

***FORECASTING* CUACA EKSTREM MENGGUNAKAN RANDOM FOREST
BERBASIS DATA PETIR DENGAN FITUR SPASIAL**

SKRIPSI

Oleh :
INTAN NURHIDAYAH
NIM. 220605110039



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

***FORECASTING* CUACA EKSTREM MENGGUNAKAN RANDOM
FOREST BERBASIS DATA PETIR DENGAN FITUR SPASIAL**

SKRIPSI

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

Oleh:
INTAN NURHIDAYAH
NIM. 220605110039

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

HALAMAN PERSETUJUAN

**FORECASTING CUACA EKTREM MENGGUNAKAN RANDOM
FOREST BERBASIS DATA PETIR DENGAN FITUR SPASIAL**

SKRIPSI

Oleh :
INTAN NURHIDAYAH
NIM. 220605110039

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 12 Juni 2026

Pembimbing I,



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

Pembimbing II,



A'la Syaqui, M.Kom
NIP. 19771201 200801 1 007

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

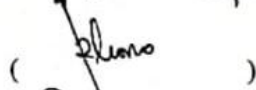
FORECASTING CUACA EKTREM MENGGUNAKAN RANDOM FOREST BERBASIS DATA PETIR DENGAN FITUR SPASIAL

SKRIPSI

Oleh :
INTAN NURHIDAYAH
NIM. 220605110039

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Supriyono, M.Kom</u> NIP. 19770103 201101 1 004	()
Anggota Penguji I	: <u>Roro Inda Melani, M.T., M.Sc</u> NIP. 19780925 200501 2 008	()
Anggota Penguji II	: <u>Dr. Agung Teguh Wibowo Almais, M.T</u> NIP. 19860301 202321 1 016	()
Anggota Penguji III	: <u>A'la Syaqui, M.Kom</u> NIP. 19771201 200801 1 007	()

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 : Intan Nurhidayah

NIM : 220605110039

Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : *Forecasting* Cuaca Ekstrem menggunakan
Random Forest Berbasis Data Petir Dengan Fitur
Spasial

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, 29 Juni 2026

Yang membuat pernyataan,



Intan Nurhidayah

NIM. 220605110039

MOTTO

“Effort Never Betrays Those Who Persevere”

HALAMAN PERSEMBAHAN

Puji syukur kehadiran Allah Subhanahu wa Ta'ala atas segala rahmat, nikmat, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Shalawat serta salam senantiasa tercurah kepada junjungan Nabi Muhammad Shallallahu 'alaihi wasallam, yang telah membawa umat manusia dari zaman kegelapan menuju zaman yang penuh dengan ilmu pengetahuan dan cahaya Islam. Dengan penuh rasa syukur, karya sederhana ini penulis persembahkan kepada mama tercinta yang senantiasa memberikan kasih sayang, doa, semangat, serta pengorbanan tanpa batas dalam setiap langkah kehidupan penulis. Terima kasih atas cinta dan ketulusan yang tidak pernah berhenti menguatkan penulis hingga mampu menyelesaikan pendidikan ini. Karya ini juga penulis persembahkan kepada bapak tercinta, yang telah bekerja keras, memberikan dukungan, nasihat, serta menjadi teladan dalam menjalani kehidupan. Terima kasih atas setiap doa, pengorbanan, dan kepercayaan yang selalu diberikan kepada penulis. Teruntuk abang dan kakak tercinta, terima kasih atas doa, dukungan, perhatian, serta semangat yang selalu diberikan. Kehadiran kalian menjadi sumber kebahagiaan dan motivasi bagi penulis untuk terus berjuang menyelesaikan setiap proses yang telah dilalui. Persembahan ini juga penulis tujukan kepada diri sendiri sebagai bentuk penghargaan atas usaha, kesabaran, dan perjuangan hingga skripsi ini dapat diselesaikan. Terima kasih kepada teman-teman seperjuangan atas dukungan, kebersamaan, dan motivasi selama masa perkuliahan. Semoga skripsi ini dapat memberikan manfaat serta menjadi langkah awal menuju masa depan yang lebih baik. Aamiin

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Subhānahu wa Ta‘ālā atas segala limpahan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“FORECASTING CUACA EKSTREM MENGGUNAKAN RANDOM FOREST BERBASIS DATA PETIR DENGAN FITUR SPASIAL”** sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad ﷺ ‘Alaihi Wasallam, beserta keluarga, sahabat, dan seluruh umatnya yang senantiasa istiqamah mengikuti ajaran beliau hingga akhir zaman.

Penyusunan skripsi ini dapat terselesaikan berkat bimbingan, bantuan, doa, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP, 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 dengan sabar, telaten, dan penuh dedikasi telah memberikan bimbingan,

arahan, serta masukan yang sangat berharga sehingga penulis dapat menyelesaikan skripsi ini.

5. A'la Syauqi, M.Kom, selaku Dosen Pembimbing II, yang telah meluangkan waktu dan pikiran untuk membimbing, mengarahkan, serta memberikan kritik dan saran yang konstruktif selama proses pengerjaan skripsi.
6. Supriyono, M.Kom dan Roro Inda Melani, M.T., M.Sc, selaku penguji yang telah memberikan masukan saran sehingga skripsi ini dapat disempurnakan.
7. Kedua orang tua tercinta, Mama dan Bapak, yang senantiasa menjadi sumber kekuatan, doa, kasih sayang, serta pengorbanan tanpa batas bagi penulis. Terima kasih atas setiap dukungan, kepercayaan, dan perjuangan yang telah diberikan sehingga penulis dapat menyelesaikan pendidikan dan skripsi ini. Semoga segala kebaikan, cinta, dan pengorbanan Mama dan Bapak senantiasa mendapat balasan terbaik dari Allah Swt. Terima kasih juga kepada Abang dan Kakak yang selalu memberikan semangat, motivasi, serta doa kepada penulis dalam setiap proses penyusunan skripsi ini.
8. Segenap dosen, staf administrasi, laboran, dan seluruh civitas akademika Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang telah memberikan ilmu, bimbingan, bantuan, serta dukungan selama penulis menempuh pendidikan hingga penyusunan skripsi ini.

9. Teman-teman PKL tercinta, Dila, Fina, Faiza, Renata, dan Arum, yang senantiasa menemani, memberikan semangat, dukungan, serta bantuan kepada penulis selama proses penyusunan skripsi. Kehadiran dan kepedulian kalian menjadi kekuatan bagi penulis untuk terus berjuang hingga skripsi ini dapat terselesaikan. Khususnya Dila, terima kasih telah menjadi teman terdekat yang selalu membantu, menemani, mendengarkan keluh kesah, memberikan semangat, serta selalu hadir di setiap proses penyusunan skripsi ini.
10. Teman-teman seperjuangan tercinta, Suci, Lay, Humay, dan Izzah, yang senantiasa memberikan dukungan, bantuan, semangat, serta motivasi kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas kebersamaan, perhatian, kepedulian, serta berbagai pengalaman yang telah dilalui bersama, baik dalam suka maupun duka. Secara khusus, penulis menyampaikan rasa terima kasih kepada Suci atas kebersamaan, dukungan, dan ketulusannya dalam menemani penulis selama proses penyusunan skripsi. Terima kasih karena selalu bersedia mendengarkan setiap keluh kesah, memberikan motivasi, serta menjadi tempat berbagi cerita dan semangat ketika penulis menghadapi berbagai tantangan hingga skripsi ini dapat diselesaikan.
11. Teman-teman orda yang telah menjadi keluarga kedua di perantauan. Terima kasih atas kebersamaan, perhatian, dan dukungan yang diberikan, sehingga penulis dapat melalui berbagai tantangan selama menempuh pendidikan jauh dari keluarga.

12. Teman-teman KKM SatyaWidya, yaitu Zubda, Farin, Nana, Alya, Keysa, Hilza, Arifah, dan Faiza, yang telah memberikan semangat, doa, dan dukungan selama proses penyusunan skripsi. Terima kasih atas kebersamaan, perhatian, serta kehadiran kalian yang selalu menjadi penyemangat bagi penulis hingga skripsi ini dapat terselesaikan.
13. Seluruh teman-teman Teknik Informatika INFINITY Angkatan 2022 yang telah memberikan dukungan dan kebersamaan selama masa perkuliahan.
14. Untuk diri sendiri, terima kasih telah bertahan dan terus berjuang hingga sejauh ini. Terima kasih atas setiap air mata, rasa lelah, dan perjuangan yang mungkin tidak selalu terlihat oleh orang lain. Terima kasih karena tidak memilih menyerah dan terus melangkah hingga skripsi ini dapat terselesaikan. Maaf karena perjalanan ini begitu panjang dan sering kali terasa berat. Maaf telah memaksamu bertahan di saat lelah, menangis di saat sendiri, dan tetap melangkah meski keadaan tidak selalu mudah. Semoga ke depannya penulis senantiasa diberikan kekuatan, keberanian, dan keteguhan hati dalam menghadapi setiap tantangan serta meraih impian yang lebih besar.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Malang, 29 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN TULISAN	iii
MOTTO	vi
HALAMAN PERSEMBAHAN	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	7
2.1 Penelitian Terkait	7
2.2 Cuaca Ekstrem	11
2.3 <i>Forecasting</i>	14
2.4 Algoritma <i>Random Forest</i>	15
2.5 Data Petir	17
BAB III METODE PENELITIAN	21
3.1 Desain Sistem	21
3.2 Pengumpulan Data	21
3.3 <i>Preprocessing</i>	23
3.3.1 <i>Data Cleaning</i>	24
3.4 <i>Split Data</i>	25
3.5 <i>Random Forest</i>	25
3.5.1 Penentuan <i>Hyperparameter Random Forest</i>	27
3.6 Evaluasi Kinerja	27
3.7 Skenario Uji Coba	29
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Hasil Uji Coba	31
4.1.1 Skenario 1A dengan Rasio 90:10	31
4.1.2 Skenario 2A dengan Rasio 80:20	34
4.1.3 Skenario 3A dengan Rasio 70:30	37
4.1.4 Skenario 4B dengan Rasio 90:10	40
4.1.5 Skenario 5B dengan Rasio 80:20	43
4.1.6 Skenario 6B dengan Rasio 70:30	46
4.2 Pembahasan Analisis Skenario	50

4.2.1 Analisis Skenario A (semua fitur).....	50
4.2.2 Analisis Skenario B (Tanpa Fitur Spasial).....	55
4.3 Perbandingan Performa Model Berdasarkan Matrik Evaluasi.....	59
4.3.1 <i>Accuracy</i>	59
4.3.2 <i>Precision</i>	61
4.3.3 <i>Recall</i>	63
4.3.4 <i>F1-Score</i>	65
4.4 Pembahasan Hasil Keseluruhan	67
4.5 Integrasi Islam.....	69
BAB V KESIMPULAN DAN SARAN	72
5.1 Kesimpulan	72
5.2 Saran.....	72
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2.1 Alur Kerja <i>Random Forest</i>	16
Gambar 3.1 Desain Sistem	21
Gambar 4.1 <i>Confusion Matrix</i> Skenario 1A Rasio 90:10	32
Gambar 4.2 <i>Confusion Matrix</i> Skenario 2A Rasio 80:20	35
Gambar 4.3 <i>Confusion Matrix</i> Skenario 3A Rasio 70:30	38
Gambar 4.4 <i>Confusion Matrix</i> Skenario 4B Rasio 90:10	41
Gambar 4.5 <i>Confusion Matrix</i> Skenario 5B Rasio 80:20	44
Gambar 4.6 <i>Confusion Matrix</i> Skenario 6B Rasio 70:30	47
Gambar 4.7 Metrik Evaluasi skenario A (semua fitur)	53
Gambar 4.8 Metrik Evaluasi Skenario B (Tanpa Fitur Spasial)	58
Gambar 4.9 Grafik Perbandingan <i>Accuracy</i> pada tiap Skenario	61
Gambar 4.10 Grafik Perbandingan <i>Precision</i> pada tiap Skenario	63
Gambar 4.11 Grafik Perbandingan <i>Recall</i> pada tiap Skenario	65
Gambar 4.12 Grafik Perbandingan <i>F1-Score</i> pada tiap Skenario	67

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 3.1 Data Petir.....	22
Tabel 3.2 Nilai <i>Hyperparameter</i>	27
Tabel 3.3 Skenario Uji Coba.....	30
Tabel 4.1 Hasil Evaluasi Pengujian Skenario 1A Rasio 90:10	33
Tabel 4.2 Hasil Evaluasi Pengujian Skenario 2A Rasio 80:20	36
Tabel 4.3 Hasil Evaluasi Pengujian Skenario 3A Rasio 70:30	39
Tabel 4.4 Hasil Evaluasi Pengujian Skenario 4B Rasio 90:10	42
Tabel 4.5 Hasil Evaluasi Pengujian Skenario 5B Rasio 80:20	45
Tabel 4.6 Hasil Evaluasi Pengujian Skenario 6B Rasio 70:30	48
Tabel 4.7 Analisis skenario A (semua fitur)	50
Tabel 4.8 Analisis skenario B (tanpa fitur spasial)	55

ABSTRAK

Nurhidayah, Intan. 2026. ***Forecasting Cuaca Ekstrem Menggunakan Random Forest Berbasis Data Petir Dengan Fitur Spasial***. 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) A'la Syauqi, M.Kom.

Kata kunci: *Forecasting Cuaca Ekstrem, Random Forest, Aktivitas Petir, Fitur Spasial.*

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Random Forest* dalam *forecasting* cuaca ekstrem berbasis data aktivitas petir menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* dengan dataset aktivitas petir BMKG Stasiun Geofisika Kelas III Malang sebanyak 2.208 data. Metode yang digunakan adalah algoritma *Random Forest* dengan memanfaatkan karakteristik aktivitas petir *Cloud to Ground Positive (CG+)*, *Cloud to Ground Negative (CG-)*, *Intra Cloud (IC)*, serta fitur spasial berupa *latitude* dan *longitude*. Pengujian dilakukan pada enam skenario menggunakan variasi fitur dan rasio pembagian data latih dan data uji (90:10, 80:20, dan 70:30), serta *hyperparameter tuning* yang dievaluasi menggunakan *confusion matrix*. Hasil penelitian menunjukkan bahwa penambahan fitur spasial meningkatkan performa model dibandingkan tanpa fitur spasial. Model terbaik diperoleh pada Skenario 1A dengan rasio 90:10 yang menghasilkan *accuracy* 98.64%, *precision* 88.46%, *recall* 100.00%, dan *F1-score* 93.87%. Hasil tersebut menunjukkan bahwa algoritma *Random Forest* berbasis data petir dan fitur spasial mampu meningkatkan kemampuan dalam *forecasting* cuaca ekstrem.

ABSTRACT

Nurhidayah, Intan. 2026. ***Forecasting Extreme Weather Using a Lightning-Data-Based Random Forest With Spatial Features***. 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) A'la Syauqi, M.Kom.

This study aims to evaluate the performance of the *Random Forest* algorithm in *forecasting* extreme weather based on lightning activity data using the metrics *accuracy*, *precision*, *recall*, and *F1-score* with a dataset of 2,208 lightning activity records from the BMKG Class III Geophysical Station in Malang. The method used was the *Random Forest* algorithm, utilizing the characteristics of *Cloud-to-Ground Positive (CG+)*, *Cloud-to-Ground Negative (CG-)*, and *Intra-Cloud (IC)* lightning activity, as well as spatial features such as *latitude* and *longitude*. Testing was conducted across six scenarios using variations in features and training-to-test data split ratios (90:10, 80:20, and 70:30), along with *hyperparameter tuning* evaluated using a *confusion matrix*. The results indicate that the inclusion of spatial features improves model performance compared to models without spatial features. The best model was obtained in Scenario 1A with a 90:10 ratio, yielding an *accuracy* of 98.64%, *precision* of 88.46%, *recall* of 100.00%, and an *F1-score* of 93.87%. These results demonstrate that the *Random Forest* algorithm based on lightning data and spatial features can improve the ability to *forecast* extreme weather.

Keywords: Extreme Weather *Forecasting*, Random Forest, Lightning Activity, Spatial Features

الملخص

نورهيدايا، إيتان. ٢٠٢٦. التنبؤ بالظواهر الجوية المتطرفة باستخدام غابات عشوائية قائمة على بيانات الصواعق ذات السمات المكانية. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية «مولانا مالك إبراهيم» في مالانغ المشرف الرئيسي: د. أغونغ تيغوه ويوو أليس، ماجستير في الهندسة. المشرف الثاني: علاء شوقي، ماجستير في علوم الحاسوب

الكلمات المفتاحية: التنبؤ بالظواهر الجوية المتطرفة، غابة عشوائية، نشاط البرق، السمات المكانية.

يهدف هذا البحث إلى تقييم أداء خوارزمية *Random Forest* في التنبؤ بالظواهر الجوية المتطرفة استنادًا إلى بيانات نشاط البرق، وذلك باستخدام مقاييس الدقة (*accuracy*) والدقة النوعية (*precision*) ومعدل الاسترجاع (*recall*) ومؤشر *F1-score*، مع استخدام مجموعة بيانات نشاط البرق الصادرة عن محطة الجيوفيزياء من الدرجة الثالثة التابعة لـ BMKG في مالانج، والتي تضم 2,208 قيمة. الطريقة المستخدمة هي خوارزمية *Random Forest* مع الاستفادة من خصائص نشاط البرق من نوع «السحابة إلى الأرض الموجبة (CG+)»، و«السحابة إلى الأرض السالبة (CG-)»، وداخل السحابة (IC)، بالإضافة إلى السمات المكانية المتمثلة في خطوط العرض وخطوط الطول. أجريت الاختبارات على ستة سيناريوهات باستخدام تباينات في السمات ونسب توزيع البيانات التدريبية وبيانات الاختبار 90:10، و80:20، و70:30، بالإضافة إلى ضبط الملمات الفائقة التي تم تقييمها باستخدام مصفوفة الارتباك. أظهرت نتائج البحث أن إضافة السمات المكانية تحسن أداء النموذج مقارنةً بعدم استخدام السمات المكانية. تم الحصول على أفضل نموذج في السيناريو 1A بنسبة 90:10، مما أدى إلى دقة بنسبة 98,64%، ودقة تحديد بنسبة 88,46%، ومعدل استرجاع بنسبة 100,00%، ودرجة F1 بنسبة 93,87%. وتشير هذه النتائج إلى أن خوارزمية *Random Forest* القائمة على بيانات الصواعق والسمات المكانية قادرة على تحسين القدرة على التنبؤ بالظواهر الجوية المتطرفة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia sebagai negara tropis yang beriklim basah dengan curah hujan tinggi serta karakter geografis yang kompleks memiliki tingkat kerentanan yang tinggi terhadap cuaca ekstrem (Tumangkeng *et al.*, 2021). Cuaca ekstrem mengacu pada kondisi cuaca dan iklim yang menunjukkan penyimpangan dari keadaan normal akibat terjadinya fenomena alam yang tidak lazim (Caraka *et al.*, 2015). Kondisi tersebut menyebabkan wilayah Indonesia rentan terhadap bencana ekstrem seperti hujan lebat, maupun petir yang kerap terjadi secara musiman (Mananohas *et al.*, 2019). Dampak yang ditimbulkan meliputi kerusakan infrastruktur, kerugian ekonomi, hingga ancaman terhadap keselamatan masyarakat (Adrianti, 2024). Oleh karena itu, diperlukan sistem *forecasting* cuaca ekstrem yang mampu memberikan informasi lebih dini sehingga dapat mendukung upaya mitigasi dan pengambilan keputusan sebelum kejadian cuaca ekstrem terjadi (Atika *et al.*, 2024).

Salah satu fenomena atmosfer yang berkaitan erat dengan kejadian cuaca ekstrem adalah aktivitas petir. Petir terbentuk akibat proses konveksi kuat pada awan cumulonimbus ketika atmosfer berada dalam kondisi labil dan memiliki kandungan uap air yang tinggi (Angga & Diansari, 2018). Aktivitas petir merupakan proses pelepasan muatan listrik yang berkaitan dengan pembentukan awan konvektif penyebab hujan lebat maupun badai petir (Paski *et al.*, 2017). Data aktivitas petir yang direkam oleh sensor berbasis darat (*ground-based*) maupun

satelit dapat dimanfaatkan untuk mengetahui jumlah sambaran petir serta karakteristik distribusinya (Mehranzmir *et al.*, 2023). Karakteristik tersebut menjadikan data petir berpotensi dimanfaatkan sebagai sumber informasi dalam membangun model *forecasting* cuaca ekstrem. Oleh karena itu, pendekatan yang banyak digunakan adalah *machine learning* karena mampu mempelajari pola dari beberapa data pada masa lampau untuk menghasilkan model *forecasting* dengan tingkat akurasi yang baik.

Dalam menghadapi fenomena alam seperti petir dan cuaca ekstrem, penting bagi kita untuk mengingat bahwa Al-Qur'an memberikan isyarat mengenai hal tersebut. Allah SWT berfirman dalam Al-Qur'an surat Ar-Ra'd (13) ayat 12–13:

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

“(12) Dialah yang memperlihatkan kepadamu kilat (untuk menimbulkan) ketakutan dan harapan (akan turun hujan) serta menjadikan awan yang berat (mendung). (13) Guruh bertasbih dengan memuji-Nya, (demikian pula) malaikat karena takut kepada-Nya. Dia (Allah) melepaskan petir, lalu menimpakannya kepada siapa yang Dia kehendaki. Sementara itu, mereka (orang-orang kafir) berbantah-bantahan tentang kekuasaan Allah, padahal Dia Maha Keras hukuman-Nya.” (QS. Ar-Ra'd: 12-13).

Menurut Tafsir Ibnu Katsir, Allah SWT menjelaskan pada ayat ini dengan memperlihatkan kilat kepada manusia sebagai sebab munculnya rasa takut terhadap bahaya petir dan harapan akan turunnya hujan yang membawa keberkahan. Allah SWT mengendalikan petir dan menimpakkannya kepada siapa pun yang dia kehendaki sebagai bentuk rahmat atau hukuman. Meski tanda kekuasaan Allah SWT tampak jelas pada fenomena kilat, guruh, dan awan mendung, manusia tetap saja membantah tentang Allah SWT, Hal ini menegaskan

bahwa segala peristiwa alam terjadi sesuai kehendak Allah SWT (Al-Sheikh, 2004).

Dari pemaparan ayat di atas, salah satu bentuk ikhtiar dalam mewujudkan kesadaran akan tanda-tanda kebesaran Allah SWT adalah dengan memanfaatkan ilmu pengetahuan dan teknologi untuk kebaikan. Dalam penelitian ini, dibangun sistem *forecasting* cuaca ekstrem dengan menerapkan algoritma *Random Forest* yang ditambahkan dengan informasi dari spasial berbasis data petir. Algoritma *Random Forest* menerapkan pemilihan fitur secara acak pada setiap node dalam pembentukan *decision tree*, sehingga mampu meningkatkan kemampuan generalisasi model dan menghasilkan akurasi yang lebih baik (Amaliah *et al.*, 2022). Hasil akhir dipilih dengan melakukan penentuan node dan memilih beberapa node untuk mendapatkan hasil akhirnya (Alita & Isnain, 2020). Penelitian yang dilakukan oleh Liu *et al.* (2022) mengusulkan integrasi fitur spasial berupa *Spatial Lag* dan *Eigenvector Spatial Filtering* (ESF) ke dalam model *machine learning* seperti *Random Forest*. Dari hasil penelitian yang dilakukan bahwa adanya penambahan spasial bisa meningkatkan performa model.

Berbagai penelitian yang menerapkan algoritma *Random Forest* memiliki performa yang baik dalam *forecasting* maupun prediksi cuaca. Dwiyanti dan Prianto (2023) memperoleh nilai akurasi sebesar 71% pada prediksi kondisi cuaca di Jakarta menggunakan *Random Forest*. Syahreza *et al.* (2024) menunjukkan bahwa *Random Forest* memiliki performa akurasi tertinggi sebesar 85.1% dibandingkan *Decision Tree*, *Gradient Boosting*, dan SVM pada prediksi kondisi cuaca. Mukeshkumar *et al.* (2025) juga melaporkan bahwa *Random Forest* mampu

mencapai akurasi sebesar 95% dalam sistem forecasting dan peringatan cuaca ekstrem berbasis data meteorologi historis. Selain itu, Sangaji dan Sutabri (2025) menunjukkan bahwa *Random Forest* memberikan performa yang baik pada forecasting curah hujan dan semakin meningkat ketika dikombinasikan dalam metode ensemble.

Meskipun demikian, sebagian besar penelitian terdahulu masih menggunakan parameter meteorologi maupun klimatologi sebagai dasar dalam membangun model *forecasting*. Penelitian yang secara khusus memanfaatkan data petir sebagai sumber informasi utama dalam *forecasting* cuaca ekstrem masih relatif terbatas. Padahal, aktivitas petir memiliki hubungan yang erat dengan pembentukan awan konvektif dan umumnya muncul sebelum terjadinya hujan lebat maupun badai petir, sehingga berpotensi menjadi indikator penting dalam mengidentifikasi potensi cuaca ekstrem.

Selain itu, penelitian terdahulu umumnya lebih berfokus pada perbandingan algoritma atau peningkatan performa model tanpa mengevaluasi pengaruh fitur spasial terhadap hasil *forecasting*. Padahal, aktivitas petir memiliki karakteristik geografis yang kuat karena distribusinya dipengaruhi oleh lokasi kejadian. Informasi spasial berupa *latitude* dan *longitude* berpotensi membantu model mengenali pola persebaran aktivitas petir pada wilayah tertentu sehingga dapat meningkatkan kemampuan model dalam melakukan *forecasting* cuaca ekstrem.

Berdasarkan kondisi tersebut, penelitian ini membuat model *forecasting* cuaca ekstrem menggunakan algoritma *Random Forest* dengan memanfaatkan

data petir yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang. Selain menggunakan parameter aktivitas petir, penelitian ini juga mengevaluasi pengaruh penambahan fitur spasial berupa *latitude* dan *longitude* melalui perbandingan dua skenario, yaitu model yang memanfaatkan seluruh fitur dan model yang tidak menggunakan fitur spasial. Melalui perbandingan tersebut, diharapkan dapat diperoleh gambaran mengenai kontribusi informasi spasial dalam meningkatkan performa algoritma *Random Forest* untuk forecasting cuaca ekstrem berbasis data petir.

1.2 Pernyataan Masalah

- a. Bagaimana penerapan algoritma *Random Forest* dalam *forecasting* cuaca ekstrem dengan menggunakan data petir?
- b. Bagaimana mengevaluasi kinerja algoritma *Random Forest* dalam *forecasting* cuaca ekstrem dengan memanfaatkan fitur spasial dari data petir?

1.3 Tujuan Penelitian

- a. Menerapkan algoritma *Random Forest* dengan memanfaatkan data petir dalam *forecasting* cuaca ekstrem.
- b. Mengevaluasi kinerja algoritma *Random Forest* dengan memanfaatkan fitur spasial pada data petir dalam *forecasting* cuaca ekstrem.

1.4 Manfaat Penelitian

- a. Memberikan kontribusi pada pengembangan metode *forecasting* cuaca ekstrem berbasis *machine learning*.
- b. Memberikan acuan bagi instansi terkait, seperti BMKG, dalam mendukung sistem peringatan dini bencana.

1.5 Batasan Masalah

- a. Data yang digunakan dalam penelitian ini terbatas pada data petir yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang selama periode Maret–Mei 2025.
- b . Penelitian ini difokuskan pada penerapan algoritma *Random Forest* sebagai metode pemodelan, tanpa melibatkan variabel atmosfer lainnya maupun membandingkannya dengan metode *machine learning* lain.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian terkait yang dirangkum pada Tabel 2.1 menunjukkan bahwa kajian mengenai prediksi maupun prakiraan cuaca telah banyak dilakukan menggunakan berbagai metode *machine learning* maupun metode statistik. Secara umum, penelitian-penelitian tersebut memanfaatkan data meteorologi, serta data cuaca historis sebagai data masukan dalam membangun model prediksi.

Dari sisi metode, sebagian besar penelitian menerapkan algoritma *Random Forest* sebagai model utama dalam melakukan pembandingan terhadap algoritma lain, seperti *Support Vector Machine* (SVM), *Decision Tree*, *Gradient Boosting*, *K-Nearest Neighbor* (KNN), *Naïve Bayes*, *Long Short-Term Memory* (LSTM), *Recurrent Neural Network* (RNN), serta metode statistik seperti ARIMA dan Regresi Nonparametrik Estimator Deret Fourier. Beberapa penelitian juga mengembangkan model melalui pendekatan *ensemble* maupun teknik penyeimbangan data, seperti SMOTE, untuk meningkatkan performa *forecasting*.

Tabel 2.1 Penelitian Terdahulu

No	Referensi	Topik	Objek	Metode	Subyek	Hasil
1	(Dwiyanti & Prianto, 2023)	Prediksi Cuaca	Data cuaca harian dari open data jakarta	<i>Random Forest</i>	Prediksi Kondisi cuaca di jakarta	Akurasi 0.71; presisi 0.71; recal 0.70; F1score 0.70; ROC-AUC 0.92
2	(Jones <i>et al.</i> , 2024)	Prakiraan Cuaca	33.746 data cuaca (Temperatur)	<i>KNN</i> , <i>Logistic Regression</i> , <i>Ensemble</i>	Pendekatan Metode <i>Ensemble</i>	<i>Naïve Bayes</i> memperoleh <i>accuracy</i> 99.00%,

No	Referensi	Topik	Objek	Metode	Subyek	Hasil
				<i>Random Forest, Naïve Bayes, MLP, dan Soft Voting Classifier</i>	<i>e Learning</i> untuk Prakiraan Cuaca menggunakan <i>Soft Voting Classifier</i>	sedangkan <i>Soft Voting Classifier</i> (<i>Random Forest–MLP–Naïve Bayes</i>) menghasilkan <i>accuracy</i> terbaik 99.03%. <i>Random Forest</i> memberikan akurasi tinggi..
3	(Amri <i>et al.</i> , 2024)	Prediksi Cuaca Ekstrem	Data curah hujan dan kecepatan angin (Jepara)	<i>Regresi Nonparametrik Estimasi Deret Fourier</i>	Pediksi cuaca ekstrem di Jepara	Nilai <i>Generalized Cross Validation</i> (GCV) = 0.236; MAPE = 1.887 menunjukkan tingkat kesalahan rendah
4	(Mujiasih, 2022)	Prakiraan Cuaca	Data sinoptik BMKG	<i>Random Forest, C4.5, Association Rule</i>	Pemanfaatan Data Mining untuk Prakiraan Cuaca	C4.5 memberikan akurasi terbaik (68.5%).
5	(Syahreza <i>et al.</i> , 2024)	Analisis Model Prediksi Cuaca	Data suhu udara, kelembapan udara, dan kecepatan angin dari BMKG Provinsi Jawa tahun 2020	<i>SVM, Gradient Boosting, Random Forest, dan Decision Tree</i>	Perbandingan beberapa algoritma prediksi cuaca	Akurasi tertinggi pada <i>Random Forest</i> sebesar 85.1%; <i>Decision Tree</i> 84.7%; <i>Gradient Boosting</i> 81.8%; <i>SVM</i> 67%.
6	(Hikmah <i>et al.</i> , 2023)	Peramalan Cuaca Ekstrem	Data cuaca ekstrem nasional	<i>ARIMA</i> dan <i>Recurrent Neural Network</i> (RNN)	Perbandingan model <i>ARIMA</i> dan RNN untuk peramalan cuaca	Model RNN memberikan MAPE 6.54 (lebih baik dari <i>ARIMA</i> 10.87), menunjukkan kemampuan prediksi tinggi
7	(Sangaji & Sutabri, 2025)	Prediksi Curah Hujan	Data curah hujan 2015–2023 (BPS)	<i>Random Forest, XGBoost, dan</i>	Prediksi curah hujan untuk	odel gabungan terbaik dengan MSE 14.125,78;

No	Referensi	Topik	Objek	Metode	Subyek	Hasil
			Sumatera Selatan)	<i>Ensemble</i> (gabungan RF 0,4 + XGB 0,6)	mitigasi kebakaran di Sumatera Selatan	RMSE 118.85; MAPE 11.87%; outperform dibanding RF dan <i>XGBoost</i> tunggal
8	(Mukeshkumar <i>et al.</i> , 2025)	Prediksi & Peringatan Cuaca Ekstrem	Data cuaca historis dan real-time dari berbagai sumber meteorologi	<i>Random Forest dan Long Short Term Memory</i> (LSTM)	Sistem peringatan cuaca ekstrem berbasis machine learning	<i>Random Forest</i> mencapai akurasi 95%, sedangkan LSTM mencapai akurasi 97%
9	(Ibrahim <i>et al.</i> , 2026)	Prediksi Cuaca	Data meteorologi BMKG Kota Semarang beberapa	<i>Random Forest, XGBoost, LightGBM, KNN</i> + <i>SMOTE</i>	Perbandingan beberapa algoritma prediksi cuaca	<i>Random Forest</i> meningkat dari <i>accuracy</i> 63.1% menjadi 77.2% setelah <i>SMOTE</i> , namun <i>LightGBM</i> memperoleh performa terbaik dengan <i>accuracy</i> 86.9%.
10	(Risanti <i>et al.</i> , 2024)	Prediksi Kondisi Cuaca	Data suhu udara, kelembapan udara, dan kecepatan angin BMKG Provinsi Jawa 2020	<i>SVM, Gradient Boosting, Decision Tree, Random Forest</i>	Perbandingan empat algoritma ML	<i>Random Forest</i> terbaik: akurasi 85.1%; <i>Decision Tree</i> 84.7%; <i>Gradient Boosting</i> 81.8%; 67%

Berdasarkan Tabel 2.1 penelitian terkait prediksi maupun prakiraan cuaca telah banyak dilakukan menggunakan berbagai metode dan jenis data. Penelitian oleh Dwiyantri dan Prianto (2023), Joses *et al.* (2024), Syahreza *et al.* (2024), Ibrahim dan Prabowo (2026), serta Risanti *et al.* (2024) memanfaatkan parameter dari data meteorologi dan parameter cuaca lainnya untuk memprediksi kondisi cuaca. Berdasarkan hasil penelitian terdahulu, algoritma *Random Forest*

memiliki kinerja yang baik dalam membangun model prediksi, dengan tingkat akurasi yang pada beberapa kasus menghasilkan nilai akurasi lebih tinggi dibandingkan metode lainnya.

Beberapa penelitian, seperti Joses *et al.* (2024), Syahreza *et al.* (2024), Ibrahim dan Prabowo (2026), serta Risanti *et al.* (2024) berfokus pada perbandingan berbagai algoritma *machine learning* untuk memperoleh model dengan performa terbaik. Selain itu, penelitian Sangaji dan Sutabri (2025) mengombinasikan *Random Forest* dan *XGBoost* melalui metode *ensemble*, sedangkan Ibrahim dan Prabowo (2026) menerapkan teknik SMOTE untuk meningkatkan kemampuan yang memiliki data yang tidak seimbang. Meskipun demikian, fokus penelitian masih didominasi oleh prediksi cuaca umum, prakiraan curah hujan, dan peringatan cuaca ekstrem yang menggunakan parameter meteorologi sebagai variabel utama.

Kajian yang secara khusus memanfaatkan data aktivitas petir sebagai sumber informasi utama dalam *forecasting* cuaca ekstrem masih relatif terbatas. Beberapa penelitian yang membahas cuaca ekstrem, seperti Amri *et al.* (2024), Hikmah *et al.* (2023), dan Mukeshkumar *et al.* (2025) menggunakan data curah hujan, kecepatan angin, maupun data meteorologi historis lainnya. Padahal, aktivitas petir merupakan salah satu fenomena atmosfer yang erat kaitannya dengan pembentukan awan konvektif dan kejadian cuaca ekstrem. Oleh karena itu, pemanfaatan karakteristik petir sebagai variabel prediktor masih memiliki peluang untuk dikembangkan lebih lanjut.

Kesenjangan penelitian lainnya terlihat pada penggunaan informasi

spasial dalam proses *forecasting*. Sebagian besar penelitian terdahulu lebih berfokus pada perbandingan algoritma maupun peningkatan performa model tanpa menganalisis pengaruh fitur spasial terhadap hasil *forecasting*. Padahal, fenomena petir memiliki karakteristik geografis yang kuat karena distribusinya dipengaruhi oleh lokasi kejadian. Informasi spasial berupa *latitude* dan *longitude* berpotensi memberikan kontribusi dalam mengenali pola kejadian cuaca ekstrem pada suatu wilayah.

Berdasarkan kesenjangan tersebut, penelitian ini menerapkan algoritma *Random Forest* untuk *forecasting* potensi cuaca ekstrem menggunakan data aktivitas petir yang diperoleh dari BMKG Stasiun Geofisika Kelas III Malang. Variabel yang digunakan meliputi CG+, CG-, IC, *latitude*, dan *longitude*. Selain membangun model *forecasting*, penelitian ini juga mengevaluasi pengaruh penggunaan fitur spasial melalui perbandingan dua skenario, yaitu menggunakan seluruh fitur dan tanpa fitur spasial. Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan model *forecasting* cuaca ekstrem berbasis data petir serta memberikan gambaran mengenai pengaruh penambahan informasi spasial terhadap kinerja algoritma *Random Forest*.

2.2 Cuaca Ekstrem

Cuaca ekstrem mengacu pada kondisi atmosfer yang mengalami perubahan secara singkat pada periode yang cepat dengan tingkat intensitas yang tinggi. Kondisi tersebut dapat meningkatkan risiko terjadinya gangguan terhadap lingkungan, infrastruktur, maupun aktivitas manusia (Eva *et al.*, 2024). Fenomena ini termasuk salah satu kejadian hidrometeorologi yang menyebabkan

berbagai bencana, seperti hujan lebat, badai petir, angin kencang, banjir, maupun tanah longsor (Lestari *et al.*, 2025). Terjadinya cuaca ekstrem dipengaruhi oleh berbagai faktor atmosfer, antara lain peningkatan suhu permukaan laut, kelembapan udara yang tinggi, serta proses konveksi vertikal yang kuat sehingga memicu terbentuknya awan konvektif (Simbolon *et al.*, 2023). Dalam penelitian ini, aktivitas petir digunakan sebagai indikator utama karena memiliki hubungan yang erat dengan pembentukan awan konvektif yang sering memicu hujan lebat dan badai lokal (Septiadi & Hadi, 2011). Oleh sebab itu, data aktivitas petir diterapkan sebagai sumber informasi untuk membangun model *forecasting* potensi cuaca ekstrem.

Dalam kajian klimatologi, kejadian cuaca ekstrem umumnya diidentifikasi menggunakan pendekatan statistik berdasarkan distribusi data historis. Salah satu metode yang sering digunakan adalah nilai ambang (*threshold*) yang ditentukan berdasarkan persentil tinggi, seperti persentil ke-90 (P90), persentil ke-95 (P95), atau persentil ke-99 (P99). Nilai ambang tersebut digunakan untuk membedakan kejadian normal dengan kejadian yang memiliki intensitas tinggi dan relatif jarang terjadi, sehingga dapat dikategorikan sebagai kejadian cuaca ekstrem.

Pada penelitian ini, persentil ke-90 (P90) digunakan sebagai nilai ambang (*threshold*) dalam proses pembentukan label kelas. Pemilihan P90 didasarkan pada pendekatan statistik yang umum digunakan dalam penelitian klimatologi untuk merepresentasikan kejadian dengan intensitas tinggi yang relatif jarang terjadi, namun memiliki dampak yang signifikan. Dengan menggunakan nilai

ambang tersebut, data yang memiliki jumlah sambaran petir (*Strike*) melebihi nilai P90 dikategorikan sebagai cuaca ekstrem (label 1), sedangkan data dengan nilai *Strike* yang kurang dari atau sama dengan nilai P90 dikategorikan sebagai tidak ekstrem (label 0).

Dalam penelitian ini, variabel *Strike* dimanfaatkan untuk menentukan tingkat intensitas aktivitas petir sekaligus membentuk label kelas. Oleh sebab itu, variabel tersebut dikeluarkan dari kumpulan fitur yang digunakan pada tahap pelatihan dan pengujian model. Langkah ini bertujuan menghindari *data leakage*, yaitu situasi ketika variabel yang menjadi dasar pembentukan target juga digunakan sebagai masukan model, sehingga hasil evaluasi berpotensi menjadi bias dan tidak mencerminkan kemampuan model dalam melakukan generalisasi terhadap data baru. Hal ini dilakukan untuk menghindari terjadinya *data leakage*, yaitu kondisi ketika informasi yang digunakan untuk membentuk variabel target juga diberikan sebagai variabel input kepada model. Jika variabel *Strike* tetap digunakan sebagai fitur masukan, model akan memperoleh informasi yang secara langsung berkaitan dengan target sehingga nilai evaluasi yang dihasilkan berpotensi menjadi terlalu tinggi dan tidak memberikan kontribusi dalam generalisasi terhadap data baru.

Pendekatan penggunaan threshold berdasarkan persentil tinggi juga telah digunakan pada penelitian (Zhang *et al.*, 2023) yang menyatakan bahwa identifikasi kejadian ekstrem dapat dilakukan menggunakan distribusi data historis dengan menetapkan nilai ambang pada persentil tinggi, seperti P90. Kejadian yang memiliki nilai melebihi ambang tersebut dikategorikan sebagai

kejadian ekstrem karena merepresentasikan kondisi yang relatif jarang terjadi namun memiliki dampak meteorologis yang signifikan. Oleh karena itu, penggunaan P90 pada penelitian ini dipandang sesuai sebagai dasar pembentukan label kelas dalam membangun model *forecasting* cuaca ekstrem berbasis aktivitas petir.

2.3 *Forecasting*

Forecasting merupakan proses analitis yang memanfaatkan informasi historis untuk menghasilkan perkiraan mengenai kondisi atau kejadian pada periode mendatang dengan menerapkan pendekatan kualitatif maupun kuantitatif (Arrijal *et al.*, 2020). Proses tersebut dilakukan dengan memanfaatkan pola, kecenderungan, dan hubungan antarvariabel pada data masa lalu untuk menghasilkan estimasi kondisi pada periode berikutnya. Meskipun hasil *forecasting* tidak dapat memberikan kepastian secara mutlak, informasi yang dihasilkan dapat digunakan sebagai dasar dalam pengambilan keputusan dan penyusunan langkah.

Perkembangan teknologi telah mendorong pemanfaatan metode *machine learning* dalam proses *forecasting*. Algoritma *machine learning* mampu mempelajari pola yang terdapat pada data historis sehingga dapat menghasilkan model yang digunakan untuk memperkirakan kondisi yang belum diketahui (Patrick *et al.*, 2025). Dibandingkan metode statistik konvensional, pendekatan ini lebih mampu menangani data dalam jumlah besar serta hubungan yang kompleks antarvariabel sehingga banyak diterapkan pada berbagai bidang, termasuk *forecasting* cuaca.

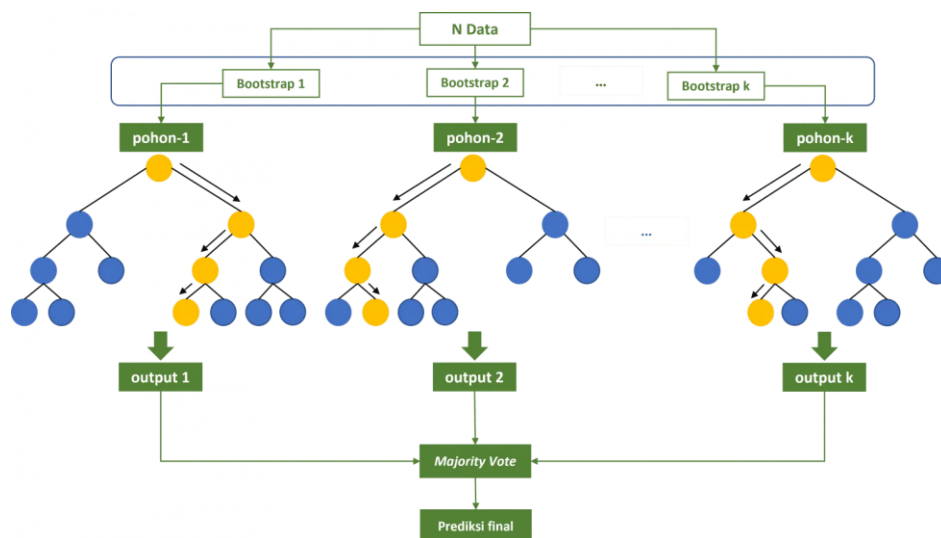
Dalam bidang meteorologi, *forecasting* cuaca bertujuan memperkirakan kondisi cuaca pada periode waktu tertentu berdasarkan data pengamatan sebelumnya. Informasi hasil *forecasting* sangat penting karena dapat dimanfaatkan sebagai dasar dalam mitigasi bencana, pengelolaan transportasi, pertanian, maupun kegiatan lain yang dipengaruhi oleh kondisi cuaca. Oleh karena itu, berbagai penelitian telah memanfaatkan algoritma *machine learning* untuk meningkatkan ketepatan hasil *forecasting* (Intan & Koswara, 2021).

Pada penelitian ini, konsep *forecasting* diterapkan dengan memanfaatkan model *Random Forest* yang dibangun dari data historis aktivitas petir. Model tersebut menghasilkan informasi mengenai potensi terjadinya cuaca ekstrem berdasarkan pola hubungan antara variabel CG+, CG-, IC, *Latitude*, dan *Longitude* dengan kejadian cuaca ekstrem pada data historis. Dengan demikian, konsep *forecasting* pada penelitian ini tidak mengacu pada *forecasting* nilai kontinu, melainkan pada pemanfaatan pola historis aktivitas petir untuk memperkirakan kemungkinan terjadinya kejadian cuaca ekstrem pada data yang belum diketahui untuk membangun model yang mampu mengenali pola kejadian cuaca ekstrem. Hasil *forecasting* tersebut diharapkan dapat mendukung penyediaan informasi dini sebagai bagian dari upaya mitigasi bencana hidrometeorologi.

2.4 Algoritma *Random Forest*

Algoritma *Random Forest* merupakan salah satu metode *machine learning* yang menerapkan pendekatan *ensemble learning* untuk menyelesaikan permasalahan klasifikasi maupun regresi. Algoritma ini diperkenalkan oleh Leo

Breiman pada tahun 2001 sebagai pengembangan dari algoritma *Decision Tree* dengan tujuan meningkatkan akurasi serta kemampuan generalisasi model (Rose & Hassen, 2019). Berbeda dengan *Decision Tree* tunggal, *Random Forest* membangun sejumlah pohon keputusan (*decision tree*) yang kemudian hasil prediksinya digabungkan menjadi satu keputusan akhir. Pendekatan *ensemble learning* tersebut memungkinkan model menghasilkan prediksi yang lebih akurat dan stabil dibandingkan penggunaan satu model saja (Siregar *et al.*, 2023). Alur kerja algoritma *Random Forest* ditunjukkan pada Gambar 2.1.



Gambar 2.1 Alur Kerja *Random Forest*

Berdasarkan Gambar 2.1 Alur kerja algoritma *Random Forest* diawali dengan pengambilan sampel data menggunakan teknik *bootstrap sampling*, yaitu pengambilan sampel secara acak dengan pengembalian (*sampling with replacement*). Selanjutnya, setiap data hasil *bootstrap* digunakan untuk membentuk *decision tree*. Pada setiap *node*, model memilih sebagian fitur secara acak, kemudian menentukan pemisahan data yang paling optimal berdasarkan

nilai *Gini Index* atau kriteria pemisahan lainnya. Hasil prediksi dari seluruh pohon kemudian digabungkan menggunakan mekanisme *majority voting* pada permasalahan klasifikasi atau *averaging* pada permasalahan regresi sehingga menghasilkan keluaran yang lebih akurat dan stabil (Nugroho, 2025).

Pendekatan tersebut menghasilkan kumpulan pohon keputusan yang beragam sehingga mampu meningkatkan kemampuan generalisasi model. Dibandingkan dengan *Decision Tree* tunggal, algoritma *Random Forest* memiliki beberapa keunggulan, di antaranya mampu mengurangi risiko *overfitting*, menangani data berdimensi tinggi, mempertahankan performa pada data yang tidak seimbang (*imbalanced data*), serta menghasilkan prediksi yang lebih stabil melalui penerapan teknik *bootstrapping* dan *averaging*. Berbagai keunggulan tersebut menjadikan *Random Forest* sebagai salah satu algoritma yang banyak digunakan dalam berbagai penelitian untuk menyelesaikan permasalahan klasifikasi maupun regresi lainnya (Efendi & Zyen, 2024); Inonu & Magda, 2025).

Pada penelitian ini, algoritma *Random Forest* digunakan untuk membangun model klasifikasi cuaca ekstrem berdasarkan data aktivitas petir. Variabel CG+, CG-, IC, *Latitude*, dan *Longitude* digunakan sebagai fitur masukan untuk mempelajari hubungan antara aktivitas petir dan kejadian cuaca ekstrem. Selanjutnya, performa model dievaluasi menggunakan *confusion matrix* melalui metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

2.5 Data Petir

Data petir merupakan kumpulan data hasil pencatatan aktivitas sambaran

petir yang memuat informasi mengenai waktu kejadian, jenis sambaran petir, serta lokasi terjadinya sambaran petir. Data tersebut digunakan untuk merepresentasikan karakteristik aktivitas kelistrikan atmosfer berdasarkan aspek temporal dan spasial sehingga dapat dimanfaatkan dalam analisis fenomena meteorologi, termasuk cuaca ekstrem (Paski *et al.*, 2017). Dalam penelitian ini digunakan beberapa atribut data petir yang dijelaskan pada Tabel 2.2.

Tabel 2.2 Definisi Atribut Data Petir

Variabel	Satuan (SI)	Definisi
<i>Datetime</i>	YYYY-MM-DD HH:MM:SS	waktu dan tanggal terjadinya sambaran petir
Strike	Jumlah	Total jumlah sambaran petir yang terdeteksi selama waktu pengamatan
CG+(0)	Jumlah	(<i>Cloud to Ground Positive</i>) sambaran petir bermuatan positif dari awan menuju permukaan bumi
CG-(1)	jumlah	(<i>Cloud to Ground Negative</i>) sambaran petir bermuatan negatif dari awan menuju permukaan bumi
IC(2)	jumlah	(<i>Intra Cloud</i>) sambaran petir yang terjadi di dalam awan tanpa mencapai permukaan bumi
<i>Latitude</i>	Derajat (°)	Posisi geografis utara–selatan lokasi sambaran petir
<i>Longitude</i>	Derajat (°)	Posisi geografis timur–barat lokasi sambaran petir.

Berdasarkan Tabel 2.2, data petir yang digunakan dalam penelitian ini terdiri atas atribut waktu kejadian (*Datetime*), jumlah total sambaran petir (*Strike*), jenis aktivitas petir yang meliputi *Cloud to Ground Positive* (CG+), *Cloud to Ground Negative* (CG-), dan *Intra Cloud* (IC), serta informasi lokasi berupa *Latitude* dan *Longitude*. Aktivitas petir dalam penelitian ini terdiri atas tiga jenis, yaitu *Cloud to Ground Positive* (CG+), *Cloud to Ground Negatif* (CG-), dan *Intra Cloud* (IC). CG+ merupakan sambaran petir bermuatan positif dari

awan menuju permukaan bumi. Meskipun jumlah kejadiannya relatif lebih sedikit dibandingkan CG-, sambaran CG+ umumnya memiliki arus listrik yang lebih besar sehingga berpotensi menimbulkan dampak yang lebih besar terhadap objek yang disambar.

Cloud to Ground Positive (CG+) merupakan jenis sambaran petir yang terjadi ketika muatan listrik positif dari bagian atas awan cumulonimbus dilepaskan menuju permukaan bumi. Sambaran ini relatif lebih jarang terjadi dibandingkan *Cloud to Ground Negatif (CG-)*, namun umumnya memiliki arus listrik dan energi yang lebih besar sehingga berpotensi menimbulkan dampak yang lebih besar, seperti kerusakan infrastruktur, kebakaran, maupun gangguan pada sistem kelistrikan.

Cloud to Ground Negatif (CG-) merupakan sambaran petir yang terjadi ketika muatan listrik negatif dari bagian bawah awan cumulonimbus mengalir menuju permukaan bumi. Jenis petir ini merupakan sambaran yang paling dominan terjadi. Penelitian Septiadi dan Hadi (2011) menunjukkan bahwa dominasi sambaran CG- berkaitan erat dengan peningkatan intensitas curah hujan akibat aktivitas konvektif yang kuat.

Intra Cloud (IC) merupakan sambaran petir yang terjadi di dalam awan, yaitu antara pusat-pusat muatan listrik positif dan negatif tanpa mencapai permukaan bumi. Aktivitas IC umumnya muncul pada fase awal perkembangan awan konvektif dan sering meningkat sebelum terjadinya hujan lebat atau badai petir. Oleh karena itu, aktivitas IC dapat dimanfaatkan sebagai salah satu indikator dalam mengidentifikasi potensi cuaca ekstrem.

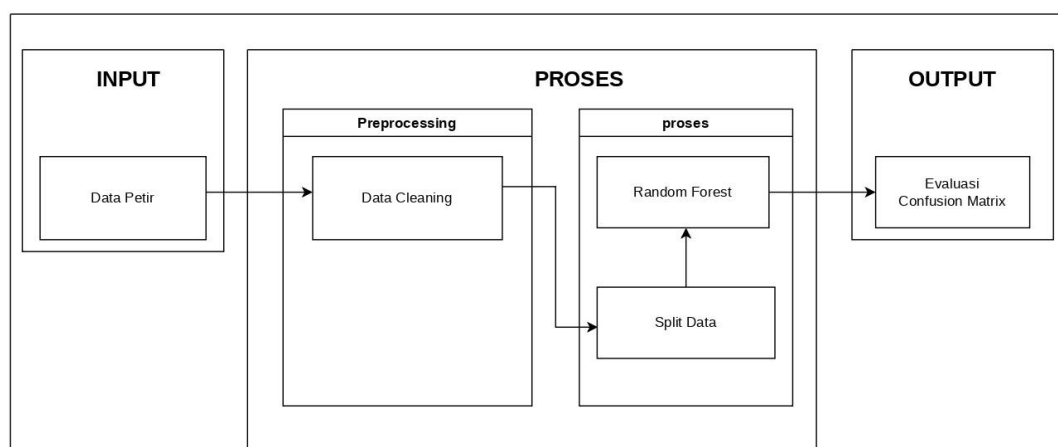
Selain karakteristik aktivitas petir, penelitian ini juga menggunakan latitude dan *longitude* sebagai representasi informasi spasial. *Latitude* (garis lintang) merupakan koordinat geografis yang menunjukkan posisi suatu lokasi terhadap garis khatulistiwa (ekuator), dengan nilai berkisar antara 90° LU hingga 90° LS. *Longitude* (garis bujur) merupakan koordinat geografis yang menunjukkan posisi suatu lokasi terhadap Meridian Utama, dengan nilai berkisar antara 180° BT hingga 180° BB. Kedua variabel tersebut menunjukkan lokasi geografis setiap kejadian sambaran petir sehingga memungkinkan analisis hubungan antara distribusi spasial aktivitas petir dengan kejadian cuaca ekstrem. Informasi spasial tersebut digunakan sebagai salah satu fitur masukan pada model *Random Forest* untuk mengetahui pengaruh lokasi terhadap proses klasifikasi cuaca ekstrem.

BAB III

METODE PENELITIAN

3.1 Desain Sistem

Subbab ini menjelaskan penerapan algoritma *Random Forest* dalam *forecasting* cuaca ekstrem berbasis data petir. Proses penelitian dimulai dari pengumpulan data sebagai dasar analisis, kemudian dilanjutkan dengan tahap preprocessing untuk menangani data hilang. Setelah itu, data dibagi menjadi data latih dan data uji (*split data*) untuk proses pembentukan dan pengujian model. Tahap akhir menggunakan algoritma *Random Forest* untuk melakukan *forecasting*, sedangkan *confusion matrix* digunakan untuk mengevaluasi performa model dalam *forecasting* cuaca ekstrem. Alur lengkap tahapan penelitian ditunjukkan pada Gambar 3.1.



Gambar 3.1 Desain Sistem

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data aktivitas petir yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)

melalui Stasiun Geofisika Kelas III Malang Karangates. Dataset mencakup hasil pengamatan selama periode Maret hingga Mei 2025 dengan interval pencatatan setiap satu jam sehingga diperoleh sebanyak 2.208 data pengamatan.

Dataset yang digunakan terdiri atas tujuh atribut, yaitu *Datetime*, *Strike*, *CG+*, *CG-*, *IC*, *Latitude*, dan *Longitude*. Atribut-atribut tersebut merepresentasikan informasi waktu pengamatan, aktivitas sambaran petir, dan koordinat geografis lokasi kejadian. Definisi masing-masing variabel telah dijelaskan pada Subbab 2.2 Data Petir. Gambaran bentuk data yang digunakan dalam penelitian disajikan pada Tabel 3.1, yang menampilkan contoh struktur dataset sebelum dilakukan tahap praproses dan pemodelan

Tabel 3.1 Data Petir

<i>Datetime</i>	<i>Strike</i>	<i>CG+ (0)</i>	<i>CG- (1)</i>	<i>IC (2)</i>	<i>Latitude</i>	<i>Longitude</i>
01-03-2025 00:00:00	3	0	0	3	- 9.219843755	114.9134671
01-03-2025 01:00:00	5	0	1	4	- 8.920031969	115.2040046
01-03-2025 03:00:00	2	0	0	2	- 9.258902952	113.9841746
01-03-2025 04:00:00	14	0	4	10	- 8.357837501	112.7476107
01-03-2025 05:00:00	94	10	52	32	- 7.570135577	109.9424452
...						
31-05-2025 19:00:00	938	629	309	162	- 8.064493201	112.6978525
31-05-2025 20:00:00	497	273	224	168	- 8.380991216	112.5935847
31-05-2025 21:00:00	196	126	70	84	- 8.69422474	112.3699785
31-05-2025 22:00:00	68	36	32	22	- 8.557175171	112.2761469

31-05-2025 23:00:00	77	22	55	33	-9.992452927	112.1810669
------------------------	----	----	----	----	--------------	-------------

Data petir yang diperoleh dari BMKG pada awalnya tersedia dalam resolusi waktu per menit berdasarkan waktu terjadinya setiap sambaran petir. Dalam penelitian ini, data tersebut terlebih dahulu diagregasi menjadi resolusi per jam dengan menjumlahkan seluruh kejadian sambaran petir yang terjadi pada setiap interval satu jam.

Proses agregasi dilakukan agar resolusi waktu data masukan sesuai dengan pembentukan label cuaca ekstrem yang juga ditetapkan pada skala waktu per jam. Keseragaman resolusi waktu antara data masukan dan target diperlukan untuk menjaga konsistensi proses pembelajaran model sehingga informasi temporal yang digunakan selama pelatihan dan pengujian tetap selaras.

Selain itu, penggunaan resolusi waktu per jam dinilai lebih representatif dalam menggambarkan perkembangan fenomena cuaca ekstrem dibandingkan data per menit yang cenderung sangat fluktuatif. Pendekatan agregasi waktu per jam juga umum diterapkan pada penelitian aktivitas petir, termasuk dalam pengolahan data *World Wide Lightning Location Network* (WWLLN), yang melakukan agregasi dan koreksi data berbasis interval waktu sebelum digunakan pada tahap analisis lebih lanjut (Kaplan & Lau, 2022).

3.3 Preprocessing

Preprocessing dilakukan untuk menyiapkan data agar sesuai dengan kebutuhan proses pembentukan model machine learning. Tahapan ini bertujuan meningkatkan kualitas data sehingga proses pelatihan model dapat berjalan

dengan lebih baik. Pada penelitian ini, *preprocessing* dilakukan melalui pembersihan data (*data cleaning*) untuk menangani nilai yang hilang (*missing values*).

3.3.1 Data Cleaning

Tahap data *cleaning* merupakan proses awal dalam menyiapkan data sebelum digunakan untuk analisis. Pada tahap ini, data petir yang diperoleh dari BMKG diperiksa secara menyeluruh untuk memastikan kualitas dan konsistensinya. Total data yang digunakan berjumlah 2.208 data, dengan 10 data yang mengalami missing value yang ditemukan pada tanggal 2 Maret dan 27 Mei 2025. Penanganan terhadap *missing values* dilakukan melalui penerapan metode interpolasi temporal yang digunakan untuk mengestimasi nilai yang tidak tersedia berdasarkan karakteristik data runtun waktu di sekitarnya (Lepot *et al.*, 2017). Metode ini diterapkan untuk memperkirakan nilai yang hilang berdasarkan pola atau tren data pada waktu sebelum dan sesudah pengamatan, sehingga kontinuitas data tetap terjaga. Dalam penelitian ini, setiap fitur diperiksa dan diperbaiki melalui proses interpolasi hingga tidak terdapat lagi nilai yang hilang. Dengan demikian, data yang digunakan sebagai input model menjadi lebih bersih dan siap digunakan dalam proses pelatihan model sehingga dapat menghasilkan model menghasilkan hasil *forecasting* yang lebih akurat.

Secara matematis, interpolasi temporal dihitung menggunakan rumus berikut:

$$X_i = (1 - \alpha)x_B + \alpha x_A \quad (3.1)$$

Keterangan :

X_i : nilai hasil interpolasi pada waktu ke- i

x_B : nilai sebelum data hilang

x_A : nilai setelah data hilang

α : faktor posisi waktu antara dua titik, bernilai antara 0–1

3.4 *Split Data*

Split data merupakan tahap penting sebelum model *Random Forest* dilatih, seluruh dataset dibagi menjadi dua himpunan data sesuai kebutuhan proses pengembangan model. Himpunan pertama berupa *training set* yang digunakan untuk mempelajari karakteristik data sehingga model mampu membentuk pola klasifikasi. Himpunan kedua berupa *testing set* yang berfungsi sebagai data evaluasi untuk mengetahui kemampuan model ketika dihadapkan pada data yang tidak digunakan selama proses pelatihan. Dalam penelitian ini, pembagian dataset diterapkan menggunakan tiga skenario, yaitu rasio 90:10, 80:20, dan 70:30

Penggunaan tiga skenario pembagian dataset tersebut bertujuan untuk mengkaji pengaruh proporsi data latih dan data uji terhadap kemampuan algoritma *Random Forest* dalam mengklasifikasikan kondisi cuaca ekstrem. Selain itu, variasi rasio tersebut juga digunakan untuk menilai kestabilan performa model ketika dilatih menggunakan jumlah data yang berbeda.

3.5 *Random Forest*

Random Forest merupakan algoritma *machine learning* berbasis pendekatan *ensemble learning* yang membangun sejumlah *decision tree* sebagai dasar dalam menghasilkan prediksi. Setiap pohon keputusan memberikan hasil klasifikasi secara independen, kemudian seluruh hasil tersebut dikombinasikan

melalui mekanisme pemungutan suara (*majority voting*) sehingga diperoleh keputusan akhir yang lebih akurat dan stabil (Haerudin *et al.*, 2025).

Berikut adalah langkah-langkah algoritma *Random Forest* (Fauzan *et al.*, 2024):

- a. Mengambil sampel dari dataset beberapa kali dengan teknik pengambilan sampel acak dengan pengembalian (*bootstrap*).
- b. Membuat pohon keputusan dengan memilih fitur yang dipilih secara acak di setiap node.

Ini dilakukan untuk meningkatkan keragaman antar pohon. Pemisahan data di setiap node didasarkan pada nilai *Gini Impurity*, dengan rumus:

$$Gini(s_i) = 1 - \sum_{i=1}^c p_i^2 \quad (3.2)$$

Keterangan:

s_i : menunjukkan node ke-i
 c : proporsi data ke-i dalam node
 p_i^2 : jumlah kelas

- c. Titik *split* terbaik ditentukan berdasarkan nilai *Gini Split*, yaitu nilai rata-rata tertimbang dari *impurity* pada dua subset hasil split:

$$Gini_{split} = \sum_{i=1}^k \left(\frac{n_i}{n} \times Gini(s_i) \right) \quad (3.3)$$

Keterangan:

n_i : jumlah data pada subset ke-i
 n : total data pada node
 $Gini(s_i)$: nilai gini dari subset ke-i

- d. Proses ini diulang-ulang untuk setiap node hingga pohon mencapai kedalaman maksimum atau node tidak dapat di *split* lebih lanjut.
- e. Untuk melakukan prediksi pada data uji, hasil dari setiap pohon digabungkan

dengan metode voting mayoritas, di mana kelas dengan suara terbanyak menjadi keputusan akhir model.

3.5.1 Penentuan *Hyperparameter Random Forest*

Penentuan *hyperparameter* pada algoritma *Random Forest* dilakukan untuk menentukan parameter awal sebelum proses pelatihan model. *Hyperparameter* merupakan parameter yang ditetapkan sebelum proses pembelajaran dimulai dan nilainya tidak berubah selama proses pelatihan berlangsung. Pemilihan *hyperparameter* yang tepat dapat memengaruhi kemampuan model dalam menghasilkan *forecasting* yang optimal.

Tabel 3.2 Nilai *Hyperparameter*

No	<i>Hyperparameter</i>	Nilai yang Digunakan	Keterangan
1	<i>n_estimators</i>	(10, 30, 50)	Menentukan jumlah pohon keputusan (<i>decision tree</i>) yang dibangun dalam model <i>Random Forest</i> .
2	<i>max_depth</i>	(10, 3, 5)	Menentukan kedalaman maksimum dari setiap pohon keputusan.
3	<i>min_samples_split</i>	(5, 10)	Menentukan jumlah minimum sampel yang diperlukan untuk melakukan pemisahan node
4	<i>min_samples_leaf</i>	(2,4)	Menentukan jumlah minimum sampel pada setiap node akhir (<i>leaf</i>).

3.6 Evaluasi Kinerja

Evaluasi kinerja dilakukan setelah model selesai dibangun untuk mengetahui kemampuan model dalam mengklasifikasikan data secara tepat. Tahap ini bertujuan menilai seberapa baik model menghasilkan prediksi dengan membandingkan hasil klasifikasi terhadap label sebenarnya. Pada penelitian ini, proses evaluasi menggunakan *Confusion Matrix*, yaitu metode yang menyajikan

hubungan antara hasil prediksi model dan kondisi aktual sehingga berbagai ukuran performa dapat dihitung. Setiap baris pada matriks menunjukkan kelas aktual, sedangkan setiap kolom merepresentasikan kelas hasil prediksi. Dari matriks tersebut dapat diketahui jumlah prediksi yang benar maupun yang salah sebagai dasar perhitungan metrik evaluasi.

1. *True Positive* (TP): merupakan jumlah data yang termasuk kelas positif dan berhasil diprediksi sebagai kelas positif oleh model.
2. *True Negative* (TN): merupakan jumlah data yang termasuk kelas negatif dan berhasil diklasifikasikan sebagai kelas negatif.
3. *False Positive* (FP): menunjukkan data yang sebenarnya berada pada kelas negatif, tetapi diprediksi sebagai kelas positif oleh model.
4. *False Negative* (FN): adalah jumlah data yang seharusnya berada pada kelas positif, tetapi diklasifikasikan sebagai kelas negatif..

Perhitungan untuk mengukur *accuracy*, *precision*, *recall* dan *F-1 Score*.

1. *Accuracy* menunjukkan tingkat ketepatan model dalam mengklasifikasikan seluruh data, baik yang termasuk kelas positif maupun kelas negatif. Nilai ini diperoleh dari perbandingan antara jumlah prediksi yang benar dengan keseluruhan data yang dievaluasi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.4)$$

2. *Precision* digunakan untuk mengukur tingkat ketepatan prediksi pada kelas positif. Metrik ini menunjukkan proporsi data yang diprediksi positif dan benar-benar termasuk ke dalam kelas positif.

$$Precision = \frac{TP}{TP+FP} \quad (3.5)$$

3. *Recall* menggambarkan kemampuan model dalam mengidentifikasi seluruh data yang benar-benar berasal dari kelas positif. Semakin tinggi nilai *recall*, semakin banyak data positif yang berhasil dikenali oleh model

$$Recall = \frac{TP}{TP+FN} \quad (3.6)$$

4. *F1-Score* merupakan metrik yang menggabungkan nilai *precision* dan *recall* dalam satu ukuran evaluasi. Nilai ini digunakan untuk memberikan gambaran yang lebih seimbang mengenai performa model, terutama ketika distribusi kelas pada dataset tidak seimbang

$$F1-Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (3.7)$$

3.7 Skenario Uji Coba

Dalam penelitian ini, pengujian dilakukan melalui enam skenario berdasarkan penggunaan fitur dan rasio pembagian data latih dan data uji. Skenario 1 hingga Skenario 3 termasuk dalam Skenario A yang menggunakan semua fitur, yaitu CG+, CG-, IC, *Latitude*, dan *Longitude*. Sementara itu, Skenario 4 hingga Skenario 6 termasuk dalam Skenario B yang menggunakan fitur tanpa spasial, yaitu CG+, CG-, dan IC, sehingga fitur *Latitude* dan *Longitude* tidak digunakan dalam proses pelatihan maupun pengujian model.

Pada masing-masing skenario digunakan tiga variasi rasio pembagian data, yaitu 90:10, 80:20, dan 70:30. Variasi rasio tersebut digunakan untuk menganalisis pengaruh proporsi data latih dan data uji terhadap performa model. Selain itu, pada setiap skenario dilakukan proses *tuning hyperparameter* untuk

memperoleh kombinasi parameter terbaik yang dapat meningkatkan kinerja algoritma *Random Forest*. Dengan demikian, enam skenario pengujian ini digunakan untuk menganalisis pengaruh penggunaan fitur spasial, variasi rasio pembagian data, serta tuning *hyperparameter* terhadap kinerja algoritma *Random Forest* dalam *forecasting* cuaca ekstrem. Skenario pengujian yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.3

Tabel 3.3 Skenario Uji Coba

Skenario	Penggunaan Fitur	Rasio Data
Skenario 1A	Semua Fitur (CG+, CG-, IC, <i>Latitude</i> , <i>Longitude</i>)	90:10
Skenario 2A	Semua Fitur (CG+, CG-, IC, <i>Latitude</i> , <i>Longitude</i>)	80:20
Skenario 3A	Semua Fitur (CG+, CG-, IC, <i>Latitude</i> , <i>Longitude</i>)	70:30
Skenario 4B	Tanpa Fitur Spasial (CG+, CG-, IC)	90:10
Skenario 5B	Tanpa Fitur Spasial (CG+, CG-, IC)	80:20
Skenario 6B	Tanpa Fitur Spasial (CG+, CG-, IC)	70:30

BAB IV

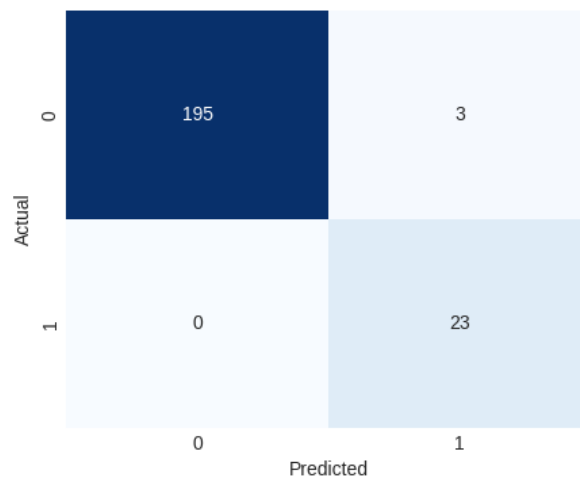
HASIL DAN PEMBAHASAN

4.1 Hasil Uji Coba

Bab ini menyajikan hasil pengujian, analisis, dan perbandingan performa model *Random Forest*. Pengujian dilakukan berdasarkan enam skenario yang telah dijelaskan pada Tabel 3.3. Skenario A menggunakan semua fitur, yaitu CG+, CG-, IC, *Latitude*, dan *Longitude*, sedangkan Skenario B menggunakan fitur tanpa spasial, yaitu CG+, CG-, dan IC. Pada masing-masing skenario diterapkan tiga variasi rasio pembagian data latih dan data uji, yaitu 90:10, 80:20, dan 70:30. Hasil pengujian kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall* dan *F1-score* untuk menganalisis pengaruh penggunaan fitur serta variasi rasio data terhadap kinerja model *Random Forest* dalam *forecasting* cuaca ekstrem. Selain itu, dilakukan perbandingan performa antar skenario untuk menentukan model dengan kinerja terbaik.

4.1.1 Skenario 1A dengan Rasio 90:10

Pengujian dilakukan pada Skenario 1A dengan rasio pembagian data 90:10 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan dpada *confusion matrix* sebagai berikut:



Gambar 4.1 *Confusion Matrix* Skenario 1A Rasio 90:10

memperlihatkan hasil confusion matrix pada Skenario 1A dengan rasio pembagian data 90:10. Visualisasi tersebut digunakan untuk mengevaluasi kemampuan algoritma dalam membedakan data yang termasuk kelas 0 dan kelas 1 melalui perbandingan antara label aktual dan hasil prediksi. Dari matriks ini dapat diketahui jumlah prediksi yang sesuai dengan kondisi sebenarnya maupun prediksi yang mengalami kesalahan.

Berdasarkan hasil evaluasi, model mampu mengidentifikasi kelas 0 dengan sangat baik. Hal ini terlihat dari nilai *True Negative* (TN) sebanyak 195, yang menunjukkan bahwa seluruh data tersebut berhasil dikenali sebagai kelas 0. Selain itu, terdapat 3 data yang termasuk *False Positive* (FP), yaitu data yang sebenarnya berada pada kelas 0 tetapi diprediksi sebagai kelas 1 oleh model. Untuk kelas 1, model menghasilkan 23 *True Positive* (TP), sehingga seluruh data yang termasuk kelas positif berhasil dikenali dengan benar. Sementara itu, nilai *False Negative* (FN) sebesar 0 mengindikasikan bahwa tidak terdapat data kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Secara keseluruhan, hasil *confusion matrix* memperlihatkan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Hal tersebut ditunjukkan oleh dominasi jumlah prediksi yang benar dibandingkan jumlah prediksi yang keliru. Kondisi tersebut berkontribusi terhadap tingginya nilai *accuracy*, *precision*, dan *recall* yang diperoleh pada skenario ini. Grafik nilai masing-masing metrik evaluasi disajikan pada bagian berikutnya. Berikut grafik *accuracy*, *precision*, *recall* dan *F1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.1 Hasil Evaluasi Pengujian Skenario 1A Rasio 90:10

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9864	0.8846	1.0000	0.9387

Tabel 4.1 menyajikan hasil evaluasi model pada Skenario 1A dengan rasio pembagian data 90:10 yang menggunakan seluruh fitur sebagai variabel masukan. Berdasarkan hasil pengujian, model menghasilkan *accuracy* sebesar 98.64%, *precision* 88.46%, *recall* 100.00%, dan *F1-Score* 93.87%. Nilai tersebut menunjukkan bahwa model mampu melakukan klasifikasi cuaca ekstrem dengan tingkat ketepatan yang sangat tinggi.

Tingginya nilai *accuracy* menunjukkan bahwa hampir seluruh data uji berhasil diklasifikasikan sesuai dengan kelas sebenarnya. Kinerja tersebut didukung oleh pemanfaatan seluruh fitur, yaitu CG+, CG-, IC, *latitude*, dan *longitude*, sehingga model dapat mempelajari karakteristik aktivitas petir beserta informasi spasial secara lebih menyeluruh. Kombinasi fitur tersebut membantu algoritma mengenali pola yang berkaitan dengan potensi terjadinya cuaca ekstrem.

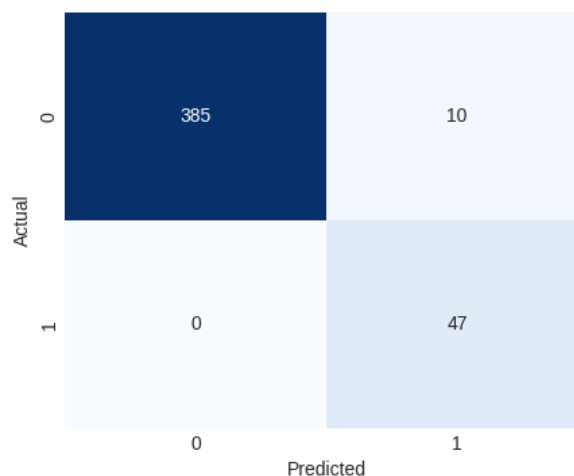
Selain itu, penggunaan rasio pembagian data 90:10 memberikan jumlah data pelatihan yang lebih banyak dibandingkan data pengujian. Kondisi tersebut

memungkinkan algoritma *Random Forest* memperoleh informasi yang lebih lengkap selama proses pembelajaran sehingga kemampuan model dalam melakukan generalisasi terhadap data baru menjadi lebih baik. Di samping itu, proses hyperparameter tuning turut berkontribusi terhadap peningkatan performa karena menghasilkan konfigurasi parameter yang paling sesuai dengan karakteristik dataset yang digunakan.

Nilai *precision* sebesar 88.46% mengindikasikan bahwa sebagian besar data yang diprediksi sebagai cuaca ekstrem benar-benar termasuk ke dalam kelas tersebut. Sementara itu, *recall* yang mencapai 100.00% menunjukkan bahwa seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model tanpa terdapat data positif yang terlewat (*false negative*). Hasil tersebut memperlihatkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi kejadian cuaca ekstrem.

4.1.2 Skenario 2A dengan Rasio 80:20

Pengujian dilakukan pada Skenario 2A dengan rasio pembagian data 80:20 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan dpada *confusion matrix* sebagai berikut:



Gambar 4.2 *Confusion Matrix* Skenario 2A Rasio 80:20

Gambar 4.2 menyajikan hasil *confusion matrix* pada Skenario 2A dengan rasio pembagian data 80:20. Visualisasi tersebut digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data ke dalam kelas 0 dan kelas 1 berdasarkan perbandingan antara label aktual dan hasil prediksi. Nilai yang terdapat pada *confusion matrix* menunjukkan jumlah prediksi yang sesuai dengan kondisi sebenarnya maupun prediksi yang mengalami kesalahan.

Berdasarkan hasil evaluasi, model mampu mengidentifikasi kelas 0 dengan sangat baik. Hal tersebut ditunjukkan oleh nilai *True Negative* (TN) sebanyak 385 data, yang mengindikasikan bahwa seluruh data tersebut berhasil diklasifikasikan sebagai kelas 0. Selain itu, terdapat 10 data yang termasuk *False Positive* (FP), yaitu data yang sebenarnya berada pada kelas 0 tetapi diprediksi sebagai kelas 1.

Pada kelas 1, model menghasilkan nilai *True Positive* (TP) sebanyak 47 data sehingga seluruh data pada kelas tersebut berhasil dikenali dengan benar. Sementara itu, nilai *False Negative* (FN) sebesar 0 menunjukkan bahwa tidak terdapat data kelas 1 yang salah diprediksi sebagai kelas 0. Hasil ini menunjukkan

bahwa model memiliki kemampuan yang sangat baik dalam membedakan kedua kelas dengan tingkat kesalahan yang relatif rendah.

Secara keseluruhan, dominasi jumlah prediksi yang benar dibandingkan prediksi yang keliru menunjukkan bahwa model memiliki performa klasifikasi yang tinggi. Selain menghasilkan nilai *accuracy* yang baik, model juga mampu mendeteksi seluruh data pada kelas positif tanpa menghasilkan *false negative*. Grafik nilai *accuracy*, *precision*, *recall*, dan *F1-score* untuk seluruh pengujian pada skenario ini disajikan pada bagian berikutnya. Berikut grafik *accuracy*, *presisi*, *recall* dan *f1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.2 Hasil Evaluasi Pengujian Skenario 2A Rasio 80:20

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9773	0.8245	1.0000	0.9038

Tabel 4.2 menyajikan hasil evaluasi kinerja model dalam Skenario 2A dengan rasio pembagian data 80:20 menggunakan semua fitur yang tersedia. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 97.73%, presisi 82.45%, *recall* 100.00%, dan *F1-Score* 90.38%. Tingginya nilai akurasi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan tepat oleh model. Performa yang baik ini dipengaruhi oleh penggunaan fitur-fitur terkait aktivitas petir (CG+, CG-, dan IC) serta fitur spasial (*latitude* dan *longitude*), yang memberikan informasi penting tentang karakteristik aktivitas petir dan lokasi kejadian, sehingga model dapat dengan efektif mengenali pola cuaca ekstrem.

Meskipun rasio pembagian data 80:20 menghasilkan jumlah data latih yang lebih sedikit dibandingkan dengan rasio 90:10, proporsi tersebut masih memadai untuk mempelajari pola historis aktivitas petir. Selain itu, penerapan kombinasi

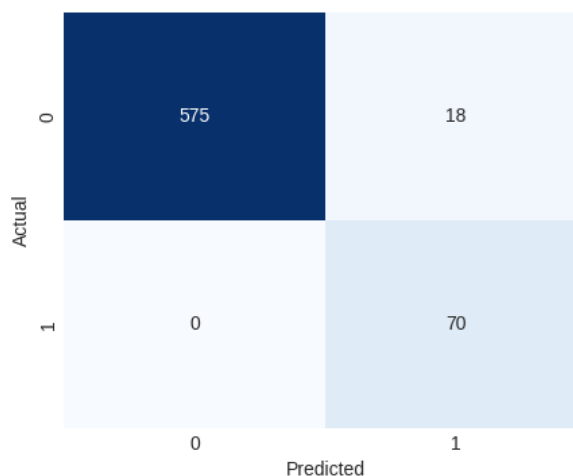
hyperparameter terbaik dari proses tuning hyperparameter juga berkontribusi dalam meningkatkan kemampuan model untuk melakukan generalisasi terhadap data uji, sehingga kualitas performa yang dihasilkan tetap tinggi.

Nilai presisi sebesar 82.45% menunjukkan bahwa sebagian besar prediksi cuaca ekstrem adalah akurat, meskipun ada beberapa *false positive*. Di sisi lain, nilai *recall* sebesar 100.00% menandakan bahwa semua kejadian cuaca ekstrem dalam data uji berhasil diidentifikasi tanpa adanya *false negative*. Hal ini menunjukkan bahwa model sangat efektif dalam mendeteksi seluruh kejadian cuaca ekstrem, sehingga risiko kejadian yang tidak terdeteksi dapat diminimalkan.

Nilai *F1-Score* yang mencapai 90.38% mencerminkan keseimbangan yang baik antara presisi dan recall. Meskipun nilai akurasi, presisi, dan *F1-Score* sedikit lebih rendah dibandingkan dengan Skenario 1A, performa model tetap sangat baik. Penurunan nilai tersebut kemungkinan disebabkan oleh berkurangnya jumlah data latih akibat peningkatan proporsi data uji, yang mengakibatkan informasi yang dipelajari oleh model menjadi sedikit lebih terbatas. Namun, algoritma *Random Forest* masih mampu melakukan *forecasting* cuaca ekstrem dengan performa tinggi pada rasio pembagian data 80:20.

4.1.3 Skenario 3A dengan Rasio 70:30

Pengujian dilakukan pada Skenario 3A dengan rasio pembagian data 70:30 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan dpada *confusion matrix* sebagai berikut:



Gambar 4.3 *Confusion Matrix* Skenario 3A Rasio 70:30

Gambar 4.3 hasil *confusion matrix* skenario 3A Rasio 70:30 menunjukkan performa model dalam memprediksi kelas 0 dan kelas 1. Nilai pada matriks menunjukkan jumlah data yang diprediksi dengan benar maupun salah oleh model. Model menunjukkan performa yang sangat baik dalam memprediksi kelas 0 dengan nilai *True Negative* (TN) sebanyak 575 data, yang berarti data pada kelas 0 berhasil diprediksi dengan benar oleh model. Nilai *False Positive* (FP) sebanyak 18 data, yang menunjukkan terdapat 18 data kelas 0 yang salah diprediksi sebagai kelas 1.

Pada kelas 1, model menghasilkan *True Positive* (TP) sebanyak 70 data, yang berarti seluruh data kelas 1 berhasil diprediksi dengan benar oleh model. Sementara itu, nilai *False Negative* (FN) sebesar 0 data, yang menunjukkan tidak terdapat data kelas 1 yang salah diprediksi sebagai kelas 0. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam membedakan kedua kelas, dengan tingkat *accuracy* yang tinggi dan jumlah kesalahan klasifikasi yang relatif rendah. Selain itu, tidak adanya nilai *false*

negative menunjukkan bahwa model mampu mendeteksi seluruh kejadian pada kelas 1 tanpa ada data yang terlewat. Berikut grafik *accuracy*, *presisi*, *recall* dan *f1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.3 Hasil Evaluasi Pengujian Skenario 3A Rasio 70:30

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9728	0.7954	1.0000	0.8860

Tabel 4.3 menunjukkan hasil evaluasi performa model pada Skenario 3A dengan rasio pembagian data 70:30 menggunakan seluruh fitur. Berdasarkan hasil pengujian, model memperoleh nilai *accuracy* sebesar 97.28%, *precision* sebesar 79.54%, *recall* sebesar 100.00%, dan *F1-score* sebesar 88.60%. Nilai *accuracy* yang tetap tinggi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model. Performa tersebut menunjukkan bahwa kombinasi fitur aktivitas petir (CG+, CG-, dan IC) serta fitur spasial (*latitude* dan *longitude*) masih mampu memberikan informasi yang relevan bagi algoritma *Random Forest* dalam mengenali pola historis cuaca ekstrem.

Meskipun demikian, nilai *accuracy*, *precision*, dan *F1-score* mengalami penurunan dibandingkan Skenario 1A dan Skenario 2A. Penurunan tersebut diduga disebabkan oleh berkurangnya proporsi data latih menjadi 70%, sehingga informasi yang digunakan model untuk mempelajari pola hubungan antara aktivitas petir, lokasi kejadian, dan kondisi cuaca ekstrem menjadi lebih terbatas. Akibatnya, kemampuan model dalam melakukan generalisasi terhadap data uji sedikit menurun, yang ditunjukkan dengan meningkatnya jumlah *false positive*.

Nilai *precision* sebesar 79.54% menunjukkan bahwa sebagian besar data yang diprediksi sebagai cuaca ekstrem merupakan prediksi yang benar, meskipun

masih terdapat beberapa data cuaca tidak ekstrem yang diprediksi sebagai cuaca ekstrem. Sementara itu, nilai *recall* sebesar 100.00% menunjukkan bahwa seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model tanpa adanya *false negative*. Hal ini menunjukkan bahwa model tetap sangat baik dalam mendeteksi seluruh kejadian cuaca ekstrem sehingga potensi kejadian yang tidak terdeteksi dapat dihindari.

Nilai *F1-score* sebesar 88.60% menunjukkan bahwa keseimbangan antara *precision* dan *recall* masih tergolong baik meskipun terjadi sedikit penurunan dibandingkan dua skenario sebelumnya. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma *Random Forest* tetap mampu melakukan *forecasting* potensi cuaca ekstrem dengan performa yang tinggi pada rasio pembagian data 70:30. Namun, hasil ini juga menunjukkan bahwa semakin besar proporsi data latih yang digunakan, semakin baik kemampuan model dalam mempelajari pola historis sehingga menghasilkan performa *forecasting* yang lebih optimal.

4.1.4 Skenario 4B dengan Rasio 90:10

Pengujian dilakukan pada Skenario 4B dengan rasio pembagian data 90:10 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan dpada *confusion matrix* sebagai berikut:

Actual	0	194	4
	1	0	23
		Predicted 0	Predicted 1

Gambar 4.4 *Confusion Matrix* Skenario 4B Rasio 90:10

Gambar 4.4 hasil *confusion matrix* skenario 4B Rasio 90:10 menunjukkan performa model dalam memprediksi kelas 0 dan kelas 1. Nilai pada matriks menunjukkan jumlah data yang diprediksi dengan benar maupun salah oleh model. Model menunjukkan performa yang sangat baik dalam memprediksi kelas 0 dengan nilai *True Negative* (TN) sebanyak 194 data, yang berarti data pada kelas 0 berhasil diprediksi dengan benar oleh model. Nilai *False Positive* (FP) sebanyak 4 data, yang menunjukkan terdapat 4 data kelas 0 yang salah diprediksi sebagai kelas 1.

Pada kelas 1, model menghasilkan *True Positive* (TP) sebanyak 23 data, yang berarti seluruh data kelas 1 berhasil diprediksi dengan benar. Sementara itu, nilai *False Negative* (FN) sebesar 0 data, yang menunjukkan tidak terdapat data kelas 1 yang salah diprediksi sebagai kelas 0. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam membedakan kedua kelas, dengan tingkat kesalahan prediksi yang rendah. Selain itu, tidak adanya nilai *false negative* menunjukkan bahwa model mampu mendeteksi seluruh data pada

kelas 1 tanpa melewatkan satu pun kejadian cuaca ekstrem. Berikut grafik *accuracy*, *presisi*, *recall* dan *f1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.4 Hasil Evaluasi Pengujian Skenario 4B Rasio 90:10

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9819	0.8518	1.0000	0.9200

Tabel 4.4 menunjukkan hasil evaluasi performa model pada Skenario 4B dengan rasio pembagian data 90:10 tanpa menggunakan fitur spasial. Berdasarkan hasil pengujian, model memperoleh nilai *accuracy* sebesar 98.19%, *precision* sebesar 85.18%, *recall* sebesar 100.00%, dan *F1-score* sebesar 92.00%. Nilai *accuracy* yang tinggi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model. Meskipun hanya menggunakan fitur aktivitas petir, yaitu CG+, CG-, dan IC, ketiga fitur tersebut mampu memberikan informasi yang kuat mengenai karakteristik cuaca ekstrem sehingga algoritma *Random Forest* masih dapat mengenali pola historis dengan baik.

Penggunaan rasio pembagian data 90:10 juga memberikan jumlah data latih yang lebih besar sehingga model memiliki lebih banyak informasi untuk mempelajari hubungan antara aktivitas petir dan kejadian cuaca ekstrem. Selain itu, penggunaan kombinasi *hyperparameter* terbaik hasil *hyperparameter tuning* membantu meningkatkan kemampuan generalisasi model terhadap data uji sehingga menghasilkan performa yang tinggi meskipun tanpa memanfaatkan fitur spasial.

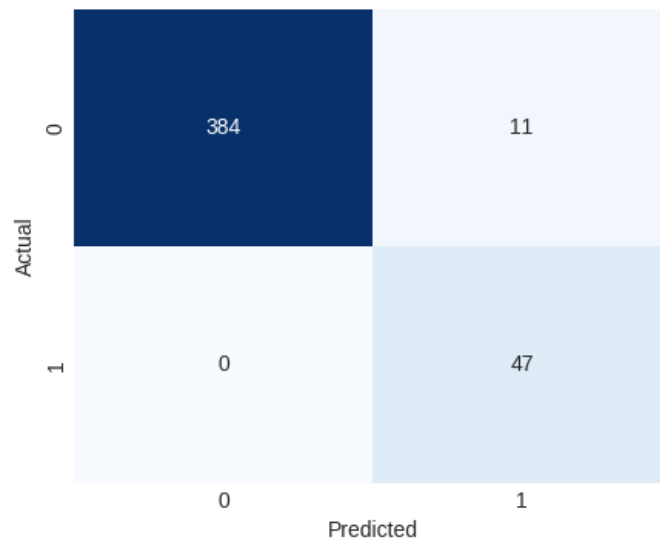
Nilai *precision* sebesar 85.18% mengindikasikan bahwa sebagian besar prediksi pada kelas cuaca ekstrem sesuai dengan kondisi sebenarnya, meskipun masih terdapat sejumlah data yang salah diprediksi sebagai cuaca ekstrem (*false*

positive). Di sisi lain, nilai *recall* sebesar 100.00% menunjukkan bahwa seluruh data yang termasuk kelas cuaca ekstrem berhasil dikenali oleh model selama proses pengujian seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model tanpa adanya *false negative*. Hal ini menunjukkan bahwa aktivitas petir memiliki hubungan yang kuat dengan kejadian cuaca ekstrem sehingga mampu dimanfaatkan sebagai indikator dalam proses *forecasting*.

Nilai *F1-score* sebesar 92.00% menunjukkan keseimbangan yang sangat baik antara *precision* dan *recall*. Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma *Random Forest* tetap mampu melakukan *forecasting* cuaca ekstrem dengan performa yang sangat baik meskipun hanya menggunakan fitur aktivitas petir. Namun, apabila dibandingkan dengan Skenario 1A yang menggunakan fitur spasial, nilai *accuracy*, *precision*, dan *F1-score* pada Skenario 4B sedikit lebih rendah. Hal ini mengindikasikan bahwa informasi lokasi (*latitude* dan *longitude*) memberikan informasi tambahan yang membantu model mengenali variasi karakteristik cuaca ekstrem di setiap wilayah sehingga meningkatkan performa *forecasting*.

4.1.5 Skenario 5B dengan Rasio 80:20

Pengujian dilakukan pada Skenario 5B dengan rasio pembagian data 90:10 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan pada *confusion matrix* sebagai berikut:



Gambar 4.5 *Confusion Matrix* Skenario 5B Rasio 80:20

Gambar 4.5 hasil *confusion matrix* skenario 5B Rasio 80:20 menunjukkan performa model dalam memprediksi kelas 0 dan kelas 1. Nilai pada matriks menunjukkan jumlah data yang diprediksi dengan benar maupun salah oleh model. Model menunjukkan performa yang sangat baik dalam memprediksi kelas 0 dengan nilai *True Negative* (TN) sebanyak 384 data, yang berarti data pada kelas 0 berhasil diprediksi dengan benar oleh model. Nilai *False Positive* (FP) sebanyak 11 data, kelas 0 yang salah diprediksi sebagai kelas 1.

Pada kelas 1, model menghasilkan *True Positive* (TP) sebanyak 47 data, yang berarti seluruh data kelas 1 berhasil diprediksi dengan benar oleh model. Sementara itu, nilai *False Negative* (FN) sebesar 0 data, yang menunjukkan tidak terdapat data kelas 1 yang salah diprediksi sebagai kelas 0. Hasil evaluasi menunjukkan bahwa model mampu membedakan data cuaca ekstrem dan tidak ekstrem dengan baik, yang ditunjukkan oleh rendahnya jumlah kesalahan prediksi. Selain itu, tidak ditemukannya nilai *false negative* mengindikasikan bahwa seluruh

data pada kelas 1 berhasil dikenali oleh model tanpa ada kejadian cuaca ekstrem yang terlewatkan sehingga model memiliki kemampuan deteksi yang sangat baik terhadap kelas positif. Berikut grafik *accuracy*, *presisi*, *recall* dan *f1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.5 Hasil Evaluasi Pengujian Skenario 5B Rasio 80:20

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9751	0.8103	1.0000	0.8952

Tabel 4.5 menunjukkan hasil evaluasi performa model pada Skenario 5B dengan rasio pembagian data 80:20 tanpa menggunakan fitur spasial. Berdasarkan hasil pengujian, model memperoleh nilai *accuracy* sebesar 97.51%, *precision* sebesar 81.03%, *recall* sebesar 100.00%, dan *F1-score* sebesar 89.52%. Nilai *accuracy* yang tinggi menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model. Performa tersebut menunjukkan bahwa fitur aktivitas petir, yaitu CG+, CG-, dan IC, masih mampu memberikan informasi yang cukup bagi algoritma *Random Forest* untuk mengenali pola historis yang berkaitan dengan kejadian cuaca ekstrem meskipun tanpa memanfaatkan fitur spasial.

Penggunaan rasio pembagian data 80:20 masih memberikan jumlah data latih yang memadai sehingga model dapat mempelajari hubungan antara aktivitas petir dan kejadian cuaca ekstrem dengan baik. Selain itu, penggunaan kombinasi *hyperparameter* terbaik hasil *hyperparameter tuning* turut meningkatkan kemampuan model dalam melakukan generalisasi terhadap data uji sehingga menghasilkan performa yang tetap tinggi.

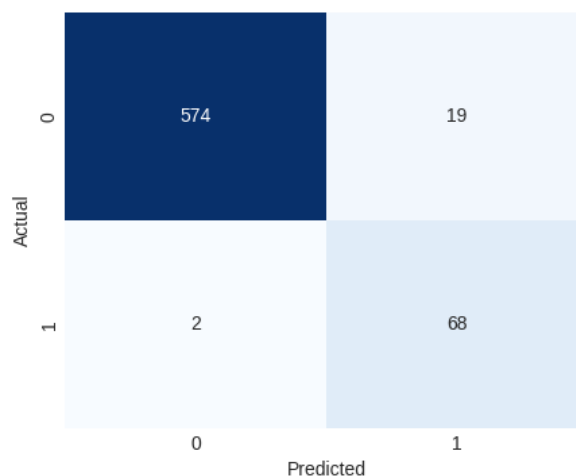
Nilai *precision* sebesar 81,03% mengindikasikan bahwa sebagian besar

prediksi yang dikategorikan sebagai cuaca ekstrem sesuai dengan kondisi sebenarnya, meskipun masih terdapat sejumlah prediksi yang tidak tepat. Di sisi lain, nilai *recall* yang mencapai 100,00% menunjukkan bahwa seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model tanpa menghasilkan *false negative*. Hasil tersebut mengindikasikan bahwa karakteristik aktivitas petir memiliki hubungan yang erat dengan kejadian cuaca ekstrem, sehingga dapat dimanfaatkan sebagai salah satu indikator dalam proses *forecasting*..

Nilai *F1-score* sebesar 89.52% menunjukkan bahwa keseimbangan antara *precision* dan *recall* masih tergolong baik. Meskipun demikian, nilai *accuracy*, *precision*, dan *F1-score* sedikit lebih rendah dibandingkan Skenario 4B. Penurunan tersebut diduga disebabkan oleh berkurangnya jumlah data latih akibat peningkatan proporsi data uji, sehingga informasi yang dipelajari model menjadi lebih terbatas. Selain itu, tidak digunakannya fitur spasial menyebabkan model hanya mengandalkan informasi aktivitas petir tanpa mempertimbangkan karakteristik lokasi kejadian. Secara keseluruhan, algoritma *Random Forest* tetap mampu melakukan *forecasting* cuaca ekstrem dengan performa yang baik meskipun hanya menggunakan fitur aktivitas petir.

4.1.6 Skenario 6B dengan Rasio 70:30

Pengujian dilakukan pada Skenario 4B dengan rasio pembagian data 90:10 dengan menerapkan proses *hyperparameter tuning*. Hasil pengujian ditunjukkan dpada *confusion matrix* sebagai berikut:



Gambar 4.6 *Confusion Matrix* Skenario 6B Rasio 70:30

Gambar 4.6 menyajikan hasil *confusion matrix* skenario 6B dengan rasio pembagian data 70:30 yang menggunakan kombinasi fitur tanpa informasi spasial. Visualisasi tersebut digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan data ke dalam kelas 0 dan kelas 1 berdasarkan perbandingan antara label aktual dan hasil prediksi. Pada kelas 0, model menghasilkan nilai *True Negative* (TN) sebanyak 574 data, yang menunjukkan bahwa sebagian besar data non-ekstrem berhasil dikenali dengan benar. Namun, masih terdapat 19 data yang termasuk *False Positive* (FP), yaitu data kelas 0 yang diprediksi sebagai kelas 1.

Pada kelas 1, model memperoleh nilai *True Positive* (TP) sebanyak 68 data, sehingga sebagian besar kejadian cuaca ekstrem berhasil diidentifikasi secara tepat. Sementara itu, nilai *False Negative* (FN) sebesar 2 menunjukkan bahwa masih terdapat dua data cuaca ekstrem yang salah diklasifikasikan sebagai kelas non-ekstrem. Secara keseluruhan, hasil *confusion matrix* memperlihatkan bahwa model tetap memiliki kemampuan klasifikasi yang baik meskipun tingkat kesalahan prediksi sedikit lebih tinggi dibandingkan skenario sebelumnya. Kondisi

tersebut menunjukkan bahwa model masih mampu mengenali sebagian besar kejadian cuaca ekstrem dengan baik, walaupun kemampuan deteksinya mengalami sedikit penurunan. Berikut grafik *accuracy*, *presisi*, *recall* dan *f1-score* dari seluruh pengujian pada skenario ini:

Tabel 4.6 Hasil Evaluasi Pengujian Skenario 6B Rasio 70:30

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0.9683	0.7816	0.9714	0.8662

Tabel 4.6 menunjukkan hasil evaluasi performa model pada Skenario 6B dengan rasio pembagian data 70:30 tanpa menggunakan fitur spasial. Berdasarkan hasil tersebut, model memperoleh nilai *accuracy* sebesar 96,83%, yang menunjukkan bahwa sebagian besar data cuaca ekstrem maupun tidak ekstrem berhasil diklasifikasikan dengan benar. Nilai *accuracy* yang tetap tinggi menunjukkan bahwa informasi yang terkandung pada fitur aktivitas petir, yaitu CG+, CG-, dan IC, masih mampu merepresentasikan pola kejadian cuaca ekstrem meskipun tanpa didukung informasi spasial.

Tabel 4.6 menyajikan hasil evaluasi model pada Skenario 6B dengan rasio pembagian data 70:30 menggunakan fitur tanpa informasi spasial. Berdasarkan hasil evaluasi, model memperoleh nilai *accuracy* sebesar 96.83%, *precision* sebesar 78.16%, *recall* sebesar 97.14%, dan *F1-score* sebesar 86.62%. Nilai tersebut menunjukkan bahwa model masih mampu menghasilkan performa klasifikasi yang baik meskipun informasi lokasi tidak digunakan sebagai fitur masukan. Hal ini mengindikasikan bahwa variabel aktivitas petir, yaitu CG+, CG-, dan IC, tetap memberikan informasi yang cukup untuk membedakan kondisi cuaca ekstrem dan tidak ekstrem.

Nilai *precision* sebesar 78.16% menunjukkan bahwa sebagian besar prediksi yang dikategorikan sebagai cuaca ekstrem sesuai dengan kondisi sebenarnya, meskipun masih terdapat beberapa prediksi yang tidak tepat. Di sisi lain, nilai *recall* sebesar 97.14% mengindikasikan bahwa hampir seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model, walaupun masih terdapat sebagian kecil data positif yang tidak berhasil terdeteksi. Kondisi tersebut menyebabkan nilai *F1-score* sedikit lebih rendah dibandingkan skenario dengan performa terbaik, karena metrik tersebut dipengaruhi oleh keseimbangan antara *precision* dan *recall*.

Nilai *recall* sebesar 97.14% menunjukkan bahwa hampir seluruh kejadian cuaca ekstrem berhasil dikenali oleh model, meskipun masih terdapat sebagian kecil data cuaca ekstrem yang tidak terdeteksi (*false negative*). Sementara itu, nilai *F1-score* sebesar 86.62% menunjukkan bahwa model masih memiliki keseimbangan yang baik antara *precision* dan *recall* dalam mengklasifikasikan kedua kelas.

Secara keseluruhan, nilai evaluasi yang tetap tinggi pada skenario ini menunjukkan bahwa aktivitas petir memiliki hubungan yang kuat dengan kejadian cuaca ekstrem sehingga mampu menjadi sumber informasi utama dalam proses *forecasting*. Namun, dibandingkan skenario yang menggunakan fitur spasial, penurunan nilai *accuracy*, *precision*, *recall*, dan *F1-score* mengindikasikan bahwa informasi lokasi kejadian turut membantu *Random Forest* mengenali pola cuaca ekstrem dengan lebih baik. Selain itu, penggunaan rasio data latih yang lebih kecil (70%) menyebabkan variasi pola yang dipelajari model menjadi lebih terbatas

sehingga kemampuan generalisasi model sedikit menurun dibandingkan rasio 90:10 dan 80:20.

4.2 Pembahasan Analisis Skenario

Pengujian dilakukan untuk menganalisis pengaruh penggunaan fitur terhadap performa model dalam melakukan peramalkan cuaca ekstrem.

4.2.1 Analisis Skenario A (semua fitur)

Hasil evaluasi model *Random Forest* pada skenario A (semua fitur) ditunjukkan pada Tabel 4.7 dengan variasi rasio *split data* dan *hyperparameter* terbaik pada setiap pengujian.

Tabel 4.7 Analisis skenario A (semua fitur)

Skenario A (Semua Fitur)	Rasio <i>Split Data</i>	<i>Best Hyperparameters</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Skenario 1A	90:10	(50, 10, 5, 2)	0.9864	0.8846	1.0000	0.9387
Skenario 2A	80:20	(10, 5, 10, 2)	0.9777	0.8245	1.0000	0.9038
Skenario 3A	70:30	(30, 3, 5, 2)	0.9728	0.7954	1.0000	0.8860

Hasil evaluasi model *Random Forest* pada Skenario A (Semua Fitur) ditunjukkan pada Tabel 4.7 dengan variasi rasio pembagian data dan kombinasi *hyperparameter* terbaik pada setiap pengujian. Pengujian ini bertujuan untuk menganalisis pengaruh penggunaan seluruh fitur terhadap performa model dalam melakukan *forecasting* potensi cuaca ekstrem.

Hasil evaluasi memperlihatkan bahwa model memberikan performa yang konsisten pada seluruh skenario pembagian data. Kinerja terbaik diperoleh pada

Skenario 1A dengan rasio 90:10, yang menghasilkan *accuracy* sebesar 98.64%, *precision* sebesar 88.46%, *recall* sebesar 100.00%, dan *F1-score* sebesar 93.87%. Nilai *recall* yang mencapai 100.00% menunjukkan bahwa tidak terdapat kejadian cuaca ekstrem yang terlewat selama proses klasifikasi. Selain itu, tingginya nilai *precision* mengindikasikan bahwa prediksi yang diberikan model untuk kategori cuaca ekstrem sebagian besar telah sesuai dengan label sebenarnya.

Berdasarkan hasil evaluasi, model mampu mempertahankan kinerja klasifikasi yang tinggi pada setiap variasi rasio pembagian data. Pada Skenario 1A dengan rasio 90:10, model memperoleh nilai *accuracy* sebesar 98.64%, *precision* sebesar 88.46%, *recall* sebesar 100.00%, dan *F1-score* sebesar 93.87%. Capaian nilai *recall* yang sempurna mengindikasikan bahwa seluruh data cuaca ekstrem pada data uji berhasil diidentifikasi tanpa adanya *false negative*. Di sisi lain, nilai *precision* yang relatif tinggi menunjukkan bahwa sebagian besar prediksi yang diklasifikasikan sebagai cuaca ekstrem sesuai dengan kondisi aktual.

Pada Skenario 2A dengan rasio 80:20, model memperoleh nilai *accuracy* sebesar 97.77%, *precision* sebesar 82.45%, *recall* sebesar 100.00%, dan *F1-score* sebesar 90.38%. Meskipun terjadi sedikit penurunan pada *accuracy*, *precision*, dan *F1-score*, model tetap mampu mendeteksi seluruh kejadian cuaca ekstrem. Penurunan tersebut terjadi karena jumlah data latih yang digunakan lebih sedikit dibandingkan rasio 90:10 sehingga pola hubungan antarvariabel yang dipelajari model menjadi lebih terbatas.

Pada Skenario 3A dengan rasio 70:30, model menghasilkan nilai *accuracy* sebesar 97.28%, *precision* sebesar 79.54%, *recall* sebesar 100.00%, dan *F1-score*

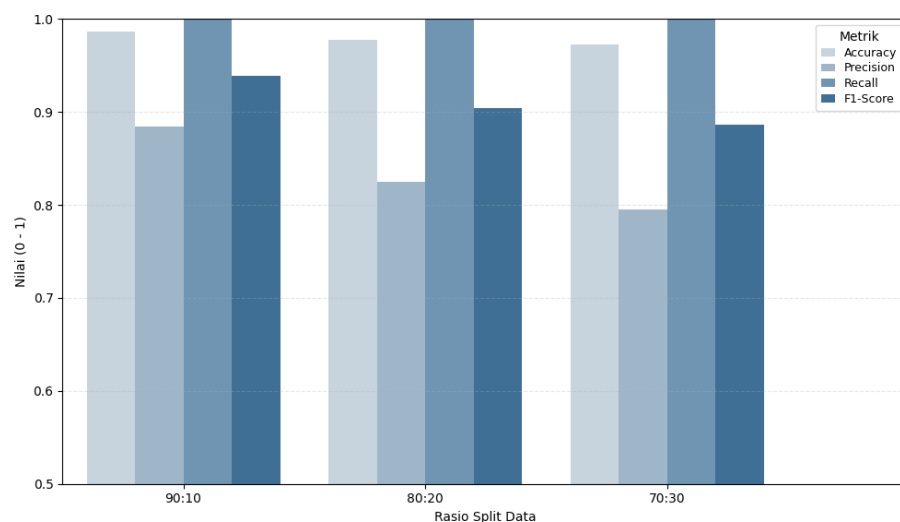
sebesar 88.60%. Nilai *precision* dan *F1-score* kembali mengalami penurunan, yang menunjukkan meningkatnya jumlah *false positive*. Kondisi tersebut menunjukkan bahwa semakin sedikit data latih yang digunakan, kemampuan model dalam membedakan karakteristik cuaca ekstrem dan tidak ekstrem menjadi sedikit menurun.

Penggunaan seluruh fitur memberikan kontribusi yang baik terhadap performa model karena model tidak hanya memanfaatkan informasi aktivitas petir berupa CG+, CG-, dan IC, tetapi juga informasi spasial berupa *latitude* dan *longitude*. Informasi spasial membantu model mengenali bahwa karakteristik cuaca ekstrem tidak hanya dipengaruhi oleh intensitas aktivitas petir, tetapi juga oleh lokasi terjadinya sambaran petir. Setiap wilayah memiliki karakteristik geografis yang berbeda, seperti topografi, ketinggian, kedekatan dengan laut, serta kondisi atmosfer lokal yang memengaruhi pembentukan awan konvektif. Perbedaan karakteristik tersebut menyebabkan hubungan antara aktivitas petir dan kejadian cuaca ekstrem dapat bervariasi pada setiap lokasi. Dengan adanya informasi koordinat geografis, model mampu mempelajari pola spasial tersebut sehingga proses klasifikasi menjadi lebih akurat dibandingkan apabila hanya menggunakan informasi aktivitas petir.

Secara pengujian pada Skenario A menghasilkan performa klasifikasi yang sangat baik pada setiap variasi rasio pembagian data. Hal tersebut terlihat dari nilai *recall* yang konsisten mencapai 100.00% pada seluruh skenario, sehingga seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali oleh model tanpa ada data positif yang terlewat.

Berdasarkan grafik pada Gambar 4.7, seluruh metrik evaluasi pada Skenario 1A, Skenario 2A, dan Skenario 3A memperlihatkan kinerja model yang tetap tinggi meskipun rasio pembagian data berbeda. Nilai *accuracy*, *precision*, dan *F1-score* mengalami penurunan secara bertahap seiring berkurangnya proporsi data latih, sedangkan nilai *recall* tetap bertahan pada angka 100,00% untuk seluruh skenario. Temuan tersebut menunjukkan bahwa model mampu mengenali seluruh data yang termasuk kategori cuaca ekstrem, sementara penambahan jumlah data latih berkontribusi terhadap peningkatan ketepatan prediksi secara keseluruhan.

Skenario 1A memberikan hasil terbaik karena memanfaatkan proporsi data latih yang paling besar dibandingkan skenario lainnya. Dengan jumlah data pelatihan yang lebih banyak, algoritma *Random Forest* memperoleh informasi yang lebih lengkap untuk mempelajari karakteristik aktivitas petir, sehingga kemampuan model dalam melakukan generalisasi terhadap data baru menjadi lebih optimal.



Gambar 4.7 Metrik Evaluasi skenario A (semua fitur)

Berdasarkan grafik pada Gambar 4.7, seluruh metrik evaluasi pada Skenario 1A, Skenario 2A, dan Skenario 3A menunjukkan performa yang sangat baik pada setiap rasio pembagian data. Nilai *accuracy*, *precision*, *recall*, dan *F1-score* berada pada kategori tinggi, yang menunjukkan bahwa model mampu mengklasifikasikan kondisi cuaca sebagai dasar *forecasting* potensi cuaca ekstrem dengan baik. Nilai *recall* yang konsisten mencapai 100.00% menunjukkan bahwa seluruh kejadian cuaca ekstrem berhasil dikenali oleh model pada ketiga skenario.

Skenario 1A menghasilkan performa terbaik karena menggunakan proporsi data latih yang paling besar, sehingga model memperoleh informasi yang lebih banyak untuk mempelajari hubungan antara aktivitas petir, lokasi kejadian, dan kondisi cuaca ekstrem. Sebaliknya, pada Skenario 2A dan Skenario 3A terlihat adanya sedikit penurunan nilai *accuracy*, *precision*, dan *F1-score* seiring bertambahnya proporsi data uji. Kondisi ini menunjukkan bahwa berkurangnya jumlah data latih menyebabkan kemampuan model dalam mempelajari variasi pola data menjadi sedikit berkurang.

Selain itu, penggunaan fitur spasial berupa *latitude* dan *longitude* turut berkontribusi terhadap tingginya performa model pada seluruh skenario. Informasi lokasi membantu model membedakan karakteristik aktivitas petir pada wilayah yang memiliki kondisi geografis berbeda. Dengan demikian, model tidak hanya memanfaatkan jumlah sambaran petir, tetapi juga mempertimbangkan lokasi terjadinya sambaran sebagai informasi tambahan dalam mengenali pola cuaca ekstrem. Hal inilah yang menyebabkan model dengan seluruh fitur mampu menghasilkan performa yang lebih baik dibandingkan model tanpa fitur spasial.

4.2.2 Analisis Skenario B (Tanpa Fitur Spasial)

Hasil evaluasi model *Random Forest* pada skenario B (tanpa fitur spasial) ditunjukkan pada Tabel 4.8 dengan variasi rasio *split data* dan *hyperparameter* terbaik pada setiap pengujian.

Tabel 4.8 Analisis skenario B (tanpa fitur spasial)

Skenario B (Tanpa Fitur Spasial)	Rasio <i>Split Data</i>	<i>Best Hyperparameters</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Skenario 4B	90:10	(50, 5, 5, 4)	0.9819	0.8518	1.0000	0.9200
Skenario 5B	80:20	(50, 5, 10, 2)	0,9751	0.8103	1.0000	0.8952
Skenario 6B	70:30	(30, 3, 5, 2)	0.9683	0.7816	0,9714	0.8662

Hasil evaluasi model *Random Forest* pada Skenario B (Tanpa Fitur Spasial) ditunjukkan pada Tabel 4.8 dengan variasi rasio pembagian data dan kombinasi *hyperparameter* terbaik pada setiap pengujian. Pengujian dilakukan untuk menganalisis pengaruh penggunaan fitur CG+, CG- , dan IC tanpa penambahan fitur spasial terhadap performa model dalam melakukan klasifikasi sebagai dasar *forecasting* cuaca ekstrem.

Berdasarkan Tabel 4.8, pada Skenario 4B dengan rasio pembagian data 90:10, model mampu mempertahankan kinerja klasifikasi yang sangat baik. Hasil evaluasi menunjukkan nilai *accuracy* sebesar 98.19%, *precision* sebesar 85.18%, *recall* sebesar 100.00%, dan *F1-score* sebesar 92.00%. Nilai *recall* yang mencapai 100.00% mengindikasikan bahwa seluruh kejadian cuaca ekstrem pada data uji berhasil dikenali tanpa adanya *false negative*. Selain itu, tingginya nilai *precision* menunjukkan bahwa sebagian besar data yang diprediksi sebagai cuaca ekstrem

sesuai dengan kondisi sebenarnya.

Pada Skenario 5B dengan rasio pembagian data 80:20, model memperoleh nilai *accuracy* sebesar 97.51%, *precision* sebesar 81.03%, *recall* sebesar 100.00%, dan *F1-score* sebesar 89.52%. Dibandingkan dengan skenario sebelumnya, terjadi sedikit penurunan pada nilai *accuracy*, *precision*, dan *F1-score*. Meskipun demikian, kemampuan model dalam mengenali seluruh kejadian cuaca ekstrem tetap terjaga, sebagaimana ditunjukkan oleh nilai *recall* yang tetap mencapai 100.00%. Kondisi tersebut mengindikasikan bahwa berkurangnya jumlah data latih menyebabkan kemampuan model dalam mempelajari variasi pola aktivitas petir menjadi sedikit berkurang.

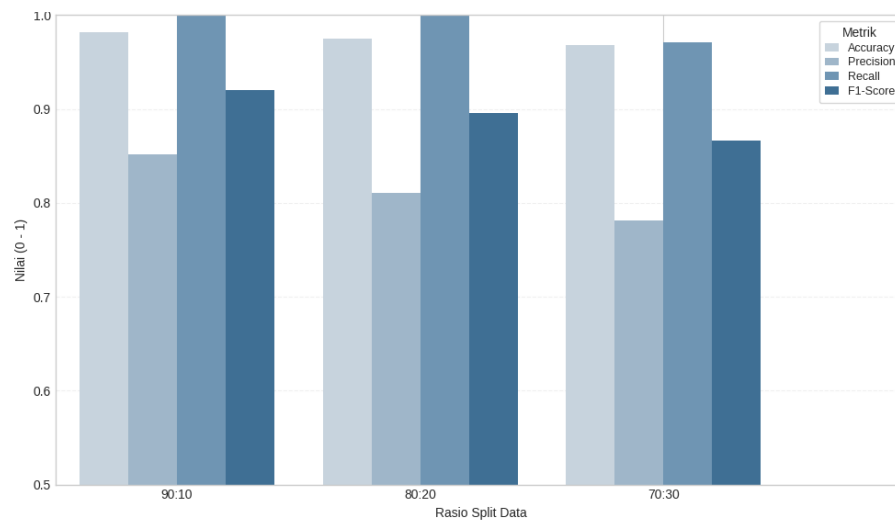
Sementara itu, pada Skenario 6B dengan rasio pembagian data 70:30, model memperoleh nilai *accuracy* sebesar 96.83%, *precision* sebesar 78.16%, *recall* sebesar 97.14%, dan *F1-score* sebesar 86.62%. Hasil tersebut menunjukkan adanya penurunan performa dibandingkan dua skenario sebelumnya. Berkurangnya proporsi data latih menyebabkan model memiliki informasi yang lebih sedikit untuk mempelajari karakteristik aktivitas petir, sehingga kemampuan klasifikasi menjadi menurun. Dampaknya, jumlah kesalahan prediksi berupa *false positive* maupun *false negative* meningkat dibandingkan pada skenario dengan rasio data latih yang lebih besar.

Hasil pengujian menunjukkan bahwa fitur aktivitas petir berupa CG+, CG-, dan IC telah mampu memberikan informasi yang kuat untuk mengenali pola kejadian cuaca ekstrem. Hal ini dapat dipahami karena aktivitas petir berkaitan erat dengan proses konveksi di atmosfer yang menjadi penyebab terbentuknya awan

cumulonimbus, hujan lebat, dan cuaca ekstrem. Oleh karena itu, meskipun tanpa informasi spasial, model masih mampu menghasilkan nilai *accuracy* di atas 96% dan nilai *recall* yang sangat tinggi pada seluruh skenario.

Namun demikian, performa model pada Skenario B tetap lebih rendah dibandingkan Skenario A yang menggunakan fitur spasial. Hal ini menunjukkan bahwa informasi aktivitas petir saja belum sepenuhnya mampu menggambarkan variasi karakteristik cuaca ekstrem pada setiap lokasi. Tanpa informasi *latitude* dan *longitude*, model hanya memanfaatkan pola aktivitas petir tanpa mempertimbangkan perbedaan kondisi geografis antarwilayah. Akibatnya, model memiliki informasi yang lebih terbatas dalam membedakan pola kejadian cuaca ekstrem yang dipengaruhi oleh faktor lokasi, sehingga nilai *accuracy*, *precision*, dan *F1-score* menjadi sedikit lebih rendah dibandingkan model yang menggunakan seluruh fitur.

Secara keseluruhan, Skenario 4B dengan rasio pembagian data 90:10 merupakan konfigurasi terbaik pada kelompok tanpa fitur spasial karena menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* tertinggi, yaitu masing-masing sebesar 98.19%, 85.18%, 100.00%, dan 92.00%. Hasil tersebut menunjukkan bahwa aktivitas petir (CG+, CG-, dan IC) merupakan faktor utama yang mampu membentuk model *forecasting* cuaca ekstrem, sedangkan informasi spasial berfungsi sebagai fitur pendukung yang mendukung model dalam mengidentifikasi karakteristik cuaca yang berbeda.



Gambar 4.8 Metrik Evaluasi Skenario B (Tanpa Fitur Spasial)

Berdasarkan grafik pada Gambar 4.8 metrik evaluasi Skenario 4B, Skenario 5B, dan Skenario 6B menunjukkan performa yang baik pada seluruh rasio pembagian data. Nilai *accuracy*, *precision*, *recall*, dan *F1-score* tetap berada pada kategori tinggi, yang menunjukkan bahwa model mampu mengklasifikasikan kondisi cuaca sebagai dasar *forecasting* cuaca ekstrem meskipun hanya menggunakan fitur aktivitas petir.

Skenario 4B menghasilkan performa terbaik karena menggunakan proporsi data latih yang paling besar, sehingga model memperoleh lebih banyak informasi untuk mempelajari hubungan antara aktivitas petir dan kejadian cuaca ekstrem. Pada Skenario 5B dan Skenario 6B terlihat adanya penurunan bertahap pada nilai *accuracy*, *precision*, dan *F1-score* seiring bertambahnya proporsi data uji. Hal tersebut menunjukkan bahwa berkurangnya proporsi data latih memengaruhi kemampuan model dalam prose mempelajari pola cuaca ekstrem secara optimal.

Meskipun demikian, tingginya nilai *recall* pada seluruh skenario

menunjukkan bahwa aktivitas petir berupa CG+, CG- , dan IC telah mampu menjadi indikator yang kuat dalam mengenali kejadian cuaca ekstrem. Namun, dibandingkan dengan Skenario A, performa model pada Skenario B masih sedikit lebih rendah. Hasil tersebut mengindikasikan bahwa penambahan informasi spasial membantu model dalam mempelajari pola kejadian cuaca ekstrem secara lebih efektif dalam mengenali perbedaan karakteristik cuaca ekstrem pada setiap lokasi. Dengan demikian, fitur aktivitas petir berperan sebagai sumber informasi utama, sedangkan fitur spasial berfungsi meningkatkan ketepatan model dalam melakukan *forecasting* cuaca ekstrem.

4.3 Perbandingan Performa Model Berdasarkan Matrik Evaluasi

Analisis ini dilakukan untuk membandingkan performa model pada setiap skenario dan rasio pembagian data berdasarkan nilai *accuracy*, *precision*, *recall* dan *F1-score*. Melalui metrik evaluasi tersebut, dapat diketahui skenario model yang memberikan hasil *forecasting* paling optimal.

4.3.1 *Accuracy*

Accuracy digunakan untuk menilai tingkat keberhasilan model berdasarkan jumlah prediksi yang benar dibandingkan dengan keseluruhan data yang dievaluasi. Gambar 4.9 memperlihatkan perbandingan nilai *accuracy* pada skenario Semua Fitur (Skenario 1A, 2A, dan 3A) dan kelompok skenario Tanpa Fitur Spasial (Skenario 4B, 5B, dan 6B) dengan tiga rasio pembagian data yang berbeda.

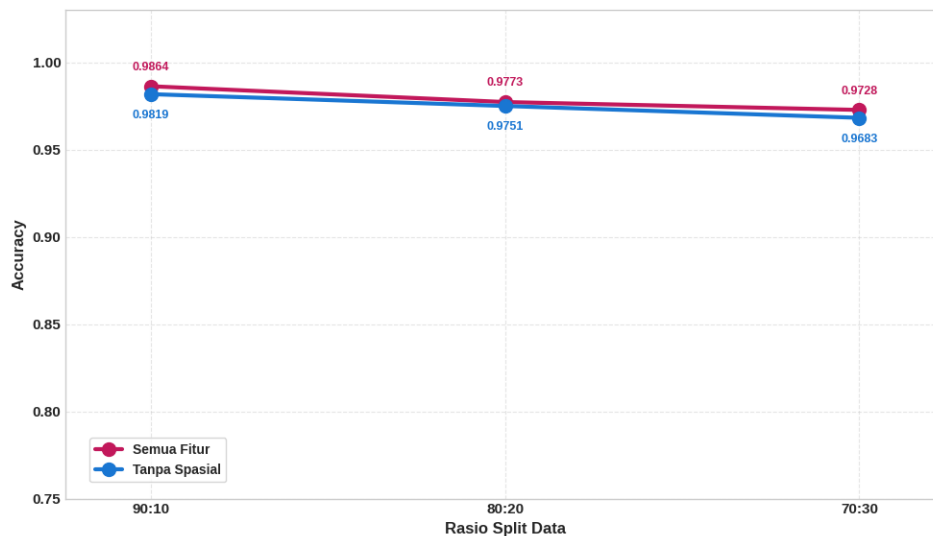
Pada kelompok skenario Semua Fitur, nilai *accuracy* tertinggi diperoleh pada Skenario 1A sebesar 98.64%. Nilai tersebut kemudian mengalami penurunan

pada Skenario 2A dan Skenario 3A. Meskipun demikian, seluruh skenario pada kelompok ini tetap menunjukkan tingkat *accuracy* yang sangat tinggi. Hasil tersebut menunjukkan bahwa penggunaan semua fitur, termasuk *latitude* dan *longitude*, membantu model dalam mengenali pola cuaca ekstrem dengan lebih baik.

Pada kelompok skenario Tanpa Fitur Spasial, nilai *accuracy* tertinggi diperoleh pada Skenario 4B sebesar 98.19%. Nilai *accuracy* kemudian cenderung menurun pada Skenario 5B dan Skenario 6B seiring meningkatnya proporsi data uji. Meskipun demikian, nilai *accuracy* pada seluruh skenario tetap berada di atas 96%. Hal ini mengindikasikan bahwa model tetap mampu menghasilkan *forecasting* cuaca ekstrem yang baik meskipun hanya memanfaatkan fitur CG+, CG-, dan IC.

Nilai *accuracy* yang tinggi pada seluruh skenario menunjukkan bahwa pada skenario Semua Fitur, penambahan fitur spasial berupa *latitude* dan *longitude* memberikan informasi lokasi yang membantu model *forecasting* data dengan lebih tepat. Sementara itu, pada skenario Tanpa Spasial, fitur CG+, CG-, dan IC tetap mampu memberikan informasi yang relevan dalam membedakan kondisi cuaca ekstrem dan tidak ekstrem, meskipun nilai *accuracy* yang dihasilkan lebih rendah dibandingkan skenario Semua Fitur.

Secara keseluruhan, nilai *accuracy* tertinggi diperoleh pada Skenario 1A (Semua Fitur) sebesar 98.64%, sedangkan nilai *accuracy* terendah diperoleh pada Skenario 6B (Tanpa Fitur Spasial) sebesar 96.83%. Berdasarkan hasil evaluasi, algoritma *Random Forest* mampu menunjukkan kemampuan yang tinggi dalam melakukan klasifikasi cuaca ekstrem.. Selain itu, penggunaan fitur spasial memberikan kontribusi positif terhadap peningkatan performa model.



Gambar 4.9 Grafik Perbandingan *Accuracy* pada tiap Skenario

4.3.2 *Precision*

Precision merupakan ukuran kemampuan model dalam menghasilkan *forecasting* cuaca ekstrem yang tepat. Nilai *precision* merepresentasikan kemampuan model dalam menghasilkan prediksi cuaca ekstrem yang benar. Gambar 4.10 memperlihatkan perbandingan nilai *precision* pada kelompok skenario Semua Fitur (Skenario 1A, 2A, dan 3A) dan kelompok skenario Tanpa Fitur Spasial (Skenario 4B, 5B, dan 6B) dengan tiga rasio pembagian data yang berbeda.

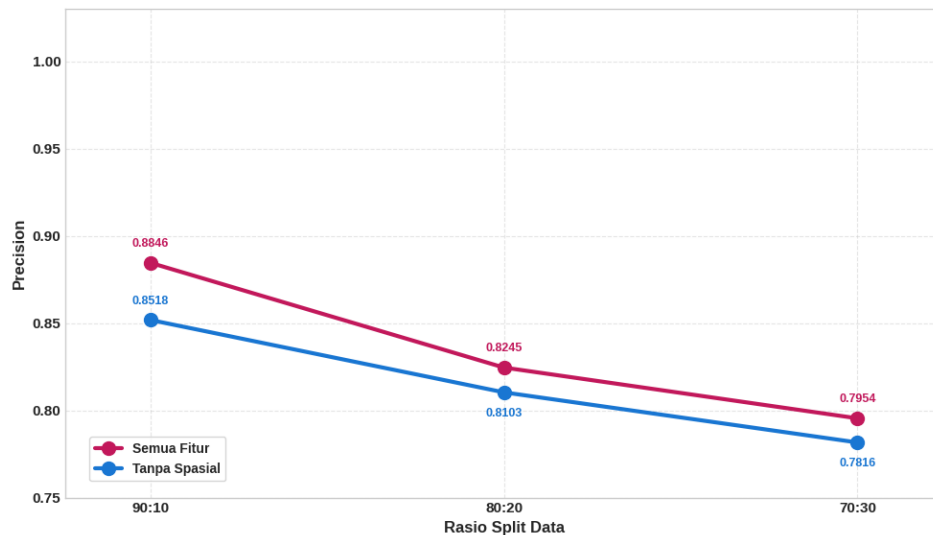
Pada kelompok skenario Semua Fitur, nilai *precision* tertinggi diperoleh pada Skenario 1A sebesar 88.46%. Nilai tersebut kemudian mengalami penurunan pada Skenario 2A dan Skenario 3A. Meskipun demikian, seluruh skenario pada kelompok ini tetap menunjukkan kemampuan yang baik dalam menghasilkan *forecasting* cuaca ekstrem yang tepat. Hasil tersebut menunjukkan bahwa penggunaan semua fitur, termasuk *latitude* dan *longitude*, membantu model dalam mengenali pola cuaca ekstrem dengan lebih baik.

Pada kelompok skenario Tanpa Fitur Spasial, nilai *precision* tertinggi diperoleh pada Skenario 4B sebesar 85.18%. Nilai *precision* kemudian cenderung menurun pada Skenario 5B dan Skenario 6B seiring meningkatnya proporsi data uji. Walaupun mengalami penurunan, nilai *precision* pada seluruh skenario tetap berada di atas 78%, sehingga menunjukkan bahwa model masih mampu menghasilkan peramal cuaca ekstrem yang cukup tepat hanya menggunakan fitur CG+, CG-, dan IC.

Nilai *precision* yang lebih tinggi pada skenario Semua Fitur menunjukkan bahwa penambahan fitur spasial berupa *latitude* dan *longitude* mampu meningkatkan ketepatan model dalam *forecasting* kejadian cuaca ekstrem. Informasi lokasi membantu model mengenali pola kejadian cuaca ekstrem dengan lebih baik sehingga *forecasting* yang dihasilkan lebih sering sesuai dengan kondisi sebenarnya. Sementara itu, pada skenario Tanpa Spasial, model tetap mampu menghasilkan nilai *precision* yang tinggi, meskipun lebih rendah dibandingkan skenario Semua Fitur. Hasil ini menunjukkan bahwa penggunaan fitur spasial mampu meningkatkan kemampuan model dalam mengidentifikasi kejadian cuaca ekstrem secara lebih tepat.

Secara keseluruhan, nilai *precision* tertinggi diperoleh pada Skenario 1A (Semua Fitur) sebesar 88.46%, sedangkan nilai *precision* terendah diperoleh pada Skenario 6B (Tanpa Fitur Spasial) sebesar 78.16%. Berdasarkan hasil pengujian pada seluruh skenario, algoritma *Random Forest* mampu mempertahankan performa klasifikasi yang tinggi, sebagaimana ditunjukkan oleh nilai metrik evaluasi yang konsisten pada setiap pengujian. Selain itu, penggunaan fitur spasial

memberikan kontribusi positif terhadap peningkatan performa model.



Gambar 4.10 Grafik Perbandingan *Precision* pada tiap Skenario

4.3.3 Recall

Recall merupakan metrik yang menunjukkan tingkat keberhasilan model dalam mengenali seluruh data yang benar-benar termasuk kategori cuaca ekstrem. Nilai *recall* menunjukkan seberapa banyak kejadian cuaca ekstrem yang berhasil diidentifikasi oleh model dari seluruh kejadian cuaca ekstrem yang terdapat dalam data. Gambar 4.11 memperlihatkan perbandingan nilai *recall* pada kelompok skenario Semua Fitur (Skenario 1A, 2A, dan 3A) dan kelompok skenario Tanpa Fitur Spasial (Skenario 4B, 5B, dan 6B) dengan tiga rasio pembagian data yang berbeda.

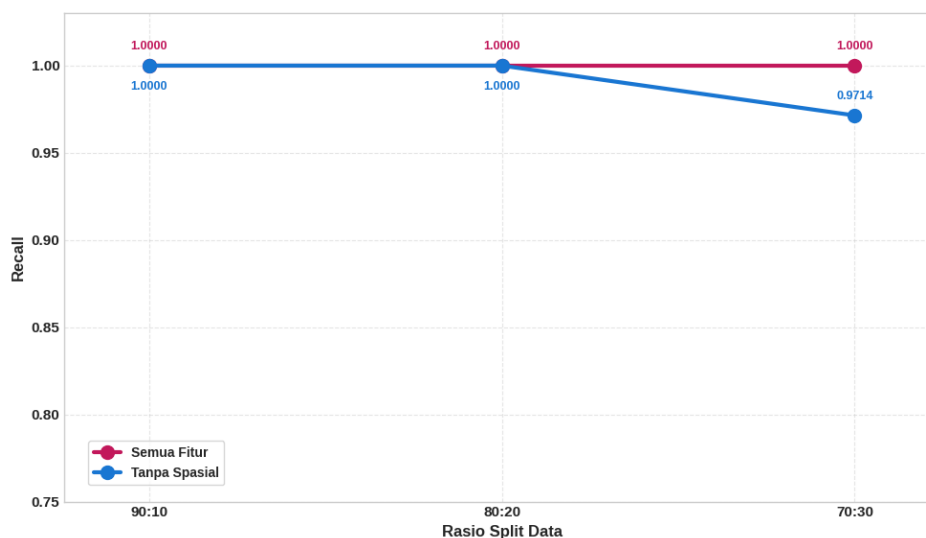
Pada kelompok skenario Semua Fitur, nilai *recall* tertinggi diperoleh pada Skenario 1A, 2A, dan 3A dengan nilai yang sama, yaitu 100.00%. Hasil tersebut menunjukkan kemampuan model dalam mengenali semua data yang termasuk kategori cuaca ekstrem pada seluruh rasio pembagian data. Kondisi ini mengindikasikan bahwa penggunaan semua fitur, termasuk *latitude* dan *longitude*,

membantu model mengenali pola cuaca ekstrem secara optimal.

Pada kelompok skenario Tanpa Fitur Spasial, nilai *recall* tertinggi diperoleh pada Skenario 4B dan Skenario 5B sebesar 100.00%. Nilai tersebut kemudian mengalami sedikit penurunan pada Skenario 6B menjadi 97.14%. Meskipun demikian, nilai *recall* pada seluruh skenario tetap berada di atas 97%. Sehingga menunjukkan bahwa model masih mampu mendeteksi sebagian besar kejadian cuaca ekstrem hanya dengan menggunakan fitur CG+, CG-, dan IC.

Nilai *recall* yang tinggi pada kedua skenario menunjukkan bahwa fitur CG+, CG-, dan IC mampu membantu model mengenali kejadian cuaca ekstrem dengan baik. Pada skenario Semua Fitur, penambahan fitur spasial berupa *latitude* dan *longitude* membantu model mendeteksi seluruh kejadian cuaca ekstrem yang terdapat dalam data. Sementara itu, pada skenario Tanpa Spasial, model tetap menunjukkan kemampuan deteksi yang sangat baik, meskipun terjadi sedikit penurunan nilai *recall* pada skenario 6B rasio 70:30.

Secara keseluruhan, nilai *recall* tertinggi diperoleh pada Skenario 1A, 2A, dan 3A (Semua Fitur) sebesar 100.00%, sedangkan nilai *recall* terendah diperoleh pada Skenario 6B (Tanpa Fitur Spasial) sebesar 97.14%. Performa model *Random Forest* pada skenario pengujian ini menunjukkan kemampuan yang baik dalam mengenali pola kejadian cuaca ekstrem. Selain itu, penggunaan fitur spasial memberikan kontribusi positif terhadap peningkatan kemampuan deteksi model.



Gambar 4.11 Grafik Perbandingan *Recall* pada tiap Skenario

4.3.4 *F1-Score*

F1-score digunakan sebagai metrik evaluasi untuk menilai performa model dengan mengombinasikan nilai *precision* dan *recall* dalam satu ukuran. Melalui metrik ini, dapat diketahui sejauh mana model mampu menjaga keseimbangan antara ketepatan prediksi dan keberhasilannya dalam mengidentifikasi kejadian cuaca ekstrem. Semakin tinggi nilai *F1-score* yang diperoleh, semakin baik kemampuan model dalam menghasilkan klasifikasi yang akurat sekaligus mengenali sebagian besar kejadian cuaca ekstrem. Gambar 4.12 memperlihatkan perbandingan nilai *F1-Score* pada kelompok skenario Semua Fitur (Skenario 1A, 2A, dan 3A) dan kelompok skenario Tanpa Fitur Spasial (Skenario 4B, 5B, dan 6B) dengan tiga rasio pembagian data yang berbeda.

Pada kelompok skenario Semua Fitur, nilai *F1-Score* tertinggi diperoleh pada Skenario 1A sebesar 93.87%. Nilai tersebut kemudian mengalami penurunan pada Skenario 2A dan Skenario 3A. Meskipun demikian, seluruh skenario pada kelompok ini tetap menunjukkan performa yang baik. Hasil tersebut menunjukkan

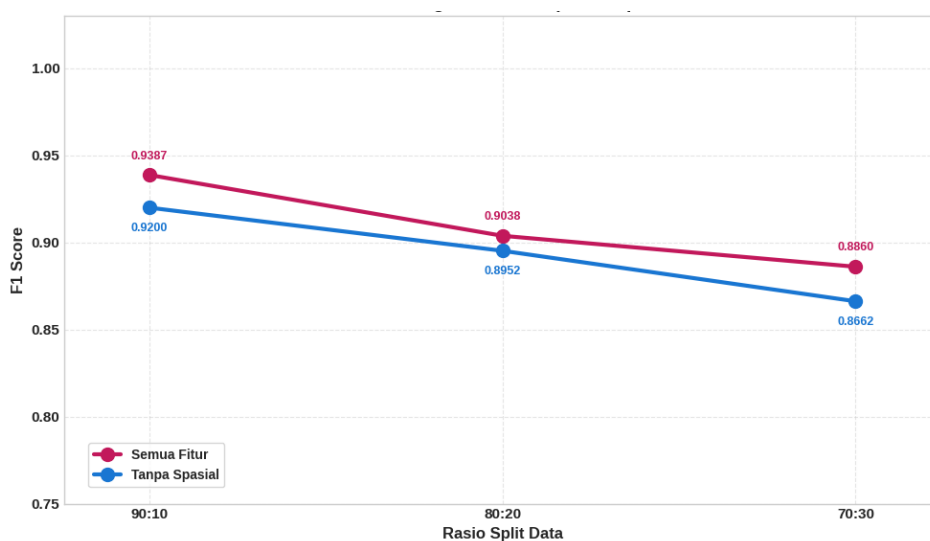
bahwa penggunaan semua fitur, termasuk *latitude* dan *longitude*, membantu model dalam mengenali pola cuaca ekstrem dengan lebih baik sehingga menghasilkan keseimbangan yang baik antara ketepatan dan kemampuan deteksi.

Pada kelompok skenario Tanpa Fitur Spasial, nilai *F1-Score* tertinggi diperoleh pada Skenario 4B sebesar 92.00%. Nilai *F1-Score* kemudian cenderung menurun pada Skenario 5B dan Skenario 6B seiring meningkatnya proporsi data uji. Walaupun mengalami penurunan, nilai *F1-Score* pada seluruh skenario tetap berada di atas 86%, yang menunjukkan bahwa model masih mampu mempertahankan keseimbangan yang baik antara ketepatan *forecasting* dan kemampuan mendeteksi kejadian cuaca ekstrem hanya dengan menggunakan fitur CG+, CG-, dan IC.

Nilai *F1-score* yang relatif tinggi pada semua skenario menunjukkan bahwa model mampu menjaga keseimbangan antara kemampuan mengenali kejadian cuaca ekstrem dan ketepatan prediksinya, sehingga performa klasifikasi yang dihasilkan tetap optimal. Hal ini menunjukkan bahwa fitur CG+, CG-, dan IC mampu membantu model mengenali pola cuaca ekstrem dengan baik. Selain itu, penambahan fitur spasial berupa *latitude* dan *longitude* memberikan informasi lokasi yang memperkuat kemampuan model, sehingga nilai *F1-Score* pada skenario Semua Fitur lebih tinggi dibandingkan skenario Tanpa Spasial.

Secara keseluruhan, nilai *F1-Score* tertinggi diperoleh pada Skenario 1A (Semua Fitur) sebesar 93.87%, sedangkan nilai *F1-Score* terendah diperoleh pada Skenario 6B (Tanpa Fitur Spasial) sebesar 86.62%. Dari hasil tersebut model *Random Forest* memiliki kemampuan yang baik dalam menghasilkan *forecasting*

cuaca ekstrem yang tepat sekaligus mendeteksi kejadian cuaca ekstrem dengan baik. Selain itu, penggunaan fitur spasial memberikan kontribusi positif terhadap peningkatan performa model secara keseluruhan.



Gambar 4.12 Grafik Perbandingan *F1-Score* pada tiap Skenario

4.4 Pembahasan Hasil Keseluruhan

Berdasarkan hasil pengujian menunjukkan bahwa variasi penggunaan fitur dan rasio pembagian data memberikan pengaruh terhadap performa algoritma *Random Forest* dalam melakukan klasifikasi kondisi cuaca sebagai dasar *forecasting* potensi cuaca ekstrem. Pengujian dilakukan pada enam skenario, yaitu Skenario 1A, 2A, dan 3A yang menggunakan seluruh fitur, serta Skenario 4B, 5B, dan 6B yang menggunakan fitur tanpa informasi spasial. Perbedaan penggunaan fitur dan pembagian rasio data latih dan data uji menghasilkan performa model yang berbeda pada setiap skenario pengujian.

Berdasarkan hasil pengujian pada kelompok skenario yang menggunakan seluruh fitur (Skenario 1A, 2A, dan 3A), model menghasilkan kinerja yang konsisten dan sangat baik pada setiap metrik evaluasi. Nilai *accuracy*, *precision*,

dan *F1-score* mengalami sedikit penurunan ketika proporsi data uji semakin besar, sedangkan nilai *recall* tetap mencapai 100% pada seluruh skenario. Berdasarkan hasil evaluasi, seluruh kejadian cuaca ekstrem pada data uji dapat dikenali dengan benar oleh model. Selain itu, penggunaan jumlah data latih yang lebih banyak memberikan informasi yang lebih lengkap selama proses pembelajaran, sehingga algoritma dapat mengenali hubungan antara aktivitas petir dan kejadian cuaca ekstrem secara lebih optimal. Kondisi tersebut berdampak pada meningkatnya kemampuan model dalam menghasilkan prediksi yang akurat.

Pada kelompok skenario Tanpa Fitur Spasial (Skenario 4B, 5B, dan 6B), model juga menghasilkan performa yang baik. Pola yang sama terlihat pada nilai *accuracy*, *precision*, dan *F1-score* yang sedikit menurun ketika proporsi data uji meningkat, sedangkan nilai *recall* tetap berada pada kategori sangat tinggi meskipun mengalami sedikit penurunan pada Skenario 6B. Hasil tersebut menunjukkan bahwa fitur aktivitas petir berupa CG+, CG-, dan IC telah mampu memberikan informasi yang kuat untuk membedakan kondisi cuaca ekstrem dan tidak ekstrem meskipun tanpa memanfaatkan informasi spasial.

Berdasarkan perbandingan kedua kelompok skenario, penggunaan fitur spasial berupa *latitude* dan *longitude* menghasilkan performa yang lebih baik dibandingkan model tanpa fitur spasial dengan performa model yang konsisten pada seluruh rasio pembagian data. Peningkatan tersebut menunjukkan bahwa lokasi kejadian petir turut memengaruhi karakteristik cuaca ekstrem. Hal ini disebabkan setiap wilayah memiliki kondisi geografis dan atmosfer yang berbeda, seperti topografi, ketinggian wilayah, kedekatan dengan laut, serta karakteristik

pembentukan awan konvektif. Informasi spasial tersebut membantu model mengenali pola distribusi aktivitas petir yang berbeda pada setiap lokasi sehingga kemampuan model dalam membedakan kejadian cuaca ekstrem menjadi lebih baik. Temuan ini sejalan dengan konsep bahwa fenomena cuaca merupakan fenomena spasial yang dipengaruhi oleh kondisi geografis suatu wilayah.

Secara keseluruhan, performa terbaik diperoleh pada Skenario 1A dengan rasio pembagian data 90:10, yaitu *accuracy* sebesar 98.64%, *precision* 88.46%, *recall* 100.00%, dan *F1-score* 93.87%. Hasil tersebut menunjukkan bahwa kombinasi informasi aktivitas petir dan fitur spasial mampu meningkatkan kemampuan model dalam mengklasifikasikan kondisi cuaca sebagai dasar *forecasting* potensi cuaca ekstrem. Dengan demikian, aktivitas petir (CG+, CG-, dan IC) berperan sebagai sumber informasi utama dalam pembentukan model, sedangkan fitur spasial (*latitude* dan *longitude*) berfungsi sebagai informasi pendukung yang dapat mengenali pola kejadian cuaca ekstrem pada lokasi berbeda sehingga memperkuat kemampuan model. Oleh karena itu, model *Random Forest* yang dibangun memiliki potensi untuk dimanfaatkan sebagai pendukung sistem *forecasting* dini cuaca ekstrem berbasis data aktivitas petir.

4.5 Integrasi Islam

Penelitian mengenai *forecasting* cuaca ekstrem menggunakan algoritma *Random Forest* memiliki keterkaitan dengan nilai-nilai Islam, khususnya dalam pemanfaatan ilmu pengetahuan dan teknologi untuk memahami fenomena alam serta mengurangi risiko bencana. Berbagai peristiwa yang terjadi di alam semesta dipandang dalam Islam sebagai ayat-ayat kauniyah yang mencerminkan kekuasaan

Allah SWT. Oleh karena itu, manusia dianjurkan untuk mengamati, memahami, dan mengambil hikmah dari setiap fenomena alam yang terjadi.

Hal tersebut sebagaimana dijelaskan dalam Al-Qur'an Surah Ar-Rum ayat 48:

اللَّهُ الَّذِي يُرْسِلُ الرِّيحَ فَتُثِيرُ سَحَابًا فَيَبْسُطُهُ فِي السَّمَاءِ كَيْفَ يَشَاءُ وَيَجْعَلُهُ كِسْفًا فَيَزِي الْوَدْقَ يَخْرُجُ مِنْ خِلَالِهِ ۚ فَإِذَا أَصَابَ بِهِ ۚ مِنْ يَشَاءُ مِنْ عِبَادِهِ ۚ إِذَا هُمْ يَسْتَبْشِرُونَ

“Allahlah yang mengirim angin, lalu ia (angin) menggerakkan awan, kemudian Dia (Allah) membentangkannya di langit menurut yang dikehendaki-Nya dan Dia menjadikannya bergumpal-gumpal, lalu engkau melihat hujan keluar dari celah-celahnya. Maka, apabila Dia menurunkannya kepada hamba-hamba-Nya yang dikehendaki-Nya, seketika itu pula mereka bergembira”. (QS Ar-Rum :48)

Ayat tersebut menjelaskan bahwa berbagai fenomena cuaca seperti angin, awan, hujan, dan perubahan kondisi atmosfer merupakan bagian dari kekuasaan Allah SWT. Fenomena alam tersebut dapat dipelajari oleh manusia melalui ilmu pengetahuan agar manusia mampu memahami pola kejadian alam dan mengambil langkah antisipasi terhadap kemungkinan bencana yang dapat terjadi.

Dalam Tafsir Ibnu Katsir dijelaskan bahwa Allah SWT menunjukkan tanda-tanda kekuasaan-Nya melalui proses terjadinya hujan dan perubahan cuaca sebagai pelajaran bagi manusia untuk berpikir dan mengambil hikmah dari setiap fenomena alam. Oleh karena itu, penelitian terkait *forecasting* cuaca ekstrem dapat menjadi salah satu bentuk ikhtiar manusia dalam memahami pola alam guna meminimalkan risiko yang ditimbulkan oleh cuaca ekstrem.

Penelitian ini menerapkan algoritma *Random Forest* sebagai metode pemodelan untuk *forecasting* cuaca ekstrem dengan memanfaatkan data petir dengan fitur spasial. Model dibangun berdasarkan pola yang dipelajari dari data

historis sehingga dapat digunakan untuk memperkirakan kemungkinan terjadinya cuaca ekstrem. Dengan adanya sistem *forecasting* tersebut, informasi cuaca dapat digunakan sebagai bentuk mitigasi dini sehingga masyarakat maupun instansi terkait dapat lebih siap menghadapi potensi bencana hidrometeorologi.

Selain itu, Islam juga mendorong manusia untuk menggunakan akal dan ilmu pengetahuan sebagaimana dijelaskan dalam QS. Al-Jatsiyah ayat 13:

وَسَخَّرَ لَكُم مَّا فِي السَّمُوتِ وَمَا فِي الْأَرْضِ جَمِيعًا مِّنْهُ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِّقَوْمٍ يَتَفَكَّرُونَ

“Dia telah menundukkan (pula) untukmu apa yang ada di langit dan apa yang ada di bumi semuanya (sebagai rahmat) dari-Nya. Sesungguhnya pada yang demikian itu ben’ar-benar terdapat tanda-tanda (kebesaran Allah) bagi kaum yang berpikir”. (QS Al- :jatsiyah 13).

Ayat tersebut menegaskan bahwa manusia diperintahkan untuk memanfaatkan ilmu pengetahuan dalam memahami fenomena alam sebagai bentuk rasa syukur kepada Allah SWT. Dalam konteks penelitian ini, penggunaan data petir, fitur spasial, dan algoritma *Random Forest* merupakan bentuk pemanfaatan ilmu pengetahuan dan teknologi untuk memberikan manfaat bagi kehidupan manusia, khususnya dalam meningkatkan kesiapsiagaan terhadap cuaca ekstrem.

Dengan demikian, penelitian ini tidak hanya memberikan manfaat akademis dalam pengembangan metode *forecasting* cuaca ekstrem berbasis algoritma *Random Forest*, tetapi juga mencerminkan nilai-nilai Islam yang menekankan pentingnya pengamatan terhadap alam sebagai bentuk ikhtiar untuk memperoleh pengetahuan sekaligus mewujudkan keselamatan dan kemaslahatan bagi masyarakat.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai *forecasting* cuaca ekstrem menggunakan algoritma *Random Forest* berbasis data petir, dapat disimpulkan bahwa model yang dibangun mampu melakukan *forecasting* potensi cuaca ekstrem dengan performa yang sangat baik. Hasil pengujian menunjukkan nilai *accuracy* sebesar 96.83%–98.64% dan *recall* sebesar 97.14%–100.00%, sehingga menunjukkan bahwa algoritma *Random Forest* efektif digunakan untuk *forecasting* cuaca ekstrem berdasarkan data aktivitas petir. Selain itu, penambahan fitur spasial berupa *latitude* dan *longitude* terbukti meningkatkan performa model pada seluruh skenario pengujian dibandingkan model tanpa fitur spasial. Performa terbaik diperoleh pada Skenario 1A dengan rasio pembagian data 90:10, yang menghasilkan *accuracy* sebesar 98.64%, *precision* sebesar 88.46%, *recall* sebesar 100.00%, dan *F1-score* sebesar 93.87%. Hasil tersebut menunjukkan bahwa informasi spasial berkontribusi dalam meningkatkan kemampuan model dalam melakukan *forecasting* cuaca ekstrem.

5.2 Saran

Penelitian ini masih memiliki beberapa keterbatasan yang dapat menjadi bahan evaluasi untuk pengembangan penelitian selanjutnya. Oleh karena itu, beberapa saran berikut diharapkan dapat menjadi acuan bagi penelitian berikutnya:

1. Menggunakan data aktivitas petir dengan rentang waktu pengamatan yang lebih panjang dan jumlah data yang lebih banyak agar model dapat mempelajari pola historis secara lebih komprehensif sehingga meningkatkan kemampuan dalam melakukan prediksi cuaca ekstrem.
2. Menerapkan dan membandingkan algoritma *machine learning* lain, seperti *Decision Tree*, *Support Vector Machine (SVM)*, *XGBoost*, atau *Neural Network*, untuk mengetahui algoritma yang memberikan performa terbaik dalam *forecasting* cuaca ekstrem berbasis data aktivitas petir.
3. Mengembangkan fitur spasial menggunakan metode transformasi data atau pengelompokan wilayah tertentu agar informasi lokasi dapat dimanfaatkan secara lebih efektif.

DAFTAR PUSTAKA

- Adrianti, L. (2024). ANALISIS INDEKS IKLIM EKSTREM UNTUK DETEKSI PERUBAHAN IKLIM DI BANTEN. *Buletin Meteorologi, Klimatologi, Dan Geofisika*, 5(4), 8–16. <https://bmkgsoft.bmkg.go.id>.
- Al-Sheikh, A. I. I. bin M. bin I. bin A. (2004). *Lubaabut Tafsir Min Ibni Katsiir* (pp. 484–487). Maktabah Darul Falah.
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi*, 8(2), 50–58. <https://doi.org/10.23960/komputasi.v8i2.2615>
- Amaliah, S. N., Muhammad, & Aswi. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variansiunm31>
- Amri, I. F., Chamidah, N., Saifudin, T., & Purwanto, D. (2024). *Prediction of extreme weather using nonparametric regression approach with Fourier series estimators Predicción de fenómenos meteorológicos extremos mediante un enfoque de regresión no paramétrica con estimadores de series de Fourier*. 1–12. <https://doi.org/10.56294/dm2024319>
- Angga, N., & Diansari, V. (2018). *Analisa Kerawanan Kejadian Petir dan Hubungan dengan Pola Curah Hujan di Wilayah Bengkulu Sepanjang Tahun 2017*. 9, 41–47.
- Arrijal et al., 2020. (2020). *Peramalan Produk Ragum Dengan Metode Causal dan Time Series TALENTA Conference Series Peramalan Produk Ragum Dengan Metode Causal dan Time Series*. 3(2). <https://doi.org/10.32734/ee.v3i2.996>
- Atika, P. Y., Alwin, A., & Sya'ban, M. B. A. (2024). Efektifitas Penerapan Sistem Peringatan Dini Dalam Upaya Pengurangan Risiko Bencana Banjir Di Kelurahan Kampung Melayu. *GEOGRAPHY: Jurnal Kajian, Penelitian Dan Pengembangan Pendidikan*, 12(2), 671. <https://doi.org/10.31764/geography.v12i2.24533>
- Caraka, R. E., Yasin, H., & Suparti. (2015). PEMODELAN TINGGI PASANG AIR LAUT DI KOTA SEMARANG DENGAN MENGGUNAKAN MAXIMAL OVERLAP DISCRETE WAVELET TRANSFORM (MODWT). *Meteorologi Klimatologi Dan Geofisika*, 2(2), 1–16.
- Dwiyanti, Z. A., & Prianto, C. (2023). *Prediksi Cuaca Kota Jakarta menggunakan Metode Random Forest : Studi Optimalitas*. 17(2), 127–137.
- Efendi, M. S., & Zyen, A. K. (2024). *Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk*. 5(1), 12–20. <https://doi.org/10.30865/resolusi.v5i1.2149>
- Fauzan, R., Iskandar, N., Gutama, D. H., Wijaya, D. P., & Danianti, D. (2024). *JUTIN: Jurnal Teknik Industri Terintegrasi Klasifikasi Menggunakan Metode Random Forest untuk Awal Deteksi Diabetes Melitus Tipe 2*. 7(3).
- Haerudin, Erick, & Aripadono, heru wijaya. (2025). *JTIM: Jurnal Teknologi Informasi dan Multimedia Perbandingan Support Vector Machine , Random Forest Classifier , dan K-Nearest Neighbour dalam Pendeteksian Anomali*. 7(1), 23–33.
- Hikmah, Asrirawan, Apriyanto, & Nilawati. (2023). *Peramalan Data Cuaca Ekstrem Indonesia Menggunakan Model ARIMA dan Recurrent Neural Network*. 5(1), 230–242.
- Ibrahim, M. W., Aji, W., & Prabowo, E. (2026). *Evaluasi Komparatif Random Forest , XGBoost , LightGBM , dan K-Nearest*. 13(1), 347–357. <https://doi.org/10.30865/jurikom.v13i1.9488>
- Inonu, O. Y., & Magda, K. (2025). *Analisis Kinerja Algoritma Random Forest Dengan Model Machine Learning Pada Dataset Penyakit Diabetes*. 15(1), 1–7.
- Intan, I., & Koswara, A. T. C. (2021). *Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning*. 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Joses, S., Yulvida, D., & Rochimah, S. (2024). *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier*. 5(1), 72–80.
- Kaplan, J. O., & Lau, K. H. (2022). *World Wide Lightning Location Network (WWLLN) Global Lightning Climatology (WGLC) and time series , 2022 update*. 5665–5670.

- Lepot, M., Aubin, J. B., & Clemens, F. H. L. R. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water (Switzerland)*, 9(10). <https://doi.org/10.3390/w9100796>
- Liu, X., Kounadi, O., & Zurita-Milla, R. (2022). Incorporating Spatial Autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features. *ISPRS International Journal of Geo-Information*, 11(4). <https://doi.org/10.3390/ijgi11040242>
- Mananohas, M. A., Bobanto, M. D., & Ferdy, F. (2019). Hubungan Cuaca dan Tanaman Pangan Menggunakan Regresi Linear di Kota Tondano. *D'CARTESIAN*, 8(2), 169. <https://doi.org/10.35799/dc.8.2.2019.24390>
- Mehranzamin, K., Pour, A. B., Abdul-Malek, Z., Afrouzi, H. N., Alizadeh, S. M., & Hashim, M. (2023). Implementation of Ground-Based Lightning Locating System Using Particle Swarm Optimization Algorithm for Lightning Mapping and Monitoring. *Remote Sensing*, 15(9). <https://doi.org/10.3390/rs15092306>
- Mujiasih, S. (2022). *PEMANFATAN DATA MINING UNTUK PRAKIRAAN CUACA. September 2011*, 189–195.
- Mukeshkumar, P., Balachandar, A., Mohanraj, K., & Vanitha, R. (2025). *EXTREME WEATHER PREDICTION AND ALERT SYSTEM WITH MACHINE LEARNING ABSTRACT*: (Issue Icaai). Atlantis Press International BV. <https://doi.org/10.2991/978-94-6463-738-0>
- Nugroho, M. W. (2025). Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 9(4), 1562–1571. <https://doi.org/10.35870/jtik.v9i4.4236>
- Patrick, M., Pattisina, K., Bayu, A., Saputra, A., Wahyu, F., Wahyono, T., Studi, P., Informatika, T., Informasi, F. T., Satya, U. K., Studi, P., Digital, B., Informasi, F. T., & Satya, U. K. (2025). *Penerapan Pembelajaran Mesin dalam Deteksi Cuaca : Sebuah Tinjauan Pustaka Sistematis Application of Machine Learning in Weather Detection : A Systematic Literature Review. 11*(April), 24–35. <https://doi.org/10.37715/juisi.v11i1.5694>
- Risanti, Indrasari, W., & Suhendar, H. (2024). Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree. *XII*, 119–128. <https://doi.org/10.21009/03.1201.fa18>
- Rose, M., & Hassen, H. R. (2019). *A Survey of Random Forest Pruning Techniques*. 99–109. <https://doi.org/10.5121/csit.2019.91808>
- Sangaji, D., & Sutabri, T. (2025). Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla. 5(1), 13–18.
- Siregar, A. P., Purba, D. P., Pasaribu, J. P., & Bakara, K. R. (2023). Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke. *Jurnal Penelitian Rumpun Ilmu Teknik*, 2(4), 155–164.
- Syahreza, A., Ningrum, N. K., & Syahrazy, M. A. (2024). Perbandingan Kinerja Model Prediksi Cuaca : Random Forest , Support Vector Regression , dan XGBoost. 8(2), 526–534. <https://doi.org/10.29408/edumatic.v8i2.27640>
- Tumangkeng, G. T., Warouw, V. R. C., & Mawara, J. M. (2021). Analisis Pengaruh Curah Hujan Terhadap Erosi Pada Tanah Tanpa Mulsa Dan Diberi Mulsa. *Cocos*, 3(3).
- Zhang, C., Xiao, C., Li, S., Ren, Y., Zhang, S., Cai, X., & Sangbu, Z. (2023). *Analysis of the Composite Risk Grade for Multi Extreme Climate Events in China in Recent 60 Years*.