



Clustering Pola Intensitas Curah Hujan Berbasis Data Radar NOAA NEXRAD Level II Menggunakan Algoritma K-Means

Silvester Kosamah^{1*}, Farizky Aulia Lubis², M. Faris Al Rafiq³, Zahira Putri Julia Daulay⁴

¹⁻⁴Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

Korespondensi penulis: kosamahsilvester58@gmail.com

Abstract. *Accurate classification of rainfall intensity patterns is important for early warning systems, hydrometeorological risk assessment, and water resource management. Surface rain gauges have limited spatial coverage, so this study uses NOAA NEXRAD Level II radar data from the KTLX station in 2023. K-Means clustering was applied to identify rainfall intensity patterns from 30 randomly selected days, with scans stratified into four daily time intervals. Seven features were extracted from each radar sweep, including reflectivity statistics, convective and stratiform ratios, and rainfall coverage. The data were normalized and balanced before clustering. The optimal cluster count was determined through a combined evaluation of the Elbow Method, Silhouette Score, and Davies-Bouldin Index, yielding K=5 as the most representative configuration. Evaluation results demonstrated a Silhouette Score of 0.3871 and a Davies-Bouldin Index of 0.8599, indicating moderate cluster cohesion that reflects the inherent overlapping nature of rainfall intensity transitions in radar reflectivity data. The clusters represent rainfall regimes from non-precipitating conditions to intense convective events. These results support the use of K-Means for automated rainfall pattern recognition and flood forecasting applications.*

Keywords: *K-Means clustering; NEXRAD Level II; Radar Reflectivity; Rainfall Intensity; Unsupervised Learning.*

Abstrak. Klasifikasi pola intensitas curah hujan berperan dalam sistem peringatan dini, penilaian risiko hidrometeorologi, dan pengelolaan sumber daya air. Data penakar hujan permukaan memiliki keterbatasan cakupan, sehingga penelitian ini menggunakan data radar NOAA NEXRAD Level II dari stasiun KTLX tahun 2023. Metode K-Means digunakan untuk mengidentifikasi pola dari 30 hari yang dipilih secara acak, dengan scan dibagi ke dalam empat interval waktu harian. Tujuh fitur diekstraksi dari setiap sapuan radar, mencakup statistik reflektivitas, rasio konvektif dan stratiform, serta cakupan hujan. Data dinormalisasi dan diseimbangkan sebelum clustering. Penentuan jumlah cluster optimal dilakukan melalui evaluasi gabungan Metode Elbow, Silhouette Score, dan Davies-Bouldin Index, yang menghasilkan K=5 sebagai konfigurasi paling representatif. Hasil evaluasi menunjukkan Silhouette Score sebesar 0.3871 dan Davies-Bouldin Index sebesar 0.8599, yang mengindikasikan kualitas pemisahan cluster yang moderat, dimana kondisi tersebut lazim dijumpai pada data reflektivitas radar akibat sifat kontinum transisi intensitas curah hujan di atmosfer. Cluster yang terbentuk merepresentasikan rezim hujan dari kondisi tanpa presipitasi hingga kejadian konvektif intensif. Hasil ini menunjukkan K-Means dapat digunakan untuk pengenalan pola hujan otomatis dan mendukung prakiraan banjir.

Kata Kunci: Clustering K-Means, NEXRAD Level II; Reflektivitas Radar; Intensitas Curah Hujan; Pembelajaran Tanpa Supervisi.

1. LATAR BELAKANG

Curah hujan dengan intensitas tinggi merupakan salah satu pemicu utama kejadian banjir bandang, tanah longsor, dan berbagai bencana hidrometeorologi lain yang menyebabkan kerugian besar di banyak wilayah. Kemampuan mendeteksi dan mengklasifikasikan pola intensitas hujan secara cepat dan akurat menjadi syarat dasar bagi sistem peringatan dini yang dapat diandalkan (Koutsovili, Tzoraki, Theodossiou, &

Tsekouras, 2023; Muñoz, Orellana-Alvear, Bendix, Feyen, & Célleri, 2021). Tanpa klasifikasi yang tepat, proses pengambilan keputusan dalam manajemen risiko bencana dan pengelolaan sumber daya air menjadi tidak efisien.

Pendekatan konvensional dalam pengukuran curah hujan bertumpu pada jaringan stasiun penakar hujan di permukaan. Metode ini memiliki batasan yang sudah lama dikenali: distribusi stasiun yang tidak merata, cakupan spasial yang terbatas, dan ketidakmampuan menangkap variabilitas hujan dalam skala mesoskala secara real-time. Di kawasan dengan topografi kompleks atau wilayah yang luas dan sparsely monitored, keterbatasan ini semakin terasa. Akibatnya, estimasi intensitas hujan berbasis stasiun sering tidak merepresentasikan kondisi aktual di lapangan (Hu et al., 2022; J. Li et al., 2023).

Radar cuaca hadir sebagai solusi observasi yang lebih komprehensif. Sistem radar mampu merekam data reflektivitas volumetrik dalam radius ratusan kilometer dengan resolusi temporal yang sangat tinggi. NOAA NEXRAD Level II, sebagai salah satu jaringan radar cuaca paling lengkap dan terdokumentasi di dunia, menyediakan data mentah berformat biner yang memuat informasi reflektivitas, kecepatan radial, dan spektral width dari berbagai sudut elevasi (Dixon & Romatschke, 2022). Data ini menjadi sumber informasi yang kaya untuk analisis struktur presipitasi.

Salah satu tantangan fundamental dalam analisis data radar adalah pemisahan antara curah hujan konvektif dan stratiform. Keduanya memiliki karakteristik fisik yang berbeda, hujan konvektif dicirikan oleh nilai reflektivitas tinggi yang terlokalisasi, sedangkan hujan stratiform menunjukkan cakupan spasial luas dengan reflektivitas yang lebih rendah dan seragam (Schumacher & Funk, 2023; Wang, Tang, Chang, & Tang, 2021). Pemisahan ini bukan sekadar kategorisasi akademis. Ia berdampak langsung pada akurasi estimasi curah hujan kuantitatif (QPE) dan interpretasi risiko banjir di hilir (Ghada et al., 2022; Mikidadi, Huang, & Bu, 2025).

Pendekatan berbasis pembelajaran mesin mulai banyak diterapkan untuk mengatasi kompleksitas ini. Berbagai algoritma, mulai dari support vector machine hingga deep learning, telah diujicobakan untuk mengekstrak pola dari data radar secara otomatis (El Hafyani, El Himdi, & El Adlouni, 2024). Di antara sekian banyak pilihan, metode clustering tanpa supervisi menawarkan keunggulan tersendiri: ia tidak memerlukan data

berlabel sebagai acuan pelatihan, sehingga lebih fleksibel diterapkan pada dataset baru yang belum memiliki ground truth. (Rahimpour, Rahimzadegan, Nosratpour, Homayouni, & Behrangi, 2025) menunjukkan bahwa pendekatan berbasis clustering mampu meningkatkan akurasi estimasi curah hujan ekstrem secara signifikan dibandingkan metode konvensional.

Algoritma K-Means adalah salah satu metode clustering yang paling banyak digunakan karena sifatnya yang sederhana, skalabel, dan deterministik. Kemampuannya mempartisi data multidimensi ke dalam kelompok-kelompok homogen secara komputasional efisien menjadikannya pilihan yang relevan untuk analisis data radar berskala besar (Miraftabzadeh, Colombo, Longo, & Foiadelli, 2023). Sejumlah studi sebelumnya telah memanfaatkan K-Means dalam konteks hidrometeorologi, namun sebagian besar berfokus pada data permukaan atau variabel cuaca umum, bukan pada fitur reflektivitas yang diekstraksi langsung dari data radar volumetrik level mentah.

Kesenjangan itulah yang menjadi titik berangkat penelitian ini. Penelitian ini mengusulkan penerapan algoritma K-Means untuk mengklasifikasikan pola intensitas curah hujan secara langsung dari fitur-fitur yang diekstraksi dari data NOAA NEXRAD Level II stasiun KTLX. Tujuh fitur kuantitatif diturunkan dari setiap sapuan radar, meliputi nilai reflektivitas maksimum, rata-rata reflektivitas piksel hujan, standar deviasi reflektivitas, rata-rata reflektivitas seluruh piksel valid, rasio piksel konvektif, rasio piksel stratiform, dan cakupan area hujan total. Penentuan jumlah cluster optimal dilakukan menggunakan kombinasi tiga metrik evaluasi yaitu Elbow Method, Silhouette Score, dan Davies-Bouldin Index. Hasil penelitian ini diharapkan dapat memberikan kontribusi metodologis bagi pengembangan sistem klasifikasi hujan otomatis berbasis radar yang dapat diadaptasi untuk keperluan operasional hidrometeorologi.

2. KAJIAN TEORITIS

Radar Cuaca dan Data NEXRAD Level II

Radar cuaca mengukur intensitas presipitasi berdasarkan nilai reflektivitas dalam satuan dBZ, yang merepresentasikan besarnya energi gelombang elektromagnetik yang dipantulkan oleh partikel hidrometeor di atmosfer. NOAA NEXRAD Level II menyimpan rekaman mentah sapuan volumetrik radar yang mencakup reflektivitas,

kecepatan radial, dan spektral width pada berbagai sudut elevasi. Data ini tersedia secara publik melalui Amazon Web Services, sehingga aksesibel untuk keperluan penelitian hidrometeorologi. Pembacaan dan pemrosesan format biner NEXRAD Level II dalam lingkungan Python difasilitasi oleh library Py-ART yang dikembangkan khusus untuk tujuan tersebut (Helmus & Collis, 2016).

Klasifikasi Hujan Konvektif dan Stratiform

Presipitasi atmosfer dibedakan menjadi dua rezim utama. Hujan konvektif memiliki intensitas tinggi, cakupan sempit, dan durasi pendek, sedangkan hujan stratiform tersebar merata dengan intensitas lebih rendah. Ambang batas 40 dBZ secara umum digunakan sebagai pemisah antara keduanya dalam domain reflektivitas radar. Pada data single-polarization, fitur berbasis reflektivitas tetap menjadi dasar utama klasifikasi (W. Li, Chen, & Han, 2024). Kualitas pemisahan kedua rezim ini berpengaruh langsung pada akurasi estimasi curah hujan kuantitatif berbasis radar (Verdelho et al., 2024).

Algoritma K-Means Clustering

K-Means bekerja dengan meminimalkan Within-Cluster Sum of Squares (WCSS) secara iteratif hingga posisi centroid konvergen. Penggunaan strategi inisialisasi K-Means++ meningkatkan stabilitas hasil dibandingkan inisialisasi acak. (Kisi et al., 2025) menunjukkan bahwa integrasi K-Means dengan model prediktif lain menghasilkan peningkatan akurasi yang konsisten dalam pemodelan rainfall-runoff. (Wani, 2024) menegaskan bahwa K-Means tetap kompetitif dibandingkan metode clustering modern selama fitur input dinormalisasi dengan baik dan jumlah cluster ditentukan berdasarkan metrik evaluasi yang tepat.

Evaluasi Kualitas Clustering

Silhouette Score mengukur kemiripan sebuah titik data terhadap clusternya sendiri dibandingkan cluster lain, dengan rentang nilai -1 hingga 1. Davies-Bouldin Index (DBI) mengukur rasio dispersi intra-cluster terhadap jarak antar-centroid, di mana nilai lebih rendah menandakan cluster yang lebih baik. (Ros, Riad, & Guillaume, 2023) mengusulkan varian partisi DBI yang lebih sensitif terhadap struktur cluster tidak seimbang, kondisi yang lazim ditemukan pada dataset radar dengan distribusi kelas

intensitas yang tidak merata. Kombinasi Elbow Method, Silhouette Score, dan DBI memberikan evaluasi yang lebih andal dibandingkan satu metrik tunggal.

Pembelajaran Mesin untuk Analisis Data Radar

Penerapan pembelajaran mesin pada data radar cuaca berkembang pesat dalam beberapa tahun terakhir. (Cuomo & Chandrasekar, 2021) menunjukkan bahwa deep learning mampu melampaui metode ekstrapolasi konvensional dalam nowcasting radar. (Mihuleț, Burcea, Mihai, & Czibula, 2023) membuktikan bahwa pendekatan ensemble menghasilkan estimasi curah hujan yang lebih stabil. (Liu, Zuo, Tan, & Liu, 2024) menyimpulkan bahwa pemilihan algoritma harus disesuaikan dengan karakteristik data dan tujuan analisis, karena tidak ada satu metode yang secara universal unggul untuk seluruh skenario QPE berbasis radar.

3. METODE PENELITIAN

Data dan Sumber data

Data yang digunakan dalam penelitian ini merupakan data radar cuaca NEXRAD Level II yang diperoleh melalui layanan penyimpanan berbasis cloud Amazon Web Services (AWS). Akses data dilakukan menggunakan library *nexradaws* di lingkungan Python, yang memungkinkan pengambilan data radar secara langsung dari arsip publik.

Data NEXRAD Level II dipilih karena memiliki resolusi tinggi dan memuat informasi reflektivitas (*reflectivity*) yang penting untuk analisis struktur awan dan intensitas presipitasi. Parameter reflektivitas (dBZ) digunakan sebagai variabel utama dalam penelitian ini karena mampu merepresentasikan karakteristik hujan, baik konvektif maupun stratiform.

Pengambilan data dilakukan untuk beberapa periode waktu yang telah ditentukan dengan interval pengamatan setiap 30 menit. Pendekatan ini digunakan untuk memastikan variasi kondisi cuaca dapat terwakili secara baik serta mengurangi potensi bias akibat pengambilan data dalam satu waktu tertentu saja.

Tabel 1. Deskripsi Dataset Radar Cuaca NEXRAD Level II (KTLX, 2023)

KOMPONEN	DESKRIPSI
Sumber Data	Data radar cuaca Nexrad II
Lokasi Radar	KTLX (Oklahoma, USA)

Parameter Utama	Reflectivity (dBZ)
Metode Akses	Library <i>nexradaws</i>
Periode Data	1 Januari 2023 – 31 Desember 2023

Teknik Sampling Data

Teknik sampling data dalam penelitian ini dirancang untuk menghasilkan dataset yang representatif serta meminimalkan bias temporal. Proses sampling dilakukan terhadap data radar cuaca NEXRAD Level II selama periode satu tahun penuh, yaitu dari 1 Januari 2023 hingga 31 Desember 2023.

Tahap pertama adalah membentuk populasi data berupa seluruh hari dalam rentang waktu penelitian. Selanjutnya, dilakukan pemilihan sampel hari menggunakan metode *random sampling* sebanyak 30 hari dari total 365 hari. Untuk memastikan reproduisibilitas, proses ini menggunakan parameter *random seed* sebesar 42.

Pada tahap berikutnya, diterapkan pendekatan *stratified temporal sampling*, yaitu pembagian data berdasarkan kelompok waktu dalam satu hari. Waktu pengamatan dibagi menjadi empat kelompok, yaitu dini hari (00.00–05.59), pagi (06.00–11.59), siang (12.00–17.59), dan malam (18.00–23.59). Pendekatan ini bertujuan untuk menangkap variasi pola hujan yang dapat berbeda pada setiap periode waktu.

Dalam setiap kelompok waktu, dilakukan proses filtering untuk menghindari redundansi data. Data radar (scan) dipilih dengan interval minimum 30 menit antar pengamatan dalam kelompok waktu yang sama. Jika terdapat beberapa scan dalam interval yang lebih rapat, maka hanya scan pertama yang dipilih, dan scan berikutnya hanya akan diambil jika selisih waktunya telah memenuhi batas interval yang ditentukan.

Setelah proses sampling selesai, data yang terpilih kemudian diunduh secara otomatis menggunakan library *nexradaws* dan disimpan dalam direktori lokal untuk tahap pengolahan selanjutnya.

Preprocessing Data

Tahap preprocessing data dilakukan untuk mengubah data radar mentah menjadi dataset numerik yang siap digunakan dalam proses clustering. Proses ini meliputi ekstraksi fitur, pembersihan data, pelabelan, balancing data, serta normalisasi.

Pada tahap awal, data radar hasil unduhan dibaca menggunakan library *Py-ART*. Dari setiap file radar, hanya data pada *sweep* pertama (elevasi terendah) yang digunakan, karena merepresentasikan kondisi hujan paling dekat dengan permukaan. Selanjutnya, nilai reflektivitas (*reflectivity*) diekstraksi dan dilakukan pembersihan dengan menghapus nilai kosong atau *masked values*.

Untuk mengidentifikasi area hujan, diterapkan *thresholding* sebesar ≥ 15 dBZ. Nilai di bawah ambang tersebut dianggap bukan hujan dan tidak digunakan dalam analisis. Berdasarkan data yang telah difilter, dilakukan perhitungan fitur statistik yang terdiri dari nilai maksimum reflektivitas (*max dBZ*), rata-rata reflektivitas (*mean dBZ*), rasio konvektif, rasio stratiform, serta cakupan area hujan (*total coverage*).

Selanjutnya, dilakukan proses pelabelan data berdasarkan waktu pengamatan (*time group*) dan intensitas hujan (*intensity class*). Pembagian waktu dilakukan ke dalam empat kelompok, yaitu dini hari, pagi, siang, dan malam. Sementara itu, klasifikasi intensitas hujan ditentukan berdasarkan nilai maksimum reflektivitas, yaitu hujan ringan (< 20 dBZ), sedang (20–40 dBZ), dan lebat (> 40 dBZ).

Untuk mengurangi bias distribusi data, dilakukan proses *balancing* dengan metode *stratified sampling*, yaitu mengambil jumlah data yang sama pada setiap kombinasi kelompok waktu dan kelas intensitas. Dalam penelitian ini, jumlah maksimum sampel yang diambil pada setiap kelompok adalah 20 data.

Tahap terakhir adalah normalisasi fitur numerik menggunakan metode *StandardScaler*, yang bertujuan untuk menyamakan skala antar fitur sehingga tidak terjadi dominasi variabel tertentu dalam proses clustering. Fitur yang dinormalisasi meliputi *max dBZ*, *mean dBZ*, *std dBZ*, *mean dBZ keseluruhan*, rasio konvektif, rasio stratiform, dan cakupan area hujan.

Dataset hasil preprocessing kemudian disimpan dalam format CSV dan digunakan sebagai input pada tahap clustering menggunakan algoritma K-Means.

Tabel 2. Deskripsi Fitur Ekstraksi Data Radar Cuaca

Fitur	Deskripsi
max_dbz	Nilai reflektivitas maksimum
mean_dbz	Rata-rata reflektivitas area hujan
Std_dbz	Standar deviasi reflektivitas
Mean_dbz_all	Rata-rata reflektivitas seluruh area

Connective_ratio	Rasio piksel konvektif (≥ 40 dBZ)
Stratiform_ratio	Rasio piksel stratiform (20-40 dBZ)
Total_coverage	Persentase area hujan terhadap total area

Penerapan Algoritma K-Means

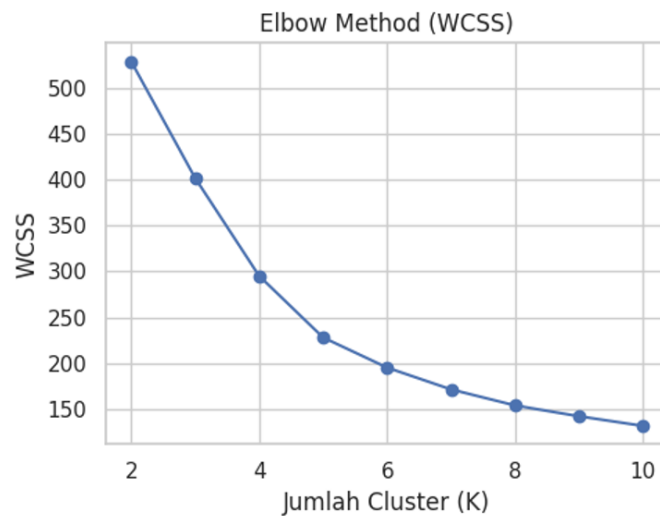
Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan pola curah hujan berdasarkan fitur numerik yang telah melalui tahap preprocessing. Sebelum proses clustering dilakukan, tahap penting yang harus ditentukan adalah jumlah cluster optimal (K), karena nilai K sangat memengaruhi hasil pengelompokan.

Penentuan jumlah cluster dilakukan menggunakan metode Elbow, yaitu dengan menganalisis nilai *Within-Cluster Sum of Squares* (WCSS) pada berbagai jumlah cluster. WCSS mengukur total variasi dalam satu cluster, di mana semakin kecil nilainya menunjukkan bahwa data dalam cluster semakin homogen. Namun, penambahan jumlah cluster secara terus-menerus akan selalu menurunkan nilai WCSS, sehingga diperlukan pendekatan visual untuk menentukan titik optimal.

Metode Elbow bekerja dengan memplot nilai WCSS terhadap jumlah cluster (K). Titik optimal ditentukan pada bagian kurva yang menunjukkan perubahan signifikan dari penurunan tajam menjadi lebih landai, yang disebut sebagai “titik siku” (*elbow point*). Titik ini menunjukkan bahwa penambahan cluster setelahnya tidak memberikan peningkatan signifikan terhadap kualitas clustering.

Dalam penelitian ini, pengujian dilakukan pada rentang $K = 2$ hingga $K = 10$. Berdasarkan hasil visualisasi grafik Elbow, diperoleh bahwa nilai K optimal berada pada $K = 5$, di mana kurva mulai menunjukkan penurunan yang lebih stabil.

Setelah jumlah cluster optimal ditentukan, algoritma K-Means diterapkan pada dataset hasil preprocessing menggunakan parameter *k-means++* untuk inisialisasi centroid, dengan maksimum iterasi sebanyak 300 dan jumlah inisialisasi ulang (*n_init*) sebanyak 10 untuk meningkatkan stabilitas hasil. Hasil clustering kemudian digunakan untuk analisis lebih lanjut pada bagian hasil dan pembahasan.



Gambar 1. Grafik Metode Elbow untuk Menentukan Jumlah Cluster Optimal (K)

4. HASIL DAN PEMBAHASAN

Hasil Preprocessing Data

Tahap preprocessing menghasilkan dataset yang telah melalui proses ekstraksi fitur, pembersihan data, serta normalisasi untuk memastikan kualitas data sebelum dilakukan clustering. Berdasarkan proses ekstraksi fitur dari data radar cuaca, diperoleh sebanyak 1247 data pengamatan (scan radar) yang siap digunakan untuk analisis lebih lanjut.

Setiap data merepresentasikan satu waktu pengamatan dengan fitur utama berupa nilai reflektivitas maksimum (*max dBZ*), rata-rata reflektivitas (*mean dBZ*), standar deviasi reflektivitas, rata-rata reflektivitas keseluruhan, rasio konvektif, rasio stratiform, serta cakupan area hujan.

Selanjutnya, dilakukan proses normalisasi terhadap seluruh fitur numerik menggunakan metode *StandardScaler*. Proses ini bertujuan untuk menyamakan skala antar fitur, sehingga tidak terjadi dominasi variabel tertentu dalam proses clustering. Sebelum normalisasi, setiap fitur memiliki rentang nilai yang berbeda, seperti nilai reflektivitas yang berada pada skala puluhan, sedangkan rasio dan cakupan berada pada rentang 0 hingga 1.

Setelah normalisasi, seluruh fitur berada dalam rentang yang sebanding dengan nilai terpusat di sekitar nol. Hal ini memungkinkan algoritma K-Means untuk bekerja

secara optimal dalam mengukur jarak antar data. Contoh data sebelum dan sesudah normalisasi ditunjukkan pada Tabel.

Tabel 3. Perbandingan Nilai Fitur Sebelum dan Setelah Normalisasi

Fitur	Sebelum Normalisasi	Setelah Normalisasi
Max_dbz	38,50	-0,46
Mean_dbz	22,66	-0,67
Std_dbz	3,41	-0,19
Convective_ratio	0	-0,51
Stratiform_ratio	1,00	0,51

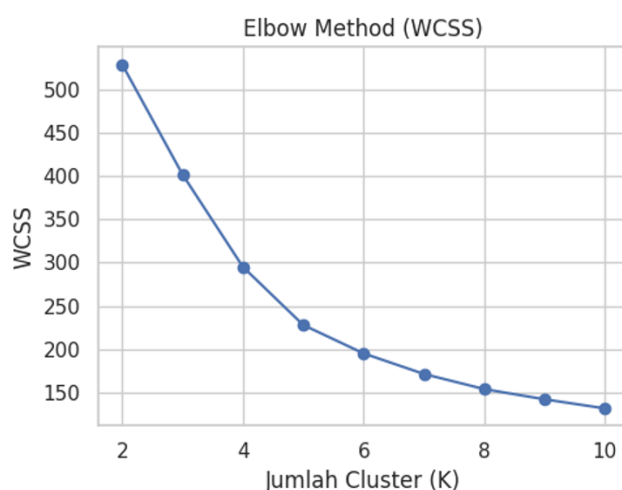
Penentuan Jumlah Cluster dengan Elbow Method

Berdasarkan hasil pengujian menggunakan Elbow Method, diperoleh hubungan antara jumlah cluster (K) dan nilai WCSS.

Grafik menunjukkan bahwa nilai WCSS mengalami penurunan yang signifikan dari K=2 hingga K=4. Setelah itu, penurunan mulai melambat, khususnya pada $K \geq 5$, yang ditandai dengan kurva yang semakin landai.

Titik perubahan pola tersebut mengindikasikan bahwa jumlah cluster optimal berada pada K=5, karena setelah nilai tersebut, penambahan cluster tidak lagi memberikan penurunan WCSS yang signifikan.

Dengan demikian, pada penelitian ini digunakan 5 cluster sebagai jumlah cluster optimal untuk proses K-Means.



Gambar 2. Grafik Elbow Method untuk Menentukan Jumlah Cluster Optimal (K = 5)

Hasil Clustering K-Means

Penerapan algoritma K-Means dengan jumlah cluster optimal K=5 menghasilkan pengelompokan data hujan ke dalam lima karakteristik yang berbeda. Setiap cluster

merepresentasikan pola distribusi reflektivitas radar yang berbeda, baik dari segi intensitas maupun cakupan area hujan.

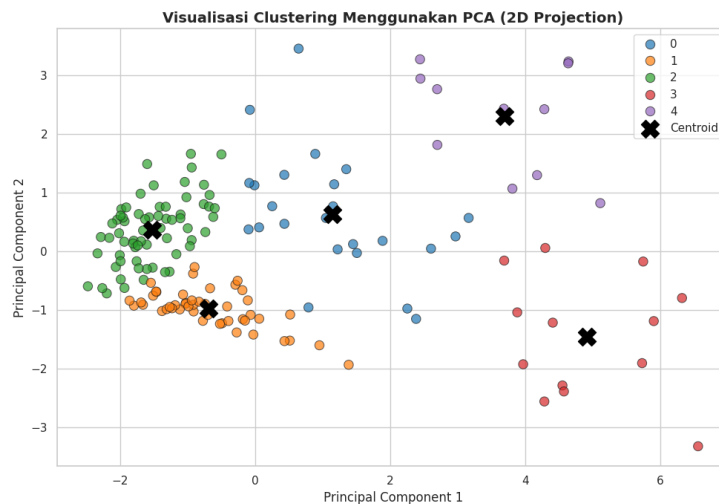
Tabel 4. Karakteristik Cluster Berdasarkan Nilai Centroid K-Means

C	Max_dbz	Mean_dbz	Std_dbz	Convective ratio	Stratiform ratio	Total coverage
0	52,12	24,79	4,47	0,01	0,99	0,12
1	39,45	24,29	4,06	0,00	1,00	0,00
2	36,44	22,25	2,23	0,00	1,00	0,03
3	59,50	27,26	7,20	0,08	0,92	0,12
4	54,77	28,00	5,41	0,03	0,97	0,45

Berdasarkan hasil profiling, terlihat bahwa masing-masing cluster memiliki karakteristik yang cukup berbeda, terutama pada nilai *max_dbz*, *mean_dbz*, dan *total_coverage*. Variasi ini menunjukkan adanya perbedaan intensitas hujan dari ringan hingga lebat.

Cluster dengan nilai reflektivitas tinggi, seperti Cluster 3 dan Cluster 4, menunjukkan karakteristik hujan dengan intensitas lebih kuat, ditandai dengan nilai *max_dbz* yang tinggi serta variasi reflektivitas yang besar. Sebaliknya, Cluster 1 dan Cluster 2 cenderung memiliki intensitas lebih rendah dengan dominasi hujan stratiform, yang terlihat dari nilai *convective_ratio* yang mendekati nol dan *stratiform_ratio* yang mendekati satu.

Dari sisi distribusi data, jumlah data pada masing-masing cluster tidak merata. Cluster dengan intensitas sedang cenderung memiliki jumlah data lebih banyak, sedangkan cluster dengan intensitas tinggi memiliki jumlah data yang lebih sedikit. Hal ini menunjukkan bahwa kejadian hujan lebat relatif lebih jarang dibandingkan hujan dengan intensitas sedang.



Gambar 3. Visualisasi Hasil Clustering K-Means Menggunakan PCA (2D Projection)

Visualisasi menggunakan reduksi dimensi PCA menunjukkan bahwa sebagian besar cluster dapat terpisah dengan cukup baik di ruang dua dimensi. Beberapa cluster terlihat membentuk kelompok yang kompak, sementara cluster lainnya memiliki penyebaran yang lebih luas, yang mengindikasikan variasi internal yang lebih tinggi dalam cluster tersebut.

Interpretasi Pola Hujan Berdasarkan Cluster

Berdasarkan hasil *clustering* yang merangkum fitur reflektivitas (dBZ), rasio konvektif-stratiform, dan luasan area (total_coverage), terbentuk lima profil cuaca yang merepresentasikan dinamika awan yang berbeda. Secara fundamental, nilai reflektivitas radar ekuivalen dengan intensitas presipitasi, di mana nilai di atas 40 dBZ umumnya menjadi batas ambang awan konvektif yang memicu hujan lebat, sementara nilai di bawahnya didominasi oleh sistem stratiform (Tomkins et al., 2023; Wang et al., 2024).

Cluster 0 (Konvektif Terisolasi): Memiliki nilai reflektivitas maksimum tinggi (52,12 dBZ) namun dengan luasan area sempit (12%). Karakteristik ini mengindikasikan sel awan konvektif lokal berumur pendek.

Cluster 1 (Gema Non-Presipitasi): Ditandai dengan cakupan area mendekati nol dan intensitas rata-rata negatif (-6,18 dBZ). Ini merupakan representasi dari *clear-air echoes* (debu/serangga) atau presipitasi yang menguap sebelum mencapai tanah.

Cluster 2 (Stratiform Stabil): Menunjukkan rata-rata dBZ terendah (22,25) dan variasi terkecil (std 2,23). Kelompok ini secara murni merepresentasikan hujan stratiform ringan-sedang yang berdurasi panjang tanpa turbulensi ekstrem.

Cluster 3 (Konvektif Kuat): Mencatat reflektivitas absolut tertinggi (Maks: 59,50 dBZ), variasi terbesar (std 7,20), dan kemunculan rasio konvektif (0,08). Pola ini menandakan fase matang Cumulonimbus penghasil hujan lebat.

Cluster 4 (Mesoscale Convective System/MCS): Memperlihatkan kombinasi intensitas puncak yang tinggi (54,77 dBZ) dengan luasan area yang sangat masif (45%). Secara spasial, ini adalah indikator kuat dari sistem *squall-line* atau badai regional yang membawa risiko banjir.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan pola intensitas curah hujan dari data radar NOAA NEXRAD Level II menggunakan algoritma K-Means dengan K=5 sebagai konfigurasi optimal. Nilai Silhouette Score sebesar 0.3871 dan Davies-Bouldin Index sebesar 0.8599 mencerminkan kualitas pemisahan cluster yang moderat, sebuah kondisi yang wajar pada data reflektivitas radar akibat sifat kontinum transisi intensitas hujan. Lima cluster yang terbentuk berhasil merepresentasikan spektrum rezim presipitasi yang berbeda, mulai dari gema non-presipitasi dan hujan stratiform stabil hingga sistem konvektif skala meso dengan cakupan area yang luas, dan mengonfirmasi kelayakan pendekatan clustering tanpa supervisi untuk pengenalan pola hujan berbasis data radar mentah. Penelitian selanjutnya disarankan memperluas pengujian ke data multistasi serta membandingkan K-Means dengan algoritma clustering lain seperti DBSCAN atau Gaussian Mixture Model.

DAFTAR REFERENSI

- Cuomo, J., & Chandrasekar, V. (2021). Use of Deep Learning for Weather Radar Nowcasting. *Journal of Atmospheric and Oceanic Technology*. <https://doi.org/10.1175/JTECH-D-21-0012.1>
- Dixon, M., & Romatschke, U. (2022). Three-Dimensional Convective–Stratiform Echo-Type Classification and Convectivity Retrieval from Radar Reflectivity. *Journal of Atmospheric and Oceanic Technology*, 39(11), 1685–1704. <https://doi.org/10.1175/JTECH-D-22-0018.1>
- El Hafyani, M., El Himdi, K., & El Adlouni, S.-E. (2024). Improving monthly precipitation prediction accuracy using machine learning models: a multi-view stacking learning technique. *Frontiers in Water*, 6. <https://doi.org/10.3389/frwa.2024.1378598>
- Ghada, W., Casellas, E., Herbinger, J., Garcia-Benadí, A., Bothmann, L., Estrella, N., ... Menzel, A. (2022). Stratiform and Convective Rain Classification Using Machine Learning Models and Micro Rain Radar. *Remote Sensing*, 14(18), 4563. <https://doi.org/10.3390/rs14184563>

- Helmus, J. J., & Collis, S. M. (2016). The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language. *Journal of Open Research Software*, 4(1), 25. <https://doi.org/10.5334/jors.119>
- Hu, C., Zhang, X., Liu, C., Liu, C., Wang, J., & Jian, S. (2022). Real-time Flood Classification Forecasting Based on k-means++ Clustering and Neural Network. *Water Resources Management*, 36(1), 103–117. <https://doi.org/10.1007/s11269-021-03014-y>
- Kisi, O., Heddami, S., Parmar, K. S., Petroselli, A., Külls, C., & Zounemat-Kermani, M. (2025). Integration of Gaussian process regression and K means clustering for enhanced short term rainfall runoff modeling. *Scientific Reports*, 15(1), 7444. <https://doi.org/10.1038/s41598-025-91339-8>
- Koutsovili, E.-I., Tzoraki, O., Theodossiou, N., & Tsekouras, G. E. (2023). Early Flood Monitoring and Forecasting System Using a Hybrid Machine Learning-Based Approach. *ISPRS International Journal of Geo-Information*, 12(11), 464. <https://doi.org/10.3390/ijgi12110464>
- Li, J., Zheng, A., Guo, W., Bandyopadhyay, N., Zhang, Y., & Wang, Q. (2023). Urban flood risk assessment based on DBSCAN and K-means clustering algorithm. *Geomatics, Natural Hazards and Risk*, 14(1). <https://doi.org/10.1080/19475705.2023.2250527>
- Li, W., Chen, H., & Han, L. (2024). Improving Explainability of Deep Learning for Polarimetric Radar Rainfall Estimation. *Geophysical Research Letters*, 51(11). <https://doi.org/10.1029/2023GL107898>
- Liu, M., Zuo, J., Tan, J., & Liu, D. (2024). Comparing and Optimizing Four Machine Learning Approaches to Radar-Based Quantitative Precipitation Estimation. *Remote Sensing*, 16(24). <https://doi.org/10.3390/rs16244713>
- Mihuleţ, E., Burcea, S., Mihai, A., & Czibula, G. (2023). Enhancing the Performance of Quantitative Precipitation Estimation Using Ensemble of Machine Learning Models Applied on Weather Radar Data. *Atmosphere*, 14(1), 182. <https://doi.org/10.3390/atmos14010182>
- Mikidadi, N. D., Huang, X., & Bu, L. (2025). The Application of the Convective–Stratiform Classification Algorithm for Feature Detection in Polarimetric Radar Variables and QPE Retrieval During Warm-Season Convection. *Remote Sensing*, 17(7), 1176. <https://doi.org/10.3390/rs17071176>
- Miraftebzadeh, S. M., Colombo, C. G., Longo, M., & Foadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, Vol. 11, pp. 119596–119633. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2023.3327640>
- Muñoz, P., Orellana-Alvear, J., Bendix, J., Feyen, J., & Céleri, R. (2021). Flood early warning systems using machine learning techniques: The case of the tomebamba catchment at the southern Andes of Ecuador. *Hydrology*, 8(4). <https://doi.org/10.3390/hydrology8040183>
- Rahimpour, M., Rahimzadegan, M., Nosratpour, R., Homayouni, S., & Behrangi, A. (2025). A Novel Machine Learning-Based Clustering-Merging Method for Improving Extreme Precipitation Estimation. *Advances in Atmospheric Sciences*, 42(8), 1693–1714. <https://doi.org/10.1007/s00376-024-4315-3>

- Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*, 528, 178–199. <https://doi.org/10.1016/j.neucom.2023.01.043>
- Schumacher, C., & Funk, A. (2023). Assessing Convective-Stratiform Precipitation Regimes in the Tropics and Extratropics With the GPM Satellite Radar. *Geophysical Research Letters*, 50(14). <https://doi.org/10.1029/2023GL102786>
- Verdelho, F. F., Beneti, C., Pavam, L. G., Calvetti, L., Oliveira, L. E. S., & Zanata Alves, M. A. (2024). Quantitative Precipitation Estimation Using Weather Radar Data and Machine Learning Algorithms for the Southern Region of Brazil. *Remote Sensing*, 16(11). <https://doi.org/10.3390/rs16111971>
- Wang, Y., Tang, L., Chang, P.-L., & Tang, Y.-S. (2021). Separation of convective and stratiform precipitation using polarimetric radar data with a support vector machine method. *Atmospheric Measurement Techniques*, 14(1), 185–197. <https://doi.org/10.5194/amt-14-185-2021>
- Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, 1–45. <https://doi.org/10.7717/PEERJ-CS.2286>