
PREDIKSI BANJIR DI KOTA MANADO MENGGUNAKAN METODE *RANDOM FOREST*

Jimmi Pirnando Sitepu¹⁾, Charles E.Mongi²⁾, Siska Ayu Widiana³⁾, Eliasta Ketaren⁴⁾

Program Studi Sistem Informasi

Fakultas Matematika dan Ilmu Pengetahuan

Alam Universitas Sam Ratulangi

Manado, Indonesia

email: jimmisitepu106@student.unsrat.ac.id¹⁾, charlesmongi@unsrat.ac.id²⁾, siskaginting@unsrat.ac.id³⁾,
eliasketaren@unsrat.ac.id⁴⁾

Abstrak

Banjir merupakan salah satu bencana alam yang sering terjadi di Kota Manado akibat curah hujan tinggi yang melebihi kapasitas penampungan air. Penelitian ini bertujuan untuk mengembangkan model prediksi banjir menggunakan algoritma Random Forest berdasarkan data cuaca historis. Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan memanfaatkan data cuaca dari BMKG Manado periode 2014-2024 yang mencakup delapan indikator cuaca. Data dibagi menjadi data latih (70%) dan data uji (30%) dengan implementasi model menggunakan 300 pohon keputusan. Hasil penelitian menunjukkan performa model yang sangat baik dengan akurasi konsisten di atas 93% dan mencapai puncak 97,97% pada April 2014. Evaluasi menggunakan confusion matrix memperlihatkan nilai precision, recall, dan F1-Score yang sempurna untuk kategori "Tidak Banjir" dan "Banjir Besar", serta nilai sangat baik untuk klasifikasi "Banjir Ringan" (98,40%) dan "Banjir Sedang" (98,59%). Kesalahan klasifikasi hanya terjadi pada kelas-kelas yang berdekatan, menunjukkan konsistensi model dalam membedakan tingkat keparahan banjir. Penelitian ini membuktikan bahwa metode Random Forest efektif untuk memprediksi kejadian banjir di Kota Manado dengan tingkat akurasi yang tinggi. Untuk penelitian selanjutnya, disarankan pengembangan model pembelajaran mesin alternatif, penambahan variabel prediktor seperti tutupan lahan dan topografi, serta pengembangan sistem peringatan dini berbasis GIS untuk meningkatkan komunikasi hasil prediksi kepada masyarakat.

Kata Kunci : Prediksi Banjir, Random Forest, Data Cuaca, Kota Manado, Machine Learning

1. Pendahuluan

Cuaca merupakan suatu kondisi udara di suatu tempat pada waktu yang relatif singkat, yang dinyatakan dengan berbagai parameter seperti curah hujan, suhu, tekanan udara, kecepatan angin, dan kelembaban udara[1]. Cuaca juga merupakan salah satu penyebab terjadinya bencana alam banjir di Indonesia yang disebabkan oleh beberapa parameter cuaca salah satunya yaitu curah hujan yang tinggi dan perubahan suhu yang tidak menentu[2]. Banjir dapat didefinisikan sebagai kondisi dimana volume air yang meluap tidak dapat ditampung oleh saluran-saluran air seperti waduk, sungai, danau, dan infrastruktur drainase lainnya[3]. Berdasarkan informasi dari website resmi BNPB (Badan Nasional Penanggulangan Bencana), Kota Manado merupakan salah satu kota yang pernah mengalami bencana banjir pada tanggal 23 Januari 2023. Banjir tersebut disebabkan oleh curah hujan yang tinggi di wilayah kota sehingga debit air di Sungai Tondano menjadi meluap.

Prediksi banjir di Kota Manado merupakan salah satu strategi mitigasi yang dapat mengatasi dampak bencana alam tersebut. Dengan memanfaatkan data cuaca historis, peneliti dapat memprediksi potensi banjir menggunakan beberapa parameter seperti curah hujan dan suhu yang terjadi setiap harinya. Data tersebut dapat diperoleh dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) Manado, dan selanjutnya akan diolah menjadi acuan untuk mendukung hasil prediksi banjir. Namun, beberapa faktor lain seperti pembangunan tanpa perencanaan yang baik dan perubahan iklim turut memperparah situasi ini, sehingga diperlukan suatu metode yang dapat memprediksi bencana alam banjir dengan akurasi yang tinggi agar upaya mitigasi dapat lebih efektif dilakukan.

Metode Random Forest salah satu algoritma pembelajaran mesin yang dapat digunakan untuk memprediksi banjir dengan akurasi tinggi. Metode ini menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang stabil dan akurat, sehingga efektif untuk digunakan dalam analisis data cuaca karena mampu bekerja dengan variabel independen yang banyak dan kompleks. Dengan memanfaatkan data cuaca seperti curah hujan dan suhu, penelitian ini bertujuan membangun model prediksi banjir di Kota Manado yang dapat membantu pihak-pihak berwenang dalam meminimalisir dampak dari banjir yang akan terjadi[4].

2. Landasan Teori Banjir

Banjir adalah salah satu bencana alam yang terjadi disebabkan oleh tingginya curah hujan (Balahanti et al., 2023). Cuaca merupakan salah satu penyebab terjadinya bencana alam banjir yang disebabkan oleh tingkat curah hujan yang tinggi sehingga intensitas air ikut meningkat yang mengakibatkan tempat penampungan air melebihi batas kapasitas[1]. Curah hujan merupakan hasil interaksi kompleks antara suhu, kelembapan, dan angin. Suhu menentukan potensi kondensasi, kelembapan menyediakan bahan baku uap air, dan angin menjadi mekanisme transportasi serta pengangkutan uap air ke atmosfer[5].

Machine Learning

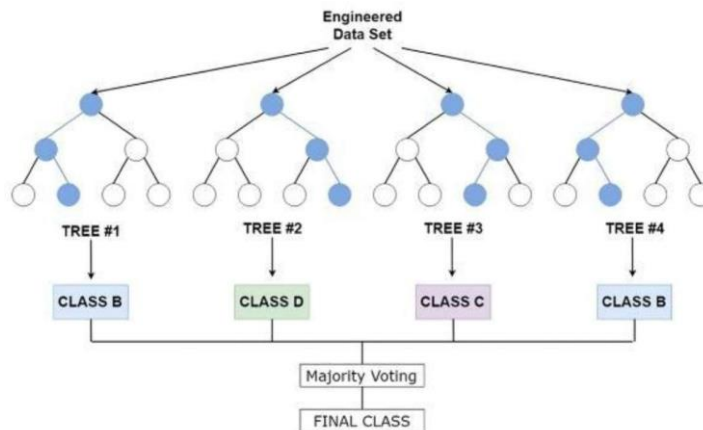
Machine Learning atau pembelajaran mesin adalah cabang dari kecerdasan buatan (Artificial Intelligence/AI) yang berfokus pada pengembangan algoritma dan model statistik yang memungkinkan sistem komputer untuk belajar dari data tanpa perlu diprogram secara eksplisit. Dengan memanfaatkan data historis, sistem ini dapat mengenali pola, membuat prediksi, dan mengambil keputusan secara otomatis[6].

Random Forest

Random Forest adalah algoritma supervised learning. Random Forest merupakan metode Klasifikasi dan regresi berdasarkan jumlah pohon Keputusan[4]. Dan merupakan salah satu algoritma dalam machine learning yang memberikan nilai akurasi terbaik, algoritma ini menggunakan pohon keputusan untuk melangsungkan proses seleksi yang dijalankan berdasarkan pada masing-masing data pada kelas yang berbeda[7].

Gambar 1 menggambarkan konsep kerja algoritma Random Forest, yang merupakan salah satu metode *ensemble learning* dalam *machine learning* yang umum digunakan untuk tugas klasifikasi maupun regresi[8]. Proses dimulai dari dataset yang telah melalui tahap rekayasa fitur (*engineered dataset*), yang kemudian digunakan sebagai input untuk membangun beberapa pohon keputusan (*decision trees*). Dalam ilustrasi ini, terdapat empat pohon keputusan (Tree 1 hingga Tree 4), di mana masing-masing pohon bekerja secara independen untuk mengenali pola dalam data. Setiap simpul (*node*) pada pohon mewakili titik pengambilan keputusan berdasarkan fitur tertentu, dengan simpul berwarna biru menunjukkan titik keputusan yang penting[8].

Setelah proses inferensi oleh masing-masing pohon, diperoleh hasil prediksi sebagai berikut: Tree 1 memprediksi kelas B, Tree 2 kelas D, Tree 3 kelas C, dan Tree 4 kembali memprediksi kelas B. Untuk menentukan prediksi akhir, algoritma menerapkan mekanisme *voting mayoritas*, di mana kelas yang paling sering dipilih oleh pohon-pohon tersebut akan menjadi hasil akhir. Dalam kasus ini, karena dua dari empat pohon memilih kelas B, maka kelas B ditetapkan sebagai kelas akhir (*final class*) dari prediksi yang dihasilkan oleh Random Forest[8].



Gambar 1. Algoritma sederhana *Random Forest*[8]

Software Matematika

Software Matematika merupakan salah satu software yang dirancang untuk menyelesaikan perhitungan numerik berbasis matriks. Software Matematika juga merupakan software yang paling efisien yang banyak diterapkan pada matematika komputasi, pengembangan dan algoritma, analisis data eksplorasi dan visualisasi[9].

Confusion matrix

Confusion matrix adalah salah satu metode evaluasi yang umum digunakan dalam Machine learning untuk menilai kinerja model klasifikasi. Confusion matrix memberikan pemahaman yang lebih mendalam tentang jenis kesalahan yang dibuat oleh model, membantu dalam menghitung metrik seperti akurasi, precision, recall, dan F1-score[10].

3. Metode Penelitian

Waktu dan Tempat Penelitian

Penelitian ini dimulai pada bulan November 2024 dimulai dengan penyusunan Proposal, tahapan perencanaan, pengumpulan data dan bahan. Penelitian ini Dilaksanakan di Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Manado. Data yang digunakan pada penelitian ini menggunakan data sekunder dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Manado.

Diagram Alur Penelitian



Gambar 2. Alur penelitian

Pada penelitian ini menggunakan beberapa langkah yang dilakukan untuk menyelesaikan masalah pada penelitian tersebut, untuk mencapai tujuan pada penelitian ini terdiri dari empat tahapan utama yaitu:

A. Pengumpulan data

Dataset yang digunakan pada penelitian ini dikumpulkan melalui metode sekunder, yang mana dataset cuaca dengan variable curah hujan, suhu, kelembaban, dan arah angin, data tersebut diperoleh langsung dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Manado.

B. Pemrosesan data

1. Praproses Data

Sebelum digunakan dalam pemodelan, data harus melalui serangkaian tahap Preprocessing untuk memastikan kualitas optimal. Tahapan ini meliputi:

- Pembersihan Data (*Data Cleaning*): Mengatasi data yang hilang atau tidak lengkap, serta menyelesaikan inkonsistensi.
- Transformasi Data: Melakukan normalisasi atau standarisasi nilai-nilai variabel untuk memastikan kesesuaian skala input pada model prediksi.
- Pemilihan Fitur (*Feature Selection*): Memilih fitur-fitur penting yang relevan terhadap prediksi kejadian banjir.

2. Pembagian Dataset

Dalam supervised learning, dataset dibagi menjadi dua bagian utama, yaitu data pelatihan (training) dan data pengujian (testing).

- Data Latih (*Training Set*) : Sebesar 70% dari keseluruhan data yang digunakan untuk melatih model.
- Data Uji (*Testing Set*) : Sebanyak 30% dari data yang digunakan untuk menguji kinerja model prediksi.

C. Menjalankan Random Forest Pada Dataset

1. Model Training (Pelatihan Model)

Metode ini peneliti menggunakan data latih (*Training data*), dimana data tersebut digunakan untuk melatih metode *Random Forest* tersebut.

2. Menjalankan metode Random Forest

Setelah data pada metode training telah terlatih, akan dilanjutkan dengan data pengujian (*Testing data*) yang akan digunakan untuk memprediksi banjir di kota Manado.

D. Evaluasi Model

Setelah proses menjalankan metode *Random Forest* selesai, dilanjutkan dengan melakukan analisis mendalam untuk memahami kekuatan dan kelemahan metode secara komprehensif. Tujuan dari analisis ini adalah untuk mengidentifikasi area di mana model bekerja dengan baik dan area di mana kinerjanya masih perlu ditingkatkan.

4. Hasil Penelitian

Hasil Pelatihan Model

A. Prediksi Data Uji

Hasil Load data					
Nama Tempat	Bulan	Tahun	Hasil Akurasi	Keterangan	Data
Bunaken	Maret	2014	92.66%	Banjir Sedang	Rata-Rata Curat
Bunaken	April	2014	92.22%	Banjir Sedang	Rata-Rata Curat
Bunaken	Agustus	2014	93.33%	Banjir Besar	Rata-Rata Curat
Bunaken	September	2014	92.30%	Banjir Besar	Rata-Rata Curat
Bunaken	Juni	2015	97.69%	Banjir Sedang	Rata-Rata Curat
Bunaken	September	2015	95.52%	Banjir Ringan	Rata-Rata Curat
Bunaken	Oktober	2015	93.17%	Banjir Ringan	Rata-Rata Curat

Gambar 3. Hasil data uji

Pada perhitungan Model prediksi banjir menggunakan data uji, hasil ditampilkan pada Gambar yang berisi hasil setelah data dilatih pada model yang menggunakan metode *Random Forest* dan menunjukkan performa yang sangat baik berdasarkan proses validasi menggunakan data uji. Seperti terlihat pada Tabel Hasil Load Data, model berhasil mencapai tingkat akurasi yang konsisten di atas 92% selama periode pengujian dari 2014 hingga 2024.

B. Prediksi data latih

Command Window	
Nama Tempat: Wonang, Bulan: April, Tahun: 2024	
- Jumlah data: 1	
- Akurasi: 92.84%	
- Status prediksi: Banjir Besar	
- Kondisi cuaca rata-rata:	
* Rata_Rata_Curah_Hujan: 607.86	
* Total_Curah_Hujan: 14540.44	
* Kelembaban_Rata_Rata: 83.28	
* Kelembaban_Minimal: 64.73	
* Angin_Dominan: 90.00	
* Frekuensi_Angin: 25.12	
* Suhu_Rata_Rata: 25.56	
* Suhu_Maksimal: 31.07	

Gambar 4. Hasil data latih

Hasil pada data latih ditampilkan pada bagian Command window dengan jumlah data yang di latih dari data set yaitu 10.164 data dengan banyak persenan yaitu 70% dari jumlah dataset menampilkan jumlah data dalam 1 baris serta akurasi perhitungan dalam satu bulan itu serta status prediksi yang dihasilkan dan keterangan data yang digunakan dijelaskan

Hasil Evaluasi Model

Metrik Evaluasi Model			
Kelas	Precision	Recall	F1-Score
Tidak Banjir	100.00%	96.55%	98.25%
Banjir Ringan	97.66%	97.66%	97.66%
Banjir Sedang	98.08%	99.35%	98.71%
Banjir Besar	100.00%	100.00%	100.00%
OOB Error:			
Waktu Training: 8.84 detik			
		Tampilkan Confusion Matrix	Tutup

Gambar 5. Evaluasi Hasil Model

Presisi (Precision): Rasio dokumen positif yang diprediksi dengan benar (TP) dengan total dokumen yang diprediksi sebagai positif (TP + FP).

$$Precision(Tidak Banjir) = \frac{96.55}{96.55} = 100.00\%$$

$$Precision(Banjir Ringan) = \frac{97.66}{101.7574} = 97.66\%$$

$$Precision(Banjir Sedang) = \frac{99.35}{101.6944} = 98.08\%$$

$$Precision(Banjir Besar) = \frac{100.00}{100.00} = 100.00\%$$

Recall (Sensitivity): Rasio dokumen positif yang diprediksi dengan benar (TP) dengan total dokumen positif sebenarnya (TP + FN).

$$Recall(Tidak Banjir) = \frac{96.55}{100.00} = 96.55\%$$

$$Recall(Banjir Ringan) = \frac{97.66}{100.00} = 97.66\%$$

$$Recall(Banjir Sedang) = \frac{99.35}{100.00} = 99.35\%$$

$$Recall(Banjir Besar) = \frac{100.00}{100.00} = 100.00\%$$

F1-Score : Ukuran gabungan dari presisi dan recall, menggunakan rumus:

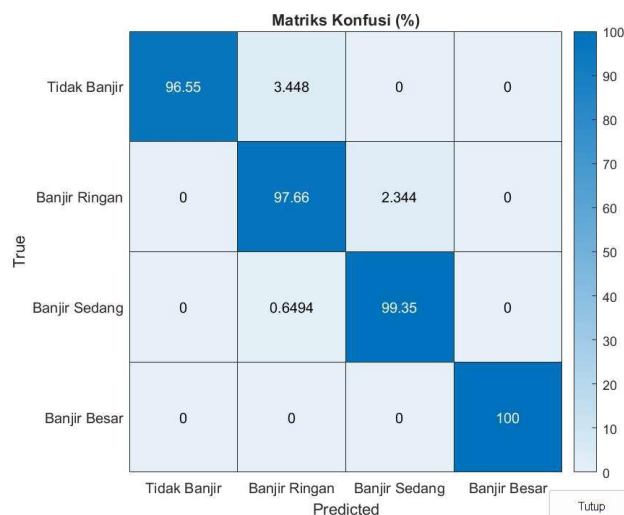
$$F1Score(Tidak Banjir) = 2 \times \frac{(1 \times 0.9655)}{(1 + 0.9655)} = 98.25\%$$

$$F1Score(Banjir Ringan) = 2 \times \frac{(0.9766 \times 0.9766)}{(0.9766 + 0.9766)} = 97.66\%$$

$$F1Score(Banjir Sedang) = 2 \times \frac{(0.9808 \times 0.9935)}{(0.9808 + 0.9935)} = 98.71\%$$

$$F1Score(Banjir Besar) = 2 \times \frac{(1 \times 1)}{(1 + 1)} = 100.00\%$$

Confusion Matrix



Gambar 6. Tampilan Confusion matrix

Matriks konfusi menunjukkan performa model klasifikasi dalam memprediksi empat kategori tingkat banjir, yaitu Tidak Banjir, Banjir Ringan, Banjir Sedang, dan Banjir Besar. Berdasarkan hasil evaluasi, model menunjukkan performa yang sangat baik dengan tingkat akurasi yang tinggi pada setiap kategori.

Hasil Prediksi

Hasil Data Predict					
Nama Tempat	Bulan	Tahun	Keterangan	Hasil Akurasi	Data
Bunaken	Januari	2025	Banjir Ringan	47.91%	Curah Hujan: 210
Bunaken	Februari	2025	Banjir Ringan	50.00%	Curah Hujan: 180
Bunaken	Maret	2025	Banjir Sedang	51.56%	Curah Hujan: 130
Bunaken	April	2025	Banjir Ringan	50.00%	Curah Hujan: 90
Bunaken	Mei	2025	Banjir Ringan	49.53%	Curah Hujan: 85
Bunaken	Juni	2025	Banjir Besar	49.16%	Curah Hujan: 75
Bunaken	Juli	2025	Banjir Sedang	50.00%	Curah Hujan: 80

Gambar 7. Tampilan Hasil Data Prediksi

5. Kesimpulan

Penelitian ini telah berhasil mengembangkan model prediksi banjir menggunakan metode *Random Forest* dengan akurasi konsisten di atas 90%, mencapai puncak 97,97%. Kategorisasi empat kelas banjir (Tidak Banjir, Banjir Ringan, Banjir Sedang, dan Banjir Besar) terbukti efektif dengan nilai *precision* dan *recall* yang sangat baik, dimana dua kategori bahkan mencapai sempurna 100%. Praproses data berhasil mengatasi permasalahan data kosong, sementara penggunaan delapan indikator cuaca memberikan basis komprehensif untuk model. Strategi pembagian dataset 70:30 menunjukkan efektivitas dalam memastikan kemampuan generalisasi model. Analisis matriks konfusi menunjukkan kesalahan klasifikasi minimal, hanya terjadi antara kelas berdekatan. Implementasi normalisasi Z-Score berhasil menyeimbangkan kontribusi seluruh fitur, dan fitur penyimpanan model meningkatkan efisiensi operasional keseluruhan sistem.

6. Daftar Pustaka

- [1] A. Pakkung, D. Hatidja, and J. Titaley, “d ’ CartesiaN Jurnal Matematika dan Aplikasi PREDIKSI CURAH HUJAN KOTA MANADO DENGAN MENGGUNAKAN METODE AUTOREGRESSIVE MOVING AVERAGE (ARMA),” *J. Mat. dan Apl.*, pp. 11–16, 2024.
- [2] M. Azizah, A. Subiyanto, S. Triutomo, and D. Wahyuni, “Pengaruh Perubahan Iklim Terhadap Bencana Hidrometeorologi di Kecamatan Cisarua - Kabupaten Bogor,” *PENDIPA J. Sci. Educ.*, vol. 6, no. 2, pp. 541–546, 2022, doi: 10.33369/pendipa.6.2.541-546.
- [3] R. Pangemanan, C. J. Supit, and J. D. Mamoto, “Prediksi Banjir Di Sungai Ranowangko Kota Tomohon,” *J. Sipil Statik*, vol. 8, no. 6, pp. 883–892, 2020.
- [4] M. Dhawangkharah and E. Riksakomara, “Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya),” *J. Tek. ITS*, vol. 6, no. 1, 2017, doi: 10.12962/j23373539.v6i1.21120.
- [5] M. D. Purnama and M. E. Mustafidah, “Relationship Between Temperature and Humidity on Rainfall,” *Eng. Math. Comput. Sci. J.*, vol. 6, no. 2, pp. 151–156, 2024, doi: 10.21512/emacsjournal.v6i2.11466.
- [6] Y. A. Saputra and D. Mahdiana, “Analisis Penerapan Machine Learning , Deep Learning , dan Data Mining dalam Prediksi Penjualan di Industri Otomotif Magister Ilmu Komputer , Fakultas Teknologi Informasi , Universitas Budi Luhur , Jakarta , Indonesia Sistem Informasi , Fakultas Teknologi Informasi , Universitas Budi Luhur , Jakarta , Indonesia Analysis of the Application of Machine Learning , Deep Learning , and Data Mining in Sales Prediction in the Automotive Industry,” vol. 5, no. 6, pp. 1743–1755, 2025.
- [7] P. Diana Nopianti, “Analisis Data Guna Meningkatkan Optimasi Perusahaan,” *J. Compr. Sci.*, vol. 1, no. 2, pp. 214–221, 2022, doi: 10.59188/jcs.v1i2.40.
- [8] G. Agung, D. Sri, A. Ningsih, and C. Pramatha, “Klasifikasi Kualitas Air Layak Minum menggunakan Algoritma Random Forest Classifier dan GridsearchCV,” vol. 12, no. 1, pp. 1197–1206, 2024.
- [9] I. Handani, “Pengaruh Media Pembelajaran Matematika meggunakan software matlab pokok bahasan matriks terhadap hasil belajar siswa tahun 2020/2021,” *Jimedu J. Ilm. Mhs. Kegur. dan ilmu Pendidik.*, vol. 2, no. 6, pp. 454–462, 2022.
- [10] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.