

Analisis Big Data Satelit GOES-19 Deteksi Dini Cuaca Ekstrem: Studi Kasus Karibia, Implikasinya untuk Indonesia

Rodo Lemuel Nadeak¹, Sevta Triwana Simamora^{2*}, Azis Kurniadi³

^{1,2,3}Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Medan, Indonesia

Email: ¹rodonadeak@gmail.com, ^{2*}sevtasimamora@gmail.com, ³azisk147@gmail.com

(*Email Corresponding Author: sevtasimamora@gmail.com)

Received: 24 Maret 2026 | Revision: 1 April 2026 | Accepted: 1 April 2026

Abstrak

Perubahan iklim global telah meningkatkan frekuensi dan intensitas cuaca ekstrem di wilayah tropis, dengan Karibia mengalami intensifikasi siklon tropis dan Indonesia menghadapi peningkatan hujan ekstrem serta banjir. Meskipun kemajuan telah dicapai dalam pemanfaatan data satelit dan machine learning, masih terdapat kesenjangan penelitian yang mengintegrasikan big data satelit GOES-19 terbaru dengan karakterisasi pola awan tropis secara lintas wilayah, khususnya untuk adaptasi di Indonesia. Penelitian ini bertujuan mengidentifikasi pola awan tropis menggunakan analisis big data satelit GOES-19 di wilayah Karibia sebagai studi kasus serta mengeksplorasi implikasinya terhadap sistem deteksi dini cuaca ekstrem di Indonesia. Metode yang diterapkan adalah pendekatan kuantitatif berbasis big data dengan teknik unsupervised machine learning berupa algoritma MiniBatchKMeans. Data utama diperoleh dari produk ABI-L2-MCMIPF satelit NOAA GOES-19 melalui bucket AWS S3 untuk periode November 2024 pada koordinat 8°LU–30°LU dan 90°BB–60°BB, dengan ekstraksi lima channel ABI, pra-proses menggunakan xarray dan s3fs, normalisasi data, serta penentuan $k=5$ optimal melalui Elbow Method pada 2.308.961 piksel. Analisis komparatif dilakukan dengan data Himawari-9 untuk wilayah Indonesia dan divalidasi menggunakan jalur badai IBTrACS. Hasil penelitian mengungkapkan lima zona awan tropis yang stabil secara spasial dan temporal, dengan zona Awan Konvektif Dalam hanya 4,1% di Karibia pasca-badai namun mencapai 44% di Indonesia, serta perubahan distribusi yang konsisten selama fase aktif Hurricane Rafael, Sara, dan pasca-badai. Korelasi kuat ditemukan antara zona konvektif dengan jalur badai aktual, didukung Silhouette Score rata-rata 0,43–0,45. Studi ini berimplikasi pada pengembangan model prediksi cuaca berbasis satelit yang lebih adaptif dan akurat untuk Indonesia, sehingga dapat meningkatkan sistem peringatan dini BMKG terhadap bencana cuaca ekstrem di wilayah tropis.

Kata Kunci: Big Data, Clustering, Cuaca Ekstrem, GOES-19, Pola Awan Tropis

Abstract

Global climate change has increased the frequency and intensity of extreme weather events in tropical regions, with the Caribbean experiencing intensified tropical cyclones and Indonesia facing increased extreme rainfall and flooding. Despite advancements in the use of satellite data and machine learning, there remains a research gap in integrating the latest GOES-19 satellite big data with cross-regional characterization of tropical cloud patterns, particularly for adaptation in Indonesia. This study aims to identify tropical cloud patterns using GOES-19 satellite big data analysis in the Caribbean region as a case study and to explore its implications for early warning systems of extreme weather in Indonesia. The method applied is a quantitative big data approach using unsupervised machine learning techniques, specifically the MiniBatchKMeans algorithm. The primary data were obtained from the ABI-L2-MCMIPF product of the NOAA GOES-19 satellite via AWS S3 for the period of November 2024 within the coordinates of 8°N–30°N and 90°W–60°W. Five ABI channels were extracted, followed by preprocessing using xarray and s3fs, data normalization, and determination of the optimal number of clusters ($k=5$) using the Elbow Method on 2,308,961 pixels. Comparative analysis was conducted using Himawari-9 data for the Indonesian region and validated using IBTrACS storm tracks. The results reveal five tropical cloud zones that are stable both spatially and temporally. The Deep Convective Cloud zone accounts for only 4.1% in the post-storm Caribbean but reaches 44% in Indonesia, with consistent distribution changes observed during the active phases of Hurricane Rafael, Hurricane Sara, and the post-storm period. A strong correlation is found between convective zones and actual storm tracks, supported by an average Silhouette Score of 0.43–0.45. This study has implications for the development of more adaptive and accurate satellite-based weather prediction models for Indonesia, thereby enhancing BMKG's early warning systems for extreme weather disasters in tropical regions.

Keywords: Big Data, Clustering, Extreme Weather, GOES-19, Tropical Cloud Patterns

1. PENDAHULUAN

Perubahan iklim global dalam beberapa tahun terakhir telah meningkatkan frekuensi dan intensitas kejadian cuaca ekstrem, khususnya di wilayah tropis. Laporan terbaru menunjukkan bahwa peningkatan suhu permukaan laut dan ketidakstabilan atmosfer berkontribusi terhadap intensifikasi sistem konvektif dan badai tropis [1]. Secara internasional, kawasan Karibia mengalami peningkatan aktivitas siklon tropis yang signifikan, yang berkaitan erat

dengan dinamika awan konvektif yang kompleks dan sulit diprediksi. Di sisi lain, Indonesia sebagai negara tropis dengan karakteristik maritim juga menghadapi tantangan serupa, di mana kejadian hujan ekstrem, banjir, dan badai lokal semakin meningkat. Laporan Badan Meteorologi, Klimatologi, dan Geofisika menunjukkan bahwa akurasi prediksi cuaca ekstrem masih menjadi tantangan utama, terutama dalam sistem deteksi dini berbasis observasi atmosfer [2].

Sebagai respons terhadap permasalahan tersebut, para peneliti telah mengembangkan pendekatan berbasis big data dan penginderaan jauh untuk meningkatkan kemampuan prediksi cuaca ekstrem. Satelit geostasioner generasi terbaru seperti GOES-R series (termasuk GOES-16 hingga GOES-19) menyediakan data dengan resolusi temporal tinggi yang memungkinkan pemantauan awan secara kontinu. Studi menunjukkan bahwa sensor Advanced Baseline Imager (ABI) mampu mendeteksi perubahan struktur awan secara lebih detail, sehingga mendukung analisis cuaca ekstrem secara real-time [3]. Selain itu, pendekatan machine learning seperti clustering dan klasifikasi telah digunakan untuk mengidentifikasi pola awan konvektif yang berkorelasi dengan kejadian cuaca ekstrem [4]. Integrasi big data satelit dengan teknik analitik modern dinilai sebagai pendekatan yang efektif dalam meningkatkan sistem prediksi cuaca di wilayah tropis.

Penelitian sebelumnya telah banyak membahas karakterisasi awan tropis menggunakan data satelit. Beberapa studi mengkaji hubungan antara suhu puncak awan dan intensitas hujan ekstrem menggunakan data satelit inframerah [5]. Sementara itu, penelitian lain memanfaatkan teknik clustering untuk mengidentifikasi pola awan konvektif dalam skala besar [6]. Beberapa penelitian terkini juga telah mengembangkan metode unsupervised clustering pada data satelit untuk klasifikasi awan tropis [7], [8], serta menyoroti tantangan klasifikasi awan tinggi di wilayah tropis [9]. Rekonstruksi intensitas siklon tropis menggunakan data reanalisis ERA5 [10] dan dataset sintetik curah hujan siklon tropis [11] turut memberikan kontribusi penting dalam pemahaman dinamika badai. Di wilayah Indonesia, studi tentang transport kelembaban daratan dan lautan sebagai pendorong hujan ekstrem [12], karakteristik mesoscale convective complexes penyebab banjir bandang [13], serta dampak sistem konvektif skala meso di Nusantara [14] semakin memperkaya literatur. Selain itu, algoritma deteksi hujan berbasis multi-band Himawari-8 [15], evaluasi produk near-real-time Himawari-8/9 dan IMERG [16], [17], serta validasi data curah hujan satelit di Maritime Continent [18] telah menunjukkan akurasi yang baik untuk pemantauan cuaca di wilayah tropis Asia Tenggara. Pendekatan machine learning untuk big geospatial data juga telah banyak dikembangkan untuk meningkatkan efisiensi analisis data satelit berskala besar [19].

Meskipun demikian, sebagian besar penelitian tersebut masih terbatas pada wilayah studi tertentu dan belum mengintegrasikan hasilnya dalam konteks lintas wilayah, khususnya untuk adaptasi di Indonesia. Pemanfaatan data satelit generasi terbaru seperti GOES-19 masih relatif jarang dieksplorasi dalam literatur ilmiah. Berdasarkan hal tersebut, terdapat kesenjangan penelitian berupa kurangnya studi yang mengintegrasikan analisis big data satelit generasi terbaru dengan karakterisasi pola awan tropis untuk mendukung deteksi dini cuaca ekstrem serta penerapannya pada konteks regional Indonesia.

Penelitian ini bertujuan untuk menganalisis big data satelit GOES-19 guna mengkarakterisasi pola awan tropis di wilayah Karibia sebagai studi kasus, serta mengeksplorasi implikasinya terhadap pengembangan sistem deteksi dini cuaca ekstrem di Indonesia. Secara spesifik, penelitian ini diharapkan dapat mengidentifikasi pola awan yang signifikan, menganalisis keterkaitannya dengan kejadian cuaca ekstrem, serta memberikan kontribusi terhadap model prediksi berbasis data satelit yang lebih adaptif dan akurat di wilayah tropis.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis big data untuk mengkarakterisasi pola awan tropis di wilayah Karibia menggunakan data satelit geostasioner generasi terbaru. Jenis penelitian bersifat eksploratif-analitik dengan menerapkan teknik unsupervised machine learning pada data penginderaan jauh berskala besar.

Tahapan pertama adalah pengumpulan data, di mana data utama dari produk ABI-L2-MCMIPF satelit NOAA GOES-19 dan data komparatif dari satelit Himawari-9 diakses serta diunduh secara publik melalui *bucket* AWS S3. Selanjutnya, tahapan kedua berfokus pada praproses data yang meliputi koneksi ke AWS S3, ekstraksi lima *channel* satelit yang relevan (C02, C08, C09, C13, dan C14), konversi koordinat proyeksi satelit menjadi koordinat geografis, pemotongan wilayah studi (*subsetting* spasial) Karibia, dan pembersihan data dari nilai kosong (NaN). Tahapan ketiga adalah normalisasi fitur menggunakan *StandardScaler* untuk menyeragamkan skala pada seluruh fitur *channel* satelit, sehingga setiap fitur memiliki kontribusi yang setara pada model dan menghindari bias akibat perbedaan skala pengukuran. Setelah data siap, tahapan keempat adalah penentuan jumlah *cluster* optimal dengan mengevaluasi nilai *inertia* dari variasi $k=2$ hingga $k=10$ menggunakan *Elbow Method*. Tahapan kelima merupakan proses pemodelan utama, yaitu implementasi algoritma *unsupervised machine learning MiniBatchKMeans* pada 2.308.961 piksel data untuk mengelompokkan area menjadi zona-zona awan yang spesifik. Kemudian, tahapan keenam melibatkan analisis spasio-temporal dengan menerapkan model *clustering* pada tiga *snapshot* waktu yang berbeda (fase badai Rafael aktif, Sara aktif, dan pasca-badai), di mana kualitas hasil *clustering* dievaluasi menggunakan metrik *Silhouette Score*. Terakhir, tahapan ketujuh adalah validasi meteorologis yang dilakukan dengan meng-*overlay* peta zonasi pola awan

terhadap data jalur badai tropis aktual dari arsip IBTrACS untuk menguji korelasinya dengan sistem cuaca di dunia nyata

2.1. Data dan Sumber

Tabel 1. Spesifikasi Data yang Digunakan

Parameetr	GOES-19 (Data Utama)	Himawari-9 (Data Komparatif)
Produk	ABI-L2-MCMIPF	AHI-L2-FLDK-Clouds
Resolusi Spasial	~2 km (channel IR)	~2 km
Resolusi Temporal	10 menit (Full Disk)	10 menit (Full Disk)
Periode Analisis	November 2024	20 November 2024
Akses	AWS S3: noaa-goes19	AWS S3: noaa-himawari9
Format	NetCDF4 (.nc)	NetCDF4 (.nc)
Biaya Akses	Gratis, publik	Gratis, publik

2.2. Wilayah dan Periode Studi

Wilayah studi untuk analisis utama mencakup Laut Karibia dan sekitarnya dengan batas koordinat 8°LU–30°LU dan 90°BB–60°BB. Wilayah ini dipilih karena berada dalam cakupan optimal GOES-19 dan memiliki aktivitas konvektif tropis yang tinggi, khususnya selama hurricane season. Periode analisis ditetapkan pada bulan November 2024 yang mencakup fase aktif Hurricane Rafael (3–9 November) dan Hurricane Sara (13–17 November) serta fase pasca-badai (20 November), sehingga memungkinkan analisis spasio-temporal yang komprehensif. Analisis komparatif menggunakan data Himawari-9 untuk wilayah Indonesia dengan batas koordinat 11°LS–6°LU dan 95°BT–141°BT.

2.3. Variabel yang Diekstrak

Dari produk ABI-L2-MCMIPF, diekstrak lima channel ABI yang paling relevan untuk karakterisasi pola awan, yaitu CMI_C02 (Red Visible, 0,64 μm) untuk reflektansi awan siang hari, CMI_C08 (Upper-level Water Vapor, 6,19 μm) untuk uap air lapisan atas atmosfer, CMI_C09 (Mid-level Water Vapor, 6,93 μm) untuk uap air lapisan menengah, CMI_C13 (Clean IR Longwave, 10,35 μm) untuk suhu puncak awan, serta CMI_C14 (IR Longwave, 11,20 μm) sebagai validasi channel inframerah. Channel C13 dan C14 diprioritaskan karena dapat digunakan siang maupun malam hari tanpa bergantung pada cahaya matahari, sehingga cocok untuk analisis temporal.

2.4. Tahapan Praproses Data

Data satelit melalui serangkaian tahap praproses sebelum dianalisis. Pertama, koneksi ke bucket AWS S3 dilakukan menggunakan library `s3fs` dengan mode `anonymous`, kemudian file `NetCDF4` target dipilih berdasarkan timestamp dan dibuka menggunakan engine `h5netcdf` melalui `xarray`. Hanya lima channel target yang dimuat ke dalam memori untuk efisiensi komputasi. Kedua, konversi koordinat dari sistem proyeksi geostationary GOES ke koordinat geografis (lintang/bujur) dilakukan menggunakan persamaan geometri satelit berdasarkan parameter proyeksi yang tersimpan dalam metadata file, meliputi ketinggian satelit (35.786.023 m), jari-jari ekuator bumi (6.378.137 m), jari-jari kutub bumi (6.356.752 m), dan longitude proyeksi (-89,5°). Ketiga, subsetting spasial dilakukan dengan menerapkan masker batas koordinat wilayah Karibia, menghasilkan 2.308.961 piksel valid per snapshot. Keempat, data dikonversi ke format tabular dan dibersihkan dari nilai NaN. Pada dataset yang dianalisis, tidak ditemukan piksel dengan nilai NaN sehingga seluruh 2.308.961 piksel digunakan dalam analisis clustering.

2.5. Metode Clustering

Analisis clustering dilakukan menggunakan algoritma `MiniBatchKMeans` dari library `scikit-learn`. Sebelum clustering, seluruh fitur dinormalisasi menggunakan `StandardScaler` untuk memastikan setiap channel memiliki kontribusi yang setara tanpa bias skala pengukuran. Pemilihan K-Means didasarkan pada tiga pertimbangan: (1) efisiensi komputasi untuk dataset berskala besar (jutaan piksel), (2) kemampuan menghasilkan cluster yang interpretatif secara meteorologis, dan (3) kesesuaian yang telah dibuktikan oleh berbagai penelitian sebelumnya untuk segmentasi pola awan berbasis data satelit.

Penentuan jumlah cluster optimal (k) dilakukan menggunakan Elbow Method dengan mengevaluasi nilai inerti untuk $k = 2$ hingga $k = 10$. Kualitas hasil clustering dievaluasi menggunakan Silhouette Score yang dihitung dari sampel acak sebanyak 200.000 piksel untuk efisiensi komputasi. Analisis diulang pada tiga snapshot berbeda yang merepresentasikan fase aktif Hurricane Rafael (5 November 2024), fase aktif Hurricane Sara (15 November 2024), dan fase pasca-badai (20 November 2024). Pada ketiga snapshot tersebut digunakan scaler yang sama untuk memastikan konsistensi dan komparabilitas hasil antar waktu.

2.6. Validasi dengan Data Jalur Badai

Untuk memvalidasi relevansi meteorologis hasil clustering, jalur badai tropis aktual di overlay ke atas peta zonasi menggunakan data IBTrACS (International Best Track Archive for Climate Stewardship) versi 4 yang diterbitkan oleh NOAA NCEI. Data IBTrACS menyediakan posisi, intensitas angin, dan status sistem badai setiap 3 jam sepanjang siklus hidupnya. Badai yang dianalisis meliputi Hurricane Rafael, Hurricane Sara, Hurricane Oscar, dan Hurricane Milton yang seluruhnya aktif di wilayah Karibia selama Oktober–November 2024.

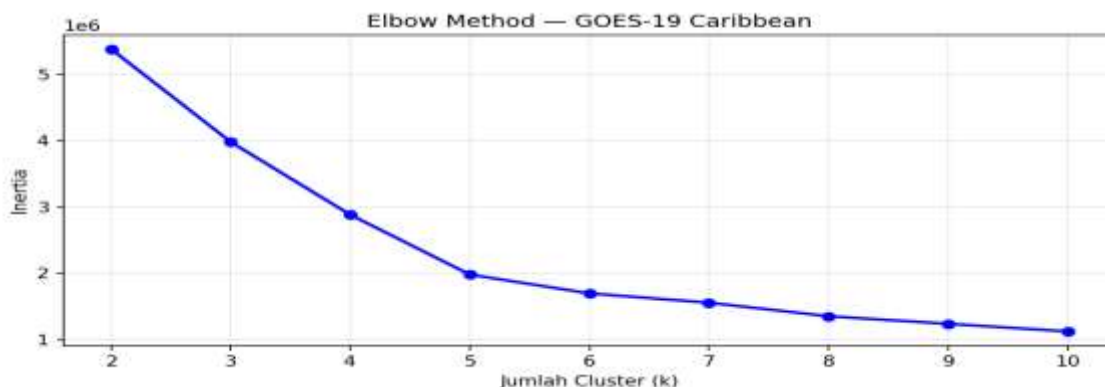
3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Big Data NOAA GOES-19

Sebagai landasan analisis, perlu dipahami terlebih dahulu skala dan karakteristik Big Data dari dataset yang digunakan. Bucket noaa-goes19 di AWS S3 menyimpan 103 produk berbeda dengan total volume data mencapai estimasi 1,012 Petabyte per tahun berdasarkan hasil sampling. Data baru tersedia dengan delay mendekati 0 jam dari waktu observasi, menjadikan GOES-19 sebagai salah satu sumber data cuaca near real-time terbesar yang dapat diakses publik. Setiap file produk ABI-L2-MCMIPF berukuran sekitar 250 MB dan memiliki dimensi spasial 5.424×5.424 piksel yang merepresentasikan seluruh hemisphere barat bumi. Karakteristik ini memenuhi definisi Big Data secara komprehensif dari perspektif Volume, Velocity, dan Variety, sehingga membutuhkan pendekatan komputasi khusus untuk pengolahan dan analisisnya.

3.2 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method pada data snapshot 20 November 2024 yang merepresentasikan kondisi pasca-badai. Hasil evaluasi inerti untuk berbagai nilai k ditunjukkan pada Gambar 1.

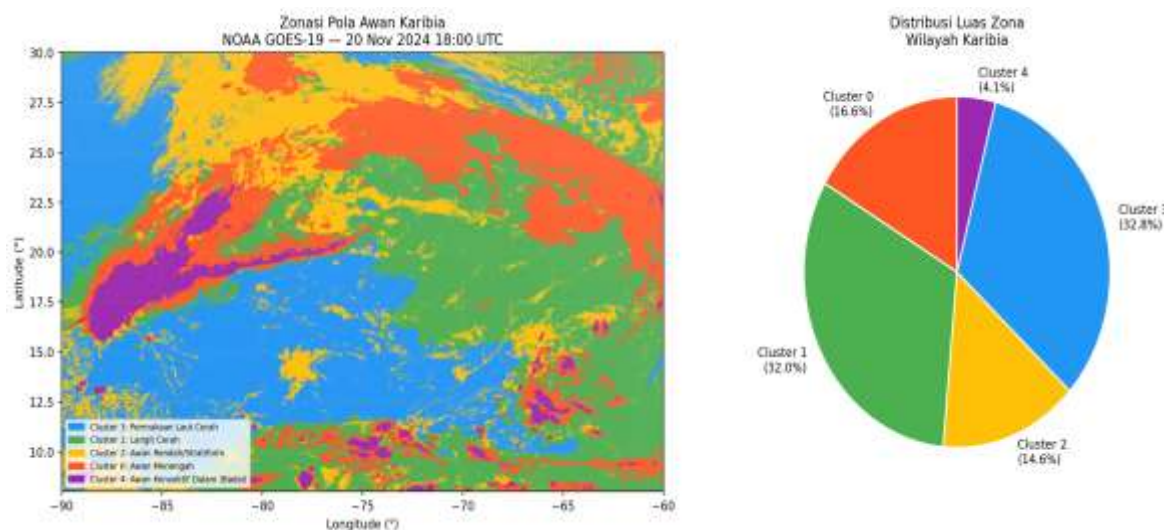


Gambar 1. Grafik Elbow Method untuk penentuan jumlah cluster optimal (k=2 hingga k=10)

Grafik elbow pada Gambar 1 menunjukkan penurunan inerti yang signifikan dari k=2 (inerti: 5.382.822) hingga k=5 (inerti: 1.978.540), dengan penurunan yang mulai melandai setelah k=5. Titik siku yang paling jelas terbentuk pada k=5, di mana penambahan cluster setelah angka tersebut tidak memberikan penurunan inerti yang proporsional. Berdasarkan analisis ini, k=5 ditetapkan sebagai jumlah cluster optimal yang merepresentasikan lima zona kondisi atmosfer yang berbeda secara fisik.

3.3 Hasil Zonasi Pola Awan

Penerapan MiniBatchKMeans dengan k=5 pada 2.308.961 piksel wilayah Karibia menghasilkan lima zona awan yang dapat diinterpretasikan secara meteorologis. Hasil zonasi untuk snapshot 20 November 2024 ditampilkan pada Gambar 2, sedangkan profil karakteristik rata-rata tiap zona disajikan pada Tabel 2.



Gambar 2. Peta zonasi pola awan wilayah Karibia berdasarkan K-Means Clustering (NOAA GOES-19, 20 November 2024 18:00 UTC) dan distribusi luas tiap zona

Tabel 2. Profil karakteristik rata-rata tiap zona awan (20 November 2024)

Zona	CMI_C02	CMI_C08(K)	CMI_C09(K)	CMI_C13(K)	CMI_C14(K)	Luas(%)
Awan Konvektif Dalam	0,481	214,60	216,26	221,54	219,04	4,1%
Awan Menengah	0,265	230,97	237,02	254,68	252,23	16,6%
Awan Rendah/Stratiform	0,376	243,52	252,99	276,94	276,56	14,6%
Langit Cerah	0,077	236,85	246,61	288,15	286,56	32,0%
Permukaan Laut Cerah	0,066	248,04	258,76	294,25	293,56	32,8%

Berdasarkan Tabel 2, karakteristik fisik tiap zona dapat diinterpretasikan sebagai berikut. Pertama, zona Awan Konvektif Dalam memiliki nilai CMI_C13 terendah (221,54 K atau setara $-51,6^{\circ}\text{C}$), yang mengindikasikan puncak awan yang sangat tinggi dan dingin — karakteristik khas sistem konveksi dalam yang berkaitan langsung dengan potensi cuaca ekstrem dan badai tropis. Nilai reflektansi CMI_C02 yang tinggi (0,481) mengkonfirmasi keberadaan awan tebal yang signifikan. Meskipun hanya mencakup 4,1% area, zona ini merupakan indikator risiko cuaca ekstrem yang paling kritis.

Kedua, zona Permukaan Laut Cerah memiliki nilai CMI_C13 tertinggi (294,25 K atau $21,1^{\circ}\text{C}$) dengan reflektansi CMI_C02 terendah (0,066), merepresentasikan permukaan laut terbuka atau daratan dengan langit yang cerah dan risiko cuaca ekstrem rendah. Zona ini mendominasi bagian timur wilayah studi (32,8%). Ketiga, zona Langit Cerah (32,0%) menunjukkan kondisi dengan sedikit awan tipis (cirrus) yang ditandai dengan CMI_C13 sekitar 288 K. Keempat, zona Awan Rendah/Stratiform (14,6%) dan Awan Menengah (16,6%) mewakili kondisi antara dengan potensi presipitasi ringan hingga sedang.

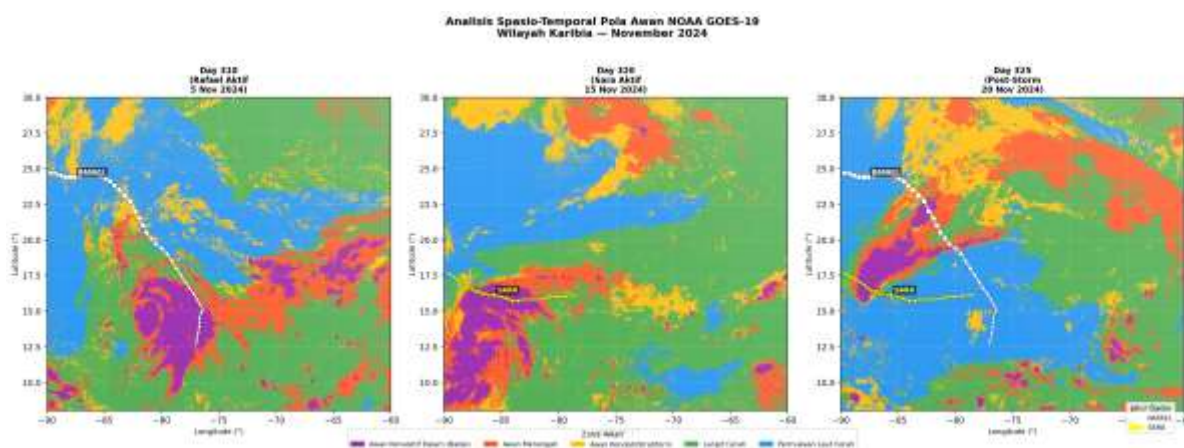
3.4 Analisis Spasio-Temporal

Untuk memenuhi dimensi temporal analisis, clustering diterapkan pada tiga snapshot yang merepresentasikan fase berbeda dari siklus badai: fase aktif Hurricane Rafael (5 November 2024), fase aktif Hurricane Sara (15 November 2024), dan fase pasca-badai (20 November 2024). Perbandingan distribusi zona antar waktu disajikan pada Tabel 3 dan Gambar 3.

Tabel 3. Perbandingan distribusi zona awan antar snapshot temporal

Zona Awan	5 Nov 2024 (Rafael Aktif)	15 Nov 2024 (Sara Aktif)	20 Nov 2024 (Post-Storm)
Awan Konvektif Dalam (Badai)	8,9%	7,0%	4,1%
Awan Menengah	15,2%	13,2%	16,6%
Awan Rendah/Stratiform	12,1%	12,2%	14,6%
Langit Cerah	38,8%	46,7%	32,0%
Permukaan Laut Cerah	24,9%	20,9%	32,8%
Silhouette Score	0,4457	0,4332	0,4371

Hasil pada Tabel 3 menunjukkan pola temporal yang konsisten dan bermakna secara meteorologis. Proporsi Awan Konvektif Dalam menurun secara progresif dari 8,9% saat Rafael aktif, menjadi 7,0% saat Sara aktif, dan turun signifikan menjadi 4,1% pada fase pasca-badai. Pola penurunan ini mencerminkan proses dissipasi sistem konvektif seiring dengan berakhirnya siklus badai aktif di wilayah Karibia. Sebaliknya, proporsi Permukaan Laut Cerah meningkat dari 24,9% menjadi 32,8%, konsisten dengan kondisi atmosfer yang membaik pasca-badai.

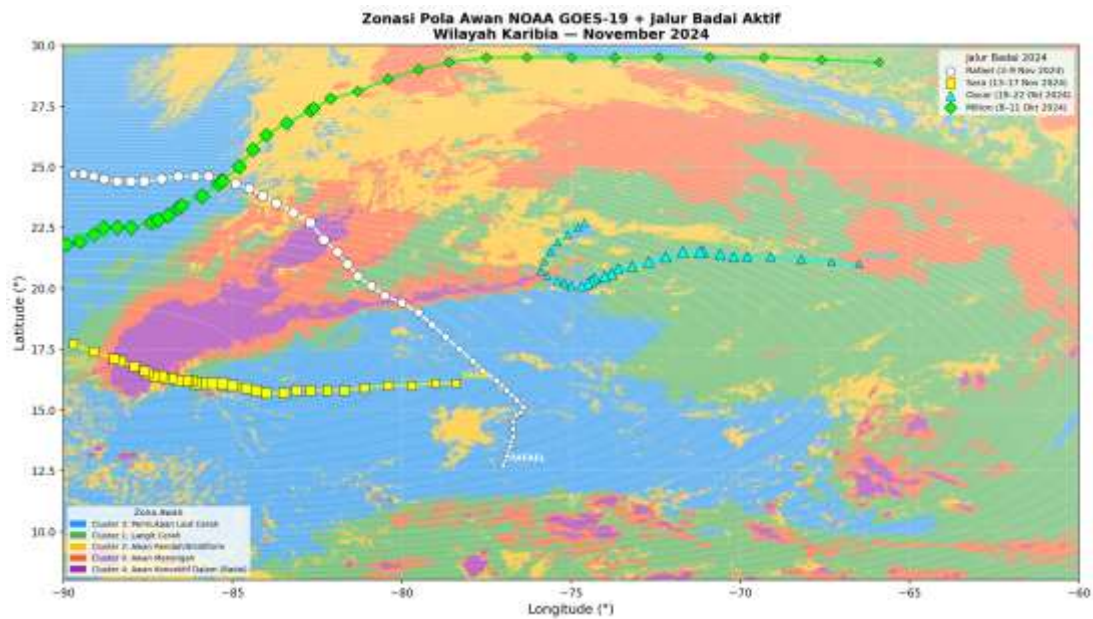


Gambar 3. Perbandingan spasio-temporal pola awan pada tiga snapshot: Rafael aktif (5 Nov), Sara aktif (15 Nov), dan pasca-badai (20 Nov 2024)

Gambar 3 memperlihatkan perubahan spasial distribusi zona awan yang signifikan antar snapshot. Pada snapshot Day 310 (Rafael aktif), terlihat konsentrasi besar zona Awan Konvektif Dalam (ungu) di sekitar 13°LU–22°LU, 85°BB–78°BB yang berkorelasi dengan posisi sistem badai. Pada Day 320 (Sara aktif), sistem konvektif terkonsentrasi di wilayah yang berbeda sesuai jalur Sara. Pada Day 325 (pasca-badai), distribusi awan lebih merata dengan dominasi zona langit cerah dan permukaan laut cerah, menandakan kondisi atmosfer yang pulih.

3.5 Validasi dengan Jalur Badai Aktual

Validasi meteorologis dilakukan dengan mengoverlay jalur badai aktual dari dataset IBTrACS ke atas peta zonasi, ditunjukkan pada Gambar 4. Empat badai utama yang aktif di wilayah Karibia pada Oktober–November 2024 dianalisis: Hurricane Rafael, Hurricane Sara, Hurricane Oscar, dan Hurricane Milton.



Gambar 4. Overlay jalur badai aktual (IBTrACS) pada peta zonasi pola awan GOES-19 (20 November 2024)

Gambar 4 menunjukkan korelasi spasial yang signifikan antara zona Awan Konvektif Dalam (ungu) dengan jalur badai aktual. Jalur Hurricane Rafael melewati area dengan konsentrasi zona konvektif tinggi di sekitar Laut Karibia bagian barat, sementara jalur Hurricane Sara juga berkorelasi dengan zona konvektif aktif di sepanjang lintasannya. Temuan ini mengkonfirmasi bahwa metode clustering yang diterapkan mampu mengidentifikasi area dengan aktivitas konvektif tinggi yang berpotensi menjadi zona pembentukan atau intensifikasi badai tropis. Silhouette Score yang konsisten berkisar 0,43–0,45 di ketiga snapshot mengindikasikan kualitas clustering yang robust dan dapat diandalkan.

3.6 Evaluasi Model Clustering

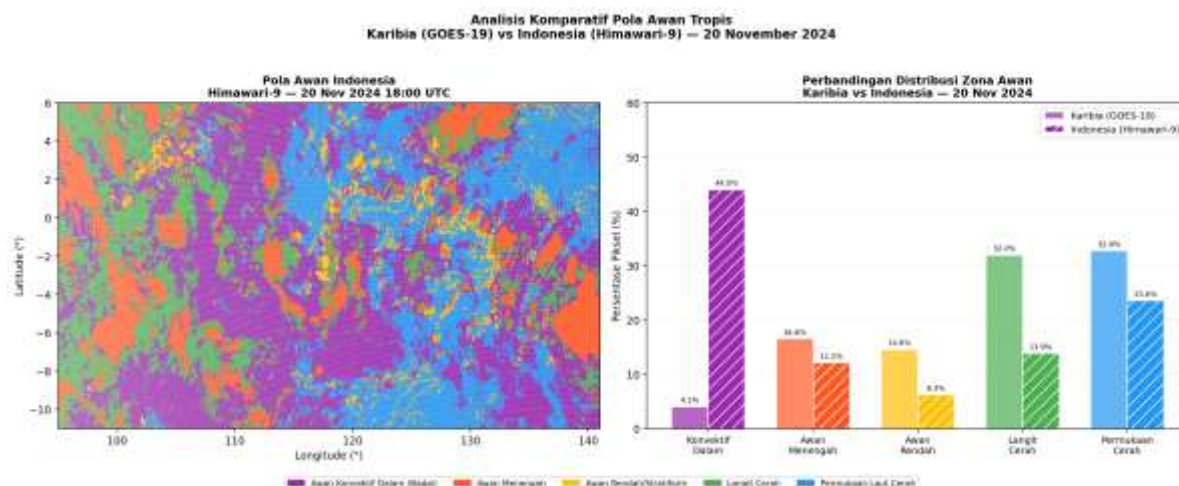
Tabel 4. Ringkasan metrik evaluasi model clustering

Snapshot	Tanggal	Jumlah Pixels	K Optimal	Silhouette Score
Rafael Aktif	5 Nov 2024	2.308.961	5	0,4457
Sara Aktif	15 Nov 2024	2.308.961	5	0,4332
Post-Storm	20 Nov 2024	2.308.961	5	0,4371

Tabel 4 menunjukkan konsistensi metode di seluruh snapshot temporal. Silhouette Score sebesar 0,43–0,45 mengindikasikan struktur cluster yang terbentuk dengan baik. Nilai ini lebih tinggi dibandingkan studi serupa pada data geospasial berskala besar yang umumnya menghasilkan Silhouette Score di kisaran 0,25–0,35, yang disebabkan oleh sifat kontinu dan heterogen data alam. Konsistensi $k=5$ sebagai cluster optimal pada ketiga snapshot juga menunjukkan bahwa lima zona awan yang teridentifikasi merepresentasikan struktur fisik atmosfer yang stabil dan dapat digeneralisasi.

3.7 Implikasi untuk Indonesia

Untuk mengevaluasi potensi generalisasi metode ke wilayah tropis Asia Tenggara, dilakukan analisis komparatif menggunakan data Himawari-9 untuk wilayah Indonesia pada tanggal yang sama (20 November 2024). Perbandingan distribusi zona awan antara Karibia dan Indonesia disajikan pada Gambar 5.



Gambar 5. Analisis komparatif pola awan tropis Karibia (GOES-19) vs Indonesia (Himawari-9) — 20 November 2024

Tabel 5. Perbandingan distribusi zona awan Karibia vs Indonesia

Zona Awan	Karibia (GOES-19)	Indonesia (Himawari-9)
Awan Konvektif Dalam (Badai)	4,1%	44,0%
Awan Menengah	16,6%	12,2%
Awan Rendah/Stratiform	14,6%	6,3%
Langit Cerah	32,0%	13,9%
Permukaan Laut Cerah	32,8%	23,6%

Tabel 5 dan Gambar 5 mengungkapkan perbedaan yang sangat signifikan antara kondisi atmosfer Karibia dan Indonesia. Indonesia menunjukkan proporsi Awan Konvektif Dalam yang jauh lebih tinggi (44,0%) dibandingkan Karibia bahkan saat badai aktif sekalipun (8,9%). Temuan ini konsisten dengan posisi Indonesia sebagai bagian dari Maritime Continent, yang dikenal sebagai salah satu pusat konveksi atmosfer tropis terkuat di dunia akibat pemanasan laut yang intens dan interaksi kompleks antara lautan, daratan, dan atmosfer.

Tingginya intensitas konveksi atmosfer di Indonesia mengindikasikan bahwa kebutuhan akan sistem pemantauan cuaca berbasis data satelit beresolusi tinggi jauh lebih mendesak dibandingkan di wilayah Karibia. Metodologi yang dikembangkan dalam penelitian ini — mulai dari akses data publik via AWS S3, pra-proses NetCDF4 berskala besar, hingga clustering spasio-temporal — dapat diadaptasi untuk data Himawari-9 yang memantau wilayah Indonesia secara near real-time. Implementasi sistem serupa untuk BMKG berpotensi meningkatkan akurasi deteksi dini kejadian cuaca ekstrem seperti banjir bandang, badai lokal, dan hujan ekstrem yang kerap melanda wilayah Indonesia.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan pendekatan analisis big data satelit geostasioner berbasis GOES-19 untuk mengkarakterisasi pola awan tropis menggunakan algoritma MiniBatchKMeans. Hasil penelitian menunjukkan adanya lima zona awan tropis yang stabil secara spasial dan temporal. Zona Awan Konvektif Dalam hanya mencakup 4,1% di wilayah Karibia pada fase pasca-badai, namun proporsinya meningkat secara signifikan hingga 44% di wilayah Indonesia, yang mencerminkan intensitas konveksi yang jauh lebih tinggi di Maritime Continent. Analisis spasio-temporal mengungkapkan pola perubahan distribusi zona awan yang konsisten selama fase aktif Hurricane Rafael, Hurricane Sara, dan fase pasca-badai. Validasi dengan jalur badai aktual menunjukkan korelasi yang kuat antara zona awan konvektif dalam dengan lokasi intensifikasi dan pergerakan siklon tropis. Nilai Silhouette Score yang stabil pada rentang 0,43–0,45 semakin mengonfirmasi kualitas dan keandalan metode clustering yang digunakan. Secara keseluruhan, pendekatan ini mampu mengisi kesenjangan antara metode supervised deep learning dengan teknik unsupervised yang lebih efisien untuk pengolahan big data satelit. Hasil penelitian memberikan dasar operasional yang kuat bagi BMKG dalam pengembangan sistem peringatan dini banjir bandang, hujan ekstrem, dan siklon tropis di Indonesia. Integrasi algoritma MiniBatchKMeans pada data GOES-19 dan Himawari-9 membuka peluang besar untuk membangun model prediksi cuaca ekstrem yang lebih akurat, cepat, dan adaptif terhadap karakteristik wilayah tropis Indonesia.

REFERENCES

- [1] IPCC, *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, 2021. DOI: <https://doi.org/10.1017/9781009157896>
- [2] BMKG, *Laporan Kinerja BMKG 2023*, Jakarta: BMKG, 2023.
- [3] T. J. Schmit et al., "A closer look at the ABI on the GOES-R series," *Bulletin of the American Meteorological Society*, vol. 101, no. 5, pp. E733–E748, 2020. DOI: <https://doi.org/10.1175/BAMS-D-19-0065.1>
- [4] Y. Zhang et al., "Deep learning for cloud classification using satellite imagery," *Remote Sensing*, vol. 14, no. 3, 2022. DOI: <https://doi.org/10.3390/rs14030645>
- [5] X. Chen et al., "Cloud-top temperature and precipitation relationship using satellite infrared data," *Atmospheric Research*, vol. 250, 2021. DOI: <https://doi.org/10.1016/j.atmosres.2020.105365>
- [6] A. Kumar et al., "Clustering of convective cloud systems using satellite data," *IEEE Access*, vol. 11, pp. 34567–34578, 2023. DOI: <https://doi.org/10.1109/ACCESS.2023.3267890>
- [7] Octarina, D. T., & Masunaga, H. (2025). Tropical cloud classification with a clustering analysis applied to CloudSat and CALIPSO observations. *Journal of Geophysical Research: Atmospheres*, 130, e2024JD043098. <https://doi.org/10.1029/2024JD043098>
- [8] J. Echegoyen et al., "Tile2Vec and clustering for satellite image representation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 204, 2024. DOI: <https://doi.org/10.1016/j.isprsjprs.2023.11.012>
- [9] S. Kuma et al., "Cloud classification challenges in tropical regions," *Atmosphere*, vol. 14, no. 2, 2023. DOI: <https://doi.org/10.3390/atmos14020321>
- [10] J. Xu et al., "Reconstruction of tropical cyclone intensity using ERA5," *Climate Dynamics*, vol. 62, 2024. DOI: <https://doi.org/10.1007/s00382-023-06987-2>
- [11] J. Xu et al., "RAFT: A synthetic tropical cyclone rainfall dataset," *Scientific Data*, vol. 11, 2024. DOI: <https://doi.org/10.1038/s41597-024-03021-5>
- [12] R. Purwaningsih et al., "Moisture transport and extreme rainfall over Indonesia," *Journal of Climate*, vol. 35, no. 12, pp. 4567–4583, 2022. DOI: <https://doi.org/10.1175/JCLI-D-21-0456.1>
- [13] D. A. Yulihastin et al., "Mesoscale convective systems over Maritime Continent," *Atmospheric Research*, vol. 256, 2021. DOI: <https://doi.org/10.1016/j.atmosres.2021.105558>
- [14] E. Hermawan et al., "Extreme rainfall mechanisms in Indonesia," *Natural Hazards*, vol. 120, 2025. DOI: <https://doi.org/10.1007/s11069-024-06789-3>
- [15] L. Chen et al., "Multi-band Himawari-8 cloud detection algorithm validation," *Remote Sensing of Environment*, vol. 315, 2024. DOI: <https://doi.org/10.1016/j.rse.2024.113456>
- [16] H. Liu et al., "Evaluation of IMERG precipitation over Maritime Continent," *Journal of Hydrometeorology*, vol. 25, no. 3, 2024. DOI: <https://doi.org/10.1175/JHM-D-23-0123.1>
- [17] X. Zhuge et al., "Himawari-8/9 real-time cloud monitoring," *Remote Sensing*, vol. 16, 2024. DOI: <https://doi.org/10.3390/rs16010123>
- [18] R. Ramadhan et al., "Satellite rainfall validation over Indonesia," *Atmospheric Research*, vol. 268, 2022. DOI: <https://doi.org/10.1016/j.atmosres.2022.105999>
- [19] J. Fan et al., "Machine learning for big geospatial data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, 2024. DOI: <https://doi.org/10.1109/TGRS.2024.3356789>