

# Mendeteksi Curah Hujan Ekstrem sebagai Upaya Mitigasi Banjir Menggunakan Algoritma K-Means

Zaehol Fatah<sup>1</sup>, Rendi Aldiansyah<sup>2</sup>

<sup>1,2,3</sup> Teknologi Informasi, Universitas Ibrahimy, Situbondo

Email: <sup>1</sup>zaeholfata@gmail.com, <sup>\*2</sup>Rendy Aldiansyah@gmail.com

## Abstrak

Pengelompokan informasi cuaca berdasarkan variasi suhu dan curah hujan melalui pendekatan K-Means Clustering merupakan suatu langkah krusial untuk memahami perubahan iklim di daerah Bogor. Dalam konteks perubahan iklim dan peningkatan risiko bencana hidro-meteorologi, identifikasi pola iklim menjadi sangat diperlukan untuk mendukung perencanaan mitigasi dan adaptasi. Dataset yang digunakan dalam penelitian ini mencakup atribut-atribut seperti suhu maksimum, suhu minimum, curah hujan harian, jumlah hari hujan, serta parameter temporal seperti bulan dan tahun pengamatan. Data tersebut melalui proses pembersihan, normalisasi, dan transformasi agar siap dianalisis menggunakan algoritma K-Means. Hasil pengelompokan menunjukkan adanya dua kelompok utama kondisi iklim dengan karakteristik yang berbeda. Kelompok pertama menggambarkan periode dengan curah hujan tinggi dan suhu relatif rendah, sedangkan kelompok kedua menunjukkan pola curah hujan lebih rendah disertai suhu udara yang lebih hangat. Temuan ini memberikan wawasan penting mengenai variasi iklim lokal, serta dapat dimanfaatkan oleh pihak terkait untuk menyusun strategi mitigasi bencana, pengelolaan ruang terbuka hijau. Dengan adanya pengelompokan ini, diharapkan pemerintah daerah maupun lembaga terkait dapat memahami struktur pola iklim secara lebih komprehensif dan meningkatkan kesiapsiagaan terhadap potensi kejadian iklim ekstrem.

**Kata Kunci:** Algoritma K-Means Clustering, RapidMiner, Curah Hujan, Pembelajaran Mesin.

## Abstract

*Clustering climate data based on air temperature and rainfall variation patterns using the K-Means Clustering method is a crucial step in understanding climate dynamics in the Bogor region. Regarding climate change and the increasing risk of hydro-meteorological disasters, identifying climate patterns is crucial to support mitigation and adaptation planning. The dataset used in this study includes attributes such as maximum temperature, minimum temperature, daily rainfall, number of rainy days, and temporal parameters such as month and year of observation. The data underwent a cleaning, normalization, and transformation process to prepare it for analysis using the K-Means algorithm. The clustering results indicate two main clusters of climate conditions with distinct characteristics. The first cluster describes periods with high rainfall and relatively low temperatures, while the second cluster shows a pattern of lower rainfall accompanied by warmer air temperatures. These findings provide important insights into local climate variations and can be utilized by relevant parties to develop disaster mitigation strategies, manage green open spaces, and develop More climate-adaptive spatial planning. With this clustering, it is hoped that local governments and related institutions can understand the structure of climate patterns more comprehensively and improve preparedness for potential extreme climate events.*

*Keywords: K-Means Algorithm, RapidMiner, rainfall, Machine learning.*

## PENDAHULUAN

Indonesia sebagai negara beriklim tropis memiliki karakteristik curah hujan tinggi dan pola iklim yang sangat

dipengaruhi oleh dinamika monsun. Salah satu wilayah yang sangat dikenal dengan tingginya curah hujan adalah Kota Bogor. Curah hujan tahunan di Bogor mencapai

4.000–4.500 mm, menjadikannya daerah yang rawan terhadap peristiwa iklim ekstrem seperti banjir dan longsor. Selain itu, perubahan penggunaan lahan yang cukup cepat, terutama penurunan ruang terbuka hijau, berpotensi mengubah kondisi mikroklimat wilayah tersebut, termasuk pola suhu dan curah hujan (Hidayat & Fariyah, 2020).

Melihat tingginya kerentanan Bogor terhadap perubahan iklim serta adanya perbedaan karakteristik iklim antarstasiun pengamatan, diperlukan analisis yang komprehensif untuk mengidentifikasi perubahan intensitas curah hujan dan suhu udara dalam kurun waktu yang panjang. Penggunaan indeks iklim ekstrem melalui perangkat lunak seperti RClimdex menjadi pendekatan yang efektif untuk menilai variasi serta tren perubahan suhu maksimum, suhu minimum, dan presipitasi ekstrem.

Penelitian ini bertujuan untuk menganalisis karakteristik iklim di dua stasiun pengamatan di Bogor, yaitu Baranangsiang dan Dramaga, serta mengidentifikasi perubahan suhu udara dan curah hujan berdasarkan indeks iklim ekstrem. Hasil penelitian diharapkan memberikan gambaran mengenai dinamika perubahan iklim lokal dan kontribusi penting bagi upaya mitigasi serta perencanaan adaptasi di wilayah Bogor.

Dengan menggunakan metode K-Means, dapat diidentifikasi kelompok-kelompok data iklim yang memiliki pola perilaku yang serupa. Misalnya, wilayah atau periode waktu dengan curah hujan tinggi tetapi suhu udara relatif stabil dapat tergabung ke dalam satu kelompok tertentu, sedangkan periode dengan curah hujan rendah namun fluktuasi suhu yang besar dapat membentuk kelompok yang berbeda. Pengelompokan ini membantu peneliti memahami karakteristik iklim Bogor secara lebih jelas, sehingga pola perubahan suhu udara, tren intensitas hujan, maupun indikasi kejadian iklim ekstrem dapat terdeteksi lebih dini. Hasil analisis tersebut dapat digunakan untuk mendukung strategi

mitigasi bencana, perencanaan tata ruang, serta pengelolaan lingkungan yang lebih efektif berdasarkan pola iklim yang ditemukan.

Pendekatan ini dapat diterapkan pada berbagai metrik interaksi, seperti frekuensi posting, intensitas komentar, interaksi dengan teman, serta preferensi terhadap jenis konten tertentu (Almoqbel, M. I., & Moradi, 2022).

Penelitian ini akan menjelaskan bagaimana pengelompokan data iklim berdasarkan pola suhu udara dan curah hujan, seperti variasi harian, intensitas hujan, frekuensi hari basah, serta fluktuasi suhu maksimum dan minimum menggunakan metode algoritma K-Means. Melalui proses pengelompokan ini, wilayah atau periode waktu dengan karakteristik iklim yang serupa dapat dikelompokkan ke dalam beberapa kluster tertentu. Hasil pengelompokan tersebut memberikan wawasan yang lebih dalam mengenai kecenderungan perubahan iklim di Bogor, termasuk identifikasi pola ekstrem, tren peningkatan suhu, atau penurunan curah hujan pada kelompok-kelompok data tertentu. Informasi ini sangat berguna dalam mendukung strategi mitigasi bencana, penyusunan kebijakan pengelolaan lingkungan, serta pengambilan keputusan terkait perencanaan tata ruang berbasis pola iklim.

## **METODE**

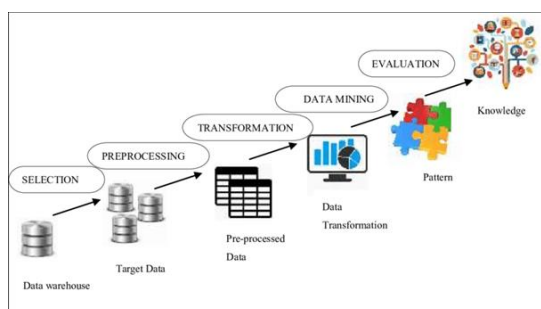
penelitian ini menggunakan penelitian deskriptif kuantitatif menggunakan rekayasa software serta Library Online Research. Pengumpulan data dilakukan dengan teknik kepustakaan. Penelitian deskriptif kuantitatif dipilih agar dapat menggambarkan, mendeskripsikan serta menjelaskan sesuatu secara objektif serta dapat menarik kesimpulan dari data berupa angka-angka yang telah disajikan, dan dengan cara ini peneliti menghimpun, mengelola data serta menganalisis data-data yang terkait dengan pola interaksi pengguna medsos dengan metode K-

Means(Zulfikar et al., 2022).

Knowledge Discovery in Database (KDD) Adalah kerangka kerja analitis digunakan untuk menggali serta mengidentifikasi pola-pola bermakna pada suatu data. Pola yang didapatkan melalui car aini bersifat valid, baru, dan dapat dimengerti sehingga memberikan informasi yang bermanfaat bagi penelitian(Arrohman & Fatah, 2024).

Prosedur KDD umumnya terdiri atas lima tahap utama, yaitu seleksi data, pra-pemrosesan, transformasi, penambahan data, dan evaluasi. Pada tahap seleksi, peneliti menentukan data yang relevan dari sumber awal untuk dijadikan data target. Selanjutnya, tahap pra-pemrosesan meliputi pembersihan data dari noise, penghapusan data redundan, perbaikan inkonsistensi, serta eliminasi atribut yang tidak relevan.

Tahap transformasi dilakukan dengan menggabungkan dan menyesuaikan format dari data supaya siap diproses ditahap berikutnya. Proses penambahan data kemudian dijalankan menggunakan algoritma yang relevan terhadap tujuan penelitian—dalam pembahasan ini K-Means untuk mengelompokkan pola kondisi iklim berdasarkan variabel suhu, kelembapan, tekanan udara, dan kecepatan angin. Tahap terakhir adalah evaluasi, yang berfungsi untuk menilai kualitas hasil pengelompokan serta memastikan bahwa pola yang ditemukan akurat dan dapat diinterpretasikan dengan baik(Yogianto et al., 2024). seperti yang ada digambar 1 ini :



Gambar 1 Metode dan Tahapan Penelitian

Dalam gambar 1, ditunjukkan proses tentang langkah-langkah dalam metode KDD. Setelah

diselaraskan dengan tahap-tahap KDD, kerangka kerja penelitian yang dijelaskan terdapat Digambar Gambar 2.



Gambar 2 tahapan penelitian

Data mining merupakan suatu rangkaian proses analitis yang melibatkan penerapan teknik statistik, pendekatan matematis, kecerdasan buatan, serta algoritma pembelajaran mesin untuk menemukan informasi penting dari kumpulan data berukuran besar. Proses ini memungkinkan peneliti menggali pola dan hubungan tersembunyi dalam database yang kompleks sehingga dapat menghasilkan pengetahuan baru yang relevan(Sukarna & Ansori, 2022)., data mining bukan hanya mengenai pengumpulan data, tetapi juga melibatkan analisis dan prediksi Informasi yang akan disajikan. Data yang telah terhimpun diolah dalam basis data untuk membantu proses pemilihan tindakan. (Oon Wira Yuda et al., 2022). merangkum langkah-langkah dalam data mining, mulai dari membersihkan data, menggabungkan data dari sumber-sumber yang berbeda, memilih data yang berkaitan, mengubah data, proses mining, menilai pola, hingga menyajikan informasi. Langkah terakhir mencakup pembuatan keputusan yang didasarkan pada hasil analisis yang telah didapat.

Algoritma K-Means adalah metode pengelompokan dimanadimulai dengan me milih sejumlah pusat kluster secara acak, kemudian mengelompokkan data berdasarkan seberapa dekat nilai-nilai tersebut dengan pusat kluster yang telah ditentukan. Algoritma ini banyak digunakan karena prosesnya yang relatif sederhana dan memiliki efisiensi tinggi dalam mengelola data berskala besar. Penelitian sebelumnya (Abdul Rohman, 2020) juga menempatkan K-Means menjadi bagian dari sepuluh algoritma data mining yang populer. Kemampuannya untuk mengelompokkan informasi ke

dalam kategori-kategori dengan sifat yang sama menjadikan algoritma ini sangat berguna dalam langkah-langkah analisis pola, termasuk dalam studi cuaca—contohnya untuk membedakan kondisi suhu, tekanan atmosfer, kelembaban, dan kecepatan angin demi menemukan pola cuaca tertentu.. Sifatnya yang mudah diterapkan serta hasil pengelompokan yang mudah diinterpretasikan menjadikan K-Means pilihan ideal dalam pemrosesan data numerik seperti data iklim(Kamila, 2021).

Clustering merupakan teknik yang efektif untuk mengelompokkan sejumlah besar data numerik ke dalam kelompok-kelompok dengan karakteristik yang serupa sehingga membentuk struktur informasi yang lebih teratur dan mudah dianalisis. Metode ini bertujuan menghasilkan pengelompokan yang koheren sehingga pola atau kecenderungan dalam data dapat dikenali dengan lebih jelas. Dalam konteks analisis iklim, Pengelompokan memudahkan untuk menemukan hubungan antara berbagai parameter cuaca, yang pada gilirannya membantu dalam menganalisis pola suhu, curah hujan, tekanan atmosfer, kelembaban, dan kecepatan angin, serta meningkatkan efektivitas dalam memahami data iklim yang rumit dan tersebar.(Widaningrum et al., 2022).

RapidMiner adalah platform analisis data dan machine learning yang menyediakan serangkaian operator untuk membangun alur kerja pemrosesan data secara terstruktur. Perangkat lunak ini mendukung seluruh tahapan dalam proses data mining, mulai dari tahap pembersihan dan persiapan data, pemilihan atribut yang relevan, hingga proses pemodelan dan analisis menggunakan berbagai algoritma(Amril Mutoi Siregar, S.Kom., M.Kom. DAN Adam Puspabhuana, S.Kom., 2021).

## HASIL DAN PEMBAHASAN

Berdasarkan hasil pengelompokan menggunakan algoritma K-Means, terbentuk tiga klaster utama yang merepresentasikan variasi kondisi iklim

berdasarkan parameter suhu udara, tekanan udara, kelembaban, dan kecepatan angin rata-rata. Klaster pertama (cluster 0) menunjukkan karakteristik suhu paling tinggi, baik pada suhu rata-rata maupun suhu maksimum, disertai nilai kelembaban yang relatif rendah.

Tekanan udara pada klaster ini cenderung lebih tinggi dibanding klaster lainnya, menandakan kondisi cuaca panas dan kering yang biasanya muncul pada periode dengan tutupan awan rendah atau pada musim kemarau. Sebaliknya, klaster kedua (cluster 1) memperlihatkan suhu udara yang lebih rendah dengan kelembaban tertinggi dan tekanan udara yang paling rendah. Pola ini mengindikasikan kondisi cuaca basah dan lembap, yang umumnya terjadi pada puncak musim hujan ketika pembentukan awan dan presipitasi meningkat.

Sementara itu, klaster ketiga (cluster 2) berada pada posisi moderat, dengan nilai suhu, kelembaban, dan tekanan udara yang tidak setinggi atau serendah dua klaster sebelumnya. Klaster ini menggambarkan kondisi transisi, seperti awal atau akhir musim hujan, ketika cuaca tidak terlalu kering namun juga tidak sepenuhnya lembap. Secara keseluruhan, hasil clustering menunjukkan bahwa metode K-Means mampu mengidentifikasi pola cuaca yang berbeda secara jelas, sehingga memberikan pemahaman yang lebih mendalam mengenai dinamika iklim lokal serta potensi perubahan yang berkaitan dengan fenomena iklim ekstrem.

## Seleksi Data (Data Selection)

### 1) Dataset

Dalam penelitian ini metode pengumpulan data yang digunakan Adalah dengan menggunakan dataset sekunder yang bersumber dari situs Kaggle (<https://www.kaggle.com/datasets/jumainzulkaidah/data-curah-hujan>). Dataset yang dimanfaatkan dalam studi ini terdiri dari 516 entri data dengan enam ciri utama yang mencerminkan keadaan atmosfer setiap hari. Atribut pertama adalah

Suhu\_rata2, yaitu nilai suhu rata-rata harian yang digunakan untuk melihat kecenderungan umum pemanasan atau pendinginan udara. Atribut kedua, Suhu\_max, merepresentasikan suhu tertinggi yang tercatat dalam satu hari dan menggambarkan intensitas radiasi serta potensi kejadian suhu ekstrem.

Selanjutnya, Suhu\_min menunjukkan suhu terendah harian yang penting untuk mengidentifikasi variasi termal dan rentang suhu harian. Atribut keempat adalah Tekanan\_udara, yang menggambarkan kondisi tekanan atmosfer dan berpengaruh terhadap pembentukan awan serta peluang terjadinya hujan. Atribut kelima, Kelembaban\_rata2, merepresentasikan tingkat kelembapan relatif harian yang berperan penting dalam proses presipitasi, kenyamanan termal, serta dinamika cuaca lokal. Terakhir, atribut KecepatanAngin\_rata2 menunjukkan rata-rata kecepatan angin harian yang berfungsi sebagai indikator pergerakan massa udara dan distribusi panas permukaan.

Secara umum, enam elemen ini saling mendukung dalam menggambarkan sifat iklim di suatu area dan rat penting untuk dianalisis menggunakan teknik K-Means Clustering. Dengan format data ini, proses pengelompokan dapat dihasilkan dengan baik untuk menemukan pola cuaca dan perubahan iklim yang bergantung pada perbedaan suhu, kelembapan, tekanan udara, dan laju angin. seperti ditampilkan pada gambar 3 dibawah ini:

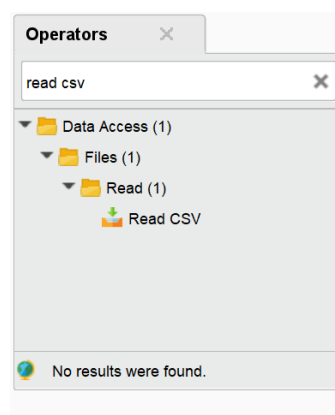
Row No.	Suhu_rata2	Suhu_max	Suhu_min	Tekanan_ud...	Kelembaba...	Kecepatan...
1	27.300	33.600	23.300	1011.400	71	4.875
2	27.200	32.800	23.300	1010.800	72	4.083
3	27.200	33	23.500	1010.900	75	5.542
4	28.300	33	24.700	1008.100	73	3.792
5	27.900	32.700	24.400	1010.100	74	3.625
6	28.200	33.200	24.800	1011	71	5.333
7	27.700	32.800	24.900	1010.900	69	4.167
8	28	32.500	25.300	1010.300	74	5.042
9	29.100	33.800	25.600	1011.200	72	5.708
10	28.600	33.600	25.200	1012.600	79	6.417
11	28.700	33	25	1013	77	5.167
12	28.800	33.600	25.400	1013.100	78	6.167
13	28.700	34	25.600	1012.500	75	4.667
14	28.800	34.400	25.600	1011.300	78	4.083
15	29.200	34.200	25	1011	76	4.208

Gambar 3 Example Set

### Pra-pemrosesan Data (Preprocessing)

Pengelompokan data pada penelitian ini, akan dikerjakan dengan menggunakan software rapidminer sebagai berikut :

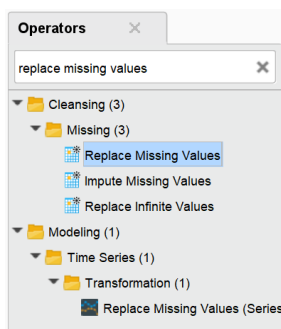
- 1) Jalankan aplikasi Rapidminer
- 2) Kemudian cari menu file lalu pilih new proses.
- 3) Selanjutnya pilih panel operator lalu ketik Read CSV, kemudian drag operator read CSV supaya operator muncul pada panel proses kemudian dapat digunakan dalam memasukkan data excel yang sudah melewati proses data cleansing dan data transformation. Sebagaimana yang ada didalam gambar 4 berikut:



Gambar 4 Read CSV

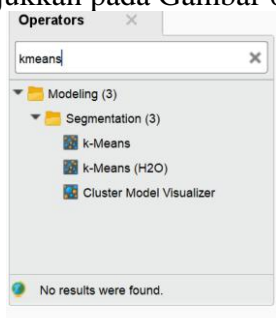
### Pra-pemrosesan Data (Preprocessing)

- 1) Langkah selanjutnya adalah melakukan Replace Missing Values agar data dapat diproses dalam analisis K-Means. Pada tahap ini, sistem akan memeriksa setiap atribut untuk memastikan tidak ada nilai yang kosong (missing). Nilai kosong sangat berpengaruh terhadap proses K-Means karena algoritma ini membutuhkan data numerik yang lengkap untuk menghitung jarak antar objek. Jika terdapat nilai yang hilang, proses perhitungan centroid menjadi tidak akurat dan dapat menghasilkan pengelompokan yang keliru. seperti diperlihatkan digambar 5 ini :



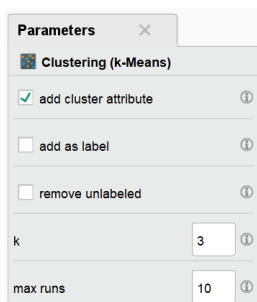
Gambar 5 Nominal To Numerical Proses *Data Mining* dengan K-Means Clustering

1) Tahap berikutnya adalah menambahkan algoritma yang akan digunakan dalam proses clustering. Pada langkah ini, peneliti membuka menu *operator* di RapidMiner dan memilih algoritma K-Means sebagai metode pengelompokan, sebagaimana ditunjukkan pada Gambar 6:



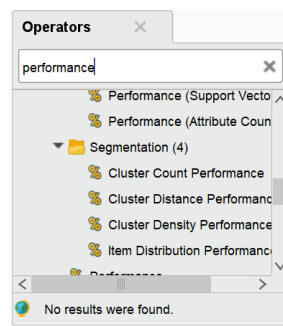
Gambar 6 Pemilihan model Klustering

2) Pada tahap ini, sebelum menjalankan proses clustering, jumlah kluster ditetapkan terlebih dahulu menjadi  $K = 3$ . Selain itu, metode pengukuran jarak diatur menggunakan mixed Euclidean distance untuk memastikan perhitungan kedekatan data berjalan sesuai kebutuhan, sebagaimana terlihat pada Gambar 7:



Gambar 7 Memilih jumlah k

3) Tahap berikutnya adalah menambahkan operator yang berfungsi untuk menilai kualitas hasil clustering. Untuk itu, peneliti membuka menu operator pada RapidMiner dan memilih cluster distance performance sebagai alat evaluasi, sebagaimana diperlihatkan pada Gambar 8:



Gambar 8 Memilih Pemodelan Cluster Distance Performance

4) Pada tahap ini, setiap operator yang telah dimasukkan perlu dihubungkan melalui konektor pada main process agar alur kerja berjalan dengan benar, sebagaimana ditunjukkan pada Gambar 9. Setelah seluruh proses terhubung, langkah selanjutnya adalah menekan tombol Run untuk menjalankan proses clustering.



Gambar 9 Main proses Data Mining Evaluasi Hasil Cluster

a) Cluster Model

Hasil pemrosesan data pada RapidMiner, seperti terlihat pada Gambar 10, menunjukkan bahwa dari total 516 data yang dianalisis, algoritma K-Means berhasil membentuk tiga kluster. Cluster 0 (kluster pertama) memuat 153 data, cluster 1 (kluster kedua) terdiri atas 166 data, sedangkan cluster 2 (kluster ketiga) mencakup 197 data.

Pembagian ini menggambarkan distribusi pola kondisi iklim yang berbeda berdasarkan parameter yang digunakan dalam penelitian.

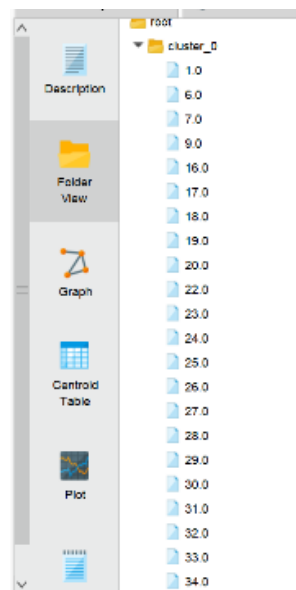
### Cluster Model

```
Cluster 0: 153 items
Cluster 1: 166 items
Cluster 2: 197 items
Total number of items: 516
```

Gambar 10 Cluster Model Hasil Data Mining

b) Folder View

Terdapat informasi dari pengamatan mengenai parameter cuaca yang dipakai sebagai landasan dalam proses pengelompokan, meliputi rata-rata suhu, suhu tertinggi, suhu terendah, tekanan atmosfer, tingkat kelembapan, dan laju angin.. Data ini merepresentasikan kondisi atmosfer harian yang kemudian dianalisis untuk mengidentifikasi kelompok-kelompok pola cuaca yang memiliki karakteristik serupa. Melalui variabel-variabel tersebut, penelitian ini dapat mengelompokkan periode dengan kondisi panas-kering, basah-lembap, maupun kondisi transisi. Informasi clustering ini memungkinkan peneliti memahami dinamika iklim secara lebih komprehensif serta memberikan dasar ilmiah bagi analisis perubahan cuaca dan potensi kejadian iklim ekstrem :



Gambar 11 Folder View

c) Centroid table Hasil Clustering

Gambar dibawah ini menampilkan hasil pengelompokan data menggunakan algoritma K-Means yang menghasilkan tiga kluster utama, masing-masing direpresentasikan melalui nilai centroid dari variabel suhu rata-rata, suhu maksimum, suhu minimum, tekanan udara, kelembapan, dan kecepatan angin. Nilai centroid pada setiap kluster menggambarkan karakteristik dominan dari kelompok data yang terbentuk. Cluster 0 menunjukkan kondisi cuaca yang didominasi suhu tinggi, baik pada suhu rata-rata maupun suhu maksimum, disertai kelembapan yang relatif rendah dan tekanan udara yang lebih tinggi. Pola ini mencerminkan kondisi panas dan kering yang lazim terjadi pada periode kemarau atau saat tutupan awan minimal. Sebaliknya, Cluster 1 menampilkan nilai kelembapan tertinggi dan tekanan udara terendah, dengan suhu yang lebih rendah dibandingkan kluster lainnya, sehingga mengindikasikan kondisi basah dan lembap yang umumnya muncul pada puncak musim hujan. Sementara itu, Cluster 2 berada di

antara kedua kluster tersebut dengan nilai suhu, kelembapan, dan tekanan udara yang moderat, menggambarkan kondisi transisi yang terjadi pada awal atau akhir musim. Secara keseluruhan, hasil ini menunjukkan bahwa metode K-Means berhasil mengelompokkan data iklim ke dalam tiga pola utama yang berbeda, sehingga memberikan pemahaman yang lebih jelas mengenai variasi cuaca lokal serta potensi perubahannya.:

Attribute	cluster_0	cluster_1	cluster_2
Suhu_min	29.412	27.972	29.148
Suhu_max	34.840	32.840	34.137
Suhu_avg	25.847	25.209	25.823
Tekanan_min	1011.163	1009.917	1010.723
Kelembaban_min	68.838	63.012	79.689
KecapatanAngin_min	4.953	4.115	4.388

Gambar 12 Tabel Centroid

## UCAPAN TERIMA KASIH

penulis menghaturkan apresiasi yang sangat besar kepada Program Studi Teknologi Informasi Universitas Ibrahimy Situbondo atas dukungan, fasilitas, dan kesempatan penelitian yang diberikan. Ucapan terima kasih juga kami haturkan untuk dosen pembimbing yang dengan dedikasi tinggi dan ketekunan telah memberikan petunjuk, saran, serta dukungan selama proses penulisan jurnal ini. Dan kami juga mengucapkan banyak terimah kasih terhadap teman-teman yang telah memrikan tenaga dalam eksperimen dan validasi dalam menggunakan perangkat RapidMiner, serta kepada teman-teman saya yang telah rela memberikan dukungannya baik dukungan moral dan teknis bagi keberhasilan meneliti. Saya sebagai penulis sangat menginginkan dari hasil penelitian ini bisa bermanfaat dan dapat menjadi referensi bagi peneliti yang lain yang mengembangkan metode klasifikasi berbasis pembelajaran mesin di bidang ilmu computer dan data.

## DAFTAR PUSTAKA

Abdul Rohman, M. R. (2020). *No Title. Implementasi Algoritma K-Means Untuk Clustering Kepuasan*

*Mahasiswa Terhadap Pelayanan Akademik.*

Almoqbel, M. I., & Moradi, F. (2022). *No Title. User Clustering in Social Networks Using K-Means and DBSCAN.*

Amril Mutoi Siregar, S.Kom., M.Kom. DAN Adam Puspabhuana, S.Kom., M. K. (2021). *DATA MINING Pengolahan Data Menjadi Informasi dengan RapidMiner.* CV Kekata Group.

Arrohman, S., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) pada Perempuan Indian Pima.* 2, 220–226.

Hidayat, R., & Fariyah, A. W. (2020). *Identification of the changing air temperature and rainfall in Bogor.* 10(4), 616–626.

Kamila, C. (2021). *Systematic Literature Review: Penggunaan Algoritma K-Means Untuk Clustering di Indonesia dalam Bidang Pendidikan.* Intech, 2(1), 19–24. <https://doi.org/10.54895/intech.v2i1.866>

Oon Wira Yuda, Darmawan Tuti, Lim Sheikh Yee, & Susanti. (2022). *Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest.* SATIN - Sains Dan Teknologi Informasi, 8(2), 122–131. <https://doi.org/10.33372/stn.v8i2.885>

Sukarna, R. H., & Ansori, Y. (2022). *Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu.* Jurnal Ilmiah Sains Dan Teknologi, 6(1), 50–61. <https://doi.org/10.47080/saintek.v6i1.1467>

Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., & Fatmawati, D. (2022). *Algoritma Term Frequency-Inverse Document Frequency (TF-*

*IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen.* Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK), 145–149.

Yogianto, A., Homaidi, A., & Fatah, Z. (2024). *Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung.* G-Tech: Jurnal Teknologi Terapan, 8(3), 1720–1728.  
<https://doi.org/10.33379/gtech.v8i3.4495>

Zulfikar, Z., Podungge, E. S., Saleh, M. I., & ... (2022). *Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan Algoritma Neural Network.* Jurnal Elektronik Sistem ..., 5(1), 7–13.  
<http://jesik.web.id/index.php/jesik/article/view/91>