

***CLUSTERING POLA SPASIAL-TEMPORAL DATA PETIR MENGGUNAKAN
PRINCIPAL COMPONENT ANALYSIS (PCA)***

SKRIPSI

Oleh :

LAILATUL FADHILLA NUR HIDAYATULLAH
NIM. 220605110002



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

***CLUSTERING POLA SPASIAL-TEMPORAL DATA PETIR
MENGUNAKAN *PRINCIPAL COMPONENT ANALYSIS* (PCA)***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
LAILATUL FADHILLA NUR HIDAYATULLAH
NIM. 220605110002

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

***CLUSTERING POLA SPASIAL-TEMPORAL DATA PETIR
MENGUNAKAN PRINCIPAL COMPONENT ANALYSIS (PCA)***

SKRIPSI

Oleh:

LAILATUL FADHILLA NUR HIDAYATULLAH
NIM. 220605110002

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 16 April 2026

Pembimbing I,

Pembimbing II,

Dr. Agung Teguh Wibowo Almais, M.T
NIP. 19860301 202321 1 016

Dr. Cahyo Crysdian, M.Cs
NIP. 19740424 200901 1 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

**CLUSTERING POLA SPASIAL-TEMPORAL DATA PETIR
MENGUNAKAN *PRINCIPAL COMPONENT ANALYSIS* (PCA)**

SKRIPSI

Oleh:

LAILATUL FADHILLA NUR HIDAYATULLAH
NIM. 220605110002

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 16 April 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. Totok Chamidy, M.Kom</u> NIP. 19691222 200604 1 001	()
Anggota Penguji I	: <u>Ahmad Fahmi Karami, M.Kom</u> NIP. 19870909 202012 1 001	()
Anggota Penguji II	: <u>Dr. Agung Teguh Wibowo Almais, M.T</u> NIP. 19860301 202321 1 016	()
Anggota Penguji III	: <u>Dr. Cahyo Crysdiyan, M.Cs</u> NIP. 19740424 200901 1 008	()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Lailatul Fadhilla Nur Hidayatullah
NIM : 220605110002
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : *Clustering* Pola Spasial-Temporal Data Petir
menggunakan *Principal Component Analysis* (PCA)

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 9 April 2026
Yang membuat pernyataan,



Lailatul Fadhilla Nur Hidayatullah
NIM. 220605110002

MOTTO

"Allah knows how hard I tried, and that is enough."

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Clustering Pola Spasial-Temporal Data Petir Menggunakan Principal Component Analysis (PCA)*” dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang terang benderang, yakni agama Islam, semoga kita termasuk golongan yang mendapatkan syafaat beliau di hari kiamat kelak. Aamiin.

Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Agung Teguh Wibowo Almais, M.T., selaku dosen pembimbing I yang telah memberikan bimbingan dengan penuh kesabaran, ketulusan, dan perhatian kepada penulis. Terima kasih atas segala arahan, masukan,

motivasi, serta dukungan yang sangat berarti dalam proses penyusunan hingga penyelesaian skripsi ini.

5. Dr. Totok Chamidy, M.Kom, selaku dosen pembimbing II yang telah memberikan arahan, masukan, dan bimbingan kepada penulis dalam proses penyusunan skripsi ini.
6. Dr. Totok Chamidy, M.Kom selaku Penguji I dan Ahmad Fahmi Karami, M.Kom selaku Penguji II yang telah memberikan kritik, saran, dan masukan yang membangun dalam pengujian serta penyempurnaan skripsi ini.
7. Segenap dosen, laboran, dan staf Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan ilmu, dukungan, serta bantuan kepada penulis selama masa studi hingga selesainya skripsi ini.
8. Kedua orang tua tercinta, serta adik-adik penulis, kakak, serta keluarga terdekat yang senantiasa memberikan doa, dukungan, semangat, dan kasih sayang yang tiada henti kepada penulis selama proses penyusunan hingga selesainya skripsi ini.
9. Teman-teman kost Griya Kost SKD 5D, khususnya Mbak Rere, Refy, dan Ines. Terima kasih selalu memberikan motivasi, dukungan serta kebersamaan selama ini.
10. Teman-teman bimbingan dan magang, yaitu Intan, Fina, Faiza, Arum, dan Renata, yang telah memberikan dukungan dan kebersamaan selama proses

penyusunan skripsi ini, khususnya Intan yang selalu hadir menemani penulis di saat dibutuhkan.

11. Teman-teman KKM Arunika, khususnya Meilinda, yang telah memberikan dukungan, kebersamaan, dan menjadi tempat berbagi cerita bagi penulis, meskipun hubungan pertemanan yang terjalin masih relatif baru.
12. Teman-teman pondok, khususnya Bela, Khadroh, Izza, Anggun, dan lainnya, yang telah menjaga kebersamaan dan pertemanan penulis sejak masa pondok hingga perkuliahan.
13. Keluarga besar “Infinity” Teknik Informatika Angkatan 2022 yang telah saling membantu, mengingatkan, serta berbagi ilmu dan pengalaman selama masa perkuliahan.
14. Seseorang berinisial J.A. yang senantiasa hadir sebagai tempat pulang bagi penulis dalam setiap lelah dan keluh, yang tidak pernah absen memberikan dukungan, menguatkan di saat sulit, serta menghadirkan ketenangan di tengah berbagai proses yang dilalui. Kehadirannya menjadi bagian penting dalam perjalanan penulis hingga skripsi ini dapat terselesaikan.
15. Diri penulis sendiri yang telah berjuang dalam diam, bertahan di tengah berbagai kesulitan, dan tidak menyerah hingga akhirnya mampu menyelesaikan skripsi ini.

Akhir kata, penulis menyadari bahwa skripsi ini masih belum sempurna. Oleh karena itu, penulis dengan terbuka menerima kritik dan saran yang membangun. Penulis berharap skripsi ini dapat memberikan manfaat bagi berbagai pihak serta berkontribusi positif dalam pengembangan ilmu pengetahuan di bidang terkait.

Malang, 9 April 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
ABSTRAK	xvi
ABSTRACT	xvii
المخلص	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
BAB II STUDI PUSTAKA	8
2.1 Principal Component Analysis (PCA)	12
2.2 PCA-Clustering	13
2.3 Petir	14
2.4 Spasial dan Temporal	16
BAB III DESAIN DAN IMPLEMENTASI SISTEM	18
3.1 Desain Sistem	18
3.1.1 Input	18
3.1.2 Rekayasa Fitur (<i>Feature Engineering</i>)	18
3.1.3 Preprocessing	20
3.1.4 Main Process.....	20
3.1.5 Output	28
3.2 Pseudocode Clustering menggunakan PCA-Clustering	28
BAB IV UJI COBA DAN PEMBAHASAN	35
4.1 Desain Uji Coba	35
4.2 <i>Preparation Data</i>	36
4.2.1 Teknik Pengumpulan Data.....	36
4.2.2 <i>Cleaning Data</i>	38
4.2.3 Rekayasa Fitur Temporal.....	40
4.3 Normalisasi Data	42
4.3.1 <i>Standard Scaler</i>	43
4.4 Hasil Analisis <i>Principal Component Analysis</i> (PCA).....	45
4.4.1 Menentukan Nilai Eigen dan Rasio Varians.....	46
4.4.2 Penentuan Komponen Utama Berdasarkan <i>Faktor Loading</i>	49
4.4.3 Karakterisasi Komponen Utama Berdasarkan <i>Faktor Loading</i>	52
4.5 Visualisasi Data Hasil PCA.....	55
4.5.1 Visualisasi PCA Dua Dimensi (PC1 Vs PC2).....	55

4.5.2 Visualisasi PCA Tiga Dimensi	57
4.6 Hasil <i>Clustering</i> Spasial Temporal Petir Berbasis PCA.....	59
4.6.1 Hasil <i>Clustering</i> Data Petir	59
4.7 Evaluasi Hasil PCA- <i>Clustering</i> Kejadian Petir	67
4.7.1 Evaluasi Internal Menggunakan Silhouette Score	67
4.7.2 Evaluasi Berdasarkan Perspektif BMKG	70
4.8 Pembahasan	71
BAB V KESIMPULAN	77
5.1 Kesimpulan.....	77
5.2 Saran	78
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3.1 Desain Sistem.....	18
Gambar 3.2 Flowchart Main Process Sistem PCA-Clustering	21
Gambar 4.1 Flowchart Desain Uji Coba PCA-Clustering	35
Gambar 4.2 Visualisasi Sebaran Data Petir pada PCA 2D	56
Gambar 4.3 Visualisasi Sebaran Data Petir pada PCA 3D	58
Gambar 4.4 Visualisasi Hasil Clustering	61

DAFTAR TABEL

Tabel 2.1 Rangkuman Penelitian Terkait.....	11
Tabel 3.1 Data Petir.....	23
Tabel 3.2 Pseudocode Algoritma Clustering	29
Tabel 4.1 Data Petir 29 Mei 2025	37
Tabel 4.2 Contoh Data Sebelum Cleaning.....	39
Tabel 4.3 Contoh Data Setelah Cleaning.....	39
Tabel 4.4 Data Waktu Kejadian Petir Sebelum Rekayasa Fitur	41
Tabel 4.5 Hasil Rekayasa Fitur Temporal Data Petir	42
Tabel 4.6 Hasil Normalisasi Data	45
Tabel 4.7 Nilai Eigen dan Rasio Varians Hasil PCA.....	47
Tabel 4.8 Faktor Loading Variabel terhadap Komponen Utama.....	50
Tabel 4.9 Karakterisasi Komponen Utama Hasil PCA.....	53
Tabel 4.10 Distribusi Jumlah Data pada Setiap Cluster.....	60
Tabel 4.11 Cuplikan Data Cluster 1 dan Identifikasi Wilayah	63
Tabel 4.12 Cuplikan Data Cluster 2 dan Identifikasi Wilayah	64
Tabel 4.13 Cuplikan Data Cluster 3 dan Identifikasi Wilayah	65
Tabel 4.14 Rekapitulasi Hasil Clustering Harian Bulan Mei 2025.....	72

ABSTRAK

Hidayatullah, Lailatul Fadhilla Nur. 2026. ***Clustering Spasial-Temporal Data Petir Menggunakan Principal Component Analysis (PCA)***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Agung Teguh Wibowo Almais, M.T (II) Dr. Cahyo Crysdiyan, M.Cs.

Kata kunci: Data Petir, Spasial Temporal, *Clustering*.

Data petir memiliki karakteristik spasial dan temporal yang kompleks, sehingga menyulitkan proses analisis dalam memahami pola sebaran dan pengelompokan kejadian petir. Oleh karena itu, diperlukan metode yang mampu mengolah data berdimensi tinggi serta mengidentifikasi pola secara efektif. Penelitian ini bertujuan untuk melakukan clustering spasial temporal pada data petir dengan memanfaatkan metode *Principal Component Analysis (PCA)* untuk reduksi dimensi data. Proses penelitian meliputi tahap pengumpulan data petir, preprocessing data, transformasi fitur spasial dan temporal menggunakan PCA, serta proses *clustering* untuk mengelompokkan data berdasarkan kemiripan karakteristik. Evaluasi hasil clustering dilakukan menggunakan *Silhouette Score* untuk mengukur kualitas pengelompokan berdasarkan tingkat kedekatan data dalam satu cluster dan tingkat pemisahan antar *cluster*. Berdasarkan hasil pengujian yang dilakukan, penerapan PCA mampu mereduksi dimensi data tanpa menghilangkan informasi penting, sehingga meningkatkan efisiensi dan kualitas proses *clustering*. Hasil evaluasi menunjukkan bahwa model *clustering* yang dihasilkan memiliki nilai *Silhouette Score* yang menunjukkan kualitas pengelompokan yang baik. Dengan demikian, metode PCA dan *clustering* dapat digunakan untuk membantu dalam memahami pola spasial temporal data petir secara lebih efektif.

ABSTRACT

Hidayatullah, Lailatul Fadhilla Nur. 2026. *Spatio-Temporal Clustering of Lightning Data Using Principal Component Analysis (PCA)*. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Agung Teguh Wibowo Almais, M.T (II) Dr. Cahyo Crysodian, M.Cs.

Lightning data exhibit complex spatio-temporal characteristics, which make the analysis process challenging in understanding the distribution patterns and grouping of lightning events. Therefore, methods capable of handling high-dimensional data and effectively identifying patterns are required. This study aims to perform spatio-temporal clustering on lightning data by utilizing Principal Component Analysis (PCA) for dimensionality reduction. The research process includes data collection, data preprocessing, transformation of spatial and temporal features using PCA, and clustering to group data based on similarity characteristics. The clustering results are evaluated using the Silhouette Score to measure the quality of clustering based on intra-cluster cohesion and inter-cluster separation. Based on the experimental results, the application of PCA is able to reduce data dimensionality without losing important information, thereby improving the efficiency and quality of the clustering process. The evaluation results indicate that the clustering model achieves a Silhouette Score that reflects good clustering quality. Therefore, the PCA and clustering methods can be effectively used to better understand the spatio-temporal patterns of lightning data.

Keywords: Lightning Data, Spatio-Temporal, Clustering.

الملخص

هداية الله، ليلة الفاضلة نور. ٢٠٢٦. التجميع الزمني المكاني لبيانات البرق باستخدام تحليل المركبات الرئيسية (PCA) بحث تخرج. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: أ (الدكتور أغونغ تيغوه ويوو ألاميس، ماجستير تقني) ب(الدكتور كاهيو كريسيديان، ماجستير علوم الحاسوب).

الكلمات المفتاحية: بيانات البرق، الزمني المكاني، التجميع.

تتميز بيانات البرق بخصائص زمانية ومكانية معقدة، مما يجعل عملية تحليلها صعبة في فهم أنماط التوزيع وتجميع أحداث البرق. لذلك، هناك حاجة إلى أساليب قادرة على معالجة البيانات عالية الأبعاد وتحديد الأنماط بشكل فعال. تهدف هذه الدراسة إلى إجراء عملية التجميع الزمني المكاني لبيانات البرق باستخدام أسلوب تحليل المركبات الرئيسية (PCA) لتقليل أبعاد البيانات. تشمل عملية البحث جمع بيانات البرق، ومعالجة البيانات الأولية، وتحويل الخصائص الزمانية والمكانية باستخدام PCA، ثم إجراء عملية التجميع لتصنيف البيانات بناءً على تشابه الخصائص. كما تم تقييم نتائج التجميع باستخدام مقياس *Silhouette Score* لقياس جودة التجميع من حيث التقارب داخل المجموعة والتباعد بين المجموعات. ناءً على نتائج الاختبارات، أظهر تطبيق PCA قدرته على تقليل أبعاد البيانات دون فقدان المعلومات المهمة، مما يساهم في تحسين كفاءة وجودة عملية التجميع. كما أظهرت نتائج التقييم أن نموذج التجميع يحقق قيمة *Silhouette Score* تدل على جودة جيدة للتجميع. وبالتالي، يمكن استخدام أسلوب PCA والتجميع لفهم الأنماط الزمانية المكانية لبيانات البرق بشكل أكثر فعالية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Petir merupakan salah satu fenomena alam yang sangat berbahaya, karena pada setiap kali terjadi sambaran, petir bisa melepaskan energi listrik dalam jumlah yang sangat besar, yang mana nilainya dapat mencapai jutaan volt (Susanto, 2018). Energi pada saat sambaran petir (baik secara langsung maupun tidak langsung) dapat menimbulkan beberapa kerusakan pada infrastruktur dan bahkan membahayakan keselamatan manusia (Wahjudi, 2014). Sambaran petir secara langsung merupakan sambaran yang langsung mengenai benda atau objek pada permukaan tanah (Patel, 2013). Sementara itu, sambaran tidak langsung merupakan sambaran yang terjadi karena induksi elektromagnetik (Zoro, 2010). Maka dari itu petir tidak sebatas fenomena alam biasa tetapi juga sebagai ancaman yang perlu dipahami secara ilmiah (Gunawan *et al.*, 2014).

Selain memiliki daya rusak yang besar, petir memiliki empat jenis pelepasan muatan listrik berdasarkan karakteristik sambarannya, yaitu sambaran dalam satu awan (*Intracloud/IC*), sambaran antar awan (*Cloud to Cloud/CC*), sambaran dari awan ke udara (*Cloud to Air/CA*), serta sambaran dari awan ke tanah (*Cloud to Ground/CG*) (Hidayat *et al.*, 2018). Dari keempat jenis petir tersebut, terdapat yang membahayakan yaitu (*Cloud to Ground/CG*), karena dapat memberikan dampak langsung terhadap aktivitas manusia dan lingkungan (Narut *et al.*, 2018).

Selain terbagi dalam empat jenis berdasarkan arah pelepasan muatannya, petir juga memiliki pola spasial temporal yang menggambarkan sebaran dan waktu

kejadian sambaran. Secara spasial distribusi sambaran petir CG terkonsentrasi pada wilayah tertentu seperti pada radius 50 km dari sensor deteksi (Pratama & Negara, 2017). Sementara itu, secara temporal, aktivitas petir cenderung meningkat ketika musim hujan dengan pola semidiurnal, yakni dua puncak pada sore hari dan dini hari, dan pada musim kemarau pola tersebut menjadi acak (Pratama & Negara, 2017). Menurut Yusnaini *et al.* (2021), menjelaskan bahwa tingkat kepadatan petir tidak hanya bervariasi menurut musim, tetapi juga oleh topografi yang ada pada area perbukitan atau dataran tinggi yang dapat memperlihatkan konsentrasi petir lebih tinggi (Yusnaini *et al.*, 2021).

Kondisi tersebut relevan dengan penelitian yang dilakukan pada wilayah kerja BMKG Stasiun Geofisika Kelas III Malang, yang mana menganalisis distribusi spasial dan temporal sambaran petir untuk melihat kecenderungan pola kejadian berdasarkan waktu serta posisi geografisnya. Dengan memahami variasi spasial-temporal tersebut, penelitian ini dapat memberikan gambaran lebih rinci mengenai karakteristik aktivitas petir pada wilayah Malang dan sekitarnya, yang berpotensi dimanfaatkan dalam analisis mitigasi dan sistem peringatan dini.

Fenomena petir juga memiliki makna dalam perspektif Islam. Sebagaimana firman Allah SWT dalam Al-Qur'an surat Ar-Ra'd [13] ayat 12-13:

هُوَ الَّذِي يُرِيكُمُ الْبَرْقَ خَوْفًا وَطَمَعًا وَيُنزِلُ السَّحَابَ الثِّقَالَ ۚ وَيُسَبِّحُ الرَّعْدُ بِحَمْدِهِ وَالْمَلَائِكَةُ مِنْ خِيفَتِهِ
وَيُرْسِلُ الصَّوَاعِقَ فَيُصِيبُ بِهَا مَنْ يَشَاءُ وَهُمْ يُجَادِلُونَ فِي اللَّهِ وَهُوَ شَدِيدُ الْمِحَالِ.

“Dialah yang memperlihatkan kilat kepadamu untuk menimbulkan rasa takut dan harapan, dan Dia menumbuhkan awan yang berat. Dan guruh bertasbih dengan memuji-Nya, demikian pula para malaikat karena takut kepada-Nya; dan Allah mengirimkan kilat, lalu mengenai siapa yang Dia kehendaki, sementara mereka

berbantah-bantahan tentang Allah; dan Dia Mahakuat lagi Mahabesar.” (QS. Ar-Ra’d: 12–13).

Menurut Tafsir Ibnu Katsir, ayat ini menjelaskan bahwa petir memiliki dua sisi makna yaitu, menimbulkan rasa takut karena dapat mendatangkan bahaya berupa sambaran yang berpotensi merusak, sekaligus menumbuhkan harapan karena kilat biasanya menandakan datangnya hujan yang membawa rahmat. Guruh bahkan turut bertasbih memuji Allah, serta para malaikat yang tunduk karena takut kepada-Nya. Hal ini menunjukkan bahwa seluruh fenomena alam sejatinya berjalan atas kehendak Allah (Fikri Ys & Taufiq, 2023).

Hal ini memberikan pemahaman bahwa petir bukan hanya fenomena alam secara ilmiah, tetapi juga tanda kebesaran Allah yang mengingatkan manusia agar selalu bersyukur dan berhati-hati. Penelitian ini, yaitu penggunaan metode *Principal Component Analysis (PCA) Clustering* pada petir merupakan bentuk pemanfaatan akal dan ilmu pengetahuan yang telah Allah anugerahkan untuk kebaikan manusia. Dengan menggabungkan analisis spasial-temporal sambaran petir dengan pendekatan ilmiah, penelitian ini dapat berkontribusi pada mitigasi bencana pengembangan sistem peringatan ini, serta perlindungan infrastruktur dan keselamatan manusia. Dalam perspektif islam, upaya tersebut juga merupakan bagian dari mendengarkan amanah untuk menjaga kehidupan dan lingkungan sebagaimana yang telah tercantum dalam Alquran (Akhyar & AN, 2020).

Beberapa penelitian terkait pemanfaatan metode analisis data dalam konteks petir telah banyak dilakukan oleh peneliti - peneliti sebelumnya. Nicora *et al.* (2024) menggunakan *Principal Component Analysis (PCA)* untuk ekstraksi fitur dari sinyal *lightning-induced voltage* pada saluran transmisi sebelum diproses

dengan neural network untuk memperkirakan lokasi sambaran dan arus puncak petir (Nicora *et al.*, 2024). Penelitian lainnya oleh Lovytaji *et al.* (2024) menggunakan PCA dengan Ellipse Fitting Method untuk menganalisis tingkat kerawanan sambaran petir di wilayah Pasuruan (Lovytaji *et al.*, 2024). Pada penelitian-penelitian yang sudah ada tidak ada yang membahas untuk mengcluster petir, sehingga penelitian ini bertujuan untuk meng-*cluster* petir menggunakan PCA berbasis data petir dari BMKG Stasiun Geofisika Kelas III Malang. Pada penelitian Almais *et al.* (2023), menjelaskan tentang penerapan PCA untuk mengcluster tingkat kerusakan bangunan pasca bencana alam (Almais *et al.*, 2023).

Adapun untuk penelitian ini berfokus pada penerapan *PCA-Clustering* dalam analisis pola spasial-temporal data petir. Pemilihan metode ini didasarkan pada karakteristik data petir yang memiliki banyak variabel dengan skala yang berbeda, seperti jumlah sambaran (*strike*), waktu kejadian (*epoch*), serta koordinat spasial (*longitude dan latitude*). Penerapan PCA memungkinkan pereduksian dimensi data yang kompleks menjadi lebih sederhana tanpa kehilangan informasi penting, sehingga hubungan antarvariabel dapat mudah dianalisis.

Selanjutnya, proses clustering dilakukan berdasarkan hasil reduksi PCA untuk mengelompokkan data yang memiliki kesamaan karakteristik spasial dan temporal secara tidak terawasi (*unsupervised*). Dengan pendekatan ini, pola kejadian petir dapat teridentifikasi secara lebih objektif dan efisien dibandingkan metode lain yang memerlukan label atau parameter awal tertentu (Gul & Rehman, 2023).

Kontribusi utama penelitian ini adalah untuk mengembangkan penerapan *PCA-Clustering* dalam mengidentifikasi pola distribusi aktivitas petir berbasis data historis. Kebaruan penelitian terletak pada pemanfaatan hasil PCA sebagai dasar pengelompokan data spasial-temporal tanpa algoritma tambahan, sebagaimana pendekatan yang diperkenalkan oleh Almais *et al* (2023). Hasil penelitian diharapkan dapat mendukung upaya mitigasi bahaya petir, membantu pengembangan sistem peringatan dini, serta menjadi landasan untuk penelitian lanjutan dalam pemodelan data spasial-temporal berbasis PCA.

1.2 Pernyataan Masalah

Berdasarkan uraian pada latar belakang, maka permasalahan yang menjadi fokus dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan *Principal Component Analysis* (PCA) untuk mereduksi dimensi dan menganalisis pola spasial-temporal petir?
2. Bagaimana hasil pengelompokan data petir yang diperoleh melalui proses *PCA-Clustering*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, penelitian ini memiliki tujuan sebagai berikut:

1. Menerapkan *Principal Component Analysis* (PCA) untuk mereduksi dimensi dan menganalisis pola spasial-temporal data petir.
2. Mengidentifikasi pola pengelompokan kejadian petir berbasis spasial-temporal.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memiliki manfaat sebagai berikut:

1. Bagi BMKG, hasil analisis pola spasial-temporal petir dengan *PCA-Clustering* dapat mendukung peningkatan akurasi sistem peringatan dini, serta membantu dalam pengelolaan dan pemetaan data petir secara lebih efisien.
2. Bagi pihak terkait, seperti pemerintah daerah, instansi penanggulangan bencana (BPBD), pengelola infrastruktur, serta sektor transportasi dan kelistrikan, penelitian ini dapat memberikan informasi mengenai pola kejadian petir yang dapat dimanfaatkan dalam upaya mitigasi bencana, perencanaan pembangunan, serta perlindungan terhadap fasilitas publik yang rentan terhadap sambaran petir. Selain itu, informasi ini juga dapat digunakan sebagai bahan sosialisasi kepada masyarakat umum, khususnya bagi kelompok yang beraktivitas di luar ruangan, agar lebih waspada terhadap potensi kejadian petir.

1.5 Batasan Masalah

Penelitian ini menetapkan beberapa batasan untuk memperjelas ruang lingkup analisis yang dilakukan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data kejadian petir yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang.
2. Jenis petir yang terdapat dalam data meliputi *Cloud to Ground (CG+)*, *Cloud to Ground (CG-)*, dan *Intracloud (IC)*.

BAB II

STUDI PUSTAKA

Penelitian yang berkaitan dengan penerapan metode *Principal Component Analysis (PCA)* dalam pengelolaan dan pelabelan data telah dilakukan sebelumnya. Salah satunya oleh Almais *et al.* (2023) yang membahas penerapan *Principal Component Analysis (PCA)* untuk pelabelan data tingkat kerusakan sektor pasca bencana alam. Umumnya, PCA banyak digunakan sebagai teknik reduksi dimensi, tetapi pada penelitian ini PCA dimanfaatkan untuk memberi label pada data yang bersifat *unsupervised*. Data yang digunakan berjumlah 216 kasus kerusakan bangunan pasca bencana di Jawa Timur, dengan lima kriteria penilaian, yaitu kondisi bangunan, struktur bangunan, fisik bangunan, fungsi bangunan, dan kondisi pendukung lainnya. Hasil analisis PCA menghasilkan dua komponen utama, yaitu PC1 dengan varian 85,17% dan PC2 dengan varian 9,36%. Data kemudian dikelompokkan menjadi tiga klaster berdasarkan koordinat hasil PCA: klaster 1 ($n < 0$) mewakili kerusakan ringan, klaster 2 ($0 \leq n < 2$) untuk kerusakan sedang, dan klaster 3 ($n \geq 2$) untuk kerusakan berat. Validasi dengan data target asli menunjukkan bahwa PC1 memiliki distribusi data yang lebih baik dibandingkan PC2. sehingga lebih sesuai digunakan untuk pelabelan. Hasil penelitian ini menunjukkan bahwa PCA tidak hanya berguna untuk reduksi fitur, tetapi juga efektif dalam pelabelan data kerusakan sektor pasca bencana. Temuan ini memberikan kontribusi penting bagi proses penentuan tingkat kerusakan secara ilmiah, yang selanjutnya dapat mendukung pengambilan keputusan, khususnya dalam distribusi bantuan pasca bencana (Almais *et al.*, 2023).

Selanjutnya, Lovytaji *et al.* (2024) mengusulkan *Ellipse Fitting Method* dengan pendekatan PCA untuk memetakan tingkat kerawanan sambaran petir pada wilayah Pasuruan. Penelitian ini menunjukkan bahwa penggunaan PCA dapat membantu menentukan konfigurasi ellipse sebagai batasan daerah rawan petir, dengan hasil yang mengindikasikan tingkat kerentanan tinggi pada Kecamatan Lumbang dan tingkat kerentanan sedang pada Kecamatan Gempol. Meskipun efektif untuk pemetaan risiko, penelitian ini lebih menitikberatkan pada identifikasi zona rawan, bukan pada analisis pengelompokan spasial-temporal petir (Lovytaji *et al.*, 2024).

Nicora *et al.* (2024) memanfaatkan PCA untuk melakukan ekstraksi fitur pada sinyal lightning-induced voltage pada saluran transmisi, yang kemudian diproses dengan Neural Network untuk memperkirakan lokasi sambaran dan arus puncak petir (Nicora *et al.*, 2024). Selanjutnya, Zhang *et al.* (2024) menggunakan PCA untuk melakukan kompresi sinyal elektromagnetik VLF pada jaringan deteksi petir jarak jauh, sehingga efisiensi transmisi data meningkat dan akurasi penentuan lokasi sambaran lebih baik (Zhang *et al.*, 2024). Sementara itu, Morgenstern *et al.* (2022) mengombinasikan PCA dengan K-Means pada data ERA5 yang mencakup variabel atmosfer dan data sambaran petir di Jerman, untuk mengidentifikasi tipe petir atau awan badai serta membandingkan pola spasial-temporal antar musim (Morgenstern *et al.*, 2022).

Dari sisi pemetaan, Budi & Pratama (2025) menerapkan metode Inverse Distance Weighted (IDW) pada data sambaran petir tipe CG di Kota Manado tahun 2023 untuk membuat peta kerapatan sambaran petir yang dapat dibandingkan

dengan pendekatan clustering spasial (Budi & Pratama, 2025). Penelitian lain oleh Stucke *et al.* (2024) menggunakan PCA dan K-Means *clustering* pada data *upward lightning* di objek tinggi kawasan Eropa, dengan tujuan membedakan tipe petir berdasarkan asal dan karakteristiknya sekaligus membentuk model identifikasi (Stucke *et al.*, 2024). Selain itu, Abriyanto & Batih (2023) menganalisis potensi ancaman sambaran petir pada titik-titik tower transmisi Air Anyir Pangkalpinang dengan data historis 2019-2021. sekaligus menilai kelayakan ekonomi pemasangan proteksi berdasarkan prioritas tower yang paling rawan (Abriyanto & Batih, 2023).

Dari perspektif kerentanan wilayah, Septiarini *et al.* (2020) mengkaji data petir Cloud to Ground (CG) tahun 2018-2020 dengan mempertimbangkan faktor kepadatan penduduk dan luasan bangunan menggunakan metode SAW, dan hasilnya menunjukkan bahwa beberapa kecamatan masuk kategori rawan tinggi (Septiarini *et al.*, 2020). Sementara itu, Seftiani *et al.* (2023) meneliti sinyal petir CG dengan sensor Universitas Andalas dan menemukan rasio PB/RS rata-rata sebesar 13,89% yang menjadi ciri khas petir CG+ meski lebih jarang terjadi namun berbahaya (Seftiani *et al.*, 2023). Kemudian, Surono *et al.* (2023) memadukan PCA, K-Means, dan RNN pada data kualitas udara, PCA berhasil mereduksi variabel menjadi tiga komponen, K-Means menghasilkan pengelompokan optimal, dan RNN menunjukkan performa prediksi dengan akurasi tinggi (Surono *et al.*, 2023).

Penelitian terbaru oleh Haljasmaa *et al.* (2024) menunjukkan penerapan PCA dan K-Means dalam pengelompokan kondisi cuaca untuk meningkatkan akurasi peramalan energi pada pembangkit listrik tenaga surya dalam kurung PLTS. Dalam penelitian tersebut menggunakan data meteorologi seperti temperatur,

kelembaban, dan angin untuk membangun model prediksi spesifik pada setiap cluster. Hasilnya, model berbasis cluster sering memberikan akurasi prediksi hingga 90%, yang lebih tinggi daripada pendekatan model umum tanpa mengelompokkan data. Penelitian ini menegaskan bahwa kombinasi PCA dengan metode clustering mampu menyederhanakan variabel kompleks sekaligus meningkatkan ketepatan analisis (Haljasmaa *et al.*, 2024).

Tabel 2.1 Rangkuman Penelitian Terkait

Referensi	Topik	Metode	Subjek
(Almais <i>et al.</i> , 2023)	Pelabelan data kerusakan pasca bencana	PCA untuk labeling & clustering	216 kasus kerusakan bangunan di Jatim
(Lovytaji <i>et al.</i> , 2024)	Pemetaan kerawanan petir	PCA & Ellipse Fitting	Wilayah Pasuruan
(Nicora <i>et al.</i> , 2024)	Estimasi lokasi & arus puncak petir	PCA & Neural Network	Data Sinyal Petir
(Zhang <i>et al.</i> , 2024)	Kompresi sinyal petir VLF	PCA untuk kompresi data	Jaringan deteksi petir jarak jauh
(Budi & Pratama, 2025)	Pemetaan kerapatan petir	Inverse Distance Weighted (IDW)	Data titik sambaran CG Manado 2023
(Stucke <i>et al.</i> , 2024)	Identifikasi upward lightning	PCA & K-means	Data petir objek tinggi di Eropa
(Abriyanto & Batih, 2023)	Analisis potensi ancaman petir	Analisis historis & ekonomi proteksi	Tower transmisi Air Anyir (2019-2021)
(Septiarini <i>et al.</i> , 2020)	Kerentanan wilayah rawan petir	Simple Additive Weighting (SAW)	Data CG & factor kepadatan penduduk
(Seftiani <i>et al.</i> , 2023)	Analisis sinyal petir CG	Analisis rasio PB/RS	Data sensor Universitas Andalas
(Surono <i>et al.</i> , 2023)	Predikis kualitas udara	PCA & K-means & RNN	Data kualitas udara
(Haljasmaa <i>et al.</i> , 2024)	Peramalan energi PLTS	PCA & K-Means	Data meteorologi
Penelitian saat ini	Clustering spasial-temporal petir	PCA & PCA Clustering	Data Petir BMKG

Secara keseluruhan, penelitian - penelitian yang telah dibahas menunjukkan kemampuan PCA dalam mereduksi dimensi data serta kontribusinya ketika PCA digabungkan dengan metode clustering untuk menghasilkan analisis yang lebih akurat. Sebagian besar penelitian tersebut berfokus pada konteks lain, seperti transmisi listrik, deteksi sinyal petir maupun *forecasting energy*, kesamaannya terletak pada pemanfaatan PCA sebagai tahap penting untuk menyederhanakan data sebelum masuk ke proses analisis selanjutnya. Keterkaitan dengan penelitian ini terletak pada penggunaan PCA untuk menyaring informasi spasial-temporal dan petir agar dapat terbentuk cluster yang lebih terstruktur. Dengan cara tersebut, penelitian ini mampu memperluas penerapan PCA-Clustering khususnya dalam konteks analisis pola petir pada wilayah Indonesia.

2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan merupakan salah satu yang banyak digunakan dalam analisis data untuk mereduksi dimensi tanpa menghilangkan informasi penting yang terkandung (Mubarok *et al.*, 2025). PCA bekerja dengan cara mentransformasi sekumpulan variabel yang saling berkorelasi menjadi kelompok variabel baru yang tidak saling berkorelasi, yang dikenal sebagai principal components (Jolliffe & Cadima, 2016).

Dalam penerapannya, PCA sering dimanfaatkan sebagai tahap preprocessing dalam analisis data untuk mengurangi beban komputasi, serta meningkatkan efektivitas metode lanjutan seperti klasifikasi atau clustering (Rosyada & Utari, 2024). Pada penelitian ini, PCA memiliki peran untuk menyaring informasi penting dari data spasial-temporal petir, sehingga pola yang kompleks

dapat diubah menjadi representasi yang lebih sederhana. Dalam proses ini, PCA tidak menghapus variabel secara langsung, melainkan mengubah bentuk representasinya menjadi beberapa komponen utama yang memuat sebagian besar variasi informasi (Mathai *et al.*, 2022). Komponen dengan kontribusi varians yang kecil dapat diabaikan karena dianggap memiliki pengaruh yang minim terhadap struktur utama data. Dengan cara ini, PCA mampu menyederhanakan data berdimensi tinggi menjadi ruang berdimensi lebih rendah tanpa kehilangan pola penting di dalamnya (Rosyada & Utari, 2024).

2.2 PCA-Clustering

Clustering merupakan metode unsupervised learning yang bertujuan mengelompokkan sekumpulan data menjadi beberapa kelompok berdasarkan tingkat kemiripan antar objek dalam grup tersebut (Serra & Tagliaferri, 2018). Tujuan utama dari proses *clustering* yaitu untuk mengidentifikasi struktur yang tersembunyi dalam data, sehingga dapat memberikan pemahaman tentang kelompok atau kategori pada data tersebut (Hendrastuty, 2024). *Clustering* memiliki banyak manfaat pada berbagai bidang, mulai dari pengenalan pola, pengolahan citra, analisis pasar, hingga studi mengenai fenomena alam. Dengan menerapkan metode ini, data berukuran besar dapat terpetakan menjadi kelompok yang lebih kecil dan lebih homogen, sehingga memudahkan proses analisis (Indraputra & Fitriana, 2020).

PCA-Clustering merupakan pendekatan analisis data berbasis *unsupervised learning* yang menggabungkan proses reduksi dimensi dengan *Principal Component Analysis* (PCA) dan pengelompokan (*clustering*) data berdasarkan pola

kesamaan hasil proyeksi PCA. Metode ini digunakan untuk menyederhanakan data berdimensi tinggi menjadi beberapa komponen utama (*principal components*) yang mampu menjelaskan sebagian besar variasi dalam data, sehingga struktur atau pola tersembunyi dapat diidentifikasi dengan lebih jelas (Almais *et al.*, 2023).

Dengan menggunakan pendekatan ini, *PCA-Clustering* mampu mengidentifikasi pola atau kelompok data yang memiliki karakteristik serupa secara spasial dan temporal tanpa memerlukan parameter awal seperti jumlah cluster. Hal ini menjadikan metode ini lebih efisien dan objektif dibandingkan metode clustering tradisional yang memerlukan asumsi awal tertentu. Selain itu, hasil pengelompokan *PCA-Clustering* dapat divisualisasikan dalam bentuk grafik dua dimensi yang memperlihatkan distribusi alami antar kelompok data, sehingga memudahkan analisis interpretatif terhadap fenomena yang dikaji, termasuk dalam penelitian ini yaitu pola spasial-temporal petir .

2.3 Petir

Petir merupakan fenomena alam berupa pelepasan muatan listrik bertegangan tinggi pada atmosfer, yang terjadi akibat perbedaan potensial antara awan dengan awan lain atau antara awan dengan bumi (Denov *et al.*, 2023). Fenomena ini termasuk dalam kajian geofisika dan meteorologi, karena berhubungan dengan dinamika awan konvektif dan cuaca ekstrem (Zhu *et al.*, 2025). Petir dapat menghasilkan energi besar dan arus puncak tinggi, yang berpotensi merusak infrastruktur dan membahayakan keselamatan manusia (Cahyo *et al.*, 2023).

Data petir adalah kumpulan informasi yang merekam karakteristik kejadian sambaran, baik secara spasial maupun temporal, pada wilayah tertentu. Informasi yang biasanya dicatat meliputi waktu kejadian (*epoch* dan *datetime*), posisi geografis (*latitude* dan *longitude*), serta jenis sambaran petir. Jenis sambaran tersebut umumnya dibedakan menjadi empat kategori, yaitu *Intracloud* (IC), *Cloud to Cloud* (CC), *Cloud to Air* (CA), dan *Cloud to Ground* (CG). Dari keempatnya, *Cloud to Ground* (CG) dianggap paling berbahaya karena memiliki dampak langsung terhadap manusia, infrastruktur, dan lingkungan (Yusnaini *et al.*, 2021).

Dari keempat jenis petir diatas, penelitian ini berfokus pada tiga jenis petir utama, yaitu CG+, CG-, dan IC. Pemilihan ketiga jenis ini didasarkan pada relevansi dan kontribusinya terhadap analisis spasial-temporal aktivitas petir. Jenis CG+ dan CG- atau petir *Cloud to Ground* merupakan sambaran yang paling berbahaya karena secara langsung berinteraksi dengan permukaan bumi dan dapat menimbulkan dampak signifikan terhadap keselamatan manusia, infrastruktur, serta lingkungan sekitar (Jiang *et al.*, 2023).

Sementara itu, jenis IC digunakan karena memiliki peran penting sebagai indikator awal dari aktivitas petir di dalam awan. Sambaran IC sering kali menjadi tahapan awal sebelum terbentuknya sambaran CG, sehingga analisis terhadap IC dapat memberikan informasi tentang dinamika konvektif dan proses pembentukan badai (Tai *et al.*, 2023). Dengan memasukkan IC ke dalam analisis bersama CG+ dan CG-, penelitian ini bisa menggambarkan aktivitas petir secara lebih komprehensif. Sebaliknya untuk dua jenis lainnya, yaitu *Cloud to Cloud* (CC), dan *Cloud to Air* (CA) tidak digunakan dalam penelitian ini karena intensitasnya relatif

kecil dan jarang tercatat secara signifikan pada sistem deteksi petir darat, serta tidak memberikan kontribusi langsung terhadap risiko permukaan (Gunawan et al., 2014).

2.4 Spasial dan Temporal

Selain ditinjau dari jenisnya, petir juga memiliki pola spasial-temporal yang menggambarkan pola penyebaran sambaran berdasarkan lokasi dan waktu kejadiannya (Pratama & Negara, 2017). Secara spasial, distribusi petir tidak merata di seluruh wilayah, melainkan terkonsentrasi pada area tertentu yang memiliki kondisi geografis tertentu seperti daerah pantai, pegunungan, dimana proses konvektif lebih aktif akibat perbedaan elevasi dan kelembapan udara (Chakraborty et al., 2021). Misalnya, wilayah selatan Jawa Timur yang memiliki aktivitas konvektif tinggi cenderung mengalami kepadatan sambaran petir lebih besar dibandingkan daerah dataran rendah di sekitarnya (Riqmawatin & Intan, 2020).

Sementara secara temporal, kejadian petir dipengaruhi oleh siklus waktu harian dan musiman. Pada skala harian, sambaran petir sering meningkat pada sore hingga malam hari akibat proses pemanasan permukaan bumi yang memicu pembentukan awan konvektif, sedangkan pada skala musiman, intensitas petir lebih sering terjadi pada musim hujan dibandingkan musim kemarau karena meningkatnya kelembapan atmosfer dan aktivitas awan cumulonimbus (Shubri et al., 2023).

Dalam penelitian ini, spasial-temporal direpresentasikan secara langsung melalui atribut pada data petir yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang. Atribut spasial ditunjukkan oleh koordinat *latitude* dan *longitude*, yang

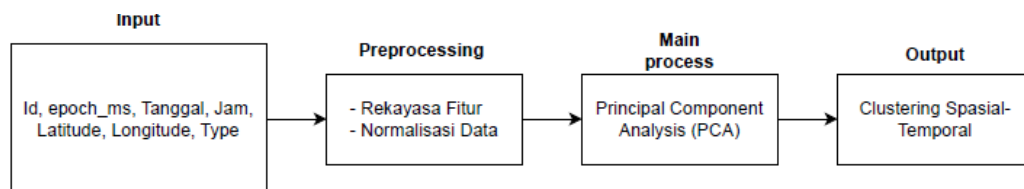
menggambarkan posisi geografis sambaran petir di permukaan bumi. Melalui data ini, pola persebaran sambaran dapat divisualisasikan dan dianalisis untuk mengetahui wilayah dengan tingkat kejadian petir yang tinggi. Sedangkan atribut temporal direpresentasikan oleh *epoch_ms*, yaitu waktu terdeteksinya sambaran petir dalam satuan milidetik sejak epoch time. Kedua atribut ini memungkinkan analisis dinamika kejadian petir dalam kurun waktu tertentu, seperti durasi, frekuensi, serta distribusi waktu sambaran. Dengan demikian, kombinasi atribut spasial dan temporal memberikan gambaran yang komprehensif mengenai pola persebaran dan aktivitas petir, yang kemudian menjadi dasar dalam proses analisis menggunakan metode *PCA-Clustering*.

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Desain Sistem

Desain sistem ini disusun secara terstruktur agar setiap tahap penelitian dapat dilaksanakan dengan optimal dan bisa menghasilkan hasil yang tepat. Gambar 3.1 berikut menampung desain sistem yang digunakan dalam analisis data petir menggunakan *Principal Component Analysis* (PCA).



Gambar 3.1 Desain Sistem

3.1.1 Input

Tahap input merupakan proses awal dalam sistem, yaitu memasukkan data petir hasil proses *cleaning* yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang. Data yang digunakan meliputi atribut waktu dan spasial, yaitu *epoch_ms*, tanggal, jam, *latitude*, *longitude*, serta *type* sebagai informasi tambahan. Atribut-atribut tersebut selanjutnya digunakan sebagai dasar dalam proses rekayasa fitur dan analisis PCA untuk mengidentifikasi pola spasial-temporal kejadian petir.

3.1.2 Rekayasa Fitur (*Feature Engineering*)

Rekayasa fitur dilakukan untuk menghasilkan representasi data yang lebih informatif dan sesuai dengan kebutuhan analisis spasial-temporal. Pada penelitian ini, data petir yang diperoleh dari BMKG memiliki atribut waktu dalam bentuk tanggal dan jam kejadian. Agar informasi temporal tersebut dapat dimanfaatkan

secara optimal dalam analisis PCA dan *clustering*, dilakukan proses rekayasa fitur terhadap variabel waktu (Tropmann-frick & Forke, 2021).

Tahap awal rekayasa fitur dilakukan dengan menggabungkan atribut tanggal dan jam menjadi satu atribut waktu utuh (*datetime*). Selanjutnya, dari atribut *datetime* tersebut diekstraksi beberapa fitur temporal, yaitu jam kejadian (*hour*) dan hari dalam satu minggu (*dayofweek*). Variabel *dayofweek* digunakan untuk merepresentasikan variasi kejadian petir berdasarkan hari, dengan rentang nilai 0 hingga 6.

Untuk merepresentasikan pola waktu harian secara lebih akurat, dilakukan transformasi siklik terhadap variabel jam kejadian menggunakan fungsi sinus dan kosinus (Indriyanti & Fajriah, 2025). Transformasi ini menghasilkan dua fitur baru, yaitu *hour_sin* dan *hour_cos*, yang dihitung menggunakan persamaan:

$$hour_sin = \sin\left(\frac{2\pi \cdot hour}{24}\right) \quad (3.1)$$

$$hour_cos = \cos\left(\frac{2\pi \cdot hour}{24}\right) \quad (3.2)$$

Transformasi siklik dilakukan untuk menghindari diskontinuitas nilai waktu, di mana jam 23 dan jam 0 secara numerik berjauhan tetapi secara konseptual berdekatan. Dengan pendekatan ini, hubungan periodik pada data waktu dapat direpresentasikan dengan lebih baik.

Hasil dari proses rekayasa fitur ini adalah penambahan variabel temporal *hour_sin*, *hour_cos*, dan *dayofweek* yang selanjutnya digunakan bersama dengan variabel spasial *latitude* dan *longitude* sebagai input dalam proses normalisasi,

PCA, dan *clustering*. Rekayasa fitur ini bertujuan untuk memastikan bahwa informasi spasial dan temporal kejadian petir dapat dianalisis secara seimbang dan terintegrasi.

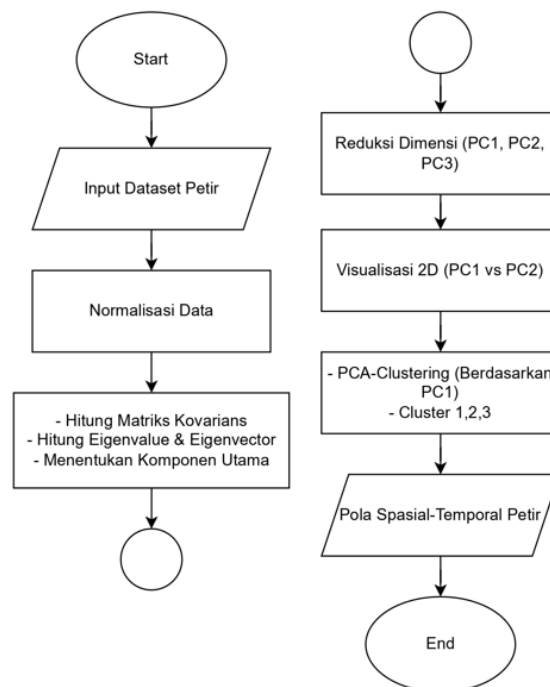
3.1.3 Preprocessing

Pada tahap *preprocessing* dilakukan rekayasa fitur dan normalisasi data. Rekayasa fitur dilakukan untuk membentuk variabel temporal baru yang lebih representatif terhadap pola waktu kejadian petir, seperti *hour*, *dayofweek*, serta *encoding siklik hour_sin* dan *hour_cos*. Tahap ini bertujuan untuk mengubah informasi waktu menjadi bentuk numerik yang dapat dianalisis secara matematis dalam proses PCA.

Setelah rekayasa fitur, dilakukan normalisasi data untuk menyamakan skala antar atribut sehingga setiap variabel memiliki kontribusi yang seimbang dalam perhitungan PCA. Proses ini dilakukan menggunakan metode *StandardScaler*. Dengan normalisasi, perbedaan rentang nilai antar fitur dapat diminimalkan sehingga tidak ada variabel yang mendominasi pembentukan komponen utama.

3.1.4 Main Process

Tahap utama dalam sistem ini adalah penerapan metode *Principal Component Analysis* (PCA) untuk mereduksi dimensi data petir dan mengekstraksi komponen utama yang merepresentasikan variasi terbesar dalam dataset. Proses ini meliputi tahap normalisasi data, perhitungan matriks kovarians, penentuan nilai eigen dan vektor eigen, hingga pembentukan komponen utama. Alur lengkap proses utama sistem ditunjukkan pada *flowchart* berikut.



Gambar 3.2 *Flowchart Main Process Sistem PCA-Clustering*

a) Normalisasi Data

Langkah awal pada proses PCA yaitu melakukan normalisasi data. Normalisasi bertujuan agar setiap variabel memiliki skala yang seimbang, sehingga tidak ada atribut yang mendominasi proses analisis. Proses ini dilakukan menggunakan metode standar seperti *StandardScaler* agar data sesuai dengan standar PCA (Almais *et al.*, 2023). *StandardScaler* menstandarisasi data dengan mengubah setiap nilai menjadi bentuk *z-score*, yaitu mengurangi rata-rata dan membaginya dengan simpangan baku. Dengan cara ini, setiap atribut memiliki rata-rata nol dan deviasi standar satu, sehingga tidak ada variabel yang mendominasi proses pembentukan komponen utama (Han *et al.*, 2022). Normalisasi dihitung dengan rumus:

$$X_{Standart} = \frac{X - \text{mean}(x)}{X(\text{standart deviasi})} \quad (3.3)$$

Dengan X adalah nilai data asli, $mean(X)$ merupakan nilai rata-rata dari data, dan $standar\ deviasi(X)$ adalah simpangan baku. Standar deviasi dihitung dengan rumus:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.4)$$

Keterangan:

- a) x_i = data ke- i
- b) n = Jumlah sampel
- c) X = simbol standar deviasi

Proses ini memiliki peran yang sangat penting dalam PCA, sebab kualitas hasil reduksi dimensi bergantung pada data yang sudah melalui tahap normalisasi secara tepat.

b) Menentukan Matriks Kovarians dan Nilai Eigen

Setelah melalui tahap normalisasi, langkah berikutnya dalam PCA adalah menghitung nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari matriks kovarians. Nilai eigen menggambarkan besarnya varians yang dijelaskan oleh setiap komponen utama, sedangkan vektor eigen menunjukkan arah dari komponen tersebut dalam ruang data. Matriks kovarians dapat dihitung dengan rumus:

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.5)$$

Keterangan:

- a) C = matriks kovarians
- b) n = jumlah sampel
- c) x_i = vector data pada titik ke- i

Sedangkan untuk menghitung nilai eigen, bisa menggunakan rumus persamaan berikut:

$$\det(C - \lambda I) = 0 \quad (3.6)$$

Keterangan:

- a) C = matriks kovarians
- b) λ = nilai eigen yang dicari
- c) I = matriks identitas dengan dimensi yang sama dengan C
- d) $\det()$ = operator determinan

Matriks kovarians digunakan untuk mengetahui tingkat hubungan antara masing-masing variabel berdasarkan data yang telah dikumpulkan. Semakin besar nilai kovarians, maka semakin kuat hubungan antara dua variabel tersebut. Berikut adalah contoh perhitungan matriks kovarians menggunakan data petir dari tabel 3.1:

Tabel 3.1 Data Petir

id	epoch_ms	tanggal	jam	latitude	longitude	type
1	1.74603E+12	4/30/2025	17:00:07	-9.219843755	114.9134671	2
2	1.74603E+12	4/30/2025	17:00:17	-7.437814564	109.2385823	0
3	1.74603E+12	4/30/2025	17:00:42	-9.233128731	114.9770453	2
4	1.74603E+12	4/30/2025	17:00:47	-5.099550413	112.1911726	2
5	1.74603E+12	4/30/2025	17:01:34	-4.980607535	112.6858848	2

a. Hitung Rata-Rata Setiap Variabel

Langkah pertama dalam menghitung matriks kovarians adalah menghitung nilai rata-rata pada setiap variabel. Berikut contoh perhitungan rata-rata pada variable *type*:

$$Mean_{type} = \frac{(2 + 0 + 2 + 2 + 2)}{5} = 1.6$$

b. Hitung Varians Setiap Kolom

Setelah memperoleh nilai rata-rata, langkah selanjutnya adalah menghitung varians masing-masing variabel menggunakan Rumus (3.6):

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Berikut contoh perhitungan varians variable *type*:

$$C_{type} = \frac{1}{5-1} [(2 - 1.6)^2 + (0 - 1.6)^2 + (2 - 1.6)^2 + (2 - 1.6)^2 + (2 - 1.6)^2]$$

$$C_{type} = \frac{0.16 + 2.56 + 0.16 + 0.16 + 0.16}{4}$$

$$C_{type} = 0.8$$

c. Hitung Kovarians antar Variabel

Langkah berikutnya adalah menghitung kovarians antar variable dengan menggunakan rumus:

$$C_{kl} = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{il} - \bar{x}_l)$$

Berikut contoh perhitungan kovarians antara variabel *type* (x_7) dan variabel latitude (x_5):

$$C_{75} = \frac{1}{4} \sum [(2 - 1.6)(-9.2198 + 7.1942) + (0 - 1.6)(-7.4378 + 7.1942) + (2 - 1.6)(-9.2331 + 7.1942) + (2 - 1.6)(-5.0996 + 7.1942) + (2 - 1.6)(-4.9806 + 7.1942)]$$

$$C_{75} = 0.1218$$

Setelah menghitung kovarians untuk seluruh pasangan variable, maka diperoleh matriks kovarians berikut:

$$\begin{bmatrix} 2.5 & 0 & 0 & \dots & 0.5 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ - & - & - & \dots & - \\ - & - & - & \dots & - \\ 0.5 & 0 & 0 & \dots & 0.8 \end{bmatrix}$$

Setelah menghitung matriks kovarians, tahap berikutnya adalah mencari nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*). Nilai eigen merepresentasikan besarnya varians yang dapat dijelaskan oleh masing-masing komponen utama, sementara vektor eigen menunjukkan arah dari komponen tersebut dalam ruang data. Nilai varians yang tinggi menunjukkan data memiliki penyebaran luas, sedangkan varians rendah menunjukkan data lebih terkonsentrasi di sekitar nilai rata-rata. Semakin besar nilai eigen, semakin besar pula proporsi keragaman data yang mampu dijelaskan oleh komponen utama terkait. Dengan demikian, nilai eigen menjadi indikator penting dalam memilih dimensi yang relevan untuk analisis lebih lanjut.

Tahap selanjutnya adalah reduksi data dengan menghitung nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari matriks kovarians yang telah diperoleh sebelumnya. Proses ini bertujuan untuk menentukan komponen utama yang mampu merepresentasikan variasi data petir dalam dimensi yang lebih rendah. Nilai eigen diperoleh dengan menyelesaikan persamaan karakteristik sebagai berikut:

$$\det(C - \lambda I) = 0$$

Sebagai contoh, setiap elemen diagonal pada matriks kovarians dikurangi dengan λ . Matriks $C - \lambda I$ dari data petir yang memiliki 7 variabel adalah:

$$C - \lambda I = \begin{bmatrix} 2.5 - \lambda & 0 & 0 & \dots & 0.5 \\ 0 & -\lambda & 0 & \dots & 0 \\ 0 & 0 & -\lambda & \dots & 0 \\ - & - & - & \dots & - \\ - & - & - & \dots & - \\ 0.5 & 0 & 0 & \dots & 0.8 - \lambda \end{bmatrix}$$

$$\det(C - \lambda I) = (-\lambda)^3 \cdot \det(A)$$

Dengan submatriks A sebagai berikut:

$$A = \begin{bmatrix} 2.5 - \lambda & 2.7 & -0.38 & 0.5 \\ 2.7 & 4.4 - \lambda & -2.35 & 0.12 \\ -0.38 & -2.35 & 5.57 - \lambda & 1.78 \\ 0.5 & 0.12 & 1.78 & 0.8 - \lambda \end{bmatrix}$$

$$\det A = \lambda^4 - 13.27\lambda^3 + 52.91\lambda^2 - 41.6\lambda$$

$$(-\lambda)^3(\lambda^4 - 13.27\lambda^3 + 52.91\lambda^2 - 41.6\lambda) = 0$$

$$\lambda_1 = 0. \lambda_2 = 0. \lambda_3 = 0. \lambda_4 = 5.57. \lambda_5 = 4.4, \lambda_6 = 2.5, \lambda_7 = 0.8$$

Setelah nilai eigen diperoleh, langkah selanjutnya adalah menghitung vektor eigen yang bersesuaian dengan menyelesaikan sistem persamaan linear:

$$(C - \lambda I)v = 0$$

Sebagai contoh, jika digunakan nilai eigen $\lambda = 2.5$ maka persamaannya:

$$\begin{bmatrix} 2.5 - 2.5 & 0 & 0 & \dots & 0.5 \\ 0 & -2.5 & 0 & \dots & 0 \\ 0 & 0 & -2.5 & \dots & 0 \\ - & - & - & \dots & - \\ - & - & - & \dots & - \\ - & - & - & \dots & - \\ 0.5 & 0 & 0 & \dots & 0.8 - 2.5 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Sebagai contoh, pada baris kedua diperoleh persamaan:

$$-2.5V_2 = 0. V_2 = 0$$

Setelah nilai eigen dan vektor eigen diperoleh, langkah selanjutnya adalah memilih komponen utama (*principal components*) yang memiliki nilai eigen lebih besar dari 1. Berdasarkan hasil perhitungan, diperoleh dua nilai eigen terbesar, sehingga dua vektor eigen yang bersesuaian digunakan sebagai basis baru.

c) Menentukan Komponen Utama

Setelah nilai eigen dan vektor eigen diperoleh, vektor eigen diurutkan berdasarkan nilai eigen terbesar. Komponen utama kemudian dibentuk dari kombinasi linier variabel asli menggunakan vektor eigen. Komponen dengan nilai eigen terbesar diprioritaskan karena mampu memuat variasi informasi paling signifikan dari data. Jumlah komponen utama yang digunakan dipilih berdasarkan variansi yang mampu merepresentasikan karakteristik utama data petir, sehingga informasi penting terkait pola spasial-temporal tetap terjaga meskipun dimensi data berkurang.

d) Membuat Grafik 2D

Jika data petir memiliki banyak variabel spasial-temporal, PCA akan mereduksi kompleksitasnya menjadi beberapa komponen utama. Dua komponen pertama (PC1 dan PC2) kemudian dipilih karena mampu merepresentasikan sebagian besar variasi pola kejadian petir yang terjadi pada data. Hasil reduksi ini divisualkan dalam bentuk *scatter plot* dua dimensi untuk memperlihatkan pola distribusi data secara lebih sederhana. Melalui grafik ini, kecenderungan pembentukan cluster maupun pemisahan antar kelompok data dapat diamati dengan lebih jelas, sehingga mendukung proses interpretasi hasil analisis.

Dengan rangkaian tahapan tersebut, PCA mampu menyederhanakan data berdimensi tinggi menjadi representasi baru yang lebih ringkas tanpa kehilangan informasi penting, proses-proses tersebut menjadikan PCA tidak hanya sebagai metode reduksi dimensi tetapi juga sebagai pondasi penting dalam mendukung analisis clustering. Dalam penelitian ini, penerapan PCA berfungsi untuk

mengekstraksi informasi utama dari data spasial temporal petir sehingga pola kejadian dapat teridentifikasi dengan lebih jelas dan terstruktur.

Setelah melakukan proses PCA, tahap selanjutnya adalah *PCA-Clustering* yaitu pemberian label pada data hasil reduksi dimensi. Label ini memiliki fungsi sebagai penanda kelompok atau *cluster* yang terbentuk, sehingga data yang telah divisualisasikan dapat dibedakan berdasarkan kelas atau kategori tertentu. Dengan label, interpretasi hasil analisis menjadi lebih mudah karena pola pengelompokan data dapat langsung dikenali melalui perbedaan warna atau simbol yang merepresentasikan masing-masing *cluster*.

3.1.5 Output

Output dari sistem ini berupa hasil pengelompokan (*clustering*) kejadian petir berdasarkan kemiripan karakteristik spasial dan temporal yang diperoleh dari hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Hasil tersebut divisualisasikan dalam bentuk grafik PCA 2D untuk memperlihatkan distribusi data dalam ruang komponen utama. Dengan demikian, sistem mampu menampilkan pola pengelompokan kejadian petir secara lebih terstruktur sehingga memudahkan analisis distribusi spasial-temporal dan interpretasi hasil penelitian.

3.2 Pseudocode Clustering menggunakan PCA-Clustering

Pseudocode clustering adalah bentuk penulisan algoritma clustering yang ditulis bukan dalam bahasa pemrograman murni, melainkan dalam bentuk deskripsi langkah-langkah yang menyerupai kode. Pseudocode berfungsi untuk menjelaskan alur proses mulai dari input data, normalisasi, perhitungan matriks kovarians, perhitungan nilai eigen dan vektor eigen, pembentukan komponen utama PCA,

hingga proses *clustering* dan visualisasi hasil. Pseudocode ini tidak langsung dieksekusi oleh komputer, namun dapat dengan mudah diterjemahkan ke dalam bahasa pemrograman.

Tabel 3.2 Pseudocode Algoritma Clustering

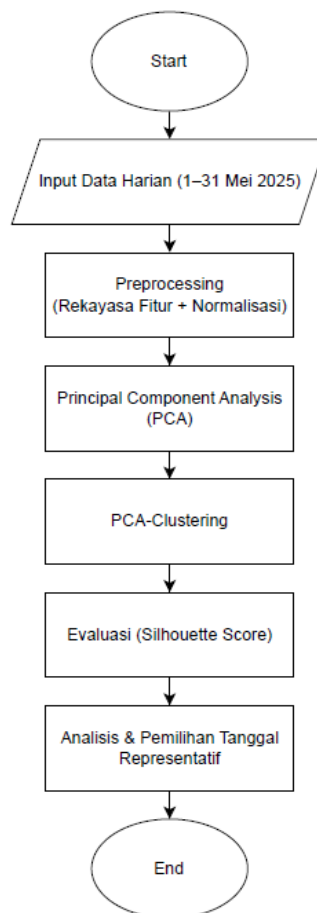
Algoritma Clustering menggunakan PCA dalam label data	
Input	Membaca data petir menggunakan fungsi <i>pd.read_csv("data_petir.csv")</i>
Process	<ul style="list-style-type: none"> • Normalisasi data menggunakan fungsi <i>StandardScaler()</i> dan <i>fit_transform()</i> • Menghitung rasio varians menggunakan fungsi <i>explained_variance_ratio_()</i> • Menghitung eigenvalues menggunakan fungsi <i>explained_variance_()</i> • Membuat komponen PCA menggunakan library <i>PCA()</i> • Membuat data frame hasil komponen PCA <i>pd.DataFrame()</i> • Visualisasi 2 dimensi data petir dengan PCA menggunakan <i>scatterplot()</i> • Membuat label menggunakan percabangan (<i>if else</i>) dan percabangan (<i>for</i>)
Output	<ul style="list-style-type: none"> • Menggabungkan data frame hasil komponen PCA dengan label menggunakan <i>concat()</i> • Hasil visualisasi dari proses pelabelan data menggunakan PCA yang divalidasi dengan data petir dari BMKG

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Desain Uji Coba

Desain uji coba pada penelitian ini bertujuan untuk menjelaskan alur eksperimen yang dilakukan dalam menerapkan metode *PCA-Clustering* terhadap data petir harian periode 1-31 Mei 2025. Setiap data harian diproses melalui tahapan *preprocessing*, reduksi dimensi menggunakan PCA, proses *clustering*, serta evaluasi menggunakan *Silhouette Score*. Alur lengkap proses uji coba ditunjukkan pada Gambar 4.1 berikut.



Gambar 4.1 *Flowchart* Desain Uji Coba *PCA-Clustering*

4.2 Preparation Data

Preparation data merupakan tahapan awal untuk menyiapkan data petir sebelum dianalisis lebih lanjut. Proses ini mencakup penginputan data, penyesuaian format temporal dan spasial, pembersihan data, serta normalisasi data. Seluruh tahapan *preprocessing* dilaksanakan menggunakan bahasa pemrograman Python pada platform Google Colab.

4.2.1 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari BMKG Stasiun Geofisika Kelas III Malang, berupa data kejadian petir untuk periode bulan Mei 2025. Data tidak digunakan dalam bentuk satu file tunggal, melainkan berasal dari kumpulan file harian yang tersimpan dalam satu folder berdasarkan bulan pengamatan. Meskipun data diperoleh melalui BMKG Stasiun Geofisika Malang, cakupan data tidak terbatas pada wilayah Malang saja, melainkan mencakup wilayah Indonesia secara nasional. Setiap file harian disimpan dalam format *Comma Separated Values* (CSV) yang merepresentasikan catatan kejadian petir pada tanggal pengamatan tersebut.

Setiap file CSV berisi informasi kejadian petir yang tercatat secara individual (*event-based*), dengan atribut utama meliputi id kejadian, *epoch time* (*epoch_ms*), tanggal, jam kejadian, koordinat geografis (*latitude* dan *longitude*), serta tipe petir. Tipe petir direpresentasikan dalam bentuk kode numerik, yaitu 0 untuk *Cloud to Ground* positif (CG+), 1 untuk *Cloud to Ground* negatif (CG-), dan 2 untuk *Intra Cloud* (IC). Contoh struktur data kejadian petir yang digunakan dalam penelitian ini ditunjukkan pada potongan data berikut. Proses pengumpulan data dilakukan

dengan memanfaatkan kumpulan file data petir harian berformat CSV yang tersedia untuk periode bulan Mei. Setiap file merepresentasikan kejadian petir pada rentang waktu tertentu dan memiliki struktur atribut yang seragam. Data tersebut kemudian digunakan sebagai input pada tahap *preprocessing* dan analisis selanjutnya.

Tabel 4.1 Data Petir 29 Mei 2025

id	epoch_ms	tanggal	jam	latitude	longitude	type
1	1.74845E+12	5/28/2025	17:00:04	-6.376008553	112.6554655	2
2	1.74845E+12	5/28/2025	17:00:11	-7.389219208	112.0637543	1
3	1.74845E+12	5/28/2025	17:00:11	-7.48637853	112.2015621	2
4	1.74845E+12	5/28/2025	17:00:32	-7.592867137	111.7753207	1
5	1.74845E+12	5/28/2025	17:00:32	-7.606525278	112.2666301	1
6	1.74845E+12	5/28/2025	17:00:39	-6.454559731	112.8058601	2
7	1.74845E+12	5/28/2025	17:00:42	-7.342066891	111.9446636	0
8	1.74845E+12	5/28/2025	17:00:45	-8.007372485	109.668349	1
9	1.74845E+12	5/28/2025	17:00:45	-7.423406107	112.0963115	2
10	1.74845E+12	5/28/2025	17:00:45	-7.346828755	112.1137935	2

Seluruh proses pembacaan data petir dilakukan menggunakan bahasa pemrograman Python dengan bantuan library pandas pada platform Google Colab. Setiap file CSV harian dibaca secara terpisah untuk mempertahankan struktur data kejadian petir pada masing-masing tanggal pengamatan. Proses ini bertujuan untuk memastikan bahwa setiap dataset harian memiliki format atribut yang konsisten sehingga dapat diproses pada tahap *preprocessing* dan dianalisis menggunakan metode *Principal Component Analysis (PCA)* dan *PCA-Clustering*.

Berdasarkan proses eksplorasi awal terhadap seluruh data harian, salah satu tanggal pengamatan dipilih untuk disajikan secara rinci pada bab hasil dan pembahasan. Pada penelitian ini, data kejadian petir tanggal 29 Mei 2025 digunakan sebagai representasi hasil analisis, sementara hasil analisis untuk tanggal-tanggal lainnya tetap disertakan pada bagian lampiran sebagai pendukung.

4.2.2 *Cleaning Data*

Proses *cleaning* data pada penelitian ini diawali dengan pengecekan struktur dan format data. Tahap ini meliputi penyesuaian penulisan nama kolom agar seragam serta memastikan tidak terdapat perbedaan format yang dapat menghambat proses pengolahan data selanjutnya.

Tahap kedua adalah pengolahan atribut waktu. Pada data mentah, informasi waktu tersimpan dalam satu kolom `datetime_utc` dengan format penulisan jam menggunakan tanda titik (.) serta menyertakan milidetik. Format ini perlu disesuaikan karena proses analisis memerlukan ekstraksi komponen waktu seperti jam, menit, dan hari dalam minggu sebagai fitur temporal. Oleh karena itu, atribut waktu tersebut dipisahkan menjadi dua kolom, yaitu tanggal dan jam, sehingga memudahkan proses ekstraksi fitur temporal yang digunakan dalam analisis selanjutnya.

Meskipun nama kolom yang digunakan adalah `datetime_utc`, waktu yang tercatat pada data telah disesuaikan dengan zona waktu Indonesia Barat (WIB, UTC+7). Dengan demikian, seluruh analisis temporal dalam penelitian ini merepresentasikan waktu lokal wilayah penelitian. Penamaan kolom tersebut merupakan format bawaan data dan tidak diubah, namun tidak mempengaruhi interpretasi waktu yang digunakan dalam analisis.

Perbedaan struktur data sebelum dan sesudah proses *cleaning* dapat dilihat pada Tabel 4.2 dan Tabel 4.3 berikut.

Tabel 4.2 Contoh Data Sebelum *Cleaning*

id	epoch_ms	datetime_utc	latitude	longitude	type
1	1.74845E+12	2025-05-28 17.00.04.000	-6.376008553	112.6554655	2
2	1.74845E+12	2025-05-28 17.00.11.000	-7.389219208	112.0637543	1
3	1.74845E+12	2025-05-28 17.00.11.000	-7.48637853	112.2015621	2
4	1.74845E+12	2025-05-28 17.00.32.000	-7.592867137	111.7753207	1
5	1.74845E+12	2025-05-28 17.00.32.000	-7.606525278	112.2666301	1

Tabel 4.3 Contoh Data Setelah *Cleaning*

id	epoch_ms	tanggal	jam	latitude	longitude	type
1	1.74845E+12	5/28/2025	17:00:04	-6.376008553	112.6554655	2
2	1.74845E+12	5/28/2025	17:00:11	-7.389219208	112.0637543	1
3	1.74845E+12	5/28/2025	17:00:11	-7.48637853	112.2015621	2
4	1.74845E+12	5/28/2025	17:00:32	-7.592867137	111.7753207	1
5	1.74845E+12	5/28/2025	17:00:32	-7.606525278	112.2666301	1

Berdasarkan kedua tabel tersebut, terlihat bahwa atribut waktu yang semula berada dalam satu kolom *datetime_utc* dengan format jam menggunakan tanda titik dan milidetik telah disesuaikan menjadi dua kolom terpisah, yaitu tanggal dan jam, dengan format penulisan waktu menggunakan tanda titik dua (:). Penyesuaian format waktu dilakukan agar data dapat diproses dengan baik menggunakan fungsi pengolahan waktu pada library *pandas* di Python. Format waktu yang lebih standar memungkinkan proses konversi ke tipe data *datetime* sehingga memudahkan ekstraksi komponen waktu seperti jam kejadian dan hari dalam minggu pada tahap rekayasa fitur.

Tahap berikutnya adalah pengecekan konsistensi data spasial. Pada tahap ini, atribut *latitude* dan *longitude* diperiksa untuk memastikan seluruh nilai berada dalam format numerik yang seragam dan tidak mengandung kesalahan penulisan atau nilai yang tidak valid. Pengecekan ini penting agar data spasial dapat merepresentasikan lokasi kejadian petir secara akurat serta dapat diproses pada tahap normalisasi dan PCA.

Tahap terakhir adalah penanganan data yang tidak lengkap, data dengan nilai kosong (*missing value*) pada variabel utama diidentifikasi dan tidak disertakan dalam proses analisis lanjutan. Setelah seluruh tahapan *cleaning* selesai, data petir memiliki struktur yang rapi, konsisten, dan siap digunakan sebagai input pada tahap rekayasa fitur, normalisasi data, serta analisis PCA dan *PCA-Clustering*.

4.2.3 Rekayasa Fitur Temporal

Rekayasa fitur merupakan tahapan lanjutan dalam *preprocessing data* yang bertujuan untuk menghasilkan variabel baru yang lebih representatif terhadap pola spasial dan temporal kejadian petir. Tahapan ini dilakukan setelah proses *cleaning data*, ketika data telah berada dalam kondisi konsisten dan siap untuk ditransformasikan lebih lanjut. Rekayasa fitur tidak mengubah nilai dasar hasil pengamatan, melainkan menurunkan informasi baru dari atribut yang telah tersedia. Pada penelitian ini, fokus rekayasa fitur pada aspek temporal data petir. Data awal hanya menyediakan informasi waktu dalam bentuk tanggal dan jam kejadian. Oleh karena itu, dilakukan transformasi terhadap atribut waktu tersebut untuk memperoleh fitur temporal yang dapat merepresentasikan karakteristik kejadian petir secara lebih informatif, khususnya dalam analisis berbasis *Principal Component Analysis (PCA)*.

Pada penelitian ini, rekayasa fitur difokuskan pada atribut waktu kejadian petir, yaitu kolom tanggal dan jam yang terdapat pada dataset awal. Informasi waktu tersebut digunakan untuk menurunkan beberapa fitur temporal yang dapat merepresentasikan pola kejadian petir dalam skala harian maupun mingguan.

Tabel 4.4 Data Waktu Kejadian Petir Sebelum Rekayasa Fitur

No	Tanggal	Jam
1	28/05/2025	17:00:04
2	28/05/2025	17:00:11
3	28/05/2025	17:00:11
4	28/05/2025	17:00:32
5	28/05/2025	17:00:32

Tabel 4.4 menunjukkan atribut waktu kejadian petir sebelum dilakukan proses rekayasa fitur. Informasi waktu masih tersimpan dalam dua kolom utama, yaitu tanggal dan jam, yang merepresentasikan hari serta waktu kejadian petir hingga tingkat detik. Dalam bentuk ini, atribut waktu belum secara langsung merepresentasikan pola temporal yang diperlukan dalam analisis numerik seperti PCA.

Untuk memperoleh representasi temporal yang lebih informatif, dilakukan proses rekayasa fitur dengan mengekstraksi beberapa informasi dari atribut tersebut. Dari atribut jam diperoleh fitur *hour* dan *minute* yang menunjukkan waktu kejadian petir dalam satu hari. Sementara itu, dari atribut tanggal dihasilkan fitur *dayofweek* yang menunjukkan hari kejadian dalam satu minggu dengan rentang nilai 0 hingga 6, di mana nilai 0 merepresentasikan hari Senin dan nilai 6 merepresentasikan hari Minggu.

Selain itu, dilakukan *cyclical* encoding terhadap variabel *hour* menggunakan fungsi sinus dan cosinus sehingga menghasilkan fitur *hour_sin* dan *hour_cos*. Pendekatan ini digunakan untuk merepresentasikan sifat periodik waktu, sehingga jam yang berdekatan secara kronologis, seperti pukul 23.00 dan 00.00, tetap memiliki kedekatan nilai secara matematis.

Tabel 4.5 Hasil Rekayasa Fitur Temporal Data Petir

No	Hour	Minute	Dayofweek	Hour_sin	Hour_cos
1	17	0	2	-0.9659	-0.2588
2	17	0	2	-0.9659	-0.2588
3	17	0	2	-0.9659	-0.2588
4	17	0	2	-0.9659	-0.2588
5	17	1	2	-0.9659	-0.2588

Tabel 4.5 menyajikan hasil rekayasa fitur temporal yang diperoleh dari transformasi atribut waktu kejadian petir. Proses ini menghasilkan beberapa fitur baru, yaitu *hour*, *minute*, *dayofweek*, serta fitur hasil *encoding* siklik berupa *hour_sin* dan *hour_cos*. Kolom *hour* dan *minute* menunjukkan waktu kejadian petir dalam satu hari, sedangkan *dayofweek* merepresentasikan hari kejadian dalam satu minggu dengan rentang nilai 0 hingga 6, di mana nilai 0 menunjukkan hari Senin dan nilai 6 menunjukkan hari Minggu. Melalui proses ini, atribut waktu yang sebelumnya tersimpan dalam bentuk tanggal dan jam telah ditransformasikan menjadi fitur temporal numerik yang lebih sesuai untuk analisis data. Fitur-fitur tersebut selanjutnya digunakan pada tahap normalisasi menggunakan metode *Standard Scaler* sebelum dilakukan analisis *Principal Component Analysis* (PCA).

4.3 Normalisasi Data

Tahap normalisasi data dilakukan untuk menyamakan skala antar variabel sebelum diterapkan analisis *Principal Component Analysis* (PCA). Data petir yang digunakan dalam penelitian ini memiliki rentang nilai yang berbeda-beda, baik pada atribut spasial seperti *latitude* dan *longitude* maupun pada atribut temporal dan kuantitatif lainnya. Perbedaan skala tersebut berpotensi menyebabkan variabel dengan nilai yang lebih besar mendominasi pembentukan komponen utama pada

proses PCA, sehingga diperlukan tahap normalisasi agar kontribusi setiap variabel tetap seimbang.

4.3.1 *Standard Scaler*

Setelah proses rekayasa fitur dilakukan pada tahap sebelumnya, data petir yang digunakan dalam penelitian ini memiliki beberapa variabel dengan satuan dan rentang nilai yang berbeda-beda. Variabel spasial seperti *latitude* dan *longitude* memiliki skala geografis, sedangkan variabel temporal seperti *hour_sin*, *hour_cos*, dan *dayofweek* berada pada skala numerik yang berbeda. Perbedaan skala tersebut dapat memengaruhi hasil analisis *Principal Component Analysis* (PCA), karena PCA sensitif terhadap besaran varians setiap variabel.

Sebelum proses normalisasi dilakukan, terlebih dahulu dilakukan pemilihan variabel yang relevan untuk analisis *Principal Component Analysis* (PCA). Data awal petir yang diperoleh dari BMKG terdiri dari beberapa atribut seperti *id*, *epoch_ms*, tanggal, jam, *latitude*, *longitude*, dan *type*. Namun, tidak seluruh atribut tersebut digunakan dalam proses normalisasi maupun analisis PCA.

Variabel *epoch_ms*, tanggal, dan jam merupakan representasi waktu mentah yang digunakan pada tahap rekayasa fitur untuk menghasilkan variabel temporal baru seperti *hour_sin*, *hour_cos*, dan *dayofweek*. Setelah proses transformasi dilakukan, variabel waktu mentah tersebut tidak lagi digunakan dalam tahap analisis karena informasi temporalnya telah direpresentasikan dalam bentuk fitur numerik yang lebih sesuai untuk analisis statistik.

Selain itu, variabel *id* hanya berfungsi sebagai penanda unik setiap kejadian petir dan tidak memiliki makna matematis dalam proses analisis, sedangkan

variabel *type* bersifat kategorikal sehingga tidak digunakan dalam pembentukan komponen utama pada PCA. Oleh karena itu, variabel yang digunakan dalam tahap normalisasi dan analisis PCA hanya mencakup variabel numerik yang merepresentasikan karakteristik spasial dan temporal kejadian petir, yaitu *latitude*, *longitude*, *hour_sin*, *hour_cos*, dan *dayofweek*.

Setelah proses seleksi variabel dilakukan, tahap selanjutnya adalah normalisasi data menggunakan metode *Standard Scaler*. Metode ini bertujuan untuk menyamakan skala seluruh variabel dengan cara mengubah setiap variabel sehingga memiliki nilai rata-rata (*mean*) sebesar 0 dan standar deviasi sebesar 1. Dengan demikian, setiap variabel memiliki kontribusi yang seimbang dalam proses pembentukan komponen utama.

Pada penelitian ini, variabel yang distandarisasi meliputi *latitude*, *longitude*, *hour_sin*, *hour_cos*, dan *dayofweek*. Pemilihan variabel tersebut didasarkan pada kebutuhan analisis *Principal Component Analysis* (PCA) yang bekerja pada variabel numerik dan sensitif terhadap perbedaan skala antar variabel. Variabel *latitude* dan *longitude* merepresentasikan dimensi spasial kejadian petir, sedangkan *hour_sin*, *hour_cos*, dan *dayofweek* merupakan representasi numerik dari transformasi waktu yang digunakan untuk menangkap pola temporal secara periodik. Oleh karena itu, variabel-variabel tersebut digunakan sebagai input dalam proses normalisasi dan analisis PCA.

Tabel 4.6 Hasil Normalisasi Data

No	Latitude	Longitude	Hour_sin	Hour_cos	Dayofweek
1	0.6692	0.3893	-0.7319	-0.3722	-0.7569
2	-0.1097	-0.0199	-0.7319	-0.3722	-0.7569
3	-1.8444	0.0754	-0.7319	-0.3722	-0.7569
4	-0.2663	-0.2194	-0.7319	-0.3722	-0.7569
5	-2.2768	0.1204	-0.7319	-0.3722	-0.7569

Berdasarkan Tabel 4.6, terlihat bahwa seluruh variabel telah berada pada skala yang sama setelah dilakukan proses standarisasi. Nilai negatif menunjukkan bahwa data berada di bawah rata-rata variabel, sedangkan nilai positif menunjukkan bahwa data berada di atas rata-rata. Dengan kondisi ini, tidak terdapat variabel yang memiliki dominasi skala tertentu, sehingga analisis PCA dapat dilakukan secara lebih objektif dan akurat. Hasil standarisasi data ini selanjutnya digunakan sebagai input utama dalam proses *Principal Component Analysis* (PCA) untuk mereduksi dimensi data dan mengidentifikasi pola dominan pada kejadian petir berbasis spasial-temporal.

4.4 Hasil Analisis *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) merupakan metode analisis yang digunakan untuk mereduksi dimensi data dengan cara mentransformasikan sejumlah variabel awal yang saling berkorelasi menjadi sekumpulan variabel baru yang saling tidak berkorelasi, yang disebut sebagai komponen utama (*principal components*). Setiap komponen utama dibentuk sebagai kombinasi linier dari variabel asal dan diurutkan berdasarkan besarnya *varians* yang dapat dijelaskan. Dengan demikian, PCA memungkinkan penyederhanaan struktur data tanpa menghilangkan informasi penting yang terkandung di dalamnya.

Pada penelitian ini, penerapan PCA pada data kejadian petir yang telah melalui proses normalisasi menggunakan metode *Standard Scaler*. Penerapan normalisasi bertujuan untuk memastikan bahwa perbedaan skala antar variabel tidak memengaruhi pembentukan komponen utama, karena PCA sangat sensitif terhadap besarnya *varians* data. Analisis PCA dilakukan untuk mengidentifikasi pola spasial-temporal kejadian petir berdasarkan kombinasi informasi lokasi dan waktu kejadian.

PCA dilakukan terhadap lima variabel input, yaitu *latitude*, *longitude*, *hour_sin*, *hour_cos*, dan *dayofweek*, yang masing-masing merepresentasikan aspek spasial dan temporal kejadian petir. Seluruh variabel tersebut telah berada pada skala yang sama dan bersifat numerik kontinu, sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses pembentukan komponen utama. Hasil transformasi PCA selanjutnya digunakan untuk menganalisis struktur *varians* data dan sebagai dasar dalam proses pengelompokan kejadian petir berbasis spasial-temporal.

4.4.1 Menentukan Nilai Eigen dan Rasio Varians

Nilai eigen dan rasio *varians* merupakan dua ukuran utama yang digunakan untuk mengevaluasi kontribusi setiap komponen utama terhadap keseluruhan variasi data. Kedua ukuran ini menjadi dasar dalam menentukan jumlah komponen utama yang dipertahankan pada proses analisis. Secara matematis, PCA dilakukan dengan menghitung matriks kovarians dari data hasil normalisasi, kemudian mencari nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari matriks tersebut. Nilai eigen merepresentasikan besarnya *varians* data yang dapat dijelaskan

oleh masing-masing komponen utama. Semakin besar nilai eigen suatu komponen, semakin besar pula informasi varians yang terkandung pada komponen tersebut.

Nilai rasio varians diperoleh dengan membagi nilai eigen dari setiap komponen utama dengan total keseluruhan nilai eigen. Rasio varians menunjukkan proporsi varians yang dijelaskan oleh masing-masing komponen utama terhadap total varians data. Selain itu, nilai *cumulative variance* merupakan penjumlahan bertahap dari rasio varians yang digunakan untuk mengetahui seberapa besar total varians yang dapat dijelaskan oleh sejumlah komponen utama pertama.

Nilai eigen dan rasio varians digunakan untuk menentukan jumlah komponen utama yang optimal dalam analisis PCA. Komponen utama dengan nilai eigen yang besar dan rasio varians yang tinggi dipertahankan karena mampu merepresentasikan sebagian besar informasi dalam data, sedangkan komponen dengan kontribusi varians yang kecil dapat diabaikan. Dengan demikian, analisis PCA dapat menyederhanakan dimensi data tanpa kehilangan informasi varians yang signifikan.

Tabel 4.7 Nilai Eigen dan Rasio Varians Hasil PCA

Komponen	Eigenvalue	Variance Ratio	Cumulative Variance
PC1	2.056387179	0.411221304	0.411221304
PC2	1.213858085	0.242738483	0.653959787
PC3	1.040044151	0.207980441	0.861940228
PC4	0.550912442	0.11016745	0.972107678
PC5	0.139480645	0.027892322	1

Tabel 4.7 diperoleh dari hasil penerapan *Principal Component Analysis* (PCA) terhadap data hasil normalisasi menggunakan *Standard Scaler*. Proses PCA diawali dengan pembentukan matriks data terstandarisasi, kemudian dihitung matriks kovarians yang merepresentasikan hubungan antar variabel. Dari matriks

kovarians tersebut, dilakukan dekomposisi eigen untuk memperoleh pasangan nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*).

Setiap komponen utama (PC) dibentuk sebagai kombinasi linier dari variabel input berdasarkan vektor eigen yang bersesuaian. Nilai eigen pada setiap PC menunjukkan besarnya varians data yang dapat dijelaskan oleh komponen utama tersebut. Karena jumlah variabel input pada penelitian ini adalah lima, maka dihasilkan lima komponen utama (PC1 hingga PC5).

Komponen utama pertama (PC1) memiliki nilai *eigenvalue* sebesar 2.0564, yang merupakan nilai eigen terbesar hasil dekomposisi matriks kovarians. Nilai ini menunjukkan bahwa PC1 mampu menjelaskan varians sebesar 2.0564 dari total varians data. Rasio varians PC1 sebesar 0.4112 diperoleh dengan membagi nilai eigen PC1 dengan total *eigenvalue* (≈ 5), sehingga PC1 menjelaskan sekitar 41.12% dari keseluruhan varians data. Hal ini menunjukkan bahwa PC1 merupakan komponen utama yang paling dominan dalam merepresentasikan pola data petir.

Komponen utama kedua (PC2) memiliki nilai *eigenvalue* sebesar 1.2139 dengan rasio varians sebesar 0.2427, yang berarti PC2 mampu menjelaskan sekitar 24,27% varians data. Jika kontribusi PC1 dan PC2 digabungkan, nilai *cumulative variance* mencapai 65,40%, yang menunjukkan bahwa dua komponen utama pertama telah mampu merepresentasikan lebih dari setengah variasi data petir.

Komponen utama ketiga (PC3) memiliki nilai *eigenvalue* sebesar 1.0400 dan rasio varians sebesar 0.2080, sehingga PC3 menjelaskan sekitar 20.80% varians data. Dengan menambahkan kontribusi PC3, nilai *cumulative variance* meningkat menjadi 86,19%, yang berarti tiga komponen utama pertama sudah cukup

representatif dalam menjelaskan sebagian besar variasi data petir berbasis spasial-temporal.

Komponen utama keempat (PC4) memiliki nilai *eigenvalue* sebesar 0.5509 dengan *variance ratio* sebesar 11.02%. Kontribusi varians PC4 relatif lebih kecil dibandingkan tiga komponen utama sebelumnya. Namun, penambahan PC4 meningkatkan nilai *cumulative variance* menjadi 97.21%, yang menunjukkan bahwa hampir seluruh varians data telah terwakili.

Komponen utama kelima (PC5) memiliki nilai *eigenvalue* terkecil, yaitu sebesar 0.1395, dengan *variance ratio* sebesar 2.79%. Kontribusi PC5 terhadap total varians data sangat kecil, sehingga komponen ini mengandung informasi tambahan yang terbatas dan umumnya tidak dipertahankan dalam analisis lanjutan.

4.4.2 Penentuan Komponen Utama Berdasarkan *Faktor Loading*

Setelah jumlah komponen utama ditentukan berdasarkan nilai *eigenvalue* dan *cumulative variance*, tahap selanjutnya adalah menentukan komponen utama. Penentuan ini dilakukan dengan menganalisis *faktor loading*, yaitu nilai yang menunjukkan besarnya kontribusi setiap variabel awal terhadap pembentukan suatu komponen utama.

Faktor loading diperoleh dari hasil dekomposisi PCA, yang merepresentasikan koefisien variabel asal pada masing-masing komponen utama. Secara matematis, *faktor loading* menunjukkan korelasi antara variabel asli dengan komponen utama. Nilai *faktor loading* yang besar, baik bernilai positif maupun negatif, menunjukkan bahwa variabel tersebut memiliki pengaruh yang kuat

terhadap komponen utama terkait, sedangkan nilai yang mendekati nol menunjukkan kontribusi yang lemah.

Pada penelitian ini, interpretasi komponen utama difokuskan pada komponen-komponen dengan kontribusi varians terbesar, yaitu PC1, PC2, dan PC3, karena ketiga komponen tersebut telah mampu menjelaskan sebagian besar variasi data petir. Dengan menganalisis *faktor loading* pada setiap komponen utama, dapat diketahui apakah suatu komponen lebih merepresentasikan aspek spasial, temporal, atau kombinasi keduanya.

Berdasarkan analisis *faktor loading*, komponen utama pertama (PC1) cenderung dipengaruhi oleh variabel-variabel temporal, sedangkan komponen utama berikutnya menunjukkan kontribusi yang lebih besar dari variabel spasial. Hal ini menunjukkan bahwa PCA mampu memisahkan pengaruh waktu dan lokasi kejadian petir ke dalam komponen utama yang berbeda, sehingga pola spasial-temporal kejadian petir dapat dianalisis secara lebih terstruktur.

Tabel 4.8 *Faktor Loading* Variabel terhadap Komponen Utama

Variabel	PC1	PC2	PC3	PC4	PC5
<i>latitude</i>	-0.201731242	0.860434052	-0.191358838	-0.42632647	-0.026866034
<i>longitude</i>	-0.245785703	-0.093366047	0.923639945	-0.278505947	0.018230708
<i>hour_sin</i>	0.746552583	0.451994675	0.335109528	0.290574553	-0.204363863
<i>hour_cos</i>	-0.712669819	0.482479856	0.181377268	0.452156148	0.148689349
<i>dayofweek</i>	0.943420396	0.166457532	0.071547804	-0.052094794	0.273044669

Tabel 4.8 menunjukkan nilai *faktor loading* masing-masing variabel terhadap lima komponen utama yang dihasilkan melalui analisis *Principal Component Analysis* (PCA). Nilai faktor loading merepresentasikan tingkat kontribusi setiap variabel terhadap pembentukan suatu komponen utama. Semakin besar nilai *faktor*

loading, semakin besar pula pengaruh variabel tersebut dalam menentukan karakteristik komponen utama yang bersangkutan.

Komponen utama pertama (PC1) didominasi oleh variabel *dayofweek* dengan nilai *loading* sebesar 0.9434, serta variabel *hour_sin* (0.7466) dan *hour_cos* (-0.7127). Secara struktural, PC1 memiliki *faktor loading* terbesar pada variabel *dayofweek*, *hour_sin*, dan *hour_cos*, yang menunjukkan bahwa komponen ini memuat informasi temporal. Namun, pada dataset ini nilai variabel temporal relatif seragam, sehingga kontribusinya dalam membedakan kejadian petir menjadi terbatas. Oleh karena itu, variasi PC1 secara praktis lebih dipengaruhi oleh perbedaan lokasi kejadian petir dibandingkan oleh faktor waktu.

Komponen utama kedua (PC2) memiliki kontribusi paling besar dari variabel *latitude* dengan nilai *loading* sebesar 0.8604, disertai kontribusi moderat dari *hour_sin* (0.4520) dan *hour_cos* (0.4825). Hal ini menunjukkan bahwa PC2 merepresentasikan pola spasial lintang yang masih dipengaruhi oleh variasi waktu kejadian. Dengan demikian, PC2 dapat diinterpretasikan sebagai komponen yang menangkap hubungan antara lokasi geografis lintang dan waktu kejadian petir.

Komponen utama ketiga (PC3) didominasi oleh variabel *longitude* dengan nilai *loading* sebesar 0.9236, yang menunjukkan bahwa PC3 terutama merepresentasikan pola spasial bujur kejadian petir. Kontribusi variabel temporal pada PC3 relatif kecil dibandingkan kontribusi *longitude*, sehingga variasi spasial arah timur barat menjadi karakteristik utama komponen ini.

Komponen utama keempat (PC4) menunjukkan kontribusi yang relatif seimbang antara variabel *hour_cos* (0.4522), *latitude* (-0.4263), dan *hour_sin*

(0.2906). Nilai loading yang tidak terlalu dominan pada satu variabel tertentu menunjukkan bahwa PC4 merepresentasikan pola variasi campuran antara aspek spasial dan temporal, dengan kontribusi varians yang lebih kecil dibandingkan komponen utama sebelumnya.

Komponen utama kelima (PC5) memiliki nilai *loading* terbesar pada variabel *dayofweek* (0.2730), namun secara keseluruhan nilai *loading* pada PC5 relatif kecil. Hal ini sejalan dengan nilai *eigenvalue* dan rasio varians PC5 yang rendah, sehingga komponen ini hanya mengandung informasi tambahan yang terbatas dan tidak digunakan dalam analisis lanjutan.

Secara keseluruhan, hasil *faktor loading* menunjukkan bahwa PC1 didominasi oleh variabel temporal, PC2 dan PC3 merepresentasikan variasi spasial, sedangkan PC4 dan PC5 memiliki kontribusi varians yang lebih kecil. Temuan ini menunjukkan bahwa PCA berhasil memisahkan pola spasial dan temporal kejadian petir ke dalam komponen utama yang berbeda, sehingga analisis pola spasial-temporal dapat dilakukan secara lebih terstruktur.

4.4.3 Karakterisasi Komponen Utama Berdasarkan *Faktor Loading*

Berdasarkan hasil analisis *Principal Component Analysis* (PCA), lima komponen utama (PC1 hingga PC5) memiliki nilai eigen, rasio varians, dan *faktor loading* yang berbeda-beda. Pada setiap komponen utama menjelaskan dimensi variasi data petir yang berbeda, baik dari aspek temporal maupun spasial.

Tabel 4.9 Karakterisasi Komponen Utama Hasil PCA

Komponen	Nilai Eigen	Rasio Varians	Variabel Dominan (<i>Faktor Loading</i> Terbesar)	Interpretasi Komponen
PC1	2.0564	0.4112	<i>dayofweek</i> (0.9434), <i>hour_sin</i> (0.7466), <i>hour_cos</i> (-0.7127)	Kombinasi fitur temporal, namun kontribusi aktual terbatas karena variasi waktu rendah
PC2	1.2139	0.2427	<i>latitude</i> (0.8604), <i>hour_cos</i> (0.4825), <i>hour_sin</i> (0.4520)	Variasi spasial berdasarkan lintang dengan pengaruh temporal
PC3	1.0400	0.2080	<i>longitude</i> (0.9236)	Variasi spasial arah bujur (timur, barat)
PC4	0.5509	0.1102	<i>hour_cos</i> (0.4522), <i>latitude</i> (-0.4263), <i>hour_sin</i> (0.2906)	Pola variasi campuran spasial temporal
PC5	0.1395	0.0279	<i>dayofweek</i> (0.2730)	Variasi tambahan dengan kontribusi sangat kecil

Tabel 4.9 menyajikan karakterisasi komponen utama hasil *Principal Component Analysis* (PCA) berdasarkan nilai eigen, rasio varians, serta variabel dominan yang memiliki *faktor loading* terbesar pada masing-masing komponen. Tabel ini bertujuan untuk mempermudah pemahaman terhadap makna setiap komponen utama tanpa harus menelaah seluruh nilai *faktor loading* secara terpisah.

Komponen utama pertama (PC1) memiliki nilai eigen dan rasio varians terbesar, sehingga merupakan komponen paling dominan dalam menjelaskan variasi data. Variabel dominan pada PC1 adalah *dayofweek*, *hour_sin*, dan *hour_cos*, yang menunjukkan bahwa komponen ini terutama merepresentasikan pola temporal kejadian petir, khususnya variasi kejadian berdasarkan hari dan jam. Hal ini mengindikasikan bahwa faktor waktu memiliki pengaruh yang sangat kuat terhadap struktur variasi data petir.

Komponen utama kedua (PC2) didominasi oleh variabel *latitude* dengan kontribusi tambahan dari variabel temporal. Kondisi ini menunjukkan bahwa PC2 merepresentasikan variasi spasial berdasarkan lintang, yang masih dipengaruhi oleh

pola waktu kejadian. Dengan kata lain, distribusi kejadian petir pada arah utara selatan memiliki keterkaitan dengan waktu terjadinya peristiwa petir.

Komponen utama ketiga (PC3) memiliki variabel dominan *longitude*, yang menunjukkan bahwa PC3 merepresentasikan variasi spasial arah bujur (timur barat). Komponen ini menangkap perbedaan lokasi kejadian petir secara horizontal, terpisah dari pengaruh temporal yang lebih dominan pada PC1.

Komponen utama keempat (PC4) memperlihatkan kontribusi yang relatif seimbang antara variabel spasial dan temporal, tanpa adanya satu variabel yang sangat dominan. Hal ini menunjukkan bahwa PC4 merepresentasikan pola variasi campuran spasial temporal dengan kontribusi varians yang lebih kecil dibandingkan komponen utama sebelumnya.

Komponen utama kelima (PC5) memiliki nilai eigen dan rasio varians yang paling kecil, dengan *faktor loading* yang relatif rendah pada seluruh variabel. Hal ini menunjukkan bahwa PC5 hanya mengandung informasi tambahan yang sangat terbatas dan tidak berperan signifikan dalam menjelaskan struktur utama variasi data petir.

Secara keseluruhan, Tabel 4.9 menunjukkan bahwa sebagian besar informasi penting dalam data petir dapat dijelaskan oleh tiga komponen utama pertama (PC1-PC3), yang masing-masing merepresentasikan pola temporal serta variasi spasial lintang dan bujur. Komponen utama selanjutnya hanya memberikan kontribusi pendukung dan tidak menjadi fokus utama dalam analisis lanjutan.

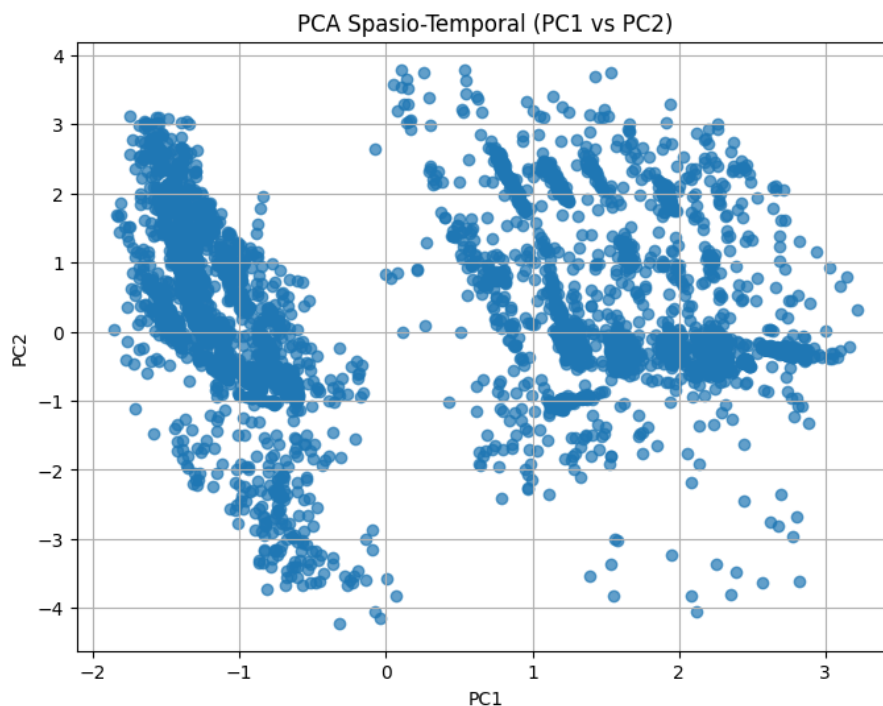
4.5 Visualisasi Data Hasil PCA

Visualisasi hasil PCA dilakukan untuk memberikan gambaran sebaran data petir pada ruang komponen utama. Melalui visualisasi ini, pola spasial-temporal kejadian petir dapat diamati secara lebih detail berdasarkan kombinasi komponen utama yang telah diperoleh.

4.5.1 Visualisasi PCA Dua Dimensi (PC1 Vs PC2)

Visualisasi hasil *Principal Component Analysis* (PCA) dua dimensi dilakukan untuk memberikan gambaran mengenai sebaran data petir setelah proses reduksi dimensi. Visualisasi ini bertujuan untuk menampilkan distribusi kejadian petir pada ruang komponen utama serta mengamati pola spasial-temporal yang terbentuk berdasarkan dua komponen utama dengan kontribusi varians terbesar, yaitu PC1 dan PC2.

Pemilihan PC1 dan PC2 didasarkan pada nilai rasio varians tertinggi yang dimiliki oleh kedua komponen tersebut, sehingga keduanya mampu menjelaskan sebagian besar variasi pada dataset. Oleh karena itu, visualisasi dua dimensi ini digunakan untuk mengamati struktur distribusi data, kecenderungan pengelompokan, serta pola hubungan antar kejadian petir sebelum dilakukan analisis lanjutan seperti *clustering*.



Gambar 4.2 Visualisasi Sebaran Data Petir pada PCA 2D

Berdasarkan Gambar 4.2, terlihat bahwa sebaran data pada ruang PC1-PC2 membentuk beberapa area dengan tingkat kepadatan yang berbeda. Pemisahan data paling jelas terlihat pada sumbu PC1, di mana titik-titik data cenderung terbagi menjadi dua kelompok besar pada sisi kiri dan kanan grafik. Hal ini menunjukkan bahwa PC1 memiliki kontribusi paling dominan dalam membedakan karakteristik kejadian petir dalam dataset, sehingga komponen ini menjadi faktor utama dalam menjelaskan variasi data.

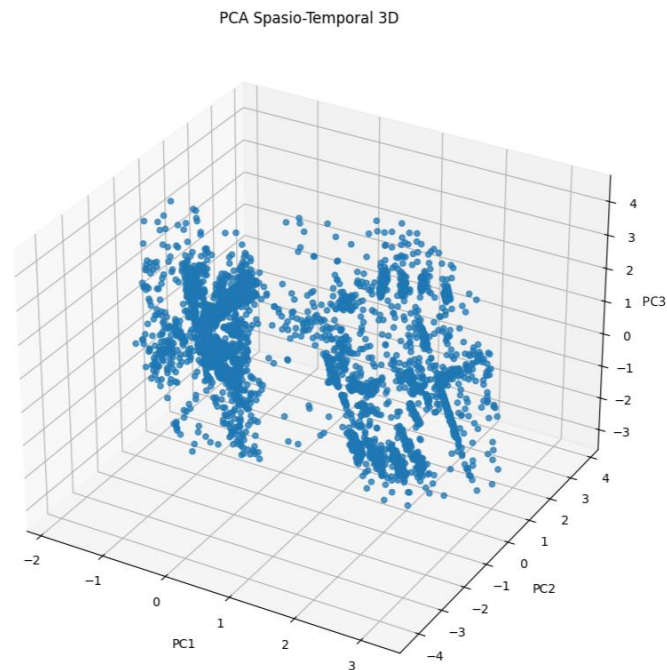
Sementara itu, PC2 menunjukkan variasi tambahan yang memperlihatkan perbedaan distribusi data secara vertikal pada grafik. Variasi pada sumbu ini menggambarkan adanya perbedaan karakteristik kejadian petir yang tidak sepenuhnya dijelaskan oleh PC1, sehingga PC2 memberikan informasi tambahan terkait variasi spasial kejadian petir.

Melalui visualisasi ini dapat diamati bahwa data tidak tersebar secara acak, melainkan menunjukkan kecenderungan membentuk beberapa kelompok pada ruang PCA. Pola sebaran tersebut mengindikasikan adanya struktur alami pada data petir, yang selanjutnya menjadi dasar dalam proses pengelompokan data menggunakan metode *clustering* pada tahap analisis berikutnya.

4.5.2 Visualisasi PCA Tiga Dimensi

Selain visualisasi dua dimensi, juga dilakukan visualisasi hasil *Principal Component Analysis* (PCA) tiga dimensi dengan menggunakan komponen utama pertama (PC1), kedua (PC2), dan ketiga (PC3). Visualisasi tiga dimensi ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap sebaran data petir, mengingat tiga komponen utama pertama telah mampu menjelaskan sebagian besar variasi data, yaitu lebih dari 80% berdasarkan nilai *cumulative variance*.

Penggunaan visualisasi tiga dimensi memungkinkan pengamatan struktur data yang tidak sepenuhnya terlihat pada visualisasi dua dimensi. Dengan menambahkan dimensi PC3, pola sebaran data petir dapat diamati secara lebih lengkap, terutama untuk melihat variasi tambahan yang tidak tertangkap oleh kombinasi PC1 dan PC2 saja.



Gambar 4.3 Visualisasi Sebaran Data Petir pada PCA 3D

Berdasarkan Gambar 4.3, terlihat bahwa sebaran data membentuk struktur yang lebih jelas ketika diamati dalam ruang tiga dimensi. Variasi sepanjang sumbu PC1 masih terlihat paling dominan, yang menunjukkan bahwa komponen utama pertama tetap memiliki kontribusi terbesar dalam menjelaskan variasi data. Sementara itu, PC2 dan PC3 memberikan pemisahan tambahan terhadap data, yang mengindikasikan adanya variasi spasial dan temporal lanjutan yang tidak sepenuhnya terwakili pada visualisasi dua dimensi.

Kontribusi PC3, yang pada analisis *factor loading* didominasi oleh variabel *longitude*, terlihat memberikan pemisahan data pada dimensi ketiga, sehingga perbedaan lokasi kejadian petir secara arah bujur dapat diamati dengan lebih baik. Hal ini menunjukkan bahwa visualisasi tiga dimensi mampu menampilkan struktur data yang lebih representatif dibandingkan visualisasi dua dimensi, khususnya

dalam analisis pola spasial-temporal kejadian petir. Secara keseluruhan, visualisasi PCA tiga dimensi ini memperkuat hasil analisis sebelumnya dan menjadi dasar yang lebih kuat untuk tahap analisis lanjutan, yaitu pengelompokan (*clustering*) kejadian petir berbasis komponen utama.

4.6 Hasil *Clustering* Spasial Temporal Petir Berbasis PCA

Tahap selanjutnya adalah pengelompokan kejadian petir berdasarkan kemiripan pola spasial-temporal yang terbentuk. Pada penelitian ini, penentuan jumlah *cluster* didasarkan pada pola distribusi nilai komponen utama pertama (PC1), yang menunjukkan adanya kecenderungan pemisahan data secara alami ke dalam beberapa kelompok. Pola tersebut mengindikasikan keberadaan batas-batas alami (*natural breaks*) yang dapat digunakan sebagai dasar pengelompokan kejadian petir. Selain itu, penentuan jumlah *cluster* juga mempertimbangkan temuan pada penelitian terdahulu yang menyatakan bahwa pengelompokan data klimatologi ke dalam tiga kelompok utama telah cukup representatif untuk menggambarkan variasi kondisi lingkungan secara umum. Pembagian tersebut umumnya merepresentasikan kondisi dengan intensitas rendah, sedang, dan tinggi. Dengan pendekatan ini, struktur *cluster* yang dihasilkan diharapkan tetap sederhana, namun mampu mencerminkan karakteristik utama variasi data petir secara spasial-temporal.

4.6.1 Hasil *Clustering* Data Petir

Hasil *clustering* menunjukkan bahwa data petir pada periode pengamatan dapat dikelompokkan ke dalam tiga *cluster* berdasarkan nilai PC1. Distribusi jumlah data pada masing-masing *cluster* ditunjukkan pada Tabel 4.10.

Pengelompokan ini didasarkan pada nilai PC1 yang, berdasarkan hasil analisis PCA sebelumnya, didominasi oleh variabel temporal yaitu *dayofweek*, *hour_sin*, dan *hour_cos*. Oleh karena itu, cluster yang terbentuk tidak hanya mencerminkan perbedaan spasial lokasi kejadian petir, tetapi juga merepresentasikan perbedaan pola waktu kejadian petir.

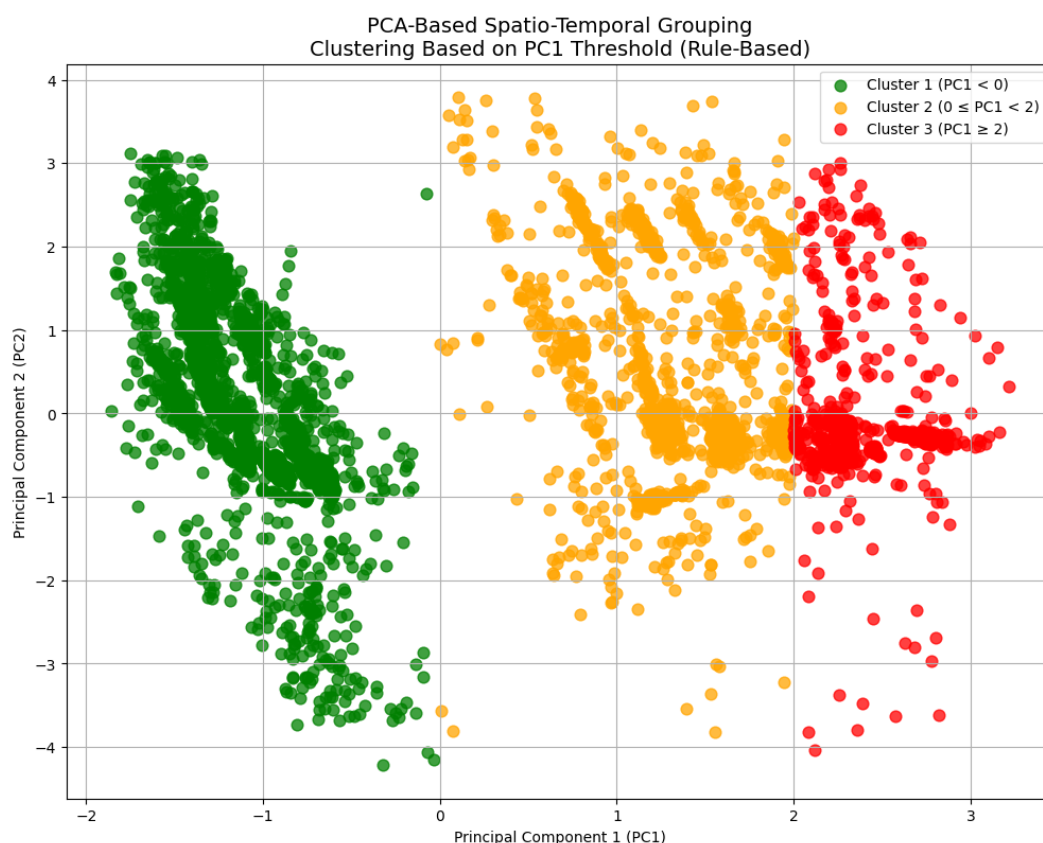
Tabel 4.10 Distribusi Jumlah Data pada Setiap Cluster

Cluster	Kriteria	Jumlah Data
Cluster 1	$PC1 < 0$	4.657
Cluster 2	$0 \leq PC1 < 2$	1.542
Cluster 3	$PC1 \geq 2$	1.128

Pada Tabel 4.10 menunjukkan bahwa data petir pada periode pengamatan terbagi ke dalam tiga *cluster* berdasarkan nilai PC1, yaitu *Cluster 1* ($PC1 < 0$), *Cluster 2* ($0 \leq PC1 < 2$), dan *Cluster 3* ($PC1 \geq 2$). Distribusi jumlah data pada masing-masing cluster ditunjukkan pada Tabel 4.8. Berdasarkan tabel tersebut, *Cluster 1* memiliki jumlah data paling besar, yaitu sebanyak 4.657 kejadian, diikuti oleh *Cluster 2* dengan 1.542 kejadian, dan *Cluster 3* dengan 1.128 kejadian. Perbedaan jumlah data ini menunjukkan bahwa sebagian besar kejadian petir berada pada rentang nilai PC1 negatif, yang mencerminkan dominasi pola tertentu pada komponen utama pertama.

Perbedaan nilai PC1 pada setiap *cluster* menunjukkan adanya variasi karakteristik kejadian petir, khususnya dari aspek temporal yang direpresentasikan oleh PC1. Hal ini menunjukkan bahwa setiap cluster memiliki kecenderungan pola waktu kejadian petir yang berbeda, baik dalam hal frekuensi kemunculan maupun distribusi waktunya.

Cluster 1 mencerminkan kelompok kejadian petir dengan pola temporal yang paling dominan selama periode pengamatan, sedangkan *Cluster 2* merepresentasikan kelompok transisi dengan pola waktu yang lebih bervariasi. *Cluster 3* menunjukkan kelompok kejadian petir dengan nilai PC1 tinggi yang jumlahnya relatif lebih sedikit, sehingga mencerminkan pola kejadian pada waktu-waktu tertentu yang lebih jarang terjadi. Untuk memperjelas hasil pengelompokan, dilakukan visualisasi *clustering* pada ruang PCA dua dimensi menggunakan komponen utama pertama (PC1) dan kedua (PC2), sebagaimana ditunjukkan pada Gambar 4.3.



Gambar 4.4 Visualisasi Hasil *Clustering*

Berdasarkan Gambar 4.3, terlihat bahwa masing-masing cluster menempati wilayah yang berbeda pada ruang PCA, dengan pemisahan yang cukup jelas sepanjang sumbu PC1. *Cluster 1* terkonsentrasi pada sisi PC1 negatif, *Cluster 2* berada pada rentang nilai PC1 menengah, sedangkan *Cluster 3* berada pada sisi PC1 positif. Pola ini menunjukkan bahwa pengelompokan berbasis ambang batas PC1 mampu memisahkan data petir secara konsisten sesuai dengan struktur variasi utama yang terbentuk dari hasil PCA, yang dalam hal ini didominasi oleh pola temporal kejadian petir.

Meskipun demikian, terdapat beberapa titik data yang secara visual terlihat berada dekat dengan *cluster* lain pada ruang PC1-PC2. Hal ini terjadi karena proses pengelompokan dalam penelitian ini ditentukan berdasarkan ambang nilai PC1 saja. Oleh karena itu, penentuan *cluster* bergantung pada posisi nilai PC1 masing-masing data terhadap batas yang telah ditentukan. Data yang berada sangat dekat dengan batas tersebut tetap dapat masuk ke *cluster* yang berbeda apabila berada pada sisi ambang PC1 yang berlainan.

Secara keseluruhan, hasil clustering berbasis PCA ini menunjukkan bahwa komponen utama pertama memiliki peran yang sangat dominan dalam membedakan pola kejadian petir. Struktur *cluster* yang terbentuk bersifat jelas dan mudah dipahami, sehingga dapat digunakan untuk mendukung analisis pola spasial-temporal kejadian petir secara lebih sistematis.

Untuk memperkuat interpretasi spasial dari hasil pengelompokan tersebut, dilakukan analisis terhadap beberapa sampel data pada masing-masing *cluster* dengan memperhatikan koordinat geografis berupa latitude dan longitude.

Identifikasi lokasi wilayah dilakukan berdasarkan posisi koordinat tersebut untuk mengetahui kecenderungan distribusi kejadian petir pada setiap *cluster* dalam wilayah penelitian. Dengan demikian, hasil *clustering* tidak hanya menunjukkan pemisahan data secara statistik pada ruang PCA, tetapi juga mencerminkan pola distribusi geografis kejadian petir pada wilayah penelitian. Meskipun dasar pengelompokan didominasi oleh aspek temporal, interpretasi spasial tetap diperlukan untuk memahami distribusi geografis dari setiap pola waktu kejadian petir. Dengan demikian, analisis pada masing-masing *cluster* dilakukan dengan mengaitkan antara karakteristik temporal dan lokasi kejadian petir.

Tabel 4.11 Cuplikan Data *Cluster 1* dan Identifikasi Wilayah

No	Latitude	Longitude	PC1	Cluster	Keterangan Wilayah
1	-7.486379	112.201562	-0.681	Cluster 1	Pasuruan, Jawa Timur
2	-7.606525	112.26663	-0.6757	Cluster 1	Probolinggo, Jawa Timur
3	-8.451237	114.102567	-0.7342	Cluster 1	Selat Bali
4	-8.623451	114.356782	-0.8121	Cluster 1	Bali Barat
5	-8.754321	114.892345	-0.7934	Cluster 1	Laut Bali
6	-8.934567	115.102345	-0.8451	Cluster 1	Bali Selatan
7	-8.712345	115.321456	-0.8012	Cluster 1	Laut Bali-arah NTB Barat

Berdasarkan Tabel 4.11, terlihat bahwa titik-titik pada *Cluster 1* memiliki nilai *longitude* yang relatif lebih besar dibandingkan *cluster* lainnya, yaitu berada pada rentang sekitar 112.2° hingga 115.3° BT. Rentang koordinat tersebut menunjukkan bahwa lokasi kejadian petir pada cluster ini berada pada wilayah bagian timur dari area penelitian. Identifikasi wilayah berdasarkan koordinat latitude dan longitude menunjukkan bahwa beberapa titik berada pada wilayah daratan Pulau Jawa, seperti Pasuruan dan Probolinggo di Provinsi Jawa Timur. Selain itu, beberapa titik lainnya berada pada wilayah yang lebih ke arah timur,

seperti Selat Bali, Bali Barat, Laut Bali, serta wilayah perairan yang mengarah ke Nusa Tenggara Barat.

Keberadaan titik-titik tersebut menunjukkan bahwa distribusi kejadian petir pada *Cluster 1* dominan berada pada wilayah timur area penelitian yang mencakup Jawa Timur, Bali, serta perairan di sekitarnya. Oleh karena itu, *Cluster 1* dapat diinterpretasikan sebagai zona aktivitas petir wilayah timur dalam distribusi spasial kejadian petir pada area pengamatan. Selain menunjukkan dominasi pada wilayah timur, *cluster* ini juga mencerminkan pola temporal yang paling sering muncul, sehingga dapat diinterpretasikan sebagai kelompok kejadian petir yang tidak hanya dominan secara spasial, tetapi juga memiliki frekuensi kejadian tinggi pada waktu-waktu tertentu.

Tabel 4.12 Cuplikan Data *Cluster 2* dan Identifikasi Wilayah

No	Latitude	Longitude	PC1	Cluster	Keterangan Wilayah
1	- 6.742315	111.984512	1.2456	Cluster 2	Laut Jawa-selatan Kalimantan
2	- 7.103214	111.875463	1.2789	Cluster 2	Rembang, Jawa Tengah
3	- 7.215643	111.954327	1.3012	Cluster 2	Tuban, Jawa Timur
4	- 6.564321	111.845236	1.3425	Cluster 2	Laut Jawa-selatan Kalimantan Tengah
5	- 7.332145	111.904321	1.2876	Cluster 2	Bojonegoro, Jawa Timur
6	- 6.754321	111.932156	1.3215	Cluster 2	Laut Jawa-tengah
7	- 6.645213	111.867453	1.2764	Cluster 2	Laut Jawa-selatan Kalimantan

Berdasarkan Tabel 4.12, terlihat bahwa titik-titik pada *Cluster 2* memiliki nilai *longitude* yang berada pada rentang sekitar 111.8° hingga 112.0° BT. Rentang koordinat tersebut menunjukkan posisi wilayah yang berada pada bagian tengah dari area penelitian. Identifikasi wilayah berdasarkan koordinat *latitude* dan *longitude* menunjukkan bahwa beberapa titik berada pada wilayah daratan Pulau

Jawa, seperti Rembang di Jawa Tengah, Tuban dan Bojonegoro di Jawa Timur. Selain itu, beberapa titik lainnya berada pada wilayah perairan Laut Jawa yang terletak di selatan Kalimantan, yang merupakan wilayah peralihan antara Pulau Jawa dan Pulau Kalimantan.

Keberadaan titik-titik pada wilayah daratan Jawa serta perairan Laut Jawa menunjukkan bahwa distribusi kejadian petir pada *Cluster 2* tidak terfokus pada satu wilayah tertentu, melainkan berada pada wilayah tengah area penelitian yang menghubungkan zona barat dan zona timur. Oleh karena itu, *Cluster 2* dapat diinterpretasikan sebagai zona tengah atau wilayah transisi dalam distribusi spasial kejadian petir pada area pengamatan. Dari sisi temporal, *cluster* ini menunjukkan pola kejadian yang lebih menyebar dan tidak terlalu dominan pada waktu tertentu, sehingga mencerminkan karakter transisi baik secara spasial maupun temporal.

Tabel 4.13 Cuplikan Data Cluster 3 dan Identifikasi Wilayah

No	Latitude	Longitude	PC1	Cluster	Keterangan Wilayah
1	- 8.203145	110.742156	2.4512	Cluster 3	Samudra Hindia, selatan Yogyakarta
2	- 7.982314	110.623451	2.3765	Cluster 3	Samudra Hindia, selatan Jawa Tengah
3	- 7.653421	110.184532	2.4681	Cluster 3	Purworejo, Jawa Tengah
4	- 7.903214	110.052341	2.5364	Cluster 3	Samudra Hindia, selatan Jawa
5	- 6.734521	109.845236	2.4923	Cluster 3	Laut Jawa, barat Jawa Tengah
6	- 5.852281	108.5414845	2.6844	Cluster 3	Laut Jawa, utara Indramayu/Cirebon (Jawa Barat)
7	- 3.942413	114.0512677	2.8276	Cluster 3	Kalimantan Selatan (Banjar / Banjarmasin)

Berdasarkan Tabel 4.13, terlihat bahwa titik-titik pada *Cluster 3* memiliki nilai *longitude* yang relatif lebih kecil dibandingkan *cluster* lainnya, yaitu berada pada rentang sekitar 108° hingga 110° BT, yang menunjukkan bahwa lokasi kejadian petir pada *cluster* ini berada pada wilayah bagian barat dari area penelitian.

Identifikasi wilayah berdasarkan koordinat *latitude* dan *longitude* menunjukkan bahwa beberapa titik berada pada wilayah perairan Samudra Hindia di selatan Pulau Jawa, seperti wilayah selatan Yogyakarta dan selatan Jawa Tengah. Selain itu, terdapat juga titik yang berada pada wilayah daratan seperti Purworejo di Jawa Tengah, serta wilayah perairan Laut Jawa bagian barat di utara Jawa Tengah dan Jawa Barat.

Selain wilayah di sekitar Pulau Jawa, terdapat pula titik yang berada pada wilayah Kalimantan Selatan, seperti daerah sekitar Banjar dan Banjarmasin. Hal ini menunjukkan bahwa distribusi kejadian petir pada *cluster* ini mencakup wilayah barat dari area penelitian yang meliputi daratan dan perairan di sekitar Pulau Jawa serta wilayah selatan Kalimantan. Oleh karena itu, berdasarkan distribusi koordinat geografis tersebut, *Cluster 3* dapat diinterpretasikan sebagai zona aktivitas petir wilayah barat dalam distribusi spasial kejadian petir pada area pengamatan.

Dari sisi temporal, *cluster* ini mencerminkan kejadian petir dengan pola waktu yang lebih spesifik dan frekuensi yang relatif rendah, sehingga menunjukkan karakteristik temporal yang berbeda dibandingkan dua *cluster* lainnya. Hal ini mengindikasikan bahwa kejadian petir pada *cluster* ini cenderung terjadi pada kondisi waktu tertentu yang lebih terbatas.

Berdasarkan identifikasi wilayah dari koordinat geografis pada masing-masing *cluster*, dapat disimpulkan bahwa hasil pengelompokan menunjukkan pola distribusi spasial-temporal kejadian petir yang cukup jelas pada wilayah penelitian. *Cluster 1* merepresentasikan zona aktivitas petir wilayah timur dengan dominasi kejadian pada pola waktu tertentu. *Cluster 2* menunjukkan zona tengah atau

wilayah transisi yang memiliki variasi baik secara spasial maupun temporal. Sementara itu, *Cluster 3* merepresentasikan zona wilayah barat dengan kejadian petir yang lebih spesifik dan relatif jarang terjadi secara temporal.

Dengan demikian, hasil *clustering* tidak hanya mampu memisahkan data berdasarkan lokasi geografis, tetapi juga mengungkap adanya perbedaan pola waktu kejadian petir, sehingga analisis yang dihasilkan mencerminkan keterkaitan antara aspek spasial dan temporal secara lebih komprehensif.

4.7 Evaluasi Hasil PCA-Clustering Kejadian Petir

Evaluasi dilakukan untuk menilai kualitas hasil PCA-Clustering kejadian petir yang telah diperoleh pada tahap sebelumnya. Pada penelitian ini, evaluasi dilakukan melalui dua pendekatan, yaitu evaluasi internal menggunakan *Silhouette Score* dan evaluasi berbasis keilmuan dengan mengacu pada karakteristik data kejadian petir yang digunakan oleh BMKG.

4.7.1 Evaluasi Internal Menggunakan Silhouette Score

Evaluasi internal PCA-Clustering dilakukan menggunakan *Silhouette Score*, yaitu metrik evaluasi internal yang digunakan untuk mengukur tingkat kekompakan data dalam satu *cluster* (*intra-cluster cohesion*) serta tingkat pemisahan antar *cluster* (*inter-cluster separation*). Nilai *Silhouette Score* berada pada rentang -1 hingga 1. dengan interpretasi sebagai berikut:

- a) Nilai mendekati 1 menunjukkan bahwa data berada dekat dengan *cluster*-nya sendiri dan jauh dari *cluster* lain (*cluster* sangat baik).
- b) Nilai mendekati 0 menunjukkan bahwa data berada di perbatasan antar *cluster*.

- c) Nilai negatif menunjukkan bahwa data lebih dekat dengan cluster lain dibanding *cluster*-nya sendiri.

Secara matematis, nilai *Silhouette* untuk setiap data ke- i dirumuskan sebagai berikut:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4.1)$$

Keterangan:

- a) (i) = nilai silhouette untuk data i
- b) (i) = rata-rata jarak antara data i dengan semua titik lain dalam *cluster* yang sama
- c) (i) = rata-rata jarak antara data i dengan semua titik dalam *cluster* terdekat selain *cluster*-nya
- d) $\max(a(i), b(i))$ = memilih nilai terbesar antara jarak dalam *cluster* sendiri dan cluster lain

Nilai *Silhouette Score* keseluruhan dari rata-rata seluruh (i) pada dataset.

Pada penelitian ini, perhitungan *Silhouette Score* dilakukan menggunakan nilai komponen utama pertama (PC1) sebagai representasi data, sesuai dengan pendekatan *clustering* yang didasarkan pada ambang batas nilai PC1. Hasil perhitungan menunjukkan nilai *Silhouette Score* sebesar 0.7664, yang termasuk dalam kategori sangat baik. Hal ini menunjukkan bahwa *cluster* yang terbentuk memiliki tingkat pemisahan yang jelas serta kekompakan data yang tinggi.

Sebagai contoh penerapan rumus *Silhouette Score* pada Persamaan (4.1), digunakan salah satu data dari hasil *clustering* yang memiliki nilai PC1 = -0.681 dan termasuk ke dalam *Cluster 1*. Perhitungan dilakukan menggunakan nilai PC1 sebagai representasi jarak antar data.

- a. Menghitung jarak dalam *cluster*

Misalkan terdapat beberapa data lain pada Cluster 1 dengan nilai PC1 sebagai berikut:

-0.681, -0.6757, -0.7342, -0.8121

Jarak antar data dihitung menggunakan selisih absolut:

$$-0.681 - (-0.6757) = 0.0053$$

$$-0.681 - (-0.7342) = 0.0532$$

$$-0.681 - (-0.8121) = 0.1311$$

Sehingga:

$$a(i) = (0.0053 + 0.0532 + 0.1311) / 3$$

$$a(i) = 0.0632$$

- b. Menghitung jarak ke *cluster* terdekat

Misalkan nilai PC1 dari beberapa data pada Cluster 2 adalah:

$$1.2456, 1.3012, 1.2789$$

Maka:

$$-0.681 - 1.2456 = 1.9266$$

$$-0.681 - 1.3012 = 1.9822$$

$$-0.681 - 1.2789 = 1.9599$$

Sehingga:

$$b(i) = (1.9266 + 1.9822 + 1.9599) / 3$$

$$b(i) = 1.9562$$

- c. Substitusi ke rumus *Silhouette*

Rumus (4.1):

$$S(i) = \frac{b_{(i)} - a_{(i)}}{\max\{a(i), b(i)\}}$$

$$S(i) = \frac{1.9562 - 0.0632}{1.9562}$$

$$S(i) = 0.9677$$

d. Interpretasi nilai

Nilai *Silhouette* sebesar 0.9677 mendekati nilai 1, yang menunjukkan bahwa data tersebut berada sangat dekat dengan *cluster*-nya sendiri serta memiliki jarak yang cukup jauh terhadap *cluster* lainnya.

Perhitungan yang sama dilakukan terhadap seluruh data dalam dataset menggunakan bantuan perangkat lunak Python. Nilai *Silhouette* dari seluruh data kemudian dirata-ratakan sehingga diperoleh nilai *Silhouette Score* keseluruhan sebesar 0.7664. Nilai ini menunjukkan bahwa struktur *cluster* yang terbentuk memiliki tingkat pemisahan antar *cluster* yang baik serta kekompakan data yang tinggi.

4.7.2 Evaluasi Berdasarkan Perspektif BMKG

Selain evaluasi internal, hasil *clustering* juga dievaluasi melalui validasi eksternal dengan menginterpretasikan hasil pengelompokan berdasarkan karakteristik kejadian petir menurut perspektif BMKG, khususnya ditinjau dari atribut waktu kejadian dan lokasi geografis. Evaluasi ini bertujuan untuk memastikan bahwa *cluster* yang terbentuk tidak hanya memiliki kualitas matematis yang baik, tetapi juga selaras dengan karakteristik distribusi petir yang dikenal dalam kajian geofisika dan meteorologi.

Dalam praktik operasional BMKG, kejadian petir umumnya dianalisis berdasarkan pola distribusi spasial dan temporal, karena aktivitas petir sering berkaitan dengan dinamika sistem konvektif, kondisi atmosfer, serta pola cuaca harian pada suatu wilayah. Oleh karena itu, interpretasi hasil *clustering* pada penelitian ini dilakukan dengan melihat apakah kelompok data yang terbentuk

menunjukkan konsentrasi kejadian pada wilayah tertentu atau pada rentang waktu tertentu.

Hasil pengelompokan menunjukkan bahwa cluster yang terbentuk mampu menggambarkan pola distribusi kejadian petir berdasarkan kemiripan spasial-temporal. Pada beberapa *cluster* terlihat adanya konsentrasi kejadian pada wilayah tertentu maupun pada periode waktu tertentu, yang sejalan dengan dinamika distribusi petir harian yang umum diamati dalam analisis BMKG.

Dengan demikian, hasil *clustering* dalam penelitian ini tidak merepresentasikan tipe petir (seperti CG+, CG-, atau IC), melainkan kelompok kejadian petir yang memiliki kemiripan pola distribusi spasial dan temporal. Hal ini sesuai dengan tujuan penelitian, yaitu mengidentifikasi pola distribusi kejadian petir berdasarkan struktur variasi data yang diperoleh dari hasil reduksi dimensi menggunakan PCA.

4.8 Pembahasan

Penelitian ini bertujuan untuk menganalisis pola spasial-temporal kejadian petir menggunakan pendekatan *Principal Component Analysis* (PCA) dan *PCA-Clustering*. Berdasarkan hasil pengolahan data, PCA terbukti mampu mereduksi kompleksitas variabel spasial dan temporal menjadi beberapa komponen utama yang tetap mempertahankan proporsi varians yang signifikan. Reduksi dimensi ini mempermudah proses identifikasi pola tanpa menghilangkan informasi utama yang terkandung dalam data.

Hasil PCA menunjukkan bahwa komponen utama pertama (PC1) memiliki kontribusi varians terbesar dibandingkan komponen lainnya. Secara struktural, PC1

memiliki faktor *loading* yang besar pada variabel temporal seperti *hour_sin*, *hour_cos*, dan *dayofweek*. Namun demikian, dalam konteks data harian yang dianalisis, variasi nilai temporal relatif seragam pada setiap hari, sehingga kontribusinya dalam membedakan kejadian petir menjadi terbatas. Dengan demikian, meskipun secara matematis PC1 memuat informasi temporal, pemisahan *cluster* dalam penelitian ini secara praktis lebih dipengaruhi oleh variasi spasial, yaitu perbedaan koordinat *latitude* dan *longitude*.

Proses *clustering* dilakukan berdasarkan hasil transformasi PCA sehingga diperoleh tiga *cluster* kejadian petir. *Cluster* yang terbentuk tidak merepresentasikan tipe petir, melainkan kelompok kejadian yang memiliki tingkat kemiripan spasial-temporal. Pada beberapa tanggal terlihat bahwa distribusi *cluster* didominasi oleh satu atau dua *cluster*, sedangkan *cluster* lainnya memiliki jumlah anggota yang sangat kecil atau bahkan tidak terbentuk. Kondisi ini menunjukkan bahwa struktur distribusi kejadian petir dapat berbeda setiap harinya. Untuk melihat variasi distribusi *cluster* dan kualitas pemisahan secara keseluruhan, dilakukan rekapitulasi hasil clustering untuk seluruh tanggal pada bulan Mei 2025.

Tabel 4.14 Rekapitulasi Hasil *Clustering* Harian Bulan Mei 2025

Tanggal	Total Data	Cluster 1 (PC1 < 0)	Cluster 2 (0 ≤ PC1 < 2)	Cluster 3 (PC1 ≥ 2)	Silhouette Score
5/1/2025	1.249	594	627	28	0.5033
5/2/2025	5.304	2.023	3.245	36	0.4145
5/3/2025	7.935	2.680	5.255	0	0.7163
5/4/2025	6.175	2.218	3.768	189	0.6453
5/5/2025	12.710	6.276	6.345	109	0.6080
5/6/2025	15.880	4.278	11.512	90	0.6892
5/7/2025	7.370	1.754	5.616	0	0.7548

Tanggal	Total Data	Cluster 1 (PC1 < 0)	Cluster 2 ($0 \leq$ PC1 < 2)	Cluster 3 (PC1 \geq 2)	Silhouette Score
5/8/2025	6.247	1.871	4.374	2	0.5582
5/9/2025	21.099	14.920	6.179	0	0.7026
5/10/2025	3.236	1.168	3.113	7	0.5503
5/11/2025	7.792	2.119	5.642	31	0.5724
5/12/2025	5.835	4.102	957	776	0.6963
5/13/2025	19.112	4.927	14.167	18	0.6086
5/14/2025	10.296	5.336	4.541	419	0.5241
5/15/2025	29.193	9.049	20.144	0	0.5893
5/16/2025	14.152	8.535	2.823	2.794	0.713
5/17/2025	5.867	1.843	4.024	0	0.7283
5/18/2025	13.863	4.879	8.984	0	0.7713
5/19/2025	7.862	4.967	1.943	952	0.6398
5/20/2025	2.306	1.169	1.008	129	0.5681
5/21/2025	3.209	1.490	1.552	167	0.5833
5/22/2025	6.330	1.921	4.313	96	0.5151
5/23/2025	15.276	3.824	11.424	28	0.6941
5/24/2025	9.788	4.191	4.873	724	0.7053
5/25/2025	10.222	3.895	6.327	0	0.8286
5/26/2025	4.522	2.752	1.481	289	0.5953
5/27/2025	2.719	1.506	875	338	0.6617
5/28/2025	4.037	2.073	1.812	152	0.5612
5/29/2025	7.327	4.657	1.542	1.128	0.7664
5/30/2025	1.863	983	741	139	0.6378
5/31/2025	21.500	516	20.983	1	0.7147

Berdasarkan Tabel 4.14. nilai *Silhouette Score* bervariasi dengan rentang antara 0.4145 hingga 0.8286. Nilai yang lebih tinggi menunjukkan pemisahan *cluster* yang lebih jelas dan struktur pengelompokan yang lebih stabil, sedangkan nilai yang lebih rendah menunjukkan adanya tumpang tindih antar *cluster*. Selain itu, distribusi jumlah anggota *cluster* juga menunjukkan perbedaan struktur spasial

antar hari, di mana pada beberapa tanggal hanya terbentuk dua *cluster* dominan, sedangkan pada tanggal lainnya ketiga *cluster* terbentuk dengan distribusi yang lebih seimbang.

Tanggal 29 Mei 2025 dipilih sebagai representasi analisis mendalam karena menunjukkan kombinasi struktur *cluster* yang jelas, distribusi anggota *cluster* yang relatif seimbang, serta nilai *Silhouette Score* yang tinggi yaitu 0.7664. Meskipun terdapat beberapa tanggal lain yang memiliki nilai *Silhouette Score* lebih tinggi, distribusi *cluster* pada tanggal-tanggal tersebut cenderung kurang merata atau menunjukkan dominasi satu *cluster* yang sangat besar. Oleh karena itu, tanggal 29 Mei dianggap paling representatif untuk menggambarkan pola pengelompokan kejadian petir yang stabil dan informatif.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan PCA dan PCA-*Clustering* efektif dalam mengidentifikasi pola distribusi spasial kejadian petir pada skala harian. Metode ini mampu menyederhanakan data berdimensi tinggi menjadi struktur yang lebih mudah dianalisis, sekaligus memberikan gambaran mengenai karakteristik pengelompokan kejadian petir berdasarkan kemiripan spasial-temporal. Hasil ini dapat menjadi dasar untuk analisis lanjutan, seperti pemodelan risiko atau identifikasi zona rawan petir di wilayah penelitian.

Keteraturan pola kejadian petir yang teridentifikasi dalam penelitian ini menunjukkan bahwa fenomena alam memiliki pola yang sistematis dan dapat dianalisis secara ilmiah. Hal ini sejalan dengan konsep *sunnatullah* dalam Islam, yaitu hukum-hukum alam yang telah ditetapkan oleh Allah SWT dan dapat

dipelajari oleh manusia melalui ilmu pengetahuan. Sebagaimana firman Allah SWT dalam QS. Al-Furqan ayat 2:

وَحَلَقَ كُلَّ شَيْءٍ فَعَدَّهُ تَقْدِيرًا ۚ

"...dan Dia menciptakan segala sesuatu, lalu menetapkannya menurut ukuran (ketentuan) yang tepat." (QS: Al-Furqan: 2)

Ayat tersebut menegaskan bahwa setiap fenomena di alam semesta, termasuk kejadian petir, telah diatur dengan ukuran dan ketentuan tertentu oleh Allah SWT. Oleh karena itu, pola-pola yang ditemukan dalam penelitian ini merupakan bagian dari keteraturan tersebut yang dapat dipahami melalui pendekatan ilmiah.

BAB V

KESIMPULAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa penerapan *Principal Component Analysis* (PCA) dan *PCA-Clustering* mampu mengidentifikasi pola spasial-temporal kejadian petir secara efektif. PCA berhasil mereduksi kompleksitas data petir yang memiliki banyak variabel menjadi beberapa komponen utama tanpa kehilangan informasi varians yang signifikan, sehingga memudahkan analisis pola kejadian petir berdasarkan dimensi waktu dan lokasi.

Hasil analisis menunjukkan bahwa komponen utama pertama (PC1) memiliki kontribusi varians terbesar dan merepresentasikan dimensi temporal kejadian petir, sedangkan PC2 dan PC3 merepresentasikan variasi spasial berdasarkan koordinat geografis. Hal ini menunjukkan bahwa faktor waktu memiliki pengaruh yang lebih dominan dibandingkan faktor lokasi dalam membentuk pola kejadian petir pada data penelitian.

Proses *PCA-Clustering* yang dilakukan berdasarkan nilai PC1 menghasilkan tiga kelompok kejadian petir dengan karakteristik temporal yang berbeda. *Cluster* dengan jumlah kejadian terbesar merepresentasikan pola kejadian petir yang paling dominan selama periode pengamatan, sedangkan *cluster* lainnya menunjukkan variasi pola kejadian yang relatif lebih jarang terjadi.

Evaluasi hasil *clustering* menggunakan *Silhouette Score* menunjukkan nilai yang tinggi, yang mengindikasikan bahwa struktur *cluster* memiliki tingkat

pemisahan antar *cluster* yang baik serta kekompakan internal yang tinggi. Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan PCA dan PCA-*Clustering* efektif dalam menganalisis pola spasial-temporal kejadian petir dan dapat memberikan gambaran mengenai kecenderungan waktu dan lokasi terjadinya petir, sehingga berpotensi mendukung analisis lanjutan dalam konteks pemahaman pola kejadian petir dan mitigasi risiko bencana.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Pengelompokan data dilakukan hanya berdasarkan komponen utama pertama (PC1), sehingga informasi dari komponen lain yang juga merepresentasikan variasi spasial belum dimanfaatkan secara optimal dalam proses *clustering*. Selain itu, penggunaan variabel dalam analisis masih terbatas pada atribut waktu dan koordinat lokasi tanpa mempertimbangkan faktor meteorologis lain yang dapat memengaruhi kejadian petir. Keterbatasan lainnya terletak pada periode data pengamatan yang relatif terbatas, sehingga pola yang dihasilkan belum sepenuhnya merepresentasikan variasi kejadian petir dalam jangka panjang. Pola kejadian petir yang teridentifikasi dalam penelitian ini juga menunjukkan adanya keteraturan fenomena alam yang sejalan dengan konsep sunnatullah dalam Islam, yaitu hukum-hukum alam yang telah ditetapkan oleh Allah SWT.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya, yaitu:

1. Penelitian selanjutnya dapat memperluas periode data pengamatan, misalnya dengan menggunakan data lintas bulan atau lintas tahun, sehingga pola spasial-

temporal kejadian petir dapat dianalisis secara lebih komprehensif dan berkelanjutan.

2. Metode *clustering* dapat dikembangkan dengan membandingkan pendekatan yang digunakan dalam penelitian ini dengan metode *clustering* lain, seperti K-Means atau DBSCAN, untuk melihat perbedaan struktur *cluster* yang dihasilkan.
3. Penelitian berikutnya dapat menambahkan variabel meteorologis lain, seperti curah hujan, suhu udara, dan kelembapan, untuk memperkaya analisis serta meningkatkan akurasi dalam mengidentifikasi pola kejadian petir.
4. Hasil penelitian ini juga dapat dikembangkan lebih lanjut dalam bentuk visualisasi peta spasial atau sistem pendukung keputusan, sehingga dapat memberikan manfaat praktis bagi instansi terkait dalam mendukung analisis dan mitigasi risiko kejadian petir.

DAFTAR PUSTAKA

- Abriyanto, H. Y., & Batih, H. (2023). Analisis Efektifitas Pemasangan Proteksi Petir Berdasarkan Tingkat Ancaman Sambaran Petir Cloud to Ground di Titik Tower Transmisi Air Anyir - Pangkalpinang. *Electrician : Jurnal Rekayasa Dan Teknologi Elektro*, 17(3), 235–241. <https://doi.org/10.23960/elc.v17n3.2486>
- Akhyar, S., & AN, A. N. (2020). Pemikiran Tafsir Sufistik Falsafi Hamzah Fansuri tentang Tarikat dan Syariat (Kajian Kitab Turast Melayu Jawi Zinatul Muwahhidin). *Al-I'jaz : Jurnal Kwahyuan Islam*, VI(1), 19–31. <https://doi.org/http://dx.doi.org/10.30821/al-i'jaz.v6i1.7437>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysdiyan, C., Tazi, I., Hariyadi, M. A., Muslim, M. A., Basid, P. M. N. S. A., Arif, Y. M., Purwanto, M. S., Parwatiningsy, D., Supriyono, & Wicaksono, H. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. *IEEE Access*, 11(March), 74590–74601. <https://doi.org/10.1109/ACCESS.2023.3275852>
- Budi, A. S., & Pratama, W. N. (2025). ANALISIS PETA KERAPATAN PETIR TIPE CLOUD TO GROUND (CG) DI WILAYAH KOTA MANADO MENGGUNAKAN METODE INVERSE DISTANCE WEIGHTING (IDW) TAHUN 2023 Pengamatan petir oleh BMKG dilakukan pertama kali pada tahun 2008 software LD-2000 . Kemudian pada tahun 2019. XIII, 13–23. <https://doi.org/10.21009/03.1301.FA02>
- Cahyo, R. H. D., Ardimas, Anggara, Y. D., Puspitasari, A., & Peredianto. (2023). Cloud to Ground Lightning Strike Density Analysis Based on Geographic Information System in Tuban Regency Area. *Journal of Physics: Conference Series*, 2623(1). <https://doi.org/10.1088/1742-6596/2623/1/012006>
- Chakraborty, R., Chakraborty, A., Basha, G., & Ratnam, M. V. (2021). Lightning occurrences and intensity over the Indian region: Long-term trends and future projections. *Atmospheric Chemistry and Physics*, 21(14), 11161–11177. <https://doi.org/10.5194/acp-21-11161-2021>
- Denov, B., Hidayat, S., Suwarno, & Zoro, R. (2023). A Method to Obtain Lightning Peak Current in Indonesia. *Energies*, 16(17). <https://doi.org/10.3390/en16176342>
- Fikri Ys, I. Al, & Taufiq, W. (2023). The Phenomenon of Thunder in the Qur'an: A Semiotic Analysis of Roland Barthes. *Jurnal Ilmu Al-Qur'an Dan Tafsir*, 2(1), 19–30. <https://doi.org/10.15575/mjiat.v2i1.20131>
- Gul, M., & Rehman, M. A. (2023). Big data: an optimized approach for cluster initialization. *Journal of Big Data*, 10(1), 1–19. <https://doi.org/10.1186/s40537-023-00798-1>
- Gunawan, T., Naomi, L., & Pandiangan, L. (2014). Analysis of the Level of Area Vulnerability To Lightning Strike Using Simple Additive Weighting in Bali Province. *Jurnal Meterologi Dan Geofisika*, 193–201. <https://doi.org/10.31172/jmg.v15i3.221>

- Haljasmaa, K. I., Bramm, A. M., Matrenin, P. V., & Eroshenko, S. A. (2024). Weather Condition Clustering for Improvement of Photovoltaic Power Plant Generation Forecasting Accuracy. *Algorithms*, 17(9). <https://doi.org/10.3390/a17090419>
- Han, J., Pei, J., & Tong, H. (2022). Data Mining Concepts and Techniques. In *Morgan Kaufmann Publishers* (Vol. 1. Issue 1). <https://share.google/iv7Zk8CU6y6yzdDy8>
- Hidayat, F. D., Muliadi, M., & Adriat, R. (2018). Karakteristik dan Hubungan Aktivitas Petir Cloud To Ground dengan Curah Hujan (Studi Kasus Kota Pontianak dan Sekitarnya). *Prisma Fisika*, 6(3), 176–183. <https://doi.org/10.26418/pf.v6i3.28906>
- Indraputra, R. ., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275–282. <https://doi.org/10.25105/jti.v10i3.8428>
- Indriyanti, P., & Fajriah, R. (2025). Kinerja Komparatif LSTM dan XGBoost untuk Peramalan Radiasi Matahari Perkotaan Tropis. *FORMAT: Jurnal Ilmiah Teknik Informatika*, 14(164), 245–252. <https://doi.org/10.22441/format>
- Jiang, R., Zhang, G., Wang, S., Xue, B., Xie, Z., Yu, T., Wang, K., Ding, J., & Zhu, X. (2023). Characteristics of cloud-to-ground lightning (CG) and differences between +CG and -CG strokes in China regarding the China National Lightning Detection Network. *Natural Hazards and Earth System Sciences*, 23(12), 3747–3759. <https://doi.org/10.5194/nhess-23-3747-2023>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). <https://doi.org/10.1098/rsta.2015.0202>
- Lovytaji, H. A., Rozikan, R., Kuncoro, D. C., & Karisma, R. D. L. N. (2024). Analyzing Lightning Strike Susceptibility Using the Elliptical Fitting Method with a Principal Component Analysis Approach. *Jurnal Varian*, 8(1), 25–38. <https://doi.org/10.30812/varian.v8i1.3183>
- Mathai, A. M., Provost, S. B., & Haubold, H. J. (2022). Multivariate statistical analysis in the real and complex domains. In *Multivariate Statistical Analysis in the Real and Complex Domains*. <https://doi.org/10.1007/978-3-030-95864-0>
- Morgenstern, D., Stucke, I., Simon, T., Mayr, G. J., & Zeileis, A. (2022). Differentiating lightning in winter and summer with characteristics of the wind field and mass field. *Weather and Climate Dynamics*, 3(1), 361–375. <https://doi.org/10.5194/wcd-3-361-2022>
- Mubarok, M. R., Harjanto, A. T. J., & Renaldy, R. (2025). PENINGKATAN PERFORMA DBSCAN DENGAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM KLASIFIKASI TINGKAT KEMISKINAN DI INDONESIA. *Jurnal Informatika Dan Teknologi Sains (JINTEKS)*, 7. 1176–1184. <https://doi.org/10.51401/jinteks.v7i3.6129>
- Narut, F., Wahid, A., & Sumawan, S. (2018). Karakterisasi Peristiwa Petir Di

- Wilayah Kota Kupang Serta Keterkaitannya Dengan Curah Hujan. *Jurnal Fisika : Fisika Sains Dan Aplikasinya*, 3(2), 110–116. <https://doi.org/10.35508/fisa.v3i2.611>
- Nicora, M., Tucci, M., Barmada, S., Brignone, M., & Procopio, R. (2024). Lightning Location and Peak Current Estimation From Lightning-Induced Voltages on Transmission Lines With a Machine Learning Approach. *IEEE Transactions on Electromagnetic Compatibility*, 66(3), 890–899. <https://doi.org/10.1109/TEMC.2024.3375452>
- Patel, K. (2013). Effect of Lightning on Building and Its Protection Measures. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8958(2), 182.
- Pratama, I. P. D., & Negara, P. K. G. A. (2017). Analisis Spasial Dan Temporal Data Lightning Detector Tahun 2009-2015 Di Stasiun Geofisika Sanglah Denpasar. In *Jurnal Meteorologi dan Geofisika* (Vol. 17. Issue 2). <https://doi.org/10.31172/jmg.v17i2.438>
- Riqmawatin, S. R., & Intan, P. K. (2020). Analisa Pemetaan Daerah Rawan Sambaran Petir Di Wilayah Kabupaten Pasuruan dengan Menggunakan Metode Simple Additive Weighting. *Jurnal Mahasiswa Matematika ALGEBRA*, 1(1), 198–210. <https://core.ac.uk/download/pdf/329047561.pdf>
- Rosyada, I. A., & Utari, D. T. (2024). Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering. *Jambura Journal of Probability and Statistics*, 5(1), 6–13. <https://doi.org/10.37905/jjps.v5i1.18733>
- Seftiani, Y. M., Hazmi, A., Novizon, N., Abadi, A., & Widia, R. (2023). Karakteristik Medan Listrik-Dekat Petir Positive Cloud to Ground. *Elektron : Jurnal Ilmiah*, 15, 1–6. <https://doi.org/10.30630/eji.15.1.334>
- Septiarini, I., Gani, U. A., & Rajagukguk, M. (2020). ANALISA PEMETAAN DAERAH RAWAN SAMBARAN PETIR DI WILAYAH KOTA PONTIANAK DENGAN MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING. 7(2), 2141–2154. <https://doi.org/https://doi.org/10.26418/j3eit.v10i1.51852>
- Shubri, A. R., Vonnisa, M., & Marzuki, M. (2023). Seasonal and Diurnal Variation of Lightning Over Java Region and Its Relationship with Aerosol. *Jurnal Penelitian Pendidikan IPA*, 9(11), 9516–9523. <https://doi.org/10.29303/jppipa.v9i11.4784>
- Stucke, I., Morgenstern, D., Zeileis, A., Mayr, G. J., Simon, T., Diendorfer, G., Schulz, W., & Pichler, H. (2024). Diagnosing upward lightning from tall objects from meteorological thunderstorm environments. *Electric Power Systems Research*, 229(January), 110199. <https://doi.org/10.1016/j.epsr.2024.110199>
- Surono, S., Goh, K. W., Onn, C. W., & Marestiani, F. (2023). Developing an optimized recurrent neural network model for air quality prediction using K-means clustering and PCA dimension reduction. *International Journal of Innovative Research and Scientific Studies*, 6(2), 330–343. <https://doi.org/10.53894/ijirss.v6i2.1427>

- Susanto, E. (2018). Penentuan Daerah Rawan Bencana Sambaran Petir Di Wilayah Kabupaten Dan Kota Bandung Jawa Barat. *Jurnal Ilmu Dan Inovasi Fisika*, 2(2), 137–144. <https://doi.org/10.24198/jiif.v2i2.19728>
- Tai, J. H., Soong, W. K., Tan, P. H., Wang, M. Y. M., & Lin, P. H. (2023). Feasibility of forecasting convective rain by diagnosing intracloud lightning jumps. *Meteorology and Atmospheric Physics*, 135(6), 1–13. <https://doi.org/10.1007/s00703-023-00997-8>
- Tropmann-frick, M., & Forke, C. (2021). Feature Engineering Techniques and Spatio-Temporal Data Processing. *Springer Nature Link*, 237–244. <https://doi.org/10.1007/s13222-021-00391-x>
- Wahjudi, D. (2014). Mengatasi Bahaya Petir dan Proteksi Petir Gedung Bertingkat. *Teodolita Jurnal Ilmu-Ilmu Teknik*, 15(2), 57–71. <https://doi.org/10.53810/jt.v15i2.140>
- Yusnaini, H., Marzuki, Muharsyah, R., Vonnisa, M., & Tangang, F. (2021). Influence of topography on lightning density in Sumatra. *Journal of Physics: Conference Series*, 1876(1). <https://doi.org/10.1088/1742-6596/1876/1/012022>
- Zhang, T., Wang, J., Ma, Q., & Fu, L. (2024). Improving the Detection Effect of Long-Baseline Lightning Location Networks Using PCA and Waveform Cross-Correlation Methods. *Remote Sensing*, 16(5). <https://doi.org/10.3390/rs16050885>
- Zhu, B., Lin, B., Liu, B., & Liu, H. (2025). Analysis of temporal and spatial distribution of CG lightning activity and its relationship with terrain features around Jiuxianshan mountain. *Journal of Atmospheric and Solar-Terrestrial Physics*, 273(November 2024), 106546. <https://doi.org/10.1016/j.jastp.2025.106546>
- Zoro, R. (2010). Induksi Dan Konduksi Gelombang Elektromagnetik Akibat Sambaran Petir Pada Jaringan Tegangan Rendah. *MAKARA of Technology Series*, 13(1), 25–32. <https://doi.org/10.7454/mst.v13i1.492>