

Implementasi Deteksi Anomali Menggunakan Isolation Forest Pada Data Curah Hujan Harian Untuk Evaluasi Ambang Hujan Ekstrem Bmkg Di Jawa Timur

Amor Maulidiyah Rizamzam Prasetya Putri¹, Mohammad Hafiyusholeh²,
Shanas Septy Prayuda³, Hani Khaulasari⁴

^{1,2,4}Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya,

³BMKG Stasiun Meteorologi Kelas 1 Juanda, Sidoarjo

¹amormaulidiyah01@gmail.com, ²hafiyusholeh@uinsa.ac.id, ³shanas.prayuda@bmkg.go.id,

⁴hanikhaulasari@gmail.com

ARTICLE INFO

Article History:

Submitted 16-07-2026

Received 19-07-2026

Published 24-07-2026

Keywords:

First Keyword;BMKG

Second keyword;Rainfall

Third Keyword;Extreme Rain

Fourth keyword;Isolation Forest

ABSTRACT

Extreme rainfall needs to be monitored because it can increase the risk of hydrometeorological disasters. The BMKG defines extreme rainfall as daily rainfall above 150 mm, although rainfall characteristics vary by region. This study applies the Isolation Forest algorithm to detect daily rainfall anomalies and compare them with BMKG thresholds. Data comes from Juanda, Kediri, and Malang stations, with a total of 1,006 observations. The model uses daily and monthly rainfall features with a contamination value of 0.05. As a result, the model detected 17 anomalies in Juanda, 17 in Kediri, and 18 in Malang. The model also identified one BMKG extreme rainfall event in Juanda (150.8 mm) as an anomaly. The evaluation yielded 100% recall for the extreme rainfall event, but this result should be interpreted with caution because there was only one event in the data. Overall, Isolation Forest has the potential to support the identification of local rainfall anomalies without replacing BMKG operational thresholds.

ABSTRAK

Curah hujan ekstrem perlu dipantau karena dapat meningkatkan risiko bencana hidrometeorologi. BMKG mendefinisikan curah hujan ekstrem sebagai curah hujan harian di atas 150 mm, meskipun karakteristik curah hujan bervariasi menurut wilayah. Studi ini menerapkan algoritma Isolation Forest untuk mendeteksi anomali curah hujan harian dan membandingkannya dengan ambang batas BMKG. Data berasal dari Stasiun Juanda, Kediri, dan Malang dengan total 1.006 observasi. Model menggunakan fitur curah hujan harian dan bulanan dengan nilai kontaminasi 0,05. Hasilnya, model mendeteksi 17 anomali di Juanda, 17 di Kediri, dan 18 di Malang. Model juga mengidentifikasi satu kejadian curah hujan ekstrem BMKG di Juanda (150,8 mm) sebagai anomali. Evaluasi menghasilkan recall 100% untuk kejadian curah hujan ekstrem, tetapi hasil ini harus ditafsirkan dengan hati-hati karena hanya ada satu kejadian dalam data. Secara keseluruhan, Isolation Forest berpotensi untuk mendukung identifikasi anomali curah hujan lokal tanpa menggantikan ambang batas operasional BMKG.

PENDAHULUAN

Curah hujan merupakan unsur cuaca yang secara langsung memengaruhi aktivitas masyarakat, lingkungan, transportasi, pertanian, dan potensi bencana hidrometeorologi. Curah hujan tinggi seringkali dikaitkan dengan risiko banjir, terutama di daerah yang rentan terhadap perubahan kondisi lingkungan dan intensitas curah hujan ekstrem (Risnawati, 2023). Perubahan iklim juga dapat meningkatkan kemungkinan terjadinya peristiwa cuaca ekstrem, termasuk curah hujan lebat, banjir, kekeringan, dan cuaca buruk (Nurlatifah, 2023). Jawa Timur dicirikan oleh curah hujan yang tidak merata karena topografi, lokasi geografis, dan dinamika atmosfer global dan regionalnya.

Variasi curah hujan di wilayah ini terkait dengan fenomena seperti ENSO, Mode Dipol, MJO, ITCZ, dan monsun (Bimaprawira, 2021). Perbedaan antara daerah pesisir, non-pesisir, dan dataran tinggi dapat memvariasikan pola curah hujan di setiap stasiun, sehingga analisis curah hujan harus mempertimbangkan karakteristik lokal. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menggunakan ambang batas curah hujan ekstrem sebagai referensi operasional, yaitu curah hujan harian yang melebihi 150 mm. Ambang batas ini sangat penting untuk menstandarisasi identifikasi kejadian curah hujan ekstrem. Namun, kejadian curah hujan yang tidak melebihi 150 mm masih dapat dianggap tidak biasa dibandingkan dengan pola historis di lokasi tertentu. Situasi ini menyoroti perlunya pendekatan berbasis data tambahan untuk menafsirkan anomali curah hujan lokal.

Deteksi anomali adalah pendekatan komputasi yang digunakan untuk mengidentifikasi data yang menyimpang dari pola mayoritas. Dalam data cuaca, anomali dapat terjadi pada nilai curah hujan yang tidak biasa dibandingkan dengan pola historis (Fatmi & Lase, 2025). Deteksi dini anomali curah hujan berguna sebagai dasar untuk mengidentifikasi kejadian curah hujan yang tidak biasa sebelum suatu wilayah mencapai kategori ekstrem nasional. Di wilayah tertentu, curah hujan yang belum melebihi 150 mm per hari masih dapat menimbulkan dampak jika terjadi di daerah dengan topografi khusus, drainase terbatas, atau kondisi tanah jenuh (Munanjar et al., 2025). Dengan deteksi anomali, hasil model dapat digunakan sebagai sinyal peringatan lokal yang melengkapi, bukan menggantikan, ambang batas ekstrem BMKG. Pendekatan ini penting karena sistem pemantauan tidak hanya membutuhkan batas nasional tetapi juga informasi tentang penyimpangan pola curah hujan di setiap wilayah.

Isolation Forest dipilih karena dapat mendeteksi anomali tanpa label awal dan menawarkan keunggulan dalam efisiensi komputasi dan stabilitas model (Fatmi & Lase, 2025). Algoritma ini juga telah digunakan untuk mengidentifikasi curah hujan dengan potensi banjir melalui pendekatan komputasi (Wijayanto et al., 2024). Relevansi Isolation Forest juga didukung oleh perkembangan metode berbasis isolasi dalam penelitian internasional. Isolation Forest telah menjadi metode deteksi anomali yang banyak digunakan karena efektivitasnya pada berbagai macam data dan skalabilitas komputasinya. Namun, metode ini masih memiliki keterbatasan ketika berurusan dengan data dengan pola kompleks atau keterpisahan non-linier, yang menyebabkan pengembangan pendekatan yang lebih canggih seperti Deep Isolation Forest (Xu et al., 2023). Hal ini menunjukkan bahwa metode berbasis Isolation Forest masih relevan dalam penelitian deteksi anomali modern. Meskipun demikian, penelitian ini masih menggunakan metode Isolation Forest dasar karena fokus penelitian adalah implementasi awal deteksi anomali curah hujan harian dan perbandingannya dengan ambang batas ekstrem BMKG.

Stasiun Juanda, Kediri, dan Malang dipilih karena mewakili karakteristik regional yang berbeda di Jawa Timur. Juanda mewakili daerah pesisir, Kediri mewakili daerah non-pesisir, dan Malang mewakili daerah dataran tinggi. Tujuan pemilihan ketiga stasiun ini adalah untuk menentukan apakah pola anomali curah hujan harian berbeda di daerah dengan karakteristik topografi yang berbeda. Pendekatan ini relevan karena studi curah hujan di Jawa Timur telah menunjukkan perbedaan karakteristik curah hujan antar wilayah (Bimaprawira, 2021). Penelitian sebelumnya telah berfokus pada prediksi curah hujan, peramalan, klasifikasi potensi banjir, dan deteksi anomali cuaca. Prediksi intensitas curah hujan di Juanda dilakukan menggunakan model berbasis pohon keputusan (Taufiqi et al., 2025), sedangkan deteksi anomali cuaca dipelajari melalui perbandingan Isolation Forest dan AutoEncoder (Fatmi & Lase, 2025). Namun, studi yang secara khusus menggunakan Isolation Forest untuk mengevaluasi kesesuaian ambang batas curah hujan ekstrem BMKG terhadap pola anomali lokal pada data curah hujan harian multi-stasiun di Jawa Timur masih terbatas.

Studi ini bertujuan untuk mengimplementasikan Isolation Forest pada data curah hujan harian dari stasiun Juanda, Kediri, dan Malang, kemudian membandingkan hasil deteksi anomali model dengan ambang batas curah hujan ekstrem BMKG. Studi ini tidak dimaksudkan untuk menggantikan standar BMKG, tetapi lebih memberikan perspektif komputasi untuk mengamati kejadian curah hujan abnormal lokal di daerah dengan karakteristik topografi yang berbeda.

TINJAUAN PUSTAKA

Curah Hujan Harian

Curah hujan harian adalah jumlah hujan yang tercatat dalam satu hari pengamatan, dinyatakan dalam milimeter. Data ini sangat penting untuk menafsirkan pola curah hujan, mengidentifikasi kejadian curah hujan lebat, dan mendukung mitigasi bencana hidrometeorologi. Informasi curah hujan yang cepat dan akurat sangat penting karena berkaitan dengan potensi banjir dan cuaca ekstrem (Risnawati, 2023). Karakteristik curah hujan di Jawa Timur bervariasi di berbagai wilayah karena topografi dan fenomena atmosfer. Variasi curah hujan di wilayah ini dipengaruhi oleh ENSO, Mode Dipol, MJO, ITCZ, dan monsun (Bimaprawira, 2021). Pengelompokan wilayah berdasarkan curah hujan juga menunjukkan bahwa wilayah di Jawa Timur dapat dikelompokkan berdasarkan karakteristik curah hujan yang serupa (Auliya, 2021).

Hujan Ekstrem

Curah hujan ekstrem adalah peristiwa curah hujan dengan intensitas sangat tinggi yang berpotensi menyebabkan banjir, tanah longsor, kerusakan infrastruktur, dan gangguan aktivitas masyarakat. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) mengkategorikan curah hujan ekstrem sebagai curah hujan yang melebihi 150 mm per hari (Nurlatifah, 2023). Risiko banjir tidak hanya dipengaruhi oleh curah hujan tetapi dapat diperkuat oleh kondisi lingkungan dan aktivitas manusia (Risnawati, 2023). Curah hujan ekstrem juga dikaitkan dengan risiko tanah longsor di daerah tropis dengan lereng curam (Naufal et al., 2026). Lebih lanjut, curah hujan dan kelembapan ekstrem dapat dikaitkan dengan masalah kesehatan lingkungan, seperti potensi demam berdarah di Kabupaten Malang (Astika Eka Kirana, 2024). Hal ini menunjukkan pentingnya menganalisis curah hujan ekstrem karena dampaknya dapat meluas ke berbagai sektor.

Isolation Forest

Pendekatan komputasi banyak digunakan dalam analisis curah hujan karena dapat menginterpretasikan pola data yang kompleks. Prediksi intensitas curah hujan di Juanda menggunakan Extra Trees, Random Forest, dan XGBoost menunjukkan potensi pembelajaran mesin untuk memodelkan intensitas curah hujan berdasarkan parameter cuaca (Taufiqi et al., 2025). Model peramalan seperti ARIMA juga digunakan untuk meramalkan curah hujan di Kota Malang (Hutabarat et al., 2025).

Metode lain seperti Extreme Learning Machine (Pramono & Soebroto, 2017), Random Forest, XGBoost (Sangaji & Sutabri, 2025), K-Means (Zaehol Fatah, 2026), dan C4.5 (Risnawati, 2023) juga telah digunakan dalam studi curah hujan dan mitigasi bencana. Studi-studi ini menunjukkan bahwa data curah hujan dapat dianalisis menggunakan pendekatan komputasi, termasuk prediksi, klasifikasi, pengelompokan, dan pendeteksian pola yang tidak biasa. Deteksi anomali adalah proses menemukan data yang berbeda dari pola mayoritas. Data dapat dibagi menjadi observasi normal dan anomali, yaitu titik data yang memiliki karakteristik berbeda dari data umum (Sunarto et al., 2021). Dalam penelitian ini, curah hujan harian yang tidak biasa diklasifikasikan sebagai anomali lokal, meskipun nilainya tidak selalu melebihi ambang batas ekstrem BMKG.

Isolation Forest bekerja berdasarkan prinsip bahwa data anomali cenderung kecil dan berbeda, sehingga lebih mudah diisolasi daripada data normal. Algoritma ini secara acak membangun sejumlah pohon isolasi, dan data dengan jalur isolasi yang lebih pendek cenderung diberi skor anomali yang lebih tinggi. Isolation Forest menawarkan keunggulan dalam efisiensi komputasi dan stabilitas model dalam deteksi anomali cuaca (Fatmi & Lase, 2025). Setiap pengamatan dalam penelitian ini direpresentasikan sebagai fitur yang berisi curah hujan harian dan bulan pengamatan, yaitu:

$$x_i = (RR_i, B_i)$$

Sebelum diproses menggunakan Isolation Forest, fitur-fitur tersebut distandarisasi untuk memastikan bahwa perbedaan skala antar variabel tidak memengaruhi proses komputasi model. Standardisasi dilakukan menggunakan persamaan berikut:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Dalam persamaan ini, z_i adalah nilai standarisasi, x_i adalah nilai awal fitur yang digunakan, μ adalah rata-rata data, dan σ adalah standar deviasi. Proses ini penting karena fitur curah hujan dan bulan memiliki satuan dan rentang nilai yang berbeda.

Dalam Isolation Forest, tingkat anomali dalam data ditentukan oleh panjang jalur isolasi. Data yang lebih mudah diisolasi akan memiliki jalur yang lebih pendek dan cenderung mencapai skor anomali yang lebih tinggi. Skor anomali dapat dinyatakan menggunakan persamaan:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

Dengan factor normalisasi:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

Dengan pendekatan:

$$H(n-1) \approx \ln(n-1) + \gamma$$

Dalam persamaan ini, $s(x, n)$ mewakili skor anomali, $E(h(x))$ mewakili panjang rata-rata jalur isolasi, dan $c(n)$ mewakili faktor normalisasi untuk jumlah sampel. Sementara itu, $H(n - 1)$ mewakili perkiraan bilangan harmonik, dan γ mewakili konstanta Euler-Mascheroni. Semakin pendek jalur isolasi rata-rata suatu kumpulan data, semakin besar kemungkinan data tersebut dikategorikan sebagai anomali. Oleh karena itu, dalam penelitian ini, nilai curah hujan harian yang berbeda dari pola mayoritas di setiap stasiun akan memiliki peluang lebih besar untuk terdeteksi sebagai anomali lokal. Dimana jika nilai skor mendekati 1 maka data menunjukkan semakin berpotensi sebagai anomali.

Penelitian Terdahulu

Penelitian curah hujan telah dilakukan secara ekstensif menggunakan pendekatan prediksi, peramalan, klasifikasi, pengelompokan, dan deteksi anomali. Prediksi intensitas curah hujan di Juanda menunjukkan kinerja yang baik menggunakan model berbasis pohon keputusan (Taufiqi et al., 2025). Peramalan curah hujan juga telah dilakukan menggunakan ARIMA dan Extreme Learning Machines (Hutabarat et al., 2025; Pramono & Soebroto, 2017). Deteksi anomali cuaca telah dipelajari menggunakan Isolation Forest dan AutoEncoder, dengan hasil yang menunjukkan efisiensi komputasi Isolation Forest yang lebih unggul (Fatmi & Lase, 2025). Identifikasi curah hujan yang berpotensi menyebabkan banjir juga telah dilakukan menggunakan kombinasi LSTM dan Isolation Forest (Wijayanto et al., 2024). Studi ini memanfaatkan fondasi tersebut dengan berfokus pada implementasi Isolation Forest untuk membandingkan hasil anomali lokal dengan ambang batas ekstrem BMKG di tiga stasiun di Jawa Timur.

METODE PENELITIAN

Data Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan komputasional berbasis pembelajaran mesin. Fokus penelitian ini adalah implementasi algoritma Isolation Forest untuk mendeteksi anomali curah hujan harian dan kemudian membandingkannya dengan ambang batas curah hujan ekstrem BMKG. Data yang digunakan adalah data curah hujan harian dari stasiun Juanda, Kediri, dan Malang. Setelah pra-pemrosesan, terkumpul 1.006 observasi: 329 dari Juanda, 331 dari Kediri, dan 346 dari Malang. Juanda mewakili daerah pesisir, Kediri mewakili daerah non-pesisir, dan Malang mewakili daerah dataran tinggi.

Variabel Penelitian

Variabel penelitian adalah komponen data yang digunakan dalam proses analisis untuk mendeteksi anomali curah hujan harian. Variabel utama dalam penelitian ini adalah curah hujan harian (RR) dan bulan pengamatan. Curah hujan harian digunakan sebagai fitur utama karena merupakan nilai yang dianalisis untuk mendeteksi anomali, sedangkan bulan pengamatan digunakan untuk memberikan konteks temporal bagi pola curah hujan. Selain itu, terdapat variabel pendukung seperti tanggal, nama stasiun, topografi regional, kategori curah hujan BMKG, label ekstrem BMKG, skor anomali, dan label anomali model. Variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1 Variabel Penelitian

Variabel	Keterangan
TANGGAL	Tanggal Pengamatan
RR	Curah Hujan Harian Dalam mm
STASIUN	Nama stasiun
TOPOGRAFI	Karakter wilayah
BULAN	Bulan pengamatan
KATEGORI_BMKG	Kategori curah hujan BMKG
is_extreme_bmkg	Label ekstrem BMKG
anomaly_score	Skor anomali model
is_extreme_model	Label anomali Isolation Forest

Berdasarkan Tabel 1 variabel RR dan BULAN digunakan sebagai fitur input dalam model Isolation Forest. Variabel is_extreme_bmkg digunakan sebagai perbandingan berdasarkan ambang batas ekstrem BMKG, sedangkan anomaly_score dan is_extreme_model merupakan output model. Dengan demikian, variabel dalam penelitian ini terdiri dari variabel input, variabel pendukung, variabel perbandingan, dan variabel output model.

Tahapan Analisi Data

Tahapan analisis data dalam penelitian ini dilakukan secara berurutan, dimulai dengan pra-pemrosesan data, penetapan kategori curah hujan BMKG, standarisasi fitur, implementasi Isolation Forest, dan evaluasi hasil model. Tahapan-tahapan ini memastikan bahwa data curah hujan harian dari setiap stasiun memiliki format yang sama dan dapat dianalisis secara komputasi.

Tahap pertama adalah pra-pemrosesan data. Data curah hujan harian dari stasiun Juanda, Kediri, dan Malang dibaca dalam format CSV. Nama kolom distandarisasi menjadi DATE dan RR, kemudian kolom tanggal dikonversi ke format tanggal. Data kosong, nilai curah hujan negatif, dan kode data yang tidak valid seperti 8888 dan 9999 dihapus dari data. Data kemudian diberi informasi tambahan seperti nama stasiun, karakteristik topografi, dan bulan pengamatan.

Setelah data bersih diperoleh, kategori curah hujan dibuat berdasarkan klasifikasi BMKG. Kategori-kategori ini digunakan sebagai perbandingan terhadap hasil deteksi anomali Isolation Forest. Oleh karena itu, tahapan analisis data dalam penelitian ini tidak hanya mencakup pembersihan data tetapi juga pembentukan variabel perbandingan dan persiapan fitur model.

Kategori Curah Hujan BMKG

Dalam penelitian ini, kategori BMKG berfungsi sebagai tolok ukur untuk membandingkan hasil deteksi anomali model. Curah hujan harian dikategorikan sebagai berawan/tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Kategori ekstrem didefinisikan sebagai curah hujan harian yang melebihi 150 mm.

Tabel 2 Kategori Curah Hujan BMKG

Kategori	Curah Hujan Harian
Berawan	0 - 5 mm/hari
Ringan	5 - 20 mm/hari
Sedang	20 - 50 mm/hari
Lebat	50 - 100 mm/hari
Sangat Lebat	100 - 150 mm/hari
Ekstrem	>150 mm/hari

Berdasarkan Tabel 2. label ekstrem BMKG dibentuk menggunakan aturan berikut:

$$Y_{BMKG} = \begin{cases} 1, & RR > 150 \\ 0, & RR \leq 150 \end{cases}$$

Label $Y_{BMKG} = 1$ menunjukkan bahwa data tersebut merupakan kejadian ekstrem berdasarkan ambang batas BMKG, sedangkan $Y_{BMKG} = 0$ menunjukkan bahwa data tersebut tidak dianggap ekstrem. Label ini tidak digunakan untuk melatih model, melainkan berfungsi sebagai perbandingan dalam mengevaluasi hasil deteksi anomali Isolation Forest.

Implementasi Isolation Forest

Model Isolation Forest diterapkan secara terpisah di setiap stasiun untuk menangkap karakteristik curah hujan lokal. Studi ini menggunakan fitur RR dan BULAN sebagai input model. Fitur RR menunjukkan nilai curah hujan harian, sedangkan fitur BULAN menunjukkan waktu pengamatan dalam bulan. Sebelum menjalankan model, kedua fitur tersebut distandarisasi menggunakan StandardScaler untuk memastikan bahwa perbedaan skala antar variabel tidak memengaruhi proses komputasi.

Studi ini menggunakan 300 estimator, atau 300 pohon isolasi, dalam algoritma Isolation Forest. Setiap pohon isolasi secara acak memisahkan data berdasarkan fitur yang digunakan. Model kemudian menghitung panjang jalur isolasi rata-rata di semua pohon. Data dengan panjang jalur yang lebih pendek lebih mudah dipisahkan dari sebagian besar data, sehingga memiliki peluang lebih besar untuk dikategorikan sebagai anomali.

Nilai kontaminasi ditetapkan pada 0,05. Nilai ini menunjukkan bahwa model diarahkan untuk mendeteksi sekitar 5% data sebagai kandidat anomali di setiap stasiun. Dengan pengaturan ini, model mendeteksi anomali berdasarkan penyimpangan dari pola curah hujan lokal, bukan berdasarkan ambang batas tetap BMKG. Oleh karena itu, hasil model digunakan sebagai perbandingan terhadap kategori ekstrem BMKG, bukan sebagai pengganti ambang batas operasional.

Evaluasi Model

Evaluasi dilakukan dengan membandingkan hasil deteksi Isolation Forest dengan ambang batas ekstrem BMKG. Evaluasi ini tidak dimaksudkan untuk mengukur akurasi prediksi curah hujan, karena model ini tidak melakukan peramalan dan tidak dilatih menggunakan label ekstrem. Model ini hanya mendeteksi data yang menyimpang dari pola mayoritas berdasarkan fitur curah hujan harian dan pengamatan selama beberapa bulan. Total kejadian ekstrem BMKG dihitung sebagai berikut:

$$Total_{BMKG} = \sum_{i=1}^N Y_{BMKG,i}$$

Menghitung total anomali:

$$Total_{IF} = \sum_{i=1}^N Y_{BMKG,i} \times Y_{IF,i}$$

Kejadian yang sama-sama terdeteksi dihitung sebagai:

$$Overlap = \sum_{i=1}^N Y_{BMKG,i} \times Y_{IF,i}$$

Selain jumlah anomali dan kejadian yang terdeteksi oleh keduanya, penelitian ini juga menghitung nilai recall untuk kejadian ekstrem BMKG. Recall digunakan untuk menentukan sejauh mana kejadian yang dianggap ekstrem berdasarkan ambang batas BMKG dapat dideteksi sebagai anomali oleh Isolation Forest. Recall lebih tepat daripada akurasi umum karena penelitian ini menggunakan metode pembelajaran tanpa pengawasan, sehingga model tidak dilatih menggunakan label ekstrem BMKG. Berikut perhitungan recall:

$$Recall = \frac{TP}{TP + FN}$$

Dalam penelitian ini, BMKG tidak digunakan sebagai label pelatihan model, tetapi hanya sebagai tolak ukur evaluatif. Oleh karena itu, nilai recall tidak diinterpretasikan sebagai akurasi prediksi, melainkan sebagai ukuran kesesuaian model untuk menangkap kejadian ekstrem berdasarkan ambang batas BMKG.

Selain recall, studi ini juga menghitung presisi komparatif terhadap ambang batas kejadian ekstrem BMKG. Presisi digunakan untuk menentukan proporsi anomali model yang juga dianggap sebagai kejadian ekstrem menurut BMKG. Namun, nilai presisi dalam studi ini tidak diinterpretasikan sebagai tingkat kesalahan model, karena Isolation Forest bertujuan untuk mendeteksi anomali lokal, bukan hanya kejadian yang melebihi ambang batas BMKG.

$$Precision = \frac{TP}{TP + FP}$$

TP menunjukkan jumlah kejadian ekstrem BMKG yang juga terdeteksi sebagai anomali oleh model. FP menunjukkan jumlah kejadian yang hanya terdeteksi oleh model tetapi tidak dianggap ekstrem menurut ambang batas BMKG. Dalam konteks studi ini, FP tidak selalu menunjukkan kesalahan model. FP dapat menunjukkan kejadian anomali lokal yang berada di bawah ambang batas kejadian ekstrem BMKG.

HASIL DAN PEMBAHASAN

Deskripsi Data

Setelah melalui tahap pra-pemrosesan, data penelitian terdiri dari 1.006 pengamatan dari tiga stasiun: Juanda, Kediri, dan Malang. Juanda mewakili daerah pesisir, Kediri mewakili daerah non-pesisir, dan Malang mewakili daerah dataran tinggi.

Tabel 3 Deskripsi Data

Stasiun	Topografi	Jumlah Data	RR Maksimum
Juanda	Pesisir	329	150,8
Kediri	Non-Pesisir	331	105,1
Malang	Dataran Tinggi	346	145,6

Berdasarkan Tabel 3, Stasiun Juanda memiliki nilai curah hujan maksimum tertinggi, yaitu 150,8 mm, sehingga melebihi ambang batas ekstrem BMKG. Stasiun Malang memiliki curah hujan maksimum sebesar 145,6 mm, mendekati ambang batas ekstrem BMKG, sedangkan Stasiun Kediri memiliki nilai maksimum sebesar 105,1 mm.

Hasil Kategori Curah Hujan BMKG

Klasifikasi curah hujan didasarkan pada kategori BMKG: berawan, ringan, sedang, lebat, sangat lebat, dan ekstrem. Kategori ekstrem didefinisikan ketika curah hujan harian melebihi 150 mm.

Tabel 4 Hasil Kategori Curah Hujan BMKG

Stasiun	Berawan	Ringan	Sedang	Lebat	Sangat Lebat	Ekstrem
Juanda	187	90	31	19	1	1
Kediri	178	108	31	13	1	0
Malang	174	129	34	8	1	0

Sebagian besar data di ketiga stasiun tersebut termasuk dalam kategori berawan dan hujan ringan. Hanya Stasiun Juanda yang mencatat satu kejadian BMKG ekstrem. Stasiun Kediri dan Malang tidak mencatat kejadian apa pun yang melebihi ambang batas 150 mm.

Hasil Deteksi Anomali Isolation Forest

Model Isolation Forest diterapkan secara terpisah pada setiap stasiun dengan nilai kontaminasi 0,05. Nilai ini berarti model diarahkan untuk mendeteksi sekitar 5% data sebagai kandidat anomali di setiap stasiun.

Tabel 5 Hasil Deteksi Anomali Isolation Forest

Stasiun	Jumlah Data	Anomali Model	RR Minimum	RR Maksimum
Juanda	329	17	150,8	150,8
Kediri	331	17	105,1	105,1
Malang	346	18	145,6	145,6

Isolation Forest mendeteksi 17 anomali di Juanda, 17 anomali di Kediri, dan 18 anomali di Malang. Jumlah ini sesuai dengan 5% dari total data untuk setiap stasiun. Nilai anomali minimum berbeda di setiap stasiun, yaitu 35,4 mm di Juanda, 30,7 mm di Kediri, dan 22,5 mm di Malang. Nilai ini tidak diinterpretasikan sebagai ambang batas ekstrem baru, melainkan sebagai nilai curah hujan terendah yang masih dianggap tidak biasa oleh model berdasarkan pola lokal.

Perbandingan Jumlah Kejadian BMKG dan Anomali Isolation Forest

Perbandingan dilakukan untuk menentukan apakah kejadian ekstrem berdasarkan BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) juga terdeteksi sebagai anomali oleh Isolation Forest. Hasil perbandingan ditunjukkan pada Tabel 6.

Tabel 6 Hasil Model

Stasiun	Ekstrem BMKG	Anomali Model	Terdeteksi Keduanya	Model Saja
Juanda	1	17	1	16
Kediri	0	17	0	17
Malang	0	18	0	18

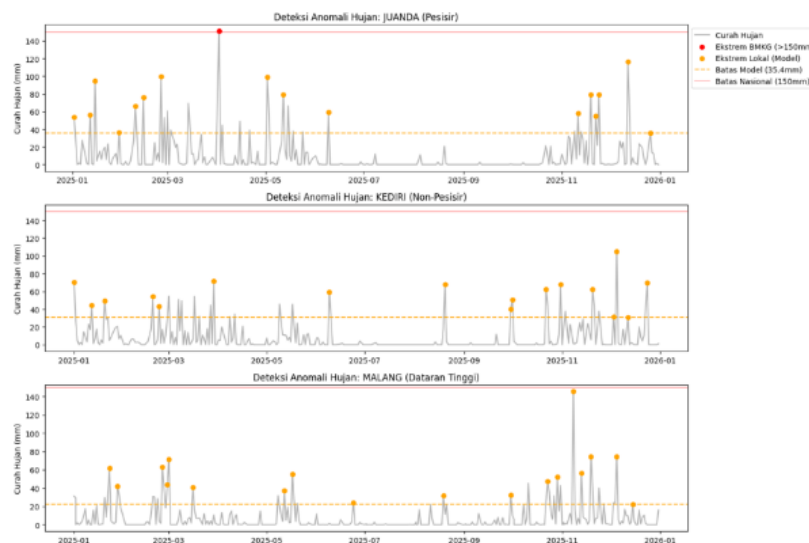
Tabel 6 menunjukkan perbandingan antara kejadian ekstrem berdasarkan ambang batas BMKG dan hasil deteksi anomali Isolation Forest. Kolom "BMKG Ekstrem" menunjukkan jumlah kejadian dengan curah hujan harian melebihi 150mm. Kolom "Anomali Model" menunjukkan jumlah kejadian yang diklasifikasikan sebagai anomali oleh Isolation Forest. Kolom "Keduanya Terdeteksi" menunjukkan jumlah kejadian yang memenuhi kategori ekstrem BMKG dan terdeteksi sebagai anomali oleh model. Kolom "Model Saja" menunjukkan kejadian yang tidak melebihi ambang batas ekstrem BMKG, tetapi dianggap tidak biasa oleh model berdasarkan pola stasiun lokal.

Satu kejadian ekstrem BMKG di Stasiun Juanda juga terdeteksi sebagai anomali oleh model. Di Stasiun Kediri dan Malang, tidak terjadi kejadian yang melebihi 150 mm, tetapi model masih mendeteksi beberapa data sebagai anomali lokal. Hal ini menunjukkan bahwa Isolation Forest tidak beroperasi berdasarkan ambang batas tetap, melainkan berdasarkan penyimpangan dari pola curah hujan masing-masing stasiun.

Berdasarkan Tabel 6, terdapat satu kejadian BMKG ekstrem dalam keseluruhan data penelitian, yaitu di Stasiun Juanda. Kejadian ini juga terdeteksi sebagai anomali oleh Isolation Forest, sehingga nilai TP adalah 1. Tidak ada kejadian BMKG ekstrem yang gagal dideteksi oleh model, sehingga nilai FN adalah 0. Dengan demikian, nilai recall untuk kejadian BMKG ekstrem adalah:

$$Recall = \frac{TP}{TP + FN}$$
$$Recall = \frac{1}{1 + 0} = 1$$

Nilai recall 100% menunjukkan bahwa semua kejadian ekstrem BMKG dalam data penelitian berhasil dideteksi sebagai anomali oleh model. Namun, nilai ini tidak menunjukkan akurasi prediksi 100%, karena Isolation Forest tidak melakukan peramalan dan tidak dilatih menggunakan label kejadian ekstrem BMKG.



Gambar 1 Visualisasi hasil deteksi anomali curah hujan harian menggunakan Isolation Forest di Stasiun Juanda, Kediri, dan Malang.

Gambar 1 memperjelas hasil yang disajikan dalam Tabel 5. Tabel 5 menunjukkan jumlah kejadian ekstrem BMKG, jumlah anomali model, kejadian yang terdeteksi oleh keduanya, dan kejadian yang hanya terdeteksi oleh model. Sementara itu, Gambar 1 menunjukkan lokasi kejadian-kejadian ini berdasarkan waktu pengamatan dan nilai curah hujan harian. Di Stasiun Juanda, satu kejadian curah hujan melebihi ambang batas 150 mm/hari dan juga terdeteksi sebagai anomali oleh Isolation Forest. Di Stasiun Kediri dan Malang, tidak ada kejadian yang melebihi ambang batas BMKG. Namun, model masih mendeteksi beberapa kejadian sebagai anomali lokal karena nilai curah hujan ini berbeda dari pola di setiap stasiun.

Hasil ini menunjukkan bahwa BMKG dan Isolation Forest memiliki basis penilaian yang berbeda. BMKG menggunakan ambang batas tetap, sedangkan Isolation Forest mendeteksi penyimpangan berdasarkan pola lokal. Dengan demikian, Gambar 1 mendukung Tabel 5. dengan menunjukkan bahwa anomali model tidak selalu sesuai dengan kejadian ekstrem BMKG. Titik dan garis pada Gambar 1 digunakan untuk membedakan antara curah hujan harian, ambang batas BMKG, kejadian ekstrem BMKG, dan anomali lokal model.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa Isolation Forest mampu mendeteksi kejadian curah hujan yang tidak biasa di setiap stasiun. Ambang batas BMKG beroperasi sebagai batas tetap, yaitu curah hujan melebihi 150 mm per hari, sedangkan Isolation Forest membaca penyimpangan data dari pola mayoritas di setiap stasiun. Perbedaan cara kerja ini berarti bahwa beberapa kejadian yang kurang dari ambang batas BMKG masih dapat dideteksi sebagai anomali oleh model.

Di Stasiun Juanda, curah hujan sebesar 150,8 mm terdeteksi sebagai kejadian ekstrem BMKG dan anomali model. Hal ini menunjukkan bahwa model mampu mengenali kejadian yang memang secara meteorologis sangat tinggi. Di Stasiun Kediri, nilai maksimum 105,1 mm tidak dianggap sebagai kejadian ekstrem BMKG tetapi masih terdeteksi sebagai anomali lokal. Di Stasiun Malang, nilai maksimum 145,6 mm juga terdeteksi sebagai anomali karena mendekati ambang batas ekstrem dan berbeda dari pola curah hujan lokal.

Nilai recall 100% menunjukkan bahwa model berhasil mendeteksi satu-satunya kejadian ekstrem BMKG dalam data penelitian. Namun, nilai ini harus ditafsirkan dengan hati-hati karena jumlah kejadian ekstrem BMKG hanya satu. Presisi komparatif sebesar 1,92% menunjukkan bahwa sebagian besar anomali model tidak termasuk dalam kategori ekstrem BMKG. Ini tidak berarti model tersebut salah, karena Isolation Forest mendeteksi penyimpangan lokal, sedangkan BMKG menggunakan ambang batas tetap 150 mm/hari.

Hasil penelitian ini juga sejalan dengan penerapan LSTM dan Isolation Forest untuk mengidentifikasi curah hujan yang berpotensi menyebabkan banjir. Dalam penelitian tersebut, Isolation Forest digunakan untuk mengidentifikasi curah hujan yang berpotensi menjadi peristiwa signifikan setelah data diproses secara komputasi. Perbedaannya adalah penelitian ini tidak memprediksi curah hujan, melainkan langsung menerapkan Isolation Forest pada data curah hujan harian historis untuk membandingkan hasil anomali model dengan ambang batas ekstrem BMKG.

Hasil ini juga menunjukkan bahwa ambang batas BMKG dan Isolation Forest memiliki fungsi yang berbeda tetapi saling melengkapi. Ambang batas BMKG tetap menjadi prosedur operasi standar, sementara Isolation Forest dapat digunakan sebagai alat komputasi untuk mengidentifikasi kejadian curah hujan yang tidak biasa berdasarkan pola historis lokal. Oleh karena itu, model ini tidak menggantikan ambang batas BMKG, tetapi dapat memberikan informasi tambahan dalam mengevaluasi curah hujan ekstrem di daerah dengan karakteristik topografi yang berbeda.

Hasil pemodelan dapat digunakan untuk menentukan batas awal anomali curah hujan lokal, tetapi belum dapat digunakan sebagai ambang batas peringatan banjir. Nilai anomali minimum menunjukkan curah hujan terendah yang dianggap tidak biasa oleh model: 35,4 mm di Juanda, 30,7 mm di Kediri, dan 22,5 mm di Malang. Namun, nilai-nilai ini tidak dapat digunakan sebagai ambang batas peringatan banjir karena penelitian ini tidak menggunakan data kejadian banjir aktual. Oleh karena itu, ambang batas peringatan awal lebih akurat ditentukan dari hasil numerik model, sedangkan Gambar 1 hanya berfungsi untuk memperjelas posisi anomali relatif terhadap ambang batas BMKG. Penelitian selanjutnya perlu menghubungkan hasil anomali dengan data banjir untuk menentukan ambang batas peringatan yang lebih valid.

KESIMPULAN

Berdasarkan hasil penelitian, algoritma Isolation Forest mampu mendeteksi anomali curah hujan harian di stasiun Juanda, Kediri, dan Malang. Model mendeteksi 17 anomali di Juanda, 17 anomali di Kediri, dan 18 anomali di Malang, dengan nilai kontaminasi 0,05. Model juga mendeteksi satu kejadian BMKG ekstrem di Juanda sebesar 150,8 mm sebagai anomali, sedangkan Kediri dan Malang tidak mengalami kejadian yang melebihi 150 mm/hari. Nilai recall 100% menunjukkan bahwa model berhasil mendeteksi satu-satunya kejadian BMKG ekstrem dalam data penelitian. Namun, nilai ini harus diinterpretasikan dengan hati-hati karena hanya ada satu kejadian BMKG ekstrem. Presisi komparatif sebesar 1,92% menunjukkan bahwa sebagian besar anomali model berada di bawah ambang batas ekstrem BMKG. Hal ini menunjukkan bahwa Isolation Forest mendeteksi penyimpangan lokal, sedangkan BMKG menggunakan ambang batas tetap.

Penelitian lebih lanjut disarankan untuk menggunakan periode data yang lebih panjang, menambahkan variabel meteorologi lainnya, dan mengkorelasikan hasil anomali dengan data kejadian banjir aktual. Penelitian di masa mendatang juga dapat membandingkan Isolation Forest dengan metode lain, seperti LOF, One-Class SVM, Extended Isolation Forest, atau AutoEncoder.

DAFTAR PUSTAKA

- Astika Eka Kirana, A. S. (2024). Prediksi Kelembaban Dan Curah Hujan Maksimum Di Kabupaten Malang Menggunakan Bivariate Extreme Value Logistic. 12(03), 465–474.
- Auliya, E. R. (2021). Pengklasteran Daerah Di Jawa Timur Berdasarkan Curah Hujan. 09(02).
- Bimaprawira, A. K. (2021). Keterkaitan Periodisitas Curah Hujan Di Daerah Pesisir Dan Pegunungan Provinsi Jawa Timur Dengan Variabilitas Cuaca Skala Global Dan Regional Relationship Of Rainfall Periodicity In Coastal Dan Mountain Areas Of East Java Province With Global And Regional. 22(2), 51–59.
- Fatmi, Y., & Lase, Y. Y. (2025). Jurnal manajemen teknologi informatika. 3(26), 237–245.
- Hutabarat, S. H., Putra, F. A. A., Informasi, S., & Malang, U. M. (2025). Perbandingan Metode Exponential Smoothing Dan Arima Pada Peramalan Curah Hujan Di Kota Malang. 9(4), 7294–7300.
- Munanjar, M. C., Ariska, J., & Safitri, I. (2025). Validasi Data Curah Hujan Maksimum Bulanan GPM. 706–711. <https://doi.org/10.33364/konstruksi/v.23-2.2708>
- Naufal, A., Brampu, F., & Fitra, J. (2026). Analisis Risiko Longsor Di Wilayah Tropis Berdasarkan Curah Hujan Dan Topografi Berbasis DEM : Studi Literatur Sistematis. 16(01), 310–324.
- Nurlatifah. (2023). Analisis Potensi Kejadian Curah Hujan Ekstrem di Masa Mendatang Sebagai Dampak dari Perubahan Iklim di Pulau Jawa Berbasis Model Iklim Regional Conformal Cubic Atmospheric Model. 21(4), 980–986. <https://doi.org/10.14710/jil.21.4.980-986>

- Pramono, A., & Soebroto, A. A. (2017). Analisis Peramalan Curah Hujan Menggunakan Metode Extreme Learning Machine (Studi Kasus : Stasiun Klimatologi Jawa Timur). 1(1), 1-7.
- Risnawati, I. (2023). Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan. 3(2), 78-84.
- Sangaji, D., & Sutabri, T. (2025). E ISSN : 2809-4069 Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla. 5(1), 13-18.
- Sunarto, M. W., Kurniawan, D., Siswanto, E., & Huda, H. I. (2021). Deteksi Anomali Menggunakan Extended Isolation Forest (Eif). 1(2), 96-111.
- Taufiqi, R. I., Pramuji, V. M., Arifudin, M., & Hari, A. (2025). Perbandingan Model Decision Tree untuk Prediksi Intensitas Hujan di. 17(1), 15-28.
- Wijayanto, A., Sugiharto, A., Santoso, R., Diponegoro, U., & Korespondensi, P. (2024). Identifikasi Dini Curah Hujan Berpotensi Banjir Menggunakan Algoritma Long Short-Term Memory (Lstm) Dan Isolation Forest Early Identification Of Rainfall With Flood Potential Using Long Short-Term Memory (Lstm) And Isolation Forest Algorithms Case Stu. 11(3). <https://doi.org/10.25126/jtiik.938718>
- Xu, H., Pang, G., Wang, Y., & Wang, Y. (2023). Deep Isolation Forest for Anomaly Detection. 1-14.
- Zaehol Fatah, R. A. (2026). Mendeteksi Curah Hujan Ekstrem sebagai Upaya Mitigasi Banjir Menggunakan Algoritma K-Means. 5(April), 243-251.