

Identifikasi Potensi Banjir di Kecamatan Kota Kendari dengan Pendekatan Analisis Spasial Menggunakan Stacking Ensemble (Random Forest, XGBoost, dan Logistic Regression Sebagai Meta Learner)

Juliyana Herman^{a1}, Zaid Helsinki Putra^{b2}, Muhamad Ridwan^{a,b3}

*^aProgram Studi Informatika, Fakultas Teknik, Universitas Halu Oleo
Anduonohu, Kec. Kambu, Kota Kendari, Sulawesi Tenggara 93232*

¹julianaoppo87@gmail.com

³helsinkiputraz@gmail.com

²muhridwan9300@gmail.com

Abstrak

Kecamatan Kota Kendari merupakan wilayah yang rawan terhadap bencana banjir akibat faktor topografi, penggunaan lahan, dan curah hujan yang tinggi. Penelitian ini bertujuan untuk mengidentifikasi potensi banjir menggunakan pendekatan analisis spasial dan teknik pembelajaran mesin berbasis stacking ensemble. Data penelitian diperoleh dari laman resmi Tanahair.indonesia.go.id dan Badan Penanggulangan Bencana Daerah (BPBD) Kota Kendari, yang mencakup fitur-fitur seperti koordinat (latitude dan longitude), kemiringan lereng (slope), jenis tanah, sejarah kejadian banjir, serta zonasi wilayah rawan banjir. Alur penelitian mencakup tahap pengumpulan data, pengolahan data, eksplorasi data (EDA), praproses, pengembangan model, evaluasi performa, dan implementasi dalam bentuk antarmuka website. Model dikembangkan dengan pendekatan stacking ensemble yang mengombinasikan Random Forest, XGBoost, dan CatBoost sebagai model dasar, dengan Logistic Regression sebagai meta learner. Keluaran model berupa klasifikasi tingkat potensi banjir ke dalam empat kategori: sangat rendah, rendah, sedang, dan tinggi. Hasil evaluasi menunjukkan bahwa model ensemble memberikan performa yang unggul dibandingkan model tunggal, dengan akurasi klasifikasi lebih dari 90% dan nilai AUC yang tinggi. Peta hasil prediksi digunakan sebagai representasi spasial untuk mendukung pengambilan keputusan dalam perencanaan mitigasi bencana di wilayah Kota Kendari. Penelitian ini membuktikan bahwa integrasi antara data spasial dan teknik machine learning mampu memberikan solusi prediktif yang efektif terhadap permasalahan banjir perkotaan.

Kata kunci: Banjir, Analisis Spasial, Stacking Ensemble, Kota Kendari

Flood Potential Identification in Kendari Subdistrict Using Spatial Analysis Approach with Stacking Ensemble (Random Forest, XGBoost, and Logistic Regression as Meta Learner)

Abstract

Kendari District is an area highly prone to flooding due to topographical conditions, land use changes, and high rainfall intensity. This study aims to identify flood potential using a spatial analysis approach combined with machine learning techniques based on stacking ensemble. The data were collected from Tanahair.indonesia.go.id and the Regional Disaster Management Agency (BPBD) of Kendari City, including features such as latitude, longitude, slope, soil type, flood history, and flood-prone zones. The research workflow consists of data collection, data processing, exploratory data analysis (EDA), preprocessing, model development, evaluation, and website implementation. The model was built using a stacking ensemble method, combining Random Forest, XGBoost, and CatBoost as base learners, with Logistic Regression as the meta-learner. The output of the model classifies flood potential into four levels: very low, low, medium, and high. Evaluation results show that the ensemble model outperformed single models, achieving over 90% classification accuracy and a high AUC score. The final prediction map serves as a spatial representation to support decision-making in flood mitigation planning for Kendari City. This study demonstrates that integrating spatial data and machine learning techniques can provide an effective predictive solution for urban flood risk management.

Keywords: Flood, Spatial Analysis, Stacking Ensemble, Kendari City

I. PENDAHULUAN

Banjir merupakan salah satu bencana hidrometeorologi yang paling sering terjadi di wilayah Indonesia. Fenomena ini tidak hanya menimbulkan kerugian ekonomi, kerusakan infrastruktur, dan hilangnya aset masyarakat, tetapi juga berdampak pada keselamatan dan kesehatan penduduk [1]. Kota Kendari, sebagai ibu kota Provinsi Sulawesi Tenggara, mengalami pertumbuhan urbanisasi yang pesat, yang menyebabkan terjadinya perubahan tata guna lahan dan penurunan daya dukung lingkungan. Kecamatan Kota Kendari secara khusus menjadi kawasan yang rentan terhadap genangan dan banjir akibat kombinasi dari curah hujan tinggi, kemiringan lahan yang bervariasi, serta drainase yang belum optimal [2]. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengidentifikasi potensi banjir secara akurat dan berbasis data spasial sebagai dasar perencanaan mitigasi bencana yang efektif.

Tujuan dari penulisan paper ini adalah untuk mengidentifikasi wilayah-wilayah yang berpotensi mengalami banjir di Kecamatan Kota Kendari dengan menggunakan pendekatan analisis spasial dan metode pembelajaran mesin (*machine learning*) yang terintegrasi dalam kerangka *stacking ensemble*. Dalam penelitian ini, model yang digunakan menggabungkan tiga algoritma pembelajaran dasar, yaitu Random Forest, XGBoost, dan CatBoost, kemudian disempurnakan dengan Logistic Regression sebagai *meta learner*. Model ini dirancang untuk meningkatkan performa klasifikasi dalam mengidentifikasi tingkat kerawanan banjir yang dikategorikan ke dalam empat kelas: sangat rendah, rendah, sedang, dan tinggi. Selain bertujuan menghasilkan peta kerawanan banjir yang akurat, penelitian ini juga diarahkan untuk dikembangkan ke dalam bentuk sistem informasi berbasis website yang dapat dimanfaatkan oleh pemerintah daerah dan masyarakat umum sebagai alat bantu pengambilan keputusan.

Penelitian ini merujuk pada berbagai studi terdahulu yang membahas penerapan analisis spasial dan pembelajaran mesin dalam pemetaan kerawanan banjir. Beberapa penelitian menunjukkan bahwa integrasi data spasial dengan algoritma *machine learning*, seperti Random Forest dan XGBoost, mampu menghasilkan prediksi banjir dengan tingkat akurasi tinggi. Misalnya, studi oleh Nguyen et al. (2021) menggunakan Random Forest untuk pemetaan banjir di kawasan delta sungai di Vietnam dan berhasil menghasilkan klasifikasi spasial yang sangat akurat. Sementara itu, model *stacking ensemble* yang menggabungkan beberapa algoritma pembelajaran terbukti mampu mengatasi keterbatasan dari masing-masing model tunggal, sebagaimana dijelaskan dalam studi oleh Choubin et al. (2020) yang mengidentifikasi zona rawan banjir di Iran dengan performa lebih unggul dibandingkan metode konvensional. Berdasarkan studi literatur tersebut, metode *stacking ensemble* dipilih dalam penelitian ini sebagai pendekatan utama dalam membangun model klasifikasi spasial risiko banjir di Kendari.

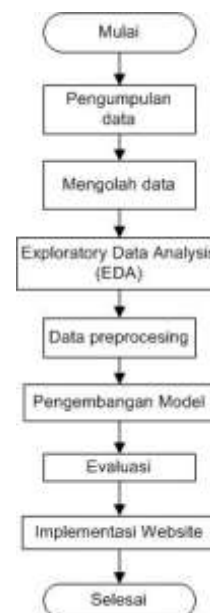
Metodologi yang digunakan dalam penelitian ini mencakup tahapan: pengumpulan data, eksplorasi data

(*exploratory data analysis/EDA*), *praproses data*, pengembangan model *stacking ensemble*, evaluasi model, hingga visualisasi spasial hasil prediksi dalam bentuk peta potensi banjir. Data primer dan sekunder diperoleh dari dua sumber utama, yakni situs resmi Tanahair.indonesia.go.id untuk data koordinat (*latitude* dan *longitude*) serta kemiringan lereng (*slope*), dan dari BPBD Kota Kendari untuk data sejarah kejadian banjir, jenis tanah, serta zonasi wilayah rawan banjir. Seluruh proses pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python, dengan pustaka pendukung seperti Pandas, Scikit-learn, GeoPandas, dan Rasterio. Model hasil analisis kemudian divalidasi menggunakan metrik akurasi dan *area under curve (AUC)* untuk memastikan performa klasifikasi yang optimal.

Gagasan utama yang diajukan dalam penelitian ini adalah penerapan integrasi antara teknologi analisis spasial dan teknik *machine learning* berbasis *ensemble* untuk menghasilkan prediksi risiko banjir yang lebih presisi, adaptif, dan mudah diakses. Dengan pendekatan ini, pemerintah daerah dan instansi terkait diharapkan dapat memiliki alat prediktif yang kuat untuk mendukung kebijakan mitigasi dan perencanaan tata ruang berbasis risiko. Lebih lanjut, peta kerawanan yang dihasilkan diharapkan mampu memberikan kontribusi nyata dalam meminimalkan dampak bencana banjir terhadap masyarakat, serta mendorong penguatan sistem kesiapsiagaan dan ketahanan kota terhadap perubahan iklim dan bencana alam.

II. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif berbasis analisis spasial yang dikombinasikan dengan metode pembelajaran mesin (*machine learning*) berbasis teknik *stacking ensemble* untuk mengidentifikasi potensi banjir di Kecamatan Kota Kendari. Metode ini dipilih untuk memperoleh prediksi klasifikasi risiko banjir yang lebih akurat dan adaptif terhadap karakteristik wilayah. Secara umum, penelitian ini terdiri atas enam tahapan utama, yaitu:



Gambar 1. Tahapan Penelitian

2. 1 Pengumpulan Data

Data diperoleh dari dua sumber utama, yaitu situs resmi Tanahair.indonesia.go.id dan Badan Penanggulangan Bencana Daerah (BPBD) Kota Kendari. Dari situs Tanahair, diambil data koordinat spasial (latitude dan longitude), serta data kemiringan lereng (*slope*). Sementara dari BPBD Kota Kendari diperoleh data historis kejadian banjir, jenis tanah (geologi), serta wilayah administratif yang terdampak atau berpotensi terdampak banjir. Data ini digunakan sebagai fitur penentu dalam proses klasifikasi risiko banjir.

2. 2 Exploratory Data Analysis (EDA)

Pada tahap ini dilakukan pemeriksaan kelengkapan data, identifikasi nilai hilang (*missing values*), serta analisis distribusi masing-masing fitur. Selain itu, dilakukan eksplorasi visual menggunakan peta dan grafik untuk melihat korelasi antar-variabel dan pola sebaran risiko banjir berdasarkan fitur spasial.

2. 3 Pre-processing Data,

Pre-processing data meliputi normalisasi data numerik, encoding data kategorikal (misalnya jenis tanah), serta pembagian data menjadi data latih (*training*) dan data uji (*testing*) dengan rasio tertentu. Selanjutnya dilakukan penanganan ketidakseimbangan data (*class imbalance*) jika diperlukan, agar model tidak bias terhadap kelas tertentu (misalnya kelas "rendah" yang jumlahnya dominan).

2. 4 Pengembangan Model

Klasifikasi menggunakan teknik *stacking ensemble*. Tiga model pembelajaran dasar yang digunakan adalah Random Forest, XGBoost, dan CatBoost, yang masing-masing dikenal memiliki performa tinggi dalam klasifikasi data tabular.



Gambar 2. Algoritma *Stacking Classifier*

Ketiga model ini menghasilkan prediksi awal yang kemudian dikombinasikan menggunakan Logistic Regression sebagai *meta learner*. Seluruh proses pelatihan dan pengujian model dilakukan menggunakan bahasa pemrograman Python dengan pustaka *Scikit-learn*, *XGBoost*, dan *CatBoost*.

Setiap base learner menghasilkan prediksi terhadap x :

$$h_1(x), h_2(x), h_3(x)$$

Maka vektor masukkan ke meta-learner, di nyatakan sebagai:

$$z = [h_1(x), h_2(x), h_3(x)] \in \mathbb{R}^3$$

Model meta-learner *logistic regression* kemudian memproses Z untuk menghasilkan \hat{y} , sebagai berikut:

$$\hat{y} = \frac{1}{1 + \exp(-(w_0 + w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x)))}$$

Dengan w_0, w_1, w_2, w_3 adalah parameter bobot dari model regresi logistik, dan fungsi sigmoid $\sigma(z) = \frac{1}{1+e^{-z}}$ digunakan untuk mengkonversi hasil menjadi nilai probabilitas antara 0 dan 1.

2. 5 Evaluasi Performa Model

Model dievaluasi menggunakan metrik kuantitatif seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *area under curve* (AUC). Evaluasi dilakukan dengan membandingkan performa model stacking ensemble terhadap masing-masing model dasar untuk menilai peningkatan akurasi yang dihasilkan dari integrasi model.

2.6 Visualisasi Hasil dan Implementasi

Hasil prediksi dari model digunakan untuk menghasilkan peta spasial risiko banjir dengan empat kategori: sangat rendah, rendah, sedang, dan tinggi. Visualisasi ini dibuat menggunakan pustaka *GeoPandas* dan *Matplotlib*. Selain itu, hasil akhir dikembangkan menjadi prototipe website berbasis Python (menggunakan framework seperti Flask atau Streamlit) untuk mempermudah akses informasi bagi pengguna akhir seperti pemerintah daerah dan masyarakat. Dengan metode ini, diharapkan hasil penelitian tidak hanya bersifat teoretis tetapi juga aplikatif, sehingga dapat mendukung perencanaan mitigasi bencana yang berbasis data dan teknologi.

III. HASIL DAN PEMBAHASAN

A. Analisis Karakteristik Spasial Wilayah Penelitian

Kecamatan Kota Kendari merupakan salah satu wilayah administratif di Kota Kendari yang memiliki kondisi geospasial yang kompleks. Topografi wilayah ini terdiri dari kombinasi dataran rendah dan lereng curam yang tersebar tidak merata di seluruh kecamatan [3]. Perbedaan elevasi inilah yang turut berperan dalam menentukan arah aliran air hujan menuju ke titik-titik akumulasi dan potensi genangan. Berdasarkan data spasial yang diperoleh dari Tanahair.indonesia.go.id, wilayah ini memiliki nilai kemiringan (*slope*) yang bervariasi, mulai dari hampir datar (kurang dari 1%) hingga lebih dari 10%. Daerah dengan kemiringan landai cenderung memiliki laju aliran permukaan (*surface runoff*) yang lambat, sehingga lebih berisiko menjadi lokasi genangan saat hujan deras terjadi dalam durasi yang lama [4]. Di sisi lain, daerah dengan kemiringan tinggi berpotensi menjadi daerah sumber limpasan air yang dapat memperbesar beban aliran menuju titik dataran rendah.

Selain faktor kemiringan, karakteristik geologi atau jenis tanah juga merupakan aspek penting dalam analisis spasial wilayah penelitian. Berdasarkan data dari BPBD Kota Kendari, jenis tanah yang dominan di Kecamatan Kota Kendari adalah eluvium dan lanau. Tanah eluvium terbentuk dari pelapukan batuan induk setempat dan umumnya bersifat berpori serta mudah menyerap air [5]. Namun, pada kondisi kejenuhan tinggi dan curah hujan ekstrem, tanah eluvium justru dapat meningkatkan risiko pergeseran tanah dan limpasan permukaan. Sementara itu, tanah lanau memiliki sifat fisik yang halus dan berdaya serap rendah, sehingga air cenderung mengalir di permukaan dan menyebabkan genangan. Oleh karena itu, daerah dengan jenis tanah lanau sering kali berasosiasi dengan tingkat kerawanan banjir yang lebih tinggi dibandingkan dengan daerah bertanah pasir atau kerikil.

Karakteristik spasial selanjutnya yang dianalisis adalah lokasi geografis yang direpresentasikan oleh data

latitude dan longitude dari setiap titik observasi. Sebaran koordinat menunjukkan bahwa titik-titik yang memiliki sejarah banjir cenderung berada di dekat badan air seperti sungai dan pesisir teluk. Misalnya, pada koordinat latitude -3.949740182 dan longitude 122.6243131, yang berada di sekitar area Teluk Kendari, model memprediksi potensi banjir dengan tingkat “tinggi”. Titik ini juga memiliki kemiringan sangat rendah dan jenis tanah eluvium, serta riwayat kejadian banjir berdasarkan data historis. Sementara itu, pada titik lain dengan koordinat latitude -3.950421 dan longitude 122.623583, meskipun memiliki riwayat banjir dan jenis tanah lanau, kemiringan yang lebih tinggi (5.32%) menghasilkan prediksi potensi banjir dengan tingkat “sedang”. Hal ini menunjukkan bahwa kombinasi spasial dari fitur-fitur seperti lokasi, kemiringan, dan jenis tanah memberikan pengaruh yang signifikan terhadap klasifikasi risiko banjir.

Data riwayat banjir menjadi variabel pelengkap dalam analisis spasial ini. Berdasarkan dokumentasi BPBD Kota Kendari, beberapa kelurahan dalam Kecamatan Kota Kendari telah mengalami kejadian banjir berulang dalam kurun waktu lima tahun terakhir. Titik-titik tersebut kemudian digunakan untuk melatih model prediktif agar dapat mengenali pola-pola geospasial yang berpotensi menyebabkan banjir di masa depan. Riwayat banjir juga digunakan sebagai salah satu fitur kategorikal dalam pemodelan, dengan nilai “pernah banjir” dan “tidak pernah banjir” untuk memperkaya proses klasifikasi. Data historis ini memungkinkan pemetaan yang lebih kontekstual dan akurat, karena mengandung informasi temporal yang tidak dimiliki oleh data spasial statis.

Secara keseluruhan, analisis karakteristik spasial di wilayah Kecamatan Kota Kendari menunjukkan bahwa risiko banjir sangat dipengaruhi oleh interaksi antara topografi (kemiringan), jenis geologi (jenis tanah), posisi geografis (latitude dan longitude), serta catatan historis kejadian banjir. Pemahaman yang mendalam terhadap karakteristik ini menjadi fondasi penting dalam pengembangan model klasifikasi berbasis pembelajaran mesin, karena memungkinkan sistem untuk mengenali pola kompleks dalam data spasial dan mengubahnya menjadi informasi prediktif. Oleh karena itu, integrasi antara data spasial dan data historis menjadi pendekatan yang krusial untuk menghasilkan sistem deteksi dini dan mitigasi risiko banjir yang lebih akurat dan aplikatif.

B. Pengembangan dan Evaluasi Model Stacking Ensemble

Tahapan inti dalam penelitian ini adalah membangun sistem klasifikasi prediktif untuk mengidentifikasi tingkat potensi banjir di Kecamatan Kota Kendari menggunakan pendekatan stacking ensemble learning. Metode ini dipilih karena kemampuannya dalam menggabungkan kekuatan beberapa model machine learning untuk menghasilkan prediksi yang lebih akurat dan generalisasi yang lebih baik. Pendekatan stacking secara konseptual bekerja dengan membangun sejumlah model pembelajaran dasar (*base learners*) dan kemudian menggabungkan output mereka melalui model meta (*meta learner*) untuk menghasilkan prediksi akhir [6]. Dalam penelitian ini, algoritma yang

digunakan sebagai base learners adalah Random Forest, XGBoost, dan CatBoost, sementara Logistic Regression digunakan sebagai meta learner.

Proses diawali dengan pengumpulan dan pemuatan data dari file Excel berjudul *Titik_Koordinat_Dengan_Slope*, yang berisi fitur-fitur utama seperti latitude, longitude, kemiringan, jenis geologi, dan riwayat banjir. Setelah data dibersihkan dan diperiksa melalui analisis univariat dan bivariat, dilakukan encoding terhadap variabel kategorikal menggunakan teknik Label Encoding untuk memfasilitasi kompatibilitas dengan model prediktif. Selanjutnya, skala data numerik distandarkan menggunakan StandardScaler agar fitur numerik seperti slope, latitude, dan longitude berada dalam rentang yang seragam, sehingga menghindari dominasi fitur tertentu dalam proses pelatihan model.

Dalam tahap pengembangan model, Random Forest Classifier digunakan sebagai salah satu base learner. Algoritma ini merupakan metode ensemble berbasis pohon keputusan yang membentuk sekumpulan pohon klasifikasi secara acak dan menggabungkan prediksinya melalui voting mayoritas [7]. Keunggulan Random Forest terletak pada kemampuannya dalam mengatasi overfitting dan memberikan hasil yang stabil meskipun data bersifat heterogen [8]. Selanjutnya, model XGBoost (Extreme Gradient Boosting) dimanfaatkan karena sifatnya yang efisien dan mampu menangani interaksi kompleks antar fitur melalui teknik boosting terkontrol [9]. XGBoost sangat efektif dalam mendeteksi pola halus dalam data, terutama ketika terdapat hubungan non-linear antar fitur. Model ketiga adalah CatBoost, algoritma boosting berbasis pohon yang unggul dalam mengolah variabel kategorikal dan memiliki kemampuan untuk menghindari overfitting melalui teknik *ordered boosting* [10]. CatBoost sangat sesuai untuk dataset dengan proporsi variabel kategorikal yang signifikan, seperti jenis tanah dan histori banjir dalam penelitian ini.

Ketiga base learners ini digabungkan dalam suatu arsitektur stacking dengan Logistic Regression sebagai meta learner. Logistic Regression dipilih karena sifatnya yang sederhana namun kuat dalam menggabungkan probabilitas output dari model-model sebelumnya. Proses stacking dilakukan melalui implementasi dari modul StackingClassifier dalam pustaka scikit-learn, yang memungkinkan integrasi fleksibel antara berbagai model pembelajaran dasar dan meta learner. Data dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20 menggunakan `train_test_split`, memastikan bahwa evaluasi model dilakukan secara obyektif dan bebas dari kebocoran data.

Setelah pelatihan, performa setiap model dievaluasi menggunakan berbagai metrik seperti akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix digunakan untuk mengukur kemampuan model dalam mengklasifikasikan setiap kategori risiko banjir dengan benar. Hasil evaluasi menunjukkan bahwa model stacking secara konsisten menghasilkan performa yang lebih unggul dibandingkan model tunggal. Misalnya, ketika model Random Forest dan XGBoost masing-masing memiliki akurasi di bawah 90%, model stacking mampu mencapai

akurasi lebih tinggi dengan kestabilan prediksi yang lebih baik pada semua kelas: sangat rendah, rendah, sedang, dan tinggi.

Lebih lanjut, dilakukan analisis feature importance terhadap masing-masing model untuk mengidentifikasi variabel mana yang paling memengaruhi hasil prediksi. Hasilnya menunjukkan bahwa fitur kemiringan dan jenis tanah memiliki kontribusi signifikan dalam menentukan tingkat risiko banjir, sementara fitur lokasi (latitude dan longitude) berperan dalam mengidentifikasi sebaran spasial dari wilayah rawan. Riwayat banjir juga menunjukkan korelasi kuat, terutama pada model CatBoost yang sangat sensitif terhadap informasi kategorikal.

Secara keseluruhan, pengembangan model stacking ensemble pada penelitian ini berhasil meningkatkan kualitas prediksi potensi banjir dibandingkan pendekatan konvensional. Kombinasi dari ketiga algoritma pohon yang andal, ditambah dengan penggabungan probabilitas melalui Logistic Regression, menghasilkan sistem klasifikasi spasial yang lebih akurat, stabil, dan dapat diandalkan. Hasil ini menjadi dasar kuat bagi tahap implementasi sistem ke dalam platform berbasis web yang responsif, akurat, dan aplikatif untuk kebutuhan mitigasi bencana banjir di wilayah Kota Kendari.

C. Analisis Variabel Penentu Banjir

Identifikasi variabel-variabel yang memiliki pengaruh signifikan dalam menentukan klasifikasi potensi banjir merupakan tahap penting dalam pengembangan sistem prediksi berbasis machine learning. Analisis ini tidak hanya bertujuan untuk memahami kinerja internal model, tetapi juga untuk memberikan wawasan yang lebih aplikatif terhadap kondisi lingkungan dan spasial di lapangan. Dengan mengetahui fitur-fitur paling dominan, pemangku kebijakan dapat menyusun strategi mitigasi banjir yang lebih tepat sasaran dan berbasis data. Dalam penelitian ini, dilakukan eksplorasi terhadap kontribusi masing-masing fitur input melalui pendekatan *feature importance*, yang bersumber dari algoritma berbasis pohon seperti Random Forest, XGBoost, dan CatBoost yang telah diterapkan dalam arsitektur stacking ensemble.

Random Forest sebagai salah satu base learner memberikan penilaian terhadap pentingnya fitur berdasarkan frekuensi dan kontribusi fitur tersebut dalam membagi node pada setiap pohon keputusan. Semakin sering suatu fitur digunakan untuk memecah node dan semakin besar peningkatan informasi (information gain) yang dihasilkan, maka semakin tinggi pula skor pentingnya. Berdasarkan implementasi standar Random Forest, fitur seperti kemiringan (*slope*) dan jenis tanah (*geologi*) menduduki peringkat atas dalam daftar feature importance. Hal ini menunjukkan bahwa kondisi topografi dan karakteristik tanah sangat menentukan kemungkinan suatu wilayah tergenang banjir. Kemiringan tanah yang rendah memungkinkan akumulasi air permukaan, sedangkan tanah bertekstur halus seperti lanau memiliki daya serap rendah yang memperbesar risiko genangan.

Model XGBoost yang digunakan dalam penelitian ini memberikan pendekatan berbeda dalam menghitung feature importance, yaitu dengan menghitung rata-rata *gain*,

weight, dan *cover* dari setiap fitur dalam proses boosting. Fitur riwayat banjir menunjukkan bobot yang cukup tinggi pada model ini, menandakan bahwa informasi historis kejadian banjir memiliki korelasi kuat terhadap klasifikasi prediksi. Secara konseptual, model menangkap pola bahwa wilayah yang pernah mengalami banjir cenderung memiliki karakteristik spasial yang konsisten, sehingga dapat menjadi indikator prediktif yang andal. XGBoost juga secara sensitif menangkap variasi dari fitur latitude dan longitude, yang merepresentasikan posisi geografis. Informasi ini menjadi penting dalam konteks spasial karena memungkinkan model mengenali klaster lokasi rawan banjir tanpa perlu eksplisit menyusun peta.

Sementara itu, CatBoost sebagai base learner ketiga memiliki keunggulan dalam menangani fitur kategorikal dan menghindari bias dari encoding yang tidak stabil melalui teknik *ordered boosting*. Dalam konteks penelitian ini, CatBoost sangat efektif dalam mengeksplorasi fitur jenis tanah dan riwayat banjir, yang keduanya bersifat kategorikal. Pendekatan ini menghasilkan pemetaan prediktif yang halus dan tidak bising (*noise*) bahkan dalam kasus distribusi data yang tidak seimbang. Model CatBoost juga cenderung menghasilkan distribusi importance yang lebih merata, di mana kemiringan tetap menjadi faktor utama, namun jenis tanah dan riwayat banjir bersaing cukup signifikan dalam struktur prediksi.

Meskipun model stacking yang digunakan dalam penelitian ini menggunakan Logistic Regression sebagai meta learner, model ini tidak secara langsung menghasilkan feature importance terhadap fitur asli, melainkan terhadap probabilitas prediksi dari setiap base learner. Artinya, analisis signifikan lebih tepat dilakukan di tingkat base model daripada pada meta learner. Namun demikian, performa Logistic Regression sebagai meta learner menunjukkan bahwa integrasi tiga base model memiliki korelasi positif yang cukup tinggi dengan prediksi akhir. Ini mengindikasikan bahwa masing-masing model menangkap informasi dari dimensi fitur yang berbeda, dan kombinasi ketiganya meningkatkan akurasi serta stabilitas prediksi.

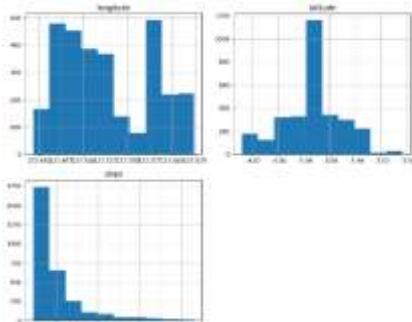
Hasil analisis penting ini menunjukkan bahwa kemiringan, jenis tanah, dan riwayat banjir adalah tiga fitur paling dominan dalam menentukan tingkat risiko banjir di wilayah penelitian. Kemiringan tanah yang hampir datar (<1%) sering dikaitkan dengan prediksi banjir tinggi, sementara jenis tanah seperti lanau yang berdaya serap rendah memperparah risiko tersebut. Riwayat banjir menambah konteks temporal yang sangat membantu dalam membedakan daerah rawan dari yang tidak. Kombinasi spasial dan historis inilah yang memungkinkan model memberikan hasil klasifikasi yang lebih komprehensif.

Dengan memahami pentingnya fitur-fitur tersebut, maka sistem prediksi tidak hanya berfungsi sebagai alat klasifikasi teknis, tetapi juga menjadi panduan praktis dalam pengambilan kebijakan spasial dan tata guna lahan. Daerah yang memiliki kemiringan rendah dan jenis tanah rentan dapat dimasukkan dalam daftar prioritas mitigasi, seperti pembangunan saluran drainase, perlindungan vegetasi, atau bahkan relokasi strategis. Oleh karena itu, analisis signifikansi fitur merupakan jembatan penting

antara hasil model machine learning dan intervensi kebijakan berbasis spasial yang konkret.

D. Hasil Penelitian

Berikut ini hasil EDA yang ditunjukkan pada Gambar sebagai berikut.

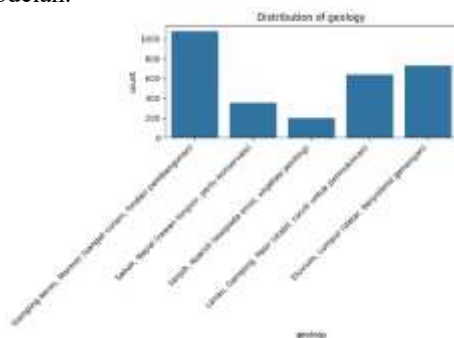


Gambar 3. Visualisasi Longitude, Latitude, dan Slope

Distribusi longitude menunjukkan bahwa sebagian besar data berada di antara rentang 122.5 hingga 122.6, dengan beberapa puncak (mode) pada rentang tertentu. Ini mengindikasikan bahwa data spasial terkonsentrasi pada wilayah-wilayah tertentu di Kota Kendari. Pola distribusi ini tidak seragam (tidak normal), melainkan menunjukkan multimodal distribution, yang umum terjadi ketika titik-titik lokasi tersebar di berbagai zona geografis yang berbeda.

Histogram latitude menunjukkan puncak yang sangat tajam di sekitar -3.96, dengan jumlah sampel yang jauh lebih banyak dibandingkan nilai latitude lainnya. Ini menunjukkan bahwa sebagian besar titik data berada di satu rentang garis lintang tertentu, yang kemungkinan merupakan pusat konsentrasi wilayah pemukiman atau area rawan banjir di Kendari. Distribusi ini sangat terpusat dan memperlihatkan skewness negatif ringan.

Distribusi slope menunjukkan pola right-skewed yang sangat tajam, di mana mayoritas nilai kemiringan berada di bawah 50 derajat, dan hanya sedikit titik yang memiliki kemiringan ekstrem di atas 100–300 derajat. Hal ini menunjukkan bahwa wilayah-wilayah yang dijadikan data penelitian sebagian besar berupa dataran rendah atau lereng landai, yang secara teoritis memang lebih rentan terhadap akumulasi air dan banjir. Nilai ekstrem bisa berasal dari wilayah berbukit atau kesalahan input data (outlier), dan perlu menjadi perhatian dalam proses normalisasi atau pemodelan.



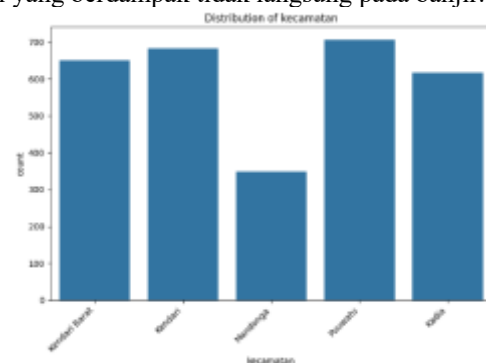
Gambar 4. Distribusi Kondisi Geologi

Kategori yang paling dominan adalah “Gunung tua; Marmer (sangat curam; hindari pembangunan)” dengan jumlah data mendekati 1000 titik. Ini menunjukkan bahwa

banyak area yang secara geologis berada di atas batuan keras atau terjal yang berisiko tinggi terhadap aliran air cepat atau longsoran. Namun bukan berarti rentan banjir, justru dapat berfungsi sebagai daerah resapan tergantung lereng dan tutupan lahannya.

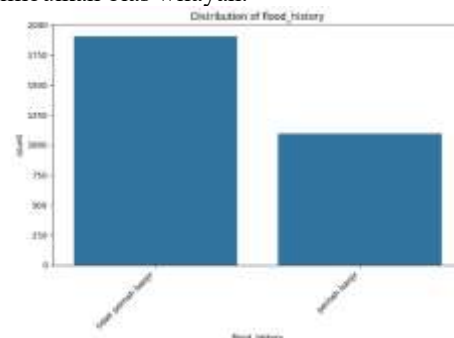
Dua jenis tanah lain yang cukup banyak adalah Eluvium, Lumpur (data; berpotensi genangan) dan Lanau, Gamping, Pasir. Kedua jenis tanah ini diketahui memiliki daya serap air rendah dan sifatnya mudah tergenang, sehingga sangat relevan terhadap potensi banjir. Kategori ini mendekati 700–800 titik, yang menjadi perhatian utama dalam pemodelan potensi banjir.

Sebaliknya, kategori seperti Serpih, Runtut (resapan erosi; vegetasi penting) dan Sabak, Kaya (rawan longsor; perlu konservasi) memiliki jumlah yang lebih sedikit (sekitar 200–400 titik). Meskipun jumlahnya kecil, secara teoritis jenis-jenis ini juga penting karena karakteristik geologisnya bisa memicu risiko lain seperti longsor atau erosi yang berdampak tidak langsung pada banjir.



Gambar 5. Distribusi Kecamatan

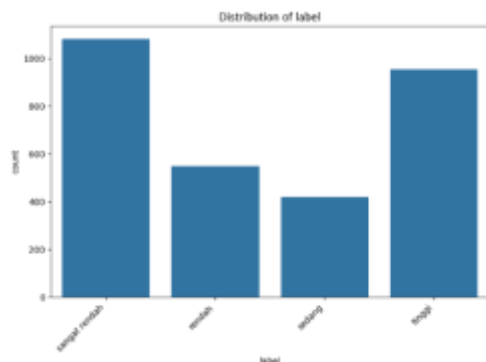
Pada grafik pertama yang menunjukkan distribusi kecamatan, terlihat bahwa jumlah data tersebar tidak merata di antara lima kecamatan yang diteliti. Kecamatan Puuwatu memiliki jumlah data terbanyak, disusul oleh Kendari dan Kendari Barat. Sementara itu, Mandonga memiliki jumlah data paling sedikit dibandingkan kecamatan lainnya. Ketimpangan jumlah ini menunjukkan bahwa data lebih banyak dikumpulkan dari kecamatan tertentu, sehingga analisis yang dilakukan perlu mempertimbangkan ketidakseimbangan tersebut agar tidak menimbulkan bias wilayah.



Gambar 6. Distribusi Riwayat Banjir

Grafik kedua memperlihatkan distribusi riwayat banjir (flood_history). Mayoritas data berasal dari daerah yang tidak pernah mengalami banjir, sedangkan sisanya berasal dari daerah yang pernah terdampak banjir. Ketimpangan ini menunjukkan bahwa persebaran data cenderung lebih besar di daerah bebas banjir, sehingga

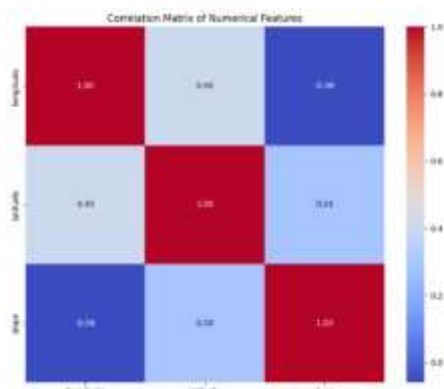
analisis yang melibatkan variabel ini juga harus memperhatikan kemungkinan adanya ketidakseimbangan kelas. Dalam konteks klasifikasi, ketimpangan ini dapat memengaruhi akurasi model jika tidak ditangani dengan benar.



Gambar 7. Klasifikasi Tingkat Banjir

Sementara itu, grafik ketiga menunjukkan distribusi label risiko, yang tampaknya mengindikasikan tingkat kerentanan atau klasifikasi tertentu. Kategori “sangat rendah” dan “tinggi” mendominasi data, sementara kategori “rendah” dan “sedang” memiliki jumlah data yang jauh lebih sedikit. Ketidakseimbangan dalam distribusi label ini menunjukkan bahwa sebagian besar wilayah dinilai sangat rendah atau tinggi tingkat risikonya, namun kategori menengah kurang terwakili secara proporsional. Hal ini penting untuk diperhatikan, khususnya dalam proses pelatihan model prediktif, karena model cenderung kesulitan dalam mengenali kelas minoritas jika tidak dilakukan penyesuaian seperti penyeimbangan data atau penggunaan metrik evaluasi khusus.

Hasil analisis bivariate divisualisasikan sebagai berikut.

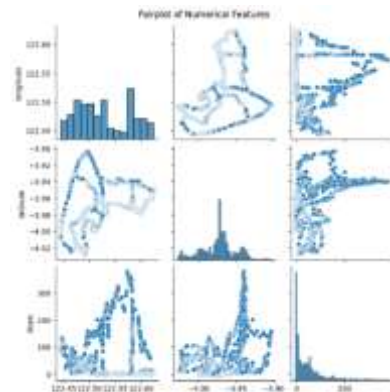


Gambar 8. Matriks Korelasi

Hasil visualisasi matriks korelasi menunjukkan hubungan antara variabel numerik longitude, latitude, dan slope. Nilai korelasi antara longitude dan latitude adalah sebesar 0,40, yang mengindikasikan adanya hubungan positif dengan kekuatan sedang. Artinya, saat nilai longitude meningkat, latitude cenderung ikut meningkat. Korelasi antara latitude dan slope bernilai 0,28, yang berarti hubungan antara keduanya bersifat lemah namun searah. Sementara itu, korelasi antara longitude dan slope sangat rendah, yaitu sebesar -0,06, yang menunjukkan hampir tidak ada hubungan linear antara keduanya. Secara keseluruhan, nilai korelasi yang rendah ini menunjukkan bahwa tidak terdapat masalah multikolinearitas, sehingga

semua variable dapat digunakan secara bersamaan dalam model tanpa memengaruhi keakuratan hasil.

Hasil analisis multivariate divisualisasikan sebagai berikut.



Gambar 9. Pairplot

Visualisasi pairplot memperlihatkan sebaran data dan pola hubungan antar variabel longitude, latitude, dan slope. Distribusi longitude dan latitude tampak membentuk pola menyebar yang menyerupai bentuk wilayah geografis, yang menandakan bahwa data bersifat spasial. Variabel slope menunjukkan distribusi yang tidak simetris dan condong ke kanan, yang berarti terdapat banyak data dengan nilai kecil, tetapi juga terdapat beberapa nilai ekstrem (outlier) dengan angka kemiringan yang tinggi. Pola sebaran antar variabel tidak menunjukkan hubungan linear yang kuat, tetapi tetap memperlihatkan keterkaitan struktur spasial yang menarik. Hasil ini mendukung temuan pada matriks korelasi bahwa hubungan antar variabel bersifat lemah hingga sedang, dan tidak ada pola linear yang dominan.

Hasil klasifikasi dan akurasi dari model yang telah ditraining ditunjukkan sebagai berikut.

Akurasi: 0.9916805324459235

Classification Report:				
	precision	recall	f1-score	support
1	1.00	1.00	1.00	215
2	1.00	1.00	1.00	111
3	0.97	0.96	0.97	61
4	0.98	0.99	0.99	194
accuracy			0.99	601
macro avg	0.99	0.99	0.99	601
weighted avg	0.99	0.99	0.99	601

Gambar 10. Evaluasi Model

Evaluasi model klasifikasi menunjukkan hasil yang sangat memuaskan. Akurasi model mencapai 99,16 persen, yang berarti bahwa hampir seluruh data diklasifikasikan dengan benar. Untuk kelas 1 dan kelas 2, precision, recall, dan f1-score masing-masing bernilai 1,00. Ini menunjukkan bahwa model berhasil mengenali semua sampel di kedua kelas tersebut tanpa kesalahan. Kelas 3 memiliki precision sebesar 0,97 dan recall sebesar 0,96, dengan f1-score 0,97, yang tetap menunjukkan performa sangat baik. Sementara itu, kelas 4 memiliki precision sebesar 0,98 dan recall sebesar 0,99, menghasilkan f1-score 0,99. Nilai rata-rata makro dan rata-rata tertimbang untuk precision, recall, dan f1-score adalah sebesar 0,99, yang mencerminkan bahwa model memiliki kinerja tinggi secara konsisten pada seluruh kelas. Dengan demikian, model

yang digunakan sangat andal dan efektif dalam melakukan klasifikasi terhadap data.

E. Visualisasi Hasil Prediksi dan Implementasi Sistem Informasi Banjir



Gambar 11. Tampilan Website

Sebagai bentuk konkret dari hasil pemodelan dan klasifikasi potensi banjir yang telah dilakukan, penelitian ini diakhiri dengan tahapan implementasi dalam bentuk sistem informasi berbasis website. Tujuan utama dari pengembangan sistem ini adalah untuk menyajikan hasil prediksi secara interaktif dan mudah dipahami oleh pengguna akhir, baik dari kalangan pemerintah daerah, lembaga penanggulangan bencana, maupun masyarakat umum. Sistem ini dikembangkan untuk menjembatani antara keluaran model berbasis pembelajaran mesin dengan kebutuhan praktis di lapangan, yakni pemetaan wilayah berdasarkan tingkat potensi banjir secara cepat dan akurat. Website ini diberi nama SIBAKAR (Sistem Informasi Banjir Kota Kendari), yang berfungsi sebagai media visualisasi sekaligus alat bantu pengambilan keputusan mitigatif dan preventif terhadap ancaman banjir.

Tampilan antarmuka sistem dibangun secara sederhana namun fungsional, dengan memfokuskan pada kemudahan penggunaan dan kejelasan informasi. Berdasarkan gambar yang telah ditampilkan sebelumnya, pada halaman utama website, pengguna dihadapkan dengan sebuah form isian yang terdiri dari enam parameter utama yang sesuai dengan fitur dalam pemodelan: Latitude, Longitude, Kemiringan Lereng (Slope), Jenis Tanah (Geologi), Riwayat Banjir, dan Nama Kecamatan. Input ini merupakan representasi spasial dan lingkungan dari titik lokasi yang akan dianalisis. Semua parameter ini telah disesuaikan dengan struktur fitur yang digunakan dalam proses training model stacking ensemble, sehingga hasil prediksi yang dihasilkan tetap berada dalam koridor konsistensi dan akurasi model.

Setelah pengguna mengisi seluruh parameter tersebut, sistem akan memproses data menggunakan model stacking ensemble yang telah dilatih sebelumnya dengan integrasi tiga algoritma terbaik, yaitu Random Forest, XGBoost, dan CatBoost, yang disatukan melalui Logistic Regression sebagai meta learner. Output dari sistem ini tidak hanya berupa klasifikasi numerik atau label semata, tetapi juga dilengkapi dengan interpretasi deskriptif yang menyatakan tingkat potensi banjir dalam kategori sangat rendah, rendah, sedang, atau tinggi. Misalnya, pada contoh tampilan yang ditampilkan dalam gambar, untuk lokasi dengan koordinat latitude -3.949740182 dan longitude 122.6243131, dengan jenis tanah eluvium dan riwayat banjir “pernah banjir”, sistem memberikan hasil prediksi berupa potensi banjir

tinggi. Sementara itu, pada titik lain dengan parameter yang sedikit berbeda namun memiliki kemiringan lebih tinggi, sistem memunculkan output potensi banjir sedang. Hal ini menunjukkan bahwa sistem tidak hanya melakukan klasifikasi sederhana, tetapi mempertimbangkan interaksi kompleks antar fitur spasial dan historis dalam memutuskan kelas risiko banjir.



Gambar 12. Hasil Prediksi

Visualisasi hasil prediksi semakin diperkuat dengan tampilan peta digital interaktif yang terintegrasi ke dalam website, memanfaatkan koordinat latitude dan longitude yang dimasukkan pengguna. Titik lokasi akan langsung ditandai pada peta, sehingga pengguna dapat dengan mudah mengetahui secara spasial lokasi yang sedang dianalisis. Fitur ini sangat berguna untuk keperluan verifikasi lapangan dan observasi distribusi spasial dari risiko banjir, terutama untuk badan pemerintah yang melakukan pemetaan wilayah rawan bencana. Dengan adanya dukungan peta, sistem ini tidak hanya bersifat prediktif tetapi juga informatif dan geografis.

Lebih jauh, sistem ini memiliki potensi besar untuk diintegrasikan dengan sistem manajemen kebencanaan yang lebih besar, seperti sistem peringatan dini (*early warning system*), perencanaan tata ruang berbasis risiko bencana, maupun edukasi masyarakat melalui informasi berbasis data. Implementasi ini sekaligus menjadi bukti bahwa pendekatan pembelajaran mesin tidak berhenti pada tahap akurasi prediksi, tetapi dapat diterjemahkan ke dalam sistem nyata yang memiliki nilai guna tinggi di masyarakat. Selain itu, sistem ini dapat dikembangkan lebih lanjut dengan fitur unggah file data masif, visualisasi heatmap risiko banjir, dan pelaporan realtime oleh masyarakat.

Secara keseluruhan, pengembangan dan visualisasi hasil prediksi dalam bentuk sistem informasi banjir berbasis web merupakan bagian integral dari penelitian ini. Dengan menggabungkan kekuatan prediksi dari model stacking ensemble dan antarmuka pengguna yang ramah, sistem ini berhasil menjawab tantangan utama dalam dunia mitigasi bencana, yaitu menjembatani hasil analisis teknis yang kompleks dengan kebutuhan masyarakat yang praktis dan cepat. Pendekatan ini diharapkan dapat menjadi model pengembangan sistem sejenis di daerah lain, terutama wilayah-wilayah dengan kerentanan hidrometeorologi tinggi di Indonesia.

IV. KESIMPULAN

Berdasarkan hasil analisis spasial dan pemodelan menggunakan pendekatan stacking ensemble dengan algoritma Random Forest, XGBoost, CatBoost, dan Logistic Regression sebagai meta learner, penelitian ini berhasil mengidentifikasi potensi banjir di Kecamatan Kota Kendari secara akurat dan sistematis dengan akurasi 99%.

Faktor utama yang berpengaruh terhadap klasifikasi risiko banjir adalah kemiringan lereng, jenis tanah, dan riwayat banjir. Visualisasi hasil prediksi yang diimplementasikan melalui sistem informasi berbasis website menunjukkan bahwa model mampu memberikan informasi prediktif yang relevan dan aplikatif. Dengan demikian, sistem ini berpotensi menjadi alat pendukung pengambilan keputusan dalam mitigasi bencana dan perencanaan tata ruang berbasis risiko banjir.

UCAPAN TERIMA KASIH

Ucapan terima kasih saya sampaikan kepada semua pihak yang telah memberikan dukungan dan kontribusi selama proses penyusunan penelitian ini. Terutama kepada Badan Penanggulangan Bencana Daerah (BPBD) Kota Kendari yang telah memberikan data historis kejadian banjir serta informasi geologi wilayah yang sangat membantu dalam analisis spasial. Saya juga mengucapkan terima kasih kepada tim pengelola portal tanahair.indonesia.go.id yang menyediakan data spasial penting seperti koordinat, kemiringan lereng, dan peta dasar. Tak lupa, apresiasi saya sampaikan kepada Rizal Adi Saputra, S.T., M.Kom. dan rekan-rekan yang senantiasa memberikan masukan, semangat, serta bantuan teknis selama proses pengolahan data, pembangunan model, hingga pengembangan sistem informasi. Tanpa dukungan dan kerja sama dari berbagai pihak, penelitian ini tidak akan terselesaikan dengan baik..

DAFTAR PUSTAKA

- [1] A. Iskandar, "ANALISIS DAMPAK BENCANA BANJIR TERHADAP EKSISTENSI KEHIDUPAN SOSIAL EKONOMI MASYARAKAT," *Jurnal Sosial Humaniora*, pp. 71-88, 2024.
- [2] Hasddin, "STUDI KARAKTERISTIK DAN WILAYAH TERDAMPAK BANJIR DI KECAMATAN MANDONGA, KOTA KENDARI," *Jurnal Pembangunan Wilayah dan Kota*, pp. 420-427, 2021.
- [3] Nursiani, "PEMETAAN JENIS TANAH DI KELURAHAN ANDUONOHU KECAMATAN POASIA," *Jurnal Penelitian Pendidikan Geografi*, pp. 55-70, 2019.
- [4] E. H. Dewantari, "Analisis Laju Erosi Berdasarkan Kemiringan Lereng Pada Area Pasca Spreading PT Angsana Jaya Energi, Tanah Bumbu, Kalimantan Selatan," *Jurnal Lingkungan Kebumian Indonesia*, pp. 1-12, 2023.
- [5] W. B. Kusuma, "Karakteristik Sifat Fisik dan Daya Dukung Tanah Endapan Aluvium Daerah Caruban," *Jurnal Nasional Pengelolaan Energi*, pp. 15-28, 2019.
- [6] B. Sunarko, "Penerapan Stacking Ensemble Learning untuk Klasifikasi Efek Kesehatan Akibat Pencemaran Udara," *Edu Komputika Journal*, pp. 56-64, 2023.
- [7] M. Fadli, "KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE," *Jurnal FT UMT*, pp. 72-81, 2023.
- [8] S. Amaliah, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, pp. 121-127, 2022.
- [9] I. G. J. Kurniarwan, "PENERAPAN MACHINE LEARNING EXTREME GRADIENT BOOSTING DALAM KLASIFIKASI POTENSI TSUNAMI BERDASARKAN DATA GEMPA BUMI," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 1-18, 2025.
- [10] J. Christian, "Implementasi Penggunaan Algoritma Categorical Boosting (Catboost) Dengan Optimisasi Hiperparameter Dalam Memprediksi Pembatalan Pesanan Kamar Hotel," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 814-825, 2022.