STUDI ABLASI XGBOOST UNTUK KLASIFIKASI RISIKO BANJIR PERKOTAAN MENGGUNAKAN KARAKTERISTIK DRAINASE DAN DATA CURAH HUJAN CHIRPS

Evan Afdrianto Kota, Mokhamad Amin Hariyadi, Cahyo Crysdian
Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang
Jl. Gajayana No. 50, Jatimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur, Indonesia
evan@unmer.ac.id

ABSTRAK

Banjir perkotaan merupakan bencana yang semakin sering terjadi di Indonesia akibat urbanisasi dan perubahan pola curah hujan. Model klasifikasi risiko banjir umumnya mengandalkan variabel topografi dan hidrologi, namun jarang mengintegrasikan kondisi infrastruktur drainase. Praktisi sering menerapkan *preprocessing* kompleks pada model XGBoost tanpa memvalidasi kontribusi individual setiap langkah, sehingga belum diketahui apakah seleksi fitur, *oversampling*, dan optimasi *hyperparameter* benar-benar meningkatkan performa. Penelitian ini mengevaluasi pengaruh tiga komponen *preprocessing* terhadap performa XGBoost melalui studi ablasi enam skenario. Klasifikasi dilakukan pada 3.471 titik drainase di Kota Malang menggunakan 53 fitur dari data drainase dan curah hujan satelit CHIRPS (2019–2025). Enam skenario menguji seleksi fitur (Pearson+VIF+RFE), BorderlineSMOTE, dan *tuning* Optuna TPE secara individual maupun kombinatif. Model *baseline* tanpa *preprocessing* mencapai performa terbaik (F1 Macro 0,8167; Akurasi 0,8763). Seluruh intervensi justru menurunkan performa. Analisis SHAP mengidentifikasi frekuensi hari hujan dan interaksi curah hujan-kemiringan sebagai prediktor utama.

Kata kunci : *klasifikasi risiko banjir, XGBoost, studi ablasi, CHIRPS, SHAP, drainase*

1. PENDAHULUAN

Kejadian banjir di kota-kota Indonesia semakin sering dalam dekade terakhir. Urbanisasi yang cepat mengurangi area resapan air, sementara perubahan pola curah hujan meningkatkan intensitas hujan ekstrem [1] . Kota Malang di Jawa Timur mencatat sejumlah insiden banjir antara 2019 hingga 2023, sebagian besar terkonsentrasi di wilayah dengan infrastruktur drainase yang sudah tua atau berkapasitas kecil [2]. Pemetaan kerentanan banjir konvensional menggunakan variabel topografi dan hidrologi seperti elevasi, kemiringan, penggunaan lahan, dan intensitas curah hujan, namun jarang memperhitungkan kondisi dan kapasitas jaringan drainase terbangun. Bersabe dan Jun (2025) [3] menemukan bahwa penambahan variabel drainase pada model banjir di Seoul meningkatkan akurasi sebesar 7,58%. Dari sisi sumber data curah hujan, CHIRPS (*Climate Hazards Group InfraRed Precipitation with Station data*) menyediakan estimasi harian pada resolusi 0,05 derajat dan telah divalidasi untuk wilayah Asia Tenggara [4]. Jang et al. (2025) [5] menunjukkan bahwa fitur statistik curah hujan menghasilkan performa lebih baik dibandingkan total kumulatif sederhana pada model *Random Forest*.

XGBoost telah menjadi algoritma populer untuk klasifikasi kerentanan banjir [6], [7]. Namun praktisi sering menerapkan XGBoost dengan *pipeline preprocessing* yang rumit tanpa memvalidasi apakah setiap langkah benar-benar meningkatkan performa. Tiga intervensi yang sering muncul dalam literatur adalah seleksi fitur, *oversampling* sintetis (SMOTE), dan optimasi *hyperparameter* [8]. Kontribusi individual dan gabungan dari ketiga teknik ini jarang diisolasi secara terkontrol, sehingga belum diketahui apakah kompleksitas *preprocessing* tersebut memang diperlukan atau justru kontraproduktif.

Penelitian ini bertujuan mengevaluasi pengaruh tiga komponen *preprocessing* (seleksi fitur, BorderlineSMOTE, dan *tuning* Optuna TPE) terhadap performa XGBoost untuk klasifikasi risiko banjir perkotaan melalui studi ablasi enam skenario. Secara spesifik, penelitian ini menjawab tiga pertanyaan: (1) bagaimana performa XGBoost pada data gabungan drainase dan CHIRPS, (2) bagaimana pengaruh seleksi fitur, SMOTE, dan Optuna secara individual dan gabungan, serta (3) fitur mana yang paling berkontribusi menurut analisis SHAP.

Untuk menjawab pertanyaan tersebut, penelitian ini mengintegrasikan data karakteristik drainase dari 3.471 titik observasi dengan statistik curah hujan satelit CHIRPS periode 2019-2025 menjadi 53 fitur prediktor. Enam skenario ablasi dirancang untuk mengisolasi pengaruh setiap komponen *preprocessing*, mulai dari model *baseline* tanpa intervensi hingga *full pipeline* yang menggabungkan ketiga komponen. Analisis SHAP digunakan untuk menginterpretasi kontribusi fitur terhadap prediksi model.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Tabel 1 merangkum penelitian terdahulu yang relevan dengan topik klasifikasi risiko banjir menggunakan *machine learning*.

Tabel 1. Literatur penelitian terdahulu

No	Peneliti	Kategori Data	Algoritma	*Preprocessing*	*Optimization*	Performa
1	Riazi *et al.* (2023) [9]	SAR, CHIRPS, Topografi, Hidrologi	RBF, BA-RBF, RC-RBF, RSS-RBF	MI, Pearson, RFE	*Default parameters*	RC-RBF AUC=0,997
2	Wang *et al.* (2023) [6]	*Urban morphology*, Topografi, Hidrologi	XGBoost	VIF, Spearman, *Outlier detection*	5-*fold* CV	SHAP mean=0,0107
3	Yuan *et al.* (2024) [7]	*Urban config*, Topografi, Hidrologi	XGBoost	Pearson, *Outlier detection*, CV	Pareto *multi-objective*	R^2=0,82
4	Qin *et al.* (2025) [8]	Urbanisasi, Topografi, Hidrologi	XGBoost, CatBoost, LightGBM	SMOTE, *One-hot encoding*, RFE	*Grid Search*, 5-*fold StratifiedKFold*	XGBoost OA=0,96; CatBoost AUC=0,85
5	Shrestha *et al.* (2025) [10]	Topografi, Hidrologi, Curah hujan	LR, XGBoost, RF	ANOVA, RFE, *Balanced sampling*	*Manual tuning*, 10-*fold Stratified* CV	LR Acc=97,92%, AUC=0,9861
6	Zhu *et al.* (2024) [11]	*Remote sensing*, *Social sensing*	XGBoost, SVC, MLP, MDL, *Ensemble* RF	Pearson, RFE, *Resampling*	Tidak disebutkan	EMF Acc=0,940
7	Bersabe & Jun (2025) [3]	Topografi, Hidrologi, Drainase	LR, RF, SVM	VIF, Pearson, *Random sampling*	Tidak disebutkan	RF Acc=0,837, AUC=0,902
8	Ren *et al.* (2024) [12]	Iklim, Geomorfik, Antropogenik	RF, XGBoost, ANN, SVM	RFE, *Random negative sampling*	5-fold CV	RF AUC=0,87
9	Goumghar *et al.* (2025) [13]	GIS, *Remote sensing*, Geomorfologi	XGBoost, CatBoost, LightGBM, HGB	MI, RFE, VIF, Pearson	GridSearchCV, 5-fold CV	HGB OA=93,1%, AUC=0,833
10	Jang *et al.* (2025) [5]	Curah hujan, Simulasi drainase	Random Forest	Normalisasi, *Chi-square*	k-fold CV	R^2=0,976, RMSE=2,74

Berdasarkan Tabel 1, mayoritas penelitian kerentanan banjir menggunakan kombinasi data topografi dan hidrologi dengan algoritma *ensemble* atau *boosting*. XGBoost menjadi model yang paling sering digunakan, baik secara tunggal [6], [7] maupun sebagai pembanding [8], [10], [12]. Dari sisi *preprocessing*, seleksi fitur dengan Pearson, VIF, dan RFE merupakan teknik yang dominan, sementara penanganan ketidakseimbangan kelas hanya diterapkan oleh sebagian kecil penelitian. Riazi *et al.* (2023) dan Jang *et al.* (2025) [9], [5] telah memanfaatkan data CHIRPS dan curah hujan sebagai variabel prediktor, namun belum mengintegrasikannya dengan data infrastruktur drainase. Bersabe dan Jun (2025) [3] menjadi satu-satunya studi yang memasukkan variabel drainase, tetapi menggunakan algoritma LR, RF, dan SVM tanpa studi ablasi. Dari keseluruhan literatur, teridentifikasi dua celah utama: (1) belum ada studi yang mengintegrasikan data karakteristik drainase dengan curah hujan satelit CHIRPS sebagai fitur prediktor pada XGBoost, dan (2) kontribusi individual dan gabungan dari seleksi fitur, SMOTE, dan *tuning hyperparameter* belum pernah diisolasi melalui studi ablasi yang terkontrol. Penelitian ini mengisi kedua celah tersebut.

2.2. XGBoost

XGBoost (*eXtreme Gradient Boosting*) adalah implementasi *gradient boosting* teregulasi yang diperkenalkan oleh Chen dan Guestrin [14]. Algoritma ini menggunakan ekspansi Taylor orde kedua dari fungsi *loss* dan menambahkan regularisasi L1/L2 pada fungsi objektifnya. Pada setiap iterasi, XGBoost mengoptimasi:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i)\right) + \Omega(f_t) \quad (1)$$

di mana $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ adalah *term* regularisasi yang mengontrol kompleksitas pohon. Keunggulan XGBoost meliputi kemampuan menangani hubungan *non-linear*, *approximate split finding* untuk skalabilitas, dan penanganan *missing values* secara otomatis.

2.3. Seleksi Fitur (*Feature Selection*)

Seleksi fitur bertujuan menghapus variabel yang redundan atau tidak relevan untuk meningkatkan performa dan interpretabilitas model. Tiga pendekatan umum digunakan: filter, *wrapper*, dan embedded. Metode filter seperti korelasi *Pearson* dan *Variance Inflation Factor* (VIF) mengeliminasi fitur berdasarkan statistik univariat tanpa melibatkan model. Metode *wrapper* seperti *Recursive Feature Elimination* (RFE) mengevaluasi subset fitur menggunakan performa model sebagai kriteria pemilihan. Li et al. [15] melakukan survei komprehensif dan menyimpulkan bahwa kombinasi beberapa tahap seleksi fitur sering menghasilkan subset yang lebih stabil dibanding tahap tunggal. Theng dan Bhoyar [16] mengkaji lebih dari dua

dekade riset seleksi fitur dan mencatat bahwa pada *dataset* berdimensi tinggi, seleksi fitur agresif kadang justru menghilangkan sinyal prediktif yang terlihat redundan secara statistik namun informatif bagi model non-linear seperti *gradient boosting*.

2.4. Penanganan Ketidakseimbangan Kelas (SMOTE)

Ketidakseimbangan kelas merupakan masalah umum pada *dataset* banjir karena titik berisiko tinggi secara alami langka. SMOTE (*Synthetic Minority Over-sampling Technique*) [17] dan variannya seperti BorderlineSMOTE menghasilkan sampel sintetis kelas minoritas. BorderlineSMOTE secara spesifik menghasilkan sampel di sepanjang batas keputusan, di mana kesalahan klasifikasi cenderung terkonsentrasi. Namun beberapa penelitian menunjukkan bahwa SMOTE tidak selalu efektif, terutama pada ketidakseimbangan moderat.

2.5. Optimasi *Hyperparameter* dengan Optuna

Optuna [18] adalah *framework* optimasi *hyperparameter* yang menggunakan *Tree-structured Parzen Estimator* (TPE). Berbeda dengan grid search atau random search, TPE membangun model probabilistik dari ruang *hyperparameter* dan secara iteratif memilih konfigurasi yang menjanjikan. Almarzooq dan Waheed [19] menunjukkan bahwa Optuna secara konsisten menghasilkan konfigurasi yang lebih baik dibandingkan metode konvensional pada berbagai domain.

2.6. SHAP untuk Interpretabilitas Model

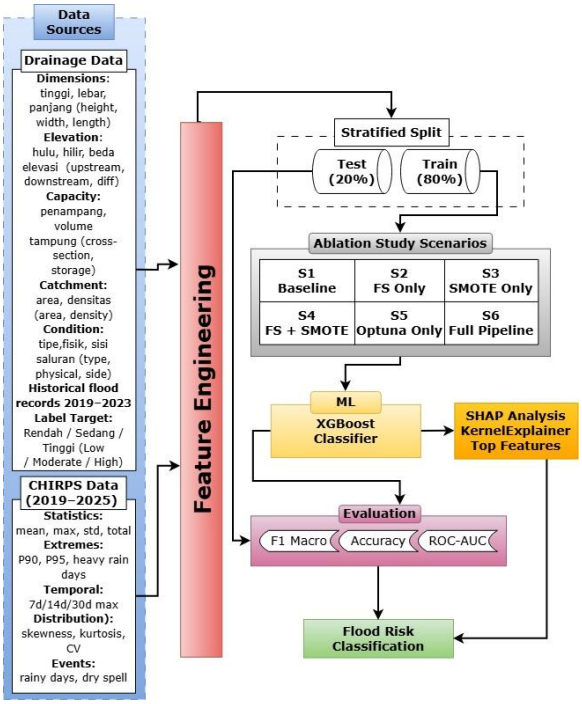
SHAP (*SHapley Additive exPlanations*) adalah metode interpretabilitas berbasis teori permainan yang menghitung kontribusi setiap fitur terhadap prediksi model. Wang et al. (2023) [6] menggunakan SHAP bersama XGBoost untuk menganalisis pengaruh konfigurasi perkotaan terhadap banjir. Qin et al. (2025) [8] mengkombinasikan XGBoost, CatBoost, dan LightGBM dengan SHAP untuk mengeksplorasi heterogenitas spasial kerentanan banjir. Aydin dan Iban [20] menerapkan SHAP pada model boosting untuk prediksi kerentanan banjir di Turki.

3. METODE PENELITIAN

3.1. Wilayah Studi dan *Dataset*

Wilayah studi adalah Kota Malang, Provinsi Jawa Timur, Indonesia. Data jaringan drainase diperoleh dari Dinas Pekerjaan Umum Kota Malang, terdiri dari 3.471 titik observasi dengan atribut dimensi pipa, jenis material, *rating* kondisi, kemiringan, konektivitas, dan karakteristik *catchment*. Data curah hujan diekstraksi dari *dataset* CHIRPS periode 2019–2025 (7 tahun, 2.557 hari) pada resolusi spasial 0,05 derajat, meliputi curah hujan rata-rata, deviasi standar, curah hujan harian maksimum, frekuensi di atas *threshold*, dan indeks variasi musiman. Total fitur yang dihasilkan adalah 53 variabel prediktor.

Label risiko banjir ditetapkan berdasarkan sistem penilaian komposit yang mempertimbangkan kapasitas drainase, catatan banjir historis, karakteristik terrain, dan paparan curah hujan, menghasilkan tiga kelas: Rendah, Sedang, dan Tinggi. Gambar 1 menunjukkan alur penelitian secara keseluruhan.



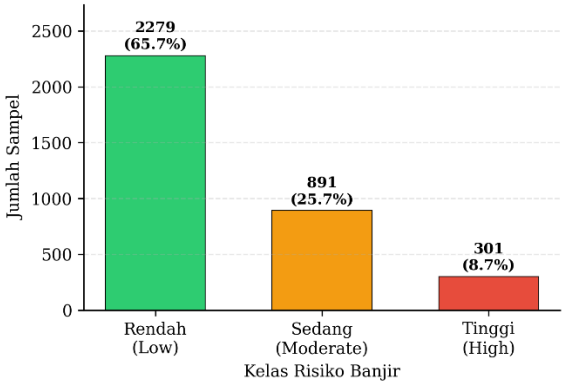
Gambar 1. Alur penelitian yang diusulkan

Tabel 2 menunjukkan distribusi kelas pada *dataset*.

Tabel 2. Distribusi kelas dataset risiko banjir

Kelas	Label	Jumlah	Proporsi (%)
Rendah	0	2.279	65,7
Sedang	1	891	25,7
Tinggi	2	301	8,7
Total		**3.471**	**100,0**

Dataset memiliki ketidakseimbangan kelas moderat. Kelas minoritas (Tinggi) menyumbang 8,7% dari seluruh observasi. Gambar 2 memvisualisasikan proporsi ketiga kelas.



Gambar 2. Distribusi kelas pada *dataset* risiko banjir

3.2. Pembagian Data

Dataset dibagi menjadi 80% data latih (2.776 sampel) dan 20% data uji (695 sampel) menggunakan *stratified random sampling*. Data uji tidak pernah terpapar perlakuan *preprocessing* apapun dan hanya digunakan untuk evaluasi akhir.

3.3. Desain Eksperimen Ablasi

Enam skenario dirancang untuk mengisolasi pengaruh tiga komponen *preprocessing*: seleksi fitur (FS), BorderlineSMOTE, dan *tuning* Optuna TPE. Tabel 3 menunjukkan matriks skenario.

Tabel 3. Desain skenario eksperimen ablasi

Skenario	Nama	FS	SMOTE	Optuna	Deskripsi
S1	*Baseline*	-	-	-	53 fitur, *hyperparameter* default
S2	FS *Only*	Ya	-	-	Corr + VIF + RFE → 15 fitur
S3	*SMOTE Only*	-	Ya	-	BorderlineSMOTE k=5
S4	FS + SMOTE	Ya	Ya	-	Seleksi fitur + *oversampling*
S5	*Optuna Only*	-	-	Ya	TPE 100 *trials*, semua fitur
S6	*Full Pipeline*	Ya	Ya	Ya	Ketiga komponen

Seleksi fitur mengikuti pipeline tiga tahap: filtering korelasi *Pearson* (*threshold* 0,95), analisis VIF (*threshold* 10), dan RFE dengan estimator XGBoost, mereduksi fitur dari 53 menjadi 15. BorderlineSMOTE dengan k=5 tetangga mengembangkan data latih dari 2.776 menjadi 5.469 sampel. Optuna TPE berjalan 100 trials dengan *5-fold stratified cross-validation.*

3.4. Metrik Evaluasi

Model dievaluasi menggunakan F1 Macro sebagai metrik utama, karena memberikan bobot setara pada setiap kelas tanpa didominasi kelas mayoritas. Metrik tambahan meliputi Accuracy dan ROC-AUC (*one-vs-rest*) [21].

$$F1_{Macro} = \frac{1}{K}\sum_{k=1}^{K} F1_k \quad (2)$$

di mana K adalah jumlah kelas dan $F1_k$ adalah F1-score untuk kelas k.

4. HASIL DAN PEMBAHASAN

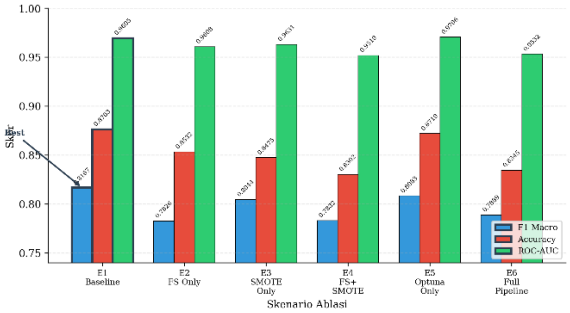
4.1. Performa XGBoost pada Enam Skenario Ablasi

Tabel 4 menyajikan hasil evaluasi XGBoost pada seluruh skenario.

Tabel 4. Hasil evaluasi XGBoost pada enam skenario ablasi

Skenario	Accuracy	F1 Macro	ROC-AUC
S1 (*Baseline*)	0,8763	0,8167	0,9695
S2 (FS *Only*)	0,8532	0,7826	0,9608
S3 (SMOTE *Only*)	0,8475	0,8044	0,9631
S4 (FS + SMOTE)	0,8302	0,7832	0,9519
S5 (*Optuna Only*)	0,8719	0,8083	0,9706
S6 (*Full Pipeline*)	0,8345	0,7889	0,9532

S1 Baseline mencapai performa terbaik secara keseluruhan dengan F1 Macro tertinggi (0,8167) dan Accuracy (0,8763). Tidak satupun intervensi *preprocessing* yang berhasil memperbaiki angka-angka ini. Gambar 3 memvisualisasikan perbandingan F1 Macro dan Accuracy antar skenario.



Gambar 3. Perbandingan F1 Macro dan Accuracy antar skenario ablasi

4.2. Analisis Pengaruh Masing-masing Komponen

Seleksi fitur menurunkan performa. F1 Macro turun dari 0,8167 (S1) menjadi 0,7826 (S2), penurunan 3,41 poin persentase. Pemotongan dari 53 ke 15 fitur menghilangkan terlalu banyak informasi, terutama statistik curah hujan CHIRPS dan metrik drainase sekunder yang ternyata membawa sinyal prediktif non-redundan.

BorderlineSMOTE juga tidak membantu. F1 Macro turun dari 0,8167 (S1) menjadi 0,8044 (S3), dan Accuracy menurun sekitar 3 poin persentase. Dengan 301 sampel kelas Tinggi, XGBoost sudah memiliki cukup contoh untuk mempelajari batas keputusan yang wajar. SMOTE menggelembungkan data latih dengan titik sintetis yang justru memperkenalkan *noise* di dekat batas kelas.

Tuning Optuna sedikit meningkatkan ROC-AUC (0,9695 → 0,9706), tetapi F1 Macro justru turun menjadi 0,8083. Konfigurasi hasil *tuning* kemungkinan mengalami sedikit *overfitting* pada *fold cross-validation.*

Pipeline lengkap (S6) justru menghasilkan performa terburuk kedua dengan F1 Macro 0,7889. Seleksi fitur menghilangkan informasi, SMOTE menambah *noise*, dan Optuna mengoptimasi pada *dataset* yang sudah terdegradasi. Menumpuk *preprocessing* tanpa validasi per langkah memperburuk keadaan secara signifikan.

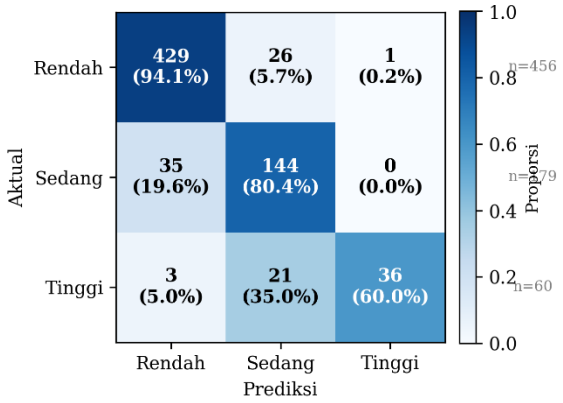
4.3. Analisis *Confusion Matrix*

Tabel 5 menyajikan *confusion matrix* untuk model terbaik (S1 *Baseline*).

Tabel 5. Confusion matrix XGBoost S1

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah (n=456)	429	26	1
Sedang (n=179)	35	144	0
Tinggi (n=60)	3	21	36

Kesalahan klasifikasi terkonsentrasi antara kelas bersebelahan. Dari 27 sampel Rendah yang salah, 26 diprediksi Sedang dan hanya 1 sebagai Tinggi. Dari 24 sampel Tinggi yang salah, 21 masuk ke Sedang dan hanya 3 ke Rendah. Pola ini wajar untuk variabel risiko ordinal dengan batas antar kelas yang gradual. Gambar 4 memvisualisasikan *confusion matrix* dalam bentuk *heatmap*.



Gambar 4. *Heatmap confusion matrix* model XGBoost S1

4.4. Analisis SHAP

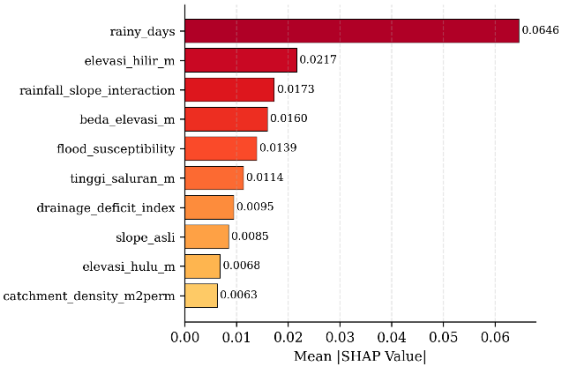
Analisis SHAP dilakukan pada model S1 menggunakan *KernelExplainer* dengan 100 sampel uji dan 50 sampel latar belakang. Tabel 6 menunjukkan 10 fitur teratas berdasarkan *mean* |SHAP *value*|.

Tabel 6. Sepuluh fitur teratas berdasarkan *mean* |SHAP *value*|

Peringkat	Fitur	*Mean* |SHAP|
1	*rainy_days*	0,066
2	*rainfall_slope_interaction*	0,021
3	elevasi_hilir_m	0,021
4	beda_elevasi_m	0,013
5	*flood_susceptibility*	0,013
6	tinggi_saluran_m	0,011
7	elevasi_hulu_m	0,010
8	slope_asli	0,008
9	*drainage_deficit_index*	0,008
10	panjang_saluran_m	0,007

Fitur *rainy_days* (jumlah hari hujan selama periode 2019-2025) mendominasi peringkat pertama dengan nilai SHAP jauh di atas fitur lainnya, mengonfirmasi bahwa frekuensi hari hujan merupakan pendorong utama risiko banjir di Kota Malang. Fitur interaksi *rainfall_slope_interaction* menempati peringkat kedua, menunjukkan bahwa kombinasi intensitas curah hujan dan kemiringan terrain memberikan sinyal prediktif yang kuat. Indeks komposit *flood_susceptibility* menempati peringkat kelima, menunjukkan bahwa penggabungan variabel drainase dan terrain ke dalam satu skor kerentanan berhasil menangkap informasi yang tidak tertangkap variabel individual.

Kehadiran fitur drainase fisik (elevasi_hilir_m, tinggi_saluran_m, beda_elevasi_m, panjang_saluran_m) dan fitur terrain (slope_asli, elevasi_hulu_m) dalam 10 besar menunjukkan bahwa data infrastruktur drainase memberikan daya prediktif yang saling melengkapi dengan data curah hujan. Gambar 5 menampilkan SHAP *summary plot* untuk model S1.



Gambar 5. SHAP *summary plot* untuk model XGBoost S1

5. KESIMPULAN DAN SARAN

Penelitian ini menguji apakah *preprocessing* yang umum digunakan benar-benar memperbaiki XGBoost untuk klasifikasi risiko banjir perkotaan. Studi ablasi enam skenario pada 3.471 titik drainase di Kota Malang menunjukkan bahwa model *baseline* dengan 53 fitur dan *hyperparameter* default mencapai performa terbaik (F1 Macro 0,8167, Akurasi 0,8763). Seleksi fitur, BorderlineSMOTE, dan *tuning* Optuna TPE seluruhnya gagal meningkatkan performa, bahkan menurunkannya ketika dikombinasikan. Analisis SHAP mengonfirmasi bahwa frekuensi hari hujan (*rainy_days*) dan fitur interaksi curah hujan-kemiringan merupakan prediktor dominan, sementara kehadiran fitur drainase dalam 10 besar memvalidasi pentingnya data infrastruktur drainase dalam pemodelan risiko banjir. Penelitian selanjutnya disarankan mengeksplorasi alternatif *deep learning* serta menguji generalisasi model pada jaringan drainase di kota lain.

DAFTAR PUSTAKA

[1] R. L. Puspasari, D. Yoon, H. Kim, and K.-W. Kim, “Machine Learning for Flood Prediction in Indonesia: Providing Online Access for Disaster Management Control,” *Economic and Environmental Geology*, vol. 56, no. 1, pp. 65–73, Feb. 2023, doi: 10.9719/EEG.2023.56.1.65.

[2] BPBD Kota Malang, "Laporan Kejadian Bencana Banjir Kota Malang 2019-2023," Kota Malang, 2023.

[3] J. T. Bersabe and B.-W. Jun, "The Machine Learning-Based Mapping of Urban Pluvial Flood Susceptibility in Seoul Integrating Flood Conditioning Factors and Drainage-Related Data," *ISPRS Int. J. Geoinf.*, vol. 14, no. 2, 2025, doi: 10.3390/ijgi14020057.

[4] C. Funk *et al.*, "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," *Sci. Data*, vol. 2, no. 1, p. 150066, 2015, doi: 10.1038/sdata.2015.66.

[5] S.-D. Jang, J.-H. Yoo, Y.-S. Lee, and B. Kim, "Flood prediction in urban areas based on machine learning considering the statistical characteristics of rainfall," *Progress in Disaster Science*, vol. 26, p. 100415, 2025, doi: 10.1016/j.pdisas.2025.100415.

[6] M. Wang *et al.*, "An XGBoost-SHAP approach to quantifying morphological impact on urban flooding susceptibility," *Ecol. Indic.*, vol. 156, p. 111137, 2023, doi: 10.1016/j.ecolind.2023.111137.

[7] H. Yuan *et al.*, "Data-driven urban configuration optimization: An XGBoost-based approach for mitigating flood susceptibility and enhancing economic contribution," *Ecol. Indic.*, vol. 166, p. 112247, 2024, doi: 10.1016/j.ecolind.2024.112247.

[8] Y. Qin, Y. Yu, J. Liu, and R. Liu, "Machine learning-based identification of key factors and spatial heterogeneity analysis of urban flooding: a case study of the central urban area of Ordos," *Sci. Rep.*, vol. 15, no. 1, p. 24749, 2025, doi: 10.1038/s41598-025-08162-4.

[9] M. Riazi *et al.*, "Enhancing flood susceptibility modeling using multi-temporal SAR images, CHIRPS data, and hybrid machine learning algorithms," *Science of The Total Environment*, vol. 871, p. 162066, 2023, doi: 10.1016/j.scitotenv.2023.162066.

[10] S. Shrestha, D. Dahal, N. Bhattarai, S. Regmi, R. Sewa, and A. Kalra, "Machine Learning-Based Flood Risk Assessment in Urban Watershed: Mapping Flood Susceptibility in Charlotte, North Carolina," *Geographies*, vol. 5, no. 3, 2025, doi: 10.3390/geographies5030043.

[11] X. Zhu, H. Guo, and J. J. Huang, "Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model," *Sustain. Cities Soc.*, vol. 108, p. 105508, 2024, doi: 10.1016/j.scs.2024.105508.

[12] H. Ren *et al.*, "Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost)," *Remote Sens. (Basel).*, vol. 16, no. 2, 2024, doi: 10.3390/rs16020320.

[13] L. Goumghar *et al.*, "Analysis of Baseline and Novel Boosting Models for Flood-Prone Prediction and Explainability: Case from the Upper Drâa Basin (Morocco)," *Earth*, vol. 6, no. 3, 2025, doi: 10.3390/earth6030069.

[14] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[15] J. Li *et al.*, "Feature Selection: A Data Perspective," *ACM Comput. Surv.*, vol. 50, no. 6, Dec. 2017, doi: 10.1145/3136625.

[16] D. Theng and K. K. Bhoyar, "Feature selection techniques for machine learning: a survey of more than two decades of research," *Knowl. Inf. Syst.*, vol. 66, no. 3, pp. 1575–1637, Dec. 2023, doi: 10.1007/s10115-023-02010-5.

[17] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[18] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.

[19] H. Almarzooq and U. bin Waheed, "Automating hyperparameter optimization in geophysics with Optuna: A comparative study," *Geophys. Prospect.*, vol. 72, no. 5, pp. 1778–1788, Jun. 2024, doi: 10.1111/1365-2478.13484.

[20] H. E. Aydin and M. C. Iban, "Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations," *Natural Hazards*, vol. 116, no. 3, pp. 2957–2991, 2023, doi: 10.1007/s11069-022-05793-y.

[21] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.