ISSN: 2477-0078

Analisis Penggunaan Model Random Forest dalam Memprediksi Resiko Banjir di Daerah Rawan Bencana Kabupaten Pamekasan

Moh. Fauzi[1], Muhsi[2], Hozairi[3]
[123]Sistem Infirmasi,Fakultas Teknik,Universitas Islam Madura
[1]masuzy33@gmail.com, [2] muhsi@uim.ac.id [,3] dr.hozairi@mail.com

ABSTRAK

Penelitian ini bertujuan untuk menganalisis efektivitas model Random Forest dalam memprediksi risiko banjir di daerah rawan bencana di Kabupaten Pamekasan. Dataset yang digunakan berasal dari BMKG dan BPBD Pamekasan, mencakup data historis terkait kejadian banjir, curah hujan, lokasi, waktu kejadian, dan musim. Proses pemodelan dimulai dari preprocessing data, pelatihan model menggunakan metode Repeated Stratified K-Fold Cross Validation, hingga optimasi model dengan hyperparameter tuning. Hasil evaluasi menunjukkan bahwa model mampu memprediksi risiko banjir dengan akurasi 82% pada data pelatihan dan 95% pada data pengujian setelah dilakukan tuning. Nilai precision dan recall yang tinggi pada kedua kelas mengindikasikan kemampuan model dalam menggeneralisasi pola dari data historis. Penelitian ini membuktikan bahwa model Random Forest dapat digunakan sebagai alat prediksi yang efektif dalam sistem peringatan dini bencana banjir berbasis data historis.
Kata Kunci: Random Forest, Prediksi Banjir, Machine Learning, Kabupaten Pamekasan, Hyperparameter Tuning

ABSTRACT

This study aims to analyze the effectiveness of the Random Forest model in predicting flood risk in disaster-prone areas of Pamekasan Regency. The dataset used is obtained from BMKG and BPBD Pamekasan, consisting of historical data related to flood events, rainfall intensity, location, time, and season. The modeling process includes data preprocessing, model training using Repeated Stratified K-Fold Cross Validation, and model optimization through hyperparameter tuning. Evaluation results show that the model can predict flood risk with 82% accuracy on training data and 95% accuracy on testing data after tuning. The high precision and recall scores for both classes indicate the model's capability to generalize patterns from historical data. This research proves that the Random Forest model can serve as an effective prediction tool in early warning systems for flood disasters based on historical data.
Keywords: *Random Forest, Flood Prediction, Machine Learning, Pamekasan Regency, Hyperparameter Tuning*

1. PENDAHULUAN

Bencana merupakan peristiwa yang mengancam kehidupan masyarakat akibat faktor alam, non-alam, atau manusia, yang dapat menimbulkan korban jiwa, kerusakan lingkungan, dan kerugian materiil (Perka BNPB No. 02 Tahun 2012). Salah satu bencana yang paling sering terjadi di Indonesia adalah banjir, yang menurut Menurut Schwab at.al (1981) dalam Somantri (2008) banjir adalah luapan atau genangan dari sungai atau badan air lainnya yang disebabkan oleh curah hujan yang berlebihan atau salju yang mencair atau dapat pula karena gelombang pasang yang membanjiri kebanyakan pada dataran banjir.[1]

Indonesia memiliki potensi banjir yang tinggi karena kondisi topografi berupa dataran rendah dan cekungan serta sebagian besar wilayahnya merupakan

ISSN: 2477-0078

lautan, sehingga curah hujan di daerah hulu dapat menyebabkan banjir di wilayah hilir [2]. Banjir menjadi bencana alam paling dominan, menyumbang dua pertiga dari total kerugian bencana di Indonesia [3]. Di Kabupaten Pamekasan sendiri, banjir terjadi setiap tahun akibat luapan sungai yang tidak mampu menampung debit air kiriman dari wilayah utara, sehingga aliran air yang mengalir ke wilayah kota mengalami akumulasi berlebih [4].

Kabupaten Pamekasan memiliki beberapa daerah yang dikenal rawan banjir, terutama saat musim hujan tiba. Namun, sejauh ini, sistem prediksi banjir yang ada masih bersifat konvensional dan belum banyak memanfaatkan teknologi kecerdasan buatan atau *machine learning*. Oleh karena itu, penelitian ini bertujuan untuk menganalisis efektivitas model Random Forest dalam memprediksi risiko banjir berdasarkan data historis dan faktor-faktor lingkungan di daerah rawan bencana di Kabupaten Pamekasan.

Berbagai penelitian sebelumnya telah menggunakan berbagai metode untuk memprediksi resiko banjir. Salah satunya [4] Penelitian yang dilakukan bertujuan untuk merancang model prediksi banjir di Kota Pontianak menggunakan metode decision tree C4.5. Proses penelitian mencakup identifikasi masalah, studi pustaka, pengumpulan dan pengolahan data, pengujian model, hingga penarikan kesimpulan. Tiga percobaan dilakukan dengan perlakuan berbeda dan evaluasi kinerja menggunakan confusion matrix. Hasil terbaik diperoleh pada percobaan kedua dengan akurasi 98%, presisi 72%, recall 76%, dan F1-score sebesar 74% pada data uji. Pada era revolusi industri 4.0, kemajuan teknologi mendorong ketergantungan manusia pada sistem informasi yang terhubung melalui internet dan kecerdasan buatan, sehingga keamanan siber menjadi aspek krusial di berbagai bidang, termasuk dalam deteksi intrusi [5]. Salah satu pendekatan yang digunakan adalah ensemble learning, yaitu metode penggabungan beberapa algoritma machine learning untuk meningkatkan akurasi prediksi [6]. Ensemble learning terdiri dari tiga jenis utama: *bagging, boosting, dan stacking, di mana stacking* unggul dalam menggabungkan model heterogen, sedangkan bagging dan boosting bekerja pada model homogen[7]. Random Forest merupakan metode ensemble berbasis bagging yang dikembangkan dari algoritma *Classification and Regression Tree* (CART), dengan teknik random feature selection dan kumpulan *decision tree* untuk klasifikasi atau prediksi data [8].

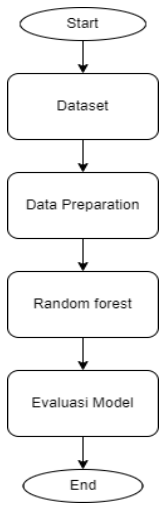
Melalui penelitian ini, diharapkan dapat diperoleh model prediksi yang mampu memberikan informasi risiko banjir secara lebih cepat dan akurat, sehingga dapat membantu pihak berwenang dalam mengambil langkah mitigasi yang tepat waktu.

2. METODOLOGI PENELITIAN

Metodologi penelitian ini menjelaskan tahapan-tahapan yang dilakukan dalam membangun model prediksi risiko banjir di Kabupaten Pamekasan menggunakan algoritma Random Forest. Pada bagian ini dijabarkan alat dan data yang digunakan, prosedur penelitian, teknik pengolahan data, serta metode evaluasi model. Penjelasan ini bertujuan untuk memberikan gambaran yang jelas mengenai pendekatan yang diterapkan, sehingga proses penelitian dapat dipahami dan direplikasi oleh peneliti lain di masa mendatang.

Prosedur pelaksanaan dalam penelitian ini mencakup serangkaian langkah sistematis, mulai dari pengumpulan data historis banjir hingga evaluasi performa model prediksi. Alur kerja penelitian ini digambarkan dalam bentuk flowchart seperti yang ditampilkan pada Gambar 1. Setiap tahapan dalam Gambar 1 akan dijelaskan secara rinci pada bagian-bagian berikutnya.

ISSN: 2477-0078



Gambar 1. Flowchart Diagram Untuk Metode Yang Diusulkan

2.1 Dataset

Teknik pengambilan data dalam penelitian ini menggunakan metode dokumentasi dengan mengumpulkan data sekunder dari instansi resmi, yaitu Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Pamekasan. Data diperoleh melalui situs resmi kedua instansi tersebut, yang mencatat secara sistematis informasi historis terkait kejadian banjir dan kondisi cuaca di wilayah Pamekasan. Penggunaan data sekunder ini dianggap valid karena berasal dari sumber terpercaya dengan pencatatan yang rutin dan akurat.

Jenis data yang digunakan berupa catatan historis kejadian banjir dan curah hujan yang meliputi beberapa atribut penting, yakni lokasi kejadian, bulan dan tahun peristiwa, besaran curah hujan dalam milimeter, serta musim pada saat kejadian. Variabel target dalam dataset ini adalah status banjir yang dibagi menjadi dua kategori, yaitu terjadi banjir dan tidak terjadi banjir. Dataset akhir terdiri dari 15 catatan, di mana 10 dataset mencerminkan peristiwa banjir dan 5 sisanya menunjukkan tidak adanya banjir, dengan setiap catatan mewakili satu kejadian pada waktu dan tempat tertentu

2.2 Data Preperation

Data preparation merujuk pada serangkaian langkah dan proses yang dilakukan untuk membersihkan, mengorganisir, dan mengubah data mentah menjadi bentuk yang lebih sesuai untuk analisis atau pemodelan. Ini melibatkan tindakan seperti penanganan nilai yang hilang, normalisasi, transformasi fitur, dan pemilihan atribut, sehingga data menjadi siap untuk dieksplorasi atau digunakan dalam pembuatan model[9]. Proses data preparation mendukung pengambilan keputusan yang akurat dan memastikan bahwa data yang digunakan dalam analisis atau pembelajaran mesin memiliki kualitas yang memadai.

2.3 Random Forest

Random Forest adalah suatu algoritma ensemble dalam machine learning yang menggabungkan hasil prediksi dari beberapa Decision Trees yang dibangun secara acak. Dengan memanfaatkan keberagaman dan kombinasi model-model ini, Random Forest dapat meningkatkan akurasi prediksi dan memiliki kemampuan yang baik untuk menangani overfitting pada data yang kompleks [10]. Secara umum, Random Forest bekerja dengan membangun sejumlah pohon keputusan (decision trees) yang dilatih pada subset data yang berbeda secara acak. Prediksi akhir untuk masalah klasifikasi diperoleh dengan menggunakan metode voting mayoritas dari semua pohon yang ada, sedangkan untuk regresi menggunakan nilai rata-rata. Formula prediksi pada Random Forest untuk klasifikasi dapat dinyatakan sebagai:

$$\hat{y} = \text{mode}(\{h_b(x)\}_{b=1}^{B}) \qquad (1)$$

Dimana:
B adalah jumlah pohon,
$h_b(x)$ adalah prediksi dari pohon ke-b, dan
$\hat{y}$ adalah hasil prediksi akhir.

ISSN: 2477-0078

2.4 Evaluasi

Confusion matrix adalah suatu metrik evaluasi dalam konteks klasifikasi pada machine learning yang menyajikan performa model dengan merinci hasil prediksi untuk setiap kelas target. Matrix ini memberikan informasi tentang jumlah true positive, true negative, false positive, dan false negative, memungkinkan analisis yang lebih mendalam terhadap sejauh mana model dapat membedakan kelas target dengan akurasi yang sesuai(I. Düntsch and G. Gediga 2020). *Confusion matrix* menjadi instrumen penting dalam mengevaluasi keandalan model dan mengidentifikasi area di mana model dapat ditingkatkan untuk meningkatkan keakuratannya.

Hasilnya akan ditampilkan dalam bentuk hasil prediksi dari akurasi yang telah ditentukan dari data latih dan data uji. Semua metrik tersebut dihitung dengan rumus di bawah ini:

. $$MSE = \sum_{i=1}^{n}(fi - yi)^2 \quad (2)$$

Keterangan:

1. N adalah jumlah data
2. Fi adalah nilai yang dikembalikan oleh model
3. Yi adalah nilai aktual untuk dala i

$$accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

$$precision = TP/(TP + FP) \quad (4)$$

$$recall = TP/(TP + FN) \quad (5)$$

$$F1\ Score = \frac{2 \text{ x (recall x precision)}}{(\text{recall+precision})} \quad (6)$$

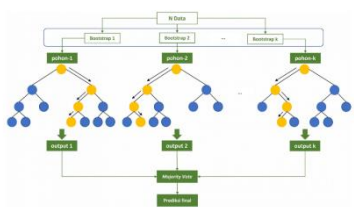
Dengan TP adalah True Positive, jumlah data positif yang diprediksi positif, TN adalah True Negative, jumlah data negatif yang diprediksi negatif, FP adalah False Positive, jumlah data positif yang diprediksi negatif, dan FN adalah False Negative, jumlah data negatif yang diprediksi positif

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk memprediksi risiko banjir di daerah rawan bencana Kabupaten Pamekasan dengan menggunakan model Random Forest. Proses penelitian dilakukan melalui beberapa tahapan, dimulai dari pengumpulan data sekunder, preprocessing data, pembagian data latih dan data uji, pelatihan model, pengujian, hingga evaluasi hasil untuk menilai performa model dalam melakukan klasifikasi risiko banjir secara akurat.

3.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset banjir Kabupaten Pamekasan 2025 yang merupakan data sekunder hasil pengumpulan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Pamekasan. Dataset ini terdiri dari 34 data historis kejadian banjir, dengan variabel-variabel seperti lokasi, bulan, tahun, curah hujan, musim, dan status banjir sebagai label klasifikasi yang dapat diakses melalui link berikut ini, https://www.bmkg.go.id , https://bpbd.pamekasankab.go.id. Visualisasi hasil dari dataset yang telah melalui tahap pemrosesan ditampilkan pada Gambar 2.



Gambar 2. Pohon keputusan random forest

Table 1 Data set

ISSN: 2477-0078

	Lokasi	..	..	..	..	Musim
0	7	..	..	..	..	0
1	9	..	..	..	..	0
2	19	..	..	..		1
3	0	..	..	..	..	2
4	5	..	..	..	..	0
…	..	..	..	..	..	….
….	..	..	..	..	..	….
….	..	..	..	..	..	….
14	3	1	2025	90	0	0

3.2 Data Preperation

Pada tahap ini, dilakukan proses persiapan data sebelum pemodelan, yang mencakup pembersihan, transformasi variabel, penambahan fitur, serta pemisahan data untuk pelatihan dan evaluasi model. Berikut adalah tahapan detailnya:

A. Pembersihan dan Standarisasi Data
Data dibaca menggunakan pustaka pandas, kemudian nama kolom dinormalisasi menjadi huruf kecil dan tanpa spasi. Selain itu, variabel kategorikal seperti lokasi dan bulan diubah menjadi format numerik menggunakan teknik label encoding. Ini penting agar model Random Forest dapat memproses variabel tersebut dengan benar.

B. Penambahan Fitur Musim
Fitur baru bernama musim ditambahkan berdasarkan informasi bulan. Kategorisasi musim terdiri dari tiga jenis, yaitu musim hujan, musim pancaroba, dan musim kemarau. Setelah diklasifikasikan, data musim diubah menjadi bentuk numerik.

C. Pemisahan Fitur dan Label
Setelah proses transformasi selesai, data dipisahkan menjadi fitur (X) dan label (y). Fitur yang digunakan meliputi lokasi, bulan, tahun, curah hujan, dan musim. Sedangkan label targetnya adalah status banjir (0 = tidak banjir, 1 = banjir).

D. Pemisahan Data dan Validasi
Data dievaluasi menggunakan metode Repeated Stratified K-Fold Cross Validation untuk menjaga distribusi kelas yang seimbang di setiap fold. Model Random Forest juga dikonfigurasi dengan jumlah estimator 300 dan pembobotan kelas otomatis.

3.3 Pelatihan Model Random Forest

Model Random Forest dilatih menggunakan data yang telah melalui tahap preprocessing sebelumnya, termasuk pembersihan data, transformasi variabel kategorikal, dan penambahan fitur musim. Model ini menggunakan 300 pohon keputusan *(estimators)* dengan parameter class_weight='balanced' untuk mengatasi potensi ketidakseimbangan kelas antara data banjir dan tidak banjir. Proses pelatihan dilakukan dengan memanfaatkan metode Repeated Stratified K-Fold Cross Validation, yaitu pembagian data ke dalam 25
fold yang diulang sebanyak 25 kali untuk memastikan hasil evaluasi yang stabil dan tidak bias terhadap distribusi kelas. Setiap iterasi pelatihan menghasilkan model yang diuji terhadap subset data yang berbeda, sehingga meminimalkan risiko overfitting dan meningkatkan generalisasi model. Pemilihan parameter seperti jumlah pohon, metode pembobotan kelas, dan jumlah fold ditentukan secara eksperimental melalui beberapa uji coba awal. Kombinasi tersebut terbukti memberikan performa yang optimal dalam klasifikasi risiko banjir berdasarkan metrik evaluasi seperti akurasi, precision, recall, dan f1-score. Dan untuk hasil rata-rata Kurasi dari 25 fold yang diulang sebanyak 25 kali adalah 84.21%

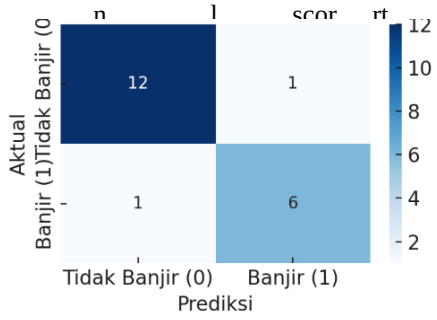
3.4 Evaluasi model

Hasil akurasi data training dan testing yang diperoleh dari penggunaan model Random Forest dengan struktur lapisan konvolusi yang telah ditentukan disajikan pada Tabel 2 dan Tabel 3.

ISSN: 2477-0078

	precision	recall	f1-score	Support
0	0.93	1.00	0.96	13
1	1.00	0.86	0.92	7
Accuracy			0.95	20
macro avg	0.96	0.93	0.94	20
weighted avg	0.95	0.95	0.95	20

	precision	recall	f1-scor	Suppo
0				
1				
Ac				
y				
ma				
av				
we				
d avg				

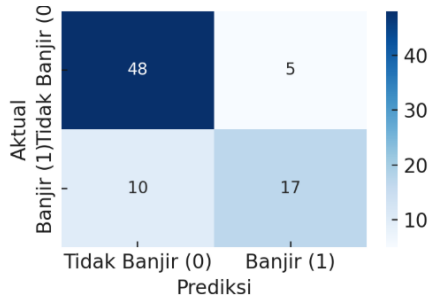


Gambar 4. Confusion Matrix – Data Testing

Selain itu, digunakan pula *confusion matrix* dan *classification report* untuk memberikan gambaran detail performa model dalam mengklasifikasikan data. Hasil evaluasi model juga ditunjukkan melalui classification report, yang memuat nilai precision, recall, f1-score, dan support untuk masing-masing kelas, seperti disajikan pada Tabel 2.

3.4.1 Sebelum Hyperprarameter Tuning

Table 2 Hasil Prediksi Pada Data Train



Gambar 3. Confusion Matrix-Data Training

Table 3 Hasil Prediksi Pada Data Testing

	precision	recall	f1-score	Support
0	0.83	0.91	0.86	53
1	0.77	0.63	0.69	27
Accuracy			0.81	80
macro avg	0.80	0.77	0.78	80
weighted avg	0.81	0.81	0.81	80

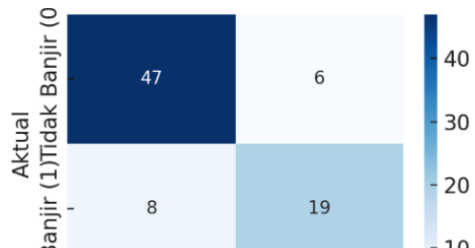
Analisis:

Pada data pelatihan, model memiliki akurasi sebesar 81% dengan performa baik pada kelas tidak banjir (0) (precision 0.83, recall 0.91) namun sedikit menurun pada kelas banjir (1) (precision 0.77, recall 0.63). Kesalahan klasifikasi terutama berasal dari kelas banjir yang diprediksi sebagai tidak banjir.

Pada data pengujian, akurasi meningkat menjadi **90%** dengan distribusi kesalahan minimal, hanya dua data yang salah klasifikasi. Nilai *macro average* dan *weighted average* yang konsisten menunjukkan kemampuan generalisasi yang baik terhadap data baru.

3.4.2 Setelah Hyperprarameter Tuning

Table 4. *Hasil Prediksi Pada Data Train*

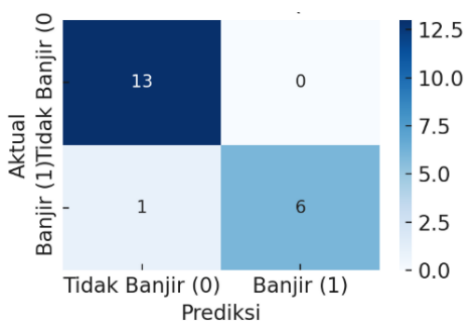


ISSN: 2477-0078

Gambar 5. Confusion Matrix Pada Data Train

Table 5. Hasil Prediksi Pada Data Testing

	precision	recall	f1-score	Support
0	0.92	0.92	0.92	13
1	0.86	0.86	0.86	7
Accuracy			0.90	20
macro avg	0.89	0.89	0.89	20
weighted avg	0.90	0.90	0.90	20



Gambar 6. Confusion Matrix Pada Data Testing

Setelah dilakukan hyperparameter tuning, akurasi model pada data pelatihan meningkat menjadi 82%, dan pada data pengujian meningkat signifikan menjadi 95%. Confusion matrix pada data uji menunjukkan bahwa model mampu mengenali seluruh data kelas tidak banjir dengan benar (recall = 1.00), serta hanya salah memprediksi 1 data kelas banjir. Performa ini tercermin pula pada classification report, di mana nilai *f1-score* untuk kedua kelas mendekati 1, menunjukkan keseimbangan yang baik antara precision dan recall.Sementara itu, hasil evaluasi pada data uji (*Tabel 5*) menunjukkan peningkatan yang signifikan, dengan akurasi mencapai 95%. Model berhasil mengklasifikasikan kelas tidak banjir dengan f1-score sebesar 0.96, dan kelas banjir dengan f1-score 0.92. Nilai *macro average* dan *weighted average* untuk f1-score yang sama-sama mencapai 0.94 dan 0.95 mengindikasikan bahwa model telah mampu menggeneralisasi dengan sangat baik pada data yang belum pernah dilihat sebelumnya.

Hyperparameter tuning memberikan peningkatan nyata terhadap kemampuan generalisasi model. Tingginya nilai f1-score pada kedua kelas, serta rendahnya jumlah kesalahan klasifikasi pada confusion matrix, menunjukkan bahwa model Random Forest sangat efektif dalam memprediksi risiko banjir berdasarkan variabel masukan seperti lokasi, bulan, tahun, curah hujan, dan musim.

4. KESIMPULAN

Penelitian ini membuktikan bahwa model Random Forest efektif dalam memprediksi risiko banjir di Kabupaten Pamekasan. Dengan memanfaatkan data historis yang mencakup variabel lokasi, bulan, tahun, curah hujan, dan musim, model mampu menghasilkan prediksi yang akurat baik pada data pelatihan maupun data pengujian. Sebelum dilakukan hyperparameter tuning, akurasi model pada data training adalah 81% dan meningkat menjadi 82% setelah tuning, sedangkan akurasi pada data testing meningkat signifikan dari 90% menjadi 95%. Selain itu, nilai f1-score pada kedua kelas (banjir dan tidak banjir) juga menunjukkan peningkatan, di mana model

ISSN: 2477-0078

pasca-tuning mampu mencapai f1-score 0.96 untuk kelas tidak banjir dan 0.92 untuk kelas banjir pada data uji.

Hasil confusion matrix mengindikasikan bahwa jumlah kesalahan klasifikasi berkurang secara signifikan setelah tuning, dengan seluruh data kelas tidak banjir pada data uji berhasil diklasifikasikan dengan benar (recall = 1.00), dan hanya satu data kelas banjir yang salah klasifikasi. Peningkatan ini menunjukkan bahwa optimasi parameter berperan penting dalam meningkatkan kemampuan generalisasi model, sehingga Random Forest dengan konfigurasi optimal sangat direkomendasikan sebagai alat bantu prediksi dalam upaya mitigasi bencana banjir, khususnya di wilayah rawan seperti Kabupaten Pamekasan.

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur, penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bantuan, dan motivasi selama proses penyusunan artikel ini.

Ucapan terima kasih secara khusus penulis sampaikan kepada:

1. Orang tua tercinta, atas doa, kasih sayang, dan dukungan yang tiada henti.
2. Dosen pembimbing, yang telah memberikan bimbingan, arahan, dan masukan berharga dalam penyusunan artikel ini.
3. Teman-teman dan rekan seperjuangan**,** atas semangat dan kebersamaan selama proses penelitian ini berlangsung.
4. Semua pihak yang tidak dapat disebutkan satu per satu, yang turut berkontribusi dalam menyelesaikan artikel ini.

Semoga segala bantuan dan kebaikan yang telah diberikan mendapatkan balasan yang setimpal dari Allah SWT.

DAFTAR PUSTAKA

[1] Somantri, Lili. 2008. Pemanfaatan Teknik Penginderaan Jauh untuk Mengidentifikasi Kerentanan dan Risiko Banjir. Jurnal Gea, Jurusan Pendidikan Geografi, 8 (2). Diperoleh pada 9 Desember 2016.

[2] Suprapto. 2011. Statistik Pemodelan Bencana Banjir Indonesia (Kejadian 2002- 2010). Jurnal Penanggulangan Bencana. 2 (2).Abdullah, A., & Koma, I. (2025). Prediksi Banjir Di Kota Pontianak Menggunakan Metode Decision Tree C4 . 5. *Justek : Jurnal Sains Dan Teknologi*, *8*(1), 40–50.

[3] Dep PU. 1994. Teknologi Pengendalian Banjir di Indonesia. Direktorat Sungai, Ditjen Pengairan

[4] Anwari, A., & Makruf, M. (2019). Pemetaan Wilayah Rawan Bahaya Banjir Di Kabupaten Pamekasan Berbasis Sistem Informasi Geografis (Sig). *Network Engineering Research Operation*, *4*(2), 117–123. https://doi.org/10.21107/nero.v4i2.127

[5] Prasetyo, B., & Trisyanti, U. (2018). Revolusi Industri 4.0 dan Tantangan Perubahan Sosial. Prosiding SEMATEKSOS 3, 22–27.

[6] Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classiffier pada Data Set Penyakit Jantung," Indonesian Journal of Data and Science (IJODAS), vol. 1, no. 3, pp. 79–86, 2020.

[7] Jan Melvin Ayu Soraya Dachi, & Pardomuan Sitompul. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam, 2(2), 87–103. https://doi.org/10.55606/jurrimipa.v2i2.

ISSN: 2477-0078

1470

[8] Samudra, A. Y. (2019). Pendekatan Random Forest untuk Model Peramalan Harga Tembakau Rajangan Di Kabupaten Temanggung. Yogyakarta: Skripsi.

[9] C. Accinelli, S. Minisi, and B. Catania, “Coverage-based Rewriting for Data Preparation,” 2020

[10] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, “Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection,” IEEE Access, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.284198