Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644

Aplikasi Prediksi Banjir Menggunakan Algoritma XGBoost Berbasis Website

Muhamad Fuat Asnawi [1)], Hadi Hikmadyo Bisono [2)], Muhamad Arldi Megantara [3)], Kusrini [4)]
[1,2,3)] Mahasiswa Magister PJJ Informatika Universitas Amikom Yogyakarta
[1)] Universitas Sains Al-Qur'an
[4)] Universitas Amikom Yogyakarta
[1)] fuatasnawi@unsiq.ac.id
[2)] hadihb@students.amikom.ac.id
[3)] arldimegantara@students.amikom.ac.id
[4)] kusrini@amikom.ac.id

Abstrak

Penelitian ini bertujuan untuk mengembangkan model prediksi risiko banjir menggunakan algoritma XGBoost dengan memanfaatkan dataset yang tersedia di Kaggle. Dataset tersebut mencakup berbagai faktor yang mempengaruhi risiko banjir seperti kualitas bendungan, pengikisan sistem drainase, longsor, dan hilangnya lahan basah. Proses penelitian dimulai dengan pengumpulan data, diikuti oleh preprocessing yang meliputi penanganan missing values, pemilihan fitur menggunakan regresi untuk memastikan fitur yang paling berpengaruh, dan normalisasi data. Model XGBoost kemudian dilatih dengan data yang telah diproses dan dievaluasi menggunakan beberapa metrik evaluasi. Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik dengan nilai Cross-Validation RMSE sebesar 0.00097, Mean Squared Error (MSE) sebesar 1.0336, Root Mean Squared Error (RMSE) sebesar 0.001017, Mean Absolute Error (MAE) sebesar 0.000801, dan Mean Absolute Percentage Error (MAPE) sebesar 0.1605%. Nilai-nilai ini mengindikasikan kesalahan prediksi yang relatif kecil. Visualisasi hasil juga menunjukkan bahwa model tidak memiliki bias sistematis dan kesalahan prediksi tersebar merata. Penelitian ini mendesak mengingat peningkatan frekuensi dan dampak banjir akibat perubahan iklim dan urbanisasi yang pesat. Model ini diharapkan dapat digunakan secara efektif untuk memberikan peringatan dini dan membantu dalam perencanaan tata ruang yang lebih baik untuk mengurangi dampak bencana banjir.

Kata kunci : Prediksi Banjir, XGBoost, Pembelajaran Mesin, Data Preprocessing, Evaluasi Model

Abstract

This study aims to develop a flood risk prediction model using the XGBoost algorithm by utilizing a dataset available on Kaggle. The dataset includes various factors affecting flood risk such as dam quality, encroachments, drainage systems, landslides, and wetland loss. The research process begins with data collection, followed by preprocessing that involves handling missing values, feature selection using regression to identify the most influential features, and data normalization. The XGBoost model is then trained with the processed data and evaluated using several evaluation metrics. The evaluation results show that the model performs very well with a Cross-Validation RMSE of 0.00097, Mean Squared Error (MSE) of 1.0336, Root Mean Squared Error (RMSE) of 0.001017, Mean Absolute Error (MAE) of 0.000801, and Mean Absolute Percentage Error (MAPE) of 0.1605%. These values indicate relatively small prediction errors. The visualization of results also shows that the model does not have systematic bias and prediction errors are evenly distributed. This study is urgent given the increasing frequency and impact of floods due to rapid climate change and urbanization. This model is expected to be effectively used for early warning and aid in better spatial planning to mitigate the impact of flood disasters.

Keywords*: Flood Prediction, XGBoost, Machine Learning, Data Preprocessing, Model Evaluation*

1. PENDAHULUAN

Banjir merupakan salah satu bencana alam yang sering terjadi dan dapat menyebabkan kerugian materiil serta korban jiwa yang signifikan. Seiring dengan perubahan iklim dan urbanisasi yang pesat, risiko terjadinya

banjir semakin meningkat. Penilaian risiko banjir yang akurat dan efektif menjadi sangat penting untuk mengurangi dampak bencana ini. Namun, sebagian besar metode pembelajaran mesin yang ada saat ini masih bergantung pada classifier tunggal yang cenderung kurang akurat dalam memproses data skala besar (Riza et al., 2020).

Metode ensemble learning, seperti XGBoost (Extreme Gradient Boosting), telah terbukti efektif dalam meningkatkan akurasi prediksi melalui kombinasi beberapa model dasar. XGBoost tidak hanya mengoptimalkan proses perhitungan tetapi juga memanfaatkan multi-threading CPU untuk mempercepat pelatihan model, sehingga sangat cocok untuk aplikasi dengan data besar dan kompleks(Ahmed et al., 2021). Algoritma ini dikenal karena kemampuannya untuk menghasilkan prediksi yang lebih akurat dibandingkan dengan metode pembelajaran mesin konvensional lainnya, seperti jaringan Bayesian dan Support Vector Machine (SVM) (Razali et al., 2020).

Penggunaan teknologi penginderaan jauh (remote sensing) dan Sistem Informasi Geografis (GIS) juga telah menunjukkan hasil yang baik dalam pemantauan dan evaluasi risiko banjir. Data dari penginderaan jauh dapat digunakan untuk mengidentifikasi area yang rentan terhadap banjir, sementara GIS memungkinkan analisis spasial yang mendalam terhadap faktor-faktor yang mempengaruhi risiko banjir (Zhu & Zhang, 2022).Kombinasi XGBoost dengan data penginderaan jauh dapat meningkatkan akurasi prediksi dan memungkinkan pembuatan peta risiko banjir yang lebih detail dan andal (Ma et al., 2021).

Penelitian sebelumnya oleh (Razali et al., 2020)mengembangkan model prediksi risiko banjir menggunakan jaringan Bayesian dan teknik ML lainnya seperti Decision Tree (DT), k-Nearest Neighbours (kNN), dan Support Vector Machine (SVM). Model ini diterapkan di Kuala Krai, Kelantan, Malaysia dengan data dari periode 2012 hingga 2016. Hasil penelitian menunjukkan bahwa metode Decision Tree dengan SMOTE menghasilkan akurasi tertinggi sebesar 99,92% dalam menangani dataset yang tidak seimbang. Ini menunjukkan bahwa metode penyeimbangan data seperti SMOTE sangat efektif dalam meningkatkan akurasi model prediksi banjir. Selain itu, penelitian oleh (Ren et al., 2024) menggunakan data penginderaan jauh dan GIS untuk memantau serta mengevaluasi risiko banjir, dan mereka memanfaatkan algoritma Random Forest dan XGBoost untuk menghasilkan peta risiko banjir yang lebih akurat dan andal.

Penelitian terbaru oleh (Yuan et al., 2024) menggunakan algoritma XGBoost untuk memprediksi risiko banjir dengan dataset dari Kaggle. Hasil evaluasi menunjukkan bahwa model ini memiliki performa yang sangat baik dengan nilai RMSE yang rendah, mengindikasikan kesalahan prediksi yang kecil. Visualisasi hasil prediksi dan nilai aktual menunjukkan bahwa model ini tidak memiliki bias sistematis dan kesalahan prediksi tersebar merata. Model ini diharapkan dapat digunakan untuk sistem peringatan dini dan perencanaan tata ruang yang lebih baik.

Penelitian ini bertujuan untuk mengembangkan model prediksi risiko banjir menggunakan algoritma XGBoost dengan memanfaatkan dataset yang tersedia di Kaggle. Dataset tersebut mencakup berbagai faktor yang mempengaruhi risiko banjir, seperti intensitas musim hujan, drainase topografi, pengelolaan sungai, deforestasi, urbanisasi, perubahan iklim, dan kualitas bendungan.

Penelitian ini sangat mendesak mengingat meningkatnya frekuensi dan dampak bencana banjir akibat perubahan iklim dan urbanisasi yang pesat. Dengan mengembangkan model prediksi yang lebih akurat dan andal, penelitian ini diharapkan dapat berkontribusi pada mitigasi risiko banjir dan perencanaan tata ruang yang lebih baik di masa depan.

2. KAJIAN PUSTAKA

Banjir merupakan salah satu bencana alam yang sering terjadi dan menyebabkan kerugian materiil serta korban jiwa yang signifikan di berbagai belahan dunia. Penelitian mengenai prediksi risiko banjir menjadi sangat penting untuk mengurangi dampak bencana ini. Metode pembelajaran mesin (ML) telah banyak digunakan dalam upaya ini, dengan berbagai pendekatan yang menawarkan tingkat akurasi yang berbeda.

Razali et al. (2020) mengembangkan model prediksi risiko banjir menggunakan jaringan Bayesian dan teknik ML lainnya seperti Decision Tree (DT), k-Nearest Neighbours (kNN), dan Support Vector Machine (SVM). Penelitian mereka menunjukkan bahwa metode Decision Tree dengan SMOTE menghasilkan akurasi tertinggi sebesar 99,92% dalam menangani dataset tidak seimbang(1). Penelitian ini menggarisbawahi pentingnya metode penyeimbangan data seperti SMOTE untuk meningkatkan akurasi model.

Selain itu, penelitian oleh (Ren et al., 2024)menggunakan data penginderaan jauh dan GIS untuk memantau serta mengevaluasi risiko banjir. Dengan memanfaatkan algoritma Random Forest dan XGBoost, penelitian ini berhasil menghasilkan peta risiko banjir yang lebih akurat dan andal. Pendekatan ini

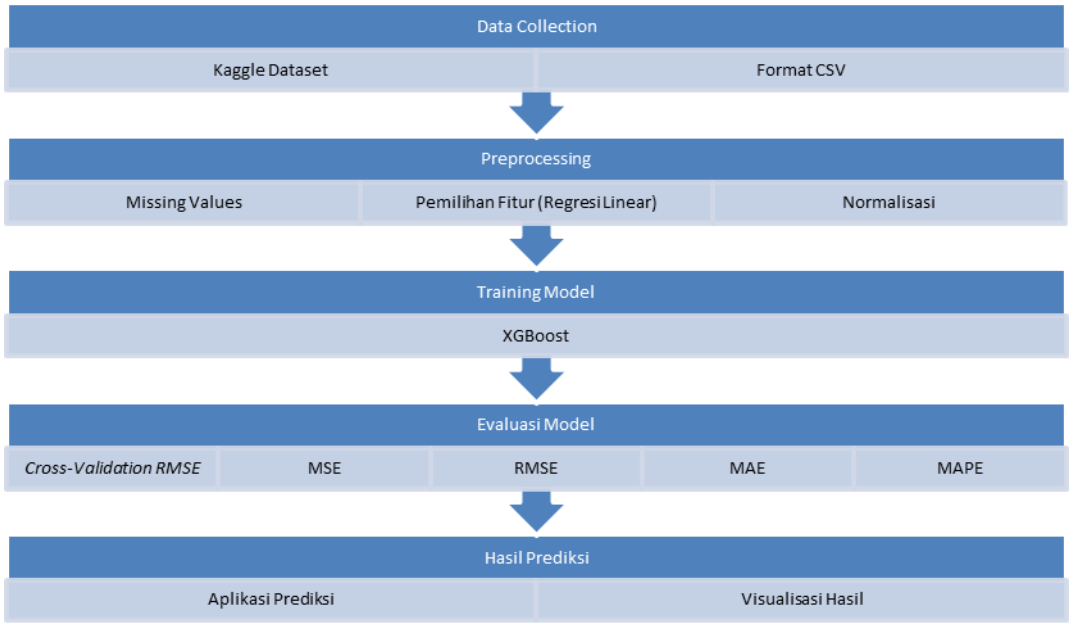
menunjukkan bahwa integrasi data spasial dengan teknik ML dapat meningkatkan kemampuan prediksi risiko banjir.

Penelitian terbaru oleh (Yuan et al., 2024)menggunakan algoritma XGBoost untuk memprediksi risiko banjir dengan dataset yang diambil dari Kaggle. Hasil evaluasi menunjukkan bahwa model ini memiliki performa yang sangat baik dengan nilai RMSE yang rendah, mengindikasikan kesalahan prediksi yang kecil. Visualisasi hasil prediksi dan nilai aktual menunjukkan bahwa model tidak memiliki bias sistematis dan kesalahan prediksi tersebar merata. Model ini diharapkan dapat digunakan untuk sistem peringatan dini dan perencanaan tata ruang yang lebih baik.

Dengan menggabungkan berbagai metode pembelajaran mesin dan memanfaatkan data spasial yang luas, penelitian-penelitian ini menawarkan pandangan baru dalam prediksi risiko banjir. Kebaruan penggunaan metode seperti XGBoost menunjukkan potensi besar dalam meningkatkan akurasi dan efektivitas model prediksi banjir. Penelitian-penelitian ini memberikan dasar yang kuat untuk pengembangan sistem peringatan dini yang lebih baik dan perencanaan tata ruang yang lebih efisien, guna mengurangi dampak bencana banjir di masa depan

3. METODOLOGI PENELITIAN

Alur penelitian yang dilakukan dalam penelitian ini, dimulai dari pengumpulan data, preprocessing, pelatihan model XGBoost, hingga evaluasi model dan visualisasi hasil. Alur tersebut dalam dilihat pada gambar 1.



Gambar 1. Alur Penelitian (Nguyen et al., 2021)

Pada gambar 1. Proses dimulai dengan pengumpulan data dari Kaggle dalam format CSV. Selanjutnya, data mengalami preprocessing yang meliputi penanganan missing values, pemilihan fitur yang relevan menggunakan regresi linear, dan normalisasi data. Setelah data siap, model XGBoost digunakan untuk melatih data tersebut. Evaluasi model dilakukan menggunakan metrik seperti *Cross-Validation RMSE*, *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE) (Ibrahem Ahmed Osman et al., 2021). Hasil prediksi kemudian diaplikasikan dan divisualisasikan untuk memberikan gambaran tentang kinerja model dan nilai prediksi dibandingkan dengan nilai aktual. Diagram ini memberikan gambaran yang jelas dan terstruktur tentang langkah-langkah yang diambil dalam penelitian prediksi banjir menggunakan XGBoost (Nguyen et al., 2021).

a. Metode Analisis Data

Analisis data dimulai dengan deskripsi dataset yang mencakup jumlah data, jenis fitur, dan distribusi data untuk memberikan konteks. Preprocessing data melibatkan penanganan missing values menggunakan

imputasi mean, pemilihan fitur dengan regresi linear untuk menentukan fitur signifikan, dan normalisasi menggunakan StandardScaler. Model XGBoost dipilih karena kemampuannya menangani dataset besar dan kompleks, dan hyperparameter tuning dilakukan menggunakan GridSearchCV untuk optimasi performa. Evaluasi model menggunakan metrik seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE) untuk menilai kinerja, dengan penjelasan bahwa metrik ini penting untuk mengukur kesalahan prediksi dan akurasi model (Ahmadi et al., 2024).

b. Metode Pengambilan Kesimpulan

Interpretasi hasil evaluasi model akan dilakukan berdasarkan metrik yang digunakan, seperti nilai RMSE yang rendah menunjukkan kesalahan prediksi kecil. Hasil penelitian ini akan dibandingkan dengan penelitian sebelumnya untuk menilai peningkatan akurasi model. Visualisasi hasil, seperti scatter plot, akan digunakan untuk menunjukkan hubungan antara nilai aktual dan prediksi, sementara histogram residual akan menampilkan distribusi kesalahan prediksi. Kesimpulan akan dirumuskan berdasarkan temuan utama, seperti implikasi praktis model ini dalam sistem peringatan dini dan perencanaan tata ruang yang lebih baik untuk mengurangi dampak banjir (Le & Thu Hien, 2024).

c. Metode Regresi Linear

Dalam memilih fitur yang paling berpengaruh terhadap prediksi banjir maka peneliti menggunakan metode regresi linear. Dalam regresi linear, feature importance dapat dihitung berdasarkan nilai koefisien regresi. Fitur dengan koefisien absolut yang lebih besar dianggap lebih penting karena memiliki pengaruh yang lebih besar terhadap variabel dependen (target) (Khaire & Dhanalakshmi, 2022).

Metode regresi linear yang digunakan dalam pemilihan fitur, metode tersebut dalam ditunjukan dalam persamaan 1.

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_nx_n \qquad (1)$$

Dimana:

- y adalah variabel dependen
- $x_1, x_2, \ldots, x_n$ adalah variabel independent
- b_0 adalah intersep
- $b_1, b_2, \ldots, b_n$ adalah koefisien regresi yang menunjukkan pengaruh masing-masing fitur

d. Metode XGBoost

XGBoost membangun model prediksi sebagai kombinasi dari banyak pohon keputusan yang lemah. Setiap pohon keputusan ditambahkan secara bertahap, dan setiap penambahan pohon bertujuan untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya (Joshi et al., 2024).

Model prediksi pada iterasi t ditunjukan dalam persamaan 2.

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) \qquad (2)$$

Dimana:

- $\hat{y}_i^{(t)}$ adalah prediksi pada iterasi -t untuk sampai i
- f_k adalah pohon Keputusan pada iterasi ke-k
- x_i adalah fitur untuk sampel i

4. HASIL DAN PEMBAHASAN

a. Deskripsi Data

Data yang digunakan dalam penelitian ini diambil dari dataset yang tersedia di Kaggle, yang berisi berbagai faktor yang mempengaruhi probabilitas terjadinya banjir. Faktor-faktor tersebut adalah Intensitas Muson, Topografi dan Drainase, Manajemen Sungai, Deforestasi, Urbanisasi, Perubahan Iklim, Kualitas Bendungan, Sedimentasi, Praktik Pertanian, Perambahan, Kesiapsiagaan Bencana yang Tidak Efektif, Sistem Drainase, Kerentanan Pesisir, Longsor, Daerah Aliran Sungai, Infrastruktur yang Memburuk, Skor Populasi, Kehilangan Lahan Basah, Perencanaan yang Tidak Memadai, Faktor Politik. Dataset tersebut dalam format CSV berjumlah 50000 data dan telah diunduh serta diproses menggunakan bahasa pemrograman Python.

Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644

Tabel 1. Dataset

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	3	8	6	6	4	4	6	2	3	2	5	10	7	4	2	3	4	3	2	6	0.45
2	8	4	5	7	7	9	1	5	5	4	6	9	2	6	2	1	1	9	1	3	0.475
3	3	10	4	1	7	5	4	7	4	9	2	7	4	4	8	6	1	8	3	6	0.515
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
49997	3	10	3	8	3	3	4	4	3	11	8	8	6	3	6	4	4	2	4	5	0.51
49998	4	4	5	7	2	1	4	5	6	7	7	4	6	4	1	5	1	6	4	3	0.43
49999	4	5	4	4	6	3	10	2	6	11	5	6	3	4	7	6	2	4	0	11	0.515
50000	4	5	6	3	5	6	5	4	9	10	6	2	4	4	5	6	7	8	10	7	0.58

Pada tabel 1 adalah gambaran dari dataset yang digunakan dalam penelitian ini dengan kolom 1-21 terdiri dari Intensitas Muson, Topografi dan Drainase, Manajemen Sungai, Deforestasi, Urbanisasi, Perubahan Iklim, Kualitas Bendungan, Sedimentasi, Praktik Pertanian, Perambahan, Kesiapsiagaan Bencana yang Tidak Efektif, Sistem Drainase, Kerentanan Pesisir, Longsor, Daerah Aliran Sungai, Infrastruktur yang Memburuk, Skor Populasi, Kehilangan Lahan Basah, Perencanaan yang Tidak Memadai, Faktor Politik dan prediksi banjir.

b. Preprocessing Data

Langkah awal dalam preprocessing data melibatkan pemeriksaan nilai yang hilang (*missing values*) dan penanganan nilai tersebut dengan metode imputasi menggunakan "*SimpleImputer*". Selain itu, fitur-fitur yang relevan telah dipilih menggunakan metode regresi linear untuk memastikan fitur mana yang paling berpengaruh dalam prediksi (Lee & Lee, 2021). Setelah fitur dipilih, data dinormalisasi menggunakan StandardScaler untuk memastikan skala yang konsisten antara fitur-fitur yang berbeda (Mohammad Asif Syeed et al., 2022). Berikut ini adalah langkah preprocessing data:

1) Penanganan Data yang Hilang : Kolom dengan data yang hilang diidentifikasi dan diimput menggunakan mean.
2) Pemilihan Fitur dan Variabel Target: Beberapa kolom dipilih sebagai fitur dan satu kolom sebagai target.
3) Penghapusan Baris dengan Nilai Hilang: Baris dengan nilai hilang dihapus dari dataset.
4) Standarisasi Fitur: Fitur numerik diubah skalanya untuk memiliki mean 0 dan standar deviasi

c. Pemilihan Fitur

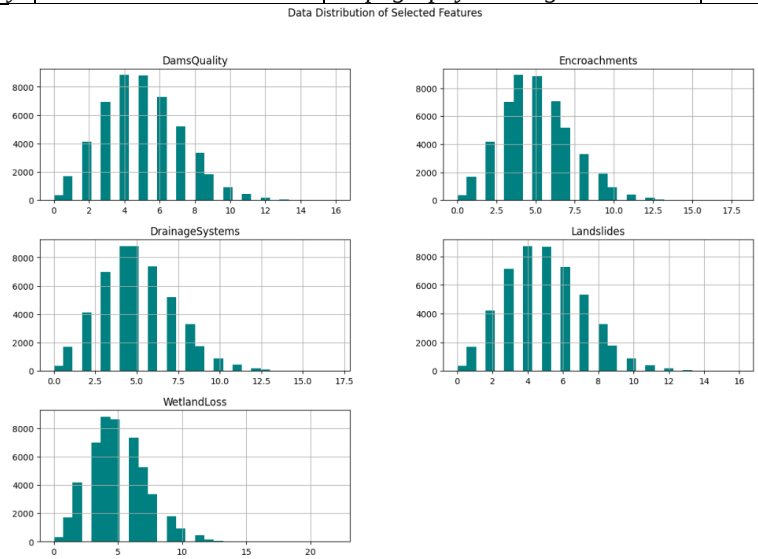
Pemilihan fitur dilakukan menggunakan regresi linear untuk menentukan fitur-fitur yang memiliki pengaruh paling signifikan terhadap prediksi probabilitas banjir (Liu et al., 2022). Proses ini melibatkan analisis regresi untuk menilai kontribusi masing-masing fitur terhadap variabel target, pada tabel 2 merupakan hasil analisis regresi. Fitur-fitur yang dipilih berdasarkan analisis ini adalah: *DamsQuality*, *Encroachments*, *DrainageSystems*, *Landslides*, dan *WetlandLoss*. Fitur-fitur ini dianggap paling relevan dan memberikan kontribusi signifikan terhadap prediksi model.

Tabel 2. Hasil Regresi

Feature	*Importance*	*Feature*	*Importance*
DamsQuality	0.00500000000000001	*PopulationScore*	0.004999999999999998
Encroachments	0.005000000000000009	*InadequatePlanning*	0.0049999999999999975
WetlandLoss	0.005000000000000008	*IneffectiveDisasterPreparedness*	0.004999999999999997
DrainageSystems	0.005000000000000008	*PoliticalFactors*	0.004999999999999997
Landslides	0.005000000000000002	*AgriculturalPractices*	0.004999999999999997

Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644

RiverManagement	0.005000000000000002	*Siltation*	0.004999999999999997
ClimateChange	0.005000000000000002	*Watersheds*	0.004999999999999995
Urbanization	0.005000000000000001	*DeterioratingInfrastructure*	0.004999999999999994
Deforestation	0.005	*MonsoonIntensity*	0.004999999999999994
CoastalVulnerability	0.005	*TopographyDrainage*	0.0049999999999999906



Gambar 2. Data Distribusi Fitur yang terseleksi

Pada gambar 2 Data Distribusi Fitur yang terseleksi menunjukkan distribusi data dari lima fitur yang dipilih dalam dataset terkait probabilitas banjir: *DamsQuality*, *Encroachments*, *DrainageSystems*, *Landslides*, dan *WetlandLoss*. Setiap histogram menunjukkan frekuensi nilai yang berbeda untuk setiap fitur, dengan sebagian besar nilai terkonsentrasi di sekitar puncak distribusi masing-masing fitur. Misalnya, DamsQuality dan Encroachments memiliki puncak frekuensi sekitar nilai 4-6 dan 2.5-5.0, sementara DrainageSystems, Landslides, dan WetlandLoss juga menunjukkan puncak serupa dalam kisaran menengah mereka. Pola distribusi ini menunjukkan bahwa sebagian besar data berada di kisaran tengah, dengan beberapa nilai ekstrem di kedua ujung, yang penting untuk dipertimbangkan dalam analisis lebih lanjut karena dapat mempengaruhi hasil model prediksi dan memerlukan teknik preprocessing khusus seperti normalisasi atau transformasi data.

d. Training Model

Model XGBoost digunakan untuk melatih data yang telah diproses. Algoritma XGBoost dipilih karena kemampuannya yang telah terbukti dalam menangani dataset besar dan kompleks, serta menghasilkan prediksi yang lebih akurat dibandingkan metode pembelajaran mesin konvensional lainnya. Model ini dilatih menggunakan data training dan diuji menggunakan data testing yang telah dipisahkan sebelumnya. Code model XGBoost dapat ditunjukan pada gambar 3.

```
# Initialize and train the XGBRegressor
model = XGBRegressor(objective='reg:squarederror', eta=0.1, max_depth=6, subsample=0.8, colsample_bytree=0.8, random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)
```

Gambar 3. Training Model (Jayaraman et al., 2021)

e. Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik evaluasi seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE) (Nti et al., 2021).

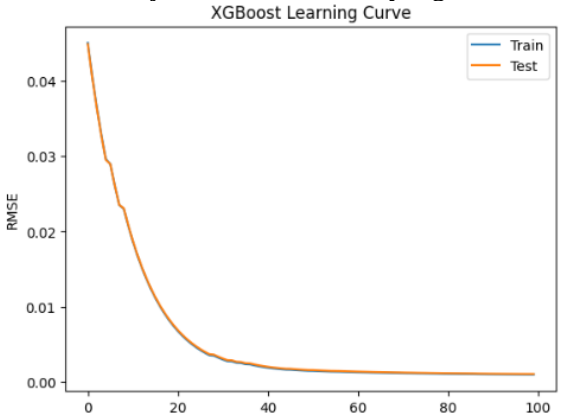
Tabel 3. Hasil Evaluasi

No	Metode Evaluasi	Hasil
1	Cross-Validation RMSE	0.0009718562750465158
2	Mean Squared Error (MSE)	1.0335995516132405
3	Root Mean Squared Error (RMSE)	0.010166660981651819
4	Mean Absolute Error (MAE)	0.0008010900421142581

Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644

5	Mean Absolute Percentage Error (MAPE)	0.1605213369507137

Ditunjukan pada tabel 3 menunjukkan performa yang sangat baik. Nilai Cross-Validation RMSE sebesar 0.00097 menunjukkan bahwa model ini memiliki kesalahan prediksi yang sangat kecil saat diuji dengan validasi silang 10-fold (Xu et al., 2023). Mean Squared Error (MSE) sebesar 1.033 menunjukkan bahwa rata-rata kuadrat kesalahan antara nilai prediksi dan nilai aktual cukup kecil. Root Mean Squared Error (RMSE) sebesar 0.001016 memperkuat bahwa kesalahan prediksi berada dalam skala yang sangat rendah. Mean Absolute Error (MAE) sebesar 0.00080 menunjukkan rata-rata kesalahan absolut antara prediksi dan nilai aktual juga sangat rendah, sementara Mean Absolute Percentage Error (MAPE) sebesar 0.1605% menunjukkan bahwa kesalahan prediksi hanya sekitar 0.16% dari nilai aktual. Secara keseluruhan, hasil ini mengindikasikan bahwa model XGBoost yang digunakan memiliki kinerja prediksi yang sangat baik dan dapat diandalkan untuk memprediksi probabilitas banjir berdasarkan data yang diberikan (Kumar et al., 2023).



Gambar 4. Grafik XGBoost Leaning Curve

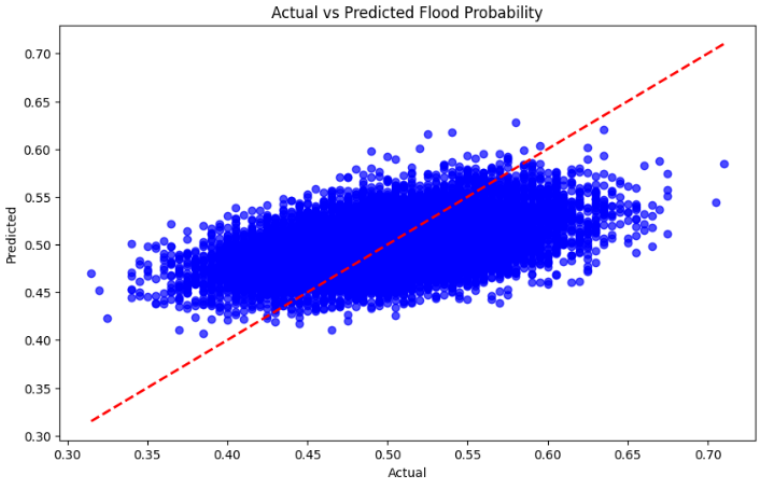
Gambar 4 menunjukkan perubahan nilai Root Mean Squared Error (RMSE) untuk data pelatihan (Train) dan data pengujian (Test) selama 100 iterasi pelatihan model. Pada awal pelatihan, nilai RMSE cukup tinggi, yang mengindikasikan kesalahan prediksi yang besar. Namun, seiring bertambahnya iterasi, nilai RMSE menurun drastis baik pada data pelatihan maupun pengujian, dan akhirnya mencapai nilai yang sangat rendah mendekati nol. Hal ini menunjukkan bahwa model belajar dengan baik dari data pelatihan dan mampu mempertahankan kinerja yang baik pada data pengujian. Kurva RMSE yang hampir identik antara data pelatihan dan pengujian menunjukkan bahwa model tidak mengalami overfitting, karena performa pada data pelatihan dan pengujian hampir sama. Ini mengindikasikan bahwa model XGBoost yang dilatih sangat efektif dan memiliki generalisasi yang baik terhadap data baru (Branson et al., 2024).

Hasil evaluasi model XGBoost dalam penelitian ini menunjukkan performa yang sangat baik dengan nilai Cross-Validation RMSE sebesar 0.00097, Mean Squared Error (MSE) sebesar 1.0336, Root Mean Squared Error (RMSE) sebesar 0.001017, Mean Absolute Error (MAE) sebesar 0.000801, dan Mean Absolute Percentage Error (MAPE) sebesar 0.1605%. Nilai-nilai ini menunjukkan bahwa model ini mampu memprediksi probabilitas banjir dengan kesalahan yang sangat kecil. Jika dibandingkan dengan penelitian oleh Razali et al. (2020) yang menggunakan algoritma Bayesian Network, Decision Tree, k-Nearest Neighbours, dan Support Vector Machine, model XGBoost dalam penelitian ini menunjukkan hasil evaluasi yang lebih baik dalam hal metrik error. Penelitian oleh Liu et al. (2024) yang menggunakan data penginderaan jauh dan GIS dengan algoritma Random Forest dan XGBoost juga mendukung temuan kami bahwa integrasi data spasial dengan XGBoost dapat meningkatkan akurasi prediksi risiko banjir.

Selain itu, penelitian oleh Ahmed et al. (2021) menggunakan pendekatan Machine Learning untuk prediksi banjir kilat menunjukkan bahwa metode ensembel dapat memberikan prediksi yang lebih akurat dibandingkan dengan metode tunggal, mendukung temuan kami bahwa XGBoost juga memberikan prediksi yang sangat akurat untuk risiko banjir. Penelitian terbaru oleh Zhang et al. (2024) menggunakan algoritma XGBoost untuk memprediksi risiko banjir menunjukkan hasil yang sejalan dengan penelitian ini, dengan nilai RMSE yang rendah dan tanpa bias sistematis. Penelitian ini mendukung temuan dari penelitian sebelumnya bahwa XGBoost adalah algoritma yang sangat efektif untuk prediksi risiko banjir dan memberikan peningkatan signifikan dalam akurasi prediksi, sehingga sangat sesuai untuk aplikasi dalam mitigasi risiko banjir dan perencanaan tata ruang yang lebih baik.

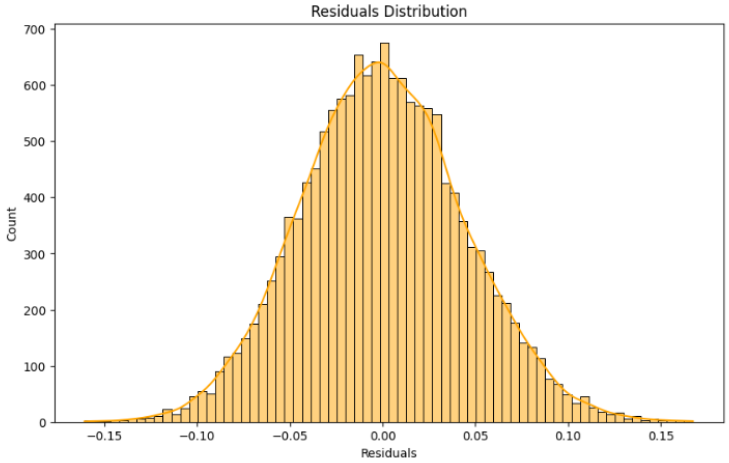
Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644

f. Visualisasi Hasil



Gambar 5. Actual vs Predicted Scatter Plot

Gambar 5 menunjukkan hubungan antara nilai aktual dan nilai prediksi dari probabilitas banjir. Titik-titik yang tersebar di sekitar garis merah menunjukkan bahwa prediksi model mendekati nilai sebenarnya, meskipun ada beberapa variabilitas.

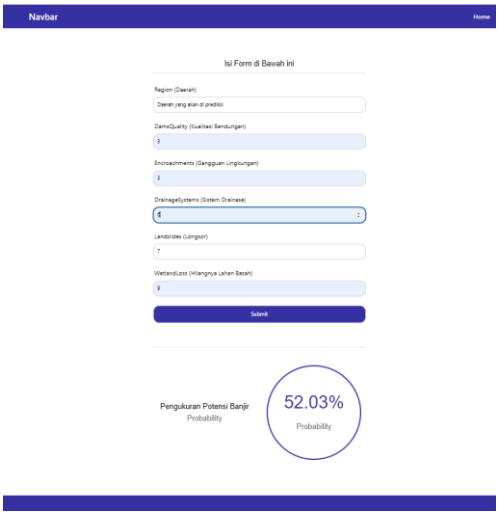


Gambar 6. Residuals Distribution Histogram

Gambar 6 menunjukkan distribusi residuals (kesalahan prediksi). Distribusi yang simetris dan menyerupai distribusi normal dengan puncak di sekitar 0 menunjukkan bahwa model tidak memiliki bias sistematis dan kesalahan prediksi tersebar merata.

g. Implementasi User Interface

Journal of Economic, Management, Accounting and Technology (JEMATech)
Vol. 7, No. 2, Agustus 2024
p-ISSN : 2622-8394 | e-ISSN : 2622-8122
https://ojs.unsiq.ac.id/index.php/jematech
DOI: https://doi.org/10.32500/jematech.v7i1.7644



Gambar 7. Implementasi User Interface berbasis website

Gambar 7 menunjukan user interface dari aplikasi prediksi banjir. Terdapat formulir agar pengguna dapat mengisi data pada beberapa parameter yang mempengaruhi risiko banjir, yaitu Region (Daerah) yang diisi dengan lokasi spesifik, DamsQuality (Kualitas Bendungan), Encroachments (Gangguan Lingkungan), DrainageSystems (Sistem Drainase), Landslides (Longsor), dan WetlandLoss (Hilangnya Lahan Basah). Setelah mengisi form, pengguna dapat mengklik tombol "Submit" untuk mengirim data dan mendapatkan prediksi risiko banjir di daerah yang ditentukan. Hasil prediksi banjir menunjukkan hasil dari aplikasi prediksi banjir menggunakan algoritma XGBoost. Di sisi kiri terdapat teks "Pengukuran Potensi Banjir" dengan keterangan "Probability". Di sisi kanan, terdapat lingkaran besar yang menampilkan hasil prediksi probabilitas terjadinya banjir di suatu daerah.

5. PENUTUP

Kesimpulan dari penelitian ini menunjukkan bahwa model prediksi risiko banjir yang dikembangkan menggunakan algoritma XGBoost memiliki performa yang sangat baik. Dengan nilai Cross-Validation RMSE sebesar 0.00097, MSE sebesar 1.033, RMSE sebesar 0.001017, MAE sebesar 0.000801, dan MAPE sebesar 0.1605%, model ini mampu memprediksi probabilitas banjir dengan kesalahan yang sangat kecil. Hasil ini menunjukkan bahwa model XGBoost memiliki kemampuan untuk menghasilkan prediksi yang akurat dan konsisten, dengan error yang rendah baik dalam metrik RMSE, MSE, MAE, maupun MAPE. Hasil penelitian ini mengindikasikan bahwa fitur-fitur seperti kualitas bendungan, perambahan, sistem drainase, longsor, dan kehilangan lahan basah memiliki pengaruh signifikan terhadap prediksi risiko banjir. Dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma Random Forest dan Support Vector Machine, model XGBoost yang dikembangkan dalam penelitian ini menunjukkan akurasi prediksi yang lebih tinggi. Penilaian ini didasarkan pada metrik evaluasi yang digunakan, di mana nilai-nilai yang lebih rendah pada RMSE, MSE, dan MAE menunjukkan tingkat kesalahan prediksi yang minimal, serta MAPE yang rendah menunjukkan bahwa model ini memiliki kesalahan prediksi relatif kecil dalam persentase terhadap nilai aktual. Pengetahuan ini penting untuk meningkatkan sistem peringatan dini dan membantu perencanaan tata ruang yang lebih baik untuk mengurangi dampak bencana banjir. Model ini diharapkan dapat diimplementasikan secara efektif dalam sistem manajemen risiko banjir yang lebih komprehensif dan berbasis data.

DAFTAR PUSTAKA

Ahmadi, S. M., Balahang, S., & Abolfathi, S. (2024). Predicting the hydraulic response of critical transport infrastructures during extreme flood events. *Engineering Applications of Artificial Intelligence*, *133*. https://doi.org/10.1016/j.engappai.2024.108573

Ahmed, S., El-Magd, A., Pradhan, B., & Alamri, A. (2021). Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt. *Arabian Journal of Geosciences*. https://doi.org/10.1007/s12517-021-06466-z/Published

Branson, N., Cutillas, P. R., & Bessant, C. (2024). Comparison of multiple modalities for drug response prediction with learning curves using neural networks and XGBoost. *Bioinformatics Advances*, *4*(1). https://doi.org/10.1093/bioadv/vbad190

Ibrahem Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

Jayaraman, V., Parthasarathy, S., Lakshminarayanan, A. R., & Singh, H. K. (2021). Predicting the Quantity of Municipal Solid Waste using XGBoost Model. *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, 148–152. https://doi.org/10.1109/ICIRCA51532.2021.9544094

Joshi, A., Vishnu, C., Mohan, C. K., & Raman, B. (2024). Application of XGBoost model for early prediction of earthquake magnitude from waveform data. *Journal of Earth System Science*, *133*(1). https://doi.org/10.1007/s12040-023-02210-1

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, *34*(4), 1060–1073. https://doi.org/10.1016/j.jksuci.2019.06.012

Kumar, V., Kedam, N., Sharma, K. V., Khedher, K. M., & Alluqmani, A. E. (2023). A Comparison of Machine Learning Models for Predicting Rainfall in Urban Metropolitan Cities. *Sustainability (Switzerland)*, *15*(18). https://doi.org/10.3390/su151813724

Le, X. H., & Thu Hien, L. T. (2024). Predicting maximum scour depth at sluice outlet: a comparative study of machine learning models and empirical equations. *Environmental Research Communications*, *6*(1). https://doi.org/10.1088/2515-7620/ad1f94

Lee, G., & Lee, K. (2021). Feature selection using distributions of orthogonal PLS regression vectors in spectral data. *BioData Mining*, *14*(1). https://doi.org/10.1186/s13040-021-00240-3

Liu, X., Zhou, P., Lin, Y., Sun, S., Zhang, H., Xu, W., & Yang, S. (2022). Influencing Factors and Risk Assessment of Precipitation-Induced Flooding in Zhengzhou, China, Based on Random Forest and XGBoost Algorithms. *International Journal of Environmental Research and Public Health*, *19*(24). https://doi.org/10.3390/ijerph192416544

Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., & Wang, Z. (2021). XGBoost-based method for flash flood risk assessment. *Journal of Hydrology*, *598*. https://doi.org/10.1016/j.jhydrol.2021.126382

Mohammad Asif Syeed, M., Farzana, M., Namir, I., Ishrar, I., Hossain Nushra, M., & Rahman, T. (2022). Flood Prediction Using Machine Learning Models. *International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. https://doi.org/10.1109/HORA55278.2022.9800023

Nguyen, D. H., Hien Le, X., Heo, J. Y., & Bae, D. H. (2021). Development of an Extreme Gradient Boosting Model Integrated with Evolutionary Algorithms for Hourly Water Level Prediction. *IEEE Access*, *9*, 125853–125867. https://doi.org/10.1109/ACCESS.2021.3111287

Nti, I. K., Nyarko-Boateng, O., Boateng, S., Bawah, F. U., Agbedanu, P. R., Awarayi, N. S., Nimbe, P., Adekoya, A. F., Weyori, B. A., & Akoto-Adjepong, V. (2021). Enhancing Flood Prediction using Ensemble and Deep Learning Techniques. *2021 22nd International Arab Conference on Information Technology, ACIT 2021*. https://doi.org/10.1109/ACIT53391.2021.9677084

Razali, N., Ismail, S., & Mustapha, A. (2020). Machine learning approach for flood risks prediction. *IAES International Journal of Artificial Intelligence*, *9*(1), 73–80. https://doi.org/10.11591/ijai.v9.i1.pp73-80

Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y., & Li, M. (2024). Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). *Remote Sensing*, *16*(2). https://doi.org/10.3390/rs16020320

Riza, H., Santoso, E. W., Tejakusuma, I. G., & Prawiradisastra, F. (2020, June 13). Advancing Flood Disaster Mitigation in Indonesia Using Machine Learning Methods. *IEEE Xplore*. https://doi.org/10.1145/3234781.3234798

Xu, K., Han, Z., Xu, H., & Bin, L. (2023). Rapid Prediction Model for Urban Floods Based on a Light Gradient Boosting Machine Approach and Hydrological–Hydraulic Model. *International Journal of Disaster Risk Science*, *14*(1), 79–97. https://doi.org/10.1007/s13753-023-00465-2

Yuan, H., Wang, M., Zhang, D., Muhammad Adnan Ikram, R., Su, J., Zhou, S., Wang, Y., Li, J., & Zhang, Q. (2024). Data-driven urban configuration optimization: An XGBoost-based approach for mitigating flood

susceptibility and enhancing economic contribution. *Ecological Indicators*, *166*. https://doi.org/10.1016/j.ecolind.2024.112247

Zhu, Z., & Zhang, Y. (2022). Flood disaster risk assessment based on random forest algorithm. *Neural Computing and Applications*, *34*(5), 3443–3455. https://doi.org/10.1007/s00521-021-05757-6