

BAB II

LANDASAN TEORI

2.1 Landasan Teori

2.1.1 Bencana Banjir

Banjir adalah keadaan suatu wilayah yang tergenang air dalam skala besar, biasanya akibat curah hujan tinggi, lokasi cekungan, buruknya resapan air, pembangunan di bantaran sungai, aliran sungai terhambat sampah, serta kurangnya tutupan lahan di hulu. Banjir berdampak signifikan seperti korban jiwa, kerusakan properti, kerugian ekonomi, lingkungan, dan krisis air bersih. (Abrar et al., 2023) (dino, 2023).

2.1.2 Prediksi Banjir

Prediksi banjir adalah proses mengestimasi atau menggambarkan kejadian banjir yang akan terjadi di masa mendatang. Prediksi banjir banyak digunakan untuk mengingatkan masyarakat dan pemerintah tentang risiko dan dampak dari bencana banjir. (FEBRIANA, 2019)

2.1.3 Prediksi Iklim

Prediksi iklim adalah upaya memperkirakan kondisi atmosfer di masa depan menggunakan data suhu, kelembapan, tekanan, angin, dan curah hujan dari stasiun cuaca, satelit, dan radar. Prediksi ini penting untuk perencanaan aktivitas sehari-hari, peringatan dini cuaca ekstrem, serta menjaga keselamatan dan pengelolaan sumber daya alam. (Siregar, Tukino, Faisal, Fauzi, & kadori, 2020).

2.1.4 Elevasi

Elevasi atau ketinggian adalah ukuran titik tertinggi dan terendah suatu lokasi dari permukaan laut. Informasi ini penting dalam data geospasial untuk berbagai aplikasi, seperti pemetaan sungai, analisis daerah rawan longsor dan banjir, serta perencanaan jalan. Lahan yang lebih rendah memiliki risiko banjir lebih tinggi, sehingga klasifikasi elevasi dapat membantu menentukan kerentanan terhadap banjir. (Badan Informasi Geospasial, 2024).

2.1.5 Badan Penanggulangan Bencana Daerah (BPBD)

BPBD (Badan Penanggulangan Bencana Daerah) adalah lembaga pemerintah daerah di Indonesia yang bertanggung jawab untuk mengelola dan mengoordinasikan upaya penanggulangan bencana di tingkat provinsi dan kabupaten/kota. Tugas utama BPBD meliputi pencegahan, kesiapsiagaan, tanggap darurat, serta pemulihan pasca-bencana. BPBD bekerja sama dengan berbagai instansi pemerintah, masyarakat, dan organisasi terkait untuk mengurangi risiko bencana, memberikan bantuan saat terjadi bencana, dan memastikan pemulihan yang tepat setelah kejadian bencana (PORTAL RESMI BADAN PENANGGULANGAN BENCANA DAERAH, 2021).

2.1.6 Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)

BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) adalah lembaga pemerintah Indonesia yang bertanggung jawab untuk memberikan informasi terkait cuaca, iklim, dan fenomena geofisika seperti gempa bumi dan tsunami. Sebagai badan resmi negara, BMKG menjalankan berbagai

tugas penting, termasuk pengamatan kondisi atmosfer melalui jaringan stasiun cuaca, satelit, dan radar untuk memantau perubahan cuaca dan pola iklim di seluruh wilayah Indonesia. Data iklim merupakan salah satu komponen penting dalam penelitian yang berkaitan dengan bencana alam, terutama banjir. BMKG menyediakan data iklim yang lengkap melalui situs web resmi mereka, yang mencakup berbagai parameter iklim seperti suhu, curah hujan, kelembapan, dan angin. Pada data iklim ini mencakup informasi mengenai parameter dari data iklim seperti

1. Suhu Udara: Suhu Maksimum (T_x) dan Suhu Minimum (T_n) untuk memahami fluktuasi suhu harian.
2. Suhu Rata-Rata (T_{avg}) untuk memberikan gambaran umum kondisi cuaca.
3. Kelembapan Udara (Relative Humidity - RH_{avg}): Diukur dalam persen (%), kelembapan tinggi sering kali menjadi tanda kemungkinan hujan lebat.
4. Curah Hujan (Rainfall - RR): Mengukur jumlah hujan yang turun dalam satu periode waktu tertentu, biasanya diukur dalam milimeter (mm).
5. Durasi Penyinaran Matahari (Sunshine Duration - ss): Menunjukkan jumlah jam sinar matahari yang diterima, yang dapat mempengaruhi tingkat penguapan.
6. Kecepatan Angin (ff_x dan ff_{avg}): Mengukur intensitas angin, yang dapat mempengaruhi distribusi awan dan pola hujan. Kecepatan Angin Maksimum (ff_x), Kecepatan Angin Rata-Rata (ff_{avg})

7. Arah Angin (Wind Direction - ddd): Menunjukkan dari mana angin bertiup, yang dapat memberi petunjuk tentang pergerakan sistem cuaca yang membawa hujan, ddd_x (Arah Angin dalam Derajat), ddd_car (Arah Angin dalam Kategori atau Karakteristik) (ZonaBMI, 2024)



Gambar 2. 1 Tampilan BMKG (BADAN METEOROLOGI, KLIMATOLOGI, DAN GEOFISIKA, 2024)

Dengan terus melakukan penelitian dan pengembangan, BMKG berupaya meningkatkan akurasi prediksi dan mitigasi risiko bencana. Informasi dari BMKG dapat diakses oleh masyarakat melalui berbagai media, termasuk situs web, aplikasi, dan saluran berita, sehingga mereka dapat merencanakan aktivitas sehari-hari dan merespons potensi bahaya dengan lebih baik. (BADAN METEOROLOGI, KLIMATOLOGI, DAN GEOFISIKA, 2024)

2.1.7 Badan Informasi Geospasial (Ina-Geospasial)

Ina-Geoportal adalah geoportal nasional yang menjadi platform penghubung berbagai instansi pemerintah, termasuk kementerian, lembaga, provinsi, dan daerah. Portal ini berfungsi sebagai simpul Jaringan Informasi

Geospasial Nasional (JIGN), berbagi dan akses data geospasial secara efektif. Pengguna dapat memanfaatkan fitur canggih seperti analisis data, *geoprocessing*, dan *geotagging*. Teknologi *mapviewer* berbasis sumber terbuka (*opensource*) mendukung kemampuan untuk "*drag and drop*" data file, memberikan pengalaman yang lebih interaktif dan mudah dalam mengelola serta memvisualisasikan data geospasial (Ina Geoportal, 2024).



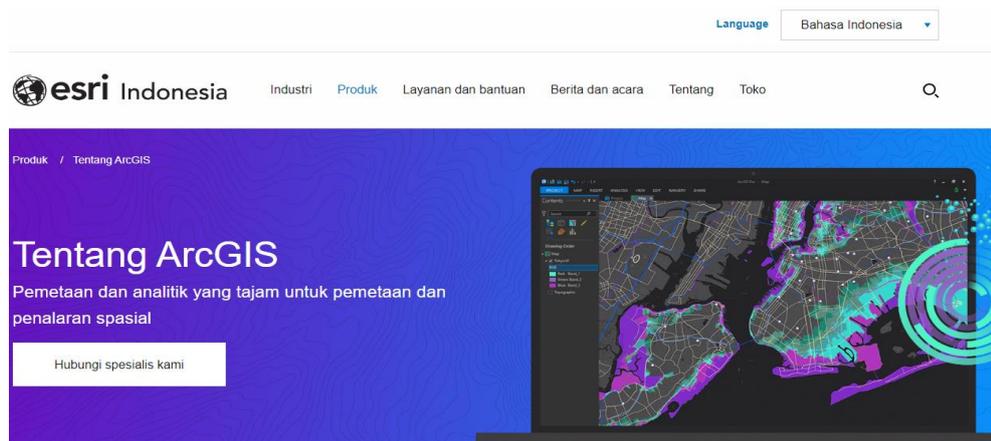
Gambar 2. 2 Tampilan Ina-Geospasial (Ina Geoportal, 2024)

2.1.8 ArcGIS

ArcGIS adalah platform sistem informasi geografis (SIG) yang menyediakan alat untuk pemetaan, analisis spasial, serta pengelolaan data geografis. ArcGIS mendukung berbagai fungsi, seperti:

1. Analisis Spasial: Menggali hubungan geografis untuk pengambilan keputusan.
2. Operasional Lapangan: Mendukung pengelolaan data di lapangan secara real-time.
3. Pemetaan & Visualisasi: Membuat peta dinamis.
4. SIG 3D: Visualisasi dan analisis data dalam 3 dimensi.

5. Penginderaan Jauh: Mengolah citra satelit.
6. Pengumpulan & Pengelolaan Data: Menyederhanakan pemrosesan data lapangan (Esri Indonesia, 2024) (Gambar 2.3).



Gambar 2. 3 Tampilan ArcGIS (Esri Indonesia, 2024)

2.1.9 Haversine Formula

Haversine formula adalah rumus yang digunakan untuk menghitung jarak antara dua titik di permukaan bumi berdasarkan garis lintang (latitude) dan garis bujur (longitude) dalam koordinat geografis. Rumus ini mempertimbangkan bentuk bumi yang bulat, sehingga lebih akurat dibandingkan perhitungan jarak euclidean biasa. Diberikan dua titik dengan koordinat:

$$x = (\text{lon}2 - \text{lon}1) * \cos((\text{lat}1 + \text{lat}2)/2) \quad (2.1)$$

$$y = (\text{lat}2 - \text{lat}1) \quad (2.2)$$

Keterangan:

x = Longitude (Lintang)

y = Latitude (Bujur)

d = Jarak

R = Radius Bumi di mana R adalah jari-jari bumi (sekitar 6.371 km)

1 derajat = 0.017 radian (dalam tiga decimal) (Abadi Nugroho, 2020).

2.1.10 *Machine learning*

Machine learning adalah aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan deteksi di masa yang akan datang. *Machine learning* terdiri dari tiga kategori, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. *Machine learning* membantu komputer belajar dari data-data sebelumnya dan mendeteksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran. *Machine learning* dapat digunakan dalam berbagai bidang, seperti bisnis, kedokteran, robotika, bahasa alami, matematika, game, persepsi, diagnosis medis, teknik, analisis keuangan, analisis sains, dan penalaran (Riadi, 2022) (Rozzi Kesuma Dinata, 2020)

2.1.11 *K-Nearest Neighbor*

K-Nearest Neighbor dikenal sebagai K -NN, sebuah algoritma *machine learning* yang digunakan untuk klasifikasi dan regresi (adminlp2m, 2023). Algoritma ini bekerja dengan mencari K titik data terdekat dengan titik data input, dan kemudian mengklasifikasikan titik data input berdasarkan mayoritas kelas dari K tetangga terdekat tersebut. Nilai K dapat diatur oleh pengguna dan dapat disesuaikan untuk meningkatkan akurasi klasifikasi.

Algoritma K-NN sering digunakan dalam berbagai bidang seperti pengenalan gambar, pemrosesan bahasa alami, dan sistem rekomendasi (Pangestu, 2018) (Cholil et al., 2021). Berikut merupakan alur perhitungan KNN:

1. Pilih nilai K: Pilih jumlah tetangga terdekat yang akan digunakan untuk pertimbangan penentuan kelas.
2. Hitung jarak: Hitung jarak dari data baru ke masing-masing data point di dataset.
3. Temukan kelas terdekat: Ambil K data dengan jarak terdekat.

Untuk menentukan seberapa dekat atau jauh suatu titik pada kelas k , menggunakan jarak *Euclidean* yaitu cara yang sering digunakan pada penelitian lainnya seperti pada penelitian (Rahayu et al., 2022), dimana dari hasil tersebut uji coba pada jarak euclidean memiliki akurasi yang paling tinggi yaitu 84% pada $K=5$ dibandingkan dengan jarak manhattan. Pada *K-Nearest Neighbor* terdapat 3 rumus fungsi jarak yaitu *Manhattan* adalah metode perhitungan jarak pada ruang jarak dengan menerapkan konsep selisih mutlak, *Minkowski* (juga disebut ruang norma L_p) adalah kelas dari berbagai fungsi jarak yang dibentuk oleh parameter p . Untuk dua sampel yang diberikan, *Euclidean* digunakan untuk menghitung jarak antara 2 titik dalam dua dimensi. Berikut rumus menghitung jarak *Manhattan* pada persamaan (2.3), *Minkowski* pada persamaan (2.4), *Euclidean* pada persamaan (2.5) (Mailagaha Kumbure & Luukka, 2022).

$$\text{Manhattan Distance} = d_{ij} \sum_k |x_{ik} - c_{jk}| \quad (2.3)$$

Keterangan pada persamaan (2.3), dimana diketahui d_{ij} adalah jarak antara kasus antara i_{th} dan j_{th} dengan semua parameternya. W merepresentasikan jumlah dari bobot. X adalah kasus yang baru dikurangi dengan C yaitu *history*.

$$\text{Minkowski Distance} = d_{Md}(X_i, X_j) = \left(\sum_{t=1}^m |x_t^i - x_t^j|^p \right)^{1/p} \text{ for } p \geq 1 \quad (2.4)$$

Keterangan pada persamaan 2.4 adalah untuk dua sampel yang diberikan X_i dan X_j dimana $X_i = \{x_1^i, x_2^i, \dots, x_m^i\} \in R^m$,

$$\text{Euclidean Distance} = d_i \sqrt{\sum_{i=1}^p (x_1 - x_2)^2} \quad (2.5)$$

Keterangan pada persamaan 2.5 sebagai berikut.

d = jarak

i = variabel data

p = dimensi data

x_1 = sampel data

x_2 = data uji

4. Tentukan kelas: Tentukan kelas dari data baru tersebut berdasarkan mayoritas dari kelas pada tetangga terdekat. KNN memiliki kelebihan seperti mudah diimplementasikan, dapat mengatasi data noisy, dan dapat menanggulangi data yang jumlahnya besar. Namun, KNN juga memiliki kekurangan seperti perlu menentukan nilai K dari parameter, dan tidak cocok untuk dataset berukuran besar atau dimensi tinggi (Rahayu et al., 2022) (JAMARI, 2022).

2.1.12 Algoritma Random Forest

Algoritma Random Forest merupakan algoritma *machine learning* yang menggabungkan hasil (*output*) dari beberapa *decision tree* untuk mencapai satu hasil yang lebih akurat. Perpaduan antara masing-masing pohon melalui pohon keputusan yang digabungkan menjadi satu model (Afdhal et al., 2022). Pada *Random Forest* terdapat dua jenis rumus, yaitu rumus untuk prediksi regresi dan klasifikasi., dimana jika masalah untuk prediksi klasifikasi dapat dilihat pada persamaan (2.6) sebagai berikut:

$$I(y) = \operatorname{argmax}_c (\sum_{n=1}^N I_{h_n}(y) = c) \quad (2.6)$$

Keterangan pada persamaan 2.6 sebagai berikut.

$I(y)$ = prediksi kelas untuk sampel uji y

C = kelas yang mungkin

$(\sum_{n=1}^N I_{h_n}(y) = c)$ = indikator apakah pohon keputusan ke- i memprediksi kelas c untuk sampel uji y

argmax_c = fungsi yang mengembalikan nilai c dengan probabilitas tertinggi

Random Forest memiliki mekanisme internal yang disebut out-of-bag (OOB) error estimate, di mana I adalah fungsi indikator dan h_n adalah pohon ke- n dari RF. Mekanisme ini memberikan estimasi dari kesalahan generalisasi sendiri. Hanya dua pertiga dari data asli yang digunakan untuk pengambilan sampel bootstrap digunakan dalam pembentukan pohon, dan satu pertiga sisanya dimasukkan ke dalam pohon yang telah dibentuk dan digunakan untuk mengevaluasi kinerjanya. Kesalahan estimasi OOB adalah

kesalahan rata-rata untuk setiap kasus pelatihan y menggunakan pohon yang tidak memasukkan y dalam sampel bootstrap-nya. Saat *Random Forest* dibuat, semua kasus pelatihan menyusuri setiap pohon, dan matriks kedekatan setiap kasus dihitung dengan menggunakan sepasang kasus yang sampai di terminal node yang sama (Jonathan, 2023)

Untuk membangun suatu pohon prediksi, perlu melakukan sampling ini merupakan salah satu aspek yang terpenting. Parameter yang lebih berpengaruh dalam hasil prediksi pada Algoritma *Random Forest* untuk pencegahan terjadi suatu *overfitting*, adalah

1. *N_Estimators* merupakan parameter dalam Algoritma *Random Forest* yang digunakan untuk menentukan jumlah suatu pohon keputusan yang akan dibentuk dalam kumpulan (ensemble).
2. *Max Depth* merupakan suatu pengukuran kedalaman pohon yang akan membagi suatu data sesuai dengan dengan fitur (variable). Dimana semakin dalam pohon tersebut, semakin kompleks keputusan yang bisa diambil.
3. *Criterion* merupakan parameter untuk mengukur atau menentukan metrik dalam memilih pemisahan untuk membangun suatu pohon.
4. *Min Samples Split* merupakan parameter yang menentukan jumlah minimum sampel pada satu simpul, sebelum dibangun agar pemisah dapat diproses.

5. Max Features merupakan parameter yang dapat mengatur jumlah maksimum fitur (*variable*) dalam membangun pohon (Afdhal et al., 2022)

2.1.13 *Confusion Matrix*

Confusion Matrix adalah alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai. Berikut akan disajikan bentuk *Confusion Matrix* pada Tabel 2.1 (Qadrini, 2023).

Tabel 2. 1 Confusion Matrix

		TRUE	FALSE
Nilai Prediksi	TRUE	TP (<i>True Positive</i>) <i>Correct result</i>	FP (<i>False Positive</i>) <i>Unexpected result</i>
	FALSE	FN (<i>False Negative</i>) <i>Missing result</i>	TN (<i>True Negative</i>) <i>Correct absence of result</i>

Pada Tabel 2.1 nilai TP (*True Positive*) dan TN (*True Negative*) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, Presisi, dan Recall. Jika label prediksi keluaran bernilai benar (*true*) dan nilai sebenarnya bernilai salah (*false*) disebut sebagai *False Positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*false*) dan nilai sebenarnya bernilai benar (*true*) maka hal ini disebut sebagai *False Negative* (FN). Berikut formulasi untuk menghitung Akurasi, Presisi, Recall dan FI-Score pada pembentukan model

klasifikasi ditunjukkan pada Persamaan (2.7), Persamaan (2.8), Persamaan (2.9), Persamaan (2.10) sebagai berikut (Qadrini, 2023):

- *Accuracy*, menggambarkan *persentase* jumlah record data yang terklasifikasi dengan benar oleh sistem menggunakan Persamaan (2.7)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.7)$$

- *Precision*, menggambarkan presentase akurasi antara data yang diminta dengan hasil klasifikasi KNN menggunakan Persamaan (2.8)

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (2.8)$$

- *Recall*, menggambarkan presentase akurasi antara data yang diminta dengan hasil KNN menggunakan Persamaan (2.9)

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.9)$$

- F1-Score, menggambarkan suatu perbandingan nilai rata-rata dari *precision* dan *recall* menggunakan Persamaan (2.10) (Martia Nanda et al., 2022).

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2.10)$$

2.2 Penelitian Terkait

2.2.1 *State of the art*

Pada penelitian ini, berdasarkan rumusan masalah dan tujuan penelitian yang telah dibuat, dilakukan *literature review* dengan implementasi deteksi bencana banjir di Kabupaten Garut, *Machine learning* serta *Algoritma Random Forest*, *K-Nearest Neighbor* dan algoritma lainnya. Beberapa penelitian sebelumnya yang dilakukan adalah berikut Tabel 2.2:

Tabel 2.2 *State of The Art*

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
1.	Mengidentifikasi Resiko Banjir Di Kabupaten Cirebon Dalam Menghadapi Curah Hujan Yang Ekstrem Menggunakan Algoritma K-Means, (Depi Novita Sari, 2024)	<i>K-Means Clustering</i>	Dalam penelitian ini hanya menggunakan variabel curah hujan untuk analisis kejadian bencana banjir. Saran untuk melibatkun variabel – variabel tambahan yang relevan	SINTA 5
2.	Klasifikasi Daerah Rawan Banjir menggunakan 10-Fold Cross Validation dan <i>K-Nearest Neighbors</i> , (Bagaskara & Hartomo, 2024)	<i>K-Nearest Neighbors</i>	Dalam penelitian ini hanya mempertimbangkan curah hujan dan diusulkan kembali pada penelitian ini akan menambahkan variabel dengan cukup sehingga bisa menghasilkan yang relevan.	SINTA 3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
3.	Flood <i>Classification</i> and prediction in South Sudan using artificial intelligence models under a changing climate, (Mohamed El-Sayed El-Mahdy F. A.-T., 2024)	<i>Artificial Neural Networks (ANN), General Regression Neural Network (GRNN), Recurrent Neural Networks (RNN), ...,K-Nearest Neighbor (KNN), Gradient Boost (GB</i>	Penelitian ini hanya menggunakan data curah hujan dari stasiun di sepanjang wilayah antara tahun 2010 dan 2019. Mungkin ada faktor lain yang mempengaruhi risiko banjir Sungai Nil, seperti tutupan lahan, tinggi permukaan air sungai, dan data elevasi daerah.	Q3
4.	Predicting real-time roadway pluvial flood risk: A hybrid <i>machine learning</i> approach coupling a graph-based flood spreading model, historical vulnerabilities, and Waze data, (Arefeh Safaei-Moghadam, 2024)	<i>XGBoost (Extreme Gradient Boosting), Random Forest, Support Vector Machines (SVM)</i>	Penelitian ini memberikan kontribusi yang berharga dalam memahami risiko banjir bandang di ruas jalan. Namun, mempertimbangkan variabel lain seperti kondisi saluran air, ketinggian permukaan di persimpangan jalan, dan topografi daerah sekitarnya dapat meningkatkan akurasi prediksi model dan kegunaan penelitian dalam penelitian.	Q1
5.	Spatial modeling of flood probability using geo-environmental Variables and <i>machine learning</i> models, case	<i>Self-Organization Map (SOM), Radial Basis Function Neural Network (RBFNN), and</i>	Ketiadaan variabel intensitas curah hujan adalah kelemahan karena curah hujan tinggi dalam waktu singkat bisa menyebabkan banjir bandang. Juga,	Q3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
	study:Tajan watershed, Iran, (Mohammadtaghi Avand, 2021)	<i>Multi-layers Perceptron (MLP)</i>	penelitian akan lebih akurat jika mencakup semua jenis penutup lahan, bukan hanya lahan pertanian dan pemukiman.	
6.	Flood susceptibility prediction using four <i>machine learning</i> techniques and comparison of their performance at Wadi Qena Basin, Egypt, (El-Haddad et al., 2021)	<i>Boosted Regression Tree (BRT), Functional Data Analysis (FDA) General Linear Model (GLM) Multivariate Discriminant Analysis (MDA)</i>	Data intensitas curah hujan belum dimasukkan dalam penelitian ini, padahal merupakan faktor utama penyebab banjir. Dengan menambahkannya, model akan lebih akurat dalam mengidentifikasi daerah rentan banjir. Selain itu, data drainase wilayah, seperti kapasitas saluran dan keberadaan sungai atau waduk, juga berpengaruh. Memasukkan data drainase akan membantu model mengidentifikasi daerah dengan drainase buruk yang lebih rentan banjir.	Q1
7.	<i>A Machine learning Based Prediction and Analysis of Flood Affected Households: A Case Study of Floods in Bangladesh</i> , (Ganguly et al., 2019)	<i>linear regression, Random Forest and artificial neural network</i>	Penelitian ini menganalisis kerusakan banjir menggunakan model <i>machine learning</i> berdasarkan data tingkat distrik dari rumah tangga yang terdampak. Elevasi sebagai variabel, elevasi berpotensi menjadi variabel	Q3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
			penting karena berkaitan dengan ketinggian banjir, akses evakuasi, dan dampaknya terhadap kerusakan.	
8.	Analisis Pengaruh Gain Rati untuk Algoritma <i>K-Nearest Neighbor</i> pada Klasifikasi Data Banjir Di Kota Samarinda (Septia Intan Permata Sari, 2023)	<i>K-Nearest Neighbor</i>	Hasilnya menunjukkan peningkatan akurasi tertinggi saat $K=7$ (5,95%), diikuti oleh $K=5$ (5,81%) dan $K=3$ (5,68%). Namun, belum ada interpretasi hasil klasifikasi untuk mendukung pengambilan keputusan mitigasi banjir di Kota Samarinda.	SINTA 5
9.	Implementasi Algoritma <i>K-Means Clustering</i> Untuk Pengelompokan Wilayah Rawan Banjir (Try Wathoriq, 2023)	<i>K-Means Clustering</i>	BPBD Kota Tangerang kesulitan dalam mengelola data banjir. Solusinya dengan menggunakan <i>K-Means Clustering</i> , untuk mengelompokkan wilayah banjir berdasarkan dampak. Kemungkinan, penelitian sebelumnya juga telah mencoba pendekatan lain seperti analisis statistik atau pemodelan spasial.	SINTA 4
10.	Flood Prediction based on Weather Parameters in Jakarta using <i>K-Nearest Neighbours Algorithm</i> , (Hariman Lumbantobing, 2024)	<i>K-Nearest Neighbours Algorithm</i>	<i>K-Nearest Neighbours (KNN)</i> adalah algoritma yang digunakan untuk memprediksi wilayah yang akan terkena dampak banjir. Hasil penelitian	SINTA 3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
			ini dengan nilai k=2 sampai k=9 menunjukkan performa terbaik pada k=7, dimana tingkat akurasi mencapai 92.25%, presisi 88.89%, recall 92.25%, dan F1-measure sebesar 89.52%.”	
11.	<i>Machine Learning-Based Daily Rainfall Data Gap-Filling Method: A Case Study of DKI Jakarta Area</i> , (Asshidiqi & Saputro, 2023)	extreme gradient boosting (XGB), <i>K-Nearest Neighbor</i> (KNN), <i>Random Forest</i> (RF), stochastic gradient descent (SGD), dan multi-layer perceptron (MLP)	Beberapa ML yang digunakan untuk gap-filling dalam penelitian ini antara lain extreme gradient boosting (XGB), <i>K-Nearest Neighbor</i> (KNN), <i>Random Forest</i> (RF), stochastic gradient descent (SGD), dan multi-layer perceptron (MLP). Hasil penelitian yang diperoleh adalah pemodelan KNN mendominasi nilai performa terbaik pada data latih, namun pada data pengujian SGD mendominasi performa terbaik dengan nilai rata-rata RMSE, MAE, dan R ² berturut-turut sebesar 5,95, 2,40, dan 0,87.	Q1
12.	Klasifikasi Cuaca Provinsi DKI Jakarta Menggunakan Algoritma <i>Random Forest</i> dengan Teknik Oversampling (Hamami & Dahlan, 2022)	<i>Random Forest</i>	Penelitian menggunakan <i>Random Forest</i> dan SMOTE untuk klasifikasi cuaca ekstrem, mencapai akurasi 70%. <i>Random Forest</i> efektif mengatasi	SINTA 4

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
			ketidakseimbangan data, namun perlu pemahaman yang lebih dalam terkait jenis data. Penelitian sebelumnya mungkin telah mencoba pendekatan alternatif.	
13.	Analisis Perbandingan Algoritma <i>Decision Tree</i> , <i>Random Forest</i> , dan <i>Naïve Bayes</i> untuk Prediksi Banjir di Desa Dayeuhkolot (Arya Darmawan et al., 2023)	<i>Decision Tree</i> , <i>Random Forest</i> , dan <i>Naïve Bayes</i>	<i>Machine learning (Decision Tree, Random Forest, dan Naïve Bayes)</i> untuk prediksi banjir di Desa Dayeuhkolot, Kabupaten Bandung. <i>Random Forest</i> memiliki performa terbaik dengan akurasi tinggi dan waktu komputasi cepat, dibandingkan dengan <i>Decision Tree</i> dan <i>Naïve Bayes</i> .	SINTA 3
14.	Analisis Perbandingan Algoritma <i>Random Forest</i> Dan <i>Naïve Bayes</i> Untuk Klasifikasi Penyelesaian Hujan Berdasarkan Iklim Di Indonesia, (Nicolaus Advendea Prakoso Indaryono, 2024)	<i>Random Forest</i> dan <i>Naïve Bayes</i>	<i>Algoritma Random Forest</i> mencapai akurasi sebesar 86.55%, sedangkan algoritma <i>Naïve Bayes</i> hanya mencapai akurasi sebesar 36.61%. Diharapkan hasil dari penelitian ini dapat digunakan sebagai referensi untuk studi literatur penelitian selanjutnya dan digunakan untuk memantau curah hujan harian di Indonesia guna mencegah bencana alam.	SINTA 3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
15.	Algoritma <i>Random Forest</i> untuk Prediksi Curah Hujan, (Primajaya & Sari, 2018)	<i>Random Forest</i>	Penelitian ini bertujuan memprediksi curah hujan untuk mengantisipasi banjir menggunakan algoritma <i>Random Forest</i> . Metode penelitian terdiri dari pengumpulan data, pengolahan data, implementasi <i>Random Forest</i> , dan analisis. Implementasi <i>Random Forest</i> dengan training set menghasilkan model dengan akurasi 71,09%, presisi 0,75, recall 0,85, f-measure 0,79, kappa 0,33, MAE 0,35, RMSE 0,46, dan ROC Area 0,78. Dengan 10-fold cross validation, akurasi meningkat menjadi 99,45%, dengan presisi 0,99, recall 0,99, f-measure 0,99, kappa 0,99, MAE 0,09, RMSE 0,14, dan ROC Area 1.	SINTA 2
16.	<i>Flood Mapping Using Random Forest and Identifying The Essential Conditioning Factors; A Case Study In Fredericton, New Brunswick, Canada</i> , (Esfandiari et al., 2020)	<i>Random Forest</i>	Penelitian ini menunjukkan bahwa faktor berkorelasi dapat mempengaruhi prediksi. Faktor terpenting adalah ketinggian, kemiringan lereng, aspek, jarak dari sungai, dan penggunaan lahan. Meskipun demikian, <i>Random</i>	Q3

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
			<i>Forest</i> tetap akurat dengan data yang berkorelasi, dan menghilangkan faktor seperti ketinggian tidak merusak akurasi model.	
17.	Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and <i>Random Forest</i> methods, (Chen et al., 2020)	naïve Bayes tree, alternating decision tree, and <i>Random Forest</i>	Hasil penelitian menunjukkan bahwa metode RF merupakan model yang efisien dan handal dalam penilaian kerentanan banjir, dengan nilai AUC, rasio prediksi positif, rasio prediksi negatif, sensitivitas, spesifisitas, dan akurasi tertinggi untuk data pelatihan (masing-masing 0,951, 0,892, 0,941, 0,945, 0,886, dan 0,915) dan data validasi (masing-masing 0,925, 0,851, 0,938, 0,945, 0,835, dan 0,890).	Q1
18.	Flood Hazard Rating Prediction for Urban Areas Using <i>Random Forest</i> and LSTM, (Kim & Kim, 2020)	<i>Random Forest</i> and LSTM	Penggunaan <i>Random Forest</i> dalam penelitian ini kemungkinan besar menguntungkan karena dapat meningkatkan akurasi klasifikasi zona bahaya banjir, menangani kompleksitas data banjir, dan mengurangi overfitting.	Q2

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
19.	Identification of the key driving factors of flash flood based on different feature <i>Selection</i> techniques coupled with <i>Random Forest</i> method, (Tan et al., 2024)	<i>Random Forest</i>	Menunjukkan bahwa <i>Random Forest</i> memiliki beberapa kelebihan dalam mengidentifikasi KDF banjir bandang dan memodelkan risiko banjir bandang. Akurasi, kemampuan prediksi, dan wawasan yang diperoleh dari metode ini membuatnya menjadi alat yang berharga untuk pengelolaan risiko banjir bandang	Q1
20.	An empirical flood fatality model for Italy using <i>Random Forest</i> algorithm. (Yazdani et al., 2023)	<i>Random Forest</i>	menunjukkan bahwa <i>Random Forest</i> memiliki beberapa kelebihan dalam memprediksi kematian akibat banjir dan mendukung pengembangan strategi mitigasi. Akurasi, kemampuan generalisasi, dan interpretasi yang mudah dari model ini menjadikannya alat yang berharga untuk meningkatkan pengelolaan risiko banjir dan menyelamatkan nyawa.	Q2
21.	A Spatial Improved-kNN-Based Flood Inundation Risk Framework for Urban Tourism under Two Rainfall Scenarios, (Liu S, 2021)	Knn	menunjukkan bahwa KNN memiliki beberapa kelebihan dalam mengevaluasi risiko FIR untuk lokasi pariwisata. Efektivitas, akurasi,	Q1

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
			integrasi data spasial, skalabilitas, dan kemampuan adaptasi menjadikan KNN sebagai alat yang berharga untuk mendukung pengambilan yang berkelanjutan dalam industri pariwisata perkotaan.	
22.	A Novel Hybrid Extreme Learning Machine Approach Improved by K Nearest Neighbor Method and Fireworks Algorithm for Flood Forecasting in Medium and Small Watershed of Loess Region, (Ren et al., 2019)	K Nearest Neighbor	menunjukkan bahwa KNN memiliki beberapa kelebihan dalam memprediksi banjir di daerah aliran sungai sedang dan kecil. Akurasi prediksi yang lebih tinggi, kinerja yang unggul di berbagai kondisi, dan kemampuan adaptasi menjadikan KNN sebagai alat yang berharga untuk mendukung pengendalian banjir dan pengurangan bencana di wilayah loess	Q1
23.	The Applicability of LSTM-KNN Model for Real-Time Flood Forecasting in Different Climate Zones in China, (M. Liu et al., 2020)	LSTM- KNN	ini mengusulkan pendekatan berbasis KNN untuk memperbarui model LSTM secara real-time. Ini menunjukkan bahwa KNN mungkin berguna untuk memasukkan data baru dan berpotensi meningkatkan akurasi prakiraan banjir.	Q1

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
24.	A Rapid Forecast Method for the Process of Flash Flood Based on Hydrodynamic Model and KNN Algorithm, (Zhou et al., 2024)	KNN	Menunjukkan bahwa KNN memiliki beberapa kelebihan dalam meningkatkan akurasi dan ketepatan waktu prediksi banjir bandang. Kecepatan komputasi yang tinggi, keandalan, dan kemudahan implementasi menjadikan KNN alat yang berharga untuk mengembangkan sistem peringatan dini dan peramalan banjir bandang yang efektif.	Q1
25.	Penerapan Data Mining Untuk Pemetaan daerah Rawan Bencana Sebagai Upaya Kesiapsiagaan Terhadap Bencana (Purwayoga et al., 2023)	<i>K-Means Clustering</i>	Hasilnya memberikan pemahaman yang lebih baik tentang pola bencana di wilayah tersebut untuk merujuk dalam program kesiapsiagaan bencana. Namun, tantangan seperti penyesuaian jumlah cluster, sensitivitas terhadap inisialisasi <i>centroid</i> , dan kecocokan algoritma dengan beragam bentuk cluster perlu diatasi.	SINTA 4
26.	Penerapan Metode <i>K-Means Clustering</i> Untuk Desa Rawan Bencana	<i>K-Means Clustering</i>	Penelitian wilayah rawan banjir di Kabupaten Purbalingga menggunakan metode <i>K-Means clustering</i> karena kecepatan, fleksibilitas, dan	SINTA 4

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
	Berdasarkan Data Kejadian Terjadinya Bencana Alam (Ramadhani et al., 2022)		kemudahan implementasinya. Namun, ada kekurangan, termasuk sensitivitas terhadap jumlah cluster awal yang harus diperhatikan.	
27.	Flood Damage Analysis Using <i>Machine learning Techniques</i> (Snehil & Goel, 2020)	<i>Random Forest</i> dan K-NN	Hasilnya menunjukkan model <i>Random Forest</i> dan K-NN memiliki kinerja terbaik. Namun, belum dipertimbangkan faktor-faktor lokal seperti siklon dan bendungan. Rekomendasi selanjutnya adalah memasukkan faktor-faktor tersebut untuk meningkatkan akurasi prediksi.	Q3
28.	Penerapan Deteksi Bencana Banjir menggunakan Metode <i>Machine learning</i> (Fatonah, 2021)	<i>Naïve Bayes</i> dan <i>Support Vector Machine (SVM)</i>	Kendala utama adalah ketidakmampuan tools untuk memprediksi hari terjadinya banjir. Rekomendasi pengembangan termasuk pemrosesan data lebih canggih dan penggunaan model prediksi yang lebih kompleks untuk meningkatkan akurasi dalam mitigasi bencana banjir.	SINTA 5
29.	Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat	<i>K-Means Clustering</i>	Penelitian menggunakan data mining dengan <i>K-Means Clustering</i> untuk mengelompokkan data bencana alam di	SINTA 5

No	Judul, Penulis, Tahun	Metode/Solusi	Hasil	Terindeks oleh
	Menggunakan Algoritma <i>K-Means Clustering</i> (Halik & Septiana, 2022)		Jawa Barat. Meskipun memberikan wawasan tentang daerah rawan bencana, kelemahannya adalah kemungkinan menghasilkan cluster yang tidak merepresentasikan keberagaman spasial data.	
30.	Perbandingan Akurasi <i>Linear Regression</i> dan <i>Support Vector Regression</i> dalam Prediksi Suhu Rata-Rata (Gideon Namlea Lesnusa1, 2024)	<i>Linear Regression</i> dan <i>Support Vector Regression</i>	Penggunaan data mingguan dapat menganalisis yang lebih rinci dan responsif terhadap perubahan cuaca, mendukung pentingnya ketelitian data dalam membangun model prediksi yang akurat.	SINTA 2
31.	Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan (Studi Kasus: Institut Teknologi Nasional Bandung)(Miftahuddin et al., 2020)	<i>Haversine</i>	Hasilnya bahwa <i>haversine</i> memiliki tingkat akurasi yang lebih tinggi daripada euclidean. Dari hal tersebut didapatkan hasil analisis bahwa perhitungan jarak pada permukaan bumi harus tetap mengikutsertakan kemiringan permukaan bumi.	SINTA 3

2.2.2 Matriks Penelitian

Tabel 2. 3 Matriks Penelitian

Judul	Permasalahan	Indikator
Pemodelan Multivariabel Dengan Pendekatan <i>Machine Learning</i> Untuk Deteksi Bencana Banjir Di Kabupaten Garut	<ol style="list-style-type: none"> 1. Bagaimana membangun pemodelan multivariabel untuk deteksi banjir dengan pendekatan <i>Machine learning</i> menggunakan algoritma Algoritma K-Nearest Neighbor dan <i>Random Forest</i>? 2. Bagaimana analisis performa dari algoritma <i>Algoritma K-Nearest Neighbor</i> dan <i>Random Forest</i> menggunakan multivariabel dalam mendeteksi banjir? 	<ol style="list-style-type: none"> a. Membangun model multivariabel untuk deteksi banjir menggunakan pendekatan <i>Machine learning</i> dengan algoritma <i>K-Nearest Neighbor</i> (KNN) dan <i>Random Forest</i> (RF). Tujuan ini menekankan pengembangan model yang memanfaatkan beberapa variabel (multivariabel) untuk memprediksi kejadian banjir. b. Menganalisis performa algoritma <i>K-Nearest Neighbor</i> dan <i>Random Forest</i> dalam memprediksi banjir menggunakan data multivariable, untuk menentukan algoritma mana yang memiliki kinerja lebih baik dalam hal akurasi, precision, recall, dan F1-score dalam mendeteksi banjir berdasarkan berbagai variabel yang relevan
Sumber Data	Metode Penelitian	
<ol style="list-style-type: none"> 1. BMKG 2. BPBD 	<ol style="list-style-type: none"> 1. Jenis Penelitian: Penelitian kuantitatif menggunakan data numerik untuk menguji hipotesis tentang deteksi dan prediksi banjir dengan machine learning, menggunakan data cuaca, bujur, lintang, elevasi, dan label potensi banjir. 	

<p>3. Data Batas Wilayah dari (Ina-Geospasial) Badan Informasi Geospasial Indonesia</p>	<p>2. Pengumpulan Data: Data diperoleh dari BMKG, BPBD, dan Batas Wilayah Ina-Geospasial untuk keperluan visualisasi.</p> <p>3. Proses Penelitian: Dimulai dengan identifikasi masalah, studi literatur, dan pengumpulan data (cuaca dari BMKG, kejadian banjir dari BPBD, dan batas wilayah dari Ina-Geospasial). Data kemudian diproses melalui pembersihan, transformasi (scaling, normalisasi), serta analisis eksplorasi dan seleksi fitur untuk memilih data cuaca yang relevan. Data dibagi menjadi data latih dan uji (70:30, 80:20, 90:10), dan model KNN dan <i>Random Forest</i> digunakan untuk klasifikasi. Evaluasi dilakukan dengan metrik akurasi, precision, recall, F1-score, dan Confusion Matrix untuk mengukur performa model dalam memprediksi kategori "banjir" dan "tidak banjir". Hasil evaluasi diharapkan dapat digunakan oleh BPBD Kabupaten Garut dalam mitigasi dan perencanaan penanganan banjir.</p>
---	--

2.3 Kebaruan Penelitian

Penelitian ini memiliki beberapa kebaruan signifikan. Pertama, penelitian ini mengusulkan penggunaan pendekatan *multivariable* dalam deteksi banjir dengan memanfaatkan metode *Machine learning* seperti *K-Nearest Neighbor* dan *Random Forest*.

Hasil penelitian ini tidak hanya memberikan kontribusi baru terhadap deteksi bahaya banjir, tetapi juga mempertimbangkan berbagai variabel tambahan yang relevan selain curah hujan. Dalam proses ini, Algoritma K-NN dan *Random Forest* digunakan dengan beberapa langkah khusus. Selain itu, penelitian ini melakukan analisis performa algoritma *K-Nearest Neighbor* dan *Random Forest* dalam memprediksi banjir menggunakan data *multivariable*, untuk menentukan algoritma mana yang memiliki kinerja lebih baik dalam hal akurasi, precision, recall, dan F1-score dalam mendeteksi banjir berdasarkan berbagai variabel yang relevan. Berdasarkan beberapa penelitian sebelumnya, *Random Forest* sering menunjukkan performa yang lebih baik dalam beberapa kasus, namun K-NN juga dapat memberikan hasil yang baik. Oleh karena itu, perbandingan performa dilakukan secara spesifik terhadap data dan tujuan yang digunakan. Penelitian ini diharapkan dapat menghasilkan model deteksi banjir yang lebih akurat dengan pendekatan multivariabel dan menggunakan algoritma *Machine learning* (*K-Nearest Neighbor* dan *Random Forest*), yang mempertimbangkan berbagai variabel relevan selain curah hujan, sehingga prediksi banjir menjadi lebih tepat dan bermanfaat bagi mitigasi bencana banjir di wilayah Kabupaten Garut.