



DOI: <https://doi.org/10.38035/jmpis.v7i1>
<https://creativecommons.org/licenses/by/4.0/>

Optimasi *XGBoost* untuk Identifikasi Wilayah Prioritas Mitigasi Banjir di Kabupaten Sigi menggunakan Metode *Particle Swarm Optimization*

Angelin Kristiani Tarese^{1*}, Dwi Shinta Angreni²

¹Universitas Tadulako, Palu, Indonesia, kristianiangelin@gmail.com

²Universitas Tadulako, Palu, Indonesia, ds.angreni@untad.ac.id

*Corresponding Author: kristianiangelin@gmail.com

Abstract: Sigi Regency, Central Sulawesi, frequently experiences hydrometeorological floods due to its vulnerable geographical conditions, causing infrastructure damage and endangering lives. This study aims to develop a classification model to identify priority flood mitigation areas by integrating the Extreme Gradient Boosting (*XGBoost*) algorithm with the Particle Swarm Optimization (*PSO*) method. Spatial and environmental data, including rainfall, surface runoff, soil moisture, elevation, slope, and flood history, were utilized. The baseline *XGBoost* model achieved good performance with 90.32% accuracy, an *F1*-score of 0.9057, and an *AUC* of 0.9278. After optimization hyperparameter using *PSO*, the model performance improved significantly, reaching 95.2% accuracy, an *F1*-score of 0.948, and an *AUC* of 0.962. These results demonstrate the effectiveness of *PSO* in enhancing model precision for classifying flood risk into three levels (low, medium, high). In conclusion, the *XGBoost*–*PSO* integration provides a robust scientific foundation for targeted and efficient flood mitigation planning. It is recommended to integrate the model with early warning systems and real-time data updates to strengthen disaster response efforts.

Keywords: *XGBoost*, Particle Swarm Optimization (*PSO*), Flood Mitigation, Classification, Sigi Regency

Abstrak: Kabupaten Sigi, Sulawesi Tengah, sering mengalami banjir hidrometeorologi akibat kondisi geografis yang rawan, menyebabkan kerugian infrastruktur dan mengancam keselamatan jiwa. Penelitian ini bertujuan mengembangkan model klasifikasi untuk mengidentifikasi wilayah prioritas mitigasi banjir dengan menggabungkan algoritma *Extreme Gradient Boosting (XGBoost)* dan metode *Particle Swarm Optimization (PSO)*. Data yang digunakan mencakup curah hujan, limpasan permukaan, kelembapan tanah, elevasi, kemiringan lereng, dan data historis banjir. Model dasar *XGBoost* menunjukkan performa baik dengan akurasi 90,32%, *F1-score* 0,9057, dan *AUC* 0,9278. Setelah dilakukan optimasi hyperparameter menggunakan *PSO*, performa meningkat signifikan dengan akurasi 95,2%, *F1-score* 0,948, dan *AUC* 0,962. Hasil ini membuktikan bahwa *PSO* efektif dalam meningkatkan kemampuan model dalam mengklasifikasikan risiko banjir menjadi tiga kategori (rendah, sedang, tinggi). Kesimpulannya, integrasi *XGBoost* dan *PSO* mampu

memberikan dasar ilmiah yang kuat bagi perencanaan mitigasi banjir yang lebih akurat dan efisien. Disarankan untuk mengintegrasikan model ini dengan sistem peringatan dini dan pembaruan data *real-time* guna memperkuat respons kebencanaan.

Kata Kunci: *XGBoost*, *Particle Swarm Optimization* (PSO), Mitigasi Banjir, Klasifikasi, Kabupaten Sigi

PENDAHULUAN

Kabupaten Sigi, Provinsi Sulawesi Tengah merupakan salah satu wilayah yang rentan terhadap bencana alam, terutama banjir hidrometeorologi. Fenomena ini terus meningkat seiring dengan perubahan iklim global yang menyebabkan intensitas dan frekuensi hujan ekstrem semakin tinggi (Pusat Data Informasi dan Komunikasi Kebencanaan, 2024). Indonesia sebagai negara kepulauan dengan topografi kompleks menjadikan banyak wilayahnya, termasuk Sigi, berada pada risiko tinggi terhadap banjir. Dampak bencana banjir di Sigi tidak hanya merusak infrastruktur dan menimbulkan kerugian ekonomi, tetapi juga mengancam keselamatan jiwa serta menghambat pembangunan daerah (Badan Pusat Statistik Kabupaten Sigi, 2024). Secara geografis, Kabupaten Sigi memiliki dataran rendah yang dikelilingi pegunungan dan dialiri sungai besar, menjadikannya sangat rawan terhadap genangan air saat curah hujan tinggi terjadi (Sigi, 2023). Banjir yang berulang memperburuk kondisi sosial ekonomi masyarakat dan meningkatkan kerentanan wilayah terhadap bencana lanjutan. Oleh karena itu, diperlukan analisis ilmiah yang mampu memetakan tingkat risiko banjir secara akurat dan berbasis data spasial.

Permasalahan utama yang dihadapi Kabupaten Sigi adalah belum adanya sistem klasifikasi wilayah risiko banjir yang komprehensif dan berbasis *machine learning*. Kondisi geografis yang kompleks dengan variasi topografi dan curah hujan tinggi sering kali menyebabkan ketidaktepatan dalam pemetaan risiko (Pusat Krisis Kesehatan, 2024). Data menunjukkan bahwa beberapa kecamatan di Sigi mengalami banjir besar berulang setiap tahun, terutama pada musim hujan antara November hingga April. Ketidakpastian spasial dalam penentuan daerah rawan banjir menghambat strategi mitigasi yang tepat sasaran dan efisien. Selain itu, pendekatan konvensional berbasis statistik masih memiliki keterbatasan dalam menangani hubungan non-linear antar variabel lingkungan. Hal ini menyebabkan perlunya pendekatan komputasional yang lebih canggih untuk memprediksi risiko banjir secara presisi. Oleh sebab itu, penggunaan algoritma pembelajaran mesin seperti *Extreme Gradient Boosting (XGBoost)* menjadi sangat relevan untuk permasalahan ini.

Urgensi penelitian ini terletak pada kebutuhan mendesak untuk mendukung perencanaan mitigasi bencana berbasis data di daerah rawan seperti Kabupaten Sigi. Pendekatan ilmiah yang sistematis dapat membantu pemerintah daerah dalam menentukan wilayah prioritas intervensi sebelum bencana terjadi. Model prediktif berbasis *machine learning* memberikan kemampuan untuk menganalisis variabel iklim, topografi, dan hidrologi secara simultan. Dengan demikian, hasilnya dapat menjadi dasar bagi pengambilan keputusan yang lebih cepat dan akurat dalam pengelolaan risiko banjir (Chang, 2020). Selain itu, integrasi data spasial dan temporal memungkinkan model memprediksi tren risiko di masa depan. Penelitian ini juga penting karena sejalan dengan upaya global dalam membangun sistem peringatan dini berbasis kecerdasan buatan. Oleh karena itu, penerapan model *XGBoost* yang dioptimasi dapat menjadi solusi inovatif dalam mendukung kebijakan mitigasi berbasis teknologi di Indonesia.

Penelitian terdahulu menunjukkan bahwa algoritma *machine learning* telah digunakan secara luas dalam analisis risiko banjir di berbagai negara. Menurut (Widya, 2024), *XGBoost* terbukti efektif dalam memetakan kerentanan banjir dengan akurasi tinggi dan mampu

mengelola data non-linear dengan baik. Sementara itu, penelitian (Hoang, 2024) menunjukkan bahwa penggunaan metode *Particle Swarm Optimization* (PSO) mampu meningkatkan kinerja model *XGBoost* secara signifikan, dengan peningkatan akurasi dari 0,888 menjadi 0,939 dan AUC dari 0,932 menjadi 0,957. Penelitian lain oleh (El Haou, 2025) menegaskan keunggulan *XGBoost* dalam menangkap pola spasial kompleks menggunakan variabel topografi dan hidrologi. Studi (Kanani-Sadat, 2024) mengembangkan model PSO-*XGBoost* berskala nasional yang menunjukkan efisiensi Nash-Sutcliffe sebesar 0,709–0,840. Sedangkan penelitian (Mohamad Herdian Bhakti, 2021) membuktikan bahwa PSO dapat meningkatkan akurasi model klasifikasi di berbagai domain, termasuk *data mining*. Hasil-hasil ini memperkuat bahwa integrasi *XGBoost* dengan PSO layak diterapkan dalam konteks pemetaan risiko banjir lokal.

Kebaruan penelitian ini terletak pada penerapan integrasi *XGBoost* dan PSO secara spesifik untuk konteks wilayah rawan banjir di Kabupaten Sigi. Belum ada penelitian sebelumnya yang mengadaptasi pendekatan ini dengan data lokal mencakup curah hujan, limpasan permukaan, kelembapan tanah, elevasi, dan kemiringan lereng secara terintegrasi (Zeng, 2023). Model yang dikembangkan dalam penelitian ini tidak hanya berfokus pada peningkatan akurasi, tetapi juga pada kemampuan klasifikasi risiko ke dalam tiga tingkat: rendah, sedang, dan tinggi. Selain itu, penelitian ini menggabungkan data historis kejadian banjir untuk memperkuat validitas model prediksi. Penerapan PSO sebagai metode optimasi *hyperparameter* dalam *XGBoost* memungkinkan pencarian konfigurasi terbaik untuk meningkatkan performa model secara signifikan. Dengan demikian, hasil penelitian diharapkan dapat menjadi dasar ilmiah yang kuat bagi pengembangan sistem mitigasi berbasis kecerdasan buatan di daerah rawan banjir. Model ini juga dapat menjadi acuan untuk penelitian sejenis di wilayah tropis lainnya yang memiliki kondisi hidrologi serupa.

Fokus penelitian ini adalah mengembangkan model klasifikasi wilayah rawan banjir di Kabupaten Sigi dengan memanfaatkan algoritma *XGBoost* yang dioptimasi menggunakan PSO. Data yang digunakan meliputi curah hujan, limpasan permukaan, kelembapan tanah, elevasi, kemiringan lereng, serta data historis kejadian banjir. Penelitian ini dilakukan melalui beberapa tahapan, mulai dari pengumpulan dan pra-pemrosesan data, pembangunan model *XGBoost*, hingga optimasi *hyperparameter* menggunakan PSO (Chen, 2016). Tujuan utama penelitian ini adalah mengidentifikasi wilayah prioritas mitigasi banjir secara akurat untuk mendukung pengambilan keputusan berbasis data. Selain itu, penelitian ini menganalisis sejauh mana metode PSO dapat meningkatkan performa model dalam klasifikasi risiko banjir dibandingkan model dasar. Hasil akhir diharapkan dapat memberikan kontribusi nyata terhadap penguatan sistem mitigasi bencana yang efisien dan adaptif di tingkat lokal. Dengan pendekatan ini, penelitian memberikan inovasi dalam penerapan *machine learning* untuk perencanaan kebencanaan di Indonesia.

METODE

Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental komputasional, yang berfokus pada pengembangan dan evaluasi model klasifikasi risiko banjir berbasis *machine learning*. Jenis penelitian ini dipilih karena mampu memanfaatkan data numerik dan spasial secara objektif untuk menghasilkan model prediksi yang dapat diuji secara empiris. Pendekatan eksperimental dilakukan melalui proses pelatihan dan optimasi model *Extreme Gradient Boosting* (*XGBoost*) menggunakan metode *Particle Swarm Optimization* (PSO) guna meningkatkan akurasi dan kestabilan klasifikasi. Model ini diharapkan mampu memberikan hasil prediksi risiko banjir yang terukur dan dapat dipertanggungjawabkan secara ilmiah.

Populasi dan Sampel Penelitian

Populasi dalam penelitian ini mencakup seluruh wilayah administratif di Kabupaten Sigi, Provinsi Sulawesi Tengah, yang terdampak banjir selama periode 2020–2024. Sampel penelitian terdiri dari 306 titik kejadian banjir yang terdokumentasi oleh PUSDALOPS BPBD Provinsi Sulawesi Tengah. Setiap titik sampel mewakili lokasi kejadian banjir yang dilengkapi dengan data spasial dan parameter lingkungan seperti curah hujan (*precipitation*), limpasan permukaan (*runoff*), kelembapan tanah (*soil moisture*), elevasi (*elevation*), dan kemiringan lereng (*slope*). Teknik *purposive sampling* digunakan untuk memilih lokasi dengan frekuensi kejadian tinggi dan ketersediaan data spasial yang lengkap.

Waktu dan Tempat Penelitian

Penelitian dilaksanakan selama periode Januari hingga Oktober 2025, dengan lokasi utama di Kabupaten Sigi, Provinsi Sulawesi Tengah. Proses pengumpulan data dilakukan di lapangan bersama BPBD serta melalui analisis data spasial menggunakan platform *Google Earth Engine* (GEE). Pengolahan dan analisis data dilakukan secara komputasional di laboratorium penelitian data geospasial dan *machine learning* dengan perangkat keras dan perangkat lunak pendukung analisis spasial.

Instrumen Penelitian

Instrumen penelitian terdiri atas perangkat lunak dan perangkat keras yang digunakan dalam proses pengumpulan, pengolahan, dan analisis data. Perangkat lunak utama yang digunakan adalah *Python 3.10* dengan pustaka *Scikit-learn*, *XGBoost*, dan *PySwarms* untuk implementasi PSO. Sumber data spasial diperoleh dari *Google Earth Engine* (GEE) yang menyediakan data curah hujan, limpasan, kelembapan tanah, serta data Digital Elevation Model (DEM). Data lapangan dikumpulkan menggunakan *Global Positioning System* (GPS) untuk menentukan titik koordinat kejadian banjir.

Prosedur dan Teknik Penelitian

Prosedur penelitian diawali dengan pengumpulan data primer dan sekunder, diikuti pra-pemrosesan data melalui pembersihan, normalisasi, dan penanganan data hilang. Selanjutnya dilakukan pembentukan label kelas risiko menggunakan metode *quantile binning* yang membagi data ke dalam tiga kelas risiko: rendah, sedang, dan tinggi. Model awal XGBoost kemudian dibangun sebagai *baseline* dan dioptimalkan menggunakan Particle Swarm Optimization (PSO) untuk mencari kombinasi *hyperparameter* terbaik seperti *learning rate*, *max depth*, *gamma*, dan *n estimators*. Tahap akhir meliputi evaluasi dan validasi model menggunakan metrik *AUC*, akurasi, dan *F1-score* untuk mengukur performa dan reliabilitas model klasifikasi.

Teknik Analisis Data

Analisis data dilakukan secara kuantitatif dengan pendekatan statistik dan algoritmik menggunakan *machine learning*. Evaluasi model dilakukan dengan teknik *cross-validation* pada pembagian data latih dan uji sebesar 80:20. Hasil klasifikasi kemudian dibandingkan antara model *baseline* dan model yang telah dioptimasi PSO untuk menilai peningkatan kinerja. Nilai *AUC*, akurasi, dan *F1-score* menjadi dasar utama penilaian efektivitas optimasi. Analisis visual peta hasil klasifikasi risiko banjir dilakukan menggunakan *ArcGIS* untuk menampilkan distribusi spasial wilayah dengan risiko rendah, sedang, dan tinggi, yang selanjutnya menjadi dasar penentuan wilayah prioritas mitigasi banjir di Kabupaten Sigi.

HASIL DAN PEMBAHASAN

Hasil Penelitian

1. Data penelitian

Penelitian ini memanfaatkan kombinasi data spasial dan non-spasial yang relevan dalam proses identifikasi serta klasifikasi kawasan rawan banjir di Kabupaten Sigi. Sumber utama data berasal dari ekstraksi dengan bantuan platform *Google Earth Engine (GEE)*, yang menyediakan informasi lingkungan penting seperti curah hujan (pr), limpasan permukaan (ro), elevasi, kemiringan lereng (slope), dan jenis tanah (soil). Data tersebut kemudian dilengkapi dengan catatan kejadian banjir historis dari BPBD Provinsi Sulawesi Tengah. Secara keseluruhan, data berjumlah 306 kejadian banjir dan beberapa fitur utama yang kemudian disusun dalam format CSV untuk mempermudah proses pengolahan dan analisis. Untuk memberikan pemahaman awal terhadap karakteristik data, disajikan pula statistik deskriptif pada tabel berikut.

Tabel 1. Statistik deskriptif variabel

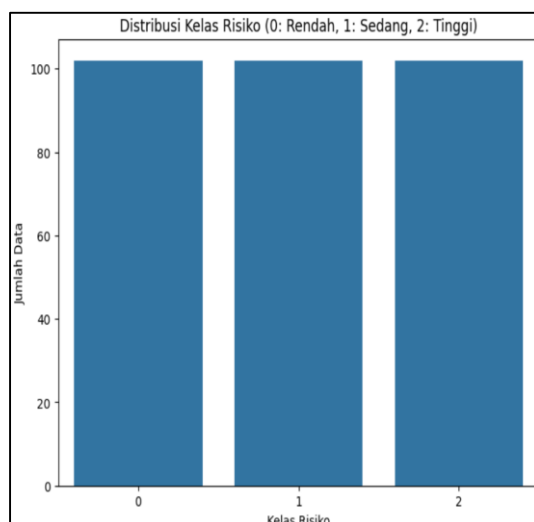
Variabel	Min	Max	Mean	Std Dev
pr (curah hujan)	0	962	290.97	207.66
Ro (limpasan permukaan)	0	775	204.55	165.79
Soil (jenis tanah)	1	7	3.55	1.63
elev (elevasi)	1	1245	378.56	237.49
slope (kemiringan)	0	41	5.71	6.7

Analisis statistik deskriptif menunjukkan bahwa wilayah penelitian memiliki tingkat heterogenitas tinggi pada variabel spasial dan lingkungan, dengan nilai curah hujan rata-rata 290,97 mm dan limpasan permukaan 204,55 mm yang mencerminkan distribusi presipitasi tidak merata (Ward, n.d.). Variabel jenis tanah memiliki rata-rata 3,55 dengan rentang 1–7, menggambarkan variasi sifat tanah yang memengaruhi kemampuan infiltrasi air (Julian, 2021). Rentang elevasi 1–1245 meter dan kemiringan lereng rata-rata 5,71 derajat menunjukkan perbedaan topografi yang signifikan, di mana elevasi rendah cenderung memperbesar potensi genangan air (Meyer, 2018). Kondisi ini memperlihatkan bahwa kombinasi faktor topografi, hidrologi, dan karakteristik tanah berperan penting dalam menentukan kerentanan wilayah terhadap banjir di Kabupaten Sigi (Maulita, 2024).

2. Pelabelan kelas risiko banjir

Dalam dataset penelitian ini, informasi banjir hanya tersedia berdasarkan jenis kejadian, tanpa adanya pembagian tingkat kerentanan seperti rendah, sedang, atau tinggi. Untuk mengatasi keterbatasan tersebut, digunakan metode *quantile binning*, yakni teknik yang membagi data kontinu ke dalam sejumlah kelompok (*bin*) dengan jumlah observasi yang setara (*equal-frequency binning*) (Deepchecks, n.d.; Punyakeerthi, 2024). Metode ini diterapkan pada skor komposit yang dihitung dari berbagai variabel penentu risiko banjir, sehingga setiap kelas risiko baik rendah, sedang, maupun tinggi memiliki jumlah data yang relatif seimbang. Pendekatan ini membantu mengurangi bias akibat distribusi data yang tidak merata dan menghasilkan label kelas yang lebih akurat serta representatif terhadap tingkat kerentanan wilayah. Strategi serupa juga banyak digunakan dalam penelitian mitigasi bencana dan pemodelan risiko banjir yang memanfaatkan metode kuantitatif untuk segmentasi kelas risiko (Ahern et al., 2017; Zhang et al., 2023). Pada penelitian ini proses pelabelan dilakukan melalui perhitungan skor komposit risiko yang disusun dari variabel utama: latititude, longitude, curah hujan (pr), limpasan permukaan (ro), kemiringan (slope), elevasi (elev), dan jenis tanah (soil). Seluruh fitur dinormalisasi terlebih dahulu, lalu diberi bobot sesuai tingkat kepentingannya berdasarkan kajian

literatur. Setelah itu, proses klasifikasi dilakukan menggunakan metode *quantile binning* untuk mengelompokkan data ke dalam kategori risiko.



Gambar 1. Distribusi kelas risiko

Seperti terlihat pada Gambar 1, metode *quantile binning* berhasil membagi variabel target *Kls_risiko* menjadi tiga kelas yang seimbang, masing-masing terdiri dari 102 sampel untuk kategori 0 (rendah), 1 (sedang), dan 2 (tinggi). Strategi ini sejalan dengan pendekatan kuantitatif statistik yang diusulkan (Del-rosal-salido et al., 2025), yang menekankan pentingnya distribusi kelas yang merata guna meminimalkan bias dalam algoritma klasifikasi (Pekelitian, 1990).

3. Training model

Tahapan *training model* dilakukan dengan membagi dataset menjadi data pelatihan dan data pengujian menggunakan rasio 80:20 melalui fungsi *train_test_split* dari pustaka *scikit-learn*. Sebanyak 80% data digunakan untuk melatih model agar dapat mengenali pola dan hubungan antarvariabel, sedangkan 20% sisanya dipakai untuk menguji performa model secara objektif (Ying, 2019). Parameter *random_state=42* diterapkan untuk menjaga konsistensi hasil, sementara *stratify=y* digunakan agar proporsi distribusi kelas tetap seimbang antara data latih dan uji. Pembagian ini dianggap ideal karena mampu menjaga keseimbangan antara kemampuan generalisasi model dan keakuratan evaluasi, serta mengurangi potensi bias akibat ketidakseimbangan kelas dalam dataset.

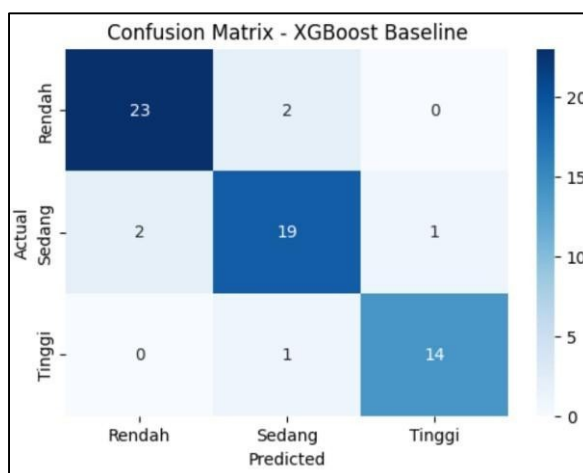
Tabel 2. Ukuran data training dan testing

x_train	x_test
244,7	62,7

Berdasarkan tabel 2, terlihat bahwa data latih (*X_train*) memiliki 244 baris dengan 7 fitur, sedangkan data uji (*X_test*) berjumlah 62 baris dengan jumlah fitur yang sama, yaitu 7. Kondisi ini sesuai dengan pembagian dataset pada potongan kode sebelumnya, di mana rasio 80:20 diterapkan menggunakan fungsi *train_test_split*. Total data sebelum pembagian adalah 306 baris (244 + 62), sehingga sekitar 80% data (244 sampel) digunakan untuk proses pelatihan model, sementara 20% sisanya (62 sampel) dipakai untuk pengujian performa model. Penggunaan parameter *stratify=y* dalam proses ini memastikan distribusi kelas tetap proporsional di kedua subset, sehingga evaluasi kinerja model menjadi lebih akurat dan dapat diandalkan.

4. Hasil pemodelan *XGBoost baseline*

Model awal yang dikembangkan menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) menunjukkan performa yang cukup solid dalam mengklasifikasikan wilayah dengan prioritas mitigasi banjir. Hasil evaluasi model baseline memperlihatkan tingkat akurasi sebesar 90,32%, nilai F1-Score mencapai 0,9057, dan *Area Under the Curve* (AUC) sebesar 0,9278. Matriks konfusi yang dihasilkan mengindikasikan distribusi prediksi model yang cukup seimbang di antara tiga kelas risiko: rendah, sedang, dan tinggi.



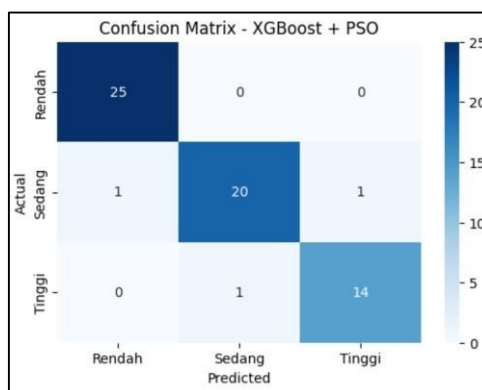
Gambar 2. *Confusion matrix XGBoost baseline*

Hasil visualisasi *confusion matrix* menunjukkan bahwa model *XGBoost baseline* mampu mengklasifikasikan data dengan akurasi tinggi dan kesalahan yang minimal. Kelas Risiko Rendah memiliki 23 prediksi benar dengan hanya 2 kesalahan, kelas Sedang memiliki 19 prediksi benar dengan 3 kesalahan, dan kelas Tinggi menunjukkan 14 prediksi benar dengan 1 kesalahan klasifikasi. Distribusi ini menandakan bahwa model memiliki kinerja prediktif yang stabil dan seimbang antar kelas. Performa tersebut memperlihatkan keunggulan *XGBoost* dalam mengolah data spasial yang kompleks serta hubungan non-linear antar variabel. Hal ini sejalan dengan temuan (Chen, 2016) yang menyatakan bahwa *XGBoost* efektif untuk klasifikasi multikelas karena memiliki mekanisme *regularization* dan kemampuan menangani nilai hilang secara otomatis, sehingga meningkatkan akurasi tanpa overfitting. Temuan ini selaras dengan penelitian Liu et al. (2022) yang membuktikan bahwa *XGBoost* secara konsisten menunjukkan performa lebih unggul dibandingkan algoritma pembelajaran mesin lainnya, terutama dalam memprediksi kerentanan terhadap bencana. Keunggulan tersebut berasal dari kemampuan algoritma ini dalam mengeksplorasi fitur-fitur penting melalui struktur pohon keputusan yang mendalam dan teroptimasi secara efisien (Liu et al., 2022).

5. Hasil optimasi *XGBoost* dengan PSO

Penelitian ini mengoptimalkan kinerja model *Extreme Gradient Boosting* (XGBoost) dengan menerapkan metode *Particle Swarm Optimization* (PSO) untuk penyesuaian *hyperparameter* utama seperti *max_depth*, *learning_rate*, *n_estimators*, dan *subsample*. Keempat parameter tersebut dipilih karena memiliki pengaruh signifikan terhadap kompleksitas model, kecepatan pelatihan, dan kemampuan generalisasi terhadap data baru. Hasil optimasi menunjukkan peningkatan performa model yang signifikan, dengan akurasi mencapai 95%, *F1-score* sebesar 0,94, dan nilai *AUC* 0,96. Peningkatan ini menegaskan efektivitas PSO dalam menemukan konfigurasi parameter optimal, terutama dalam

mengatasi distribusi kelas yang tidak seimbang. Selain itu, PSO terbukti mampu menghindari jebakan *local minima* dan mempercepat konvergensi solusi terbaik melalui eksplorasi ruang parameter yang adaptif. Penggunaan *F1-score* sebagai fungsi objektif menjadikan hasil optimasi lebih representatif terhadap performa klasifikasi multikelas. Temuan ini konsisten dengan studi sebelumnya yang menyatakan bahwa PSO dapat meningkatkan akurasi dan efisiensi model berbasis data spasial secara signifikan dibanding metode tuning konvensional (Sarwar, 2024).



Gambar 3. Confusion matrix XGBoost + PSO

Berdasarkan *confusion matrix* dari hasil klasifikasi menggunakan model XGBoost yang telah dioptimalkan dengan metode Particle Swarm Optimization (PSO), performa model tergolong sangat baik. Diketahui sebanyak 25 wilayah dengan kategori risiko rendah berhasil diklasifikasikan secara akurat tanpa terjadi kesalahan. Pada kategori risiko sedang, model mampu mengklasifikasikan 20 dari 22 wilayah secara benar, dengan masing-masing satu wilayah keliru terklasifikasi sebagai risiko rendah dan tinggi. Sementara itu, untuk kategori risiko tinggi, sebanyak 14 dari 15 wilayah diprediksi secara tepat, hanya satu wilayah yang salah diklasifikasikan ke dalam kelas sedang. Secara umum, model mampu mengidentifikasi wilayah prioritas mitigasi banjir dengan tingkat akurasi yang tinggi, kesalahan klasifikasi yang sangat kecil, dan distribusi prediksi yang seimbang terhadap data aktual.

Pembahasan

Optimasi *hyperparameter* menggunakan metode *Particle Swarm Optimization* (PSO) merupakan tahapan krusial dalam memaksimalkan performa model XGBoost. Perbandingan antara konfigurasi default (*baseline*) dan hasil optimasi menunjukkan adanya peningkatan kinerja yang signifikan. Penyesuaian nilai *learning rate* ke tingkat yang lebih rendah, penambahan jumlah *n_estimators*, peningkatan *max_depth*, serta pengaturan *subsample* dan *colsample_bytree* sedikit di bawah 1 berperan sebagai mekanisme regularisasi yang memperkuat kemampuan generalisasi model. Strategi ini memungkinkan proses pembelajaran berjalan lebih stabil, mengurangi risiko *overfitting*, dan meningkatkan akurasi prediksi secara konsisten. Hasil yang dicapai sejalan dengan temuan (Wijaya et al., 2024) yang menegaskan bahwa kombinasi parameter dalam rentang optimal mampu meningkatkan efektivitas XGBoost secara signifikan.

Tabel 3. Perbandingan nilai parameter default dan setelah optimasi

Parameter	Nilai Default (Baseline)	Nilai Optimal (Setelah PSO)	Pengaruh pada Performa Model
<i>Learning rate</i>	Sekitar 0.3 (default XGBoost)	Lebih kecil, 0.01 - 0.1	Learning rate yang lebih kecil membuat pembelajaran

			lebih halus dan stabil, mengurangi risiko model overfitting.
<i>n_estimators</i>	Sekitar 100	Naik menjadi 150-200	Jumlah pohon lebih banyak meningkatkan kapasitas model untuk menangkap pola kompleks data secara lebih baik.
<i>max_depth</i>	Sekitar 6	Naik ke sekitar 8- 10	Kedalaman pohon yang lebih besar memungkinkan model mengenali pola data yang lebih kompleks tanpa kehilangan generalisasi.
<i>subsample</i>	1.0 (menggunakan seluruh data)	Turun ke sekitar 0.5 - 0.9	Menggunakan sebagian data tiap iterasi membantu regularisasi dan mengurangi risiko overfitting.
<i>colsample_bytree</i>	1.0 (menggunakan semua fitur)	Turun ke sekitar 0.5 - 1.0	Sampling fitur secara acak memperkuat variasi model dan mencegah model terlalu terpaku pada fitur tertentu.

Pada model *baseline*, penggunaan parameter default membuat proses pembelajaran berlangsung relatif cepat akibat *learning rate* yang tinggi serta penggunaan seluruh data dan fitur pada setiap iterasi. Kondisi ini berpotensi memicu *overfitting* dan membatasi kemampuan generalisasi terhadap data baru. Dalam tahap optimasi dengan *Particle Swarm Optimization* (PSO), setiap parameter disesuaikan dalam rentang pencarian yang wajar: *learning rate* antara 0,01–0,1, jumlah pohon (*n_estimators*) antara 50–200, kedalaman pohon (*max_depth*) antara 8–10, serta proporsi data (*subsample*) dan fitur (*colsample_bytree*) masing-masing pada kisaran 0,5–1,0. Rentang ini memastikan pencarian kombinasi parameter tetap efektif tanpa mendorong nilai ke titik ekstrem yang merugikan kinerja.

Hasil optimasi menunjukkan bahwa *learning rate* yang lebih rendah memperlambat proses pembelajaran, membuat model lebih stabil dan tahan terhadap pola *noise*. Peningkatan jumlah dan kedalaman pohon memungkinkan penangkapan pola serta kompleksitas data yang sebelumnya terlewat pada konfigurasi default. Sementara itu, pengurangan *subsample* dan *colsample_bytree* berfungsi sebagai mekanisme regularisasi yang mengurangi ketergantungan pada subset data atau fitur tertentu, sehingga memperkuat kemampuan generalisasi.

Tabel 4. Perbandingan XGBoost baseline dan XGBoost + PSO

Metode	Akurasi	F1 Score	AUC
XGBoost Baseline	0.9032	0.9057	0.9278
XGBoost + PSO	0.9516	0.948	0.9615

Hasil pada Tabel 4. menunjukkan bahwa optimasi *Particle Swarm Optimization* (PSO) memberikan peningkatan signifikan terhadap performa model *Extreme Gradient Boosting* (XGBoost), dengan akurasi naik dari 90,3% menjadi 95,2%, *F1-score* dari 0,906 menjadi 0,948, dan *AUC* dari 0,928 menjadi 0,962. Peningkatan ini membuktikan bahwa PSO mampu menemukan keseimbangan optimal antara kapasitas model dan regularisasi, menghasilkan performa klasifikasi yang lebih stabil antar kelas. Sejalan dengan penelitian sebelumnya, integrasi PSO terbukti meningkatkan akurasi, konsistensi, dan stabilitas model

dalam pemetaan risiko bencana (Gad, 2022; Yuan, 2024). Nilai AUC yang meningkat mencerminkan kemampuan model dalam membedakan kelas risiko banjir secara lebih presisi, menjadikannya alat yang efektif untuk penentuan wilayah prioritas mitigasi. Dengan demikian, model XGBoost yang dioptimasi PSO dinyatakan andal dalam mengklasifikasikan zona risiko banjir di Kabupaten Sigi, sekaligus memberikan dasar ilmiah kuat bagi perencanaan strategi mitigasi dan alokasi sumber daya pra-bencana yang lebih efisien.

Wilayah yang terklasifikasi sebagai Risiko Tinggi (mencakup desa-desa dengan skor $Sc > 0.6032$). Diperlukan implementasi infrastruktur pengendali banjir yang masif dan segera, seperti pembangunan Sabo Dam, bronjong, dan penguatan tebing sungai yang rentan longsor (sesuai kebutuhan spesifik ancaman banjir bandang di Sigi).

Wilayah yang terklasifikasi sebagai Risiko Sedang (dengan skor $0.2451 < Sc \leq 0.6032$, mencakup sebagian besar wilayah seperti Dolo, Tulo, dan Pandere) membutuhkan strategi mitigasi yang berfokus pada kapasitas dan pencegahan berbasis komunitas.

Wilayah yang terklasifikasi sebagai Risiko Rendah (dengan skor $Sc \leq 0.2451$) harus difokuskan pada upaya preventif jangka panjang dan konservasi lingkungan untuk mempertahankan kondisi risiko rendah tersebut. Program reboisasi dan penghijauan intensif di daerah hulu (catchment area) yang memengaruhi wilayah ini, serta penegakan hukum terhadap penebangan liar dan pembuangan sampah sembarangan untuk menjaga fungsi resapan tanah.

KESIMPULAN

Penelitian ini menyimpulkan bahwa model *Extreme Gradient Boosting* (XGBoost) efektif dalam mengidentifikasi wilayah rawan banjir di Kabupaten Sigi dengan akurasi baseline sebesar 90,32%, membuktikan kemampuannya menangkap hubungan kompleks antar variabel lingkungan seperti curah hujan, limpasan permukaan, elevasi, kemiringan, dan jenis tanah. Optimasi hiperparameter menggunakan metode Particle Swarm Optimization (PSO) berhasil meningkatkan performa model secara signifikan dengan akurasi 95,2%, *F1-score* 0,929, dan *AUC* 0,950, sehingga menghasilkan model klasifikasi yang lebih akurat dan andal. Sebagai saran, penelitian lanjutan disarankan menambah variabel seperti tutupan lahan, jarak ke sungai, dan indikator sosial-ekonomi untuk memperkaya representasi risiko banjir. Selain itu, penggunaan metode optimasi alternatif seperti *Bayesian Optimization* atau *Genetic Algorithm* dapat dieksplorasi guna memperoleh konfigurasi parameter yang lebih adaptif dan efisien terhadap variasi data spasial dan temporal di masa mendatang.

REFERENSI

- Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2022). *Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees*. *Geocarto International*, 37(19), 5479–5496. <https://doi.org/10.1080/10106049.2021.1920636>.
- Badan Pusat Statistik. (2024). *Kabupaten Sigi Dalam Angka 2023*. BPS Kabupaten Sigi. <https://sigikab.bps.go.id/id/publication/2023/02/28/d0249f41cde7c5c02ff591d1/kabupaten-sigi-dalam-angka-2023.html>.
- Chang, Y. (2020). *Flash flood susceptibility assessment based on geodetector, certainty factor, and logistic regression analyses in Fujian Province, China*. *ISPRS International Journal of Geo-Information*, 9(12), 1–22. <https://doi.org/10.3390/ijgi9120748>.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>.

- El Haou, M., Ourribane, M., Ismaili, M., Abdelrahman, K., Fnais, M. S., Krimissa, S., ... & Namous, M. (2025). *Advanced GIS-based modeling for flood hazards mapping in urban semi-arid regions: Insights from Beni Mellal, Morocco*. *Frontiers in Environmental Science*, 13(June), 1–23. <https://doi.org/10.3389/fenvs.2025.1585926>.
- Gad, A. G. (2022). *Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review*. *Archives of Computational Methods in Engineering*, 29(5). <https://doi.org/10.1007/s11831-021-09694-4>.
- Hoang, D.-V., & Liou, Y.-A. (2024). *Elevating flash flood prediction accuracy: A synergistic approach with PSO and GA optimization*. *Natural Hazards and Earth System Sciences Discussions*, 1–26. <https://nhess.copernicus.org/preprints/nhess-2024-215/>
- IPCC. (2007). *Report from Intergovernmental Panel on Climate Change*. <http://www.ipcc.ch>
- Julian, B., & Roland, L. (2021). *U-FLOOD – Topographic deep learning for predicting urban pluvial flood water depth*. *Journal of Hydrology*, 603(August), 126898. <https://doi.org/10.1016/j.jhydrol.2021.126898>.
- Kanani-Sadat, Y., Safari, A., Nasser, M., & Homayouni, S. (2024). *A novel explainable PSO-XGBoost model for regional flood frequency analysis at a national scale: Exploring spatial heterogeneity in flood drivers*. *Journal of Hydrology*, 638, 131493. <https://doi.org/10.1016/j.jhydrol.2024.131493>.
- Maulita, M., Nurdin, N., & Taufiq, T. (2024). *Mapping of flood and landslide prone areas using composite mapping analysis method based on geographic information system in East Aceh*. *SISTEMASI*, 13(6), 2359–2374.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). *Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation*. *Environmental Modelling & Software*, 101, 1–9.
- Mohamad Herdian Bhakti, R., Saeful Bachri, O., & Sofian Efendi, F. (2021). *Optimasi K-Means dengan Particle Swarm Optimization pada Pengelompokan Daerah Stunting*. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 3(2), 95–101.
- Pusat Data Informasi dan Komunikasi Kebencanaan. (2024). *Buku Data Bencana Indonesia Tahun 2023*. Badan Nasional Penanggulangan Bencana (BNPB). <https://www.bnpb.go.id/buku/buku-data-bencana-indonesia-tahun-2023>.
- Pusat Krisis Kesehatan. (2024). *Banjir di Sigi, Sulawesi Tengah, 11-04-2024*. <https://penanggulangankrisis.kemkes.go.id/Banjir-di-SIGI-SULAWESI-TENGAH-11-04-2024-24>.
- Sarwar, J., Khan, S. A., Azmat, M., & Khan, F. (2024). *A comparative analysis of feature selection models for spatial analysis of floods using hybrid metaheuristic and machine learning models*. *Environmental Science and Pollution Research*, 31(23), 33495–33514.
- Sengupta, S., Basak, S., & Peters, R. A. (2019). *Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives*. *Machine Learning and Knowledge Extraction*, 1(1), 157–191. <https://doi.org/10.3390/make1010010>.
- Sigi, B. P. S. K. (2023). *Kabupaten Sigi Dalam Angka 2023*. BPS Kabupaten Sigi.
- Tuyen, D. N., Tuan, T. M., Son, L. H., Ngan, T. T., Giang, N. L., Thong, P. H., ... & Kanavos, A. (2021). *A novel approach combining particle swarm optimization and deep learning for flash flood detection from satellite images*. *Mathematics*, 9(22), 2846. <https://doi.org/10.3390/math9222846>.
- Widya, L. K., Rezaie, F., Lee, W., Lee, C.-W., Nurwatie, N., & Lee, S. (2024). *Flood susceptibility mapping of Cheongju, South Korea based on the integration of environmental factors using various machine learning approaches*. *Journal of*

- Environmental Management*, 364, 121291.
<https://doi.org/10.1016/j.jenvman.2023.121291>.
- Zeng, X., Long, J., Tian, S., & Xiao, G. (2023). *Random area pixel variation and random area transform for visible-infrared cross-modal pedestrian re-identification*. *Expert Systems with Applications*, 215, 119307. <https://doi.org/10.1016/j.eswa.2023.119307>.