

ANALISIS POLA DATA KLIMATOLOGI POTENSIAL DENGAN PRINCIPAL COMPONENT ANALYSIS

Syahreni ¹, Agung Teguh Wibowo Almais ², Fresy Nugroho ³, Cahyo Crysdian ⁴,
Sri Harini ⁵, Ahmad Zarkoni ⁶, Ma'muri ⁷, Tomy Ivan Sugiharto ⁸

^{1,2,3,4,5,6} Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
Jalan Gajayana No. 50, Jatimulyo, Kecamatan Lowokwaru, Kota Malang, Jawa Timur, Indonesia

^{7,8} Badan Meteorologi, Klimatologi dan Geofisika
Jl. Zentana No. 33, Desa Ngijo, Kecamatan Karangploso, Kabupaten Malang, Jawa Timur, Indonesia
240605210002@student.uin-malang

ABSTRAK

Perubahan iklim global menyebabkan pola curah hujan di berbagai wilayah menjadi semakin kompleks dan sulit diprediksi, termasuk di Jawa Timur. Kompleksitas tersebut menimbulkan tantangan dalam analisis klimatologi karena data curah hujan bersifat multivariat dan memiliki tingkat variabilitas tinggi. Hal ini menyebabkan proses identifikasi faktor dominan yang memengaruhi variasi curah hujan menjadi sulit dilakukan secara langsung. Penelitian ini bertujuan untuk menyederhanakan data curah hujan yang kompleks melalui penerapan metode Principal Component Analysis (PCA) guna memperoleh representasi data yang lebih efisien tanpa kehilangan informasi penting. Data penelitian diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Jawa Timur, khususnya dari Stasiun Sumberpucung dan Karangploso selama periode Januari hingga Desember 2023. Tahapan penelitian meliputi pembersihan data, normalisasi menggunakan Standard Scaler, serta analisis reduksi dimensi menggunakan PCA. Hasil penelitian menunjukkan bahwa dua komponen utama (PC1 dan PC2) mampu menjelaskan 73,33% total variansi data, dengan PC1 sebesar 42,14% dan PC2 sebesar 31,19%. Hasil ini membuktikan bahwa PCA efektif dalam mereduksi enam variabel klimatologis menjadi dua komponen representatif, sehingga dapat digunakan sebagai pendekatan awal yang kuat dalam analisis dan visualisasi pola curah hujan di wilayah Malang.

Kata kunci : Klimatologi, Jawa Timur, PCA, Curah Hujan, unsupervised learning

1. PENDAHULUAN

Perubahan iklim global secara signifikan memengaruhi pola curah hujan di berbagai wilayah Indonesia [1]. Fenomena ini terlihat nyata di wilayah Malang yang memiliki karakteristik iklim tropis dengan pola musim yang jelas, yaitu musim hujan yang berlangsung dari November hingga April dan diikuti oleh musim kemarau [2]. [3]. Keanekaragaman topografi yang mencakup dataran rendah hingga daerah pegunungan turut berperan dalam menentukan sebaran dan intensitas curah hujan di wilayah ini [4]. Akibatnya, curah hujan tidak hanya memengaruhi ekosistem alami, tetapi juga berdampak besar terhadap sektor pertanian, pengelolaan sumber daya air, serta upaya mitigasi bencana [5].

Variasi spasial dan temporal curah hujan menjadi tantangan dalam analisis data klimatologis [6], [7]. Oleh karena itu, diperlukan pemantauan intensif oleh lembaga resmi seperti Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk menyediakan data yang akurat. Kompleksitas pola curah hujan ini mendorong penerapan pendekatan komputasi cerdas agar dapat menghasilkan representasi data yang lebih bermakna [8]. Salah satu pendekatan yang banyak digunakan adalah *Principal Component Analysis* (PCA) [9]. yaitu teknik reduksi dimensi yang berfungsi menyaring variabel paling berpengaruh dalam suatu himpunan data [10].

Dalam penelitian klimatologi, PCA telah terbukti menjadi kerangka analisis yang andal untuk menyederhanakan data kompleks [11]. PCA berperan penting dalam menyaring informasi utama dari variabel cuaca yang saling berkorelasi, sehingga mampu mengurangi redundansi data dan menekankan faktor dominan.

Meskipun metode *Principal Component Analysis* (PCA) [6] telah banyak diterapkan dalam penelitian klimatologi, penerapannya untuk analisis curah hujan di Indonesia masih terbatas. [13] Sebagian besar studi sebelumnya lebih berfokus pada prediksi berbasis deret waktu, sementara penggunaan PCA untuk reduksi dimensi data curah hujan di tingkat regional masih jarang dilakukan. Kondisi ini menyebabkan analisis curah hujan sering kali kurang efisien karena tingginya korelasi antarvariabel dan keterbatasan data spasial. Penelitian ini bertujuan menerapkan metode PCA dalam mereduksi [12] dimensi data curah hujan multivariat dari Stasiun Klimatologi Sumberpucung dan Karangploso guna memperoleh representasi data yang lebih sederhana dan informatif. Data klimatologis BMKG Jawa Timur tahun 2023 yang mencakup enam variabel utama—curah hujan harian, curah hujan maksimum, suhu udara, evaporasi, lama penyinaran matahari, dan kelembaban udara—diproses melalui tahap pembersihan, normalisasi dengan *Standard Scaler*, serta analisis PCA. Hasil reduksi dimensi kemudian divisualisasikan dalam

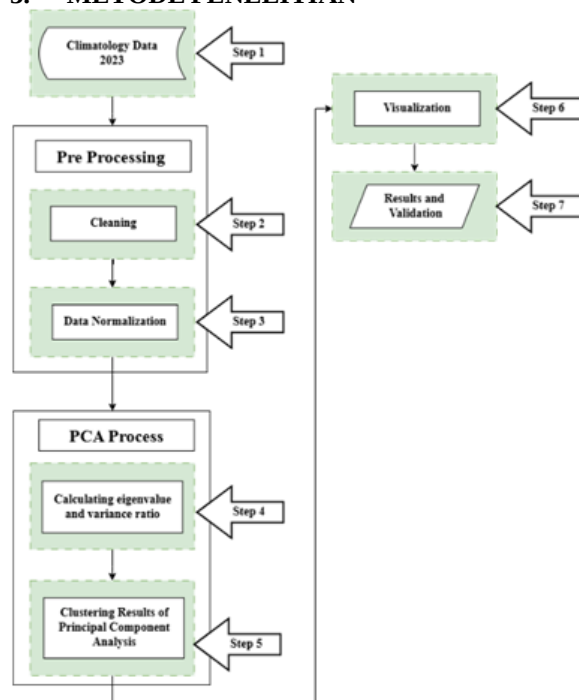
ruang dua dan tiga dimensi untuk mengidentifikasi pola curah hujan dominan di wilayah Malang.

2. TINJAUAN PUSTAKA

Menurut Almais et al., *Principal Component Analysis* (PCA) merupakan metode ekstraksi fitur yang efektif dalam meminimalkan kompleksitas data [12]. Selanjutnya, menurut Mpanza et al., kombinasi antara PCA dan *Self-Organizing Map* (SOM) terbukti efektif dalam proses klasifikasi curah hujan [14]. Namun demikian, sebagian besar penelitian terdahulu lebih berfokus pada aspek prediksi berbasis deret waktu (*time series*) atau klasifikasi konvensional [15]. Dalam penelitiannya, Hendrawati et al. menjelaskan bahwa teknik *clustering* digunakan untuk menganalisis data curah hujan deret waktu di Jawa Barat melalui pendekatan ARIMA dan *Model-Based Clustering* [16]. Menurut Lima et al., karakterisasi curah hujan ekstrem di Brasil dilakukan dengan metode PDF dan *clustering* yang berfokus pada kejadian hujan ekstrem di Rio de Janeiro [17]. Handhayani et al. meneliti pola iklim di Sumatra menggunakan *Agglomerative Clustering* berdasarkan data meteorologi harian dari 17 kota [18]. Menurut Urgiles et al, penelitiannya menganalisis pola spasial-temporal curah hujan ekstrem di Andes melalui *Clustering Analysis* [19].

Sementara itu, Wu et al. melakukan penelitian tentang regionalisasi curah hujan dan analisis multiskala dengan menggabungkan metode PCA, *Multivariate SOM*, dan *Wavelet* menggunakan data klimatologi ERA5 di Shanxi, China [20].

3. METODE PENELITIAN



Gambar 1. Desain alur sistem reduksi dimensi data curah hujan menggunakan PCA.

Metode penelitian ini dibagi menjadi dua bagian utama, yaitu tahap persiapan materi (data) dan metodologi yang diusulkan. Bagian metodologi menjelaskan penerapan *Principal Component Analysis* (PCA) untuk mereduksi dimensi variabel curah hujan. Data yang digunakan diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Jawa Timur, khususnya dari Stasiun Klimatologi Sumberpucung dan Karangploso. Dataset tersebut mencakup periode 1 Januari hingga 31 Desember 2023, dengan total 713 entri data dan 6 variabel.

Langkah-langkah *Principal Component Analysis* (PCA) yang dilakukan adalah sebagai berikut:

- Data diperoleh dari situs resmi BMKG Jawa Timur untuk periode yang telah ditentukan.
- Proses ini melibatkan penghapusan kesalahan, inkonsistensi, dan nilai yang hilang guna memastikan kualitas data.
- Normalisasi dilakukan untuk memastikan seluruh variabel memiliki skala yang seragam dan dapat dibandingkan, sesuai dengan standar yang disyaratkan dalam metode PCA [21]. Proses ini mengubah setiap nilai fitur agar memiliki rata-rata 0 dan simpangan baku 1 [22]. Dalam penelitian ini, normalisasi dilakukan menggunakan metode *Standard Scaler* (SS) dengan rumus sebagai berikut:

$$X_{standard} = \frac{x - \mu}{\sigma} \quad (1)$$

Di mana X merupakan nilai data asli, sedangkan μ , adalah nilai rata-rata dari seluruh data pada satu fitur, yang dihitung menggunakan rumus sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

σ merupakan simpangan baku, yaitu akar kuadrat dari varians (s^2). Varians dihitung dengan rumus sebagai berikut:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} \quad (3)$$

- Setelah proses normalisasi, langkah selanjutnya adalah menghitung matriks kovarians dari data untuk memahami hubungan antarvariabel. PCA beroperasi dengan mengidentifikasi *eigenvektor* dan *eigenvalue* dari matriks kovarians tersebut [21]. *Eigenvektor* menentukan arah dari komponen utama baru, sedangkan *eigenvalue* menunjukkan besarnya variansi yang dijelaskan oleh masing-masing komponen utama. Hubungan tersebut dijelaskan melalui persamaan *eigenvalue* berikut:

$$C \cdot v = \lambda \cdot v \quad (4)$$

Di mana C adalah matriks kovarians dari data yang telah dinormalisasi, v merupakan *eigenvektor* dari matriks kovarians, dan λ adalah *eigenvalue* yang bersesuaian.

- Setelah komponen utama diperoleh, dilakukan analisis untuk memahami struktur data yang telah direduksi [23].
- Hasil PCA kemudian divisualisasikan untuk menampilkan sebaran data dalam ruang komponen

utama yang baru, sehingga pola-pola dalam data menjadi lebih mudah dikenali [24].

4. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan proses mulai dari persiapan data hingga analisis komponen utama (Principal Component Analysis/PCA) untuk reduksi dimensi data klimatologis.

4.1. Persiapan dan Standarisasi Data

Variabel penelitian yang digunakan dalam analisis, termasuk variabel curah hujan, disajikan pada Tabel 1.

Tabel 1. Persiapan

Variabel	Keterangan
Location	Titik Pengumpulan Data
Average Temperature	Suhu harian rata-rata maksimum dan minimum
Rainfal	Total curah hujan dalam 24 jam
Rainfall (07.00)	Curah hujan harian tertinggi selama pengamatan
Sunlight	Lama paparan sinar matahari harian
Evaporation	Jumlah air yang menguap dari permukaan

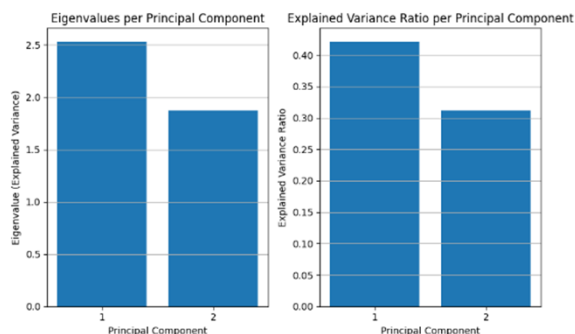
Setelah tahap persiapan data, dilakukan proses standarisasi untuk menyamakan bobot setiap fitur dalam pembentukan Komponen Utama (Principal Components/PCs). Normalisasi menggunakan pustaka StandardScaler pada Python menghasilkan nilai-nilai baru yang siap diproses oleh metode PCA.

4.2. Nilai Rasio Variansi dan Eigenvalue

Python, nilai eigenvalue dan rasio variansi untuk setiap Komponen Utama (Principal Component/PC) telah ditentukan. Hasil perhitungan tersebut disajikan pada Tabel 2 dan divisualisasikan pada Gambar 2.

Tabel 2. Variabel Curah Hujan

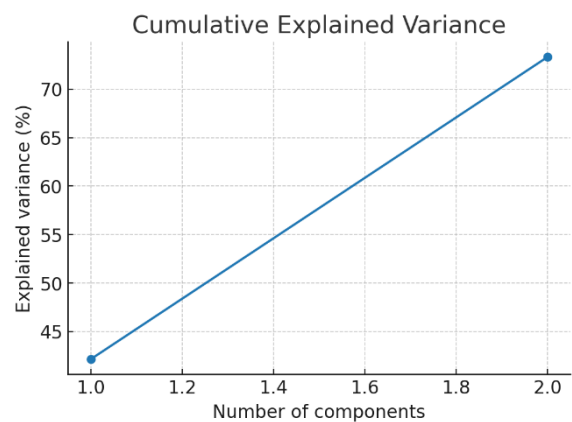
Jumlah komponen	Nilai Eigen	Rasio Varians (%)
1.0	2.53	42.14
2.0	1.87	31.19



Gambar 2. Grafik Eigenvalue dan Rasio Variansi

4.3. Jumlah Komponen Utama (Principal Component/PC)

Penentuan jumlah komponen utama (PCs) didasarkan pada nilai eigenvalue dan rasio variansi. Komponen utama pertama (PC1) memiliki nilai eigenvalue terbesar yaitu 2,53 dan menjelaskan 42,14% dari total variansi. Komponen utama kedua (PC2) memiliki nilai eigenvalue sebesar 1,87 dan menjelaskan 31,19% dari total variansi. Secara kumulatif, seperti yang ditunjukkan pada Gambar 3, kedua komponen utama tersebut mampu menjelaskan 73,33% dari keseluruhan variasi data. Hal ini menunjukkan bahwa dua komponen tersebut sudah cukup untuk mewakili sebagian besar informasi dalam data, sehingga memungkinkan proses reduksi dimensi yang efektif tanpa kehilangan informasi penting secara signifikan.



Gambar 3. Grafik Nilai Rasio Variansi Kumulatif PC1 dan PC2

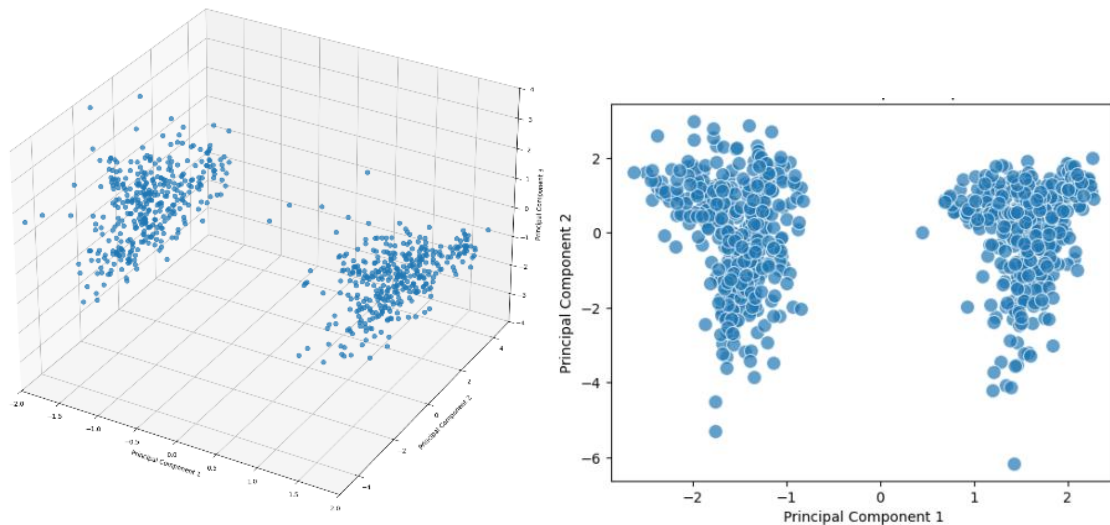
Gambar 2 (Scree Plot) tidak hanya menampilkan besarnya nilai eigenvalue, tetapi juga memberikan dasar visual yang kuat dalam pemilihan dua komponen utama. Kriteria Kaiser-Guttman merekomendasikan untuk mempertahankan komponen dengan nilai eigenvalue lebih besar dari 1 (Tripathi and Dev Garg, 2021). Baik PC1 (2,53) maupun PC2 (1,87) memenuhi kriteria tersebut, sedangkan komponen berikutnya memiliki nilai eigenvalue di bawah 1, yang menunjukkan bahwa komponen tersebut hanya menangkap variansi yang lebih kecil dibandingkan satu variabel asli. Adanya bentuk “elbow” yang jelas pada grafik setelah PC2 menunjukkan bahwa penambahan komponen ketiga dan seterusnya hanya akan memberikan peningkatan variansi total yang kecil dan tidak signifikan. Oleh karena itu, pemilihan dua komponen utama dianggap sebagai keseimbangan optimal antara kesederhanaan model (hanya dua dimensi) dan retensi informasi (73,33%).

4.4. Visualisasi Komponen Utama

Visualisasi Komponen Utama (Principal Components) dilakukan untuk memahami sebaran data pada PC1 dan PC2. Sebaran data yang dihasilkan dapat divisualisasikan dalam bentuk dua dimensi (2D)

maupun tiga dimensi (3D) seperti yang ditunjukkan pada Gambar 4. Berdasarkan gambar tersebut, sebaran data pada PC1 dan PC2 menunjukkan adanya

pengelompokan (clustering) pada titik-titik tertentu, yang mengindikasikan keberadaan struktur atau pola laten dalam data setelah dilakukan reduksi dimensi.



Gambar 4. Visualisasi 3D dan 2D Sebaran Data pada Komponen Utama

4.5. Interpretasi Komponen Utama dan Hubungannya dengan Variabel Iklim

Setelah komponen utama (Principal Components/PCs) berhasil diidentifikasi, langkah penting selanjutnya adalah menginterpretasikan makna fisik dari PC1 dan PC2. Interpretasi ini dilakukan dengan menganalisis faktor pemuatan (loading factor) dari setiap variabel klimatologis asli terhadap kedua komponen utama tersebut. Nilai loading factor yang tinggi (mendekati +1 atau -1) menunjukkan adanya pengaruh kuat dari variabel tersebut dalam pembentukan suatu komponen utama.

Berdasarkan hasil analisis loading factor (yang dapat disajikan dalam tabel tambahan), diketahui bahwa PC1 didominasi oleh variabel Curah Hujan, Curah Hujan (07.00), dan Evaporasi. Hal ini menunjukkan bahwa PC1 secara efektif merepresentasikan dimensi “Kelembapan dan Intensitas Presipitasi.” Nilai loading positif pada ketiga variabel tersebut mengindikasikan bahwa hari-hari dengan curah hujan tinggi cenderung bertepatan dengan nilai evaporasi yang tinggi fenomena yang konsisten dengan kondisi setelah hujan, di mana sinar matahari yang muncul meningkatkan penguapan dari permukaan yang basah [3]. Temuan ini sejalan dengan penelitian yang dilakukan oleh Yan (2024), yang juga menemukan adanya korelasi kuat antara kandungan kelembapan tanah (sebagai proksi kelembapan udara) dan suhu permukaan pada komponen utama pertama mereka [11].

Sementara itu, PC2 menunjukkan loading positif yang sangat kuat pada variabel Lama Penyinaran Matahari dan loading negatif pada variabel Suhu Rata-rata. Konfigurasi ini mengidentifikasi PC2 sebagai dimensi “Kondisi Energi Permukaan dan Keterbukaan Atmosfer.” Hubungan yang berbanding terbalik antara lama penyinaran matahari dan suhu rata-rata pada PC2

mungkin tampak tidak intuitif pada awalnya, tetapi dapat dijelaskan melalui dinamika lokal. Hari-hari dengan lama penyinaran matahari yang sangat tinggi di wilayah Malang yang umumnya terjadi pada musim kemarau dapat secara paradoks memiliki suhu rata-rata yang lebih rendah dibandingkan hari-hari berawan pada musim peralihan. Hal ini disebabkan oleh tutupan awan pada siang hari yang dapat menahan panas dan mencegah penurunan suhu signifikan pada malam hari [2]. Selain itu, topografi Karangploso yang bervariasi dan berbukit-bukit dapat menciptakan mikroklimat, di mana suhu permukaan tidak selalu berkorelasi secara linear dengan lama penyinaran matahari. Hasil ini memperkuat temuan Geng (2022) mengenai pentingnya menganalisis hubungan non-linear antarvariabel cuaca sebelum melakukan reduksi dimensi [21].

4.6. Implikasi Praktis dari Pengurangan Dimensionalitas

Keberhasilan dalam mereduksi enam variabel menjadi dua komponen utama (Principal Components/PCs) memiliki implikasi praktis yang signifikan bagi penelitian klimatologi dan peramalan cuaca. Pertama, penyederhanaan data ini sangat bermanfaat untuk proses pemodelan prediktif selanjutnya. Algoritma pembelajaran mesin seperti Support Vector Machine (SVM) atau Long Short-Term Memory (LSTM) untuk prediksi curah hujan dapat dilatih dengan lebih efisien dan cepat menggunakan dua fitur yang tidak berkorelasi (PC1 dan PC2), dibandingkan dengan enam fitur asli yang kemungkinan memiliki tingkat multikolinearitas tinggi [6], [10]. Kedua, visualisasi 2D dan 3D dari kedua komponen utama (Gambar 4) memungkinkan analisis untuk secara visual mengidentifikasi pola atau kluster tertentu. Misalnya, pengelompokan titik pada

kuadran tertentu dapat merepresentasikan hari-hari dengan karakteristik musim hujan ekstrem, musim kemarau, atau periode peralihan. Mengidentifikasi pola-pola tersebut secara manual dari data berdimensi enam hampir tidak mungkin dilakukan. Kemampuan PCA untuk “meringkas” data kompleks menjadi bentuk yang dapat divisualisasikan merupakan kontribusi metodologis yang penting, sebagaimana ditekankan oleh Susanti (2023) dalam konteks pemetaan kerentanan penyakit [24].

Meskipun hasil penelitian ini sangat menjanjikan, terdapat beberapa keterbatasan yang perlu diakui. Pertama, analisis hanya mencakup data dari dua stasiun pengamatan dalam satu tahun (2023). Untuk memperoleh pemahaman yang lebih kuat dan bersifat umum mengenai pola curah hujan di wilayah Malang, penelitian selanjutnya sebaiknya memperluas cakupan data secara spasial (menambah jumlah stasiun) dan temporal (menggunakan data deret waktu 10–30 tahun). Data jangka panjang memungkinkan analisis terhadap variasi antar-tahun serta pengaruh fenomena skala besar seperti ENSO terhadap komponen utama curah hujan [25], [26]. Kedua, PCA merupakan metode linear, sedangkan hubungan antarvariabel klimatologis sering kali bersifat nonlinear. Untuk menangani kompleksitas tersebut, dapat digunakan teknik reduksi dimensi nonlinear seperti Kernel PCA atau t-Distributed Stochastic Neighbor Embedding (t-SNE), yang kemudian dibandingkan kinerjanya dengan PCA linear [21]. Ketiga, interpretasi terhadap komponen utama bersifat statistik dan memerlukan validasi serta pemahaman klimatologis dari para ahli di bidang terkait [4]. Penelitian selanjutnya dapat mengintegrasikan hasil PCA dengan analisis cuaca sinoptik untuk memberikan interpretasi fisik yang lebih komprehensif terhadap PC1 dan PC2.

Dengan mempertimbangkan berbagai implikasi dan keterbatasan tersebut, dapat disimpulkan bahwa PCA terbukti menjadi langkah awal (preprocessing) yang sangat kuat dan penting dalam rangkaian analisis data klimatologis modern. Dua komponen utama yang dihasilkan tidak hanya berfungsi sebagai representasi ringkas data, tetapi juga mengandung makna klimatologis yang aplikatif untuk berbagai bidang, termasuk pertanian, pengelolaan sumber daya air, dan kesiapsiagaan bencana.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa Principal Component Analysis (PCA) merupakan pendekatan yang efektif untuk menganalisis data curah hujan, dengan berhasil mereduksi enam variabel klimatologis menjadi dua komponen utama (PC1 dan PC2). Kedua komponen ini secara bersama-sama mampu menjelaskan 73,33% dari total varians data, yang membuktikan bahwa PCA mampu menyederhanakan dataset kompleks tanpa kehilangan informasi yang signifikan. Hasil visualisasi menunjukkan adanya distribusi data yang terstruktur dalam ruang komponen

utama, mengonfirmasi bahwa PCA tidak hanya berfungsi sebagai teknik reduksi dimensi, tetapi juga memberikan dasar yang kuat untuk analisis pola selanjutnya. Namun demikian, penelitian ini masih memiliki keterbatasan dalam jumlah data, yang hanya bersumber dari dua stasiun curah hujan selama satu tahun pengamatan. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan data secara spasial dan temporal, serta menambahkan variabel klimatologis tambahan guna memperoleh representasi yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] D. Bhattacharyya and U. Saha, “Deep learning application for disaggregation of rainfall with emphasis on preservation of extreme rainfall characteristics for Indian monsoon conditions,” *Stoch. Environ. Res. Risk Assess.*, vol. 37, no. 3, pp. 1021–1038, 2023, doi: 10.1007/s00477-022-02331-x.
- [2] B. Irawan, Wirawan, B. A. Ikawanty, and A. Takwim, “Analysis of the Season Effect on Energy Generated From Hybrid Pv/Wt in Malang Indonesia,” *Eastern-European J. Enterp. Technol.*, vol. 5, no. 8–119, pp. 70–78, 2022, doi: 10.15587/1729-4061.2022.266082.
- [3] A. Suharyanto, A. Maulana, D. Suprayogo, Y. P. Devia, and S. Kurniawan, “Land surface temperature changes caused by land cover/ land use properties and their impact on rainfall characteristics,” *Glob. J. Environ. Sci. Manag.*, vol. 9, no. 3, pp. 353–372, 2023, doi: 10.22035/gjesm.2023.03.01.
- [4] Yulianingsani *et al.*, “Spatio-temporal of climate extreme rainfall variability over the Nusantara Capital City of Indonesia (IKN),” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1490, no. 1, p. 012050, Apr. 2025, doi: 10.1088/1755-1315/1490/1/012050.
- [5] A. Anshuka, R. Chandra, A. J. V. Buzacott, D. Sanderson, and F. F. van Ogtrop, “Spatio temporal hydrological extreme forecasting framework using LSTM deep learning model,” *Stoch. Environ. Res. Risk Assess.*, vol. 36, no. 10, pp. 3467–3485, 2022, doi: 10.1007/s00477-022-02204-3.
- [6] M. Billah, M. N. Adnan, M. R. Akhond, R. R. Ema, M. A. Hossain, and S. Md. Galib, “Rainfall prediction system for Bangladesh using long short-term memory,” *Open Comput. Sci.*, vol. 12, no. 1, pp. 323–331, 2022, doi: 10.1515/comp-2022-0254.
- [7] B. Suprpty, A. S. Wiguna, A. Wajiansyah, and R. Malani, “Dataset transformation using hybrid method of polar-based cartesian and image filtering technique for annual rainfall clustering,” in *2021 I tations (ISITIA)*, 2021, pp. 210–215. doi: 10.1109/ISITIA52817.2021.9502202.
- [8] A. T. W. Almais *et al.*, “Principal Component Analysis-Based Data Clustering for Labeling of

- Level Damage Sector in Post-Natural Disasters,” *IEEE Access*, vol. 11, pp. 74590–74601, 2023, doi: 10.1109/ACCESS.2023.3275852.
- [9] A. T. W. Almaïs *et al.*, “Characterization of Structural Building Damage in Post-Disaster using GLCM-PCA Analysis Integration,” *IEEE Access*, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3469637.
- [10] M. M. Hassan *et al.*, “Machine Learning-Based Rainfall Prediction: Unveiling Insights and Forecasting for Improved Preparedness,” *IEEE Access*, vol. 11, no. September, pp. 132196–132222, 2023, doi: 10.1109/ACCESS.2023.3333876.
- [11] Y. Yan, X. Le, T. Yang, and H. Yu, “Interpretable PCA and SVM-Based Leak Detection Algorithm for Identifying Water Leakage Using SAR-Derived Moisture Content and InSAR Closure Phase,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 15136–15147, 2024, doi: 10.1109/JSTARS.2024.3443127.
- [12] A. T. W. Almaïs *et al.*, “Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters,” *IEEE Access*, vol. 11, no. March, pp. 74590–74601, 2023, doi: 10.1109/ACCESS.2023.3275852.
- [13] C. W. W. Ng, M. Usman, and H. Guo, “Spatiotemporal pore-water pressure prediction using multi-input long short-term memory,” *Eng. Geol.*, vol. 322, 2023, doi: 10.1016/j.enggeo.2023.107194.
- [14] M. A. T. Mpanza and N. F. Tandon, “Further probing the mechanisms driving projected decreases of extreme precipitation intensity over the subtropical Atlantic,” *Clim. Dyn.*, vol. 59, no. 11–12, pp. 3317–3341, 2022, doi: 10.1007/s00382-022-06268-3.
- [15] 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 J. Ilmu Mat. dan Terap.*, vol. 18, no. 2, pp. 0837–0848, 2024, doi: 10.30598/barekengvol18iss2pp0837-0848.
- [16] T. Hendrawati, A. H. Wigena, I. M. Sumertajaya, B. Sartono, A. A. Pravitasari, and M. H. Asnawi, “The ensemble distance on model-based clustering for regions clustering based on rainfall: The case of rainfall in West Java Indonesia,” *Int. J. Data Netw. Sci.*, vol. 8, no. 2, pp. 1187–1196, 2024, doi: 10.5267/j.ijdns.2023.11.015.
- [17] A. O. Lima, G. B. Lyra, M. C. Abreu, J. F. Oliveira-Júnior, M. Zeri, and G. Cunha-Zeri, “Extreme rainfall events over Rio de Janeiro State, Brazil: Characterization using probability distribution functions and clustering analysis,” *Atmos. Res.*, vol. 247, 2021, doi: 10.1016/j.atmosres.2020.105221.
- [18] C. S. de Almeida *et al.*, *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title*, vol. 5, no. 1. 2016. [Online]. Available: <https://revistas.ufrj.br/index.php/rce/article/download/1659/1508%0Ahttp://hipatiapress.com/hpjournals/index.php/qre/article/view/1348%5Chttp://www.tandfonline.com/doi/abs/10.1080/09500799708666915%5Chttps://mckinseyonsociety.com/downloads/reports/Educa>
- [19] G. Urgilés, R. Céleri, J. Bendix, and J. Orellana-Alvear, “Identification of spatio-temporal patterns in extreme rainfall events in the Tropical Andes: A clustering analysis approach,” *Meteorol. Appl.*, vol. 31, no. 5, 2024, doi: 10.1002/met.70005.
- [20] L.-H. Hsu *et al.*, “Analysis of the Interdecadal and Interannual Variability of Autumn Extreme Rainfall in Taiwan Using a Deep-Learning-Based Weather Typing Approach,” *Asia-Pacific J. Atmos. Sci.*, vol. 59, no. 2, pp. 185–205, 2023, doi: 10.1007/s13143-022-00303-3.
- [21] Z. Geng, X. Duan, Y. Han, F. Liu, and W. Xu, “Novel variation mode decomposition integrated adaptive sparse principal component analysis and its application in fault diagnosis,” *ISA Trans.*, vol. 128, pp. 21–31, 2022, doi: <https://doi.org/10.1016/j.isatra.2021.11.002>.
- [22] T. Gu, “Detection of Small Floating Targets on the Sea Surface Based on Multi-Features and Principal Component Analysis,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 809–813, 2020, doi: 10.1109/LGRS.2019.2935262.
- [23] Erlin, A. Yuniarta, L. A. Wulandhari, Y. Desnelita, N. Nasution, and Junadhi, “Enhancing Rice Production Prediction in Indonesia Using Advanced Machine Learning Models,” *IEEE Access*, vol. 12, pp. 151161–151177, 2024, doi: 10.1109/ACCESS.2024.3478738.
- [24] T. Susanti, H. Susetya, P. Widayani, Y. Fitria, and G. T. Pambudi, “Risk factors, logistic model, and vulnerability mapping of lumpy skin disease in livestock at the farm level in Indragiri Hulu District, Riau Province, Indonesia, in 2022,” *Vet. World*, vol. 16, no. 10, pp. 2071–2079, Oct. 2023, doi: 10.14202/vetworld.2023.2071-2079.
- [25] Mislan and A. T. R. Dani, “Navigating Samarinda’s climate: A comparative analysis of rainfall forecasting models,” *MethodsX*, vol. 14, p. 103080, Jun. 2025, doi: 10.1016/j.mex.2024.103080.
- [26] D. Y. Primartono, R. Hidayat, Supari, R. Prasetya, and M. Taufik, “A spatially distributed rainfall dataset for West Java, Indonesia,” *Data Br.*, vol. 62, p. 111974, Oct. 2025, doi: 10.1016/j.dib.2025.111974.