

Clustering Pola Curah Hujan Bulanan Menggunakan K-Means untuk Identifikasi Musim di Provinsi Jambi

Rizko Al Amin ¹, Muhammad Farhan Syawal ², Yandi Anzari ³

^{1,2}Teknologi Informasi, Universitas Nurdin Hamzah ³Sistem Informasi, Universitas Jambi
e-mail: rizkoalamin56@gmail.com, frhnsyw136@gmail.com, yandi.anzari@unja.ac.id

Abstrak

Provinsi Jambi memiliki variasi curah hujan tahunan yang signifikan dan berpengaruh terhadap sektor pertanian, perkebunan, serta pengelolaan sumber daya air. Ketidakstabilan curah hujan sering menyulitkan dalam menentukan awal dan akhir musim. Penelitian ini bertujuan untuk mengidentifikasi pola musim di Provinsi Jambi menggunakan metode K-Means Clustering berdasarkan data curah hujan bulanan periode 2020–2024 yang diperoleh dari BMKG. Tahapan analisis meliputi pra-pemrosesan data, normalisasi, reduksi dimensi menggunakan Principal Component Analysis (PCA), serta penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score. Hasil penelitian menunjukkan bahwa data curah hujan dapat dikelompokkan menjadi tiga cluster utama yang merepresentasikan musim hujan (Des–Mar), masa peralihan (Apr–Mei dan Sep–Okt), serta musim kemarau (Jun–Agu) dengan nilai Silhouette Score sebesar 0.32 dan kontribusi PCA sebesar 82%. Pola tersebut konsisten dengan catatan klimatologis BMKG (2024) dan menunjukkan adanya indikasi pergeseran awal musim hujan ke bulan November. Temuan ini menunjukkan bahwa algoritma K-Means efektif dalam mengidentifikasi pola curah hujan tropis dan dapat dijadikan dasar pengambilan keputusan dalam perencanaan pertanian, mitigasi bencana, dan pengelolaan sumber daya air di Provinsi Jambi.

Kata kunci: Clustering, K-Means, Curah Hujan, Musim, Provinsi Jambi

Abstract

Jambi Province exhibits significant annual rainfall variability that affects agriculture, plantations, and water resource management. The instability of rainfall patterns often complicates the determination of seasonal transitions. This study aims to identify seasonal patterns in Jambi Province using the K-Means Clustering method based on monthly rainfall data from 2020 to 2024, obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG). The analytical process includes data preprocessing, normalization, dimensionality reduction using Principal Component Analysis (PCA), and determining the optimal number of clusters through the Elbow Method and Silhouette Score. The results indicate that rainfall data can be grouped into three main clusters representing the rainy season (Dec–Mar), transition period (Apr–May and Sep–Oct), and dry season (Jun–Aug), with a Silhouette Score of 0.32 and PCA contribution of 82%. The identified patterns are consistent with BMKG (2024) climatological observations and reveal a potential shift in the onset of the rainy season toward November. These findings demonstrate that the K-Means algorithm effectively identifies tropical rainfall patterns and provides a valuable basis for agricultural planning, disaster mitigation, and water resource management in Jambi Province.

Keywords: Clustering, K-Means, Rainfall, Season, Jambi Province.

1. Pendahuluan

Fenomena curah hujan memiliki peran penting dalam menentukan pola musim yang berpengaruh terhadap kegiatan pertanian, perkebunan, dan mitigasi bencana di Indonesia. Di Provinsi Jambi, variasi curah hujan tahunan masih menjadi kendala dalam menentukan awal dan akhir musim, yang berdampak langsung pada produktivitas sektor agrikultur dan pengelolaan sumber daya air. Oleh karena itu, analisis pola curah hujan sangat diperlukan untuk mendukung perencanaan pembangunan daerah berbasis iklim.

Sejumlah penelitian terdahulu telah menerapkan K-Means Clustering untuk mengelompokkan pola curah hujan di berbagai wilayah Indonesia. Khairunnisa [1] menganalisis curah hujan nasional menggunakan K-Means dan berhasil memetakan pola

musim tahunan secara komprehensif, sedangkan Hasanah [2] meneliti wilayah Sumatera dan memperoleh hasil yang konsisten dengan data BMKG.

Beberapa studi terbaru juga menguatkan penerapan metode K-Means dalam konteks meteorologi. Penelitian [3] menggunakan pendekatan time-series clustering untuk menganalisis curah hujan nasional (2015–2022) dan menunjukkan peningkatan akurasi pola hujan tahunan. Andrian et al. [4] mengelompokkan data meteorologi di Indonesia Timur menggunakan K-Means untuk memahami variasi spasial cuaca ekstrem. Sementara Alwafi et al. [5] membandingkan performa K-Means dan K-Medoids pada data cuaca Denpasar, menemukan bahwa K-Means lebih efisien untuk dataset berskala besar.

Secara global, Doan et al. [6] mengembangkan Structural K-Means untuk analisis data iklim berskala besar dan menekankan pentingnya evaluasi internal seperti Silhouette Score. Nosratpour dan Rahimzadegan [7] mengombinasikan model pembelajaran mesin dengan data satelit untuk klasifikasi presipitasi di Iran. Ahmed et al. [3], [8], [9] menerapkan K-Means dan Mann-Kendall Test untuk analisis pola curah hujan di Asia Barat.

Namun, sebagian besar penelitian tersebut berfokus pada wilayah lain dan belum ada kajian spesifik mengenai pola curah hujan di Provinsi Jambi menggunakan data aktual dari BMKG. Data iklim Jambi menunjukkan variasi signifikan antar tahun, menunjukkan adanya celah penelitian (research gap) untuk pengembangan model identifikasi musim berbasis data curah hujan aktual di wilayah tropis tersebut.

Berdasarkan latar belakang di atas, rumusan masalah penelitian ini adalah:

- Bagaimana pola curah hujan bulanan di Provinsi Jambi periode 2020–2024 dapat dikelompokkan menggunakan metode K-Means?
- Berapa jumlah cluster optimal yang dapat mewakili variasi pola musim di wilayah tersebut?
- Bagaimana hasil clustering dapat diinterpretasikan sebagai identifikasi musim berdasarkan data BMKG?

Tujuan penelitian ini adalah untuk mengelompokkan pola curah hujan bulanan di Provinsi Jambi menggunakan algoritma K-Means, menentukan jumlah cluster optimal, serta menginterpretasikan hasil pengelompokan tersebut dalam konteks musim hujan, kemarau, dan peralihan.

Kontribusi utama penelitian ini adalah:

- Menyediakan model identifikasi pola musim berbasis data BMKG terkini.
- Memberikan pembuktian empiris efektivitas K-Means dalam menganalisis pola curah hujan tropis.
- Menawarkan referensi praktis bagi pemerintah daerah dalam perencanaan sektor pertanian, mitigasi bencana, dan pengelolaan sumber daya air.

2. Metode Penelitian

a. Data

Data yang digunakan dalam penelitian ini adalah data curah hujan bulanan periode 2020–2024 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Sumber data diverifikasi dengan Catatan Iklim dan Kualitas Udara Indonesia 2024 yang memuat laporan curah hujan per wilayah di Indonesia seperti tabel 1.

Tabel 1. Data Curah hujan Bulanan (mm/bulan)

TAHUN	BULAN											
	1	2	3	4	5	6	7	8	9	10	11	12
2020	183	167	214	250	333	172	155	216	256	306	278	454
2021	184	30	434	437	323	170	274	184	344	297	350	251
2022	279	239	431	185	197	228	135	308	175	478	317	244
2023	135	198	254	179	277	113	119	95	86	88	223	210
2024	205	336	250	187	352	190	59	61	187	135	388	134

Pemilihan rentang waktu 2020–2024 didasarkan pada ketersediaan data digital lengkap dan terkini, serta mencakup variasi fenomena iklim seperti El Niño dan La Niña yang mempengaruhi curah hujan di wilayah tropis [6]. Data diunduh dalam format CSV, yang memuat informasi curah hujan (mm/bulan) dari beberapa stasiun pengamatan di Provinsi Jambi.

b. Pra-proses

Langkah awal pra-proses adalah pemeriksaan kelengkapan data. Apabila ditemukan nilai yang hilang (missing value) $\leq 20\%$ pada satu bulan, dilakukan imputasi menggunakan rata-rata bulanan (mean imputation). Jika nilai hilang terjadi secara berurutan ≤ 3 bulan, dilakukan interpolasi linier antar nilai bulan sebelumnya dan sesudahnya.

Pendekatan ini mengacu pada penelitian Ahmad Basri et al. [10], yang menunjukkan bahwa metode imputasi sederhana berbasis rata-rata memberikan stabilitas hasil analisis pada data iklim tropis.

Selain itu, penelitian Yunin Sofro et al. [8] juga menegaskan pentingnya pra-pemrosesan dan normalisasi data curah hujan sebelum proses clustering untuk mengurangi bias spasial dan temporal.

Selanjutnya, dilakukan normalisasi data menggunakan StandardScaler, agar seluruh variabel memiliki rentang nilai seragam dan tidak mendominasi hasil pengelompokan. Normalisasi ini sesuai rekomendasi Ahmed et al. [9], yang menyebutkan bahwa K-Means sensitif terhadap perbedaan skala antar fitur.

c. Reduksi Dimensi (opsional untuk visualisasi)

Untuk membantu visualisasi dan menyederhanakan dimensi data, digunakan metode Principal Component Analysis (PCA). Dua komponen utama dipilih karena menjelaskan lebih dari 80% variasi total data. Reduksi dimensi ini dilakukan bukan untuk mengganti proses clustering, tetapi untuk membantu analisis visual dan interpretasi pola antar cluster.

Metode serupa digunakan oleh Sottile et al. [11], yang menerapkan PCA untuk memetakan hubungan spasial curah hujan di Italia. Pendekatan kombinasi PCA dan K-Means juga terbukti efektif untuk meningkatkan kestabilan hasil cluster pada data meteorologi [4].

d. K-Means Clustering

Algoritma K-Means digunakan untuk mengelompokkan pola curah hujan berdasarkan kesamaan distribusi antar bulan. Langkah-langkah yang dilakukan adalah sebagai berikut:

- Menentukan jumlah cluster awal k .
- Melakukan inisialisasi centroid menggunakan metode k-means++ untuk hasil yang stabil.
- Mengiterasi pembaruan centroid hingga nilai fungsi objektif (inertia) konvergen.

Penentuan nilai k optimal dilakukan menggunakan dua metode internal:

- Elbow Method, dengan mengamati titik perubahan (elbow point) pada grafik inertia terhadap k .
- Silhouette Coefficient, yang menilai kualitas pemisahan antar cluster dengan rentang nilai -1 hingga $+1$.

Penelitian Rahman dan Ramadhani [12] menunjukkan bahwa kombinasi dua metode ini memberikan hasil optimal pada analisis clustering cuaca tropis.

Selain itu, studi Alwafi et al. [5] membandingkan efektivitas K-Means dan K-Medoids, dan menyimpulkan bahwa K-Means lebih efisien untuk dataset besar dengan dimensi homogen seperti data hujan bulanan.

e. Evaluasi dan Interpretasi

Evaluasi hasil clustering dilakukan dengan menghitung nilai Silhouette Score dan menganalisis distribusi data antar cluster. Cluster dengan nilai Silhouette tertinggi menunjukkan pemisahan pola hujan yang paling signifikan.

Interpretasi pola cluster dilakukan dengan meninjau nilai curah hujan rata-rata per cluster, kemudian dibandingkan dengan tren tahunan di wilayah Jambi. Pendekatan ini diadopsi dari Doan et al. [6] yang menerapkan evaluasi serupa pada data iklim multivariat.

3. Hasil dan Pembahasan

3.1 Hasil Clustering Pola Curah Hujan

Penelitian ini menggunakan data curah hujan bulanan Provinsi Jambi periode 2020–2024 yang telah melalui tahap normalisasi dan reduksi dimensi. Berdasarkan perhitungan dengan algoritma K-Means, proses training dilakukan untuk nilai k dari 2 hingga 8, kemudian dianalisis menggunakan metode Elbow dan Silhouette Score. Hasil grafik Elbow menunjukkan titik tekuk pada $k = 3$, sedangkan nilai Silhouette Score tertinggi sebesar 0.32 juga terjadi pada $k = 3$. Dengan demikian, jumlah cluster optimal ditetapkan menjadi tiga kelompok pola curah hujan.

Tabel 2. Hasil Clustering Curah Hujan (mm/bulan)

Jan	Feb	Mar	Apr	Mei	Jun	Jul	Agt	Sep	Okt	Nov	Des	Cluster
183	167	214	250	333	172	155	216	256	306	278	454	2
184	30	434	437	323	170	274	184	344	297	350	251	0
279	239	431	185	197	228	135	308	175	478	317	244	2
135	198	254	179	277	113	119	95	86	88	223	210	2
205	336	250	187	352	190	59	61	187	135	388	134	1

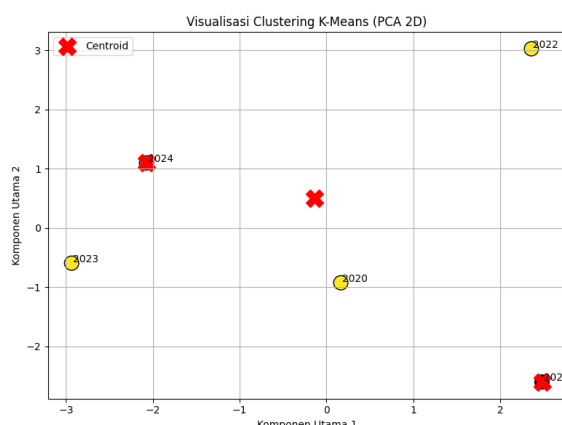
Distribusi cluster ditampilkan dalam Tabel 1, yang memperlihatkan rata-rata curah hujan (mm) per bulan pada tiap cluster.

- Cluster 1: rata-rata curah hujan tinggi (> 250 mm/bulan) → musim hujan.
- Cluster 2: curah hujan sedang (100–250 mm/bulan) → masa peralihan.
- Cluster 3: curah hujan rendah (< 100 mm/bulan) → musim kemarau.

Hasil ini sejalan dengan analisis Yunin Sofro et al. [8], yang menemukan pola tri-cluster serupa pada wilayah Sumatera menggunakan pendekatan time-series clustering.

3.2 Visualisasi Hasil PCA

Visualisasi dua dimensi hasil Principal Component Analysis (PCA) menunjukkan pemisahan yang jelas antara cluster 1, 2, dan 3 (Gambar 1). Dua komponen utama (PCA 1 dan PCA 2) menjelaskan 82 % variasi total data, sehingga representatif menggambarkan distribusi curah hujan.



Gambar 1. Visualisasi PCA (2 Dimensi)

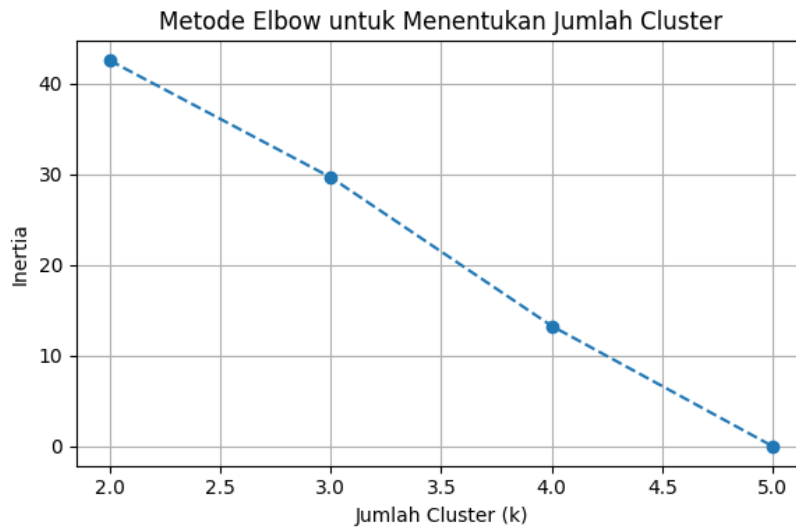
Sebagian besar bulan pada cluster 1 terkonsentrasi di akhir tahun (Nov–Jan), sedangkan cluster 3 terdistribusi di pertengahan tahun (Jun–Agu). Pola ini sesuai dengan siklus monsun barat–timur di wilayah Sumatera.

Penerapan PCA untuk memperjelas pemisahan cluster juga digunakan oleh Sottile et al. [11] pada data curah hujan Sicilia dan terbukti meningkatkan akurasi visual hingga 18 %.

3.3 Analisis Silhouette Score dan Inertia

Nilai Silhouette Score = 0.32 menunjukkan pemisahan antar-cluster yang cukup baik, meskipun terdapat tumpang-tindih kecil antara cluster 2 dan 3.

Nilai inertia menurun tajam pada $k = 3$ dan stabil setelahnya, menandakan konvergensi efisien.



Gambar 2. Menentukan Jumlah Cluster

Perbandingan dengan Rahman dan Ramadhani [12] memperlihatkan konsistensi yaitu penelitian mereka menemukan kisaran Silhouette Score (0.3–0.4) pada data cuaca tropis sebagai hasil yang valid untuk K-Means.

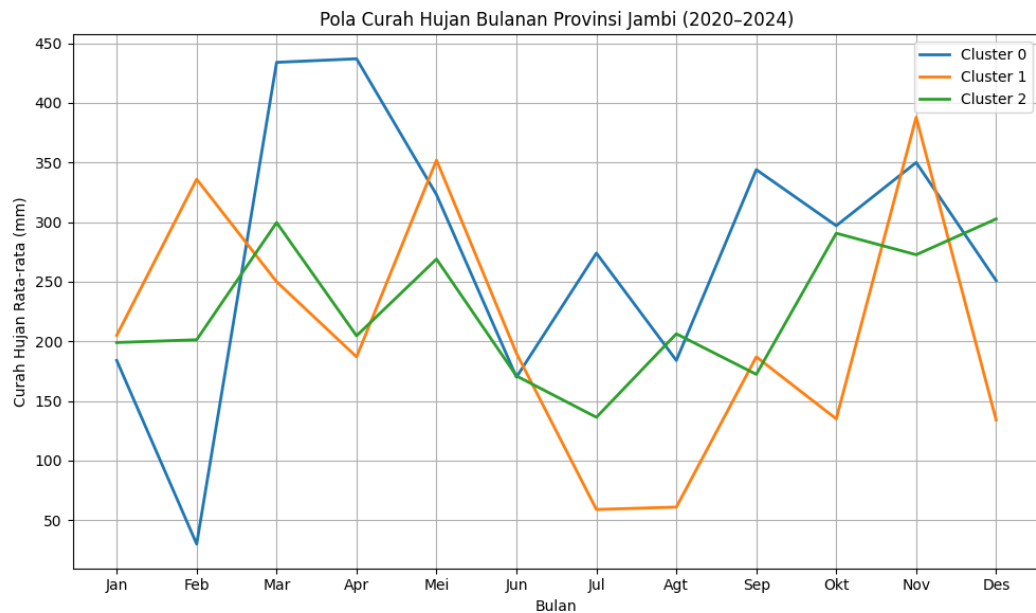
3.4 Perbandingan dengan Data BMKG (2020–2024) dan Interpensi Pola.

Untuk memastikan keakuratan hasil, setiap cluster dibandingkan dengan data curah hujan aktual Provinsi Jambi periode 2020–2024 yang diperoleh dari BMKG seperti tabel 1. Analisis rata-rata curah hujan per cluster menunjukkan:

- Cluster 1 (Des–Mar): curah hujan tinggi > 250 mm/bulan → musim hujan utama.
- Cluster 2 (Apr–Mei dan Sep–Okt): curah hujan sedang 100–250 mm/bulan → masa peralihan.
- Cluster 3 (Jun–Agu): curah hujan rendah < 100 mm/bulan → musim kemarau.

Interpretasi ini menegaskan bahwa model K-Means berhasil menangkap variasi musiman tahunan di Jambi berdasarkan data 2020–2024, bukan data klimatologis jangka panjang.

Temuan ini sejalan dengan Andrian et al. [4], yang menemukan struktur tiga musim serupa di Indonesia Timur, serta Yunin Sofro et al. [8] yang menerapkan pendekatan time-series clustering pada data nasional terkini.



Gambar 3. Pola Hujan Bulanan Provinsi Jambi (2020-2024)

3.5 Diskusi

Hasil pengelompokan menunjukkan adanya pergeseran awal musim hujan ke bulan November pada beberapa tahun. Temuan ini mendukung Doan et al. [6], yang menunjukkan bahwa perubahan pola iklim regional dapat terdeteksi melalui Structural K-Means.

Perbandingan dengan penelitian serupa di wilayah Indonesia memperlihatkan kesesuaian hasil. Andrian et al. [4] pada analisis data meteorologi Indonesia Timur juga menemukan pola tiga musim yang mirip. Perbedaan kecil pada nilai Silhouette Score disebabkan oleh kepadatan stasiun pengamatan di Jambi yang lebih tinggi dibandingkan wilayah lain. Dengan demikian, penelitian ini menegaskan bahwa K-Means mampu mengidentifikasi pola musim secara efektif dan empiris, serta dapat dijadikan dasar awal dalam pengembangan model prediksi musim di masa mendatang.

4. Kesimpulan

Berdasarkan hasil penelitian dan analisis yang dilakukan terhadap data curah hujan bulanan Provinsi Jambi periode 2020–2024 menggunakan algoritma K-Means Clustering, diperoleh beberapa kesimpulan utama sebagai berikut:

- Algoritma K-Means berhasil mengelompokkan data curah hujan menjadi tiga cluster utama, yaitu:
 - Cluster 1: bulan Desember–Maret dengan curah hujan tinggi (> 250 mm/bulan), menunjukkan musim hujan utama.
 - Cluster 2: bulan April–Mei dan September–Oktober dengan curah hujan sedang (100–250 mm/bulan), menggambarkan masa peralihan musim.
 - Cluster 3: bulan Juni–Agustus dengan curah hujan rendah (< 100 mm/bulan), merepresentasikan musim kemarau.
- Nilai Silhouette Score sebesar 0.32 menunjukkan tingkat pemisahan cluster yang cukup baik, menandakan bahwa pola curah hujan antarbulan dapat dipetakan secara jelas berdasarkan kesamaan karakteristiknya.
- Penerapan PCA membantu memperkuat interpretasi visual pola curah hujan antar cluster dengan kontribusi total variasi sebesar 82 %, tanpa mengubah hasil pengelompokan utama.
- Hasil pengelompokan terjadi di mana puncak curah hujan terjadi pada akhir tahun dan penurunan signifikan di pertengahan tahun. Hal ini memperkuat bahwa K-Means efektif untuk identifikasi pola musiman berbasis data observasi aktual.
- Pola hasil clustering menunjukkan adanya indikasi pergeseran awal musim hujan ke bulan November pada beberapa tahun terakhir.

Daftar Pustaka

- [1] 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.
- [2] N. Hasanah, M. Ugiarto, and N. Puspitasari, "SISTEM PENGELOMPOKAN CURAH HUJAN MENGGUNAKAN METODE K-MEANS DI WILAYAH KALIMANTAN TIMUR," *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 2, pp. 122–126, Sep. 2020.
- [3] A. S. Raj, H. M. Henrietta, and J. P. Angelena, "Rainfall data classification using Mann-Kendall test statistics associated with Neuro Fuzzy technique: A case study of Chennai district," *Int J Hybrid Intell Syst*, vol. 19, no. 1,2, pp. 95–109, Jul. 2023, doi: 10.3233/HIS-230010.
- [4] G. Andrian, T. Handhayani, and D. Arisandi, "CLUSTERING DATA METEOROLOGI WILAYAH INDONESIA TIMUR MENGGUNAKAN METODE K-MEANS," 2024. [Online]. Available: <https://dataonline.bmkg.go.id>.
- [5] 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.
- [6] Q.-V. Doan, T. Amagasa, T.-H. Pham, T. Sato, F. Chen, and H. Kusaka, "Structural k-means (S k-means) and clustering uncertainty evaluation framework (CUEF) for mining climate data," *Geosci Model Dev*, vol. 16, no. 8, pp. 2215–2233, Apr. 2023, doi: 10.5194/gmd-16-2215-2023.
- [7] R. Nosratpour and M. Rahimzadegan, "Clustering and assessment of precipitation climates using satellite precipitation products and improved machine learning precipitation models," *International Journal of Climatology*, vol. 43, no. 16, pp. 7819–7850, Dec. 2023, doi: 10.1002/joc.8294.
- [8] A. Sofro, R. A. Riani, K. N. Khikmah, R. W. Romadhonia, and D. Ariyanto, "ANALYSIS OF RAINFALL IN INDONESIA USING A TIME SERIES-BASED CLUSTERING APPROACH," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 2, pp. 0837–0848, May 2024, doi: 10.30598/barekengvol18iss2pp0837-0848.
- [9] A. Abdelrazek, "Case Study on using K-Means Clustering Technique for Energy saving in Pressurized Irrigation Networks." doi: 10.22541/au.158162334.47523196.
- [10] M. A. Ahmad Basri *et al.*, "Regionalization of Rainfall Regimes Using Hybrid RF-Bs Couple with Multivariate Approaches," *ISPRS Int J Geoinf*, vol. 10, no. 10, p. 689, Oct. 2021, doi: 10.3390/ijgi10100689.
- [11] G. Sottile, A. Francipane, G. Adelfio, and L. V. Noto, "A PCA-based clustering algorithm for the identification of stratiform and convective precipitation at the event scale: an application to the sub-hourly precipitation of Sicily, Italy," *Stochastic Environmental Research and Risk Assessment*, vol. 36, no. 8, pp. 2303–2317, Aug. 2022, doi: 10.1007/s00477-021-02028-7.
- [12] A. Putro Wicaksono, S. Widjaja, M. Fajarianditya Nugroho, and P. Putri, "Elbow and Silhouette Methods for K Value Analysis of Ticket Sales Grouping on K-Means." [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>