Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 4, No. 8, Agustus 2024, Hal. 369-379
DOI: https://doi.org/10.52436/1.jpti.599
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis Komparatif Linear Regression, Random Forest, dan Gradient Boosting untuk Prediksi Banjir

Ika Maulita[*1], Chyntia Raras Ajeng Widiawati[2], Arif Mu'amar Wahid[3]

[1]Jurusan Fisika, Fakultas MIPA, Universitas Jenderal Soedirman, Indonesia
[2]Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia
[3]Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Indonesia
Email: [1]ika.maulita@unsoed.ac.id, [2]chyntiaraw@amikompurwokerto.ac.id, [3]arif@amikompurwokerto.ac.id

Abstrak

Penelitian ini mengevaluasi tiga model machine learning—Linear Regression, Random Forest Regressor, dan Gradient Boosting Regressor—untuk memprediksi probabilitas banjir di India, dengan tujuan meningkatkan akurasi prediksi dan mendukung strategi mitigasi risiko banjir. Kinerja model dievaluasi menggunakan metrik Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R^2. Hasil penelitian menunjukkan bahwa Linear Regression dan Gradient Boosting Regressor memiliki kinerja yang hampir setara, dengan MAE dan RMSE yang kompetitif. Namun, Linear Regression sedikit unggul dalam menjelaskan variabilitas probabilitas banjir berdasarkan nilai R^2. Sebaliknya, Random Forest Regressor menunjukkan kinerja yang lebih rendah, yang kemungkinan disebabkan oleh overfitting atau kurang optimalnya penyetelan parameter. Penelitian ini memberikan kontribusi penting terhadap peningkatan akurasi sistem peringatan dini dan pengelolaan risiko banjir berbasis data. Dengan menganalisis faktor-faktor utama yang memengaruhi probabilitas banjir, penelitian ini menawarkan wawasan yang dapat mendukung perencanaan intervensi yang lebih efektif, seperti pengelolaan sungai yang lebih baik dan perencanaan tata ruang perkotaan yang adaptif. Saran untuk penelitian mendatang meliputi eksplorasi algoritma tambahan, termasuk pendekatan pembelajaran mendalam, penerapan rekayasa fitur lanjutan, serta optimalisasi model menggunakan alat Automated Machine Learning (AutoML). Temuan ini berkontribusi pada pengembangan metode prediksi banjir yang lebih akurat dan efisien, serta memperkuat upaya mitigasi risiko banjir di masa depan.

Kata kunci: *gradient boosting, model machine learning, prediksi probabilitas banjir, regresi linier, random forest*

Comparative Analysis of Linear Regression, Random Forest, and Gradient Boosting Algorithms in Flood Probability Prediction

Abstract

This study evaluates three machine learning models—Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor—for predicting flood probability in India. Model performance was assessed using Mean Absolute Error, Root Mean Squared Error, and the coefficient of determination R^2. The results indicate that Linear Regression and Gradient Boosting Regressor performed comparably, with competitive MAE and RMSE values. However, Linear Regression slightly outperformed Gradient Boosting in explaining flood probability variability based on the R^2 score. Conversely, Random Forest Regressor exhibited weaker performance, suggesting overfitting or the need for further parameter tuning. Overall, the moderate R^2 values across all models highlight significant opportunities to improve predictive accuracy and understanding of the key factors influencing floods. This study identifies several limitations, including data quality variability and challenges in capturing the complexity of interactions between environmental and human factors affecting floods. Nevertheless, the findings provide valuable insights for enhancing early warning systems and flood risk mitigation strategies. By integrating in-depth feature analysis, this research establishes a strong foundation for data-driven flood risk management. Future research recommendations include exploring additional machine learning algorithms, such as deep learning approaches, applying advanced feature engineering techniques, and utilizing automated tools like AutoML to optimize model performance. These findings contribute to the development of more accurate and efficient flood prediction methods and offer opportunities to strengthen risk mitigation and disaster management strategies in the future.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Keywords: *flood probability prediction, gradient boosting, linear regression, machine learning models, random forest*

1. PENDAHULUAN

Banjir merupakan salah satu bencana alam yang paling merusak di dunia, dengan dampak yang meluas dan memengaruhi jutaan orang setiap tahunnya [1]. Dampaknya mencakup hilangnya nyawa manusia, kerusakan besar pada infrastruktur penting, serta gangguan ekonomi yang signifikan. Kekuatan destruktif banjir tampak dalam berbagai bentuk, termasuk air yang mengalir deras dan cepat menggenangi area luas, merendam rumah, bisnis, serta seluruh komunitas. Arus yang kuat sering kali menyebabkan kerusakan struktural, membuat bangunan tidak dapat dihuni dan jalan menjadi tidak dapat dilalui. Selain itu, layanan esensial seperti listrik, air, dan infrastruktur komunikasi sering terganggu, memperburuk tantangan yang dihadapi masyarakat terdampak. Secara ekonomi, banjir memiliki dampak yang besar. Kerusakan pada tanaman, ternak, dan infrastruktur pertanian sering kali menyebabkan gangguan dalam produksi pangan dan kelangkaan pasokan. Biaya perbaikan dan rekonstruksi infrastruktur yang rusak dapat mencapai angka yang astronomis, membebani anggaran pemerintah dan mengalihkan sumber daya dari layanan penting lainnya [2]. Komunitas global telah mengakui urgensi untuk menangani masalah banjir. Berbagai strategi telah diterapkan oleh pemerintah, organisasi internasional, dan lembaga non-pemerintah, seperti pembangunan struktur pengendalian banjir, penerapan sistem peringatan dini, penguatan ketahanan masyarakat, dan promosi praktik penggunaan lahan yang berkelanjutan [3].

Mengatasi tantangan banjir memerlukan pendekatan yang terintegrasi, melibatkan kolaborasi antara pemerintah, komunitas, dan berbagai pemangku kepentingan. Investasi dalam langkah-langkah pengurangan risiko banjir dapat meminimalkan dampak manusia dan ekonomi dari bencana ini, sekaligus membangun masyarakat yang lebih tangguh. Seiring meningkatnya volatilitas fenomena iklim, kejadian banjir diperkirakan akan semakin sering dan intens. Hal ini menekankan kebutuhan mendesak untuk menerapkan metode prediksi yang lebih canggih [4]. Model hidrologi tradisional untuk peramalan banjir sering kali membutuhkan data yang ekstensif dan sumber daya komputasi yang besar. Namun, kemunculan machine learning menawarkan alternatif yang layak, dengan kemampuan menganalisis dataset yang luas dan memberikan prediksi dengan presisi dan efisiensi tinggi. Kemajuan terkini dalam machine learning telah didokumentasikan secara luas, dengan berbagai metodologi yang digunakan untuk memprediksi probabilitas banjir di berbagai wilayah. Random Forest classification telah diterapkan pada skala spasial yang luas, seperti yang terlihat dalam penelitian di Texas [5]. Selain itu, logistic regression telah digunakan untuk menghasilkan prediksi probabilitas banjir yang mendukung penilaian risiko praktis [6]. Machine learning juga memainkan peran penting dalam mengembangkan teknik peramalan banjir, meningkatkan akurasi prediksi, dan efektivitas sistem peringatan dini. Sebagai contoh, penelitian [7] berhasil mengintegrasikan model machine learning dengan peta genangan banjir berbasis satelit dalam kerangka operasional, menunjukkan kemampuan machine learning untuk secara akurat menggambarkan luasan banjir.

Lebih jauh lagi, penelitian [8] mengusulkan pendekatan machine learning inovatif untuk memprediksi level air sungai dengan menggabungkan berbagai teknik machine learning guna memperkuat sistem peringatan dini. Penelitian lainnya [9] menunjukkan kekuatan model ensemble yang menggabungkan prakiraan meteorologi, pemodelan hidrologi, dan machine learning untuk memberikan perspektif multidimensional dalam peramalan banjir. Selain itu, integrasi machine learning dengan sistem informasi geografis dalam peramalan banjir real-time menunjukkan potensi transformasional machine learning dalam mengatasi keterbatasan data dan meningkatkan akurasi prediksi [10]. Dalam konteks eksplorasi model machine learning, beberapa penelitian telah mendokumentasikan keunikan pendekatan dalam memodelkan banjir. Sebagai contoh, penelitian [11] membandingkan empat model machine learning – multilayer perceptron, logistic regression, support vector machine, dan random forest – untuk pemodelan kerentanan banjir bandang, memberikan informasi tentang karakteristik kinerja dan kesesuaian model untuk skenario tertentu. Di sisi lain, penelitian [8] menunjukkan potensi pendekatan ensemble dalam meningkatkan akurasi prediksi dengan memanfaatkan berbagai teknik seperti radial basis function neural networks, adaptive neuro-fuzzy inference systems, support vector machines, dan long short-term memory networks.

Pendekatan machine learning telah menunjukkan potensi besar dalam analisis data geofisika dan bencana alam, seperti yang dibahas dalam penelitian [12] yang mengevaluasi performa model Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron untuk memprediksi magnitudo gempa berdasarkan data kedalaman dan geolokasi. Studi ini menunjukkan kekuatan pendekatan berbasis data dalam menangkap pola kompleks untuk meningkatkan akurasi prediksi fenomena alam. Selain itu, penelitian [13] menggunakan analisis second vertical derivative data gravitasi untuk menginterpretasikan struktur bawah permukaan daerah Lembang, yang memberikan wawasan mendalam tentang hubungan antara struktur geologi

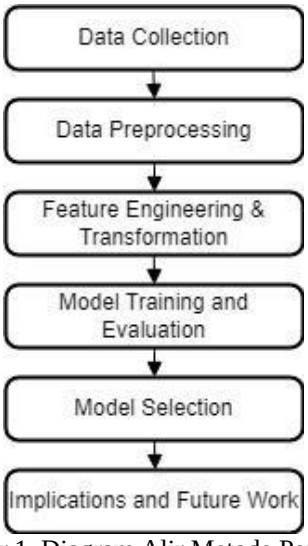
p-ISSN: 2775-4227
e-ISSN: 2775-4219

dan potensi bencana. Kedua studi ini menggarisbawahi pentingnya eksplorasi dan penerapan model machine learning serta analisis berbasis data untuk memahami dan memitigasi risiko bencana, termasuk dalam konteks prediksi banjir yang menjadi fokus penelitian ini.

Berbagai penelitian menunjukkan efektivitas algoritma machine learning seperti artificial neural networks (ANN), support vector machines (SVM), dan random forests dalam menangani berbagai aspek prediksi banjir, termasuk pemetaan kerentanan dan peramalan debit air. Sebagai contoh, integrasi ML dengan metode hidrologi tradisional, seperti yang terlihat pada model Informer, telah menunjukkan hasil yang menjanjikan dalam meningkatkan akurasi prediksi debit air banjir [14]. Selain itu, teknik ML terbukti efektif dalam mengelola ketidakpastian terkait pemetaan kerentanan banjir, menawarkan alternatif yang hemat biaya dibandingkan survei lapangan yang memakan waktu dan sumber daya [15]. Kemajuan terbaru semakin menekankan potensi ML, khususnya dalam kerangka kerja prediksi banjir real-time yang memanfaatkan data historis banjir dan berbagai variabel lingkungan [16]. Integrasi data penginderaan jauh dengan model ML juga terbukti sangat bermanfaat di wilayah dengan keterbatasan data, meningkatkan kemampuan pemetaan dan pemantauan banjir [17]. Kemampuan adaptasi dan kekuatan prediktif ML menjadikannya komponen yang tak tergantikan dalam strategi manajemen risiko banjir modern, memungkinkan tindakan proaktif untuk mengurangi dampak bencana banjir [18].

Meskipun telah terjadi banyak kemajuan dalam penerapan model machine learning untuk prediksi banjir, bidang ini tetap menjadi area penelitian yang terus berkembang. Para peneliti terus mengeksplorasi efektivitas berbagai algoritma machine learning untuk meningkatkan akurasi prediksi dan efisiensi sistem peringatan dini. Dalam konteks ini, penelitian ini berkontribusi pada pengisian kesenjangan pengetahuan penting dengan mengevaluasi dan membandingkan tiga model machine learning, yaitu Linear Regression, Random Forest, dan Gradient Boosting. Ketiga model ini dipilih untuk mengidentifikasi pendekatan yang paling efektif dalam memprediksi probabilitas banjir, mempertimbangkan kinerja mereka dalam menangkap pola yang kompleks dalam data terkait banjir. Tujuan utama penelitian ini adalah untuk memberikan analisis komparatif yang komprehensif terhadap model machine learning yang dipilih, dengan mengevaluasi akurasi prediktifnya berdasarkan metrik evaluasi yang relevan seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan skor R^2. Selain itu, penelitian ini bertujuan untuk mengeksplorasi pentingnya fitur-fitur tertentu sebagai prediktor utama dalam kejadian banjir, memberikan wawasan mendalam tentang faktor-faktor lingkungan, manusia, dan geografis yang berkontribusi terhadap risiko banjir. Kontribusi penelitian ini tidak hanya terletak pada perbandingan model machine learning untuk prediksi banjir, tetapi juga pada wawasan yang diberikan terkait pengaruh berbagai fitur terhadap probabilitas banjir. Analisis fitur ini dapat menjadi dasar bagi pengembangan strategi mitigasi risiko banjir yang lebih terarah, seperti peningkatan manajemen sungai, perencanaan tata ruang perkotaan yang lebih berkelanjutan, dan kebijakan pengelolaan lingkungan yang lebih adaptif.

2. METODE PENELITIAN



Gambar 1. Diagram Alir Metode Penelitian

Metodologi yang digunakan dalam penelitian ini dirancang untuk secara mendalam mengevaluasi efektivitas berbagai model machine learning dalam memprediksi probabilitas banjir. Setiap langkah dalam

p-ISSN: 2775-4227
e-ISSN: 2775-4219

proses metodologi ini dirancang untuk memastikan alur yang jelas dari pengumpulan data hingga evaluasi model. Untuk memberikan kejelasan lebih lanjut dan memandu pembaca dalam memahami proses metodologi ini secara visual, sebuah diagram alur disertakan. Diagram ini, yang ditampilkan pada Gambar 1, memberikan gambaran rinci dari metodologi penelitian yang diterapkan dalam penelitian ini.

2.1. Deskripsi Dataset

Dataset yang dianalisis dalam penelitian ini berfokus pada peristiwa banjir di India, yang bersumber dari Kaggle. Dengan lebih dari 50.000 entri yang mencakup berbagai faktor terkait banjir di seluruh India, dataset ini menawarkan cakupan temporal dan spasial luas, sehingga menjadi landasan yang kokoh untuk analisis mendalam tentang probabilitas banjir dan faktor-faktor yang memengaruhinya di kawasan India. Dataset ini mencakup 21 kolom, yang masing-masing mewakili faktor yang dapat memengaruhi probabilitas banjir. Kolom-kolom ini mencakup representasi numerik dari Monsoon Intensity, yang mengukur tingkat keparahan musim hujan—sebuah aspek penting dalam dinamika perubahan banjir di India. Selain itu, metrik seperti Topography Drainage dan River Management menilai kemampuan drainase alami dan terkelola yang penting untuk penilaian risiko banjir. Dataset ini juga mengevaluasi secara kuantitatif modifikasi yang dilakukan manusia terhadap lanskap, seperti Deforestation dan Urbanization, yang dapat memperburuk risiko banjir. Faktor lingkungan yang lebih luas dan aspek infrastruktur juga tercakup melalui variabel seperti Climate Change dan Dams Quality.

Indikator lainnya, seperti Agricultural Practices, Encroachments, dan Wetland Loss, mencerminkan perubahan penggunaan lahan yang dapat memengaruhi kerentanan terhadap banjir. Skor yang berkaitan dengan Ineffective Disaster Preparedness dan Inadequate Planning memberikan informasi tentang tingkat kesiapsiagaan dan kecukupan perencanaan terhadap risiko banjir, yang menyoroti area yang rentan. Selain itu, faktor-faktor seperti Landslides, Watersheds, dan Coastal Vulnerability menekankan kerentanan geografis dan lingkungan yang dapat memengaruhi terjadinya banjir. Variabel-variabel ini dipilih dengan cermat untuk memberikan gambaran yang komprehensif mengenai berbagai pengaruh terhadap probabilitas banjir, mencakup kondisi iklim, elemen topografi, intervensi manusia, dan isu-isu tata kelola. Pemilihan ini bertujuan untuk memfasilitasi analisis yang lebih mendalam terhadap faktor-faktor beragam yang berkontribusi terhadap risiko banjir, sehingga mendukung pengembangan strategi manajemen dan mitigasi banjir yang efektif. Dalam dataset ini, Flood Probability didefinisikan sebagai angka dalam bentuk desimal yang mewakili kemungkinan terjadinya banjir di suatu area tertentu. Skor probabilitas ini, yang dihitung berdasarkan fitur-fitur di atas, berfungsi sebagai variabel dependen dalam analisis kami. Dengan mengukur probabilitas banjir pada skala 0 hingga 1, dataset ini memfasilitasi pemeriksaan yang lebih mendalam terhadap risiko banjir, memungkinkan penerapan model machine learning untuk memprediksi kemungkinan banjir dengan presisi yang lebih tinggi.

2.2. Prapemrosesan Data

2.2.1. Penanganan Nilai Hilang dan Deteksi Outlier

Mengingat cakupan temporal dan spasial dataset yang luas, keberadaan data yang hilang merupakan tantangan yang sudah diperkirakan sebelumnya. Strategi yang diterapkan untuk menangani nilai hilang dilakukan melalui dua pendekatan. Pertama, teknik imputasi digunakan untuk fitur yang memiliki data hilang kurang dari 5% dari total observasi. Untuk variabel kontinu, imputasi nilai rata-rata diterapkan, sedangkan untuk variabel kategorikal, mode digunakan sebagai pengganti data yang hilang. Kedua, catatan dengan informasi yang hilang dalam jumlah signifikan, khususnya jika lebih dari 30% data pada suatu catatan tidak tersedia, dikeluarkan dari analisis untuk menjaga integritas dataset. Outlier diidentifikasi menggunakan metode Interquartile Range (IQR), yang dikenal efektif dalam mendeteksi data yang menyimpang secara signifikan dari keseluruhan dataset. Sebagai contoh, fitur WetlandLoss, yang memiliki nilai maksimum jauh di atas persentil ke-75, diperiksa secara khusus untuk outlier. Nilai outlier dibatasi hingga 1,5*IQR di atas kuartil ketiga dan di bawah kuartil pertama untuk meminimalkan pengaruh negatifnya terhadap performa model.

2.2.2. Normalisasi dan Transformasi Data

Berdasarkan statistik deskriptif, distribusi fitur seperti MonsoonIntensity dan Urbanization relatif merata, sehingga proses skala secara menyeluruh dianggap tidak diperlukan untuk sebagian besar variabel. Namun, untuk memastikan keseragaman dan meningkatkan konvergensi model, Min-Max Scaling diterapkan secara selektif pada fitur yang menunjukkan variabilitas signifikan dalam rentang dan distribusinya. Metode ini mengubah nilai fitur menjadi skala antara 0 dan 1, sehingga mempermudah pelatihan dan perbandingan model. Transformasi data juga diterapkan pada fitur tertentu yang menunjukkan distribusi miring, seperti yang diamati pada RiverManagement dan Deforestation. Transformasi logaritmik digunakan untuk menormalkan distribusi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

fitur-fitur ini, sehingga meningkatkan kemampuan model untuk menangkap pola dan mempelajari data dengan lebih efektif. Langkah ini penting untuk mengatasi kemiringan distribusi ke arah nilai yang lebih tinggi pada RiverManagement dan penyebaran luas yang diamati pada Deforestation, memastikan model dapat menginterpretasikan dan mempelajari fitur-fitur ini dengan lebih akurat. Dengan penerapan langkah-langkah prapemrosesan ini, dataset menjadi lebih siap untuk analisis dan pemodelan, dengan potensi bias dari nilai hilang dan outlier yang diminimalkan serta distribusi data yang lebih seragam untuk meningkatkan kinerja model.

2.2.3. Pemilihan Fitur

Berdasarkan analisis awal terhadap outlier menggunakan box plot, yang mengidentifikasi outlier signifikan pada fitur seperti MonsoonIntensity dan Deforestation, pendekatan yang lebih terperinci diadopsi untuk menangani outlier tersebut. Mengingat potensi pengaruhnya terhadap akurasi prediksi banjir, outlier diperiksa dengan cermat untuk menentukan apakah mereka merepresentasikan skenario ekstrem tetapi masuk akal atau merupakan kesalahan data. Data outlier yang mencerminkan kejadian langka namun penting, seperti musim hujan yang sangat intens, dipertahankan untuk memastikan model dapat belajar dari spektrum penuh kondisi yang menyebabkan banjir. Namun, untuk outlier yang tampak sebagai kesalahan pencatatan atau inkonsistensi data, dilakukan pembatasan nilai berdasarkan aturan 1,5*IQR atau mempertimbangkan penghapusannya setelah tinjauan menyeluruh. Berdasarkan informasi yang diperoleh dari Analisis Data Eksploratif, beberapa fitur menunjukkan karakteristik yang memerlukan transformasi atau rekayasa tambahan untuk mengoptimalkan kegunaannya dalam pemodelan prediktif. Secara khusus, fitur seperti RiverManagement dan Deforestation, yang memiliki distribusi miring, menjalani transformasi logaritmik. Transformasi ini penting untuk menormalkan data, sehingga meningkatkan kemampuan model untuk menginterpretasikan fitur-fitur ini dengan lebih akurat dan efektif.

Selain itu, dilakukan rekayasa fitur untuk menyempurnakan analisis lebih lanjut, terutama pada fitur seperti MonsoonIntensity dan Urbanization. Dengan mengamati pola korelasi dan distribusi fitur-fitur ini, interaksi antar variabel dikembangkan, seperti interaksi antara tingkat urbanisasi dan intensitas musim hujan. Pendekatan metodologis ini dirancang untuk menangkap efek gabungan dari variabel-variabel tersebut terhadap probabilitas banjir, memberikan pemahaman yang lebih dalam tentang hubungan kompleks yang mungkin hanya terungkap sebagian melalui analisis satu variabel. Augmentasi dataset strategis ini memperkaya input model dan bertujuan meningkatkan akurasi prediktif terkait kejadian banjir. Dalam analisis fitur terfokus yang dilakukan dalam penelitian ini, fitur-fitur dipilih dengan cermat untuk analisis mendalam berdasarkan variabilitasnya, korelasi dengan variabel target, keberadaan outlier, serta representasi yang komprehensif terhadap faktor lingkungan, manusia, geografis, dan infrastruktur. Pendekatan yang seimbang ini memastikan model tidak hanya didasarkan pada data, tetapi juga mencerminkan sifat kompleks dan multifaset dari kejadian banjir.

Untuk faktor lingkungan, perhatian difokuskan pada MonsoonIntensity dan Deforestation karena dampaknya yang langsung terhadap probabilitas banjir. Elemen-elemen ini menyoroti bagaimana kondisi alami dapat secara signifikan memperburuk risiko banjir. Dalam hal pengaruh manusia dan perencanaan, fitur seperti Urbanization dan RiverManagement dianalisis untuk perannya dalam mengubah dinamika banjir. Analisis ini menegaskan efek signifikan dari aktivitas manusia dan upaya manajemen yang efektif terhadap modifikasi kerentanan banjir. Selain itu, fitur geografis seperti TopographyDrainage diteliti karena perannya yang penting dalam akumulasi dan aliran air, yang merupakan faktor kritis dalam menilai area berisiko banjir. Analisis komprehensif ini memastikan pemahaman yang mendalam tentang berbagai dinamika, sehingga memungkinkan prediksi yang lebih akurat dan strategi mitigasi yang efektif untuk kejadian banjir. Pendekatan yang disempurnakan untuk pemilihan fitur, prapemrosesan data, dan analisis ini secara langsung menginformasikan proses pengembangan model. Dengan menangani outlier secara bijaksana, mentransformasi dan merekayasa fitur untuk lebih mencerminkan hubungan mendasar, serta berfokus pada subset fitur yang sangat relevan, kami meningkatkan ketahanan dan akurasi model machine learning. Metodologi ini tidak hanya memfasilitasi pemahaman yang lebih mendalam tentang faktor-faktor yang memengaruhi probabilitas banjir, tetapi juga selaras dengan tujuan kami untuk mengidentifikasi dan membandingkan efektivitas model Linear Regression, Random Forest, dan Gradient Boosting dalam memprediksi kejadian banjir.

2.3. Pelatihan dan Pengujian Model

2.3.1. Linear Regression

Linear Regression adalah salah satu algoritma machine learning yang paling dasar dan banyak digunakan, terutama karena kesederhanaannya dalam memodelkan hubungan linear antara variabel dependen (target) dan variabel independen (prediktor). Algoritma ini beroperasi berdasarkan prinsip bahwa variabel-variabel tersebut menunjukkan korelasi linear, sehingga memungkinkan prediksi variabel target berdasarkan kombinasi linear dari

p-ISSN: 2775-4227
e-ISSN: 2775-4219

prediktor. Metode ini sangat berguna untuk memahami dampak langsung dari setiap fitur terhadap hasil yang diamati. Linear Regression dipilih sebagai model dasar (baseline) dalam penelitian ini karena sifatnya yang mudah diinterpretasikan dan aplikasinya yang sederhana. Kesederhanaan ini memberikan tolok ukur yang jelas untuk membandingkan kinerja model yang lebih kompleks. Selain itu, transparansinya dalam mengungkap hubungan antara variabel menjadikannya alat yang sangat berharga untuk analisis awal, sekaligus menetapkan dasar bagi pemodelan yang lebih rumit.

2.3.2. Random Forest Regressor

Random Forest Regressor menggunakan pendekatan pembelajaran ensemble dengan membangun sejumlah pohon keputusan (decision trees) selama fase pelatihan dan membuat prediksi dengan menggabungkan hasil dari pohon-pohon tersebut. Metode ini meningkatkan akurasi dan ketahanan prediksi dengan merata-ratakan hasil dari banyak pohon, sehingga mengurangi risiko overfitting yang sering terjadi pada pohon keputusan tunggal. Setiap pohon dalam hutan dibangun dari subset data dan fitur yang dipilih secara acak, memastikan keberagaman model dan memperkuat keseluruhan algoritma. Pemilihan Random Forest dalam penelitian ini didasarkan pada ketahanannya terhadap overfitting dan kemampuannya dalam memodelkan hubungan kompleks, non-linear, serta interaksi antar fitur. Kemampuan algoritma ini untuk menangani data numerik dan kategorikal, serta memberikan informasi tentang pentingnya fitur, sangat sejalan dengan tujuan penelitian. Mengingat kompleksitas dataset dan interaksi antara faktor lingkungan serta manusia yang memengaruhi probabilitas banjir, Random Forest sangat cocok untuk menangkap hubungan non-linear ini tanpa memerlukan prapemrosesan data yang ekstensif.

2.3.3. Gradient Boosting Regressor

Gradient Boosting Regressor didasarkan pada teknik boosting, sebuah metode ensemble yang secara berurutan mengonversi weak learners menjadi strong learners. Setiap pohon dalam urutan tersebut bertujuan untuk memperbaiki kesalahan dari pendahulunya, dengan model memberikan bobot lebih pada data yang sulit diprediksi. Proses koreksi iteratif ini menghasilkan model prediktif yang sangat akurat, yang mampu menangani berbagai kompleksitas dan kerumitan data. Gradient Boosting Regressor dipilih karena kinerjanya yang luar biasa dalam menangani dataset yang kompleks dan efektivitasnya di berbagai tugas prediksi. Fleksibilitasnya, yang terlihat dari parameter yang dapat disesuaikan secara rinci untuk mengoptimalkan kinerja, serta kemampuannya dalam menghadapi data yang hilang, menjadikannya pilihan menarik untuk penelitian ini. Kemampuan Gradient Boosting untuk memodelkan pola dan interaksi kompleks dalam dataset, terutama ketika menggunakan varian yang lebih maju seperti XGBoost, LightGBM, atau CatBoost, memperkuat kesesuaiannya dalam memprediksi probabilitas banjir di tengah kompleksitas dataset yang digunakan.

2.4. Metrik Evaluasi

2.4.1. Mean Absolute Error (MAE)

MAE adalah metrik yang digunakan untuk mengukur rata-rata besarnya kesalahan dalam prediksi tanpa memperhatikan arah kesalahan tersebut. MAE dihitung sebagai rata-rata dari nilai absolut perbedaan antara nilai yang diprediksi dan nilai aktual dari variabel target. Metrik ini menawarkan cara yang sederhana dan mudah diinterpretasikan untuk mengevaluasi akurasi model, memberikan informasi jelas tentang sejauh mana prediksi menyimpang, rata-rata, dari hasil aktual. Secara matematis, MAE dirumuskan sebagai:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}|\mathrm{y_i} - \widehat{y_i}| \qquad (1)$$

Di mana (n) adalah jumlah observasi, (y_i) adalah nilai aktual dari variabel target pada observasi ke-i, dan $(\widehat{y_i})$ adalah nilai prediksi pada observasi ke-i. MAE dipilih sebagai metrik evaluasi utama karena sifatnya yang langsung dan mudah dipahami. Karakter linear MAE memastikan bahwa semua kesalahan memiliki bobot yang sama, memberikan gambaran yang jernih dan tidak bias tentang dampak rata-rata kesalahan prediksi pada dataset. Selain itu, karena MAE mengukur kesalahan dalam skala yang sama dengan variabel target, metrik ini memungkinkan perbandingan akurasi model yang bermakna. Hal ini sangat relevan untuk mengevaluasi kinerja model dalam memprediksi probabilitas banjir, di mana pemahaman yang intuitif terhadap perbedaan absolut antara hasil prediksi dan kejadian aktual sangat penting, terutama dalam konteks pengelolaan risiko banjir dan kesiapsiagaan bencana.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.4.2. Root Mean Squared Error (RMSE)

RMSE adalah metrik yang mengukur akar kuadrat dari rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual. Berbeda dengan MAE, RMSE memberikan penalti yang lebih besar pada kesalahan yang lebih besar, karena proses mengkuadratkan kesalahan sebelum rata-rata memperbesar dampak dari perbedaan besar. Karakteristik ini menjadikan RMSE metrik yang lebih sensitif terhadap kinerja model, terutama dalam konteks di mana kesalahan besar tidak diinginkan. Secara matematis, RMSE dirumuskan sebagai:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \widehat{\mathrm{y}_\mathrm{i}})^2} \qquad (2)$$

Di mana (n) adalah jumlah observasi, (y_i) adalah nilai aktual dari variabel target pada observasi ke-i, dan $(\widehat{y_i})$ adalah nilai prediksi pada observasi ke-i. Pemilihan RMSE sebagai metrik evaluasi didorong oleh kemampuannya untuk menyoroti model yang secara efektif meminimalkan kesalahan prediksi besar. Dalam prediksi banjir, di mana kesalahan signifikan dalam memperkirakan probabilitas banjir dapat memiliki konsekuensi yang serius, penekanan RMSE pada kesalahan besar sangat penting. RMSE memberikan pandangan yang lebih mendalam tentang kinerja model, memastikan model tidak hanya mencapai kesalahan rata-rata yang rendah tetapi juga menghindari kesalahan besar yang dapat menyebabkan kurangnya kesiapan dalam menghadapi peristiwa banjir.

2.4.3. Skor R^2

Skor R^2, atau koefisien determinasi, mengukur proporsi variansi dalam variabel target yang dapat dijelaskan oleh variabel independen dalam model. Nilai R^2 berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan bahwa model secara sempurna menjelaskan variabilitas variabel target terhadap rata-ratanya. R^2 memberikan informasi tentang kemampuan eksplanatori model, mencerminkan seberapa baik variabel independen menangkap dinamika variabel target. Secara matematis, R^2 dirumuskan sebagai:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (3)$$

Di mana (n) adalah jumlah observasi, (y_i) adalah nilai aktual dari variabel target pada observasi ke-i, $(\widehat{y_i})$ adalah nilai prediksi pada observasi ke-i dan $(\overline{y})$ adalah rata-rata nilai aktual dari variabel target. R^2 dipilih sebagai metrik evaluasi utama untuk menilai sejauh mana model cocok dengan data, memberikan ukuran perbandingan akurasi dan efisiensi prediktif antar model. Nilai R^2 memberikan pandangan holistik tentang kinerja model melampaui sekadar tingkat kesalahan, membantu peneliti menentukan model mana yang paling efektif menangkap pola dan hubungan dalam dataset. Metrik ini membimbing dalam memilih model yang paling tepat untuk memprediksi probabilitas banjir.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Kinerja Model

Penelitian ini mengevaluasi tiga model machine learning yang berbeda untuk memprediksi probabilitas banjir: Linear Regression, Random Forest Regressor, dan Gradient Boosting Regressor. Kinerja masing-masing model dinilai menggunakan metrik MAE, RMSE, dan skor R^2. Berikut adalah ringkasan hasil analisis

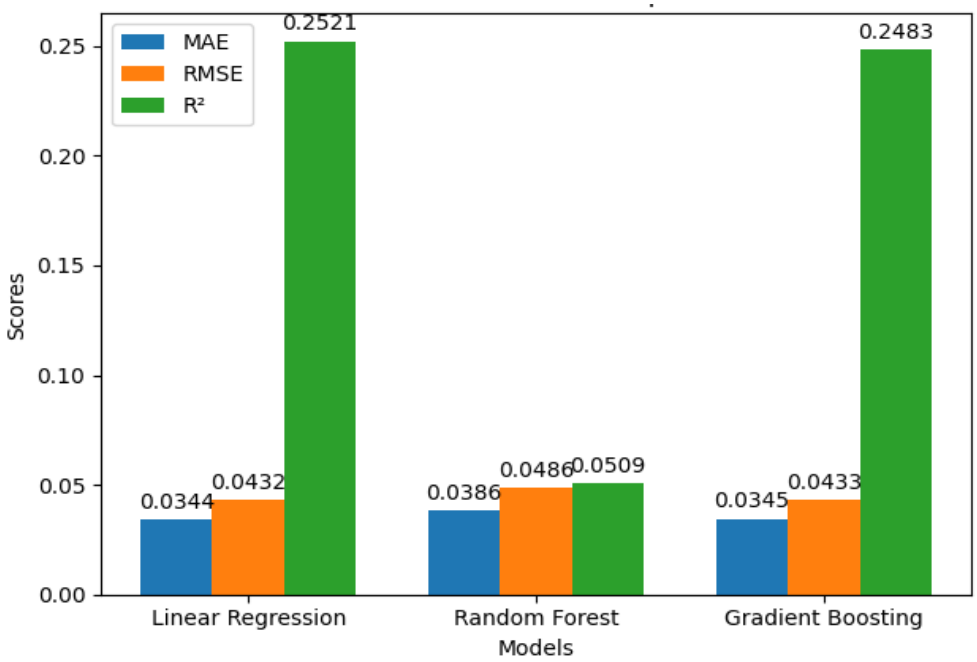
Tabel 1. Hasil Kinerja Model

Model	MAE	RMSE	R^2
Linear Regression	0.0344	0.0432	0.2521
Random Forest Regressor	0.0386	0.0486	0.0509
Gradient Boosting Regressor	0.0345	0.0433	0.2483

Dalam analisis perbandingan model prediksi banjir, Linear Regression dan Gradient Boosting Regressor menunjukkan kinerja yang hampir sebanding, terutama terlihat dari nilai MAE dan RMSE mereka. Metrik-metrik ini menunjukkan bahwa kedua model memiliki kemampuan yang mirip dalam memprediksi probabilitas banjir dengan akurasi yang tinggi. Namun, Linear Regression sedikit mengungguli Gradient Boosting dalam hal skor R^2, yang mengukur proporsi variansi dalam variabel dependen yang dapat diprediksi oleh variabel independen. Keunggulan kecil ini menunjukkan bahwa Linear Regression lebih efektif dalam menangkap

p-ISSN: 2775-4227
e-ISSN: 2775-4219

variabilitas mendasar dalam probabilitas banjir. Perbandingan kinerja model ini ditampilkan dalam Gambar 2 di bawah ini.



Gambar 2. Perbandingan Kinerja Model

Sebaliknya, Random Forest Regressor menunjukkan kinerja yang lebih lemah berdasarkan metrik yang sama. Hal ini dapat menunjukkan beberapa potensi masalah. Pertama, kinerja yang tertinggal ini dapat mengindikasikan bahwa model terlalu terfokus pada data pelatihan, menangkap noise alih-alih pola dasar, sehingga gagal melakukan generalisasi secara efektif terhadap data baru yang belum pernah dilihat sebelumnya. Overfitting adalah tantangan umum pada Random Forest, terutama ketika model dilatih dengan terlalu banyak pohon (trees) atau pohon yang terlalu dalam yang secara sempurna mencocokkan data pelatihan tetapi berkinerja buruk pada data validasi atau pengujian. Kedua, kinerja yang lebih rendah ini juga mengindikasikan perlunya penyetelan parameter (parameter tuning). Random Forest memiliki beberapa parameter seperti jumlah pohon (number of trees), kedalaman maksimum pohon (max depth), jumlah sampel minimum untuk pemisahan (min samples split), dan jumlah sampel minimum untuk daun (min samples leaf). Penyesuaian yang cermat diperlukan untuk menghindari overfitting dan mengoptimalkan kinerja.

3.2. Pembahasan

Hasil penelitian ini menunjukkan perbedaan signifikan dalam kinerja tiga model machine learning—Linear Regression, Random Forest Regressor, dan Gradient Boosting Regressor—dalam memprediksi probabilitas banjir. Berdasarkan nilai Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), Linear Regression dan Gradient Boosting Regressor menunjukkan performa yang hampir sebanding. Keduanya mampu memberikan prediksi dengan tingkat kesalahan rata-rata yang rendah, mencerminkan kemampuan yang baik dalam memodelkan pola probabilitas banjir. Namun, Linear Regression sedikit unggul dalam hal koefisien determinasi R^2, yang mengukur proporsi variansi variabel dependen yang dapat dijelaskan oleh variabel independen. Keunggulan kecil ini menunjukkan bahwa Linear Regression lebih efektif dalam menangkap variabilitas mendasar dalam data probabilitas banjir, meskipun pendekatannya lebih sederhana dibandingkan dengan model berbasis ensemble seperti Gradient Boosting. Sebaliknya, Random Forest Regressor menunjukkan kinerja yang lebih rendah dibandingkan dua model lainnya. MAE dan RMSE-nya lebih tinggi, sedangkan nilai R^2 jauh lebih rendah, yaitu 0.0509. Hal ini mengindikasikan bahwa Random Forest kurang mampu menangkap pola data secara efektif, yang mungkin disebabkan oleh overfitting. Masalah overfitting sering kali terjadi pada Random Forest ketika model terlalu fokus pada data pelatihan, menangkap noise alih-alih pola yang mendasari. Selain itu, parameter default yang digunakan dalam penelitian ini dapat menjadi faktor penyebab kinerja yang kurang optimal. Parameter seperti jumlah pohon (number of trees), kedalaman maksimum pohon (max depth),

p-ISSN: 2775-4227
e-ISSN: 2775-4219

dan ukuran minimum sampel (min samples split dan min samples leaf) membutuhkan penyesuaian lebih lanjut untuk meningkatkan performa model.

Performa moderat yang ditunjukkan oleh ketiga model mengindikasikan bahwa meskipun model machine learning dapat memprediksi probabilitas banjir dengan akurasi tertentu, terdapat ruang yang cukup besar untuk peningkatan. Salah satu pendekatan adalah dengan melakukan penyetelan parameter (hyperparameter tuning) secara lebih menyeluruh, terutama pada Random Forest dan Gradient Boosting. Teknik seperti grid search atau randomized search dapat digunakan untuk menemukan kombinasi parameter yang optimal. Selain itu, eksplorasi model alternatif seperti algoritma pembelajaran mendalam (deep learning) dapat menjadi solusi untuk menangkap hubungan non-linear dan kompleks antar variabel yang mungkin tidak teridentifikasi dengan baik oleh model regresi konvensional atau ensemble. Pendekatan lain yang dapat dipertimbangkan adalah penambahan fitur baru atau rekayasa fitur (feature engineering), seperti interaksi antar variabel atau transformasi variabel non-linear, untuk meningkatkan kemampuan model dalam memahami kompleksitas data. Hasil penelitian ini memiliki implikasi penting bagi pengelolaan risiko banjir, khususnya dalam mendukung sistem peringatan dini yang lebih akurat. Linear Regression dan Gradient Boosting Regressor yang menunjukkan kinerja kompetitif dapat digunakan sebagai dasar untuk sistem prediksi risiko banjir berbasis data yang lebih efisien dan andal. Namun, untuk penerapan yang lebih luas, peningkatan akurasi prediksi melalui eksplorasi model tambahan, optimalisasi parameter, dan pengayaan fitur menjadi langkah yang krusial.

3.3. Keterbatasan

Meskipun penelitian ini memberikan informasi kemampuan prediksi model yang dipilih, terdapat beberapa keterbatasan seperti dataset yang berfokus pada wilayah India dengan keragaman dan kompleksitas peristiwa banjir di wilayah ini menimbulkan tantangan. Variabilitas dalam data dapat memengaruhi pelatihan model dan kemampuannya untuk melakukan generalisasi. Selain itu, mungkin ada prediktor yang terlewatkan yang berpotensi meningkatkan kinerja model. Pengaruh faktor lingkungan dan manusia terhadap probabilitas banjir sangat kompleks, sehingga menangkap dinamika ini secara akurat masih menjadi tantangan. Kemudian, setiap model memiliki kekuatan dan kelemahan. Sebagai contoh, asumsi linearitas pada Linear Regression dan kecenderungan Random Forest untuk overfitting dalam kondisi tertentu dapat membatasi akurasi prediksi. Sementara itu, Gradient Boosting yang memiliki performa tinggi memerlukan penyetelan parameter yang cermat untuk menghindari overfitting dan dapat memerlukan sumber daya komputasi yang besar.

3.4. Saran untuk Penelitian Selanjutnya

Mengatasi keterbatasan ini membuka peluang penelitian lebih lanjut di masa depan. Upaya pengumpulan data yang lebih baik dapat meningkatkan pelatihan model dan akurasinya, terutama untuk wilayah atau periode yang kurang terwakili. Selain itu, eksplorasi fitur tambahan, penerapan teknik rekayasa fitur lanjutan, dan penyelidikan terhadap penggunaan algoritma machine learning alternatif dapat menghasilkan informasi baru. Selanjutnya, proses penyetelan parameter yang lebih menyeluruh, termasuk pemanfaatan alat otomatis seperti Automated Machine Learning (AutoML), berpotensi meningkatkan kinerja model secara signifikan. AutoML memungkinkan eksplorasi parameter dan algoritma secara otomatis dan efisien, sehingga menghasilkan model yang lebih optimal tanpa membutuhkan intervensi manual yang besar. Hasil dari penelitian ini berkontribusi pada perkembangan pengetahuan tentang aplikasi machine learning dalam prediksi banjir. Dengan menjelaskan kekuatan dan kelemahan berbagai pendekatan pemodelan dalam konteks ini, penelitian ini membantu membuka jalan menuju metodologi peramalan banjir yang lebih akurat dan andal. Ini juga memberikan dasar yang kuat untuk penelitian di masa depan dalam meningkatkan sistem peringatan dini dan mitigasi risiko banjir.

4. KESIMPULAN

Penelitian ini mengevaluasi tiga model machine learning—Linear Regression, Random Forest Regressor, dan Gradient Boosting Regressor—untuk memprediksi probabilitas banjir di India. Hasilnya menunjukkan bahwa Linear Regression dan Gradient Boosting Regressor memiliki kinerja yang kompetitif, dengan nilai MAE dan RMSE yang hampir setara. Namun, Linear Regression sedikit lebih unggul dalam menjelaskan variabilitas probabilitas banjir berdasarkan nilai R^2. Sebaliknya, Random Forest Regressor menunjukkan kinerja yang lebih rendah, yang mengindikasikan kemungkinan overfitting atau perlunya penyetelan parameter lebih lanjut. Secara keseluruhan, nilai R^2 yang moderat pada semua model mengungkapkan peluang besar untuk meningkatkan akurasi prediksi dan pemahaman tentang faktor-faktor utama yang memengaruhi banjir. Hasil penelitian ini memberikan wawasan penting bagi pengelolaan risiko banjir dan perencanaan kebijakan, khususnya dalam mendukung pengembangan sistem peringatan dini yang lebih akurat dan strategi mitigasi banjir yang lebih efektif. Model terbaik yang diidentifikasi, seperti Linear Regression dan Gradient Boosting Regressor, dapat

p-ISSN: 2775-4227
e-ISSN: 2775-4219

diimplementasikan untuk memperbaiki analisis risiko banjir, terutama dalam pengelolaan sumber daya air dan perencanaan tata ruang perkotaan. Peluang untuk penelitian lanjutan meliputi eksplorasi algoritma pembelajaran mendalam (deep learning) yang memiliki potensi untuk menangkap pola kompleks antar variabel, serta integrasi data spasial dan data real-time untuk meningkatkan akurasi prediksi. Selain itu, penggunaan Automated Machine Learning (AutoML) dapat mengoptimalkan kinerja model dengan mempermudah eksplorasi parameter dan algoritma secara otomatis. Dengan pendekatan ini, penelitian di masa depan dapat lebih mendalami hubungan non-linear dan kompleks antar variabel, serta memperluas cakupan geografis untuk mencakup wilayah lain yang rentan terhadap banjir. Kesimpulannya, penelitian ini tidak hanya memberikan dasar ilmiah yang kuat untuk pengembangan prediksi banjir berbasis data, tetapi juga menawarkan panduan praktis bagi pengelolaan risiko bencana yang lebih baik.

DAFTAR PUSTAKA

[1] S. N. Jonkman dan J. K. Vrijling, "Loss of Life Due to Floods," *J. Flood Risk Manag.*, 2008, doi: 10.1111/j.1753-318x.2008.00006.x.

[2] S. N. Jonkman, M. Kok, dan J. K. Vrijling, "Flood Risk Assessment in the Netherlands: A Case Study for Dike Ring South Holland," *Risk Anal.*, 2008, doi: 10.1111/j.1539-6924.2008.01103.x.

[3] Y. Qian, Y. Wang, dan N. Li, "Extreme Flood Disasters: Comprehensive Impact and Assessment," *Water*, 2022, doi: 10.3390/w14081211.

[4] B. Roy, J. U. Khan, A. K. M. Saiful Islam, K. Mohammed, dan Md. J. U. Khan, "Climate-Induced Flood Inundation for the Arial Khan River of Bangladesh Using Open-Source SWAT and HEC-RAS Model for RCP8.5-SSP5 Scenario," *Sn Appl. Sci.*, 2021, doi: 10.1007/s42452-021-04460-4.

[5] W. H. Mobley, A. Sebastian, R. Blessing, W. E. Highfield, L. Stearns, dan S. D. Brody, "Quantification of Continuous Flood Hazard Using Random Forest Classification and Flood Insurance Claims at Large Spatial Scales: A Pilot Study in Southeast Texas," *Nat. Hazards Earth Syst. Sci.*, 2021, doi: 10.5194/nhess-21-807-2021.

[6] J.-Y. Lee dan B.-H. Kim, "Scenario-Based Real-Time Flood Prediction With Logistic Regression," *Water*, 2021, doi: 10.3390/w13091191.

[7] S. Nevo *dkk.*, "Flood Forecasting With Machine Learning Models in an Operational Framework," *Hydrol. Earth Syst. Sci.*, 2022, doi: 10.5194/hess-26-4013-2022.

[8] A. Faruq, S. F. M. Hussein, A. Marto, dan S. S. Abdullah, "Flood River Water Level Forecasting Using Ensemble Machine Learning for Early Warning Systems," *Iop Conf. Ser. Earth Environ. Sci.*, 2022, doi: 10.1088/1755-1315/1091/1/012041.

[9] J. Liu *dkk.*, "Editorial: Spatial Modelling and Failure Analysis of Natural and Engineering Disasters Through Data-Based Methods," *Front. Earth Sci.*, 2022, doi: 10.3389/feart.2022.1000540.

[10] J. PUNGCHING dan S. PILAILAR, "Developing a Flood Forecasting System With Machine Learning and Applying to Geographic Information System," *Geogr. Tech.*, 2022, doi: 10.21163/gt_2023.181.01.

[11] Y. Chen, X. Zhang, K. Yang, S. Zeng, dan A. Hong, "Modeling Rules of Regional Flash Flood Susceptibility Prediction Using Different Machine Learning Models," *Front. Earth Sci.*, 2023, doi: 10.3389/feart.2023.1117004.

[12] I. Maulita dan A. Wahid, "Prediksi Magnitudo Gempa Menggunakan Random Forest, Support Vector Regression, XGBoost, LightGBM, dan Multi-Layer Perceptron Berdasarkan Data Kedalaman dan Geolokasi (Predicting Earthquake Magnitude Using Random Forest, Support Vector Regression, XGBoost, LightGBM, and Multi-Layer Perceptron Based on Depth and Geolocation Data)," *J. Pendidik. Dan Teknol. Indones.*, vol. 4, hlm. 221–232, Mei 2024, doi: 10.52436/1.jpti.470.

[13] I. Maulita, N. R. Prasetyaningsih, U. Pratiwi, dan A. Azimi, "ANALISIS SECOND VERTICAL DERIVATIVE DATA GRAVITASI UNTUK MENGINTERPRETASIKAN STRUKTUR BAWAH PERMUKAAN DAERAH LEMBANG," *J. Ilmu Fis. Dan Ter.*, vol. 11, no. 2, Art. no. 2, Okt 2024, doi: 10.21831/fisika.

[14] Y. Xu, "Flood Forecasting Method and Application Based on Informer Model," *Water*, vol. 16, no. 5, hlm. 765, 2024, doi: 10.3390/w16050765.

[15] S. Hitouri, "Flood Susceptibility Mapping Using SAR Data and Machine Learning Algorithms in a Small Watershed in Northwestern Morocco," *Remote Sens.*, vol. 16, no. 5, hlm. 858, 2024, doi: 10.3390/rs16050858.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[16] R. Kondo, B. Du, Y. Narusue, dan H. Morikawa, "Machine Learning Framework Supervised by Hydraulic Mechanical Models for Real-Time Pluvial Flood Prediction," *J. Inf. Process.*, vol. 31, no. 0, hlm. 256–264, 2023, doi: 10.2197/ipsjjip.31.256.

[17] G. K. Wedajo, "Integrating Satellite Images and Machine Learning for Flood Prediction and Susceptibility Mapping for the Case of Amibara, Awash Basin, Ethiopia," *Remote Sens.*, vol. 16, no. 12, hlm. 2163, 2024, doi: 10.3390/rs16122163.

[18] S. Janizadeh *dkk.*, "Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran," *Sustainability*, vol. 11, no. 19, hlm. 5426, 2019, doi: 10.3390/su11195426.