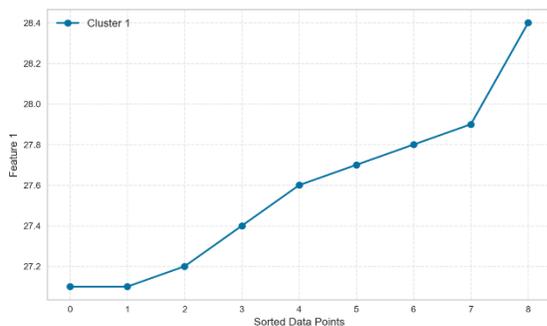
Gambar 9. Pola data temperatur maximum cluster 1.

Pola data dari parameter temperatur maximum cluster 2 dapat dilihat pada gambar berikut.



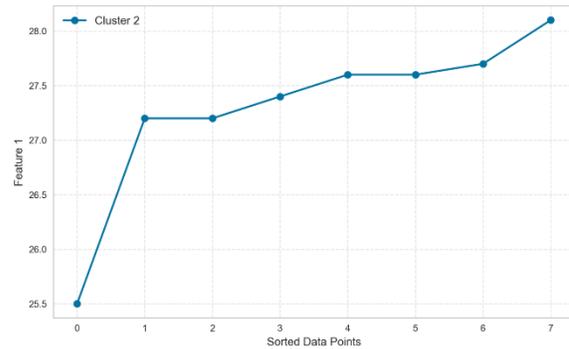
Gambar 10. Pola data temperatur maximum cluster 2.

Parameter selanjutnya yang dilakukan percobaan adalah temperatur rata - rata. Pola data dari parameter temperature rata - rata cluster 1 dapat dilihat pada gambar berikut.



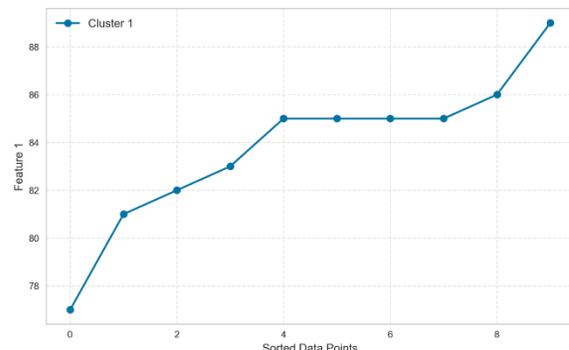
Gambar 11. Pola data temperatur rata- rata cluster 1.

Pola data dari parameter temperatur rata- rata cluster 2 dapat dilihat pada gambar berikut.



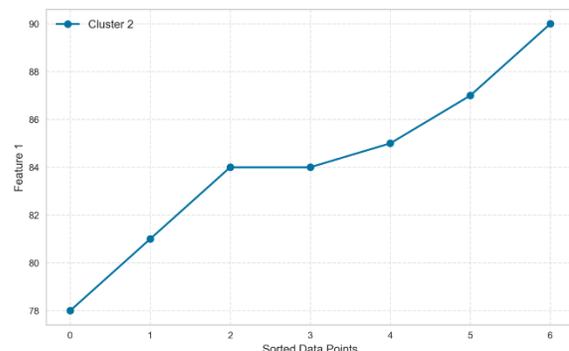
Gambar 12. Pola data temperatur rata- rata cluster 2.

Parameter selanjutnya yang dilakukan percobaan adalah temperatur rata - rata. Pola data dari parameter kelembapan rata - rata cluster 1 dapat dilihat pada gambar berikut.



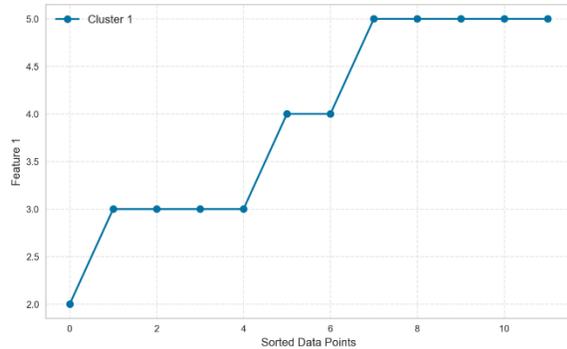
Gambar 13. Pola data kelembapan rata- rata cluster 1.

Pola data dari parameter temperatur rata- rata cluster 2 dapat dilihat pada gambar berikut.



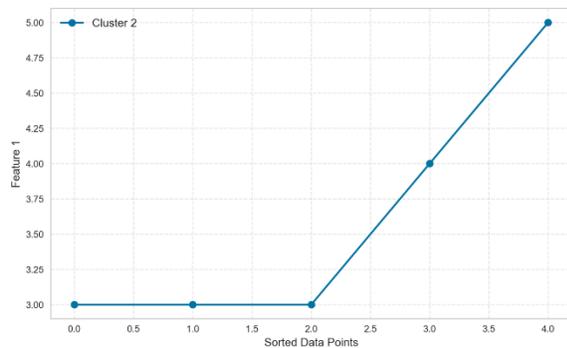
Gambar 14. Pola data kelembapan rata- rata cluster 2.

Parameter selanjutnya yang dilakukan percobaan adalah kecepatan angin. Pola data dari parameter kecepatan angin cluster 1 dapat dilihat pada gambar berikut.



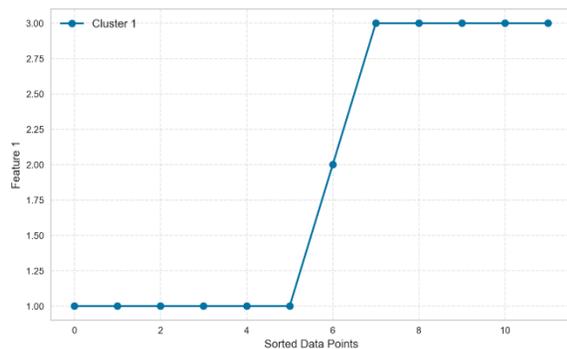
Gambar 15. Pola data kecepatan angin cluster 1.

Pola data dari parameter kecepatan angin cluster 2 dapat dilihat pada gambar berikut.



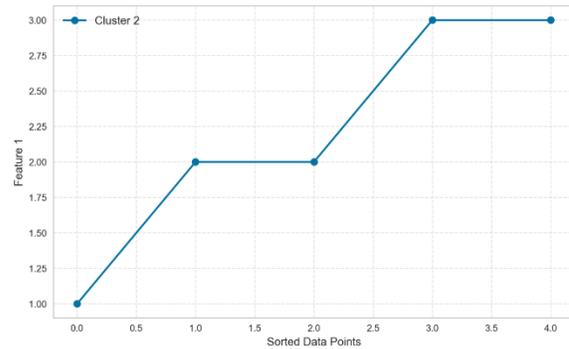
Gambar 16. Pola data kecepatan angin cluster 2.

Parameter selanjutnya yang dilakukan percobaan adalah kecepatan angin rata - rata. Pola data dari parameter kecepatan angin rata - rata cluster 1 dapat dilihat pada gambar berikut.



Gambar 17. Pola data kecepatan angin rata – rata cluster 1.

Pola data dari parameter kecepatan angin rata – rata cluster 2 dapat dilihat pada gambar berikut.



Gambar 18. Pola data kecepatan angin rata – rata cluster 2.

4. Kesimpulan

Dari percobaan clustering yang dilakukan pada 17 kota di Pulau Kalimantan didapatkan skor silhouette coefficient sebesar 0.09480563804974917. Hasil pengujian yang didapat jika diartikan dalam interpretasi skor silhouette yang terdapat pada tabel 1, skor silhouette yang didapat masuk kedalam kategori tidak ada keterkaitan antara objek dan kelompok. Hal ini disebabkan oleh algoritma yang digunakan yaitu K-Medoid kurang cocok untuk melakukan clustering pada data meteorologi. Data meteorologi mempunyai outlier yang cukup banyak, karena cuaca bisa berubah sewaktu – waktu secara tiba – tiba.

REFERENSI

- [1] Sukma Sindi, Weni Ratnasari Orktapia Ningse, Irma Agustika Sihombing, Fikrul Ilmi R. H. Zer, Dedy Hartama. 2020, "ANALISIS ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA," *Jurnal Teknologi Informasi*, pp. 166-173.
- [2] Anggoro Eko Wicaksono. 2016, "IMPLEMENTASI DATA MINING DALAM PENGELOMPOKAN DATA PESERTA DIDIK DI SEKOLAH UNTUK MEMPREDIKSI CALON PENERIMA BEASISWA DENGAN MENGGUNAKAN ALGORITMA K-

- MEANS (STUDI KASUS SMAN 16 BEKASI)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 21, no. 3.
- [3] Muliono, Rizki Sembiring, Zulfikar. 2019, "DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT TRIDARMA PENGAJARAN DOSEN," *Journal of Computer Engineering, System and Science*, vol. 4, no. 2.
- [4] Keyvan Golalipour, Ebrahim Akbari, Seyed Saeed Hamidi, Malrey Lee, Rasul Enayatifar. 2021, "From clustering to clustering ensemble selection: A review," *Engineering Applications of Artificial Intelligence*, vol. 104.
- [5] Muhammadin Hamid, S.Si, M.Si dan Dr. Indri Dayana, N.Si. 2022, *Meteorologi, Jawa Barat: Guepedia*.
- [6] Evi Rahmah, Elin Haerani, Alwis Nazir, Siti Ramadhani. 2022, "Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang)," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 3.
- [7] Noor Kamal Kaur, Usvir Kaur, Dr.Dheerendra Singh. 2014, "K-Medoid Clustering Algorithm- A Review," *International Journal of Computer Application and Technology*, vol. 1, no. 1.
- [8] Muh Arifandi, Arief Hermawan, Arief Hermawan, Donny Avianto. 2021, "IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK CLUSTERING WILAYAH TERINFEKSI KASUS COVID-19 DI DKI JAKARTA," *Jurnal Teknologi Terapan*, vol. 7, no. 2.
- [9] Norazam Arbin, Nur Suhailayani Suhaimi, Nurul Zafirah Mokhtar, Zalinda Othman. 2015, "Comparative Analysis between K-Means and K-Medoids for Statistical Clustering," *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, pp. 117-121.
- [10] Isna Rifa and Hasih Pratiwi and Respatiwulan Respatiwulan. 2020, "CLUSTERING OF EARTHQUAKE RISK IN INDONESIA USING K-MEDOIDS AND K-MEANS ALGORITHMS," *MEDIA STATISTIKA*, vol. 13, no. 2.
- [11] Mario Anggara, Herry Sujaini, Helfi Nasution. 2016, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness," *Jurnal Sistem dan Teknologi Informasi*, vol. 4, no. 1.
- [12] Peter J. Rousseeuw. 1987, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65.

Jordi Pradipta Kusuma, mahasiswa S1, program studi Teknik Informatika Universitas Tarumanagara.