

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Bencana alam seperti banjir sering melanda Kabupaten Garut, Jawa Barat, yang rentan karena kondisi geografisnya berupa pegunungan dan aliran sungai yang kompleks. Tingginya curah hujan, lemahnya filtrasi tanah, serta penyumbatan sungai akibat sedimentasi dan sampah menjadi faktor utama penyebab banjir di wilayah ini. (Hasanah et al., 2021). Banjir di Garut menyebabkan kerugian material, merusak infrastruktur dan lahan pertanian, serta mengancam keselamatan jiwa warga setempat (Portal Jabar Prov Goid, 2022). Oleh karena itu, sangat penting untuk memiliki pendeteksi yang efektif karena dapat memberikan informasi yang berharga untuk mengurangi dampak terhadap korban bencana (Vega Purwayoga, 2023). Untuk mengantisipasi banjir atau kekeringan bergantung dengan deteksi banjir menggunakan data historis pada intensitas curah hujan di suatu daerah (Sandiwarno, 2024). Pendeteksi banjir menggunakan Machine Learning, khususnya dalam klasifikasi, dengan algoritma *K-Nearest Neighbor* (KNN) dan *Random Forest* yang efektif untuk dataset besar dan kompleks, seperti data cuaca dan curah hujan.

Pendekatan ini fokus pada identifikasi banjir secara multivariable, melampaui pendekatan penelitian sebelumnya pada penelitian (Depi Novita Sari, 2024) hanya menggunakan variabel curah hujan untuk analisis kejadian bencana banjir, yang kedepannya melibatkan variabel – variabel tambahan yang relevan, pada penelitian

(Bagaskara & Hartomo, 2024) hanya mempertimbangkan curah hujan dan diusulkan kembali untuk menambahkan variabel dengan cukup sehingga bisa menghasilkan yang relevan. Pada penelitian (Mohamed El-Sayed El-Mahdy F. I.-T., 2024). Penelitian dari (Cumel, 2022) Hasil dari Algoritma KNN merupakan Algoritma yang paling bagus dalam deteksi banjir, diperoleh hasil akhir yaitu Algoritma KNN (*K-Nearest Neighbor*) lebih akurat dibandingkan Algoritma Naive Bayes dalam melakukan deteksi banjir dengan nilai akurasi 88.94%, error 11.06%.. Pada Algoritma *Random Forest* dari penelitian (Arya Darmawan et al., 2023) menunjukkan akurasi dan performa yang paling baik dibandingkan dengan *Decision Tree* dan *Naïve Bayes*. *Random Forest* memiliki nilai akurasi sebesar 99,05%, precision sebesar 97,91%, recall sebesar 99,18%, dan f1-score sebesar 98%. Penelitian ini mengidentifikasi dua gap utama. Pertama, gap terkait analisis multivariabel, yang menggabungkan berbagai variabel iklim dan kejadian banjir secara bersamaan, untuk memahami hubungan antar faktor-faktor tersebut. Kedua, gap terkait kontribusi analisis performa model, yang mengevaluasi peran masing-masing variabel iklim dalam meningkatkan akurasi model prediksi banjir, guna memberikan rekomendasi yang lebih efektif untuk mitigasi risiko banjir. Namun, penelitian ini bertujuan untuk menjawab rumusan masalah: bagaimana membangun model deteksi banjir berbasis multivariabel dan algoritma mana yang lebih unggul dalam generalisasi data.

Penelitian ini dilakukan melalui pengumpulan data cuaca dari BMKG dan data banjir dari BPBD, diikuti integrasi, pembersihan, serta transformasi data. Data tersebut kemudian digunakan untuk melatih algoritma KNN dan RF, yang hasilnya

dievaluasi menggunakan akurasi, precision, recall, dan F1-score. Tujuan penelitian ini adalah untuk mengembangkan model prediksi banjir yang akurat, membandingkan performa algoritma, serta memberikan solusi mitigasi risiko banjir berbasis teknologi.

Diharapkan hasil penelitian ini dapat menjadi acuan bagi pemerintah daerah dan masyarakat dalam merencanakan mitigasi risiko banjir serta meningkatkan kesiapsiagaan menghadapi bencana di Kabupaten Garut.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah dijelaskan, terdapat rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana membangun pemodelan multivariabel untuk deteksi banjir dengan pendekatan *Machine learning* menggunakan algoritma *Algoritma K-Nearest Neighbor* dan *Random Forest*?
2. Bagaimana analisis performa dari algoritma *Algoritma K-Nearest Neighbor* dan *Random Forest* menggunakan multivariabel dalam mendeteksi banjir?

## **1.3 Tujuan Penelitian**

Berdasarkan latar belakang masalah yang telah dijelaskan, terdapat tujuan penelitian ini adalah sebagai berikut:

1. Membangun model multivariabel untuk deteksi banjir menggunakan pendekatan *Machine learning* dengan algoritma *K-Nearest Neighbor* (KNN) dan *Random Forest* (RF). Tujuan ini menekankan pengembangan model yang

memanfaatkan beberapa variabel (multivariabel) untuk memprediksi kejadian banjir.

2. Menganalisis 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.

#### **1.4 Manfaat Penelitian**

Terdapat manfaat penelitian yang dilakukan sebagai berikut:

1. Mengimplementasikan multivariabel pada pendekatan *machine learning* dengan menggunakan algoritma Algoritma K-Nearest Neighbor dan *Random Forest* untuk deteksi bencana banjir
2. Berpartisipasi dalam ruang lingkup penelitian tentang bencana dengan mengimplementasikan pendekatan dengan *Machine learning*.

#### **1.5 Batasan Masalah**

Terdapat batasan penelitian yang dilakukan sebagai berikut:

1. Data yang digunakan merupakan data hasil penelitian melalui Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), Badan Penanggulangan Bencana Daerah (BPBD), dan Data Batas Wilayah dari (Ina-Geospasial) Badan Informasi Geospasial Indonesia.
2. Metode yang digunakan dalam implementasi dengan *Machine learning* menggunakan *Algoritma K-Nearest Neighbor* dengan batasan nilai  $k = 1-10$  dan *Random Forest*  $n\_estimator = 10$  dan  $100$ .

3. Data peta koordinat yang digunakan data koordinat Stasiun Geofisika Bandung dan Stasiun Meteorologi Kertajati, serta koordinat wilayah Kabupaten Garut.
4. Periode data cuaca dari BMKG yang digunakan adalah dari 1 Januari 2022 sampai 7 Juli 2024 dan periode data administrasi daerah banjir yang digunakan adalah dari Januari 2022 sampai Juli 2024.