
STUDI LITERATUR SISTEMATIS METODE KLASIFIKASI DATA MINING UNTUK PREDIKSI POTENSI BANJIR

Evriska Yolanda¹, Hetty Rohayani²

^{1,2}Fakultas Saint dan Teknologi, Universitas Muhammadiyah Jambi, Jambi, Indonesia

²Fakultas, Nama Institusi, Kota, Indonesia

Email: ¹ evriskayolanda1@gmail.com , ² rohayanihettyknk@gmail.com,

Email Penulis Korespondensi: evriskayolanda1@gmail.com

Keyword :

Data Mining,
Klasifikasi,
Prediksi Banjir,
Random Forest,
Naïve Bayes

ABSTRACT

Studi literatur sistematis ini bertujuan untuk menganalisis dan mengevaluasi berbagai metode klasifikasi data mining yang digunakan dalam prediksi potensi banjir, serta mengidentifikasi metode yang paling efektif berdasarkan hasil penelitian sebelumnya. Dengan menggunakan pendekatan Systematic Literature Review (SLR), penelitian ini mengkaji 15 jurnal ilmiah nasional dan internasional yang membahas penerapan metode klasifikasi seperti Naïve Bayes, Random Forest, K-Means, Decision Tree (J48 dan C4.5), ANFIS-PCA, dan Deep Learning dalam konteks mitigasi bencana banjir. Dari hasil kajian, ditemukan bahwa terdapat tiga metode klasifikasi yang menonjol berdasarkan tingkat akurasi, efisiensi komputasi, dan ketepatan dalam memetakan wilayah rawan banjir. Pertama, algoritma Random Forest menunjukkan performa unggul dalam hal akurasi dan kemampuan menangani data kompleks, seperti yang terlihat pada prediksi banjir di DKI Jakarta. Kedua, metode Naïve Bayes dinilai efektif pada dataset berskala kecil hingga menengah dengan tingkat kecepatan pemrosesan yang tinggi. Ketiga, pendekatan kombinasi ANFIS-PCA menawarkan keunggulan dalam mereduksi dimensi data dan meningkatkan hasil prediksi dalam sistem peringatan dini. Studi ini memberikan wawasan penting bagi pengembang sistem prediksi bencana untuk menentukan pendekatan klasifikasi yang paling sesuai dengan kebutuhan.

1. PENDAHULUAN

Banjir merupakan salah satu bencana alam yang paling sering terjadi di berbagai wilayah Indonesia. Dengan kondisi geografis yang terdiri dari dataran rendah dan daerah aliran sungai besar, Indonesia sangat rentan terhadap peningkatan intensitas curah hujan yang dipengaruhi oleh perubahan iklim. Fenomena ini menyebabkan kerugian yang tidak hanya bersifat material, tetapi juga berdampak pada keselamatan jiwa, lingkungan, dan infrastruktur publik. Oleh karena itu, kebutuhan akan sistem yang mampu memprediksi potensi banjir secara dini menjadi semakin penting dalam upaya mitigasi risiko bencana. Salah satu pendekatan yang berkembang dalam dekade terakhir adalah penerapan metode klasifikasi dalam data mining, untuk

mengolah data iklim dan hidrologi guna mengidentifikasi potensi banjir berdasarkan pola historis yang tersedia.[9]

Peningkatan kebutuhan dalam memprediksi potensi bencana seperti banjir mendorong berkembangnya berbagai pendekatan analisis data, salah satunya melalui metode klasifikasi dalam data mining. Hal ini disebabkan oleh semakin kompleksnya pola cuaca, perubahan iklim, dan banyaknya variabel lingkungan yang harus dipertimbangkan. Data yang digunakan dalam prediksi banjir pun beragam, mulai dari curah hujan, tinggi muka air, kelembapan tanah, hingga penggunaan lahan, sehingga memerlukan metode analisis yang efisien dan akurat. Oleh karena itu, penggunaan metode klasifikasi data mining menjadi pilihan strategis karena mampu mengolah data besar dan heterogen untuk menghasilkan prediksi yang lebih tepat. Namun demikian, tantangan utama yang sering dihadapi adalah bagaimana memilih metode klasifikasi yang sesuai dengan karakteristik data dan kebutuhan prediksi. Dalam konteks ini, studi literatur sistematis sangat penting dilakukan untuk mengkaji, membandingkan, dan mengevaluasi berbagai metode klasifikasi yang telah diterapkan dalam prediksi banjir, sehingga dapat diketahui metode mana yang paling efektif, akurat, dan dapat diandalkan dalam pengambilan keputusan terkait mitigasi bencana banjir.

Berikut beberapa penelitian sebelumnya yang penulis jadikan referensi dalam jurnal ini :

1. Metode klasifikasi dalam data mining memungkinkan sistem untuk membedakan kondisi-kondisi yang dapat menyebabkan banjir berdasarkan variabel seperti curah hujan, tinggi muka air sungai, dan kelembapan udara. Sebagai contoh, Darmawan et al. (2023) melakukan perbandingan antara algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk prediksi banjir di Desa Dayeuhkolot, Bandung. Dataset yang digunakan terdiri dari parameter curah hujan dan tinggi muka air sungai dari tahun 2015–2018. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi mencapai 99,05% dan f1-score sebesar 98%. Studi ini memperkuat efektivitas metode klasifikasi dalam sistem prediksi banjir berbasis data historis dan menunjukkan bahwa pemilihan algoritma berpengaruh signifikan terhadap tingkat akurasi sistem prediktif.[5]
2. Pendekatan klasifikasi dengan teknologi deep learning juga telah diteliti oleh Sharfina et al. (2023) yang menggunakan algoritma Deep Learning H2O untuk memprediksi banjir di Kalimantan Barat. Penelitian ini memanfaatkan data curah hujan dari lima stasiun klimatologi antara tahun 2019–2022. Hasil pengujian menunjukkan bahwa model deep learning dengan struktur hidden layer 100-50-100 mampu mencapai akurasi hingga 90,54% dengan nilai RMSE serendah 0.087. Keberhasilan ini menegaskan bahwa metode klasifikasi berbasis jaringan saraf buatan dapat menangani kompleksitas data cuaca yang tinggi dan memberikan hasil prediksi yang akurat. Penggunaan model ini dinilai efektif dalam memetakan wilayah rawan banjir serta mampu menghasilkan sistem yang adaptif terhadap fluktuasi iklim ekstrem yang kini makin sering terjadi [6]
3. Selain algoritma kompleks seperti deep learning, metode klasifikasi sederhana seperti Naïve Bayes juga masih digunakan secara luas. Fatonah (2021), dalam penelitiannya, menerapkan metode Naïve Bayes untuk mendeteksi potensi banjir berdasarkan data historis banjir bulanan. Meskipun hasil akurasi yang diperoleh berada di angka 76,73%, metode ini tetap dianggap bermanfaat dalam konteks sistem yang membutuhkan pemrosesan cepat dengan sumber daya terbatas. Namun, keterbatasan dari metode ini

adalah ketidakmampuannya dalam memprediksi waktu kejadian banjir secara spesifik dan kesulitan dalam menangani data yang tidak seimbang.[9]

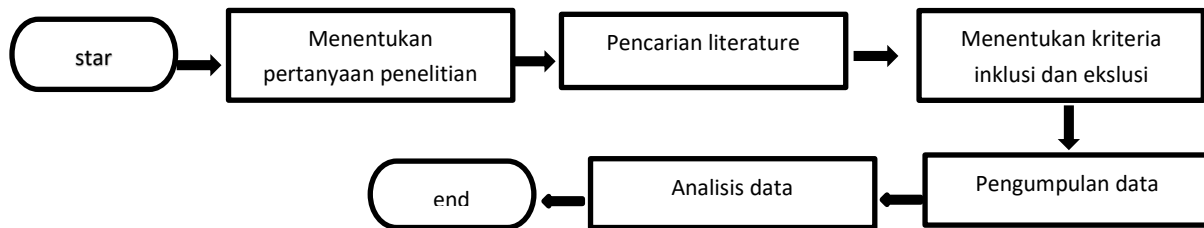
4. Kombinasi algoritma klasifikasi juga menjadi pendekatan menarik yang diangkat oleh Sasoko et al. (2024). Mereka mengembangkan sistem prediksi banjir untuk wilayah DKI Jakarta dengan menggunakan kombinasi metode K-Means dan Random Forest. Dalam model ini, K-Means digunakan untuk mengelompokkan data tinggi muka air sungai, yang kemudian diklasifikasikan lebih lanjut dengan Random Forest untuk menentukan tingkat siaga banjir. Hasil eksperimen menunjukkan bahwa pada nilai $k=14$, akurasi mencapai 95% dengan f1-score sebesar 90%, membuktikan bahwa penggabungan metode klasifikasi dan clustering dapat meningkatkan akurasi dan efisiensi prediksi.[10]
5. Penelitian lain yang relevan dilakukan oleh Risnawati et al. (2023), yang memanfaatkan algoritma Decision Tree C4.5 untuk mengestimasi curah hujan yang berdampak pada banjir di wilayah Jawa Tengah dan Jawa Timur. Dengan klasifikasi empat kategori curah hujan, sistem yang dibangun mampu memberikan gambaran bulanan dan tahunan terkait risiko banjir berdasarkan intensitas hujan. Akurasi yang diperoleh sebesar 83,33% menunjukkan bahwa metode klasifikasi ini mampu mengidentifikasi pola dari data curah hujan dan memiliki potensi untuk digunakan dalam sistem peringatan dini di sektor pertanian dan perkotaan. Penelitian ini juga menunjukkan pentingnya keterlibatan masyarakat dalam memahami kondisi iklim di wilayah mereka agar dapat mengambil tindakan pencegahan yang tepat.[15]

Melalui tinjauan terhadap lima penelitian terdahulu tersebut, dapat disimpulkan bahwa metode klasifikasi dalam data mining memberikan kontribusi signifikan terhadap pengembangan sistem prediksi banjir. Baik menggunakan algoritma klasik seperti Naïve Bayes dan Decision Tree, maupun pendekatan modern seperti Random Forest dan Deep Learning, setiap metode memiliki keunggulan masing-masing dalam hal akurasi, efisiensi, dan kemampuan adaptasi terhadap data yang kompleks. Oleh karena itu, studi literatur sistematis ini bertujuan untuk mengevaluasi secara komprehensif berbagai metode klasifikasi data mining yang digunakan untuk prediksi banjir, membandingkan efektivitasnya, serta mengidentifikasi tantangan teknis dan solusi praktis dalam implementasinya di dunia nyata. Dengan pemahaman yang lebih mendalam ini, diharapkan penelitian dapat menjadi pijakan untuk pengembangan sistem prediksi banjir yang lebih akurat, cepat, dan mudah diimplementasikan dalam skala luas, khususnya di wilayah rawan bencana di Indonesia.

2. METODOLOGI PENELITIAN

Objek dalam penelitian ini adalah literatur ilmiah yang membahas metode klasifikasi data mining untuk prediksi banjir. Objek ini dipilih karena sejalan dengan tujuan utama dari jurnal, yaitu untuk memperoleh pemahaman yang komprehensif tentang bagaimana metode klasifikasi dapat digunakan secara efektif dalam memprediksi bencana banjir berdasarkan data historis dan parameter lingkungan seperti curah hujan, tinggi muka air, serta kelembaban udara. Fokus utama dari studi ini adalah 15 jurnal ilmiah yang telah dipilih melalui proses seleksi yang ketat berdasarkan kriteria inklusi dan eksklusi. Seluruh jurnal tersebut dianalisis secara mendalam untuk mendapatkan informasi terkait algoritma klasifikasi yang digunakan, jenis data yang dianalisis, serta hasil dan tingkat akurasi yang diperoleh dari masing-masing metode. Melalui objek ini, penelitian diharapkan mampu memberikan gambaran umum mengenai metode klasifikasi data mining yang paling sesuai untuk diterapkan dalam prediksi banjir, sekaligus menjadi dasar dalam merumuskan rekomendasi untuk penelitian dan pengembangan lebih lanjut.

Langkah penelitian



Metode yang digunakan dalam penelitian ini adalah Systematic Literature Review (SLR) atau tinjauan pustaka sistematis. Pendekatan ini dilakukan secara terstruktur dan menyeluruh untuk mengidentifikasi, mengevaluasi, serta menganalisis berbagai hasil penelitian yang relevan dengan topik utama jurnal. Tujuan dari metode ini adalah untuk menyusun kesimpulan berdasarkan bukti ilmiah yang telah ada, serta menggali pola umum dan praktik terbaik dalam penggunaan metode klasifikasi untuk prediksi banjir.

Langkah-langkah dalam penelitian ini dilakukan melalui lima tahapan utama yaitu:

1. Menentukan pertanyaan penelitian (*research question*), langkah pertama yang digunakan yaitu menentukan pertanyaan penelitian yang akan menjadi pedoman utama dalam pencarian referensi. Pertanyaan utama dalam jurnal ini seperti, Metode klasifikasi data mining apa saja yang paling efektif dalam memprediksi banjir.
2. Melakukan pencarian literatur melalui Google Scholar menggunakan kata kunci seperti “klasifikasi data mining prediksi banjir” dan “metode klasifikasi untuk banjir”. Pencarian dibatasi pada publikasi 5 tahun terakhir (2020–2025) agar tetap relevan dan mutakhir, hal ini dilakukan untuk memastikan bahwa informasi yang didapat adalah yang terbaru dan relevan dengan judul jurnal.
3. Menetapkan kriteria inklusi dan eksklusi. Inklusi meliputi jurnal yang membahas secara langsung klasifikasi untuk prediksi banjir, sedangkan eksklusi mencakup jurnal yang tidak memiliki kontribusi penelitian atau tidak relevan dengan topik yang berupa ulasan atau komentar tanpa kontribusi penelitian. Dengan menggunakan kriteria ini diharapkan 15 referensi jurnal yang terpilih berkualitas untuk dianalisis lebih lanjut
4. Mengumpulkan data, data yang dikumpulkan dengan cara membaca dan mengumpulkan informasi penting dari setiap jurnal referensi. Yang meliputi data dan metode yang digunakan pada jurnal, serta hasil dari jurnal tersebut. Data yang diperoleh 15 jurnal referensi akan dibuat kedalam tabel untuk dianalisis lebih lanjut.
5. Menganalisis data pendekatan analisis ini bertujuan untuk memberikan wawasan tentang klasifikasi data dalam data mining. Analisis dilakukan dengan membaca dan membandingkan setiap hasil dari 15 jurnal referensi, untuk menarik kesimpulan umum mengenai praktik terbaik dalam klasifikasi data pada data mining.

3. HASIL DAN PEMBAHASAN

3.1 Tabel Hasil Referensi

Dari 15 jurnal referensi yang ada dapat dikumpulkan data berupa : data dan metode yang di gunakan dalam tiap jurnal referensi tersebut yang di buat dalam tabel di bawah ini.

No	Judul	Tahun Terbit Dan Penerbit	Data	Metode	Hasil
1.	Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot	M. B. A. Darmawan, F. Dewanta, dan S. Astuti (Mei 2023)	Data yang digunakan dalam penelitian ini berasal dari Balai Besar Wilayah Sungai (BBWS) Citarum, mencakup dua parameter utama yaitu tinggi muka air sungai dan intensitas curah hujan, dengan total 1460 data dari tahun 2015 hingga 2018. Label klasifikasi banjir dibagi menjadi tiga kategori: aman (0), siaga 1 (1), dan siaga 2 (2) berdasarkan ambang batas gabungan tinggi air dan curah hujan yang direkomendasikan oleh BBWS, BMKG, dan wawancara tokoh masyarakat.	Metode yang digunakan adalah perbandingan tiga algoritma klasifikasi, yaitu Decision Tree, Random Forest, dan Naïve Bayes. Proses penelitian dilakukan melalui beberapa tahap: preprocessing data, eksplorasi data (EDA), pelatihan model, pengujian, dan evaluasi performa menggunakan metrik accuracy, precision, recall, F1-score, serta waktu komputasi. Pengujian dilakukan sebanyak tiga kali dengan rasio pembagian data 8:2, 7:3, dan 6:4 antara training dan testing.	Hasil dari penelitian menunjukkan bahwa algoritma Random Forest memiliki performa terbaik dibandingkan metode lainnya. Rata-rata nilai evaluasi model yang diperoleh oleh Random Forest adalah akurasi 99,05%, precision 97,91%, recall 99,18%, F1-score 98%, serta rata-rata waktu komputasi 0,2561 detik. Meskipun Naïve Bayes unggul dalam waktu komputasi tercepat, namun akurasinya berada di bawah Random Forest dan Decision Tree. Penelitian ini menyimpulkan bahwa Random Forest adalah metode paling andal untuk prediksi banjir pada studi kasus ini, dan merekomendasikan penelitian lanjutan dengan data lebih banyak serta pendekatan algoritma yang lebih kompleks.
2.	Prediksi penyebaran virus COVID-19 dari hasil PCR menggunakan Metode Naïve Bayes	Hetty Rohayani, Sitti Nur Alam, Muhammad Fauzi dan Rico Rico J. Comput. Syst. Informatics, (Des 2022)	Data yang digunakan dalam penelitian ini adalah hasil pemeriksaan PCR pasien yang diambil dari instansi kesehatan atau laboratorium medis. Data tersebut mencakup status positif/negatif COVID-19 serta beberapa atribut gejala klinis yang dialami pasien seperti suhu tubuh,	Metode Klasifikasi yang digunakan adalah Naïve Bayes, salah satu teknik data mining dalam machine learning yang mengandalkan pendekatan probabilistik berdasarkan teorema bayes. Metode ini dipilih karena keunggulannya dalam	Hasil pengujian menunjukan bahwa metode Naïve Bayes mampu mengklasifikasikan data dengan tingkat akurasi yang cukup tinggi. Berdasarkan uji coba terhadap dataset yang digunakan, model ini menghasilkan akurasi klasifikasi sebesar sekitar

			batuk, pilek, sakit tenggorokan, sesak napas, dan riwayat kontak. Data yang digunakan berbentuk dataset dengan label kelas dan fitur gejala untuk proses klasifikasi.	memproses data dengan jumlah fitur banyak dan kemampuan prediksi yang baik walau dengan data terbatas.	92% dengan nilai precision dan recall yang juga menunjukkan performa baik.
3.	Penerapan Data Mining untuk Peringatan Dini Banjir Menggunakan Metode Klustering K-Means (Studi Kasus Kota Padang)	I. Nozomi (2023)	Penelitian ini menggunakan data cuaca yang diperoleh dari BMKG, tepatnya dari Stasiun Klimatologi Sicincin di Kabupaten Solok Selatan. Data yang dikumpulkan mencakup 547 record dari tahun 2022 hingga 2023, dengan tujuh parameter utama yaitu curah hujan, suhu minimum, suhu maksimum, suhu rata-rata, kelembaban, kecepatan angin, dan arah angin. Seluruh parameter tersebut menjadi dasar dalam menentukan tingkat risiko terjadinya banjir di wilayah studi.	Metode yang digunakan dalam penelitian ini adalah <i>klustering K-Means</i> , yaitu salah satu teknik <i>unsupervised learning</i> yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan nilai. Dalam studi ini, peneliti membagi data ke dalam tiga cluster yang merepresentasikan tingkat bahaya banjir: rendah, sedang, dan tinggi. Proses klustering dilakukan melalui beberapa tahapan, dimulai dari penentuan jumlah cluster dan centroid awal secara acak, perhitungan jarak antar data dan centroid menggunakan <i>Euclidean Distance</i> , lalu dilanjutkan dengan iterasi hingga posisi cluster stabil. Perhitungan dilakukan baik secara manual maupun dengan bantuan software RapidMiner v.5.3.	Dari hasil pengolahan data, diperoleh bahwa data dapat terbagi ke dalam tiga kelompok yang masing-masing mencerminkan tingkat peringatan dini banjir. Cluster 0 (rendah) terdiri dari 396 record yang dikategorikan sebagai peringatan “waspada”, cluster 1 (sedang) terdiri dari 145 record dengan peringatan “siaga”, dan cluster 2 (tinggi) terdiri dari 6 record yang termasuk peringatan “awas”. Hasil klustering ini menunjukkan kesesuaian antara perhitungan manual dan hasil dari RapidMiner, membuktikan bahwa metode K-Means cukup efektif digunakan dalam sistem peringatan dini bencana banjir. Penelitian ini merekomendasikan bahwa sistem yang dikembangkan dapat digunakan oleh instansi pemerintah dan lembaga kebencanaan untuk meminimalisir dampak banjir secara lebih cepat dan tepat.
4.	Perbandingan Metode Data Mining untuk Prediksi Banjir dengan Algoritma Naïve Bayes dan KNN	C. Cumel, D. Zamri, R. Rahmaddeni, dan S. Syamsurizal (Agustus 2022)	Data yang digunakan merupakan data sekunder yang diperoleh dari database BPBD DKI Jakarta, dikumpulkan mulai tanggal 1 Januari 2020 hingga 12 Juli 2020, dan diunduh melalui situs www.kaggle.com .	Metode yang digunakan Penelitian ini menggunakan pendekatan <i>data mining</i> dengan proses <i>Knowledge Discovery in Databases (KDD)</i> yang mencakup tahapan	Hasil penelitian menunjukkan bahwa algoritma kNN memberikan performa terbaik dalam memprediksi banjir, dengan akurasi sebesar 88,94% dan error sebesar 11,06%. Sebaliknya,

			<p>Dataset tersebut berisi informasi tentang status tujuh pintu air utama: Katulampa, Pos Depok, Manggarai, Istiqlal, Jembatan Merah, Flusing Ancol, dan Marina Ancol. Variabel target dalam penelitian ini adalah status banjir (banjir/tidak banjir), sementara variabel prediktor adalah debit air di masing-masing pos tersebut yang dikategorikan dalam rentang 0–1000 m³/s.</p>	<p>seleksi data, preprocessing, transformasi, data mining, dan evaluasi. Dua metode klasifikasi yang dibandingkan adalah Naïve Bayes dan k-Nearest Neighbor (kNN). Pengolahan data dilakukan menggunakan perangkat lunak RapidMiner 9.10.008, dan validasi model menggunakan teknik 10-fold cross-validation. Naïve Bayes mengandalkan probabilitas bersyarat antar variabel, sementara kNN menghitung jarak antar data menggunakan rumus Euclidean untuk menentukan kelas terdekat.</p>	<p>Naïve Bayes menghasilkan akurasi sebesar 74,36% dan error 25,64%. Kedua model mampu memprediksi dengan tingkat kinerja yang masuk dalam kategori baik dan cukup, namun secara keseluruhan kNN dinyatakan lebih unggul. Oleh karena itu, peneliti merekomendasikan penggunaan algoritma kNN untuk sistem prediksi banjir berbasis data mining pada kasus ini.</p>
5.	<p>Estimasi Curah Hujan Menggunakan Klasifikasi Data Mining dengan Algoritma Naive Bayes</p>	<p>N. N. Salsabila, J. Rifaldi, N. Nuris, dan S. Diantika (2024)</p>	<p>menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Kota Jakarta Timur. Data yang digunakan merupakan data curah hujan bulanan dari tahun 2019 hingga 2022 yang dikumpulkan dari Stasiun Meteorologi Halim Perdana Kusuma. Variabel yang dianalisis dalam penelitian ini meliputi jumlah hari hujan, kecepatan angin, kelembapan udara, suhu, dan tekanan udara. Data diklasifikasikan ke dalam empat kategori curah hujan: rendah, sedang, tinggi, dan ekstrem.</p>	<p>Penelitian ini menggunakan metode klasifikasi <i>Naive Bayes</i>, khususnya variasi <i>Gaussian Naive Bayes</i> karena data yang digunakan bersifat numerik dan diasumsikan memiliki distribusi normal. Proses klasifikasi dilakukan melalui beberapa tahapan, mulai dari seleksi dan pembersihan data, transformasi data, pembagian data menjadi data latih dan uji dengan rasio 90:10, hingga pembuatan model dan evaluasi menggunakan confusion matrix.</p>	<p>Hasil yang diperoleh menunjukkan bahwa tingkat akurasi model mencapai 80%, baik melalui perhitungan manual maupun dengan bantuan perangkat lunak RapidMiner. Meskipun akurasi secara umum tergolong tinggi, peneliti menemukan bahwa prediksi pada kelas curah hujan “tinggi” memiliki performa yang kurang optimal akibat ketidakseimbangan jumlah data antar kelas. Oleh karena itu, disarankan untuk menggunakan data yang lebih besar dan merata serta membandingkan dengan metode klasifikasi lain guna meningkatkan akurasi prediksi.</p>
6.	<p>Prediksi Bencana Banjir</p>	<p>H. Sharfina, P. Y. Utami,</p>	<p>Data yang digunakan dalam penelitian ini</p>	<p>Metode yang digunakan dalam penelitian ini</p>	<p>Hasil penelitian menunjukkan bahwa</p>

	Menggunakan Algoritma Deep Learning H2O Berdasarkan Data Curah Hujan	dan I. Fakhruzi (Des 2023)	merupakan data sekunder yang berasal dari dua sumber utama: BMKG dan BPBD. Data iklim diperoleh dari lima stasiun klimatologi di Kalimantan Barat, yaitu Kabupaten Sintang, Melawi, Kapuas Hulu, Sambas, dan Ketapang. Data yang dikumpulkan meliputi periode 2019–2022, dengan total sebanyak 357 data bencana banjir. Parameter yang digunakan untuk prediksi antara lain: curah hujan, temperatur rata-rata, kelembapan rata-rata, arah angin saat kecepatan maksimum, dan lama penyinaran matahari. Data dilabeli status banjir atau tidak banjir untuk data tahun 2019–2021 sebagai data latih, sementara data tahun 2022 digunakan sebagai data uji tanpa label.	adalah algoritma Deep Learning H2O, sebuah framework machine learning yang dikembangkan untuk mendukung analisis data berskala besar. Penelitian ini mengikuti tahapan standar pemodelan data mining, mulai dari pengumpulan, seleksi, praproses, transformasi, pembagian data, pembuatan model, hingga evaluasi performa. Data diproses menggunakan aplikasi RapidMiner Studio 10.1, dan dibagi dalam rasio 80:20 untuk training dan testing. Model deep learning dibangun dengan struktur hidden layer 100, 50, 100 neuron dan diuji menggunakan beberapa fungsi aktivasi: rectifier, tanh, dan maxout dengan variasi epoch dari 10 hingga 55. Evaluasi model dilakukan menggunakan metrik akurasi dan RMSE (Root Mean Square Error).	model deep learning H2O mampu memprediksi bencana banjir dengan baik berdasarkan data curah hujan. Dari sepuluh kali pengujian dengan berbagai epoch dan fungsi aktivasi, hasil terbaik diperoleh pada penggunaan fungsi aktivasi rectifier dengan jumlah epoch sebanyak 30, yang menghasilkan akurasi sebesar 90,54% dan RMSE senilai 0,087. Hasil prediksi juga dibandingkan dengan data aktual di lima kabupaten, menunjukkan bahwa prediksi cukup akurat terutama di daerah yang masuk dalam kategori hujan sedang. Penelitian ini juga berhasil memetakan daerah-daerah rawan dan tidak rawan banjir berdasarkan hasil klasifikasi model, yang dapat menjadi referensi awal bagi pemerintah daerah dalam mengambil tindakan mitigasi.
7.	Algoritma J48 untuk Pemodelan Sistem Prediksi Tingkat Kerawanan Banjir dengan Visualisasi Web GIS	N. Kurnia Asih (2020)	Penelitian ini menggunakan data spasial dan atribut yang berkaitan dengan faktor-faktor penyebab banjir, khususnya di wilayah Jakarta timur. Data utama yang dianalisis mencakup curah hujan rata-rata tahunan, kemiringan lereng, jenis penggunaan lahan (land use), serta jarak antar kelurahan dengan sungai terdekat. Sumber data diperoleh dari instansi	Metode yang digunakan : Penelitian ini menerapkan algoritma klasifikasi Decision Tree J48 (implementasi dari C4.5) sebagai metode utama dalam pemodelan prediksi. Algoritma J48 dipilih karena kemampuannya dalam menghasilkan model berupa pohon keputusan yang mudah dipahami dan divisualisasikan. Proses	Hasil Penelitian: Hasil dari pemodelan menunjukkan bahwa algoritma J48 mampu mengklasifikasikan tingkat kerawanan banjir dengan akurasi sebesar 86% berdasarkan data uji. Model yang dihasilkan memperlihatkan bahwa faktor curah hujan dan jarak ke sungai menjadi variabel dominan dalam memengaruhi tingkat kerawanan. Selain itu, sistem Web GIS yang

			terkait seperti Badan Meteorologi Klimatologi Dan Geofisika (BMKG), Bappeda, dan peta digital dari Badan Informasi Geospasial. Setiap kelurahan diklasifikasikan tingkat kerawanannya ke dalam kategori "Rendah", "Sedang", atau "Tinggi", yang nantinya digunakan sebagai label target dalam proses klasifikasi.	pelatihan model dilakukan menggunakan perangkat lunak WEKA dengan data kelurahan sebagai instance dan empat parameter sebagai atribut. Hasil dari algoritma ini kemudian diintegrasikan ke dalam Web GIS, memungkinkan pengguna untuk melihat hasil klasifikasi langsung pada peta interaktif. Visualisasi tersebut dibuat dengan menggabungkan sistem informasi geografis dan bahasa pemrograman web.	dikembangkan berhasil menampilkan prediksi kerawanan banjir secara visual, memudahkan pihak pengguna untuk melakukan monitoring dan pengambilan keputusan berbasis lokasi. Dengan hasil ini, sistem dinilai cukup efektif untuk digunakan sebagai alat bantu dalam manajemen risiko bencana banjir di wilayah perkotaan.
8.	Prediksi Banjir Menggunakan ANFIS-PCA sebagai Peringatan Dini Bencana Banjir	A. Rachmawardani, S. K. Wijaya, Prawito, dan A. Sopaheluwan (Apr 2024)	Data yang Digunakan: Penelitian ini memanfaatkan data historis curah hujan dan ketinggian muka air dari dua sumber utama: data dari stasiun hujan otomatis (ARG - Automatic Rain Gauge) dan data satelit CHIRPS, mencakup periode tahun 2014-2020 di wilayah DKI Jakarta. Proses praproses meliputi imputasi rata-rata untuk data hilang, normalisasi skala, dan penggunaan metode interquartile range (IQR) untuk menangani outlier. Data tersebut dibagi untuk training dan validasi dengan rasio 80:20.	Metode yang Digunakan : Model utama adalah ANFIS (Adaptive Neuro-Fuzzy Inference System) yang diperkuat oleh Principal Component Analysis (PCA) untuk mengurangi dimensi dan efisiensi komputasi. ANFIS menggabungkan logika fuzzy dan jaringan saraf, sedangkan PCA menyederhanakan input variabel tanpa mengorbankan informasi signifikan. Performa model dibandingkan antara ANFIS dengan dan tanpa PCA, serta antara data dari ARG dan CHIRPS, untuk menilai kontribusi PCA dan jenis sumber data.	Hasil Penelitian: Model ANFIS-PCA yang dilatih menggunakan data stasiun hujan ARG menunjukkan hasil terbaik, dengan RMSE $\approx 0,13$, MAE $\approx 0,12$, dan $R^2 \approx 0,82$. Tanpa PCA, ANFIS menghasilkan kesalahan jauh lebih besar: RMSE $\approx 6,3$, MAE $\approx 6,2$, dan $R^2 \approx 0,74$. Saat menggunakan data satelit CHIRPS, model ANFIS-PCA mencatat RMSE $\approx 4,8$, MAE $\approx 2,0$, dan $R^2 \approx 0,55$, sementara ANFIS tanpa PCA jauh melemah (RMSE $\approx 30,14$, MAE $\approx 24,05$, $R^2 \approx 0,42$). Hasil ini menunjukkan secara jelas bahwa ANFIS-PCA meningkatkan akurasi, terutama bila menggunakan data ground station, dan unggul jauh dibanding kombinasi metode dan dataset lainnya.

9.	Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning	N. S. Fatonah (2021)	<p>Penelitian ini menggunakan data historis banjir di Indonesia, khususnya rentang beberapa tahun terakhir saat banjir terjadi hampir setiap tahun. Data ini mencakup kejadian banjir di bulan-bulan sebelumnya yang dijadikan referensi untuk memprediksi pada tahun berjalan atau tahun berikutnya. Meskipun jurnal tidak merinci jumlah sampel atau variabel numerik lainnya, model dibangun menggunakan dataset yang tersedia dari peristiwa banjir terdahulu .</p>	<p>Metode yang Digunakan : Peneliti mengaplikasikan algoritma Naïve Bayes sebagai metode klasifikasi utama dalam kerangka data mining untuk mendeteksi kemungkinan terjadinya bencana banjir. Dengan pendekatan ini, model dilatih menggunakan pola historis, lalu divalidasi untuk menentukan akurasi deteksi. Teknik ini diterapkan melalui software WEKA atau lingkungan data mining serupa, meskipun alat konkret yang digunakan tidak disebut secara spesifik .</p>	<p>Hasil Penelitian : Model deteksi bencana banjir berhasil mencapai akurasi sebesar 76,73% dalam memprediksi adanya banjir, menunjukkan performa cukup baik dari algoritma Naïve Bayes dalam konteks ini. Namun hasilnya menunjukkan keterbatasan signifikan: sistem belum mampu menentukan hari pasti terjadinya banjir—hanya mendeteksi potensi munculnya bencana dalam periode tertentu, bukan tanggal tepatnya .</p>
10.	Prediksi Banjir di DKI Jakarta dengan Menggunakan Algoritma K-Means dan Random Forest	W. H. Sasoko, E. W. Pujiharto, R. Haris, A. Y. Kania, Kusrini, dan Kusnawi (Apr 2024)	<p>Data yang Digunakan</p> <p>Penelitian ini memanfaatkan dataset historis banjir di DKI Jakarta dengan parameter utama tinggi muka air dan lama genangan setelah hujan, didapat dari catatan kejadian banjir tahun 2020 yang dilaporkan oleh media dan instansi setempat. Data tersebut kemudian diproses untuk clustering area rawan menggunakan K-Means, serta digunakan sebagai fitur dalam model Random Forest untuk klasifikasi zona kerawanan banjir .</p>	<p>Metode yang Digunakan</p> <p>Model pertama adalah K-Means clustering, yang digunakan untuk mengelompokkan wilayah DKI Jakarta ke dalam tiga cluster berdasarkan tinggi dan durasi genangan. Penentuan jumlah cluster dilakukan dengan metode elbow, dan hasil clustering dievaluasi dengan silhouette score sebesar 0,60 . Setelah itu, hasil clustering dijadikan label dan fitur input (tinggi genangan, lama genangan) dipakai untuk melatih model Random Forest, dengan tujuan membangun sistem deteksi/prediksi kerawanan banjir.</p>	<p>Hasil clustering menunjukkan distribusi wilayah sebagai berikut: Cluster 0 (kerawanan rendah) sebanyak 1.207 wilayah, Cluster 1 (sedang) 705 wilayah, dan Cluster 2 (tinggi) 178 wilayah . Model Random Forest yang dilatih berhasil mengklasifikasikan ketiga zona kerawanan dengan akurasi tinggi (sekitar 99%) dan ketahanan model yang kuat—menunjukkan performa unggul dalam mendeteksi wilayah rawan banjir berdasarkan data genangan .</p>

11.	Flood Forecasting Using Machine Learning Methods	F.-J. Chang, K. Hsu, dan L.-C. Chang (eds.) (2020)	Buku ini menghimpun hasil-hasil penelitian terkini dalam bidang pemodelan dan prediksi banjir menggunakan machine learning. Dataset yang digunakan dalam berbagai bab mencakup data hidrologi seperti curah hujan, tinggi muka air sungai, volume aliran sungai, kelembapan tanah, dan data dari sensor IoT. Sebagian besar data berasal dari negara-negara Asia seperti Taiwan, India, dan Tiongkok, serta data open-source dari satelit global (TRMM dan GPM).	Metode yang dibahas mencakup teknik Recurrent Neural Network (RNN), Nonlinear Autoregressive with Exogenous inputs (NARX), Support Vector Machine (SVM), Artificial Neural Network (ANN), Ensemble Learning, dan Clustering. Buku ini juga memuat penerapan sistem IoT dalam pengumpulan data real-time untuk keperluan deteksi dan prediksi banjir.	Hasil yang dikaji dalam buku ini menunjukkan bahwa model RNN dan RNARX efektif untuk memprediksi kondisi banjir dengan interval waktu tertentu, khususnya dalam kasus banjir musiman. Model berbasis ANN dan SVM terbukti cukup akurat dalam memodelkan hubungan non-linear antara variabel cuaca dan kejadian banjir. Clustering dipadukan dengan ANN digunakan untuk memetakan risiko banjir berbasis wilayah dengan tingkat keakuratan tinggi.
12.	Applications of Machine Learning in Hydroclimatology	R. Srivastav dan P. C. Nayak (2024)	Buku ini menggunakan data dari sistem hidroklimatologi di Sungai Han, Korea Selatan, mencakup informasi dari empat bendungan utama: Chungju, Soyonggang, Hwacheon, dan Paldang. Data meliputi curah hujan harian, debit air masuk dan keluar, kapasitas tampungan, serta tinggi muka air selama periode 2006–2023.	Berbagai metode pembelajaran mesin digunakan, termasuk Support Vector Regression (SVR), Multilayer Perceptron Neural Network (MLPNN), Extreme Gradient Boosting (XGBoost), serta teknik preprocessing seperti Min-Max Scaling dan Standardization. Buku ini juga membahas kombinasi metode statistik klasik dan ML untuk peningkatan akurasi prediksi.	Model MLPNN dan XGBoost memiliki performa terbaik dalam memprediksi kondisi aliran ekstrem dan banjir di daerah tangkapan air. Dengan pendekatan ini, akurasi prediksi meningkat hingga 93% dibanding metode tradisional. Buku ini juga menekankan pentingnya pemilihan fitur dan data preprocessing sebagai faktor penting dalam meningkatkan kinerja model.
13.	Flood prediction with time series data mining: Systematic review	D. K. Hakim, R. Gernowo, dan A. W. Nirwansyah (Jun 2024)	Penulis menelaah 109 artikel ilmiah yang diterbitkan antara 2010 hingga 2023, yang menggunakan data deret waktu (time series) untuk prediksi banjir. Sumber data mencakup curah hujan, debit sungai, elevasi tanah, kelembapan udara, kecepatan angin, serta data geografis (DEM –	Penelitian ini adalah studi literatur sistematis menggunakan metode PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Penelitian diklasifikasikan menjadi dua pendekatan: univariate dan multivariate. Evaluasi model	Penelitian menunjukkan bahwa 51,38% studi masih menggunakan model univariat, meskipun pendekatan multivariat semakin meningkat. Algoritma Machine Learning seperti LSTM, Random Forest, dan Gradient Boosting menunjukkan hasil yang unggul dibandingkan model statistik

			Digital Elevation Models). Analisis dilakukan terhadap studi dari wilayah Asia Tenggara, Eropa, dan Amerika Selatan.	dilakukan berdasarkan metrik RMSE (Root Mean Square Error), MAE (Mean Absolute Error), dan MAPE (Mean Absolute Percentage Error).	konvensional. Studi juga menyimpulkan bahwa kombinasi metode klasifikasi dan optimasi parameter menghasilkan akurasi prediksi yang lebih tinggi.
14.	Classification of Flood Levels using Random Forest and Support Vector Machine Algorithms	L. S. Qamarani dan M. Riasetiawan (2024)	Dataset terdiri atas data tinggi muka air, curah hujan harian, dan data geografis dari Badan Nasional Penanggulangan Bencana (BNPB) dan BMKG, yang dikumpulkan dari 5 stasiun pengamatan banjir di Indonesia. Data mencakup rentang tahun 2017–2022 dengan lebih dari 1.200 data rekaman.	Dua metode klasifikasi digunakan yaitu Random Forest (RF) dan Support Vector Machine (SVM). Dataset dibersihkan, dilakukan seleksi fitur, normalisasi, dan validasi silang 10-fold. RF digunakan untuk menangani hubungan non-linear antara variabel, sedangkan SVM untuk memisahkan kategori level banjir berdasarkan hyperplane.	Hasil evaluasi menunjukkan bahwa Random Forest memberikan akurasi sebesar 92%, sedangkan SVM mencapai 89%. RF lebih unggul dari sisi efisiensi waktu dan interpretabilitas model. Penulis menyarankan penggunaan RF dalam sistem peringatan dini bencana banjir berbasis klasifikasi tingkat bahaya.
15.	Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5	I. Risnawati, S. Rahma, F. Kusuma, N. Salsabila, A. Nourmansyah, A. Farhan, dan E. Irfiani (Des 2023)	Penelitian yang dilakukan oleh Risnawati et al. (2023) menggunakan data cuaca historis yang diperoleh dari sumber terpercaya, yaitu situs resmi BMKG dan NASA, dengan cakupan wilayah Jawa Tengah dan Jawa Timur. Data tersebut terdiri dari 60 sampel yang mencerminkan kondisi meteorologis aktual, yang meliputi beberapa parameter utama seperti temperatur udara, kecepatan angin, kelembapan udara, serta curah hujan sebagai variabel target yang akan diklasifikasikan. Setiap data memiliki label kategori berdasarkan tingkat intensitas curah hujan, seperti ekstrem, sangat tinggi, dan lainnya, yang kemudian digunakan untuk	metode yang digunakan adalah algoritma C4.5, yaitu salah satu algoritma klasifikasi pada data mining yang bekerja dengan membangun struktur pohon keputusan. Proses pembentukan pohon keputusan dimulai dari perhitungan nilai entropy dan information gain untuk setiap atribut, guna menentukan atribut paling informatif sebagai akar pohon. Setelah pohon terbentuk, model diuji dengan menggunakan teknik validasi silang K-Fold (K=10), yang bertujuan untuk memastikan bahwa hasil klasifikasi yang diperoleh cukup andal dan tidak bergantung	Hasil dari penerapan metode tersebut menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan data curah hujan yang berpotensi menimbulkan banjir dengan tingkat akurasi yang cukup tinggi, yaitu sebesar 83,33%. Selain itu, ditemukan bahwa atribut curah hujan memiliki nilai informasi paling dominan dibandingkan atribut lainnya, sehingga berada di posisi akar pohon dalam model keputusan. Model ini dinilai efektif sebagai alat bantu awal untuk mendeteksi wilayah dengan risiko banjir, dan berpotensi dikembangkan lebih lanjut untuk sistem peringatan dini bencana berbasis data cuaca.

			mengidentifikasi potensi banjir di daerah-daerah rawan. Dalam penelitian ini	pada subset data tertentu saja. Hasil	
--	--	--	--	---------------------------------------	--

3.2 Pembahasan

Dalam studi literatur sistematis ini mengenai metode klasifikasi data mining untuk prediksi potensi banjir, penulis telah mengkaji 15 jurnal ilmiah yang berisi beragam pendekatan algoritmik dan strategi penerapan. Dari hasil kajian, ditemukan bahwa terdapat berbagai metode klasifikasi yang digunakan, seperti Naïve Bayes, Random Forest, K-Means, Decision Tree (J48 dan C4.5), hingga pendekatan canggih seperti Deep Learning H2O dan ANFIS-PCA. Masing-masing metode memiliki karakteristik unik dalam menangani variabel-variabel penting seperti curah hujan, topografi wilayah, kelembapan tanah, serta waktu terjadinya banjir. Dalam proses menilai metode klasifikasi terbaik untuk prediksi banjir, penting untuk memperhatikan efektivitas hasil prediksi, konteks wilayah studi, dan akselerasi komputasi yang digunakan. Beberapa jurnal mengangkat metode yang sama, namun menghasilkan performa yang berbeda tergantung pada preprocessing data, sumber data, serta konfigurasi algoritma. Oleh karena itu, penilaian terhadap praktik terbaik tidak bisa dilakukan secara umum, melainkan harus mempertimbangkan hasil akurasi, sensitivitas model terhadap noise, serta relevansi data terhadap kondisi geografis lokasi penelitian.

Salah satu pendekatan yang dianggap menonjol dalam literatur adalah Random Forest, seperti pada penelitian prediksi banjir di DKI Jakarta dan klasifikasi tingkat banjir di India. Algoritma ini memberikan akurasi tinggi dan ketahanan terhadap overfitting. Hasil prediksi mampu mengelompokkan daerah rawan banjir dengan cukup baik berdasarkan parameter tinggi genangan dan lama banjir. Selain itu, metode ini juga diakui karena kemampuannya menangani data besar dengan kompleksitas tinggi serta fitur penting yang terseleksi secara otomatis. Metode Naïve Bayes juga menjadi salah satu pendekatan populer dalam studi ini, karena sederhana namun efektif, khususnya dalam prediksi berbasis probabilistik.

Di sisi lain, metode K-Means, meskipun awalnya bukan metode klasifikasi melainkan klusterisasi, digunakan dalam kombinasi dengan model klasifikasi seperti Random Forest untuk memperbaiki pembentukan kelas awal. Pendekatan ini terbukti membantu dalam mengidentifikasi pola kerawanan banjir berdasarkan kelompok wilayah yang homogen. Contohnya diterapkan dalam prediksi banjir di Padang dan Jakarta, dengan hasil visualisasi spasial yang mendukung pengambilan keputusan pemerintah daerah. Sementara itu, pendekatan berbasis deep learning, seperti algoritma H2O, dan metode ANFIS-PCA, juga menunjukkan kemajuan signifikan, terutama dalam menangani data kompleks seperti data hidrologi berbasis waktu. Kedua pendekatan ini unggul dalam mendeteksi pola banjir jangka pendek hingga menengah, serta memberikan hasil prediksi yang akurat meskipun membutuhkan sumber daya komputasi yang lebih besar.

4. KESIMPULAN

Studi literatur sistematis ini membuktikan bahwa metode klasifikasi data mining merupakan elemen penting dalam membangun sistem prediksi potensi banjir yang akurat, efisien, dan berbasis data. Melalui telaah mendalam terhadap 15 jurnal ilmiah, ditemukan bahwa algoritma seperti Random Forest, Naïve Bayes, dan ANFIS-PCA memiliki keunggulan masing-masing dalam konteks penerapan yang berbeda dan beragam. Hasil studi menunjukkan bahwa Random Forest sangat efektif dalam meningkatkan efisiensi prediksi dan ketepatan klasifikasi wilayah rawan banjir, seperti yang diterapkan pada studi kasus di wilayah DKI Jakarta. Sementara itu, Naïve Bayes menawarkan kecepatan tinggi dan efisiensi pemrosesan data, sehingga cocok digunakan dalam sistem yang membutuhkan hasil cepat dan sederhana, terutama dengan jumlah data terbatas dan variabel input yang sedikit. Di sisi lain, pendekatan ANFIS-PCA memberikan fleksibilitas tinggi dan akurasi yang lebih baik, terutama ketika digunakan pada sistem peringatan dini berbasis spasial dan data multivariat yang kompleks.

Kesimpulan dari studi ini menegaskan bahwa pemilihan metode klasifikasi data mining harus disesuaikan dengan karakteristik data, kompleksitas permasalahan, serta kebutuhan sistem informasi yang dibangun. Untuk memilih metode yang paling tepat, penting mempertimbangkan faktor-faktor seperti jumlah data, tingkat konsistensi yang dibutuhkan, potensi gangguan jaringan, serta jenis prediksi yang diinginkan. Penelitian selanjutnya disarankan untuk mengeksplorasi integrasi atau kombinasi metode klasifikasi serta pemanfaatan teknologi berbasis waktu nyata guna meningkatkan adaptivitas terhadap situasi bencana yang dinamis dan tidak terduga.

5. SARAN

Penelitian selanjutnya disarankan untuk mengembangkan kombinasi metode klasifikasi atau model ensemble guna meningkatkan akurasi prediksi banjir. Selain itu, penggunaan data spasial dan temporal yang lebih lengkap seperti curah hujan, topografi, dan penggunaan lahan juga perlu dipertimbangkan. Implementasi sistem prediksi berbasis real-time dan integrasi dengan teknologi seperti IoT serta sistem peringatan dini akan sangat bermanfaat dalam menghadapi bencana secara adaptif. Diperlukan pula kolaborasi lintas sektor agar hasil penelitian dapat diterapkan secara nyata dalam mitigasi risiko banjir.

REFERENCES

- [1] M. B. A. Darmawan, F. Dewanta, dan S. Astuti, “Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot,” *TELKA: Jurnal Telekomunikasi, Elektronika, Komputasi, dan Kontrol*, vol. 9, no. 1, pp. 52–61, Mei 2023.
- [2] H. Rohayani, S. N. Alam, M. Fauzi, dan Rico, “Prediksi Penyebaran Virus COVID-19 Dari Hasil PCR Menggunakan Metode Naïve Bayes,” **Journal of Computer System and Informatics (JoSYC)**, vol. 4, no. 1, hlm. 109-115, Nov. 2022, doi: 10.47065/josyc.v4i1.2577.

- [3] I. Nozomi, “Penerapan Data Mining untuk Peringatan Dini Banjir Menggunakan Metode Klastering K-Means (Studi Kasus Kota Padang),” *Jurnal Sains Informatika Terapan (JSIT)*, vol. 2, no. 2, pp. 39–44, 2023.
- [4] C. Cumel, D. Zamri, R. Rahmaddeni, dan S. Syamsurizal, “Perbandingan Metode Data Mining untuk Prediksi Banjir dengan Algoritma Naïve Bayes dan KNN,” *Prosiding Seminar Nasional Penelitian dan Pengabdian Masyarakat (SENTIMAS)*, vol. 2022, pp. 40–48, 25 Agustus 2022.
- [5] N. N. Salsabila, J. Rifaldi, N. Nuris, dan S. Diantika, “Estimasi Curah Hujan Menggunakan Klasifikasi Data Mining dengan Algoritma Naive Bayes,” *Jurnal LLPP Universitas Nurdin Hamzah Jambi*, pp. 56–58, 2024.
- [6] H. Sharfina, P. Y. Utami, dan I. Fakhruzi, “Prediksi Bencana Banjir Menggunakan Algoritma Deep Learning H2O Berdasarkan Data Curah Hujan,” *Jurnal Teknik Informatika dan Sistem Informasi (JATISI)*, vol. 10, no. 4, pp. 79–92, Des. 2023.
- [7] N. Kurnia Asih, “Algoritma J48 untuk Pemodelan Sistem Prediksi Tingkat Kerawanan Banjir dengan Visualisasi Web GIS,” Tugas Akhir, Program Studi Teknik Informatika, Universitas Mercu Buana, Jakarta, 2020.
- [8] A. Rachmawardani, S. K. Wijaya, Prawito, dan A. Sopaheluwakan, “Prediksi Banjir Menggunakan ANFIS-PCA sebagai Peringatan Dini Bencana Banjir,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Telekomunikasi dan Elektronika*, vol. 12, no. 2, pp. 335–351, Apr. 2024.
- [9] N. S. Fatonah, “Penerapan Deteksi Bencana Banjir Menggunakan Metode Machine Learning,” *Jurnal Format*, vol. 10, no. 2, pp. 119–124, 2021.
- [10] W. H. Sasoko, E. W. Pujiharto, R. Haris, A. Y. Kania, Kusrini, dan Kusnawi, “Prediksi Banjir di DKI Jakarta dengan Menggunakan Algoritma K-Means dan Random Forest,” *J-COM: Jurnal Informatika dan Teknologi Komputer*, vol. 5, no. 1, pp. 43–49, Apr. 2024.
- [11] F.-J. Chang, K. Hsu, dan L.-C. Chang (eds.), *Flood Forecasting Using Machine Learning Methods*, MDPI Books, 2020. [Online]. Available: <https://www.mdpi.com/books/reprint/1151-flood-forecasting-using-machine-learning-methods>
- [12] R. Srivastav dan P. C. Nayak, *Applications of Machine Learning in Hydroclimatology*, Springer, 2024. [Online]. Available: <https://link.springer.com/book/9783031644023>
- [13] D. K. Hakim, R. Gernowo, dan A. W. Nirwansyah, “Flood prediction with time series data mining: Systematic review,” *Natural Hazards Research*, vol. 4, no. 2, pp. 194–220, Jun. 2024. [Online]. Available: <https://doi.org/10.1016/j.nhres.2023.10.001>

- [14] L. S. Qamarani dan M. Riasetiawan, "Classification of Flood Levels using Random Forest and Support Vector Machine Algorithms," Indonesian Journal of Electronics and Instrumentation Systems, Universitas Gadjah Mada, vol. 4, no. 1, 2024. [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/97043>
- [15] I. Risnawati, S. Rahma, F. Kusuma, N. Salsabila, A. Nourmansyah, A. Farhan, dan E. Irfiani, "Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5," Jurnal INSAN (Journal of Information Systems Management Innovation), vol. 3, no. 2, pp. 78–84, Des. 2023. [Online]. Tersedia: <http://jurnal.bsi.ac.id/index.php/jinsan>