JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Perbandingan Performa Algoritma Random Forest dan XGBoost dalam Memprediksi Hujan di Area Gunung Ungaran

Arizal Irsyad Imanullah, Ahmad Zainul Fanani*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]arizalirsyad08@gmail.com, [2*]a.zainul.fanani@dsn.dinus.ac.id
Email Penulis Korespondensi: a.zainul.fanani@dsn.dinus.ac.id
Submitted **14-12-2025;** Accepted **27-01-2026;** Published **28-02-2026**

Abstrak

Aktivitas pendakian di Gunung Ungaran sering kali terkendala oleh perubahan cuaca yang ekstrem dan sulit diprediksi, yang berpotensi membahayakan keselamatan para pendaki. Salah satu tantangan utama dalam pengembangan model prediksi hujan otomatis di wilayah ini adalah ketidakseimbangan kelas (*class imbalance*) pada data historis meteorologi, di mana jumlah hari tanpa hujan jauh mendominasi kejadian hujan. Kondisi ini sering kali menyebabkan model machine learning menjadi bias ke arah kelas mayoritas dan gagal mendeteksi kejadian hujan yang sebenarnya (*false negative*). Penelitian ini bertujuan untuk memecahkan masalah tersebut melalui analisis komparatif kinerja dua algoritma ensemble populer, yakni Random Forest dan Extreme Gradient Boosting (XGBoost). Penelitian secara spesifik menginvestigasi dampak penerapan teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk menyeimbangkan distribusi data latih guna meningkatkan akurasi deteksi kelas minoritas. Menggunakan dataset harian reanalisis ERA5 periode 2019–2023 dengan variabel input meliputi suhu, kelembapan, tekanan udara, dan kecepatan angin, model dilatih dan divalidasi menggunakan metode time-based split dengan rasio 80:20. Evaluasi kinerja dilakukan secara mendalam menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bukti empiris yang tegas bahwa penerapan SMOTE memberikan dampak paling optimal pada algoritma XGBoost. Model kombinasi XGBoost-SMOTE berhasil mencatatkan kinerja terbaik dengan akurasi 80,50% dan F1-score 83,23%, mengungguli model Random Forest yang tertahan pada akurasi 78,21%. Kesimpulannya, integrasi metode boosting dengan teknik resampling data terbukti sangat efektif meningkatkan reliabilitas prediksi hujan di wilayah dengan topografi kompleks.

Kata Kunci: Prediksi Hujan; Random Forest; XGBoost; Smote; Gunung Ungaran

Abstract

Hiking activities in Mount Ungaran are frequently hindered by extreme and unpredictable weather changes, which potentially endanger the safety of hikers. One of the primary challenges in developing an automated rainfall prediction model for this region is the class imbalance in historical meteorological data, where the number of non-rainy days significantly dominates rainfall events. This condition often causes machine learning models to become biased toward the majority class, leading to a failure in detecting actual rainfall events (false negatives). This study aims to address this issue through a comparative analysis of the performance of two popular ensemble algorithms, namely Random Forest and Extreme Gradient Boosting (XGBoost). Specifically, this research investigates the impact of applying the Synthetic Minority Oversampling Technique (SMOTE) to balance the training data distribution in order to enhance minority class detection accuracy. Using the ERA5 reanalysis daily dataset for the 2019–2023 period with input variables including temperature, humidity, air pressure, and wind speed, the models were trained and validated using a time-based split method with an 80:20 ratio. Performance evaluation was conducted comprehensively using accuracy, precision, recall, and F1-score metrics. The results provide strong empirical evidence that the application of SMOTE yields the most optimal impact on the XGBoost algorithm. The combined XGBoost-SMOTE model successfully achieved the best performance with an accuracy of 80.50% and an F1-score of 83.23%, outperforming the Random Forest model which remained at an accuracy of 78.21%. In conclusion, the integration of boosting methods with data resampling techniques proves to be highly effective in improving rainfall prediction reliability in regions with complex topography.

Keywords: Rainfall Prediction; Random Forest; XGBoost; Smote; Mount Ungaran

1. PENDAHULUAN

Indonesia dikenal memiliki beragam destinasi wisata alam, dan salah satunya adalah aktivitas pendakian gunung. Gunung Ungaran, yang terletak di Kabupaten Semarang, Jawa Tengah, menjadi salah satu pilihan favorit para pendaki berkat panorama alamnya yang masih terjaga serta jalur pendakian yang tergolong ramah bagi pendaki pemula maupun yang sudah berpengalaman. Namun, kondisi cuaca yang sulit diprediksi menjadi salah satu tantangan utama dalam kegiatan pendakian. Cuaca ekstrem seperti hujan deras, kabut pekat, atau angin kencang dapat mengancam keselamatan para pendaki. Oleh sebab itu, informasi prakiraan cuaca yang akurat sangat diperlukan untuk mendukung keamanan dan kenyamanan selama melakukan aktivitas pendakian [1].

Di era digital saat ini, kemajuan teknologi kecerdasan buatan (*Artificial Intelligence*) memungkinkan proses pengolahan data cuaca dilakukan secara lebih cerdas dan prediktif. Salah satu pendekatan yang dapat diterapkan adalah penggunaan algoritma machine learning, seperti Random Forest, yang dikenal mampu bekerja secara efektif dalam menyelesaikan tugas klasifikasi dan prediksi berdasarkan data historis [2]. Berdasarkan hasil perbandingan model machine learning dalam memprediksi suhu, XGBoost secara konsisten tampil sebagai model dengan kinerja paling unggul serta menunjukkan akurasi dan stabilitas yang lebih baik dalam mengolah data cuaca yang kompleks dibandingkan dengan Random Forest [3].

Beberapa penelitian terdahulu telah membuktikan efektivitas Random Forest dalam prediksi cuaca. Penelitian oleh Taqiyuddin dan Sasongko menunjukkan bahwa penggunaan Random Forest dalam memprediksi cuaca di Kabupaten

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Sleman memberikan hasil akurasi yang cukup tinggi. Berdasarkan hasil pengujian, model tersebut menghasilkan akurasi yang cukup tinggi dengan nilai R^2 sebesar 0.691 dan MSE sebesar 0.0097. Namun, penelitian ini belum menerapkan teknik penyeimbangan data (data balancing) secara khusus dan hanya terbatas pada penggunaan satu algoritma *bagging* tanpa membandingkannya dengan metode boosting seperti XGBoost8. Hal ini membuka peluang untuk pengembangan lebih lanjut guna melihat efektivitas model komparatif dalam menangani variabilitas data cuaca yang kompleks [4].

Penelitian terbaru yang dilakukan oleh Sangaji dan Sutabri mengevaluasi kinerja algoritma machine learning untuk memprediksi pola curah hujan di Sumatera Selatan. Studi ini membandingkan algoritma Random Forest, XGBoost, dan Model Gabungan (*Ensemble*). Hasil penelitian menunjukkan bahwa algoritma XGBoost memiliki kinerja yang lebih baik dibandingkan Random Forest secara tunggal, dengan nilai *R-squared* (R^2) sebesar 0,905 berbanding 0,892. Namun, akurasi tertinggi dicapai oleh Model Gabungan (dengan bobot 0,6 XGBoost dan 0,4 Random Forest) yang menghasilkan nilai R^2 sebesar 0,918 serta nilai kesalahan (*error*) terendah dengan MAPE sebesar 11,87%1111. Temuan ini mengindikasikan bahwa teknik boosting dan pendekatan ensemble sangat efektif dalam meningkatkan akurasi prediksi pada data iklim yang kompleks [3].

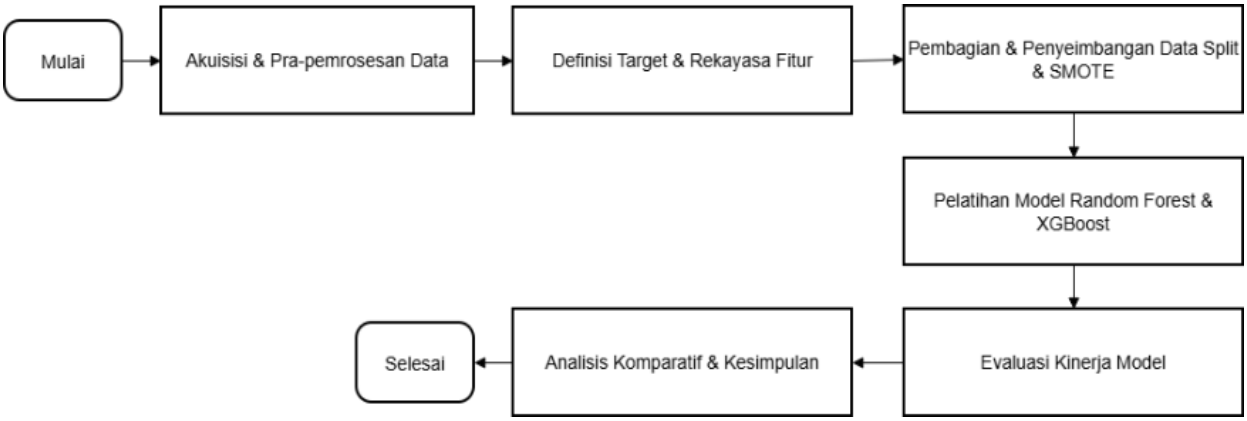
Penelitian lain yang membandingkan *XGBoost*, *Random Forest*, dan *Support Vector Regression* (SVR) untuk prediksi curah hujan menemukan bahwa hasil pengujian, algoritma XGBoost terbukti memberikan tingkat akurasi terbaik dibandingkan model lainnya. Hal ini ditunjukkan dengan nilai koefisien determinasi (R^2) tertinggi yang mencapai 0,8183, serta tingkat kesalahan yang paling rendah dengan nilai *Mean Squared Error* (MSE) sebesar 0,2278 dan *Mean Absolute Error* (MAE) sebesar 0,3744111. Kinerja ini mengungguli Random Forest yang mencatatkan skor R^2 sebesar 0,8068 dan SVR yang memperoleh skor R^2 sebesar 0,79872. Temuan ini mengindikasikan bahwa XGBoost memiliki kemampuan generalisasi yang lebih baik dan lebih efektif dalam menangani kompleksitas serta outlier pada data cuaca [5]. Lebih lanjut, penerapan spesifik XGBoost untuk peramalan pola curah hujan bulanan juga terkonfirmasi efektif dalam menangkap pola data deret waktu, khususnya pada kondisi ekstrem, di mana performa model terbaik dicapai saat menggunakan alokasi data latih yang lebih besar [6]. Meskipun demikian, terdapat kesenjangan (gap) penelitian yang signifikan di mana belum ada studi yang secara spesifik mengintegrasikan analisis komparatif antara algoritma Random Forest dan XGBoost dengan teknik penyeimbangan data SMOTE untuk memprediksi hujan harian di wilayah dengan karakteristik topografi kompleks seperti Gunung Ungaran.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan mengevaluasi model peramalan hujan harian dengan melakukan analisis komparatif antara dua algoritma ensemble learning terkemuka: Random Forest dan XGBoost [7] [8]. Untuk mencapai tujuan tersebut, penelitian ini memanfaatkan data historis cuaca dari dataset reanalisis global ERA5 yang diperkaya melalui rekayasa fitur temporal (*time-series feature engineering*), termasuk fitur lag (*data historis*) dan rolling aggregates (tren cuaca). Mengingat potensi ketidakseimbangan kelas antara hari hujan dan tidak hujan, penelitian ini juga menerapkan teknik oversampling SMOTE [9] untuk meningkatkan kemampuan model dalam mengenali kelas minoritas [10].

Kebaruan penelitian ini terletak pada tiga aspek utama: (1) fokus pada peramalan (forecasting) kondisi hujan hari berikutnya, bukan sekadar klasifikasi; (2) penerapan rekayasa fitur temporal untuk menangkap dinamika cuaca yang sering diabaikan; dan (3) analisis komparatif langsung antara Random Forest dan XGBoost untuk menentukan model mana yang paling unggul dalam konteks ini. Diharapkan, hasil perbandingan ini tidak hanya memberikan model yang lebih andal untuk keselamatan pendaki di Gunung Ungaran, tetapi juga menyajikan wawasan metodologis yang berharga mengenai efektivitas *feature engineering* temporal dan pemilihan model ensemble untuk peramalan cuaca di lingkungan pegunungan tropis.

2. METODOLOGI PENELITIAN

Metode penelitian ini disusun berdasarkan tahapan kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM), sehingga proses yang dilakukan berlangsung secara terstruktur mulai dari pengumpulan data hingga evaluasi model [11]. Tahapan penelitian yang diterapkan dalam studi ini dijelaskan melalui alur pada gambar berikut.



Gambar 1. Alur Penelitian

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penelitian ini dimulai dengan pengolahan data yang mencakup teknik penyeimbangan SMOTE, dilanjutkan dengan pelatihan model Random Forest dan XGBoost. Selanjutnya, dilakukan evaluasi dan analisis komparatif untuk menyimpulkan model mana yang memiliki kinerja terbaik dari keduanya.

2.1 Pengumpulan dan Pra-pemrosesan Data

Penelitian ini menggunakan data cuaca historis yang berasal dari dataset reanalisis global ERA5. Dataset ERA5 merupakan keluaran dari European Centre for Medium-Range Weather Forecasts (ECMWF) dan diakses melalui *Copernicus Climate Data Store* (CCDS) [12]. Untuk memberikan gambaran awal mengenai struktur dan karakteristik data mentah yang diperoleh dari dataset ERA5, cuplikan lima baris pertama data meteorologi ditampilkan pada Tabel 1. Tabel ini menyajikan variabel atmosfer dalam format aslinya sebelum melalui proses konversi.

Tabel 1. Data Meteorologi

No	Waktu (valid_time)	u10	v10	d2m (K)	t2m (K)	msl (Pa)	tp	Latitude	Longitude
1	2019-01-01 00:00:00	0.2811	0.8540	294.39	295.45	101263.44	0.000121	-7.25	110.25
2	2019-01-01 01:00:00	0.3162	0.1236	294.98	296.11	101370.44	0.000028	-7.25	110.25
3	2019-01-01 02:00:00	0.0962	-0.1920	295.17	299.05	101350.44	0.000061	-7.25	110.25
4	2019-01-01 03:00:00	0.1526	-0.4728	296.26	299.79	101352.19	0.000209	-7.25	110.25
5	2019-01-01 04:00:00	0.6155	-0.8560	295.03	300.54	101278.19	0.000299	-7.25	110.25

Seperti yang terlihat pada Tabel 1, data awal memiliki resolusi per jam dengan unit standar meteorologi. Tahap pra-pemrosesan selanjutnya sangat krusial untuk mengonversi variabel-variabel tersebut seperti t2m menjadi Celcius dan tp menjadi milimeter serta mengagregasikannya menjadi format harian agar sesuai dengan target prediksi hujan harian. Tabel 1 tersebut mencakup berbagai variabel meteorologi seperti temperatur 2 meter (t2m), titik embun 2 meter (d2m), komponen angin U dan V pada ketinggian 10 meter (u10, v10), tekanan permukaan rata-rata laut (msl), serta total presipitasi (tp), sementara itu, Latitude menunjukkan garis lintang lokasi pengamatan yang menyatakan posisi wilayah terhadap garis khatulistiwa dan menjadi penentu letak geografis data meteorologi tersebut. [13]. Data mentah tersedia dalam resolusi per jam untuk satu titik koordinat geografis yang merepresentasikan wilayah Gunung Ungaran selama periode 1 Januari 2019 hingga 31 Desember 2023.

Tahap pra-pemrosesan data bertujuan untuk mengubah data awal ERA5 per jam menjadi format data harian yang siap digunakan untuk pemodelan. Proses ini diawali dengan konversi unit dan penanganan waktu, di mana kolom waktu dikonversi ke format datetime dan ditetapkan sebagai indeks time-series, sementara variabel seperti temperatur (Kelvin ke Celsius) dan total presipitasi (meter ke milimeter) dikonversi ke unit yang lebih umum. Curah hujan interval per jam dihitung dari data total presipitasi kumulatif. Selanjutnya, dilakukan agregasi harian dengan mengumpulkan data per jam menjadi data harian menggunakan fungsi statistik yang relevan, seperti nilai rata-rata (*mean*), minimum (*min*), dan maksimum (*max*) untuk temperatur harian, total (*sum*) untuk curah hujan harian, serta nilai rata-rata (*mean*) dan maksimum (*max*) untuk kecepatan angin harian. Terakhir, dilakukan penanganan nilai hilang (*missing values*) yang mungkin muncul setelah agregasi menggunakan metode interpolasi linier, diikuti dengan pengisian maju (*forward fill*) dan mundur (*backward fill*) untuk memastikan tidak ada data yang hilang.

2.2 Definisi Target dan Rekayasa Fitur

Target peramalan dalam penelitian ini adalah kejadian hujan pada hari berikutnya. Sebuah variabel biner baru, akan_hujan_besok, dibuat untuk tujuan ini. Variabel tersebut bernilai 1 jika total curah hujan pada hari berikutnya (t+1) melebihi ambang batas 1 mm, dan bernilai 0 jika tidak. Pemilihan ambang batas 1 mm didasarkan pada definisi umum hari hujan. Dengan mendefinisikan target sebagai kondisi hari berikutnya, penelitian ini secara jelas berfokus pada masalah peramalan (*forecasting*).

2.3 Pembagian dan Penyeimbangan Data

Untuk meningkatkan kemampuan model dalam menangkap pola temporal data cuaca, dilakukan rekayasa fitur temporal. Proses data mining ini bertujuan mengekstrak informasi historis [14]. Fitur-fitur baru dibentuk untuk memperkuat kemampuan model mengenali pola temporal. Lag Features dibuat dengan menambahkan nilai variabel cuaca utama seperti suhu, curah hujan, dan kecepatan angin dari 1, 3, dan 7 hari sebelumnya agar model memiliki "ingatan" terhadap kondisi sebelumnya. Rolling Aggregates dihitung sebagai rata-rata dan maksimum variabel cuaca dalam jendela waktu 3, 7, dan 14 hari terakhir guna menangkap tren cuaca jangka pendek hingga menengah. Sementara itu, *Time Features* (*Cyclical*) diperoleh dengan mengubah informasi waktu seperti bulan dan hari menjadi bentuk siklik menggunakan fungsi sinus dan kosinus agar model dapat memahami pola musiman. Setelah fitur dibuat, data yang mengandung nilai NaN akibat proses shift dan rolling dihapus.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Dataset yang telah diperkaya dengan fitur kemudian dibagi menjadi dua bagian: 80% data digunakan sebagai data pelatihan (*training set*), sementara 20% sisanya dijadikan data pengujian (*testing set*). Proses pembagian ini dilakukan secara terstratifikasi berdasarkan variabel target (akan_hujan_besok) agar proporsi antara kelas hujan dan tidak hujan tetap konsisten pada kedua set data. Mengingat adanya potensi ketimpangan jumlah antara hari hujan dan tidak hujan, teknik oversampling SMOTE (*Synthetic Minority Over-sampling Technique*) hanya diterapkan pada data pelatihan [15]. SMOTE menghasilkan sampel sintetis untuk kelas minoritas yang dalam penelitian ini kemungkinan merupakan kelas 'Hujan' hingga jumlahnya seimbang dengan kelas mayoritas, yang bertujuan mencegah model menjadi bias dan meningkatkan kemampuannya dalam mengenali kejadian hujan [16].

Penelitian ini menerapkan dan membandingkan kinerja dua algoritma ensemble learning yang populer untuk tugas klasifikasi. Algoritma pertama adalah Random Forest, sebuah metode berbasis *bagging* yang membangun banyak pohon keputusan secara independen dan menggabungkan hasilnya melalui voting [17], dengan parameter class_weight='balanced' digunakan untuk memberi bobot lebih pada kelas minoritas [18]. Adapun rumus model random forest sebagai berikut:

$$\hat{y} = \text{mode}\big(h_1(x), h_2(x), \dots, h_T(x)\big) \quad (1)$$

Rumus tersebut menunjukkan bahwa prediksi akhir $\hat{y}$ diperoleh dari hasil voting mayoritas (mode) seluruh pohon keputusan yang ada di dalam model Random Forest. Setiap fungsi $h_T(x)$ merepresentasikan prediksi dari pohon keputusan ke-t terhadap data masukan x. Jumlah pohon keputusan dalam model dinyatakan dengan T. Dengan menggabungkan banyak pohon keputusan yang dibangun dari subset data dan fitur yang berbeda, Random Forest mampu mengurangi varians model dan meningkatkan kestabilan serta akurasi prediksi [19].

Algoritma kedua adalah XGBoost (*Extreme Gradient Boosting*), sebuah metode berbasis boosting yang membangun pohon keputusan secara sekuensial untuk memperbaiki kesalahan prediksi dari pohon sebelumnya, dikenal karena regularisasi dan efisiensinya. [8]. Kedua model tersebut dilatih dengan memanfaatkan data pelatihan yang telah diseimbangkan terlebih dahulu menggunakan teknik SMOTE. Dan rumus model XGBoost sebagai berikut:

$$\widehat{y_i} = \sum_{t=1}^{T} f_t(x_i) \quad (2)$$

Rumus tersebut menyatakan bahwa nilai prediksi y_i untuk data ke-*i* diperoleh dari penjumlahan kontribusi seluruh pohon Keputusan $f_t(x_i)$ yang dibangun secara bertahap. Setiap pohon baru ditambahkan untuk memperbaiki kesalahan prediksi dari model sebelumnya melalui mekanisme boosting. Jumlah total pohon dinyatakan dengan T sedangkan x_i merepresentasikan data masukan ke-*i* Pendekatan ini memungkinkan XGBoost menghasilkan model dengan akurasi tinggi melalui proses optimasi bertahap yang efisien.

2.4 Pelatihan Model Random Forest & XGBoost

Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan pustaka Scikit-Learn dan XGBoost. Agar eksperimen dapat direproduksi (reproducible), parameter acak (random state) dikunci pada nilai 42 untuk semua tahapan proses, termasuk pembagian data dan inisialisasi model. Konfigurasi hiperparameter spesifik diterapkan secara berbeda pada kedua algoritma untuk memastikan kinerja yang optimal [20]. Untuk algoritma Random Forest, model dibangun dengan konfigurasi standar yang terdiri dari 100 pohon keputusan (n_estimators=100) serta menggunakan kriteria Gini Impurity untuk mengukur kualitas pemisahan pada setiap node. Pada model ini, tidak diterapkan pembatasan kedalaman pohon (max_depth=None), yang memungkinkan pohon tumbuh secara maksimal hingga seluruh daun menjadi murni atau berisi kurang dari sampel minimum pemisahan [9].

Sementara itu, algoritma XGBoost (Extreme Gradient Boosting) dikonfigurasi menggunakan booster berbasis pohon (gbtree) dengan metrik evaluasi Logarithmic Loss (eval_metric='logloss') untuk memantau kinerja model selama proses pelatihan. Model ini menggunakan laju pembelajaran (learning rate) standar sebesar 0.3 dan kedalaman maksimum pohon (max_depth) 6, sebuah konfigurasi yang bertujuan menyeimbangkan antara kompleksitas model dan risiko overfitting. Pemilihan kedua algoritma ini didasarkan pada karakteristiknya yang saling melengkapi, di mana Random Forest memanfaatkan teknik *bagging* untuk mengurangi varians, sedangkan XGBoost menggunakan teknik boosting untuk mengurangi bias. Pendekatan komparatif ini didukung oleh studi terbaru yang menunjukkan bahwa kombinasi kedua metode tersebut sangat efektif untuk menangani karakteristik data meteorologi yang kompleks [21].

2.5 Evaluasi Kinerja Model

Kinerja kedua model dievaluasi pada set data pengujian menggunakan metrik standar yang berasal dari *confusion matrix*. *Confusion matrix* merepresentasikan hasil klasifikasi ke dalam empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Metrik yang digunakan meliputi Akurasi, yaitu persentase keseluruhan prediksi yang benar [22]. Presisi, yaitu proporsi prediksi positif yang benar dari seluruh prediksi positif, berperan penting dalam meminimalkan terjadinya alarm hujan palsu [23]. Sementara itu, Recall (Sensitivitas), yaitu proporsi kejadian positif aktual yang berhasil terdeteksi, menjadi indikator penting agar model tidak melewatkan hari yang sebenarnya terjadi hujan [24]. F1-Score merupakan rata-rata harmonik dari presisi dan recall. Selain itu, *Confusion Matrix* digunakan untuk menunjukkan detail prediksi yang benar maupun salah pada setiap kelas, sementara Kurva ROC

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

(AUC) dipakai untuk mengevaluasi kemampuan model dalam membedakan masing-masing kelas. Analisis *Feature Importance* juga dimanfaatkan untuk mengidentifikasi fitur prediktor yang memiliki pengaruh paling besar. Seluruh metrik tersebut dibandingkan untuk menentukan model yang mampu memberikan performa peramalan paling optimal terhadap kondisi hujan harian di Gunung Ungaran.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN} \quad (6)$$

Kurva ROC menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). AUC merupakan luas area di bawah kurva ROC dan menunjukkan kemampuan model dalam membedakan kelas hujan dan tidak hujan.

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

$$AUC = \int_0^1 TPR(FPR)\ d(FPR) \quad (9)$$

Feature Importance digunakan untuk mengidentifikasi fitur yang paling berpengaruh dalam model prediksi. Nilai ini menunjukkan kontribusi relatif setiap fitur prediktor terhadap hasil klasifikasi.

$$FI_j = \frac{1}{T}\sum_{t=1}^{T} I_{j,t} \quad (10)$$

2.6 Analisis Komparatif & Kesimpulan

Tahap akhir dalam diagram alur adalah "Analisis Komparatif & Kesimpulan". Pada tahap ini, kinerja model Random Forest dan XGBoost dibandingkan secara head-to-head untuk menentukan model terbaik. Analisis difokuskan pada delta (selisih) peningkatan performa khususnya pada metrik *Recall* dan *F1-Score* setelah penerapan SMOTE. Model yang menunjukkan keseimbangan terbaik antara kemampuan mendeteksi hujan (Recall tinggi) dan ketepatan prediksi (Presisi stabil) akan direkomendasikan sebagai model paling robust untuk diterapkan di wilayah Gunung Ungaran.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Model

Penelitian ini menggunakan data meteorologi harian dari wilayah Gunung Ungaran yang meliputi suhu udara, kelembapan, tekanan udara, curah hujan, serta kecepatan angin sebagai variabel input untuk membangun model prediksi kejadian hujan dengan dua kelas, yakni Hujan dan Tidak Hujan. Sebelum digunakan, data menjalani serangkaian tahap pra-pemrosesan, termasuk pembersihan nilai kosong, normalisasi, serta pemisahan data menjadi data pelatihan dan data pengujian dengan rasio 80:20. Untuk mengatasi ketidakseimbangan jumlah antara kelas hujan dan tidak hujan, metode *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan agar distribusi data menjadi lebih seimbang, sehingga penelitian dilakukan dalam dua skenario, yaitu tanpa SMOTE dan dengan SMOTE. Model yang diterapkan adalah Random Forest dan XGBoost, yang dievaluasi menggunakan empat metrik utama akurasi, presisi, recall, dan F1-score yang menggambarkan ketepatan, sensitivitas, serta keseimbangan performa model dalam mengidentifikasi kedua kelas tersebut. Tabel berikut menyajikan hasil pemisahan data menjadi data latih dan data uji serta perubahan distribusi kelas sebelum dan sesudah penerapan metode *Synthetic Minority Oversampling Technique* (SMOTE). Sebelum dilakukan pelatihan model, distribusi kelas target diperiksa untuk mengidentifikasi tingkat ketidakseimbangan data. Perbandingan jumlah sampel data latih antara skenario awal (imbalanced) dan setelah penerapan teknik penyeimbangan data disajikan secara rinci pada Tabel 2.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 2. Data Sebelum dan Setelah SMOTE

Tahapan Data	Jumlah Data	Jumlah Fitur	Proporsi kelas 1	Proporsi kelas 2
Data Latih (sebelum SMOTE)	1.743	55	0,5754	0,4246
Data Uji	436	55	-	-
Data Latih (setelah SMOTE)	2.006	55	0,50	0,50

Berdasarkan data pada Tabel 2, terlihat jelas bahwa sebelum SMOTE, proporsi kelas minoritas (Hujan) hanya sebesar 42,46%. Penerapan SMOTE pada data latih berhasil menyeimbangkan distribusi kelas menjadi proporsi ideal 50:50 (0,50 banding 0,50), dengan total data latih meningkat menjadi 2.006 sampel. Penyeimbangan ini bertujuan agar model tidak memprioritaskan kelas mayoritas (bias) selama proses pembelajaran.

Penerapan metode SMOTE pada data latih berhasil menyeimbangkan distribusi kelas antara kejadian hujan dan tidak hujan, sehingga model dapat dilatih secara lebih optimal tanpa bias terhadap kelas mayoritas. Berdasarkan Tabel 3, pada skenario pertama (tanpa penggunaan SMOTE), model Random Forest menghasilkan akurasi sebesar 78,44% dengan nilai presisi 81,27%, recall 81,27%, dan F1-score 81,27%, yang menunjukkan performa yang cukup stabil dalam mendeteksi dan memprediksi kelas hujan. Di sisi lain, XGBoost memberikan hasil yang sedikit lebih unggul, dengan akurasi 78,67%, presisi 80,38%, recall 83,26%, dan F1-score 81,80%, menandakan kemampuannya yang lebih mumpuni dalam mempelajari pola data yang lebih kompleks. Setelah penerapan SMOTE, model Random Forest mencatat akurasi sebesar 78,21%, presisi 81,20%, recall 80,87%, serta F1-score 81,03%, yang menunjukkan adanya peningkatan yang relatif minim. Sebaliknya, model XGBoost mengalami peningkatan yang jauh lebih signifikan, dengan akurasi mencapai 80,50%, presisi 82,42%, recall 84,06%, dan F1-score 83,23%. Temuan ini mengindikasikan bahwa penerapan SMOTE mampu meningkatkan kemampuan XGBoost dalam mengenali kelas hujan yang sebelumnya menjadi kelas minoritas pada data pelatihan, tanpa mengurangi ketepatan prediksi secara keseluruhan.

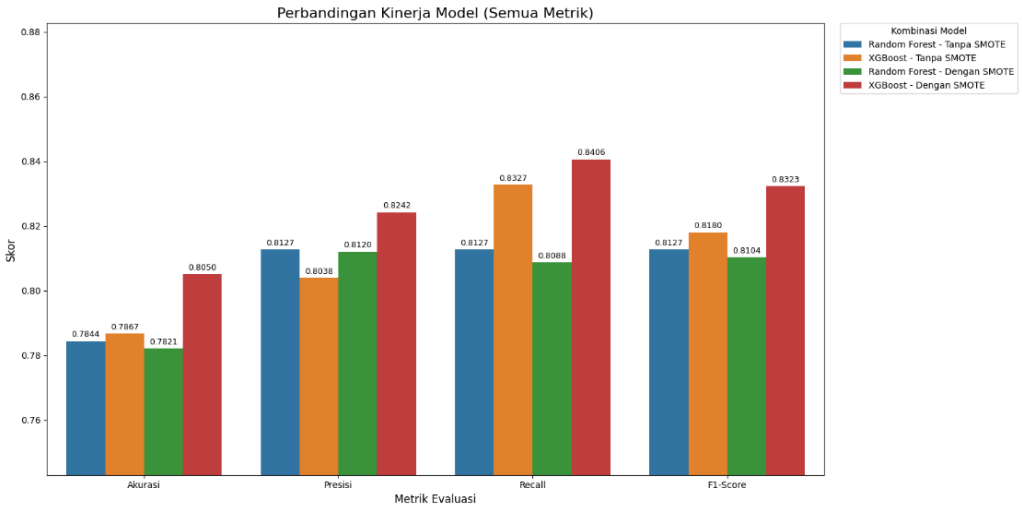
Untuk mengukur efektivitas model secara kuantitatif, dilakukan evaluasi kinerja menggunakan empat metrik utama: Akurasi, Presisi, Recall, dan F1-Score. Ringkasan hasil pengujian komparatif antara algoritma Random Forest dan XGBoost pada kedua skenario (dengan dan tanpa SMOTE) dirangkum dalam Tabel 3.

Tabel 3. Perbandingan Hasil Pengujian Model Random Forest dan XGBoost

Kondisi Data	Model	Akurasi	Presisi	Recall	F1-Score
Tanpa SMOTE	Random Forest	0.7844	0.8127	0.8127	0.8127
Tanpa SMOTE	XGBoost	0.7866	0.8038	0.8326	0.8180
Dengan SMOTE	Random Forest	0.7821	0.8120	0.8087	0.8103
Dengan SMOTE	XGBoost	0.8050	0.8242	0.8406	0.8323

Data pada Tabel 3 menunjukkan pola yang signifikan: XGBoost dengan SMOTE mencatatkan kinerja terbaik di seluruh metrik, mencapai akurasi 80,50% dan F1-Score 83,23%. Temuan penting dari tabel ini adalah lonjakan nilai Recall pada XGBoost setelah SMOTE menjadi 84,06%, yang mengindikasikan bahwa model ini jauh lebih sensitif dalam mendeteksi kejadian hujan dibandingkan Random Forest yang performanya cenderung stagnan meskipun data telah diseimbangkan.

Selain hasil numerik, visualisasi performa model juga disajikan untuk memberikan gambaran yang lebih jelas. Grafik perbandingan akurasi antara Random Forest dan XGBoost ditunjukkan pada Gambar 2, yang memperlihatkan bahwa XGBoost secara konsisten memiliki akurasi lebih tinggi pada kedua skenario.

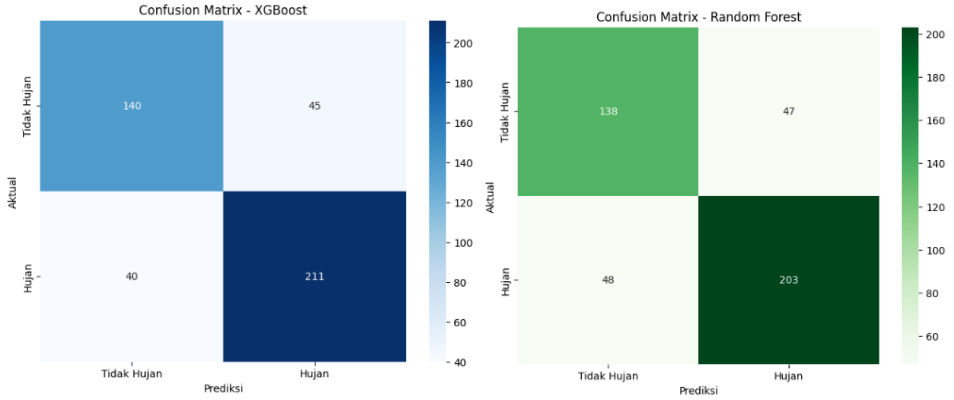


Gambar 2. Grafik Perbandingan Akurasi Model Random Forest dan XGBoost

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Visualisasi pada Gambar 2 mempertegas dominasi XGBoost (batang berwarna merah) dibandingkan model lainnya. Grafik tersebut menunjukkan bahwa peningkatan performa akibat penerapan SMOTE pada XGBoost terlihat konsisten pada setiap kategori metrik. Hal ini mengonfirmasi hipotesis bahwa algoritma boosting lebih adaptif dalam memanfaatkan data sintetis untuk memperbaiki batas keputusan (decision boundary) dibandingkan algoritma *bagging*.

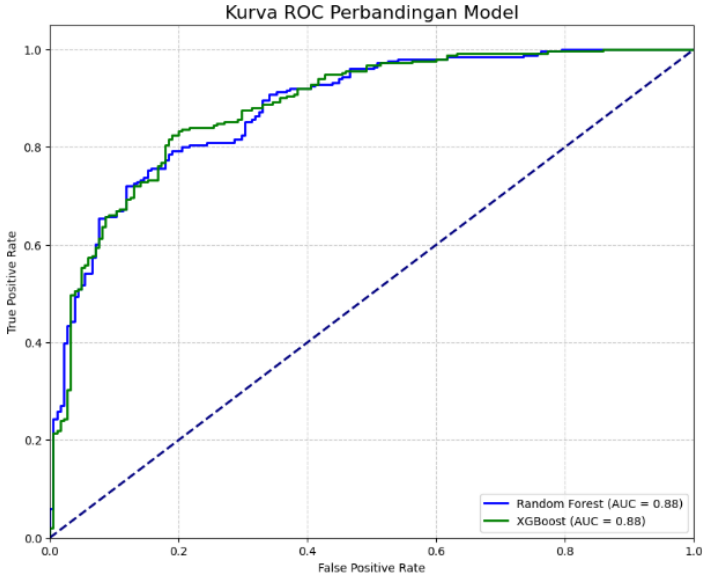
Selanjutnya, hasil confusion matrix untuk masing-masing model menunjukkan pola prediksi yang berbeda. Pada model XGBoost dengan SMOTE, jumlah kesalahan klasifikasi lebih sedikit dibandingkan dengan model lainnya, yang menunjukkan bahwa metode ini lebih efektif dalam mengenali data hujan dan tidak hujan secara seimbang. Untuk melihat lebih dalam detail kesalahan prediksi baik False Positive maupun False Negative distribusi hasil klasifikasi untuk model terbaik dipetakan menggunakan Confusion Matrix yang disajikan pada Gambar 3.



Gambar 3. Confusion Matrix Model Random Forest dan XGBoost Setelah SMOTE

Mengacu pada Gambar 3, khususnya pada matriks model XGBoost, terlihat bahwa jumlah prediksi yang benar (*True Positive* dan *True Negative*) lebih dominan dibandingkan kesalahan klasifikasi. Yang paling krusial, XGBoost mampu menekan angka False Negative (gagal mendeteksi hujan) lebih baik daripada Random Forest, yang berarti model ini memiliki risiko lebih kecil dalam konteks sistem peringatan dini bagi pendaki Gunung Ungaran.

Selain itu, kurva *Receiver Operating Characteristic* (ROC) juga digunakan untuk mengukur kemampuan model dalam membedakan antara dua kelas. Berdasarkan hasil evaluasi, nilai *area under the curve* (AUC) untuk model XGBoost maupun Random Forest sama-sama mencapai 0.88, yang menunjukkan performa sangat baik. Temuan ini mengindikasikan bahwa kedua model memiliki kemampuan diskriminatif yang setara dalam membedakan kelas target. Perbandingan kurva ROC untuk kedua model ditampilkan pada Gambar 4.



Gambar 4. Kurva ROC Model Random Forest dan XGBoost

Sebagaimana ditunjukkan oleh Gambar 4, kedua model menunjukkan performa yang sangat baik dengan nilai Area Under Curve (AUC) yang identik sebesar 0,88. Meskipun nilai AUC setara, bentuk kurva menunjukkan bahwa XGBoost memiliki laju True Positive Rate yang sedikit lebih stabil di awal kurva, mendukung temuan sebelumnya bahwa model ini lebih andal dan seimbang secara keseluruhan dibandingkan Random Forest.

Berdasarkan keseluruhan hasil pengujian dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa penerapan metode *Synthetic Minority Oversampling Technique* (SMOTE) memberikan pengaruh positif terhadap peningkatan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9416
Hal 192-201
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

performa model klasifikasi, khususnya pada algoritma XGBoost. Hasil eksperimen menunjukkan bahwa setelah diterapkan SMOTE, model XGBoost mengalami peningkatan kinerja yang cukup signifikan, ditunjukkan oleh kenaikan nilai akurasi dari 78,67% menjadi 80,50%, serta peningkatan nilai F1-score dari 0,8180 menjadi 0,8323. Peningkatan ini mengindikasikan bahwa model menjadi lebih seimbang dalam mengenali kedua kelas, terutama dalam mengidentifikasi kejadian hujan yang sebelumnya termasuk dalam kelas minoritas.

Sementara itu, algoritma Random Forest menunjukkan performa yang relatif stabil baik sebelum maupun sesudah penerapan SMOTE. Meskipun terdapat sedikit peningkatan pada beberapa metrik evaluasi, perubahan yang terjadi tidak sebesar yang dialami oleh XGBoost. Hal ini menunjukkan bahwa Random Forest memiliki ketahanan yang cukup baik terhadap ketidakseimbangan data, namun pada saat yang sama juga kurang responsif terhadap teknik oversampling yang bertujuan untuk menyeimbangkan distribusi kelas. Dengan kata lain, Random Forest cenderung mempertahankan pola pembelajaran awalnya meskipun proporsi kelas dalam data latih mengalami perubahan.

Secara konseptual, perbedaan respons kedua algoritma terhadap SMOTE dapat dijelaskan melalui mekanisme pembelajaran masing-masing. XGBoost menggunakan pendekatan boosting yang membangun model secara bertahap dengan berfokus pada kesalahan klasifikasi pada iterasi sebelumnya, sehingga model ini lebih adaptif dalam mempelajari pola dari data hasil oversampling. Kondisi ini memungkinkan XGBoost memanfaatkan data sintetis yang dihasilkan oleh SMOTE secara lebih efektif untuk meningkatkan kemampuan klasifikasi pada kelas minoritas. Sebaliknya, Random Forest yang mengandalkan pendekatan *bagging* dan pengambilan sampel acak cenderung kurang sensitif terhadap perubahan distribusi kelas, sehingga dampak SMOTE terhadap peningkatan performanya menjadi relatif terbatas.

Berdasarkan hasil tersebut, kombinasi antara algoritma XGBoost dan metode SMOTE dapat dianggap sebagai model yang paling optimal dalam penelitian ini. Kombinasi ini terbukti mampu meningkatkan kemampuan model dalam mengenali kejadian hujan tanpa mengorbankan akurasi pada kelas tidak hujan. Temuan ini menegaskan bahwa teknik penyeimbangan data memiliki peran penting dalam pengembangan model prediksi cuaca, khususnya pada wilayah dengan distribusi data yang tidak seimbang seperti Gunung Ungaran. Dengan demikian, penerapan SMOTE dapat direkomendasikan sebagai langkah praproses yang efektif untuk meningkatkan kinerja model prediksi hujan berbasis machine learning.

3.2 Pembahasan

Penelitian ini berhasil menghasilkan model prediksi hujan yang dibangun menggunakan dua algoritma *ensemble machine learning*, yaitu Random Forest dan XGBoost, melalui dua pendekatan pembelajaran: menggunakan data asli (tanpa SMOTE) serta data yang telah diseimbangkan menggunakan SMOTE. Hasil analisis menunjukkan bahwa kedua model mampu memprediksi kejadian hujan di wilayah Gunung Ungaran dengan tingkat akurasi yang cukup baik. Namun, performa keduanya tidak identik. XGBoost terbukti memberikan hasil yang lebih unggul dibandingkan Random Forest, terutama setelah proses penyeimbangan kelas dilakukan dengan SMOTE.

Perbedaan performa ini dapat dipahami dari cara kerja kedua algoritma tersebut. Random Forest membuat sejumlah pohon keputusan secara acak dan kemudian menggabungkan prediksinya melalui metode *bagging* untuk mengurangi variansi sekaligus meminimalkan risiko overfitting. Di sisi lain, XGBoost menerapkan teknik boosting, di mana model dibangun secara bertahap, dan setiap tahap berusaha memperbaiki kesalahan prediksi dari tahap sebelumnya. Proses gradient boosting menjadikan XGBoost lebih peka terhadap pola data yang kompleks dan bersifat non-linear karakteristik yang umum terdapat pada data meteorologi seperti suhu, kelembapan, tekanan udara, dan curah hujan. Dengan demikian, hasil penelitian ini memperlihatkan bahwa XGBoost lebih mampu menangkap hubungan antar variabel cuaca dibandingkan Random Forest, sehingga menghasilkan prediksi yang lebih stabil dan akurat.

Penerapan SMOTE (*Synthetic Minority Oversampling Technique*) terbukti berpengaruh positif terutama pada model XGBoost. Data cuaca umumnya bersifat tidak seimbang, di mana kejadian "tidak hujan" lebih sering terjadi daripada "hujan". Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas, sehingga meskipun akurasi tinggi, kemampuan mendeteksi hujan aktual bisa rendah. Dengan SMOTE, distribusi kelas menjadi seimbang, dan model dapat belajar dari representasi yang lebih kaya untuk kelas hujan. Hal ini terbukti meningkatkan nilai recall dan F1-score secara signifikan, terutama pada XGBoost. Secara teoretis, hasil ini sesuai dengan penjelasan [10], bahwa SMOTE efektif memperluas ruang fitur kelas minoritas sehingga memperbaiki sensitivitas model terhadap kelas yang jarang muncul.

Temuan penelitian ini juga didukung oleh hasil-hasil studi sebelumnya [4] dalam jurnal Prediksi Cuaca Kabupaten Sleman Menggunakan Algoritma Random Forest melaporkan bahwa Random Forest memiliki performa baik dengan nilai R^2 sebesar 0,691 dan MAE sebesar 0,060. Namun, penelitian tersebut berfokus pada regresi curah hujan, bukan klasifikasi kejadian hujan. Sementara itu, penelitian ini menggunakan pendekatan klasifikasi dengan XGBoost, yang terbukti lebih unggul dalam mengenali peristiwa biner "hujan" atau "tidak hujan", terutama setelah penerapan SMOTE untuk menyeimbangkan data. Dengan demikian, penelitian ini memperluas pendekatan sebelumnya melalui kombinasi metode boosting dan resampling untuk meningkatkan akurasi pada data meteorologi yang tidak seimbang.

Penelitian [2] dalam Prediksi Cuaca Kota Jakarta Menggunakan Random Forest menemukan bahwa algoritma tersebut mampu mencapai akurasi 0,71 dan ROC-AUC 0,92, menandakan kemampuannya dalam mengenali pola cuaca perkotaan. Namun, penelitian itu belum mempertimbangkan ketidakseimbangan data, sehingga kejadian langka seperti hujan ekstrem berpotensi diabaikan. Hasil penelitian ini menegaskan bahwa tanpa teknik penyeimbangan seperti SMOTE, performa model dalam mendeteksi kelas minoritas menurun. Setelah oversampling, model XGBoost menunjukkan peningkatan recall yang signifikan, yang penting bagi sistem peringatan dini cuaca.

Penelitian lain oleh [22] Dalam studi Analisis Model Prediksi Cuaca Menggunakan SVM, Gradient Boosting, Random Forest, dan Decision Tree, dilaporkan bahwa Random Forest mencapai akurasi tertinggi sebesar 85,1% pada data BMKG Jawa yang relatif seimbang. Berbeda dengan penelitian tersebut, studi ini menggunakan data dengan ketidakseimbangan kelas yang tinggi di wilayah Gunung Ungaran dan menunjukkan bahwa kombinasi XGBoost dan SMOTE mampu memberikan performa yang lebih stabil dibandingkan Random Forest. Temuan ini menjadi pembeda utama (GAP) sekaligus menegaskan kebaruan penelitian ini, bahwa peningkatan kinerja model tidak hanya ditentukan oleh pemilihan algoritma, tetapi juga oleh integrasi antara metode data balancing dan boosting untuk meningkatkan sensitivitas terhadap peristiwa meteorologis yang jarang terjadi.

Selain membuktikan efektivitas metode yang digunakan, penelitian ini juga memiliki makna geografis dan praktis yang penting. Wilayah Gunung Ungaran dengan topografi kompleks dan variasi mikroklimat yang tinggi memerlukan model yang mampu menangkap dinamika non-linear, seperti XGBoost, yang dapat menyesuaikan keputusan terhadap anomali lokal. Hal ini menunjukkan bahwa pendekatan machine learning lebih adaptif dibandingkan model statistik konvensional yang cenderung mengasumsikan hubungan linear antar variabel cuaca. Secara praktis, peningkatan recall pada XGBoost dengan SMOTE berperan penting dalam sistem peringatan dini, karena model dengan recall tinggi dapat mengurangi kesalahan deteksi hujan (false negative), sehingga meningkatkan kesiapsiagaan dan efisiensi dalam mitigasi bencana maupun pengelolaan pertanian.

Selain itu, penelitian ini menegaskan bahwa keberhasilan penerapan machine learning pada data meteorologi tidak hanya bergantung pada kompleksitas algoritma, tetapi juga pada strategi pra-pemrosesan data yang tepat. Oversampling dengan SMOTE yang dilakukan secara hati-hati terbukti meningkatkan kinerja model tanpa menyebabkan overfitting. Hal ini menunjukkan pentingnya keseimbangan antara metode pembelajaran dan kualitas data input.

Secara keseluruhan, penelitian ini memberikan kontribusi yang berarti baik secara teoretis maupun praktis. Dari aspek teoretis, temuan penelitian ini menegaskan bahwa penanganan ketidakseimbangan data merupakan elemen yang sangat penting dalam membangun model prediksi cuaca yang robust dan dapat diandalkan. Secara praktis, temuan ini menawarkan pendekatan yang dapat diterapkan oleh lembaga meteorologi dan kebencanaan untuk mengembangkan sistem prediksi hujan berbasis data historis yang lebih akurat, terutama di wilayah dengan topografi kompleks seperti Gunung Ungaran. Dengan demikian, kombinasi XGBoost dan SMOTE terbukti sebagai metode paling efektif dalam meningkatkan performa klasifikasi hujan serta memperkuat dan melengkapi temuan penelitian terdahulu dalam bidang prediksi cuaca berbasis machine learning di Indonesia.

4. KESIMPULAN

Berdasarkan analisis komparatif yang mendalam terhadap prediksi hujan harian di kawasan dengan topografi kompleks seperti Gunung Ungaran, penelitian ini menyimpulkan bahwa algoritma *Extreme Gradient Boosting* (XGBoost) menunjukkan superioritas kinerja yang signifikan dibandingkan Random Forest, terutama ketika diintegrasikan dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi masalah ketidakseimbangan data meteorologi. Temuan empiris membuktikan bahwa meskipun kedua model memiliki kemampuan dasar yang kompetitif, penerapan SMOTE memberikan dampak transformatif spesifik pada XGBoost, yang mencatatkan kinerja terbaik dengan akurasi mencapai 80,50% dan F1-score sebesar 83,23%, mengungguli model Random Forest yang kinerjanya cenderung stagnan di angka 78,21% meskipun telah melalui proses penyeimbangan data yang sama. Disparitas hasil ini menegaskan bahwa mekanisme boosting yang bekerja secara iteratif pada XGBoost jauh lebih adaptif dalam mengeksploitasi informasi dari data sintetis untuk memperbaiki sensitivitas model terhadap kelas minoritas dibandingkan mekanisme *bagging* pada Random Forest. Secara praktis, kontribusi utama penelitian ini adalah memvalidasi bahwa integrasi antara algoritma gradient boosting dan strategi resampling data merupakan pendekatan paling efektif untuk meminimalkan kegagalan deteksi hujan (*false negative*), yang merupakan aspek krusial dalam sistem peringatan dini demi keselamatan pendakian. Guna meningkatkan generalisasi model di masa depan, disarankan agar penelitian selanjutnya memperluas cakupan area pengamatan serta mengeksplorasi penerapan arsitektur *Deep Learning* seperti LSTM yang potensial untuk menangkap dinamika temporal cuaca secara lebih presisi.

REFERENCES

[1] R. Torhino and N. Andono, “Penerapan Algoritma Random Forest dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca,” Technology and Science (BITS), vol. 6, no. 3, pp. 1688–1699, 2024, doi: 10.47065/bits.v6i3.6404.

[2] Z. A. Dwiyanti and C. Prianto, “Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest,” Jurnal Tekno Insentif, vol. 17, no. 2, pp. 127–137, Oct. 2023, doi: 10.36787/jti.v17i2.1136.

[3] D. Sangaji and T. Sutabri, “Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla,” Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), vol. 5, no. 1, pp. 13–18, Apr. 2025, doi: 10.55382/jurnalpustakaai.v5i1.905.

[4] M. Taqiyuddin and T. Bayu Sasongko, “Prediksi Cuaca Kabupaten Sleman Menggunakan Algoritma Random Forest,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 8, no. 3, p. 1683, Jul. 2024, doi: 10.30865/mib.v8i3.7897.

[5] A. Syahreza, N. K. Ningrum, and M. A. Syahrazy, “Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost,” Edumatic: Jurnal Pendidikan Informatika, vol. 8, no. 2, pp. 526–534, Dec. 2024, doi: 10.29408/edumatic.v8i2.27640.

[6] A. Fauziah, H. Hermanto, and M. Akbar Sukmarini, "Extreme Gradient Boosting pada Peramalan Pola Curah Hujan Bulanan Kabupaten Banyuwangi," 2024.
[7] A. Hot Iman, F. Ready Permana, G. Putro Wardana, R. Kemmy Rachmansyah, and M. Mega Santoni, "Perbandingan Algoritma Klasifikasi Random Forest dan Extreme Gradient Boosting pada Dataset Cuaca Provinsi DKI Jakarta Tahun 2018," 2022, [Online]. Available: https://katalog.data.go.id/dataset/data-prakiraan-cuaca-wilayah-provinsi-dki-jakarta-tahun-2018.
[8] A. Suaif and E. Sylvianti Rahayu, "ANALISIS FAKTOR DAN POLA KEJADIAN BANJIR DI BANDAR LAMPUNG MENGGUNAKAN ARIMA, RANDOM FOREST, DAN XGBOOST," 2025.
[9] R. Irfannandhy, L. B. Handoko, and N. Ariyanto, "Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes," Edumatic: Jurnal Pendidikan Informatika, vol. 8, no. 2, pp. 714–723, Dec. 2024, doi: 10.29408/edumatic.v8i2.27990.
[10] R. K. Rachmansyah and R. Astriratma, "Implementasi Algoritma Extra Trees Untuk Klasifikasi Cuaca Provinsi Dki Jakarta Dengan Oversampling SMOTE," 2023. [Online]. Available: https://pypi.org.
[11] M. Alfian Fadillah and E. Dewi Sri Mulyani, "Komparasi Algoritma C4.5, Naive Bayes, K-Nearest Neighbor, Random Forest Untuk Prediksi Faktor Penyebab Penyakit Diabetes A B S T R A K I N F O A R T I K E L," 2024. [Online]. Available: https://ejournal.upi.edu/index.php/IJDB
[12] H. Hersbach et al., "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.
[13] J. Zhang et al., "Evaluation of Surface Relative Humidity in China from the CRA-40 and Current Reanalyses," Adv Atmos Sci, vol. 38, no. 11, pp. 1958–1976, Nov. 2021, doi: 10.1007/s00376-021-0333-6.
[14] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," JOINS (Journal of Information System), vol. 5, no. 1, pp. 10–17, May 2020, doi: 10.33633/joins.v5i1.2760.
[15] M. Dewi et al., "JIP (Jurnal Informatika Polinema) PENERAPAN SMOTE-NCL UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT JANTUNG KORONER," 2023.
[16] F. Fadmadika, H. H. Handayani, T. Al Mudzakir, and J. Indra, "PENGARUH SMOTE TERHADAP PERFORMA ALGORITMA RANDOM FOREST DAN ALGORITMA GRADIENT BOOSTING DALAM MEMPREDIKSI PENYAKIT STROKE," Jurnal Teknik Informasi dan Komputer (Tekinkom), vol. 7, no. 2, p. 837, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1575.
[17] D. Husen, D. Sandi, and S. Bumbungan, "Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma Random Forest Classifier," vol. 16, no. 1, 2022, [Online]. Available: https://journal.uniku.ac.id/index.php/ilkom
[18] G. Fibarkah et al., Prediksi Curah Hujan di Kabupaten Rembang dengan Model Random Forest, vol. 4, no. 1. 2023. [Online]. Available: https://webapi.bps.go.id/v1/api/list/model/data/lang/ind/domain/3317/var/488
[19] T. Mahesti, K. D. Hartomo, and S. Y. J. Prasetyo, "Penerapan Algoritma Random Forest dalam Menganalisa Perubahan Suhu Permukaan Wilayah Kota Salatiga," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 6, no. 4, p. 2074, Oct. 2022, doi: 10.30865/mib.v6i4.4603.
[20] M. Hermansyah, A. Saikhu, and B. Amaliah, "PEMODELAN DATA RADIOSONDE MENGGUNAKAN STACKING ENSEMBLE UNTUK KLASIFIKASI HUJAN," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 10, no. 2, pp. 1678–1687, May 2025, doi: 10.29100/jipi.v10i2.7806.
[21] D. D. N. Cahyo and A. Sunyoto, "Analisis Perbandingan Klasifikasi dalam Data Mining pada Prediksi Hujan dengan menggunakan Algoritma LSTM dan GRU," Jurnal Sains dan Informatika, vol. 11, no. 1, pp. 40–49, Jun. 2025, doi: 10.34128/jsi.v11i1.1212.
[22] W. Indrasari and H. Suhendar, "ANALISIS MODEL PREDIKSI CUACA MENGGUNAKAN SUPPORT VECTOR MACHINE, GRADIENT BOOSTING, RANDOM FOREST, DAN DECISION TREE," Prosiding Seminar Nasional Fisika (E-Journal), vol. XII, 2024, doi: 10.21009/03.1201.FA18.
[23] G. Ashari Rakhmat and W. Mutohar, "MIND (Multimedia Artificial Intelligent Networking Database Prakiraan Hujan menggunakan Metode Random Forest dan Cross Validation," Journal MIND Journal | ISSN, vol. 8, no. 2, pp. 173–187, 2023, doi: 10.26760/mindjournal.v8i2.173-187.
[24] N. A. Prakoso Indaryono, "ANALISA PERBANDINGAN ALGORITMA RANDOM FOREST DAN NAÏVE BAYES UNTUK KLASIFIKASI CURAH HUJAN BERDASARKAN IKLIM DI INDONESIA," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 9, no. 1, pp. 158–167, Feb. 2024, doi: 10.29100/jipi.v9i1.4421.