

PEMODELAN DATA RADIOSONDE MENGGUNAKAN STACKING ENSEMBLE UNTUK KLASIFIKASI HUJAN

Muhammad Hermansyah^{*1)}, Ahmad Saikhu²⁾, Bilqis Amaliah³⁾

1. Departemen Informatika, Institut Teknologi Sepuluh November, Surabaya, Indonesia
2. Departemen Informatika, Institut Teknologi Sepuluh November, Surabaya, Indonesia
3. Departemen Informatika, Institut Teknologi Sepuluh November, Surabaya, Indonesia

Article Info

Kata Kunci: data radiosonde; klasifikasi hujan; pembelajaran mesin; stacking ensemble

Keywords: *machine learning; precipitation classification; radiosonde data; stacking ensemble*

Article history:

Received 2 May 2025
Revised 15 May 2025
Accepted 19 May 2025
Available online 21 May 2025

DOI :

<https://doi.org/10.29100/jupi.v10i2.7806>

* Corresponding author.

Corresponding Author

E-mail address:

6025232007@student.its.ac.id

ABSTRAK

Perubahan iklim telah meningkatkan frekuensi dan intensitas kejadian cuaca ekstrem di wilayah tropis seperti Indonesia, sehingga menimbulkan tantangan dalam pemanfaatan data observasi meteorologi untuk mitigasi bencana hidrometeorologis. Data observasi permukaan sering kali kurang mampu merepresentasikan dinamika vertikal atmosfer dalam analisis kejadian cuaca ekstrem, seperti hujan sedang hingga lebat. Penelitian ini bertujuan mengembangkan model klasifikasi intensitas hujan berbasis data observasi udara atas dari radiosonde dengan pendekatan *stacking ensemble*, yang mengintegrasikan algoritma *Random Forest*, *XGBoost*, *LightGBM*, dan *SVM*, serta menggunakan *HistGradientBoosting* sebagai *meta-learner*. Untuk mengatasi ketidakseimbangan kelas antara kondisi berawan-hujan ringan dan hujan sedang-lebat, diterapkan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*. Evaluasi performa dilakukan menggunakan metrik *precision*, *recall*, *F1-score*, dan kurva *precision-recall*. Hasil menunjukkan bahwa model *stacking ensemble* memberikan performa terbaik dengan nilai *precision* sebesar 0,9084, *F1-score* 0,8718, dan *average precision* untuk kelas hujan sedang-lebat sebesar 0,949, melampaui seluruh model individual. Temuan ini menegaskan keunggulan integrasi data atmosfer vertikal dan pendekatan *multi-algorithm machine learning* dalam mendeteksi hujan intensitas sedang hingga lebat secara lebih akurat. Model ini memiliki potensi tinggi untuk diimplementasikan dalam sistem peringatan dini cuaca ekstrem, khususnya di wilayah tropis dengan keterbatasan data observasi permukaan.

ABSTRACT

Climate change has increased the frequency and intensity of extreme weather events in tropical regions such as Indonesia, posing challenges in utilizing meteorological observation data for hydrometeorological disaster mitigation. Surface observation data cannot often represent the vertical dynamics of the atmosphere in the analysis of extreme weather events, such as moderate to heavy rain. This research aims to develop a rain intensity classification model based on upper air observation data from radiosonde with a stacking ensemble approach, which integrates Random Forest, XGBoost, LightGBM, and SVM algorithms, and uses HistGradientBoosting as a meta-learner. To address the class imbalance between cloudy-light rain and moderate-heavy rain conditions, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. Performance evaluation used precision, recall, F1-score, and precision-recall curve metrics. The results show that the ensemble stacking model performs best with a precision value of 0.9084, an F1-score of 0.8718, and an average precision for the moderate-heavy rain class of 0.949, surpassing all individual models. This finding confirms the superiority of the integration of vertical atmospheric data and the multi-algorithm machine learning approach in more accurately detecting moderate to heavy intensity rain. This model has a high potential to be implemented in extreme weather early warning systems, especially in tropical regions with limited surface observation data.

I. PENDAHULUAN

Perubahan iklim global telah menyebabkan peningkatan frekuensi dan intensitas kejadian hujan lebat, terutama di wilayah tropis seperti Indonesia [1]. Fenomena ini berdampak terhadap peningkatan risiko bencana hidrometeorologis, seperti banjir dan tanah longsor. Bahkan dalam beberapa tahun terakhir

mengakibatkan kerugian besar terhadap infrastruktur, lingkungan, dan keselamatan masyarakat [2]. Oleh karena itu, kemampuan untuk mengklasifikasikan kondisi cuaca berdasarkan data observasi menjadi aspek penting dalam mendukung sistem peringatan dini dan mitigasi bencana.

Selama ini, data observasi permukaan dan citra satelit merupakan sumber utama dalam sistem prediksi operasional. Namun, data tersebut memiliki keterbatasan dalam menangkap struktur vertikal atmosfer dan dinamika pembentukan awan konvektif [3]. Sebaliknya, data pengamatan udara atas (*upper air*) seperti radiosonde menyediakan profil vertikal atmosfer yang lebih detail meliputi suhu, kelembapan, dan tekanan pada berbagai lapisan yang sangat berguna dalam mendeteksi dinamika pembentukan awan konvektif hingga terjadinya hujan [4]. Namun, pemanfaatan radiosonde untuk klasifikasi hujan ekstrem masih sangat terbatas, terutama di wilayah tropis. Sebagian besar studi sebelumnya lebih menitikberatkan pada data permukaan atau citra satelit dengan resolusi vertikal dan temporal yang rendah [5][6][7]. Bahkan penggabungan data radiosonde dalam sistem klasifikasi berbasis pembelajaran mesin masih jarang dilakukan untuk deteksi dini hujan ekstrem di wilayah tropis.

Meskipun kemajuan dalam pengumpulan data atmosfer dan metode pembelajaran mesin telah membuka peluang baru, klasifikasi intensitas hujan berdasarkan data radiosonde menghadapi sejumlah tantangan. Salah satu kendala utama adalah kompleksitas struktural dan hubungan nonlinier antara parameter atmosfer dengan kejadian hujan di permukaan sehingga sulit ditangkap dalam model statistik konvensional [8]. Selain itu, ketidakseimbangan distribusi data antara kondisi cuaca umum (berawan, hujan ringan) dan kejadian ekstrem (hujan sedang-lebat) menyebabkan penurunan akurasi model pada kelas minoritas [9]. Dalam upaya mitigasi bencana, kesalahan dalam mengidentifikasi hujan berdampak terhadap efektivitas sistem peringatan dini. Oleh karena itu, diperlukan pendekatan yang tidak hanya mampu mengatasi ketidakseimbangan kelas, tetapi juga mampu mengenali pola *nonlinier* data atmosfer vertikal yang kompleks [10]. Untuk menjawab tantangan tersebut, penelitian ini menggunakan pendekatan *stacking ensemble* yang mengintegrasikan beberapa model pembelajaran mesin dengan arsitektur berbeda, seperti *Random Forest*, *XGBoost*, *LightGBM*, dan *SVM* ke dalam satu model *meta learner*. Pendekatan ini tidak hanya mampu menangkap pola interaksi nonlinier antar parameter atmosfer vertikal yang secara individu memiliki korelasi rendah terhadap curah hujan, tetapi juga unggul dibandingkan model tunggal yang cenderung terbatas dalam mengenali pola multivariat kompleks. Dengan menyatukan kekuatan dari beberapa *tree-based algorithms* dan *kernel-based algorithms*, *stacking* membentuk keputusan kolektif yang lebih adaptif terhadap kompleksitas atmosfer tropis. Pendekatan ini terbukti memberikan performa klasifikasi terbaik dalam studi ini, melampaui semua model tunggal yang diuji.

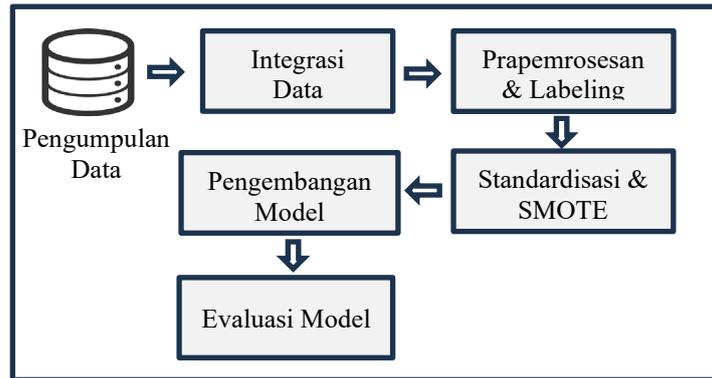
Dari sisi operasional, radiosonde dilaksanakan secara rutin pada pukul 00 dan 12 UTC di berbagai stasiun pengamatan [11]. Hal ini menjadikan model klasifikasi berbasis data radiosonde sangat potensial untuk diterapkan dalam sistem peringatan dini berbasis observasi aktual. Terlebih di wilayah Indonesia dengan keterbatasan akses radar cuaca, sistem klasifikasi ini dapat memberikan tambahan informasi yang kritis bagi pengambilan keputusan oleh BMKG serta instansi terkait seperti BPBD dalam upaya mitigasi bencana.

Secara teoretis, *stacking ensemble* cocok untuk memodelkan proses hierarkis pembentukan hujan yang kompleks di atmosfer [12]. Proses ini melibatkan tahapan mulai dari pemanasan permukaan, ketidakstabilan atmosfer (*CAPE*, *CIN*), pembentukan awan konvektif, hingga intensitas hujan yang masing-masing dipengaruhi oleh variabel atmosfer vertikal dengan hubungan *nonlinier* [13][14]. Dengan menggabungkan prediksi dari model-model yang memiliki kekuatan berbeda dalam menangkap pola-pola tersebut, *stacking* memungkinkan integrasi informasi dari berbagai sudut pandang sehingga mampu merepresentasikan dinamika fisis atmosfer secara lebih menyeluruh dibanding model tunggal.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi intensitas hujan berbasis pendekatan *stacking ensemble* dengan mengintegrasikan berbagai algoritma pembelajaran mesin pada data observasi atmosfer vertikal dari radiosonde. Fokus utama adalah membedakan kondisi cuaca berdasarkan intensitas curah hujan, sehingga dapat mendukung upaya mitigasi bencana dan peringatan dini [15][16]. Selain itu, penelitian ini bertujuan untuk mengevaluasi kinerja model dengan metrik yang sesuai untuk dengan data tidak seimbang, seperti *precision*, *recall*, *F1-score*, serta *precision-recall curve* dan *average precision* per kelas untuk memastikan kemampuan deteksi yang optimal terhadap kelas minoritas (hujan sedang-lebat) sebagai target utama.

II. METODE

Penelitian ini dirancang sebagai studi kuantitatif yang bertujuan untuk membangun model klasifikasi intensitas hujan berbasis pendekatan *stacking ensemble learning* [12]. Proses analisis (Gambar 1) mencakup tahapan mulai dari pengumpulan dan integrasi data, prapemrosesan data, penanganan ketidakseimbangan kelas, pemodelan klasifikasi menggunakan algoritma *ensemble*, hingga evaluasi kinerja model berdasarkan metrik yang relevan untuk klasifikasi dua kelas yang tidak seimbang. Seluruh tahapan dilakukan secara terstruktur menggunakan perangkat lunak pemrograman *Python* dengan pustaka *scikit-learn*, *xgboost*, *lightgbm*, dan *imbalanced-learn*.



Gambar 1. Alur Diagram Penelitian

A. Pengumpulan dan Integrasi Data

Langkah pertama dimulai dengan proses pengumpulan dan integrasi data meteorologis yang relevan untuk membangun model klasifikasi intensitas hujan berbasis pembelajaran mesin. Data diperoleh dari Badan Meteorologi Klimatologi dan Geofisika (BMKG), yaitu pengamatan udara atas dari radiosonde dan data curah hujan harian dari stasiun pengamatan permukaan. Untuk menjaga konsistensi dan keterpaduan temporal, kedua jenis data diselaraskan berdasarkan tanggal pengamatan. Proses integrasi ini menghasilkan dataset yang menggambarkan kondisi vertikal atmosfer yang terkait langsung dengan intensitas hujan harian di wilayah studi, yakni Kota Tarakan, Kalimantan Utara, selama periode observasi Januari 2015 hingga Desember 2024.

B. Prapemrosesan Data

Setelah data berhasil dikumpulkan dan diintegrasikan, langkah selanjutnya adalah proses prapemrosesan untuk memastikan bahwa dataset memiliki kualitas dan struktur yang sesuai untuk pelatihan model. Tahapan ini mencakup serangkaian proses, yaitu dengan konversi tipe data dari format *non-numerik* menjadi *numerik* untuk memastikan kompatibilitas dengan algoritma pemodelan [17]. Selanjutnya dilakukan identifikasi dan penanganan terhadap nilai yang hilang (*missing values*), serta pembersihan data dari entri yang tidak valid atau anomali yang dapat mengganggu performa model [18]. Untuk menjaga konsistensi dalam pemrosesan lebih lanjut, nama-nama variabel diubah agar bebas dari simbol-simbol khusus yang dapat menyebabkan kesalahan sintaksis. Proses prapemrosesan ini sangat penting karena kualitas input data secara langsung dapat mempengaruhi kemampuan generalisasi model klasifikasi [19].

C. Penetapan Label Klasifikasi

Langkah berikutnya dalam proses pemodelan adalah penetapan label klasifikasi yang menjadi target prediksi dalam pembelajaran mesin. Klasifikasi intensitas hujan diformulasikan sebagai masalah klasifikasi biner, dengan mendefinisikan dua kategori berdasarkan ambang nilai curah hujan harian. Kelas pertama (label 0) merepresentasikan kondisi berawan hingga hujan ringan dengan curah hujan harian tercatat di bawah 20 mm. Sementara itu, kelas kedua (label 1) mewakili kondisi hujan sedang hingga lebat dengan curah hujan sama dengan atau melebihi 20 mm. Ambang ini dipilih berdasarkan kriteria klimatologis yang umum digunakan dalam klasifikasi intensitas hujan di wilayah Indonesia [20]. Nilai curah hujan aktual kemudian dikonversi menjadi label biner sebagai variabel target. Sementara itu, seluruh fitur prediktor lainnya (variabel-variabel meteorologis dari data radiosonde) digunakan sebagai input model. Tahap ini menentukan struktur target dalam pelatihan model klasifikasi dan memastikan bahwa fokus utama model diarahkan pada pemisahan dua kondisi cuaca yang secara operasional memiliki dampak yang berbeda, terutama dalam konteks peringatan dini bencana hidrometeorologis.

D. Standardisasi dan Penanganan Ketidakseimbangan Data

Sebelum dilakukan pelatihan model, diperlukan transformasi lanjutan terhadap fitur *prediktor* untuk meningkatkan stabilitas dan efisiensi algoritma pembelajaran mesin. Tahapan ini dimulai dengan standardisasi seluruh variabel *numerik* menggunakan metode *z-score* melalui *StandardScaler* sehingga setiap fitur memiliki distribusi dengan nilai rata-rata nol dan standar deviasi satu [21]. Proses ini juga penting untuk menghilangkan skala yang tidak seragam antar fitur karena beberapa algoritma, seperti *Support Vector Machine (SVM)* dan *Gradient Boosting*, sensitif terhadap perbedaan skala input. Selanjutnya, untuk mengatasi ketidakseimbangan kelas pada klasifikasi kejadian hujan sedang-lebat (kelas minoritas) yang jarang terjadi, diterapkan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* [22]. Teknik ini menghasilkan sampel sintetis dari kelas minoritas berdasarkan

interpolasi data aktual, sehingga distribusi antar kelas menjadi lebih seimbang tanpa kehilangan representasi dari distribusi asli data. Untuk membuat satu sampel sintetik, misalnya dari kelas minoritas dipilih secara acak x_1 dan salah satu tetangga terdekatnya x_2 juga dipilih. Kemudian diambil sebuah nilai acak λ antara 0 dan 1. Sampel sintetik baru dihitung menggunakan persamaan (1). Kombinasi antara standardisasi dan penyeimbangan data ini tidak hanya meningkatkan performa klasifikasi, tetapi juga membantu model dalam mendeteksi kejadian hujan sedang-lebat sesuai tujuan utama dari penelitian ini.

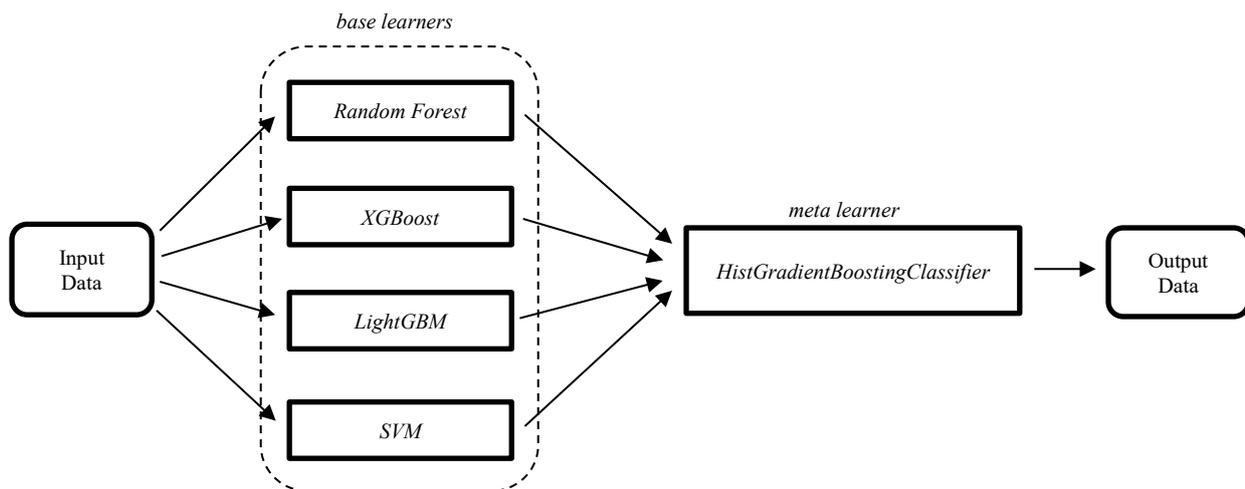
$$x_{new} = x_1 + \lambda \times (x_2 - x_1) \tag{1}$$

E. Pembagian Data

Setelah melalui proses standardisasi dan penyeimbangan kelas, tahap berikutnya adalah pembagian dataset menjadi subset data latih dan data uji untuk mendukung proses pelatihan dan evaluasi model secara valid. Dataset dibagi menggunakan metode *stratified random split*, dengan proporsi 80% sebagai data pelatihan dan 20% sebagai data pengujian. Teknik *stratification* digunakan untuk memastikan bahwa distribusi kelas tetap proporsional dalam kedua subset. Pemisahan ini juga dilakukan secara acak dengan penetapan nilai *random seed* yang konsisten. Pembagian data yang sistematis ini untuk menghindari data *leakage* dan *overfitting*, serta untuk memperoleh estimasi performa model yang lebih representatif [23]. Dengan demikian, langkah ini menjadi komponen penting dalam membangun keakuratan validasi model klasifikasi yang terpercaya.

F. Pengembangan Model Klasifikasi

Tahap inti dari penelitian ini adalah pengembangan model klasifikasi intensitas hujan berbasis pendekatan *stacking ensemble* yang dirancang untuk memanfaatkan keunggulan dari beberapa algoritma *machine learning* [12]. Dalam arsitektur model pada Gambar 2, terdapat empat algoritma digunakan sebagai *base learners* yaitu *Random Forest*, *XGBoost*, *LightGBM*, dan *Support Vector Machine (SVM)*. Algoritma ini dipilih berdasarkan kinerjanya dalam menangani data nonlinier [24]. Setiap *base learner* menghasilkan klasifikasi terhadap kelas target yang digunakan sebagai *input* bagi *meta learner* yakni *HistGradientBoostingClassifier*. Algoritma boosting modern ini mampu mengelola data heterogen secara efisien. Teknik *stacking* memungkinkan sistem pembelajaran tingkat kedua (*meta*) untuk mengevaluasi dan menggabungkan kelemahan serta kekuatan klasifikasi dari masing-masing model tunggal, sehingga memperbaiki performa keseluruhan [25]. Pendekatan ini juga diimplementasikan menggunakan skema *cross-validation* ($k=5$) untuk menjaga generalisasi model dengan konfigurasi *stack_method='predict_proba'* yang secara khusus dipilih untuk meningkatkan performa klasifikasi. Dengan desain tersebut, model *stacking* diharapkan mampu mengidentifikasi pola kompleks dalam data atmosfer vertikal secara lebih efektif dibandingkan model tunggal.



Gambar 2. Arsitektur Model

G. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk mengukur efektivitas pendekatan *stacking ensemble* dalam mengklasifikasikan hujan berdasarkan data atmosfer vertikal. Berdasarkan permasalahan klasifikasi dalam penelitian ini yang bersifat tidak seimbang, maka digunakan serangkaian metrik evaluasi yang mencerminkan performa model khususnya pada kelas minoritas (hujan sedang-lebat). Metrik utama yang digunakan meliputi

precision, *recall*, dan *F1-score*, yang memberikan gambaran keseimbangan antara ketepatan dan sensitivitas klasifikasi [26]. *Precision* (Persamaan (2)) didefinisikan sebagai rasio antara jumlah *true positives* terhadap jumlah *true positives* dan *false positives*. Metrik ini mengukur kemampuan model dalam mengidentifikasi kelas positif secara benar.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

Recall (Persamaan (3)) adalah rasio antara jumlah *true positives* terhadap jumlah *true positives* dan *false negatives*, yang mencerminkan kemampuan model dalam menangkap semua yang relevan dari kelas positif.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ negatives} \quad (3)$$

Pada dasarnya *precision* maupun *recall* tidak memperhitungkan *true negatives* dalam perhitungannya. Hal ini sesuai dengan fokus utama pada klasifikasi kelas minoritas yang sangat penting dalam penanganan dataset tidak seimbang. Kurva *precision–recall* yang memetakan *precision* pada sumbu-y dan *recall* pada sumbu-x, memberikan evaluasi yang lebih informatif terhadap performa algoritma dalam kondisi ketidakseimbangan kelas. Hal tersebut telah dibuktikan dalam berbagai studi sebelumnya [27][28].

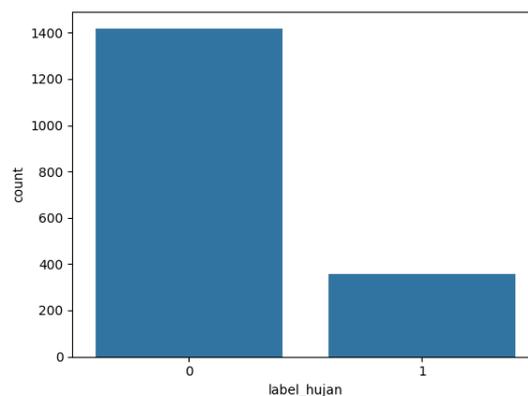
$$F1_score = \frac{2 \times (precision \times recall)}{precision + recall} \quad (4)$$

Precision dan *recall* sering kali menunjukkan hubungan *trade-off*, yaitu peningkatan salah satu metrik dapat menurunkan metrik lainnya [29]. Oleh karena itu, pemilihan evaluasi model harus mempertimbangkan prioritas tujuan penelitian. Misalnya dalam sistem pendeteksi cuaca ekstrem hujan sedang-lebat, penting untuk meminimalkan baik alarm palsu (*false positives*) maupun kejadian yang terlewat (*false negatives*). Untuk mengatasi *trade-off* ini, maka digunakan *F1-score* (Persamaan (4)) yang merupakan rata-rata harmonik antara *precision* dan *recall*. Metrik ini memberikan ukuran gabungan tunggal yang bernilai antara 0 hingga 1, dan sangat berguna ketika keseimbangan antara *precision* dan *recall* menjadi prioritas utama. Dengan pendekatan evaluasi yang beragam ini, diharapkan performa model tidak hanya tinggi berdasarkan angka, tetapi juga relevan secara praktis.

III. HASIL DAN PEMBAHASAN

A. Exploratory Data Analysis (EDA)

Eksplorasi data dilakukan untuk memahami karakteristik distribusi target dan fitur prediktor yang diperoleh dari pengamatan udara atas (radiosonde). Distribusi label klasifikasi pada gambar 3 menunjukkan adanya ketidakseimbangan kelas yang signifikan, sebanyak 1.419 sampel ($\pm 79,9\%$) dikategorikan sebagai cuaca berawan–hujan ringan (< 20 mm), dikodekan sebagai kelas 0. Sementara hanya 357 sampel ($\pm 20,1\%$) tergolong hujan sedang hingga lebat (≥ 20 mm), dikodekan sebagai kelas 1. Ketimpangan ini menunjukkan pentingnya penerapan teknik penyeimbangan kelas seperti SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menghindari bias model terhadap kelas mayoritas.



Gambar 3. Distribusi Kelas

Analisis statistik deskriptif terhadap parameter utama radiosonde (Tabel I) menunjukkan adanya variasi vertikal pada atmosfer tropis. Indeks konvektif seperti CAPE (*Convective Available Potential Energy*) memiliki nilai rata-rata sebesar 1719,71 J/kg, dengan maksimum mencapai 5075 J/kg. Sebaliknya, nilai rata-rata CIN (*Convective Inhibition*) berada pada -42,75. Nilai CAPE yang tinggi umumnya mengindikasikan potensi pembentukan awan konvektif dan hujan lebat. Namun CAPE di wilayah tropis sering kali bernilai tinggi secara umum sehingga variasinya terhadap curah hujan aktual menjadi kurang signifikan dalam skala harian, khususnya CAPE yang besar tanpa kehadiran mekanisme pemicu (*lifting*) seperti *konvergensi*. Sedangkan nilai CIN menggambarkan hambatan awal yang harus diatasi agar parcel udara dapat naik. Dalam atmosfer tropis yang relatif stabil dan lembap, CIN yang terlalu besar dapat menghambat proses konvektif meskipun nilai CAPE signifikan. Sebaliknya, CIN yang terlalu rendah dapat memicu konveksi prematur yang tidak cukup kuat membentuk hujan signifikan. Oleh karena itu, CAPE dan CIN harus ditafsirkan secara bersamaan sebagai satu sistem energi potensial dan hambatan, bukan sebagai prediktor tunggal. Hal ini menjelaskan mengapa korelasi linier antara kedua indeks tersebut terhadap intensitas hujan cenderung rendah, namun tetap relevan secara fisis.

TABEL I
 STATISTIK DESKRIPTIF PARAMETER UTAMA ATMOSFER RADIOSONDE

Variable	Count	Mean	Std	min	25%	50%	75%	max
Showalter index	1776	-0.34	1.86	-6.08	-1.47	-0.54	0.53	12.67
Lifted index	1776	-3.66	1.43	-7.8	-4.62	-3.74	-2.76	2.44
SWEAT index	1776	220.6	26.66	25	207.6	222.6	236.39	355.77
K index	1776	35.05	5.91	-6.3	33.9	36.5	38.3	43.3
Convective Available Potential Energy	1776	1719.71	858.23	1.35	1102.5	1680.44	2280.36	5075.25
Convective Inhibition	1776	-42.75	38.62	-304.14	-57.97	-33.66	-15.89	0
Mean mixed layer mixing ratio	1776	18.86	1.3	11.17	18.09	18.95	19.69	23.03
1000 hPa to 500 hPa thickness	1776	5799.81	25.62	4951.77	5792	5801	5809	5847
Precipitable water mm for entire sounding	1776	60.14	6.49	28.92	56.43	61.13	64.49	76.24

Indeks konvektif lainnya seperti *Lifted Index (LI)* menghitung perbedaan suhu parcel udara dengan lingkungan pada level 500 hPa, di mana nilai negatif menunjukkan ketidakstabilan. Namun karena parcel udara tropis cenderung hangat dan lembap, nilai LI negatif sering kali terjadi sepanjang tahun sehingga indeks ini memiliki sensitivitas rendah terhadap variasi hujan harian. Selanjutnya *K Index* yang menggabungkan suhu dan kelembapan di lapisan 850–700–500 hPa dan umumnya digunakan untuk memprediksi badai di wilayah lintang sedang. Di wilayah tropis yang umumnya lembap sepanjang tahun, nilai *K Index* cenderung tinggi secara konsisten, sehingga kurang efektif untuk membedakan hari dengan atau tanpa hujan lebat. Sementara itu, *SWEAT Index (Severe Weather Threat Index)* yang dirancang untuk lingkungan dengan variasi *wind shear* besar, yang umum terjadi di lintang sedang. Di atmosfer tropis yang memiliki shear vertikal lemah dan stabil, nilai *SWEAT* yang tinggi tidak selalu berkorelasi dengan hujan lebat, karena faktor angin vertikal tidak dominan dalam pembentukan hujan tropis.

Secara umum, korelasi linier (Tabel II) antara indeks konvektif dengan intensitas hujan cenderung rendah, dengan nilai tertinggi hanya 0,08 pada *precipitable water*. Hal ini memperkuat pemahaman bahwa hubungan antara parameter konveksi vertikal dan curah hujan di wilayah tropis bersifat kompleks, nonlinear, dan multivariabel. Indeks-indeks tersebut tidak dapat berdiri sendiri dalam menjelaskan kejadian hujan, karena masing-masing hanya mewakili satu aspek dari dinamika atmosfer yang lebih luas. Temuan ini dapat dikaitkan dengan sejumlah teori konveksi tropis, seperti *Moist Convective Adjustment* dan *Hot Tower Hypothesis* [30][31]. Teori-teori tersebut menjelaskan bahwa pembentukan hujan konvektif di wilayah tropis tidak hanya bergantung pada ketidakstabilan atmosfer (misalnya CAPE tinggi), tetapi juga pada keberadaan kelembapan vertikal yang cukup, level pemicu (seperti LFC), serta dinamika skala meso seperti *konvergensi*. Oleh karena itu, rendahnya korelasi linier antara parameter seperti CAPE, CIN, dan indeks konveksi lainnya terhadap intensitas hujan dalam studi ini tidak bertentangan dengan teori konveksi tropis, melainkan menguatkan pemahaman bahwa proses fisis pembentukan hujan bersifat nonlinear dan interdependen. Hal ini menjustifikasi pemilihan pendekatan *stacking ensemble* dapat menangkap interaksi kompleks antar variabel atmosfer dibandingkan model statistik linier konvensional.

TABEL II
 KORELASI TERHADAP LABEL_HUJAN

Variabel	Korelasi	Variabel	Korelasi
Precipitable water mm for entire sounding	0.08	1000 hPa to 500 hPa thickness	-0.01
K index	0.06	Convective Inhibition	-0.01
Pres hPa of the Lifted Condensation Level	0.06	Bulk Richardson Number	-0.01
Totals totals index	0.05	Bulk Richardson Number using CAPV	-0.01
Vertical totals index	0.05	CINS using virtual temperature	-0.02
SWEAT index	0.05	Convective Available Potential Energy	-0.02
Cross totals index	0.04	Equivalent potential temp K of the LCL	-0.02
LIFT computed using virtual temperature	0.02	CAPE using virtual temperature	-0.02
Temp K of the Lifted Condensation Level	0.02	LFCT using virtual temperature	-0.03
Lifted index	0.01	Level of Free Convection	-0.03
Mean mixed layer mixing ratio	0	Showalter index	-0.05
		Mean mixed layer potential temperature	-0.09

B. Analisis Perbandingan Metrik

Hasil evaluasi performa model klasifikasi (Tabel III) menunjukkan bahwa pendekatan *Stacking Ensemble* secara konsisten mengungguli model-model tunggal dalam hal *F1-score*, yang menjadi fokus utama dalam permasalahan klasifikasi tidak seimbang. Model *Stacking* mencatat *F1-score* tertinggi sebesar 0.8718, yang mengindikasikan keseimbangan optimal antara *precision* (0.9084) dan *recall* (0.8380). Hal ini menunjukkan bahwa *Stacking* mampu secara efektif mendeteksi hujan sedang–lebat dengan tingkat kesalahan minimum baik dari sisi *false positives* dan *false negatives*. Di sisi lain, model *LightGBM* dan *Random Forest* juga menunjukkan performa yang baik dengan *F1-score* masing-masing 0.8364 dan 0.8439, mencerminkan kekuatan model berbasis *decision tree* dalam menangani data multivariat dan nonlinearitas atmosferik. *XGBoost* berada sedikit di bawah dengan *F1-score* 0.8291, sedangkan model *SVM* menunjukkan performa paling rendah dengan *F1-score* 0.6877, yang mencerminkan keterbatasannya dalam memodelkan kompleksitas pola vertikal atmosfer pada data radiosonde.

TABEL III
 PERFORMA MODEL

Model	Precision	Recall	F1
Random Forest	0.7987	0.8944	0.8439
Xgboost	0.7980	0.8627	0.8291
Lightgbm	0.8026	0.8732	0.8364
SVM	0.6229	0.7676	0.6877
Stacking	0.9084	0.8380	0.8718

Analisis mendalam terhadap metrik evaluasi *precision*, *recall*, dan *F1-score* memberikan pemahaman yang lebih lengkap terhadap kekuatan serta keterbatasan masing-masing model klasifikasi intensitas hujan. Model *Stacking Ensemble* mencatat *precision* tertinggi sebesar 0.9084. Hal ini menunjukkan kemampuan luar biasa dalam meminimalkan *false positives* yang sangat penting dalam konteks sistem peringatan dini agar tidak menghasilkan alarm palsu. Namun demikian, *recall* (0.8380) sedikit lebih rendah dibandingkan *Random Forest* (0.8944), *Xgboost* (0.8627), dan *LightGBM* (0.8732). Hal ini menunjukkan bahwa ketiga model tersebut memiliki kemampuan lebih baik dalam menangkap seluruh kejadian hujan, meskipun dengan risiko menghasilkan lebih banyak *false positives*. Di sisi lain, model *SVM* menunjukkan *precision* dan *recall* yang terendah (0.6229 dan 0.7676). Hal ini berdampak langsung pada performa keseluruhan model sebagaimana terlihat dari nilai *F1-score* yang paling rendah (0.6877).

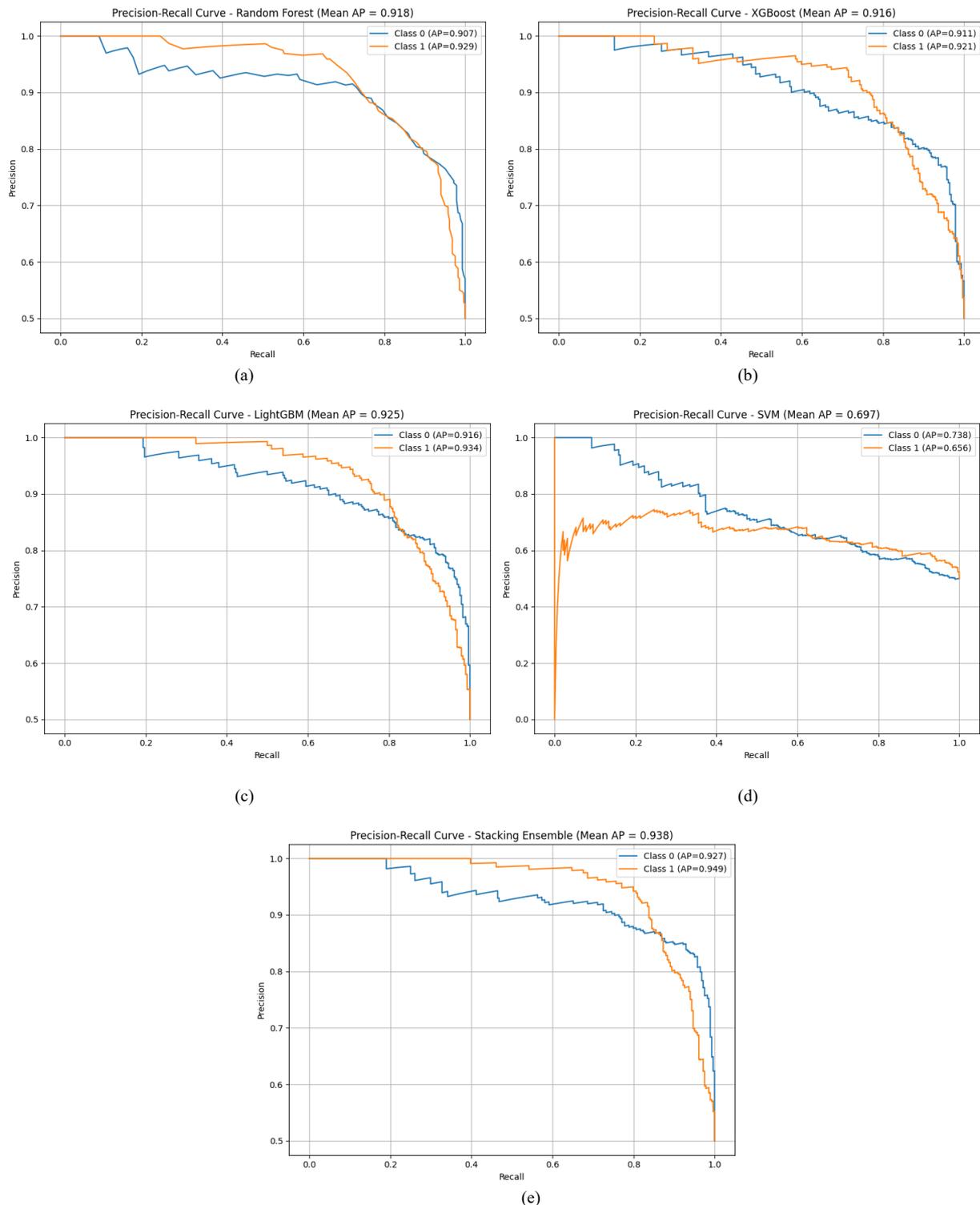
Trade-off antara *precision* dan *recall* ini menegaskan pentingnya pemilihan metrik evaluasi yang sesuai dengan kebutuhan aplikasi. Dalam konteks klasifikasi hujan yang berkaitan pada keselamatan publik, maka pendekatan yang mampu menjaga keseimbangan antara keduanya menjadi sangat krusial. Oleh karena itu, *F1-score* digunakan sebagai metrik utama karena mencerminkan harmonisasi antara kemampuan mendeteksi kejadian aktual dan menghindari kesalahan deteksi.

Analisis terhadap kelebihan dan keterbatasan masing-masing model klasifikasi mengungkapkan perbedaan signifikan dalam hal ketahanan terhadap ketidakseimbangan data serta kecermatan dalam mendeteksi kejadian hujan sedang hingga lebat. Model *Random Forest*, *LightGBM*, dan *XGBoost* menunjukkan performa yang bersaing, dengan *F1-score* yang relatif tinggi dan seimbang antara *precision* dan *recall*. Keunggulan utama dari ketiga model ini terletak pada kemampuannya menangani fitur *non-linier* serta ketahanan terhadap *overfitting*, khususnya dalam konteks data atmosfer vertikal yang kompleks. Namun demikian, model-model ini masih menghasilkan *trade-off* yang cukup jelas antara *recall* dan *precision*, sehingga kurang ideal dalam skenario yang membutuhkan tingkat akurasi dan sensitivitas yang sama tinggi. Model *SVM*, meskipun secara teoritis efektif untuk klasifikasi biner, terbukti memiliki keterbatasan dalam menangani data dengan distribusi tidak seimbang. Terlihat pada metrik evaluasi yang paling rendah di antara seluruh model. Ini menunjukkan bahwa *kernel-based* metode ini kurang mampu memetakan pola kompleks dari fitur atmosfer vertikal terhadap intensitas hujan. Sebaliknya, model *Stacking Ensemble* menunjukkan keunggulan dengan menggabungkan kekuatan prediktif dari model-model dasar, menghasilkan *precision* tertinggi (0.9084) sekaligus menjaga *F1-score* paling unggul (0.8718). Pendekatan *ensemble* ini mengurangi kelemahan individu setiap model dan meningkatkan generalisasi klasifikasi. Temuan ini menegaskan bahwa kombinasi model dapat memberikan keuntungan substansial dalam sistem klasifikasi cuaca berbasis data atmosfer vertikal.

C. Evaluasi Precision-Recall Curve (PRC)

Evaluasi performa model melalui *Precision-Recall Curve (PRC)* pada gambar 4 menunjukkan variasi kemampuan masing-masing algoritma dalam mendeteksi kelas minoritas (hujan sedang-lebat). Model *LightGBM* mencatat kinerja terbaik dengan *Mean Average Precision (AP)* sebesar 0.925, diikuti oleh *Random Forest* ($AP = 0.918$) dan *XGBoost* ($AP = 0.916$). Ketiganya menunjukkan *precision* dan *recall* yang tinggi secara konsisten pada kedua kelas dan mencerminkan kemampuan untuk menjaga *trade-off* yang optimal dalam klasifikasi intensitas hujan. Secara visual, kurva PRC dari ketiga model tersebut relatif stabil dan mendekati titik (1,1), yang menandakan performa klasifikasi sangat baik pada berbagai ambang batas keputusan (*decision thresholds*). Sementara itu, model *SVM* memperlihatkan performa yang lebih rendah dengan *mean AP* sebesar 0.697, serta menampilkan kurva yang cenderung fluktuatif khususnya pada kelas hujan sedang-lebat (kelas 1). Hal ini mengindikasikan bahwa *SVM* kurang mampu mempertahankan *precision* pada tingkat *recall* yang tinggi, yang sangat krusial dalam deteksi peristiwa kejadian hujan sedang-lebat. Secara keseluruhan, hasil ini memperlihatkan bahwa algoritma berbasis *ensemble tree* seperti *LightGBM*, *Random Forest*, dan *XGBoost* memiliki keunggulan dalam menangani ketidakseimbangan kelas dan memaksimalkan *precision* serta *recall* dalam konteks klasifikasi hujan berbasis data radiosonde. Evaluasi *Precision-Recall Curve (PRC)* pada model *stacking ensemble* menunjukkan performa yang terbaik dibandingkan seluruh model tunggal, dengan nilai *mean Average Precision (mean AP)* mencapai 0.938. Capaian ini menunjukkan kekuatan sinergi pada model *base learners* (*Random Forest*, *XGBoost*, dan *LightGBM*) yang diintegrasikan melalui *meta-learner* untuk mengoptimalkan keseimbangan antara *precision* dan *recall*. Pada grafik kurva untuk kelas 1 (kejadian hujan sedang-lebat) menunjukkan kestabilan dan konsistensi yang sangat baik, dengan nilai AP sebesar 0.949. Performa ini lebih tinggi dibandingkan model tunggal manapun. Kurva tersebut mendekati area ideal dengan nilai *precision* terbaik (>0.9) hingga pada tingkat *recall* yang sangat tinggi. Hal ini menandakan efektivitas model dalam menangkap sebanyak mungkin peristiwa hujan sedang-lebat dengan kesalahan minimum.

Dibandingkan dengan studi sebelumnya, pendekatan yang digunakan dalam penelitian ini menunjukkan keunggulan baik secara metodologis maupun hasil. Pratama et al. [32] menggunakan *Support Vector Machine (SVM)* dan teknik *Recursive Feature Elimination (RFE)* pada data permukaan untuk klasifikasi intensitas hujan, dengan capaian *F1-score* sebesar 0,81. Studi lain oleh Gu et al. [12] menerapkan pendekatan *stacking ensemble* untuk prediksi hujan bulanan di Lembah Taihu, Tiongkok dan menunjukkan peningkatan performa dibanding model tunggal, namun tidak menggunakan data vertikal atmosfer. Sementara itu, penelitian ini tidak hanya memanfaatkan data radiosonde resolusi vertikal tinggi untuk klasifikasi hujan harian, tetapi juga mencatatkan *F1-score* sebesar 0,8718 dan *average precision* sebesar 0,949 untuk kelas hujan sedang-lebat. Dengan demikian, studi ini memberikan kontribusi baru dalam mengintegrasikan data observasi udara atas dengan pendekatan pembelajaran mesin berbasis *ensemble* untuk mendukung sistem peringatan dini cuaca ekstrem di wilayah tropis.



Gambar 4. Precision-Recall Curve (a)Random Forest, (b)XGBoost, (c)LightGBM, (d)SVM, (e)Stacking

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi intensitas hujan harian berbasis data atmosfer vertikal (radiosonde) menggunakan pendekatan *stacking ensemble* yang terdiri dari kombinasi model dasar (*Random Forest*, *XGBoost*, *LightGBM*, dan *SVM*) dengan *meta-learner HistGradientBoostingClassifier*. Dalam menghadapi masalah ketidakseimbangan kelas, khususnya mendeteksi kejadian hujan sedang hingga lebat (≥ 20 mm), model yang diusulkan menunjukkan performa terbaik dibandingkan model-model tunggal. Hal ini dibuktikan melalui evaluasi menggunakan *precision*, *recall*, *F1-score*, serta *precision-recall curve* dan *average precision* per kelas, di mana *stacking ensemble* mencatatkan *precision* tertinggi (0,9084), *F1-score* tertinggi (0,8718), dan *average*

precision kelas hujan sedang-lebat mencapai 0,949. Temuan ini menegaskan bahwa integrasi pembelajaran mesin multi algoritma dengan data radiosonde dapat meningkatkan deteksi kejadian hujan sedang-lebat secara akurat.

Secara praktis, pendekatan ini memiliki potensi tinggi untuk diimplementasikan dalam sistem peringatan dini cuaca ekstrem di wilayah tropis, seperti yang dikelola oleh BMKG, khususnya di daerah dengan keterbatasan data permukaan namun memiliki cakupan radiosonde yang baik. Meskipun demikian, penelitian ini memiliki keterbatasan pada aspek generalisasi spasial dan interpretabilitas model. Oleh karena itu, disarankan agar studi lanjutan dilakukan untuk menguji kinerja model ini di wilayah lain dengan karakteristik atmosfer yang berbeda, serta mengintegrasikan metode *explainable*.

DAFTAR PUSTAKA

- [1] I. Iskandar *et al.*, *Weather and Climate Extreme Events in a Changing Climate*. IPCC_AR6_WGI_Chapter11, 2021. doi: 10.1017/9781009157896.013.1514.
- [2] D. D. Putra, B. F. Sofwany, H. Zulfinanda, and I. Kamaruddin, "Flash flood (Δ) risk and damage assessment in batu, East Java," *J. Teknosains*, vol. 12, no. 1, p. 72, 2022, doi: 10.22146/teknosains.78483.
- [3] Z. Song, W. Bai, Y. Zhang, Y. Wang, X. Xu, and J. Xin, "Evaluation of Satellite-Derived Atmospheric Temperature and Humidity Profiles and Their Application as Precursors to Severe Convective Precipitation," *Remote Sens.*, vol. 16, no. 24, 2024, doi: 10.3390/rs16244638.
- [4] F. M. Mashao *et al.*, "An Appraisal of the Progress in Utilizing Radiosondes and Satellites for Monitoring Upper Air Temperature Profiles," pp. 1–31, 2024.
- [5] B. Bochenek and Z. Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives," *Atmosphere (Basel)*, vol. 13, no. 2, pp. 1–16, 2022, doi: 10.3390/atmos13020180.
- [6] W. Han *et al.*, "Targeted sounding observations from geostationary satellite and impacts on high impact weather forecasts," *Sci. China Earth Sci.*, vol. 68, no. 4, pp. 963–976, 2025, doi: 10.1007/s11430-024-1489-5.
- [7] M. Putra, M. S. Rosid, and D. Handoko, "A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods," *Signals*, vol. 5, no. 3, pp. 542–561, 2024, doi: 10.3390/signals5030030.
- [8] H. Zhang, Y. Liu, C. Zhang, and N. Li, "Machine Learning Methods for Weather Forecasting: A Survey," *Atmosphere (Basel)*, vol. 16, no. 1, pp. 1–34, 2025, doi: 10.3390/atmos16010082.
- [9] X. Gao *et al.*, "A Comprehensive Survey on Imbalanced Data Learning," pp. 1–27, 2025, [Online]. Available: <http://arxiv.org/abs/2502.08960>
- [10] O. S. Riza and A. Nuryadi, "Bibliometric Study : Rainfall Classification - Prediction using Machine Learning Methods," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 14, no. 2, pp. 206–218, 2023, doi: 10.31849/digitalzone.v14i2.16618.
- [11] S. I. Rani, P. Sharma, J. P. George, and M. Das Gupta, "Assimilation of individual components of radiosonde winds: An investigation to assess the impact of single-component winds from space-borne measurements on NWP," *J. Earth Syst. Sci.*, vol. 130, no. 2, 2021, doi: 10.1007/s12040-021-01604-3.
- [12] J. Gu, S. Liu, Z. Zhou, S. R. Chalov, and Q. Zhuang, "A Stacking Ensemble Learning Model for Monthly Rainfall Prediction in the Taihu Basin, China," *Water (Switzerland)*, vol. 14, no. 3, pp. 1–20, 2022, doi: 10.3390/w14030492.
- [13] P. I. Khan *et al.*, "Observed Climatology and Trend in Relative Humidity, CAPE, and CIN over India," *Atmosphere (Basel)*, vol. 13, no. 2, pp. 1–13, 2022, doi: 10.3390/atmos13020361.
- [14] L. N. Zhang, D. J. S. Gianotti, and D. Entekhabi, "Land Surface Influence on Convective Available Potential Energy (CAPE) Change during Interstorms," *J. Hydrometeorol.*, vol. 24, no. 8, pp. 1365–1376, 2023, doi: 10.1175/JHM-D-22-0191.1.
- [15] B. Golding, *Towards the "Perfect" Weather Warning: Bridging Disciplinary Gaps through Partnership and Communication*. 2022.
- [16] W. F. P. Qatrinnada, E. Hidayah, G. Halik, and R. U. A. Wiyono, "A literature review: rainfall thresholds as flash flood monitoring for an early warning system," *Water Pract. Technol.*, vol. 19, no. 11, pp. 4486–4498, 2024, doi: 10.2166/wpt.2024.271.
- [17] P. Frederick, J. C. Finley, and C. Magalis, "A Quantitative Analysis for Non-Numeric Data," *Int. J. Quant. Qual. Res. Methods*, vol. 11, no. 1, pp. 1–11, 2023, doi: 10.37745/ijqqm13/vol11n1111.
- [18] A. Deji, C. Steve, and D. John, "A Systematic Literature Review on Data Cleaning Techniques and Their Impact on Machine Learning," no. December 2023, 2025.
- [19] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Orient. J. Comput. Sci. Technol.*, vol. 13, no. 0203, pp. 78–81, 2021, doi: 10.13005/ojst13.0203.03.
- [20] I. Gustari, T. W. Hadi, S. Hadi, and F. Renggono, "Akurasi Prediksi Curah Hujan Harian Operasional Di Jabodetabek : Perbandingan Dengan Model Wrf," *J. Meteorol. dan Geofis.*, vol. 13, no. 2, pp. 119–130, 2012, doi: 10.31172/jmg.v13i2.126.
- [21] C. Andrade, "Z Scores, Standard Scores, and Composite Test Scores Explained," *Indian J. Psychol. Med.*, vol. 43, no. 6, pp. 555–557, 2021, doi: 10.1177/02537176211046525.
- [22] A. S. Hussein, T. Li, C. W. Yohannese, and K. Bashir, "A-SMOTE: A new preprocessing approach for highly imbalanced datasets by improving SMOTE," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 2, pp. 1412–1422, 2019, doi: 10.2991/ijcis.d.191114.002.
- [23] S. Aymaz, *Unlocking the power of optimized data balancing ratios: a new frontier in tackling imbalanced datasets*, vol. 81, no. 2. Springer US, 2025. doi: 10.1007/s11227-025-06919-2.
- [24] R. Natras, B. Soja, and M. Schmidt, "Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting," *Remote Sens.*, vol. 14, no. 15, pp. 1–34, 2022, doi: 10.3390/rs14153547.
- [25] Aditya Gumilar, Sri Suryani Prasetyowati, and Yuliant Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 3, pp. 481–490, 2022, doi: 10.29207/resti.v6i3.4142.
- [26] A. de la Cruz Huayanay, J. L. Bazán, and C. M. Russo, "Performance of evaluation metrics for classification in imbalanced data," *Comput. Stat.*, vol. 40, no. 3, pp. 1447–1473, 2024, doi: 10.1007/s00180-024-01539-5.
- [27] L. Mdegela, E. Municio, Y. De Bock, E. Luhanga, J. Leo, and E. Mannens, "Extreme Rainfall Event Classification Using Machine Learning for Kikuletwa River Floods," *Water (Switzerland)*, vol. 15, no. 6, pp. 1–14, 2023, doi: 10.3390/w15061021.
- [28] J. Cook, "When to consult precision-recall curves," no. 1, pp. 131–148, 2020, doi: 10.1177/1536867X20909693.
- [29] Z. Erenel, "Improving the precision-recall trade-off in undersampling-based binary text categorization using unanimity rule," vol. 22, 2013, doi: 10.1007/s00521-012-1056-5.
- [30] F. Ahmed, Á. F. Adames, and J. D. Neelin, "Deep convective adjustment of temperature and moisture," *J. Atmos. Sci.*, vol. 77, no. 6, pp. 2163–2186, 2020, doi: 10.1175/JAS-D-19-0227.1.
- [31] J. Pilewskie, G. Stephens, H. Takahashi, and T. L'Ecuyer, "A Multi-satellite Perspective on 'Hot Tower' Characteristics in the Equatorial Trough Zone," *Surv. Geophys.*, vol. 45, no. 6, pp. 1933–1958, 2024, doi: 10.1007/s10712-024-09868-2.
- [32] A. R. I. Pratama, S. A. Latipah, and B. N. Sari, "Optimasi Klasifikasi Curah Hujan Menggunakan Support Vector Machine (Svm) Dan Recursive Feature Elimination (Rfe)," *JIFI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 2, pp. 314–324, 2022, doi: 10.29100/jifi.v7i2.2675.