

Analisis Intensitas Kenaikan Muka Air Sungai Bahalayung dan Sungai Balimau Menggunakan Metode Random Forest Classifier

Shofia Zulfa¹, Windarsyah², Finki Dona Marleny³

^{1,2,3} S1 Informatika, Universitas Muhammadiyah Banjarmasin, Indonesia

Email: ¹shofia_zulfa_2255201110018@umbjm.ac.id, ²windarsyah@umbjm.ac.id,

³finkidona@umbjm.ac.id

Korespondensi: shofia_zulfa_2255201110018@umbjm.ac.id

Info Artikel	Abstrak
Kata Kunci: Kenaikan Muka Air, <i>Random Forest Classifier</i> Intensitas Banjir Mitigasi Banjir Klasifikasi	Kenaikan muka air yang tidak terpantau di Kabupaten Hulu Sungai Selatan kerap menimbulkan banjir yang merugikan permukiman dan lahan pertanian sekitar sungai. Penelitian ini bertujuan menganalisis intensitas kenaikan muka air Sungai Bahalayung dan Sungai Balimau menggunakan algoritma <i>Random Forest Classifier</i> berbasis data hidrologi harian periode 2022–2025. Data diklasifikasikan ke dalam tiga kategori intensitas yaitu rendah, sedang, dan tinggi dengan rekayasa fitur berbasis riwayat ketinggian air dan evaluasi model menggunakan skema pembagian data kronologis. Akurasi terbaik dicapai sebesar 81,44% untuk model gabungan dan 82,82% untuk model Sungai Balimau, dengan fitur rasio ketinggian air mingguan dan volatilitas jangka pendek sebagai prediktor paling dominan. Sungai Bahalayung teridentifikasi lebih rawan lonjakan mendadak sehingga menjadi prioritas utama dalam upaya mitigasi banjir di wilayah tersebut..
Keywords: Water Level Rise <i>Random Forest Classifier</i> Flood Intensity Flood Mitigation Classification	Unmonitored water level surges in Hulu Sungai Selatan Regency frequently trigger floods that threaten residential areas and agricultural land along the riverbanks. This study aims to analyze the intensity of water level rise in Bahalayung River and Balimau River using a Random Forest Classifier algorithm based on daily hydrological data from 2022 to 2025. The data were classified into three intensity categories is low, moderate, and high through water level history-based feature engineering and chronological data splitting for model evaluation. The best accuracy reached 81.44% for the combined model and 82.82% for the Balimau River model, with the weekly water level ratio and short-term volatility features identified as the most dominant predictors. Bahalayung River was found to be more susceptible to sudden surges, making it the primary focus for flood mitigation efforts in the region..
JuKSIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License	
	

1. PENDAHULUAN

Banjir merupakan salah satu bencana alam yang paling sering terjadi di Indonesia. Berdasarkan Data Informasi Bencana Indonesia (DIBI) yang dikelola oleh Badan Nasional Penanggulangan Bencana (BNPB), frekuensi kejadian banjir menunjukkan tren yang terus meningkat dalam satu dekade terakhir, dari sekitar 525 kejadian pada tahun 2015 hingga hampir 1.800 kejadian pada tahun 2021 [1]. Meskipun sempat menurun menjadi 1.532 kejadian pada tahun 2022 dan 1.251 kejadian pada tahun 2023, angka tersebut kembali naik menjadi 1.457 kejadian pada tahun 2024. Tingginya frekuensi ini menempatkan banjir sebagai bencana yang tidak hanya bersifat musiman, tetapi juga berulang dengan skala yang semakin sulit diprediksi. Salah satu faktor yang memperparah dampak banjir adalah kenaikan muka air sungai yang berlangsung cepat dan tidak terpantau secara memadai, sehingga waktu yang tersedia untuk melakukan respons dan evakuasi menjadi sangat terbatas.

Kalimantan Selatan merupakan salah satu provinsi yang secara rutin menghadapi permasalahan banjir setiap tahunnya[2]. Kondisi geografis provinsi ini yang sebagian besar terdiri dari dataran rendah dan kawasan rawa menjadikannya sangat rentan terhadap kenaikan muka air sungai yang cepat, terutama pada periode curah hujan tinggi. Berdasarkan Dokumen Kajian Risiko Bencana Nasional Provinsi Kalimantan Selatan 2022–2026, banjir tercatat sebagai jenis bencana yang paling dominan dan berulang di wilayah ini, dengan sejarah kejadian yang terdokumentasi sejak

tahun 2009 [3]. Sungai-sungai yang mengalir di wilayah ini memiliki karakteristik yang sangat dipengaruhi oleh dinamika curah hujan di bagian hulu, sehingga perubahan intensitas hujan dalam waktu singkat dapat menghasilkan kenaikan muka air yang signifikan di bagian hilir dan memperluas dampak genangan ke permukiman warga di sekitarnya[4].

Kabupaten Hulu Sungai Selatan menjadi salah satu daerah di Kalimantan Selatan yang secara konsisten terdampak banjir setiap musim penghujan. Wilayah ini dialiri oleh sejumlah sungai yang pergerakannya sangat dipengaruhi oleh curah hujan di bagian hulu[5]. Sungai Bahalayung yang melintas di Desa Lungau, Kecamatan Kandangan, berada di wilayah dengan satuan geomorfologi dataran rendah aluvial yang bersumber dari pegunungan di bagian timur kabupaten, sehingga perubahan curah hujan di hulu dapat dengan cepat berdampak pada fluktuasi muka air di bagian hilir. Sementara itu, Sungai Balimau yang melintas di Desa Balimau, Kecamatan Kalumpang, berada pada wilayah yang secara geografis berbatasan langsung dengan Kabupaten Tapin dan merupakan bagian dari sistem aliran Sungai Kalumpang yang menjadi cabang dari Sungai Amandit[6]. Desa Balimau sendiri tercatat mengalami genangan hampir satu bulan penuh pada awal tahun 2025 akibat tingginya intensitas curah hujan, yang mengganggu mobilitas warga dan aktivitas ekonomi setempat [7]. Meskipun keduanya berada dalam satu kabupaten yang sama, perbedaan posisi geografis dan sistem aliran kedua sungai ini menciptakan pola respons yang tidak identik terhadap curah hujan, sehingga pemahaman yang spesifik terhadap masing-masing sungai menjadi penting sebagai dasar penanganan banjir yang lebih tepat sasaran.

Balai Wilayah Sungai Kalimantan III telah mengumpulkan data ketinggian air dan curah hujan dari kedua sungai tersebut secara rutin selama beberapa tahun terakhir. Data yang terkumpul sebenarnya menyimpan informasi yang cukup untuk memahami bagaimana pola kenaikan muka air terbentuk dan faktor apa yang mendahuluinya. Sayangnya, potensi tersebut belum dimanfaatkan secara maksimal karena analisis yang dilakukan masih terbatas pada pencatatan nilai harian tanpa pengolahan lanjutan yang lebih mendalam. Dengan *volume* data yang terus bertambah setiap harinya, pendekatan manual tidak lagi memadai untuk menangkap pola-pola perubahan yang tersembunyi di balik ribuan baris data tersebut. Akibatnya, pemahaman tentang karakteristik kenaikan muka air di kedua sungai masih belum cukup kuat untuk dijadikan landasan dalam pengambilan keputusan penanggulangan banjir yang lebih terencana dan tepat sasaran[8].

Perkembangan *machine learning* membuka peluang baru dalam analisis data hidrologi yang sebelumnya sulit dilakukan secara manual. Setiawan *et al.* menunjukkan bahwa Random Forest dapat digunakan untuk mengklasifikasikan tingkat risiko banjir ke dalam tiga kategori berbasis data ketinggian muka air dan curah hujan pada Sungai Kampar, dengan sistem yang mampu memberikan peringatan dini sebelum kondisi kritis tercapai[9]. Selain itu, Hermawan *et al.* menunjukkan bahwa model berbasis *Random Forest* dapat diterapkan untuk sistem prediksi ketinggian air sungai secara *real-time* dengan tingkat akurasi yang baik pada data sungai di Kalimantan Barat [10]. Purwati dan Pristyanto juga menegaskan bahwa *Random Forest* secara konsisten mengungguli metode klasifikasi lain dalam konteks prediksi banjir di Indonesia [11]. Penerapan *Random Forest* pada konteks hidrologi sungai di wilayah Kalimantan Selatan juga telah dibuktikan oleh Latifah *et al.*, yang berhasil membangun sistem prediksi kualitas air Sungai Barito di Pulau Curiak dengan akurasi 75% dan precision mencapai 94,23%, mengonfirmasi bahwa algoritma ini mampu bekerja secara efektif pada data sungai lokal meskipun dengan parameter yang terbatas[12]. Temuan-temuan tersebut secara keseluruhan memperkuat relevansi penggunaan algoritma *Random Forest* untuk permasalahan klasifikasi berbasis data historis, termasuk dalam konteks analisis perubahan muka air sungai.

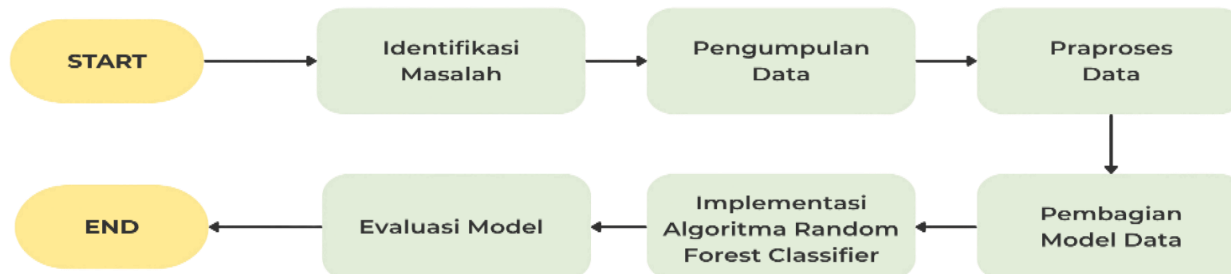
Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk membandingkan intensitas kenaikan muka air antara Sungai Bahalayung di Desa Lungau dan Sungai Balimau di Desa Balimau, Kabupaten Hulu Sungai Selatan, menggunakan algoritma *Random Forest Classifier*. Data ketinggian air dan curah hujan harian dari kedua sungai selama periode 2022 hingga 2025 digunakan sebagai dasar analisis. Intensitas kenaikan muka air diklasifikasikan ke dalam tiga kategori yaitu rendah, sedang, dan tinggi berdasarkan distribusi perubahan harian muka air masing-masing sungai. Melalui perbandingan hasil klasifikasi antara kedua sungai, penelitian ini bertujuan mengidentifikasi sungai mana yang lebih rawan mengalami kenaikan muka air secara tajam dan mendadak, sehingga hasilnya dapat dijadikan dasar rekomendasi prioritas penanggulangan banjir yang lebih tepat sasaran di Kabupaten Hulu Sungai Selatan.

2. METODOLOGI PENELITIAN

2.1 Lokasi dan Sumber Data

Penelitian menggunakan data sekunder hidrologi harian dari Balai Wilayah Sungai (BWS) Kalimantan III, meliputi data ketinggian muka air (TMA) dan curah hujan (CH) pada dua lokasi: Sungai Bahalayung di Desa Lungau dan Sungai Balimau di Desa Balimau, Kabupaten Hulu Sungai Selatan. Periode data mencakup 1 Januari 2022 hingga 31 Desember 2025, dengan total 2.922 observasi (masing-masing sungai 1.461 hari). Pengumpulan data dilakukan

melalui prosedur permohonan resmi kepada pihak BWS Kalimantan III dengan mengajukan formulir permintaan akses data. Data selanjutnya diterima dalam format digital dan dilakukan pencatatan ulang serta pembersihan data ke dalam format *spreadsheet* sebelum digunakan dalam analisis.



Gambar 1. Diagram alir proses pengambilan data

2.2 Deskripsi Data

Data yang digunakan merupakan data sekunder harian ketinggian muka air (TMA) dan curah hujan (CH) dari dua lokasi pemantauan, yaitu Sungai Bahalayung di Desa Lungau dan Sungai Balimau di Desa Balimau, Kabupaten Hulu Sungai Selatan, yang diperoleh dari Balai Wilayah Sungai (BWS) Kalimantan III. Periode data mencakup 1 Januari 2022 hingga 31 Desember 2025, dengan total 2.922 observasi (masing-masing sungai memiliki 1.461 hari pengamatan).

Statistik deskriptif tahunan dari kedua sungai secara agregat disajikan pada Tabel 1. Berdasarkan klasifikasi status air absolut (Surut < 2,50 m; Normal 2,50-2,78 m; Pasang > 2,78 m), seluruh data didominasi oleh kondisi Pasang. Namun, variabel target utama dalam penelitian ini bukanlah nilai absolut TMA, melainkan intensitas perubahan harian muka air (DELTA_TMA) yang mencerminkan laju kenaikan atau penurunan muka air, yang selanjutnya dikategorikan menjadi tiga kelas, yaitu rendah, sedang, dan tinggi berdasarkan distribusi kuantil data latih.

Tabel 1. Statistik Ketinggian Air dan Curah Hujan per Tahun

Tahun	Jumlah Hari(persungai)	KA Rata-rata (m)	CH Rata-rata (mm/hari)
2022	365	3,41	7,08
2023	365	2,67	6,27
2024	366	2,95	7,30
2025	365	2,98	5,81
Total	1.461	3,00	6,61

Penelitian ini dilaksanakan melalui enam tahapan utama, meliputi: (1) pengumpulan data TMA dan CH dari BWS Kalimantan III (2) rekayasa fitur berupa lag TMA, volatilitas, rasio lonjakan, curah hujan kumulatif, serta komponen temporal (3) pembagian data secara kronologis 80% data awal sebagai latih dan 20% akhir sebagai uji (4) pelabelan intensitas kenaikan muka air (DELTA_TMA) menjadi tiga kategori (rendah, sedang, tinggi) berdasarkan persentil ke-33 dan ke-66 yang dihitung dari data latih per sungai (5) pelatihan model *Random Forest Classifier* dengan penalaan hiperparameter menggunakan *TimeSeriesSplit* serta (6) evaluasi model menggunakan metrik akurasi, *F1-score*, dan kurva ROC-AUC.

2.3 Praproses Data dan Rekayasa Fitur

Data mentah yang diperoleh terlebih dahulu diurutkan berdasarkan tanggal secara kronologis. Tidak ditemukan nilai kosong (*missing value*) pada kedua variabel utama selama periode penelitian. Selanjutnya, dilakukan rekayasa fitur untuk menghasilkan variabel-variabel prediktor yang akan digunakan sebagai input model. Berdasarkan hasil eksperimen sebelumnya, fitur yang paling berpengaruh adalah riwayat ketinggian air (*lag*), volatilitas perubahan muka air, rasio lonjakan, serta curah hujan kumulatif dan ekstrem, sehingga seluruh fitur tersebut digabungkan dalam penelitian ini.

Variabel target (Y) dalam penelitian ini adalah intensitas kenaikan muka air (intensitas) yang dikategorikan menjadi tiga kelas, yaitu rendah, sedang, dan tinggi. Kategorisasi ini didasarkan pada nilai DELTA_TMA (perubahan ketinggian air hari ini terhadap hari sebelumnya), dengan ambang batas persenti ke-33 dan ke-66 yang dihitung secara

terpisah dari data latih masing-masing sungai.

Tabel 2 Table Variabel

Variabel	Jenis	Keterangan
TMA_LAG1, LAG2, LAG3, LAG7	Input (X)	Riwayat ketinggian air 1, 2, 3, dan 7 hari sebelumnya
DELTA_TMA_LAG1	Input (X)	Perubahan ketinggian air 1 hari sebelumnya
DELTA_TMA_ROLL3	Input (X)	Rata-rata perubahan ketinggian air 3 hari terakhir
DELTA_TMA_STDS	Input (X)	Standar deviasi perubahan ketinggian air 5 hari terakhir
TMA_ROLL3_MEAN, TMA_ROLL3_STD	Input (X)	Rata-rata dan standar deviasi TMA 3 hari terakhir
TMA_ROLL7_MEAN, TMA_RANGE7	Input (X)	Rata-rata TMA 7 hari terakhir
TMA_RATIO	Input (X)	Rasio TMA hari ini terhadap rata-rata 7 hari terakhir
CURAH HUJAN	Input (X)	Curah hujan hari ini
CH_ROLL3, CH_ROLL7	Input (X)	Rata-rata curah hujan 3 dan 7 hari terakhir
CH_KUMULATIF7	Input (X)	Total curah hujan 7 hari terakhir
CH_MAX7	Input (X)	Curah hujan maksimum dalam 7 hari terakhir
CH_RATIO	Input (X)	Rasio curah hujan hari ini terhadap rata-rata 7 hari terakhir
BULAN, MUSIM_ENC	Input (X)	Komponen waktu (bulan dan musim)
SUNGAI_ENC	Input (X)	Encoding nama sungai (khusus Model A)
INTENSITAS	Output (Y)	Kategori intensitas kenaikan muka air (rendah/sedang/tinggi)

2.4 Pembagian Data

Pembagian data dilakukan secara kronologis berdasarkan urutan tanggal untuk menjaga integritas temporal data deret waktu, sekaligus menghindari kebocoran informasi (*data leakage*) yang dapat terjadi pada pembagian acak[13]. Seluruh tanggal unik diurutkan, kemudian 80% data paling awal digunakan sebagai data latih (*training set*) dan 20% data paling akhir sebagai data uji (*testing set*). Berdasarkan perhitungan, titik potong (*split date*) jatuh pada 16 Maret 2025.

Pendekatan ini menghasilkan komposisi data yang seimbang untuk kedua sungai. Masing-masing sungai memiliki 1.163 hari data latih dan 291 hari data uji, atau total 2.326 baris data latih dan 582 baris data uji dari keseluruhan 2.908 observasi setelah rekayasa fitur.

Tabel 3. Pembagian Data Latih dan Data Uji

Subset	Periode	Jumlah Data (per sungai)	Total Data	Proporsi
Data Latih	1 Jan 2022 – 15 Mar 2025	1.163	2.326	80%
Data Uji	16 Mar 2025 – 31 Des 2025	291	582	20%
Total	1 Jan 2022 – 31 Des 2025	1.454	2.908	100%

2.5 Evaluasi Model

Penelitian ini merupakan permasalahan klasifikasi multi-kelas (tiga kategori intensitas), evaluasi kinerja model tidak menggunakan metrik regresi seperti MAE atau RMSE, melainkan metrik berbasis confusion matrix yang menilai ketepatan model dalam memprediksi kategori. Metrik utama yang digunakan meliputi akurasi, *presisi*, *recall*, dan *F1-score*, serta kurva ROC-AUC dan kurva *Precision-Recall* untuk mengevaluasi kemampuan diskriminasi model terhadap setiap kelas.

a. Confusion Matrix dan Metrik Turunannya

Confusion matrix adalah tabel yang membandingkan kelas aktual dengan kelas prediksi. Dari matriks ini, dihitung empat matrik utama untuk setiap kelas (di mana TP = *True Positive*, TN = *True Negative*, FP = *False Positive*, FN = *False Negative*):

1. Akurasi: Proporsi total prediksi yang benar terhadap seluruh data.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2. *Presisi*: Kemampuan model untuk tidak salah mengklasifikasikan kelas tertentu. Mengukur seberapa banyak prediksi positif yang benar.

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

3. *Recall*: Kemampuan model untuk menangkap seluruh instance dari kelas tertentu. Mengukur seberapa banyak data positif aktual yang berhasil diprediksi benar.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. *F1-Score*: Rata-rata harmonik dari presisi dan recall, yang memberikan keseimbangan antara keduanya. Sangat berguna ketika data tidak seimbang.

$$F1 = 2 \times \frac{TP}{TP + FN} \quad (4)$$

Karena terdapat tiga kelas (rendah, sedang, tinggi), perhitungan dilakukan secara *One-vs-Rest* (OvR) untuk setiap kelas, kemudian dirata-ratakan. Pada penelitian ini, *F1-Macro* (rata-rata sederhana F1 dari ketiga kelas tanpa mempertimbangkan jumlah sampel per kelas) dipilih sebagai metrik utama dalam proses hyperparameter tuning agar kelas tinggi (yang paling kritis untuk mitigasi banjir) tidak terabaikan meskipun jumlah sampelnya lebih kecil.

b. Kurva ROC dan AUC

Kurva ROC (*Receiver Operating Characteristic*) menggambarkan kinerja klasifikasi pada berbagai ambang batas, dengan memplot *True Positive Rate* (TPR) terhadap *False Positive Rate* (FPR). Nilai AUC (*Area Under Curve*) menunjukkan kemampuan model dalam membedakan antar kelas, di mana nilai mendekati 1 menunjukkan performa sangat baik. Untuk kasus multi-kelas, digunakan pendekatan *One-vs-Rest* (OvR) sehingga dihasilkan satu kurva ROC untuk masing-masing kelas terhadap gabungan kelas lainnya[14].

c. Kurva Precision-Recall

Kurva *Precision-Recall* digunakan sebagai metrik tambahan, terutama karena kemungkinan ketidakseimbangan distribusi kelas pada data uji. Kurva ini lebih informatif dibandingkan ROC pada kasus dengan kelas minoritas yang kritis, karena berfokus pada performa kelas positif (dalam hal ini kelas tinggi). Nilai *Average Precision* (AP) dihitung sebagai area di bawah kurva *precision-recall* untuk masing-masing kelas.

d. Validasi Silang Berbasis Waktu

Untuk memastikan stabilitas dan generalisasi model, proses validasi tidak menggunakan *cross-validation* acak biasa (seperti *KFold*), melainkan *TimeSeriesSplit* dengan 5 lipatan. Skema ini memastikan bahwa data pelatihan selalu berada di sebelum data validasi secara kronologis, sehingga menghindari kebocoran informasi dari masa depan ke masa lalu dan menghasilkan evaluasi yang lebih realistis untuk data deret waktu[15]. Skor rata-rata *F1-Macro* dari kelima lipatan digunakan sebagai indikator stabilitas model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Hyperparameter Tuning

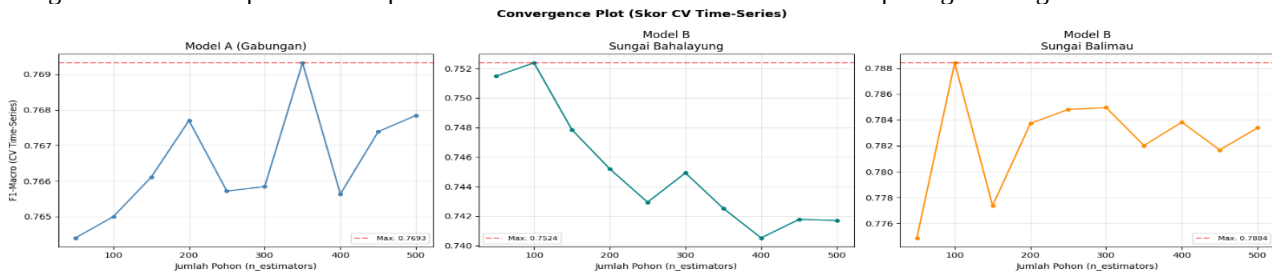
Proses tuning dilakukan menggunakan *RandomizedSearchCV* dengan skema validasi silang berbasis waktu (*TimeSeriesSplit*, *n_splits*=5) dan metrik optimasi *F1-Macro*. Pendekatan ini dipilih untuk memastikan proses tuning mengikuti urutan kronologis data sehingga tidak terjadi kebocoran informasi dari masa depan ke masa lalu. Hasil tuning untuk masing-masing model disajikan pada Tabel 4

Tabel 4. Hasil *hyperparameter tuning* terbaik per model

Model	n_estimator s	max_dept h	min_samples _leaf	min_samples_ split	max_featur es	F1-Macro CV
Model A (Gabungan)	300	None	2	6	None	0,7658
Model B – Bahalayung	100	None	5	6	sqrt	0,7524
Model B – Balimau	200	None	2	4	sqrt	0,7837

Convergence plot pada Gambar 1 memperlihatkan skor *F1-Macro CV* terhadap jumlah pohon. Model A mencapai nilai optimal pada *n_estimators*=350 (*F1-Macro CV*=0,7693), Model B Bahalayung pada *n*=100 (0,7524), dan Model B Balimau pada *n*=100 (0,7884). Secara umum kurva menunjukkan kestabilan setelah *n*=200, yang

mengonfirmasi bahwa penambahan pohon di atas nilai tersebut tidak memberikan peningkatan signifikan.



Gambar 2 Convergence Plot skor CV Time-Series per model

3.2 Evaluasi Kinerja Model

Evaluasi dilakukan pada data uji yang mencakup 20% periode paling akhir (sejak 16 Maret 2025), yang belum pernah dilihat oleh model selama pelatihan maupun tuning. Hasil evaluasi disajikan pada Tabel 5.

Tabel 5. Tabel Evaluasi Kinerja Model

Model	Akurasi	F1-Macro	Precision	Recall	F1-Score	AUC	AP
Model A (Gabungan)	81,44%	81,58%	0,90	0,79	0,84	0,958	0,867
Model B - Bahalayung	76,29%	76,27%	0,90	0,60	0,72	0,935	0,815
Model B - Balimau	82,82%	82,64%	0,90	0,88	0,89	0,952	0,902

Model B Sungai Balimau mencapai akurasi tertinggi sebesar 82,82% dengan F1-Macro 0,8264. Model A Gabungan memperoleh akurasi 81,44% dengan F1-Macro 0,8158, sedangkan Model B Sungai Bahalayung mencatat akurasi terendah sebesar 76,29%. Secara keseluruhan, dua dari tiga model telah melampaui ambang batas akurasi 80%.

Kelas SEDANG secara konsisten memiliki nilai F1-Score terendah di seluruh model. Hal ini disebabkan oleh ambiguitas batas kelas: nilai delta TMA di perbatasan RENDAH-SEDANG atau SEDANG-TINGGI lebih sulit dibedakan secara deterministik oleh model dibandingkan kelas ekstrem.

3.3 Analisis Kinerja per Model dan Implikasinya terhadap Mitigasi Banjir

Berdasarkan hasil evaluasi pada data uji, ketiga model menunjukkan performa yang bervariasi dengan karakteristik masing-masing yang relevan untuk aplikasi peringatan dini banjir.

a. Model A Gabungan

Model A mencapai akurasi 81,44% dengan F1-Macro 81,58%. Recall kelas TINGGI sebesar 0,79 menunjukkan bahwa model mampu mendeteksi 79 dari 100 kejadian kenaikan ekstrem, sementara precision kelas TINGGI sebesar 0,90 mengindikasikan tingkat kesalahan prediksi positif yang rendah. Keunggulan utama Model A terletak pada kesederhanaan implementasi karena hanya memerlukan satu model untuk dua sungai sekaligus, sehingga lebih praktis diterapkan apabila salah satu sungai kekurangan data historis[16]. Dengan performa yang kompetitif, Model A dapat difungsikan sebagai model cadangan atau digunakan pada lokasi sungai baru yang belum memiliki data historis yang memadai.

b. Model B Sungai Bahalayung

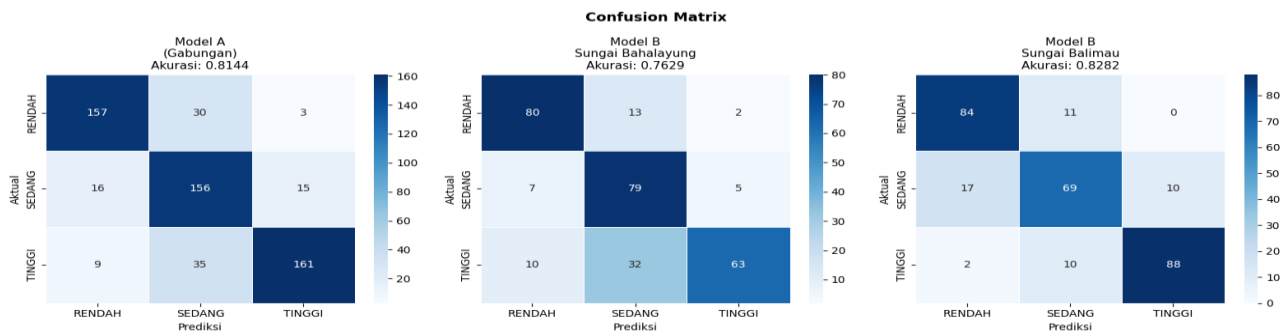
Model B Sungai Bahalayung mencatat akurasi 76,29% dengan recall kelas TINGGI hanya 0,60, yang berarti 40 dari 100 kejadian kenaikan ekstrem tidak terdeteksi oleh model. Tingkat miss-detection sebesar 40% ini berpotensi mengurangi efektivitas sistem peringatan dini, mengingat kejadian kenaikan TMA yang tidak terdeteksi secara tepat waktu dapat mempersempit waktu respons evakuasi masyarakat[17]. Performa yang lebih rendah ini berkaitan dengan karakteristik hidrologis Sungai Bahalayung yang memiliki variabilitas kenaikan TMA lebih tinggi (standar deviasi DELTA_TMA $\approx 0,54$ m) dibandingkan Sungai Balimau ($\approx 0,25$ m), sehingga batas antar kelas intensitas menjadi lebih kabur dan lebih sulit dipisahkan oleh model [18]. Meskipun demikian, precision kelas TINGGI tetap tinggi (0,90), yang berarti prediksi TINGGI yang dihasilkan model tetap dapat diandalkan ketepatan-nya.

c. Model B Sungai Balimau

Model B Sungai Balimau menunjukkan performa terbaik di antara seluruh model yang diuji, dengan akurasi 82,82% dan F1-Macro 82,64%. Recall kelas TINGGI sebesar 0,88 berarti 88 dari 100 kejadian kenaikan ekstrem berhasil terdeteksi, menjadikan model ini paling andal untuk keperluan deteksi dini. Nilai Average Precision (AP)

kelas TINGGI sebesar 0,902 mengonfirmasi bahwa model mampu mempertahankan presisi tinggi pada berbagai ambang keputusan, sehingga meminimalkan *false alarm* yang berlebihan, kondisi yang apabila terlalu sering terjadi dapat menurunkan kepercayaan masyarakat terhadap sistem peringatan. Dengan demikian, Model B Sungai Balimau layak menjadi model utama dalam sistem peringatan dini banjir di wilayah Sungai Balimau, Kabupaten Hulu Sungai Selatan.

3.4 Analisis Confusion Matrix dan Classification Report

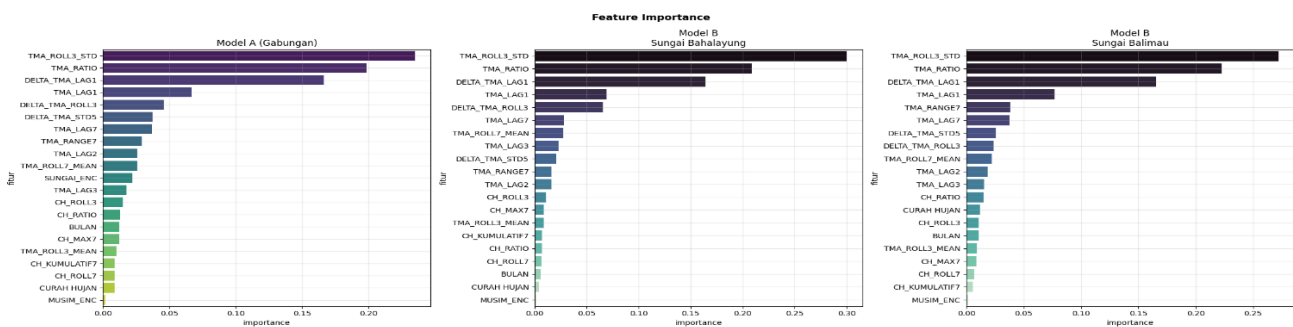


Gambar 3 Confusion Matrix ketiga model

Berdasarkan confusion matrix pada Gambar 3, kesalahan klasifikasi pada ketiga model secara konsisten terpusat di kelas SEDANG. Hampir tidak ada kesalahan silang langsung antara kelas RENDAH dan TINGGI, yang menunjukkan model mampu membedakan dua kondisi ekstrem dengan baik. Kondisi ini merupakan konsekuensi struktural dari pembagian label berbasis kuantil, di mana batas kelas SEDANG dengan kelas di sekitarnya tidak memiliki pemisah yang tegas secara hidrologis.

Yang paling kritis untuk keperluan penanggulangan banjir adalah perbandingan *recall* kelas TINGGI antar sungai. Model B Sungai Balimau berhasil mendeteksi 88 dari 100 kejadian intensitas tinggi (*recall* 0,88), sementara Model B Sungai Bahalayung hanya mendeteksi 63 dari 105 kejadian (*recall* 0,60). Perbedaan ini mencerminkan karakteristik hidrologis Sungai Bahalayung yang lebih volatil, sehingga batas antara intensitas SEDANG dan TINGGI lebih sulit ditangkap model. Peningkatan *recall* kelas TINGGI pada Sungai Bahalayung menjadi prioritas utama pengembangan model selanjutnya.

3.5 Analisis Feature Importance



Gambar 4 Feature Importance

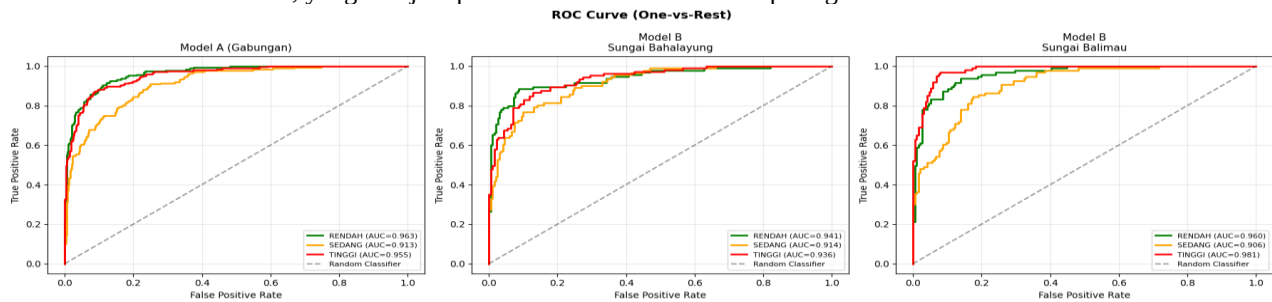
Analisis *feature importance* menunjukkan bahwa TMA_RATIO menjadi fitur paling dominan di seluruh model dengan nilai kepentingan 0,191-0,216, melampaui TMA_LAG1 yang mendominasi pada eksperimen sebelumnya. Temuan ini mengonfirmasi bahwa lonjakan relatif TMA terhadap *baseline* mingguan memberikan sinyal yang lebih tajam tentang potensi kenaikan ekstrem dibandingkan nilai ketinggian air secara absolut ketika TMA_RATIO melampaui 1,2, probabilitas kelas TINGGI meningkat drastis.

DELTA_TMA_STD5 dan CH_MAX7 juga menunjukkan kepentingan yang signifikan, menegaskan bahwa volatilitas perubahan TMA dan intensitas curah hujan puncak merupakan dua faktor kunci dalam mendeteksi kenaikan ekstrem. Sebaliknya, MUSIM_ENC dan SUNGAI_ENC secara konsisten mencatat kepentingan terendah di bawah 0,03, mengindikasikan bahwa kategori musim kalender tidak informatif dibandingkan dinamika air itu sendiri hal yang relevan mengingat curah hujan di Kalimantan Selatan dapat terjadi sepanjang tahun. Temuan ini memberikan arahan bahwa sistem peringatan dini banjir akan lebih efektif jika berbasis fitur dinamis *real-time* dibandingkan variabel

musiman yang statis.

3.6 Analisis ROC-AUC

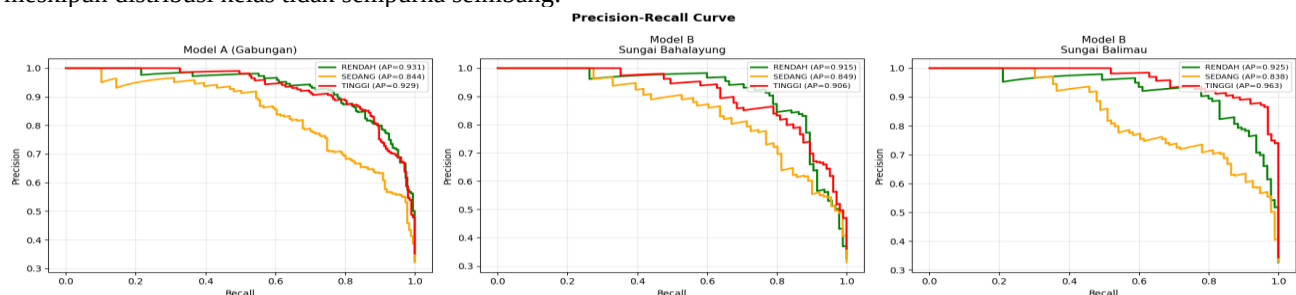
Hasil evaluasi kurva ROC pada Gambar 4 menunjukkan bahwa model memiliki daya pisah yang baik antartetiga kelas intensitas. Kelas TINGGI secara konsisten menghasilkan AUC di atas 0,93 pada seluruh model, dengan nilai tertinggi diraih Model B Balimau (0,981), diikuti Model A Gabungan (0,955) dan Model B Bahalayung (0,936). Tingginya AUC pada kelas TINGGI menegaskan bahwa model relatif andal dalam mengidentifikasi kejadian kenaikan muka air ekstrem, yang menjadi prioritas utama dalam sistem peringatan dini berbasis data.



Gambar 5 ROC Curve (One-vs-Rest) ketiga model

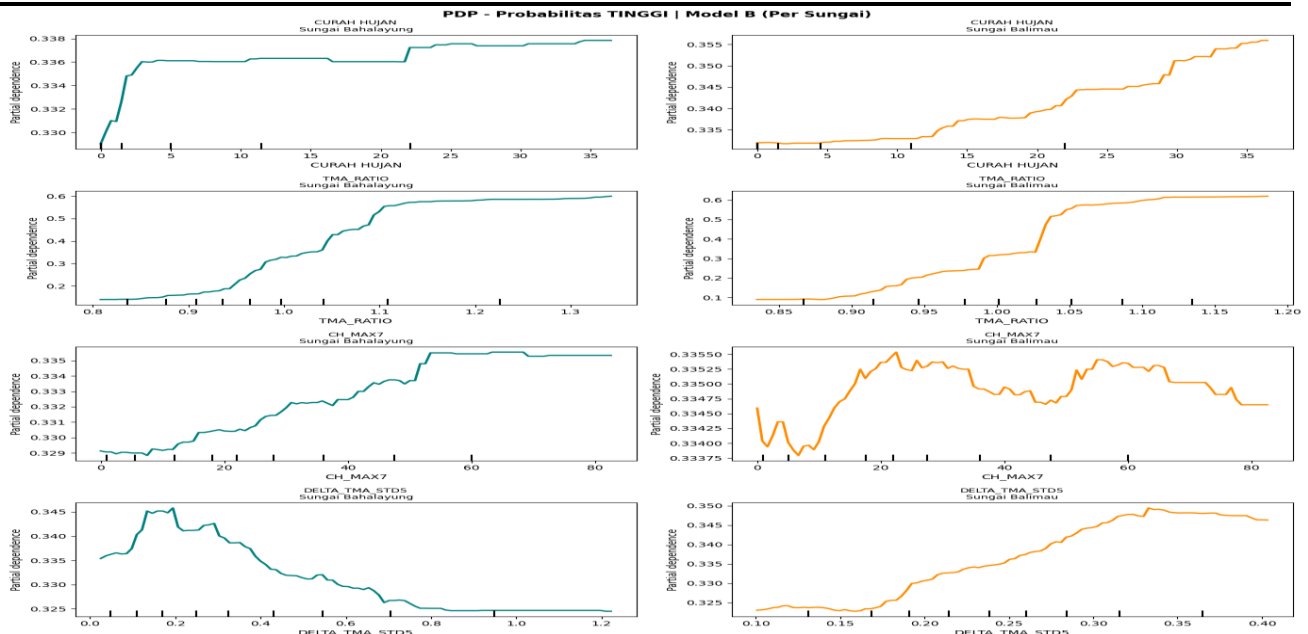
3.7 Analisis ROC-AUC

Kurva *Precision-Recall* pada Gambar 6 menghasilkan *Average Precision* (AP) kelas TINGGI sebesar 0,929 (Model A), 0,906 (Model B Bahalayung), dan 0,963 (Model B Balimau). Nilai AP yang konsisten tinggi mengonfirmasi bahwa model tidak hanya akurat secara keseluruhan, tetapi juga terpercaya dalam mendeteksi kejadian ekstrem meskipun distribusi kelas tidak sempurna seimbang.



Gambar 6 Precision-Recall Curve ketiga model

3.8 Feature Importance dan Partial Dependence Plot



Gambar 7. Partial Dependence Plot probabilitas kelas TINGGI per sungai

Partial Dependence Plot (PDP) pada Gambar 7 memperlihatkan bagaimana masing-masing fitur mempengaruhi probabilitas kelas TINGGI secara berbeda antara Sungai Bahalayung dan Sungai Balimau. Pada fitur CURAH HUJAN, Sungai Bahalayung menunjukkan lonjakan probabilitas yang tajam hanya pada curah hujan 0-5 mm/hari kemudian mendatar, mengindikasikan bahwa bahkan hujan ringan sudah cukup memicu kenaikan ekstrem. Berbeda dengan Sungai Balimau yang responsnya terus meningkat secara bertahap hingga curah hujan tertinggi, menunjukkan bahwa sungai ini membutuhkan akumulasi hujan yang lebih besar sebelum mencapai intensitas tinggi.

Untuk TMA_RATIO, kedua sungai sama-sama menunjukkan peningkatan probabilitas TINGGI yang signifikan seiring naiknya rasio, namun ambang kritisnya berbeda. Sungai Bahalayung mulai naik tajam di kisaran TMA_RATIO 0,9-1,1, sementara Sungai Balimau baru merespons signifikan di atas 1,0. Perbedaan ini konsisten dengan karakteristik Bahalayung yang lebih sensitif terhadap lonjakan relatif TMA. Pada CH_MAX7, Sungai Bahalayung memperlihatkan tren kenaikan yang lebih mulus dan konsisten, sementara Sungai Balimau menunjukkan pola yang fluktuatif tanpa tren jelas. Ini mengindikasikan bahwa curah hujan puncak mingguan lebih prediktif untuk Bahalayung dibanding Balimau.

Yang paling kontras adalah DELTA_TMA_STD5. Sungai Bahalayung menunjukkan hubungan negatif semakin tinggi volatilitas perubahan TMA, probabilitas TINGGI justru menurun, mengindikasikan bahwa pada Bahalayung, volatilitas tinggi lebih mencerminkan fluktuasi normal daripada kondisi ekstrem. Sebaliknya, Sungai Balimau menunjukkan hubungan positif yang jelas, di mana peningkatan volatilitas diikuti naiknya probabilitas TINGGI secara konsisten. Perbedaan arah respons ini memperkuat temuan bahwa kedua sungai memiliki dinamika hidrologis yang berbeda secara fundamental, sehingga model terpisah per sungai lebih tepat dibandingkan model gabungan.

3.9 Pembahasan

Hasil penelitian ini menunjukkan bahwa *Random Forest Classifier* mampu mengklasifikasikan intensitas kenaikan muka air sungai dengan tingkat akurasi yang memadai untuk keperluan penanggulangan banjir. Dua dari tiga model melampaui ambang akurasi 80%, yakni Model A Gabungan sebesar 81,44% dan Model B Sungai Balimau sebesar 82,82%. Capaian ini sebanding dengan penelitian Bagas *et al.* yang melaporkan akurasi *Random Forest* sebesar 99,05% pada klasifikasi banjir Sungai Citarum, meskipun penelitian tersebut menerapkan pembagian data secara acak tanpa mempertahankan urutan temporal sehingga berpotensi menghasilkan estimasi performa yang lebih optimistik dibandingkan kondisi pengujian yang sesungguhnya [19].

Perbandingan antar sungai dalam penelitian ini mengungkap perbedaan yang cukup mendasar antara Sungai Bahalayung dan Sungai Balimau. Model B Sungai Balimau secara konsisten unggul di seluruh metrik evaluasi, terutama pada recall kelas TINGGI yang mencapai 0,88 dibandingkan hanya 0,60 pada Sungai Bahalayung. Perbedaan ini tidak lepas dari perbedaan karakteristik hidrologis kedua sungai standar deviasi DELTA_TMA Sungai Bahalayung sekitar 0,54 m hampir dua kali lipat Sungai Balimau yang hanya 0,25 m. Mdegela *et al.* menyimpulkan bahwa tingginya variabilitas distribusi data pada kelas yang berdekatan menjadi faktor utama yang menekan performa klasifikasi,

khususnya pada kelas peralihan yang batasnya tidak tegas secara statistik [20].

Fitur TMA_RATIO yang terbukti paling dominan dalam penelitian ini dengan nilai kepentingan 0,191-0,216 menawarkan perspektif yang berbeda dari studi-studi sebelumnya yang umumnya mengandalkan nilai absolut ketinggian air sebagai prediktor utama Islam *et al.* mengemukakan bahwa fitur yang menormalisasi nilai terhadap konteks historis lokal memiliki daya prediktif yang lebih stabil lintas periode waktu, terutama pada sungai yang mengalami pergeseran musiman dalam pola ketinggian airnya [21]. Sementara itu, penggunaan CH_MAX7 sebagai representasi curah hujan puncak mingguan yang dalam penelitian ini terbukti lebih informatif dibandingkan fitur rata-rata sejalan dengan argumen Painter *et al.* bahwa sistem peringatan dini yang andal membutuhkan indikator yang peka terhadap kondisi ekstrem sesaat, bukan hanya kondisi rata-rata dalam suatu periode pengamatan [22].

Hasil analisis PDP memperlihatkan bahwa respons kedua sungai terhadap fitur-fitur kunci tidak selalu searah. Pada DELTA_TMA_STD5 misalnya, Sungai Bahalayung menunjukkan hubungan negatif semakin tinggi volatilitas perubahan TMA, probabilitas kelas TINGGI justru menurun sementara Sungai Balimau memperlihatkan hubungan yang sebaliknya. Perbedaan arah respons ini menjadi justifikasi empiris bahwa model terpisah per sungai lebih tepat dibandingkan model gabungan yang dipaksa menemukan satu pola umum untuk dua sungai dengan dinamika yang berbeda secara fundamental.

Secara keseluruhan, penelitian ini memberikan landasan empiris bagi instansi terkait untuk menyusun strategi penanggulangan banjir yang terdiferensiasi berdasarkan karakteristik masing-masing sungai. Sungai Balimau dengan pola kenaikan yang lebih teratur dapat dikelola dengan sistem pemantauan berbasis ambang nilai TMA_RATIO, sementara Sungai Bahalayung yang lebih volatil memerlukan pendekatan yang lebih adaptif dan kemungkinan membutuhkan data tambahan untuk meningkatkan kemampuan deteksi pada kelas intensitas tinggi di masa mendatang.

4. KESIMPULAN

Penelitian ini berhasil menganalisis intensitas kenaikan muka air Sungai Bahalayung dan Sungai Balimau di Kabupaten Hulu Sungai Selatan ke dalam tiga kategori rendah, sedang, dan tinggi menggunakan algoritma *Random Forest Classifier* berbasis data hidrologi harian periode 2022-2025. Dari ketiga model yang dibangun, model khusus Sungai Balimau mencatatkan akurasi tertinggi sebesar 82,82% karena pola kenaikan airnya yang relatif stabil sehingga batas antar kategori lebih mudah dikenali, model gabungan kedua sungai menyusul dengan akurasi 81,44%, namun sedikit terbatas karena model harus mempelajari dua karakteristik hidrologi yang berbeda secara bersamaan. Adapun model khusus Sungai Bahalayung memperoleh akurasi 76,29% akibat fluktuasi airnya yang jauh lebih ekstrem dan tidak beraturan sehingga polanya lebih sulit dipelajari oleh model. Untuk penelitian selanjutnya, disarankan menambahkan data seluruh sungai di Provinsi Kalimantan Selatan agar model dapat menangkap pola banjir lintas wilayah secara lebih menyeluruh.

UCAPAN TERIMA KASIH

Terima kasih kepada Balai Wilayah Sungai (BWS) Kalimantan III atas ketersediaan data curah hujan dan ketinggian muka air Sungai Bahalayung, serta kepada dosen pembimbing dan Program Studi S1 Informatika, Universitas Muhammadiyah Banjarmasin, atas bimbingan dan dukungan yang diberikan selama penelitian ini berlangsung.

DAFTAR PUSTAKA

- [1] DIBI BNPB, "Data Informasi Bencana Indonesia," Badan Nasional Penanggulangan Bencana.
- [2] A. UMMIYATI, W. A. PUTRI, V. S. NUR AZIZA, A. A. LESMANA, and K. N. AZIZ, "Analisis Pengaruh Curah Hujan terhadap Frekuensi Kejadian Banjir di Kabupaten Kotabaru, Kalimantan Selatan," *Progressive Physics Journal*, vol. 5, no. 1, p. 393, Jun. 2024, doi: 10.30872/ppj.v5i1.1266.
- [3] BNPB, "Dokumen Kajian Risiko Bencana Nasional Provinsi Kalimantan Selatan 2022–2026," Badan Nasional Penanggulangan Bencana.
- [4] D. T. Prasasti, E. Suhartanto, and L. Prasetyorini, "Aplikasi Penginderaan Jauh dan SIG untuk Pemetaan Daerah Rawan Banjir Sebagai Upaya Mitigasi Bencana di DAS Petung," *Jurnal Teknologi dan Rekayasa Sumber Daya Air*, vol. 4, no. 02, pp. 1400–1414, Aug. 2024, doi: 10.21776/ub.jtresda.2024.004.02.141.
- [5] Badan Penanggulangan Bencana Daerah Kab. Hulu Sungai Selatan, "DATA BENCANA BANJIR DI KABUPATEN HULU SUNGAI SELATAN," Satu Data Indonesia. Accessed: Jul. 06, 2026. [Online]. Available: <https://data.go.id/dataset/dataset/data-bencana-banjir-dikabupaten-hulu-sungai-selatan>

- [6] Admin HSS, “Letak Geografis Sungai,” Pemerintah Kabupaten Hulu Sungai Selatan. Accessed: Jul. 09, 2026. [Online]. Available: <https://pembkab.hulusungaiselatankab.go.id/>
- [7] Balimau HSS, “Banjir Melanda Dua Desa di Kawasan Balimau dan Karang Paci, Kalumpang,” GoIndonesia.
- [8] M. Rizal H, F. M Sabir, and M. Mursalim, “SISTEM PERINGATAN DINI BANJIR BERBASIS SUPPORT VECTOR MACHINE YANG DIOPTIMALKAN DENGAN PARTICLE SWARM OPTIMIZATION,” *Journal of Information System, Applied, Management, Accounting and Research*, vol. 10, no. 1, p. 155, Feb. 2026, doi: 10.52362/jisamar.v10i1.2260.
- [9] D. Setiawan, R. Noratama Putri, Salamun, Luluk Elvitaria, and Liza Trisnawati, “Pengembangan Arsitektur Edge Computing untuk Early Warning System (EWS) Banjir Berbasis Multi-Sensor Menggunakan Model Hybrid LSTM–Random Forest,” *JEKIN - Jurnal Teknik Informatika*, vol. 6, no. 1, pp. 82–88, Feb. 2026, doi: 10.58794/jekin.v6i1.2028.
- [10] E. Hermawan, S. Darmawan Panjaitan, E. Faja Ripanti, J. H. Hadari Nawawi, P. Tenggara, and K. Pontianak Kalimantan Barat, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Prediksi Banjir Rob Kota Pontianak Berbasis Machine Learning Menggunakan Framework Streamlit”.
- [11] S. E. Purwati and Y. Pristyanto, “Model Random Forest and Support Vector Machine for Flood Classification in Indonesia,” *sinkron*, vol. 8, no. 4, pp. 2261–2268, Oct. 2024, doi: 10.33395/sinkron.v8i4.13973.
- [12] J. Latifah, F. D. Marleny, A. A. Ningrum, and M. Munsyi, “SISTEM PREDIKSI KUALITAS AIR SUNGAI PADA PULAU CURIAK MENGGUNAKAN ALGORITMA RANDOM FOREST,” *Technologia : Jurnal Ilmiah*, vol. 16, no. 3, p. 503, Jul. 2025, doi: 10.31602/tji.v16i3.19329.
- [13] A. Soleimani, S. H. Amin, F. Baki, and B. Shah, “Forecasting weekly sales using lag-based feature engineering and a bagged tree ensemble,” *Machine Learning with Applications*, vol. 24, Jun. 2026, doi: 10.1016/j.mlwa.2026.100922.
- [14] D. Nababan, C. T. Nadeak, L. Rassiyanti, F. Farid, and K. Kunci, “Universitas Multi Data Palembang | 781 5 th MDP STUDENT CONFERENCE (MSC) Klasifikasi Multikelas Varietas Kacang Kering Menggunakan Metode Hybrid SVM Berbasis DAG.”
- [15] S. Sami, M. Hajian, and T. Nazghelichi, “A Time-Driven Approach Leveraging Universally Accessible Features for Photovoltaic Power Forecasting,” *Iranica Journal of Energy and Environment*, vol. 17, no. 2, pp. 382–396, 2026, doi: 10.5829/ijee.2026.17.02.13.
- [16] M. Gava, P. M. Biron, and T. Buffin-Bélanger, “A random forest machine learning model to detect fluvial hazards,” *River Res. Appl.*, vol. 40, no. 10, pp. 1837–1854, Dec. 2024, doi: 10.1002/rra.4353.
- [17] A. A. Shah, A. Ullah, N. A. Khan, A. Khan, M. A. U. R. Tariq, and C. Xu, “Community social barriers to non-technical aspects of flood early warning systems and NGO-led interventions: The case of Pakistan,” *Front. Earth Sci. (Lausanne)*, vol. 11, 2023, doi: 10.3389/feart.2023.1068721.
- [18] G. Man, Q. Yun, Z. Qilin, L. Yuyong, and Z. Zhuoshi, “Surface water quality classification and prediction model based on multiple machine learning algorithms,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-23775-5.
- [19] M. Bagas, A. Darmawan, F. Dewanta, and S. Astuti, “Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot Comparative Analysis of Decision Tree, Random Forest, and Naïve Bayes Algorithm for Flood Prediction at Dayeuhkolot Village,” *TELKA*, vol. 9, no. 1, pp. 52–61, 2023.
- [20] 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, Mar. 2023, doi: 10.3390/w15061021.
- [21] M. T. Islam *et al.*, “Enhancing monthly extreme water level predictions in a flood-prone river basin using regression-based machine learning,” *H2Open Journal*, vol. 8, no. 4, pp. 291–312, Jul. 2025, doi: 10.2166/h2oj.2025.016.
- [22] P. Painter, K. Semmens, K. Maxfield, C. Cattoën, and R. H. Carr, “Designing Effective Flood Early Warning Systems: A Review of Barriers, Best Practices, and Key Characteristics,” Dec. 01, 2025, *John Wiley and Sons Inc.* doi: 10.1111/jfr3.70145.