

## Analisis Klasterisasi dan Prediksi Hujan Lebat Petir Menggunakan Model K-Means, Probit, Logit

### *Clustering Analysis and Heavy Thunderstorm Prediction Using K-Means, Probit, dan Logit*

Yussi Dwi Hastuti<sup>1\*</sup>, Arief Wibowo<sup>2</sup>

<sup>1,2</sup>Program Studi Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Raya Ciledug Raya, Jakarta Selatan 12260, Indonesia.

\*Corresponding author: [2311602037@student.budiluhur.ac.id](mailto:2311602037@student.budiluhur.ac.id)

#### ABSTRAK

DOI:  
[10.30595/jrst.v10i1.27174](https://doi.org/10.30595/jrst.v10i1.27174)

#### Article information:

Received:  
26/06/2025

Revised:  
13/12/2025

Accepted:  
22/01/2026

Cuaca di Indonesia sering dipengaruhi oleh fenomena musiman, seperti puncak musim hujan, yang memiliki pola sangat dinamis. Kondisi tersebut kerap menimbulkan cuaca ekstrem, seperti hujan lebat disertai petir, yang berdampak pada aktivitas masyarakat sehari-hari. Oleh karena itu, penting untuk mengidentifikasi pola cuaca dan memprediksi potensi hujan petir secara akurat, baik untuk mitigasi risiko maupun perencanaan aktivitas masyarakat. Penelitian ini bertujuan memprediksi hujan petir pada hari ke-8 serta selama tujuh hari ke depan. Selain itu, penelitian juga bertujuan mengidentifikasi provinsi yang berpotensi mengalami hujan petir pada periode tersebut. Metode yang digunakan adalah klasterisasi K-Means dengan teknik Elbow untuk menentukan jumlah kluster optimal. Prediksi cuaca dilakukan menggunakan model logit dan probit dengan ambang batas (*threshold*) 0,2. Hasil penelitian menunjukkan jumlah kluster optimal adalah lima. Prediksi untuk hari ke-8 mengindikasikan terdapat dua kluster yang berpotensi mengalami hujan petir dengan probabilitas masing-masing sebesar 0,2 dan 0,3. Prediksi selama tujuh hari ke depan menunjukkan sebanyak 14 provinsi berpotensi mengalami hujan petir dengan tingkat probabilitas bervariasi antara 0,01 hingga 0,11. Penelitian ini memberikan gambaran yang jelas mengenai potensi hujan petir di berbagai wilayah Indonesia. Dengan memahami pola cuaca ini, diharapkan masyarakat dapat mempersiapkan diri dengan lebih baik dan mengurangi risiko yang mungkin timbul akibat cuaca ekstrem.

**Kata Kunci:** Hujan Petir, Klasterisasi, K-Means, Logit, Probit.

#### ABSTRACT

*Weather in Indonesia is often influenced by seasonal phenomena, such as the peak of the rainy season, which exhibits highly dynamic patterns. These conditions frequently lead to extreme weather events, such as heavy rain accompanied by thunderstorms, affecting daily community activities. Therefore, it is essential to identify weather patterns and accurately predict the potential for thunderstorms, both for risk mitigation and activity planning. This study aims to predict thunderstorms on the 8th day and over the following seven days. Furthermore, it seeks to identify which provinces are likely to experience thunderstorms during this period. The method employed is K-Means clustering with the Elbow technique to determine the optimal number of*

clusters. Weather prediction is then performed using logit and probit models with a threshold of 0.2. The results indicate that the optimal number of clusters is five. Predictions for the 8th day show that two clusters have the potential to experience thunderstorms, with probabilities of 0.2 and 0.3, respectively. Forecasts for the next seven days reveal that 14 provinces are likely to experience thunderstorms, with probabilities ranging from 0.01 to 0.11. This study provides a clear overview of thunderstorm potential across various regions in Indonesia. By understanding these weather patterns, communities are expected to better prepare and reduce risks associated with extreme weather conditions.

**Keywords:** Clustering, K-Means, Logit, Probit, Thunderstorms.

## 1. PENDAHULUAN

Pola cuaca di Indonesia sangat dinamis dan sering kali dipengaruhi oleh fenomena musiman seperti puncak musim hujan. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) secara rutin mengeluarkan peringatan mengenai potensi cuaca ekstrem, terutama selama puncak musim hujan yang biasanya terjadi pada bulan Desember hingga Februari (Fitria, 2024; Luthfiarta, Febriyanto, Lestiawan, & Wicaksono, 2020; Putratama, 2024). Dalam konteks ini, penting untuk mengidentifikasi pola cuaca dan melakukan prediksi potensi hujan petir secara tepat, baik untuk mitigasi risiko maupun perencanaan aktivitas masyarakat.

Penelitian terkait cuaca telah dilakukan dengan menggunakan analisis regresi linear berganda (Liyew & Melese, 2021; Luthfiarta et al., 2020), Naïve Bayes dan C4.5 (Yani, Aradea, & Mubarak, 2022), K-Means (Dinar Ajeng, Irwansyah, & Rina, 2021), K-Means dan Naïve Bayes (Raghavendran, 2019). K-Means juga digunakan untuk mengklasifikasikan dan memprediksi cuaca dengan menggunakan data histori cuaca (Amin, Rusydiyah, & Azizah, 2025; Kalude, Titaley, & Kalua, 2022; Neumann, Turowski, Mikut, Hagenmeyer, & Ludwig, 2023; Zhan, Qin, (Leon) Wang, & Hassan, 2025), serta K-Means dan Entropi untuk klusterisasi (Chusyairi, 2023). Salah satu metode yang digunakan dalam klusterisasi adalah metode *elbow* dalam K-Means, yang membantu menentukan jumlah kluster optimal dengan menganalisis titik "elbow" pada plot jumlah kluster terhadap total *within sum of squares* (Cui, 2020; Shi et al., 2021). Beberapa studi telah menerapkan algoritma K-Means untuk analisis pola cuaca, misalnya dengan mengelompokkan data cuaca harian di Kota Palembang menggunakan algoritma K-Means dan menemukan bahwa pendekatan ini efektif dalam mengidentifikasi pola karakteristik cuaca

berdasarkan kesamaan atribut meteorologi (Khairunnisa & Jambak, 2022).

Selain menggunakan klusterisasi, analisis regresi juga digunakan dalam prediksi cuaca. Regresi Logit menggunakan fungsi logistik, sedangkan regresi Probit menggunakan distribusi normal kumulatif. Keduanya banyak digunakan dalam studi meteorologi. Penelitian terbaru menunjukkan bahwa model regresi ini dapat secara efektif memprediksi kejadian cuaca ekstrem dengan menganalisis berbagai parameter atmosfer. Sebagai contoh, model Logit dan Probit telah digunakan untuk memprediksi kemungkinan terjadinya badai petir berdasarkan faktor-faktor seperti suhu, kelembapan, dan kecepatan angin (David, Alonso-Montesinos, Le Gal La Salle, & Lauret, 2023; Pan, Porth, & Li, 2022).

Artikel ini bertujuan untuk menganalisis data cuaca harian yang mencakup variabel kelembapan, suhu, dan kecepatan angin untuk setiap provinsi di Indonesia. Melalui pendekatan data mining, metode klusterisasi K-Means digunakan untuk mengelompokkan provinsi ke dalam lima kluster berdasarkan karakteristik cuaca yang mirip. Selain itu, kode cuaca akan dikategorikan ulang, di mana hujan lebat dan petir diberi nilai 1, sedangkan kondisi lainnya diberi nilai 0.

Hasil klusterisasi ini kemudian menjadi dasar untuk prediksi probabilitas kejadian hujan petir menggunakan model regresi probit dan logit yang belum pernah dilakukan sebelumnya dalam memprediksi cuaca. Regresi probit dan logit digunakan untuk menganalisis dependen variabel yang bersifat kualitatif (Ramadhani, Putri, & Raihani, 2022). Kedua model ini sangat cocok untuk data bertipe kategorikal, seperti dalam penelitian prediksi hujan petir (1 untuk true dan 0 untuk false).

Analisis dilakukan dengan fokus pada rata-rata kejadian per provinsi. Pertanyaan utama yang ingin dijawab dalam artikel ini adalah provinsi mana yang memiliki potensi hujan petir pada hari ke-8 dan selama 7 hari ke depan. Untuk menjalankan metode *K-Means* serta regresi Probit dan Logit, digunakan bahasa pemrograman Python. Python dipilih karena kemampuannya dalam menyediakan *library* yang mendukung berbagai teknik pemodelan statistik dan *machine learning*.

Melalui pendekatan ini, diharapkan penelitian ini dapat memberikan wawasan lebih mendalam mengenai pola cuaca ekstrem di Indonesia selama puncak musim hujan, sehingga mendukung pengambilan keputusan berbasis data untuk berbagai pihak, termasuk BMKG, pemerintah daerah, dan masyarakat luas.

## 2. METODE PENELITIAN

Metodologi penelitian ini menggunakan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang merupakan kerangka kerja standar untuk proyek data mining (Schröer, Kruse, & Gómez, 2021). CRISP-DM terdiri dari enam fase utama: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Setiap fase memiliki tujuan dan langkah-langkah spesifik yang harus dilakukan untuk mencapai hasil yang diinginkan. Namun, penelitian ini tidak mencakup tahap *Deployment* karena artikel ini berfokus pada penerapan metode dan analisis data.

### 2.1 Business Understanding

Pada fase ini, tujuan utama adalah memahami tujuan bisnis dan menentukan masalah yang akan dipecahkan. Dalam konteks penelitian ini, tujuan bisnis adalah untuk mengidentifikasi dan memprediksi cuaca, khususnya potensi hujan petir, pada hari ke-8 dan selama 7 hari ke depan.

Penelitian ini bertujuan untuk memberikan informasi yang akurat dan dapat diandalkan mengenai kondisi cuaca ekstrem, yang sangat penting bagi berbagai pihak. Dengan prediksi yang tepat, diharapkan dapat membantu dalam mitigasi risiko bencana, perencanaan aktivitas sehari-hari, serta pengambilan keputusan yang lebih baik.

Untuk mencapai tujuan ini, penelitian akan menggunakan data cuaca harian yang mencakup variabel kelembapan, suhu, dan kecepatan angin untuk setiap provinsi di Indonesia. Data ini akan dianalisis menggunakan metode klusterisasi *K-Means* untuk mengelompokkan provinsi berdasarkan karakteristik cuaca yang mirip. Selanjutnya, model regresi probit dan logit akan digunakan untuk memprediksi probabilitas kejadian hujan lebat petir.

Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai pola cuaca ekstrem di Indonesia selama puncak musim hujan, sehingga mendukung pengambilan keputusan berbasis data yang lebih efektif

### 2.2 Data Understanding

Pada tahap ini, kita akan memahami struktur dan karakteristik data yang digunakan dalam penelitian ini. Dataset yang digunakan merupakan dataset yang direkam per jam mulai dari tanggal 11/12/2024 jam 00.00.00 sampai dengan tanggal 21/12/2024 jam 00.00.00 bersumber dari dataset Prakiraan Cuaca Terbuka Badan Meteorologi, Klimatologi, dan Geofisika (Kecamatan) dengan alamat web data.bmkg.go.id terdiri dari berbagai jumlah entri yang direkam pada level Kecamatan kemudian diagregasikan ke level Provinsi. **Tabel 1** menampilkan ringkasan jumlah entri setiap provinsi di Indonesia dan masing - masing provinsi memiliki atribut/variabel sebanyak 11 dengan penjelasan mengenai masing-masing variabel di dalam dataset sebagai berikut:

- *id\_lokasi*: Kolom ini berisi ID unik untuk setiap lokasi pengamatan cuaca. ID ini memiliki tabel referensi yang berisi informasi detail mengenai lokasi, seperti nama provinsi, kabupaten dan kecamatan.
- *waktu*: Kolom ini mencatat waktu pengamatan cuaca dalam format tanggal (YYY-MM-DD HH:MM:SS).
- *suhu\_avg\_min*: Kolom ini mencatat suhu rata-rata minimum harian. Namun, data hanya tersedia di jam 12.00 saja sehingga terdapat banyak nilai yang kosong (missing values) dalam kolom ini.
- *suhu\_avg\_max*: Kolom ini mencatat suhu rata-rata maksimum harian, dengan kondisi yang sama seperti kolom *suhu\_avg\_min*.

- kelembapan\_avg\_min: Kolom ini mencatat kelembapan rata-rata minimum harian, dengan kondisi yang sama seperti kolom suhu\_avg\_min.
- kelembapan\_avg\_max: Kolom ini mencatat kelembapan rata-rata maksimum harian, juga dengan kondisi yang sama seperti kolom suhu\_avg\_min.
- kelembapan(%): Kolom ini mencatat kelembapan harian dalam persen, dengan data yang lengkap.
- suhu(celcius): Kolom ini mencatat suhu harian dalam derajat Celsius, dengan data yang lengkap.
- kd\_cuaca: Kolom ini berisi kode cuaca yang mengindikasikan kondisi cuaca tertentu. Kode ini memiliki tabel referensi yang menjelaskan arti dari setiap kode cuaca.
- arah\_angin: Kolom ini mencatat arah angin harian. Arah angin memiliki tabel referensi yang menjelaskan arah angin dalam bentuk deskriptif.
- kecepatan\_angin: Kolom ini mencatat kecepatan angin harian dalam satuan meter per detik (m/s), dengan data yang lengkap.

Dataset mengandung tipe data numerik dan kategori dimana variable dengan tipe data numerik adalah id\_lokasi, kelembapan(%), suhu(celcius), kd\_cuaca, dan kecepatan angin. Sedangkan untuk variable dengan tipe data kategori adalah waktu, suhu\_avg\_min, suhu\_avg\_max, kelembapan\_avg\_min, kelembapan\_avg\_max, arah\_angin, kecamatan, kabupaten, dan provinsi.

Dengan memahami struktur dan karakteristik data ini, kita dapat melakukan analisis lebih lanjut untuk mengidentifikasi pola cuaca dan memprediksi potensi hujan petir. Langkah selanjutnya adalah melakukan pemilihan variabel yang akan digunakan dan melakukan transformasi data yang diperlukan untuk analisis lebih lanjut.

### 2.3 Data Preparation

Pada tahap ini, dilakukan beberapa langkah untuk mempersiapkan data agar siap digunakan dalam analisis lebih lanjut. Langkah-langkah yang dilakukan meliputi: pertama load data \*csv per provinsi dan \*csv referensi lokasi menjadi dataframe jadi total dataframe yang terbentuk sebanyak 35. Selanjutnya melakukan konversi kolom waktu ke dalam format *datetime*

untuk memudahkan analisis berbasis waktu. Hal ini penting untuk memastikan bahwa data waktu dapat diolah dengan benar dalam analisis statistik dan prediksi.

**Tabel 1** Jumlah Entri per Provinsi

No.	Provinsi	Jml Entri	No.	Provinsi	Jml Entri
1	Aceh	36504	18	Kep. Riau	8541
2	Bali	9594	19	Lampung	28080
3	Banten	19188	20	Maluku	14859
4	Bengkulu	16146	21	Maluku Utara	14157
5	Yogyakarta	9477	22	Nusa Tenggara Barat	14742
6	DKI Jakarta	5850	23	Nusa Tenggara Timur	28259
7	Gorontalo	9711	24	Papua	43759
8	Jambi	16029	25	Papua Barat	21294
9	Jawa Barat	76284	26	Riau	18837
10	Jawa Tengah	68796	27	Sulawesi Barat	8658
11	Jawa Timur	81783	28	Sulawesi Selatan	38142
12	Kalimantan Barat	21060	29	Sulawesi Tengah	19422
13	Kalimantan Selatan	38610	30	Sulawesi Tenggara	24570
14	Kalimantan Tengah	17784	31	Sulawesi Utara	20709
15	Kalimantan Timur	12636	32	Sumatera Barat	22815
16	Kalimantan Utara	6786	33	Sumatera Selatan	28431
17	Bangka Belitung	7020	34	Sumatera Utara	48672

Kedua kode cuaca (kd\_cuaca) dikonversi berdasarkan referensi yang ada dengan membuat sebuah *function* yang diberi nama *categorize\_weather*. Cerah/Clear Skies: Kode 0 dan 100 menunjukkan kondisi cuaca cerah tanpa awan. Cerah Berawan/Partly Cloudy: Kode 1, 2, 101, dan 102 menunjukkan kondisi cuaca cerah dengan sedikit awan. Berawan/Mostly Cloudy: Kode 3 dan 103 menunjukkan kondisi cuaca berawan. Berawan Tebal/Overcast: Kode 4 dan 104 menunjukkan kondisi cuaca berawan tebal. Udara Kabur/Haze: Kode 5 menunjukkan kondisi udara kabur. Asap/Smoke: Kode 10 menunjukkan kondisi cuaca berasap. Kabur/Fog: Kode 45 menunjukkan kondisi cuaca berkabut. Hujan

Ringan/*Light Rain*: Kode 60 menunjukkan kondisi hujan ringan. Hujan Sedang/*Rain*: Kode 61 menunjukkan kondisi hujan sedang. Hujan Lebat/*Heavy Rain*: Kode 63 menunjukkan kondisi hujan lebat. Hujan Lokal/*Isolated Shower*: Kode 80 menunjukkan kondisi hujan lokal. Hujan Petir/*Severe Thunderstorm*: Kode 95 dan 97 menunjukkan kondisi hujan petir yang parah (Badan Meteorologi, 2024). Kolom kategori\_cuaca juga ditambahkan untuk menyimpan hasil konversi ini.

Ketiga data cuaca digabungkan dengan tabel referensi lokasi berdasarkan kolom id\_lokasi menjadi kolom kecamatan, kabupaten, provinsi di *dataframe* per provinsi. Tabel referensi lokasi berisi informasi detail mengenai lokasi, seperti nama kecamatan, kabupaten, dan provinsi. Penggabungan ini memungkinkan analisis cuaca per provinsi. Penggabungan dataset provinsi dengan referensi lokasi dapat dilakukan menggunakan fungsi *merge* pada *pandas*. Langkah awal pilih kolom yang dibutuhkan [id\_lokasi, kecamatan, kabupaten, provinsi] dari *dataframe* referensi lokasi. Lalu gunakan *Left Join* untuk menggabungkan dataset provinsi dengan referensi lokasi.

Keempat menambahkan kolom hujan\_petir untuk mengindikasikan kejadian hujan lebat dan petir. Nilai 1 diberikan untuk kode cuaca 63, 95, dan 97, atau kategori\_cuaca Hujan Lebat/*Heavy Rain* dan Hujan Petir/*Severe Thunderstorm* sedangkan nilai 0 diberikan untuk kondisi cuaca lainnya. Hal ini memudahkan identifikasi dan analisis kejadian hujan lebat dan petir.

Kelima menggabungkan semua dataset provinsi menjadi satu tabel yang komprehensif. Tabel ini mencakup semua variabel cuaca harian, informasi lokasi, kategori cuaca yang telah dikonversi dan hujan\_petir. Tabel ini siap digunakan untuk analisis lebih lanjut, termasuk klusterisasi dan prediksi cuaca.

Setelah dilakukan persiapan data, langkah selanjutnya adalah menganalisis statistik deskriptif dari data yang sudah diolah. Variabel yang digunakan untuk analisis statistik deskriptif meliputi provinsi, kelembapan(%), suhu(celcius), kecepatan\_angin, dan hujan\_petir. Hasil dari analisis statistik deskriptif ini disajikan pada **Tabel 2**, yang memberikan gambaran umum mengenai distribusi dan karakteristik data

cuaca di setiap provinsi. Analisis ini penting untuk memahami pola dasar dalam data sebelum melanjutkan ke tahap analisis yang lebih mendalam

**Tabel 2** Analisis Statistik Deskriptif

	mean	std	min	max
kelembapan(%)	88.75	8.98	43.0	100.0
suhu(celcius)	24.53	3.1	6.0	35.0
kecepatan_angin	6.42	5.83	0.0	79.0
hujan_petir	0.02	0.13	0.0	1.0

Langkah terakhir dalam tahap data preparation adalah analisis korelasi untuk variable kelembapan(%), suhu(celcius), kecepatan\_angin, dan hujan\_petir dengan menghitung korelasi per provinsi sehingga didapatkan hasil untuk 34 provinsi. **Tabel 3** adalah contoh hasil analisis korelasi yang menunjukkan hubungan antara variabel-variabel tersebut di salah satu provinsi.

**Tabel 3** Hasil Analisis Korelasi

Korelasi untuk Provinsi: Aceh				
	ax	bx	cx	dx
ax	1.0	-0.61	-0.4	0.01
bx	-0.61	1.0	0.49	-0.0
cx	-0.4	0.49	1.0	0.0
dx	0.01	-0.0	0.0	1.0

\* ax: kelembapan(%), bx: suhu(celcius), cx: kecepatan\_angin, dx: hujan\_petir

## 2.4 Modeling

Pada tahap ini, dilakukan pemilihan dan penerapan teknik pemodelan yang sesuai untuk mencapai tujuan penelitian. Dalam konteks penelitian ini, tujuan utama adalah memprediksi potensi hujan petir pada hari ke-8 dan selama 7 hari ke depan berdasarkan data cuaca harian. Berikut adalah langkah-langkah yang dilakukan dalam tahap pemodelan:

### 2.4.1 Clustering/klusterisasi

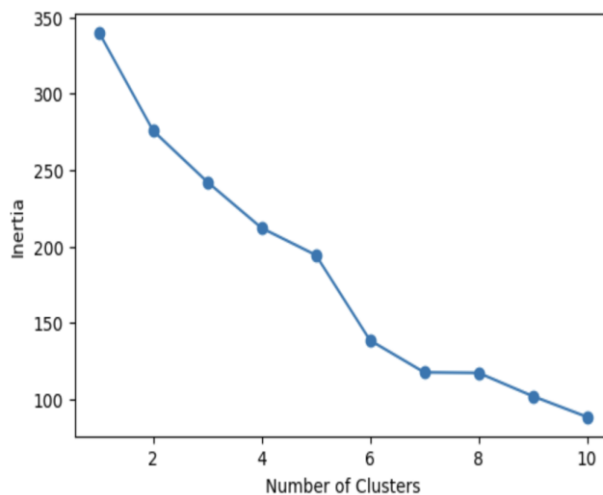
Dalam tahap ini menggunakan algoritma *K-Means* dimana *K-Means* mampu menghasilkan cluster yang lebih konsisten (Cahyanto, Chrismanto, & Sebastian, 2020), *K-Means* digunakan untuk mengelompokkan provinsi berdasarkan karakteristik cuaca yang mirip. Dengan menggunakan metode elbow untuk menentukan jumlah kluster optimal. Langkah pertama dalam proses *clustering* adalah dengan menstandarisasikan data dengan menggunakan

library *StandardScaler* yang berfungsi untuk mengubah data agar memiliki distribusi mean = 0 dan std = 1 melalui persamaan (1)

$$Z = \frac{(x-u)}{s} \quad (1)$$

Dimana  $x$  merupakan nilai yang ingin distandarkan,  $u$ (*mean*) dari seluruh data yang ingin distandarkan, dan  $s$ (*standar deviation*) standar deviasi dari kolom yang ingin distandarkan(Learn, 2025).

Setelah data distandarkan, langkah selanjutnya adalah menentukan jumlah kluster dengan menggunakan metode elbow. Pertama, tentukan rentang jumlah kluster yang akan diuji, kemudian lakukan pengulangan untuk setiap jumlah kluster dalam rentang tersebut dan simpan hasil pengulangan ke dalam sebuah daftar. Hasil grafik elbow terdapat pada **Gambar 1**.



**Gambar 1** Elbow Method

Untuk menentukan jumlah kluster yang optimal, dilakukan percobaan dengan berbagai

nilai  $k$  (jumlah kluster) mulai dari 1 hingga 10. Metode elbow digunakan untuk menganalisis inerti (total jarak antara titik data dan pusat kluster). Grafik elbow diplot untuk memvisualisasikan hubungan antara jumlah kluster dan inerti, dan jumlah kluster optimal ditentukan berdasarkan perubahan inerti yang signifikan. Dalam hal ini, berdasarkan grafik,  $k=5$  dipilih sebagai jumlah kluster yang optimal.

Setelah menentukan jumlah kluster optimal ( $k=5$ ), algoritma *K-Means* diterapkan pada data yang telah distandarisasi. Setiap provinsi kemudian diberikan label kluster yang mencerminkan kelompok yang paling mirip berdasarkan fitur yang telah dianalisis. Tujuan dari klasterisasi ini adalah untuk menyusun provinsi-provinsi ke dalam kluster-kluster berdasarkan kesamaan karakteristik cuaca.

Hasil klasterisasi kemudian ditampilkan dalam bentuk tabel yang berisi informasi tentang rata-rata, minimum, dan maksimum dari kelembapan, suhu, kecepatan angin, serta jumlah kejadian hujan petir untuk setiap provinsi. Kolom kluster menunjukkan kluster tempat provinsi tersebut ditempatkan. Terakhir, hasil klasterisasi ditampilkan menggunakan *tabulate*, yang menyajikan data dalam format tabel yang rapi dan mudah dibaca. Tabel ini menyajikan setiap provinsi beserta statistik yang dihitung dan kluster tempat provinsi tersebut dikelompokkan.

Selanjutnya, penulis menghitung statistik deskriptif per kluster. Hasil perhitungan ini disajikan dalam bentuk **Tabel 4** yang menunjukkan rata-rata, minimum, dan maksimum dari variabel kelembapan, suhu, kecepatan angin, serta jumlah kejadian hujan petir untuk setiap kluster. Tabel ini memberikan gambaran umum mengenai karakteristik cuaca di setiap kluster.

**Tabel 4** Statistik Deskriptif

kluster	KL* mean	KL* min	KL* max	Suhu mean	Suhu min	Suhu max	Kec* angin mean	Kec* angin min	Kec* angin max	Hujan petir count
0	90.45	89.76	91.85	24.9	24.27	25.37	5.44	4.5	7.25	6613.0
1	88.14	85.32	91.01	25.29	24.05	25.87	5.16	3.4	6.93	2562.0
2	87.88	84.81	90.44	25.5	23.52	26.58	10.36	8.02	13.05	1469.0
3	88.25	86.69	89.24	23.36	21.52	24.46	4.94	3.78	6.71	831.0
4	88.59	88.15	89.67	23.55	22.84	24.4	6.03	4.82	7.59	2389.0

\* Kl: Kelembapan, Kec:Kecepatan

Selain itu, data provinsi per kluster juga ditampilkan dalam **Tabel 5** yang terpisah. Tabel ini menyajikan informasi mengenai provinsi-provinsi yang termasuk dalam masing-masing kluster, sehingga memudahkan dalam melihat distribusi provinsi berdasarkan kesamaan karakteristik cuaca.

Dengan menampilkan statistik per kluster dan data provinsi per kluster, penulis dapat memberikan wawasan yang lebih mendalam mengenai pola cuaca di berbagai wilayah di Indonesia.

**Tabel 5** Anggota Kluster

kluster	provinsi
0	Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan
1	DI Yogyakarta, Gorontalo, Jambi, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Lampung, Maluku Utara, Riau, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sumatera Selatan
2	Bali, Banten, Bengkulu, DKI Jakarta, Kep. Bangka Belitung, Kep. Riau, Maluku, Nusa Tenggara Barat, Nusa Tenggara Timur
3	Papua, Papua Barat, Sulawesi Selatan
4	Aceh, Jawa Barat, Sulawesi Utara, Sumatera Barat, Sumatera Utara

2.4.2 Regresi

Sebagaimana telah disampaikan pada bagian sebelumnya, penelitian ini bertujuan untuk memprediksi cuaca, khususnya potensi hujan petir, pada hari ke-8 dan selama 7 hari ke depan. Oleh karena itu, Hal pertama yang harus dilakukan adalah mengidentifikasi variabel – variabel yang akan digunakan di dalam model regresi beserta karakteristiknya. Indikator kode cuaca yang telah dikelompokkan menjadi 2 kategori, yaitu 1 untuk cuaca hujan lebat dan petir, dan 0 untuk cuaca lainnya, dipilih sebagai variable dependent atau variabel target (Y). Selanjutnya, 3 variabel dipilih sebagai variabel independen (variabel X), terdiri dari kelembapan (%), suhu (celcius), dan kecepatan angin. Ketiga variabel X tersebut merupakan variabel numerik. Mengingat variabel target Y merupakan variabel kategori bernilai 1 atau 0, maka model regresi yang tepat digunakan sesuai dengan karakter

tersebut adalah model Logit dan Probit, dimana penulis akan memodelkan probabilitas kejadian hujan lebat dan petir atau tidak.

Dalam persamaan matematika, model Logit dapat digambarkan melalui persamaan berikut:

$$P(Y = 1|X) = \frac{1}{1+e^{-z}} \quad (2)$$

Dimana  $z = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3$  atau  $z$  merupakan kombinasi linier dari variabel prediktor X, sementara  $\beta_0$  adalah konstanta dan  $\beta_1, \beta_2, \beta_3$  adalah koefisien regresi untuk masing – masing variabel prediktor. Model ini menggunakan fungsi logistik untuk mengubah kombinasi linear dari variabel prediktor menjadi probabilitas yang berada dalam rentang 0 hingga 1. Dalam konteks prediksi cuaca, model Logit dapat digunakan untuk memprediksi probabilitas kejadian hujan petir berdasarkan variabel-variabel cuaca seperti kelembapan, suhu, dan kecepatan angin ( $P(Y = 1|X)$ ). Model ini sangat berguna karena dapat memberikan estimasi probabilitas yang dapat digunakan untuk pengambilan keputusan berbasis risiko.

Sementara itu, model Probit digunakan untuk memodelkan probabilitas kejadian suatu peristiwa biner (misalnya, hujan petir atau tidak) dengan menggunakan fungsi distribusi normal kumulatif. Persamaan dari model Probit adalah:

$$P(Y = 1|X) = \Phi(\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3) \quad (3)$$

Dimana  $\Phi$  adalah fungsi distribusi normal kumulatif. Perbedaan dengan model Logit hanya terletak pada asumsi distribusi yang digunakan dimana model Probit menggunakan fungsi distribusi normal kumulatif untuk mengubah kombinasi linear dari variabel prediktor menjadi probabilitas yang berada dalam rentang 0 hingga 1.

Langkah pertama dilakukan dengan membuat proses iterasi untuk setiap kluster dan melakukan merge data yang bertujuan untuk mengestimasi model regresi logit dan probit pada setiap kluster. Regresi dilakukan atas variable target Y yaitu hujan\_petir terhadap variabel X yang terdiri dari kelembapan (%), suhu (celcius), dan kecepatan

angin. Konstanta ditambahkan ke dalam model regresi.

Model logit dibuat dengan menggunakan *library* Statsmodels (sm.logit) dimana model ini digunakan untuk mengevaluasi hubungan antara fitur dan target secara logistic. Sedangkan Model probit menggunakan *library* Statsmodels (sm.probit).

Setelah memodelkan regresi Langkah selanjutnya adalah membuat prediksi cuaca untuk hari ke 8 dan 7 hari ke depan dengan menggunakan model regresi probit dan logit. Untuk memprediksi hari ke -8. Data cuaca dari 7 hari terakhir diambil untuk melakukan prediksi hari ke-8. Konstanta ditambahkan ke dalam data ini untuk keperluan prediksi. Model yang telah dilatih kemudian digunakan untuk memprediksi probabilitas kejadian hujan petir pada hari ke-8. lalu **menentukan threshold** rata-rata hujan\_petir per provinsi dan rata-rata keseluruhan dihitung untuk menentukan *threshold*. Prediksi (1 = ekstrim, 0 = bukan ekstrim) ditentukan berdasarkan threshold ini.

## 2.5 Evaluation

Dari estimasi model Logit dan Probit, diperoleh koefisien untuk masing-masing variabel prediktor, yaitu kelembapan (%), suhu (celcius), dan kecepatan angin. Koefisien ini menunjukkan seberapa besar pengaruh setiap variabel terhadap probabilitas kejadian hujan petir. Misalnya, koefisien positif pada variabel kelembapan menunjukkan bahwa peningkatan kelembapan cenderung meningkatkan probabilitas hujan petir. Sebaliknya, koefisien negatif pada variabel suhu menunjukkan bahwa peningkatan suhu cenderung menurunkan probabilitas hujan petir.

Berdasarkan hasil regresi tersebut, prediksi cuaca untuk hari ke-8 dan 7 hari ke depan dibuat. Model Logit dan Probit yang telah dilatih menggunakan data historis digunakan untuk memprediksi probabilitas kejadian hujan petir di setiap provinsi. Prediksi ini memberikan informasi penting yang dapat digunakan oleh BMKG, pemerintah daerah, dan masyarakat luas untuk perencanaan dan mitigasi risiko cuaca ekstrem. Prediksi ini juga membantu dalam pengambilan keputusan yang lebih baik terkait aktivitas sehari-hari dan kesiapsiagaan terhadap potensi bencana.

Hasil prediksi tersebut kemudian dievaluasi dengan membandingkannya dengan berita cuaca yang dimuat di [meteo.bmkg.go.id](http://meteo.bmkg.go.id) untuk tanggal yang bersangkutan. Evaluasi ini penting untuk memastikan akurasi dan keandalan model prediksi yang digunakan. Dengan membandingkan hasil prediksi dengan data aktual, penulis dapat mengidentifikasi kekuatan dan kelemahan model, serta melakukan perbaikan yang diperlukan untuk meningkatkan performa prediksi di masa mendatang. Evaluasi ini juga memberikan umpan balik yang berharga untuk pengembangan model prediksi cuaca yang lebih baik dan lebih akurat.

## 3. HASIL DAN PEMBAHASAN

Pada bab ini, akan dijelaskan hasil dari klusterisasi, regresi, dan prediksi yang telah dilakukan dalam penelitian ini. Hasil-hasil ini memberikan wawasan mendalam mengenai pola cuaca di Indonesia selama puncak musim hujan dan membantu dalam pengambilan keputusan berbasis data.

### 3.1. Klusterisasi

Hasil klusterisasi menggunakan metode K-Means menunjukkan bahwa provinsi-provinsi di Indonesia dapat dikelompokkan ke dalam lima klaster berdasarkan karakteristik cuaca yang mirip. Setiap klaster memiliki karakteristik unik yang tercermin dari rata-rata, minimum, dan maksimum nilai kelembapan, suhu, kecepatan angin, serta jumlah kejadian hujan petir. Tabel 6 hasil klusterisasi menampilkan informasi detail mengenai setiap klaster, termasuk provinsi-provinsi yang termasuk dalam masing-masing klaster. Klusterisasi ini membantu dalam memahami pola cuaca yang berbeda di berbagai wilayah Indonesia dan memberikan dasar untuk analisis lebih lanjut. Dari **Tabel 4** didapat kesimpulan sebagaimana pada **Tabel 6**.

Dengan hasil ini, kita dapat melihat perbedaan karakteristik cuaca di setiap klaster, yang membantu dalam memahami pola cuaca di berbagai wilayah Indonesia dan memberikan dasar untuk analisis lebih lanjut. Berdasarkan Tabel 5, diketahui bahwa Klaster 0 terdiri dari 4 provinsi, sedangkan Klaster 1 terdiri dari 13 provinsi.

**Tabel 6** Hasil klusterisasi

Kluster	Hasil
Kluster 0	<ul style="list-style-type: none"> <li>Kelembapan: Rata-rata 90.45%, dengan minimum 89.76% dan maksimum 91.85%.</li> <li>Suhu: Rata-rata 24.9°C, dengan minimum 24.27°C dan maksimum 25.37°C.</li> <li>Kecepatan Angin: Rata-rata 5.44 m/s, dengan minimum 4.5 m/s dan maksimum 7.25 m/s.</li> <li>Jumlah Hujan Petir: 6613 kejadian.</li> <li>Karakteristik: Klaster ini memiliki kelembapan yang sangat tinggi dan suhu yang relatif stabil, dengan kecepatan angin sedang.</li> </ul>
Kluster 1	<ul style="list-style-type: none"> <li>Kelembapan: Rata-rata 88.14%, dengan minimum 85.32% dan maksimum 91.01%.</li> <li><b>Suhu:</b> Rata-rata 25.29°C, dengan minimum 24.05°C dan maksimum 25.87°C.</li> <li>Kecepatan Angin: Rata-rata 5.16 m/s, dengan minimum 3.4 m/s dan maksimum 6.93 m/s.</li> <li>Jumlah Hujan Petir: 2562 kejadian.</li> <li>Karakteristik: Klaster ini memiliki kelembapan yang tinggi dan suhu yang sedikit lebih tinggi dibandingkan klaster lainnya, dengan kecepatan angin sedang.</li> </ul>
Kluster 2	<ul style="list-style-type: none"> <li>Kelembapan: Rata-rata 87.88%, dengan minimum 84.81% dan maksimum 90.44%.</li> <li>Suhu: Rata-rata 25.5°C, dengan minimum 23.52°C dan maksimum 26.58°C.</li> <li>Kecepatan Angin: Rata-rata 10.36 m/s, dengan minimum 8.02 m/s dan maksimum 13.05 m/s.</li> <li>Jumlah Hujan Petir: 1469 kejadian.</li> <li>Karakteristik: Klaster ini memiliki kelembapan yang tinggi dan suhu yang relatif tinggi, dengan kecepatan angin yang jauh lebih tinggi dibandingkan klaster lainnya.</li> </ul>
Kluster 3	<ul style="list-style-type: none"> <li>Kelembapan: Rata-rata 88.25%, dengan minimum 86.69% dan maksimum 89.24%.</li> <li>Suhu: Rata-rata 23.36°C, dengan minimum 21.52°C dan maksimum 24.46°C.</li> <li>Kecepatan Angin: Rata-rata 4.94 m/s, dengan minimum 3.78 m/s dan maksimum 6.71 m/s.</li> <li>Jumlah Hujan Petir: 831 kejadian.</li> <li>Karakteristik: Klaster ini memiliki kelembapan yang tinggi dan suhu yang lebih rendah dibandingkan klaster lainnya, dengan kecepatan angin yang relatif rendah.</li> </ul>
Kluster 4	<ul style="list-style-type: none"> <li>Kelembapan: Rata-rata 88.59%, dengan minimum 88.15% dan maksimum 89.67%.</li> <li>Suhu: Rata-rata 23.55°C, dengan minimum 22.84°C dan maksimum 24.4°C.</li> <li>Kecepatan Angin: Rata-rata 6.03 m/s, dengan minimum 4.82 m/s dan maksimum 7.59 m/s.</li> <li>Jumlah Hujan Petir: 2389 kejadian.</li> <li>Karakteristik: Klaster ini memiliki kelembapan yang tinggi dan suhu yang relatif rendah, dengan kecepatan angin sedang.</li> </ul>

**Tabel 7** Hasil Regresi

Klaster	Probit				Logit			
	const	Kel*(%)	Suhu (celcius)	Kec* angin	const	Kel*(%)	Suhu (celcius)	Kec* angin
0	0.3998	-0.0062	-0.0734	0.0152	2.0430	-0.0153	-0.1761	0.0343
1	-5.1515	0.0241	0.0303	-0.0064	-12.6734	0.0686	0.0857	-0.0158
2	-8.5235	0.0370	0.1023	0.0238	-19.4469	0.0879	0.2448	0.0619
3	-2-8939	0.0122	-0.0191	-0.0482	-6.4099	0.0354	-0.0474	-0.1504
4	-7.2623	0.0300	0.1003	-0.0134	-17.7778	0.0797	0.2685	-0.0365

\*Kel: Kelembapan, Kec:Kecepatan

### 3.2. Regresi

Regresi Probit dan Logit digunakan untuk memodelkan probabilitas kejadian hujan petir berdasarkan variabel-variabel cuaca seperti kelembapan, suhu, dan kecepatan angin. Dari estimasi model, diperoleh koefisien untuk masing-masing variabel yang menunjukkan pengaruhnya terhadap probabilitas hujan petir. Hasil regresi menunjukkan bahwa kelembapan memiliki pengaruh positif terhadap probabilitas

Berdasarkan analisis data cuaca dari tabel yang diberikan:

#### 1) Klaster 0:

- Probit: Kelembapan memiliki pengaruh negatif kecil (-0.0062), suhu memiliki pengaruh negatif (-0.0734), dan kecepatan angin memiliki pengaruh positif kecil (0.0152).
- Logit: Kelembapan memiliki pengaruh negatif kecil (-0.0153), suhu memiliki pengaruh negatif signifikan (-0.1761), dan kecepatan angin memiliki pengaruh positif kecil (0.0343).

#### 2). Klaster 1:

- Probit: Kelembapan memiliki pengaruh positif kecil (0.0241), suhu memiliki pengaruh positif kecil (0.0303), dan kecepatan angin memiliki pengaruh negatif kecil (-0.0064).
- Logit: Kelembapan memiliki pengaruh positif sedang (0.0686), suhu memiliki pengaruh positif sedang (0.0857), dan kecepatan angin memiliki pengaruh negatif kecil (-0.0158).

#### 3). Klaster 2:

- Probit: Kelembapan memiliki pengaruh positif sedang (0.0370), suhu memiliki pengaruh positif signifikan (0.1023), dan kecepatan angin memiliki pengaruh positif kecil (0.0238).

hujan petir, sementara suhu memiliki pengaruh negatif. Kecepatan angin juga menunjukkan pengaruh yang signifikan. Model regresi ini memberikan dasar yang kuat untuk melakukan prediksi cuaca yang akurat. Koefisien masing – masing variabel independent hasil regresi model ditampilkan pada **Tabel 7** dengan variabel independen (dari kiri ke kanan) terdiri dari: konstanta (const), kelembapan, suhu (Celsius), dan kecepatan angin.

- Logit: Kelembapan memiliki pengaruh positif signifikan (0.0879), suhu memiliki pengaruh positif signifikan (0.2448), dan kecepatan angin memiliki pengaruh positif sedang (0.0619).

#### 4). Klaster 3:

- Probit: Kelembapan memiliki pengaruh positif kecil (0.0122), suhu memiliki pengaruh negatif kecil (-0.0191), dan kecepatan angin memiliki pengaruh negatif sedang (-0.0482).
- Logit: Kelembapan memiliki pengaruh positif kecil (0.0354), suhu memiliki pengaruh negatif kecil (-0.0474), dan kecepatan angin memiliki pengaruh negatif signifikan (-0.1504).

#### 5). Klaster 4:

- Probit: Kelembapan memiliki pengaruh positif sedang (0.0300), suhu memiliki pengaruh positif signifikan (0.1003), dan kecepatan angin memiliki pengaruh negatif kecil (-0.0134).
- Logit: Kelembapan memiliki pengaruh positif signifikan (0.0797), suhu memiliki pengaruh positif signifikan (0.2685), dan kecepatan angin memiliki pengaruh negatif kecil (-0.0365).

**Tabel 8** Prediksi hari ke-8

Logit			Probit		
Klaster	Prediksi hari ke-8	Probability	Klaster	Prediksi hari ke-8	Probability
0	1.0	0.03	0	1.0	0.03
1	1.0	0.02	1	1.0	0.02
2	0.0	0.00	2	0.0	0.00
3	0.0	0.00	3	0.0	0.00
4	0.0	0.00	4	0.0	0.00

Prediksi Hujan Petir untuk 7 Hari ke Depan per Provinsi:

	provinsi	prediksi_7_hari	predicted_probs
0	Aceh	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
1	Kep. Bangka Belitung	[1, 1, 1, 0, 1, 1, 1]	[0.04, 0.08, 0.05, 0.01, 0.02, 0.03, 0.03]
2	Bali	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
3	Banten	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
4	Bengkulu	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
5	Gorontalo	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
6	DKI Jakarta	[0, 0, 0, 0, 1, 1, 0]	[0.0, 0.0, 0.0, 0.0, 0.02, 0.02, 0.0]
7	Jambi	[1, 1, 1, 0, 0, 0, 0]	[0.02, 0.02, 0.02, 0.01, 0.01, 0.01, 0.01]
8	Jawa Barat	[0, 0, 0, 0, 0, 0, 0]	[0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01]
9	Jawa Tengah	[1, 1, 1, 0, 1, 1, 1]	[0.05, 0.05, 0.04, 0.01, 0.03, 0.04, 0.02]
10	Jawa Timur	[1, 1, 0, 0, 1, 1, 1]	[0.02, 0.03, 0.01, 0.0, 0.03, 0.05, 0.02]
11	DI Yogyakarta	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.0, 0.0, 0.0, 0.01, 0.0]
12	Kalimantan Barat	[1, 1, 1, 1, 1, 1, 1]	[0.03, 0.03, 0.02, 0.02, 0.02, 0.02, 0.02]
13	Kalimantan Selatan	[1, 1, 1, 1, 1, 1, 1]	[0.06, 0.06, 0.11, 0.08, 0.05, 0.06, 0.06]
14	Kalimantan Tengah	[1, 1, 1, 1, 1, 1, 1]	[0.04, 0.03, 0.04, 0.04, 0.04, 0.04, 0.04]
15	Kalimantan Timur	[1, 1, 1, 0, 1, 1, 1]	[0.03, 0.04, 0.02, 0.01, 0.03, 0.04, 0.02]
16	Kalimantan Utara	[1, 1, 1, 0, 1, 1, 1]	[0.03, 0.03, 0.02, 0.01, 0.03, 0.03, 0.02]
17	Kep. Riau	[1, 1, 1, 1, 1, 1, 1]	[0.03, 0.04, 0.03, 0.03, 0.04, 0.03, 0.11]
18	Lampung	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
19	Maluku	[1, 0, 0, 0, 0, 0, 0]	[0.02, 0.01, 0.02, 0.01, 0.01, 0.01, 0.01]
20	Maluku Utara	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
21	Nusa Tenggara Barat	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01]
22	Nusa Tenggara Timur	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01]
23	Papua	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.0, 0.0, 0.01, 0.01, 0.0]
24	Papua Barat	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
25	Riau	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
26	Sulawesi Barat	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.01, 0.01, 0.0, 0.0]
27	Sulawesi Selatan	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
28	Sulawesi Tengah	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
29	Sulawesi Tenggara	[0, 0, 0, 0, 0, 1, 0]	[0.0, 0.01, 0.0, 0.0, 0.01, 0.02, 0.01]
30	Sulawesi Utara	[1, 1, 1, 0, 0, 0, 1]	[0.04, 0.07, 0.04, 0.02, 0.01, 0.01, 0.02]
31	Sumatera Barat	[1, 1, 1, 0, 0, 1, 1]	[0.02, 0.02, 0.02, 0.01, 0.02, 0.02, 0.02]
32	Sumatera Selatan	[1, 1, 1, 0, 1, 1, 1]	[0.02, 0.02, 0.02, 0.01, 0.03, 0.03, 0.03]
33	Sumatera Utara	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]

**Gambar 2** Hasil Prediksi 7 Hari ke Depan dengan Model Logit

Prediksi Hujan Petir untuk 7 Hari ke Depan per Provinsi:

	provinsi	prediksi_7_hari	predicted_probs
0	Aceh	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
1	Kep. Bangka Belitung	[1, 1, 1, 0, 1, 1, 1]	[0.04, 0.07, 0.05, 0.01, 0.02, 0.03, 0.03]
2	Bali	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
3	Banten	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
4	Bengkulu	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
5	Gorontalo	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
6	DKI Jakarta	[0, 0, 0, 0, 1, 1, 0]	[0.0, 0.0, 0.0, 0.0, 0.02, 0.02, 0.0]
7	Jambi	[1, 1, 1, 0, 0, 0, 0]	[0.02, 0.02, 0.02, 0.01, 0.01, 0.01, 0.01]
8	Jawa Barat	[0, 0, 0, 0, 0, 0, 0]	[0.02, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01]
9	Jawa Tengah	[1, 1, 1, 0, 1, 1, 1]	[0.04, 0.05, 0.04, 0.01, 0.03, 0.04, 0.02]
10	Jawa Timur	[1, 1, 0, 0, 1, 1, 1]	[0.02, 0.03, 0.01, 0.0, 0.03, 0.05, 0.02]
11	DI Yogyakarta	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.0, 0.0, 0.0, 0.01, 0.0]
12	Kalimantan Barat	[1, 1, 1, 1, 1, 1, 1]	[0.03, 0.03, 0.02, 0.02, 0.02, 0.02, 0.02]
13	Kalimantan Selatan	[1, 1, 1, 1, 1, 1, 1]	[0.06, 0.06, 0.11, 0.09, 0.05, 0.06, 0.06]
14	Kalimantan Tengah	[1, 1, 1, 1, 1, 1, 1]	[0.04, 0.03, 0.04, 0.04, 0.04, 0.04, 0.04]
15	Kalimantan Timur	[1, 1, 1, 0, 1, 1, 1]	[0.03, 0.04, 0.02, 0.01, 0.04, 0.04, 0.02]
16	Kalimantan Utara	[1, 1, 1, 0, 1, 1, 1]	[0.03, 0.03, 0.02, 0.01, 0.03, 0.03, 0.03]
17	Kep. Riau	[1, 1, 1, 1, 1, 1, 1]	[0.03, 0.04, 0.03, 0.03, 0.04, 0.03, 0.09]
18	Lampung	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
19	Maluku	[1, 0, 0, 0, 0, 0, 0]	[0.02, 0.01, 0.02, 0.01, 0.01, 0.01, 0.01]
20	Maluku Utara	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
21	Nusa Tenggara Barat	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01]
22	Nusa Tenggara Timur	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.0, 0.01, 0.01, 0.01, 0.01, 0.01]
23	Papua	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.01, 0.0, 0.0, 0.01, 0.01, 0.0]
24	Papua Barat	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
25	Riau	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
26	Sulawesi Barat	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.01, 0.01, 0.0, 0.0]
27	Sulawesi Selatan	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
28	Sulawesi Tengah	[0, 0, 0, 0, 0, 0, 0]	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]
29	Sulawesi Tenggara	[0, 0, 0, 0, 0, 1, 0]	[0.01, 0.01, 0.0, 0.0, 0.01, 0.02, 0.01]
30	Sulawesi Utara	[1, 1, 1, 0, 0, 0, 1]	[0.04, 0.07, 0.04, 0.02, 0.01, 0.01, 0.02]
31	Sumatera Barat	[1, 1, 1, 0, 0, 1, 1]	[0.02, 0.02, 0.02, 0.01, 0.02, 0.02, 0.02]
32	Sumatera Selatan	[1, 1, 1, 0, 1, 1, 1]	[0.02, 0.02, 0.02, 0.01, 0.03, 0.03, 0.03]
33	Sumatera Utara	[0, 0, 0, 0, 0, 0, 0]	[0.01, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]

Gambar 3 Hasil Prediksi 7 Hari ke Depan dengan Model Probit

### 3.2. Prediksi

Hasil prediksi cuaca per klaster hari ke-8 disajikan pada table 8, diikuti dengan analisis prediksi cuaca selama tujuh hari ke depan dengan model logit pada Gambar 2 dan model probit pada Gambar 3. Setiap prediksi dianalisis berdasarkan parameter-parameter cuaca utama seperti suhu, kelembapan, dan kecepatan angin.

Tabel 8 menunjukkan hasil prediksi kejadian hujan petir menggunakan model logit dan probit. Angka 1.0 pada kolom Prediksi hari ke-8 berarti kejadian hujan petir, sementara angka 0.0 berarti tidak ada hujan petir. Hasil menunjukkan bahwa klaster 0 pada hari ke-8 diprediksi akan mengalami hujan petir dengan probabilitas sebesar 0.03. Sementara itu, Klaster 1 juga pada

hari ke-8 diprediksi akan mengalami hujan petir dengan probabilitas sebesar 0.02. Untuk Klaster 2, 3, dan 4, diprediksi tidak akan terjadi hujan petir.

Hasil prediksi cuaca selama 7 hari ke depan menggunakan model probit dan logit menunjukkan bahwa sebanyak 14 provinsi berpotensi mengalami hujan petir. Probabilitas terjadinya hujan petir di provinsi-provinsi tersebut berkisar antara 0.01 hingga 0.11, dengan probabilitas tertinggi sebesar 0.11. Analisis ini memberikan gambaran yang jelas mengenai potensi cuaca ekstrem di berbagai wilayah, yang dapat digunakan untuk perencanaan dan pengambilan keputusan yang lebih baik.

**Tabel 9** Evaluasi kejadian hujan petir periode 22/12/2024 s.d. 28 Desember 2024

No	Provinsi/tanggal	22	23	24	25	26	27	28
0	Aceh	0	0	1	1	0	0	1
1	Kep. Bangka Belitung	1	1	1	0	1	1	1
2	Bali	1	1	0	0	1	1	1
3	Banten	1	1	1	1	1	1	1
4	Bengkulu	1	1	1	1	0	1	1
5	Gorontalo	0	0	1	0	0	0	0
6	DKI Jakarta	1	1	0	0	1	1	0
7	Jambi	1	1	1	0	1	1	0
8	Jawa Barat	1	1	1	1	1	1	1
9	Jawa Tengah	1	1	1	1	1	1	1
10	Jawa Timur	1	1	1	1	1	1	1
11	DI Yogyakarta	1	1	1	1	0	1	1
12	Kalimantan Barat	1	1	1	1	1	1	1
13	Kalimantan Selatan	1	1	1	1	1	1	1
14	Kalimantan Tengah	1	1	1	1	1	1	1
15	Kalimantan Timur	1	1	1	0	1	1	1
16	Kalimantan Utara	1	1	1	1	0	1	1
17	Kep. Riau	1	1	1	1	1	1	1
18	Lampung	1	1	1	1	1	1	1
19	Maluku	0	0	1	1	0	0	1
20	Maluku Utara	1	1	1	0	0	0	1
21	Nusa Tenggara Barat	1	1	1	1	1	1	1
22	Nusa Tenggara Timur	1	1	1	1	1	1	1
23	Papua	1	1	1	1	1	1	1
24	Papua Barat	1	1	1	1	0	1	1
25	Riau	1	1	0	0	1	1	1
26	Sulawesi Barat	0	1	1	1	0	1	0
27	Sulawesi Selatan	1	1	1	1	1	1	1
28	Sulawesi Tengah	1	1	1	1	0	1	1
29	Sulawesi Tenggara	1	1	1	0	0	1	1
30	Sulawesi Utara	1	1	1	1	0	1	1
31	Sumatera Barat	1	1	1	1	1	1	1
32	Sumatera Selatan	1	1	1	1	1	1	1
33	Sumatera Utara	1	1	1	1	1	0	1

Hasil prediksi di atas dievaluasi dengan membandingkannya dengan berita cuaca yang dimuat di [meteo.bmkg.go.id](http://meteo.bmkg.go.id) dan berdasarkan laporan cuaca historis dan catatan peristiwa dari berbagai sumber berita (seperti BMKG, Antara, dan media lokal) yang melaporkan adanya aktivitas petir/kilat selama hujan di wilayah

provinsi tersebut untuk tanggal yang bersangkutan. Dengan data dari periode 11/12/2024 s.d. 21/12/2024, maka periode evaluasi yang digunakan adalah dari 22/12/2024 s.d. 28 Desember 2024. **Tabel 9** berisi angka 1 apabila di tanggal bersangkutan terjadi hujan petir, dan isi angka 0 apabila tidak ada hujan petir.

**Tabel 10** Metrik Evaluasi (Rata-rata 34 Provinsi)

Metrik	Nilai (Estimasi)	Interpretasi
Akurasi ( <i>Accuracy</i> )	52 %	Model menebak dengan benar sekitar separuh dari total kejadian.
Presisi ( <i>Precision</i> )	88 %	Saat model memprediksi hujan petir, kemungkinan besarnya memang terjadi (sangat kuat di Kalimantan/Sumatera).
<i>Recall</i> ( <i>Sensitivity</i> )	41 %	Model banyak melewati kejadian hujan petir yang sebenarnya terjadi (terutama di Pulau Jawa).
<i>F1-Score</i>	56 %	Keseimbangan antara Presisi dan <i>Recall</i> berada di tingkat moderat.

Selanjutnya, evaluasi akurasi prediksi vs kejadian aktual secara ringkas ditampilkan pada **Tabel 10** dan menunjukkan beberapa hal menarik. Pertama, Model sangat baik dalam mengenali pola di wilayah yang memiliki "persistensi" tinggi seperti Kalimantan dan Kepulauan Bangka Belitung. Model secara konsisten memberikan angka 1 dan secara aktual memang terjadi hujan petir terus-menerus. Kedua, model gagal menangkap fenomena hujan petir di Pulau Jawa (Jawa Barat, Banten, DKI, DIY). Pada periode tersebut, wilayah ini sedang masuk puncak musim hujan (Monsun Asia), namun model justru banyak memprediksi angka 0 (tidak ada hujan petir).

Secara keseluruhan, *F1-Score* model berada di angka 56%. Ini menunjukkan model sudah cukup baik dalam memprediksi wilayah tertentu (seperti Kalimantan), namun masih sering meleset di wilayah lain (seperti Jawa), sehingga keseimbangannya belum maksimal. Ke depan, penelitian lebih lanjut masih diperlukan untuk meningkatkan akurasi prediksi, misalnya dengan menambahkan variabel *dummy* musim atau variabel *lagging*.

#### 4. KESIMPULAN

Hasil klusterisasi menggunakan metode K-Means menunjukkan bahwa provinsi-provinsi di Indonesia dapat dikelompokkan ke dalam lima klaster berdasarkan karakteristik cuaca yang mirip. Setiap klaster memiliki karakteristik unik yang tercermin dari rata-rata, minimum, dan maksimum nilai kelembapan, suhu, kecepatan angin, serta jumlah kejadian hujan petir. Klusterisasi ini membantu dalam memahami pola cuaca yang berbeda di berbagai wilayah Indonesia.

Berdasarkan data historis, hasil klusterisasi, dan model prediksi yang digunakan, kami berhasil mengidentifikasi pola cuaca di berbagai wilayah Indonesia. Prediksi cuaca untuk hari ke-8 menunjukkan bahwa beberapa klaster berpotensi mengalami hujan petir dengan probabilitas yang bervariasi. Selain itu, prediksi cuaca selama tujuh hari ke depan menunjukkan bahwa sebanyak 14 provinsi berpotensi mengalami hujan petir, dengan probabilitas berkisar antara 0.01 hingga 0.11. Hasil analisis ini memberikan gambaran yang jelas mengenai kondisi cuaca yang diharapkan dan dapat digunakan sebagai dasar untuk perencanaan dan pengambilan keputusan di berbagai sektor. Dengan memahami pola cuaca ini, diharapkan masyarakat dapat mempersiapkan diri dengan lebih baik dalam menghadapi kondisi cuaca yang akan datang. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam memahami karakteristik cuaca di Indonesia dan menyediakan informasi yang berguna untuk berbagai keperluan praktis. Kami berharap hasil ini dapat menjadi acuan bagi penelitian lebih lanjut dan membantu dalam upaya mitigasi risiko cuaca ekstrem di masa depan.

#### DAFTAR PUSTAKA

- Amin, F. M., Rusydiyah, E. F., & Azizah, A. N. (2025). Personalized Library Book Recommendations Using K-Means Clustering and Association Rules. *Journal of Scientometric Research*, 14(1), 32-45.
- Badan Meteorologi, K. dan G. (2024). Data Prakiraan Cuaca Terbuka BMKG (Kecamatan). Retrieved from <https://data.bmkg.go.id/csv/>

- Cahyanto, R., Chrismanto, A. R., & Sebastian, D. (2020). Pengelompokan Komentar Dataset Sentipol dengan Modified K-Means Clustering. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(3), 531–540.
- Chusyairi, A. (2023). Clustering Data Cuaca Ekstrem Indonesia dengan K-Means dan Entropi. *Journal of Informatics and Communication Technology (JICT)*, 5(1), 1–10.
- Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method, 5–8.
- David, M., Alonso-Montesinos, J., Le Gal La Salle, J., & Lauret, P. (2023). Probabilistic Solar Forecasts as a Binary Event Using a Sky Camera. *Energies*, 16(20), 1–18.
- Dinar Ajeng, K., Irwansyah, S., & Rina. (2021). Rain Prediction Clustering in Australia Using the K-Means Algorithm in the WEKA and RStudio Application. *Portal Journal UPN Yogyakarta*, (November), 13–2021.
- Fitria, N. I. (2024). Prospek Cuaca Mingguan Periode 03–09 Desember 2024: Siaga Puncak Musim Hujan: Potensi Cuaca Ekstrem Melanda Sejumlah Wilayah Indonesia. *www.bmkg.go.id*. Retrieved January 12, 2025, from <https://www.bmkg.go.id/cuaca/prospek-cuaca-mingguan/prospek-cuaca-mingguan-periode-03-09-desember-2024-siaga-puncak-musim-hujan-potensi-cuaca-ekstrem-melanda-sejumlah-wilayah-indonesia>
- Kalude, U. F., Titaley, J., & Kalua, A. L. (2022). Analisis Clustering Perubahan Cuaca dan Iklim di Kota Manado Menggunakan Metode K-Means, 01(02), 1–12.
- Khairunnisa, S., & Jambak, M. I. (2022). Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca. *Jurnal Media Informatika Budidarma*, 6(4), 2352.
- Learn, S. (2025). StandardScaler. Retrieved December 15, 2025, from <https://scikit-learn.org/1.5/modules/generated/sklearn.preprocessing.StandardScaler.html>
- Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8(1). Springer International Publishing. Retrieved from <https://doi.org/10.1186/s40537-021-00545-4>
- Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10–17.
- Neumann, O., Turowski, M., Mikut, R., Hagenmeyer, V., & Ludwig, N. (2023). Using weather data in energy time series forecasting: the benefit of input data transformations. *Energy Informatics*, 6(1). Springer International Publishing. Retrieved from <https://doi.org/10.1186/s42162-023-00299-8>
- Pan, Q., Porth, L., & Li, H. (2022). Assessing the Effectiveness of the Actuaries Climate Index for Estimating the Impact of ExtremeWeather on Crop Yield and Insurance Applications. *Sustainability (Switzerland)*, 14(11).
- Putratama, R. (2024). BMKG: Cuaca Ekstrem Masih Berpotensi Hingga 8 Maret 2024. Retrieved January 12, 2025, from <https://www.bmkg.go.id/siaran-pers/bmkg-cuaca-ekstrem-masih-berpotensi-hingga-8-maret-2024>
- Raghavendran, R. V. (2019). K- Means Clustering and Naive Bayes Algorithms, (June).
- Ramadhani, A., Putri, D. R., & Raihani, R. (2022). Perbandingan Keakuratan Model Logit Dan Probit Dalam Mengidentifikasi Faktor-Faktor Yang Memengaruhi Persentase Penduduk Miskin di Kalimantan. *Prosiding Seminar Nasional Matematika dan Statistika*, 387–407. Retrieved from <http://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/850>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181(2019), 526–534. Elsevier B.V. Retrieved from

<https://doi.org/10.1016/j.procs.2021.01.199>

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Eurasip Journal on Wireless Communications and Networking*, 2021(1). Springer International Publishing. Retrieved from <https://doi.org/10.1186/s13638-021-01910-w>

Yani, V. I., Aradea, A., & Mubarok, H. (2022). Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(3), 607–619.

Zhan, D., Qin, S., (Leon) Wang, L., & Hassan, I. G. (2025). Weather clustering for machine learning-based hourly building energy prediction models at design phase. *Energy and Buildings*, 329(January), 115308. Elsevier B.V. Retrieved from <https://doi.org/10.1016/j.enbuild.2025.115308>