

BAB I

PENDAHULUAN

A. Latar Belakang

Indonesia merupakan negara kepulauan dengan iklim tropis yang sangat rentan terhadap bencana hidrometeorologi, salah satunya adalah banjir [1]. Menurut data BNPB, banjir menjadi salah satu bencana paling sering terjadi di berbagai wilayah Indonesia, terutama saat musim hujan. Kombinasi curah hujan tinggi, pola cuaca yang tidak menentu, serta kondisi geografis yang kompleks menyebabkan banyak wilayah, khususnya Daerah Istimewa Yogyakarta (DIY) dan Jawa Barat, mengalami banjir secara berkala. Fenomena perubahan iklim global turut memperparah situasi ini, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menyatakan bahwa dampak perubahan iklim sudah semakin nyata dirasakan di Indonesia, ditandai dengan meningkatnya frekuensi kejadian cuaca ekstrem [2] Dalam laporan resmi yang dikutip oleh CNN Indonesia, BMKG juga menekankan adanya korelasi antara peningkatan suhu global dengan tingginya intensitas hujan dan cuaca ekstrem lainnya [3] Di sisi lain, kajian oleh Akbar dan Sanjaya menegaskan pentingnya metode yang adaptif, seperti *machine learning*, untuk mengatasi tantangan prediksi pada data cuaca yang bersifat tidak seimbang dan sangat bervariasi [4]. Oleh karena itu, pengembangan sistem prediksi cuaca dan banjir menjadi kebutuhan penting dalam mendukung upaya mitigasi risiko bencana di Indonesia.

Selain bencana banjir, masyarakat Indonesia juga bergantung pada informasi cuaca harian, baik untuk transportasi, hingga kegiatan aktivitas sehari-hari [5]. Menurut penelitian oleh Syahreza dkk, informasi cuaca yang akurat, baik dalam bentuk prediksi *real-time* maupun prakiraan cuaca mingguan, sangat membantu masyarakat dalam merencanakan kegiatan secara lebih tepat dan efisien [5]. Namun demikian, data cuaca dan banjir di Indonesia masih tersebar di berbagai sumber, seperti BMKG, lembaga

pemerintah daerah, hingga layanan cuaca secara daring. Kondisi ini menimbulkan tantangan dalam pemanfaatan data secara menyeluruh, sehingga dibutuhkan proses integrasi dan analisis lanjutan agar data tersebut dapat dimanfaatkan secara optimal dalam sistem peringatan dini maupun pengambilan keputusan berbasis data.

Kemajuan teknologi kecerdasan buatan, khususnya dalam bidang *machine learning*, telah membuka peluang baru dalam pengembangan sistem prediksi cuaca dan banjir yang lebih adaptif dan akurat. Salah satu pendekatan yang terbukti efektif adalah penggunaan model *ensemble* seperti *Random Forest*, *XGBoost*, dan *LightGBM*. Menurut tinjauan literatur oleh Mosavi dkk, model-model ini menunjukkan performa yang unggul dalam menangani data lingkungan yang kompleks dan bersifat non-linier, terutama dalam konteks prediksi banjir dan cuaca ekstrem[6]. Dengan menggabungkan data historis dari BMKG [7] serta data *real-time* dari *API OpenWeatherMap*, kini dimungkinkan untuk memproses dan menampilkan informasi prediksi cuaca (suhu, curah hujan, kelembapan udara, kecepatan angin) serta risiko banjir secara otomatis. Pengembangan aplikasi *web* yang terintegrasi dengan *backend Flask* juga memastikan hasil prediksi bisa diakses oleh pengguna dan mudah dipahami.

Melalui penelitian ini, dikembangkan aplikasi prediksi cuaca dan banjir berbasis *web* yang tidak hanya memprediksi risiko banjir menggunakan *ensemble voting machine learning*, tetapi juga memberikan informasi prediksi cuaca *real-time*, dan mingguan untuk berbagai provinsi di Indonesia dan juga memberikan informasi prediksi banjir untuk wilayah DIY dan Jawa Barat. Sistem ini diharapkan dapat menjadi kontribusi nyata untuk membantu masyarakat dan pihak terkait dalam mengambil langkah antisipasi terhadap risiko bencana, sekaligus memenuhi kebutuhan informasi cuaca yang relevan.

B. Rumusan Masalah

1. Bagaimana membangun sistem prediksi cuaca dan banjir *real-time* menggunakan *machine learning* pada wilayah Daerah Istimewa Yogyakarta (DIY) dan Jawa Barat?
2. Bagaimana perbandingan performa algoritma *ensemble Random Forest*, *XGBoost*, dan *LightGBM* dalam memprediksi risiko banjir pada data tidak seimbang?
3. Algoritma *machine learning* manakah yang memberikan hasil prediksi terbaik dalam prediksi cuaca dan prediksi banjir?

C. Batasan Masalah

1. Penelitian prediksi banjir hanya difokuskan pada wilayah Daerah Istimewa Yogyakarta (DIY) dan Jawa Barat, karena wilayah provinsi Jawa Barat berdasarkan statistik BNPB tahun 2018-2023 merupakan wilayah dengan jumlah kejadian banjir paling tinggi di Indonesia, sehingga menjadi prioritas penelitian ini. Sedangkan provinsi DIY dipilih sebagai wilayah pembanding karena merupakan memiliki jumlah kejadian banjir yang jauh lebih sedikit. Pemilihan ini bertujuan untuk mengevaluasi kinerja model prediksi pada dua karakter wilayah yang berbeda, satu dengan frekuensi banjir tinggi (Jawa Barat) dan satu dengan frekuensi rendah (DIY). Data cuaca yang digunakan dalam rentang waktu tahun 2018–2023.
2. Penelitian prediksi cuaca hanya terbatas sampai pada tingkat provinsi.
3. Prediksi banjir dibatasi pada klasifikasi *biner* (“AMAN” atau “WASPADA BANJIR”) tanpa memperkirakan tingkat keparahan, durasi, penyebab atau kedalaman banjir.
4. Fitur untuk prediksi cuaca meliputi suhu, curah hujan, kelembapan udara, dan kecepatan angin, dengan cakupan prediksi mencakup kondisi *real-time* dan prakiraan cuaca mingguan.

5. Data diambil dari sumber resmi: BMKG, BNPB, *OpenWeatherMap*, dan *OpenStreetMap*.

D. Tujuan Penelitian

1. Membangun sistem prediksi cuaca secara *real-time*, dan per minggu untuk wilayah Indonesia serta prediksi banjir secara *real-time* berbasis *machine learning* yang mengintegrasikan data historis dari BMKG dan data *real-time* dari API *OpenWeatherMap* untuk wilayah Daerah Istimewa Yogyakarta (DIY) dan Jawa Barat.
2. Menganalisis dan membandingkan performa algoritma *ensemble* (*Random Forest*, *XGBoost*, dan *LightGBM*) dalam melakukan prediksi risiko banjir pada data yang tidak seimbang.
3. Mengidentifikasi algoritma *machine learning* yang paling akurat untuk prediksi cuaca dan prediksi banjir pada wilayah studi, berdasarkan hasil evaluasi metrik performa model.

E. Metode Penelitian

Penelitian ini menggunakan pendekatan yang meliputi pengumpulan data, *preprocessing* data, perancangan dan pelatihan model *machine learning*, evaluasi performa model, serta implementasi ke dalam aplikasi *web*.

1. Pengumpulan Data

Data yang digunakan terdiri dari:

- Data historis cuaca harian (curah hujan, kelembapan, kecepatan angin, suhu, elevasi, *latitude*, *longitude* dan atribut lingkungan lain) periode 2018–2023, diperoleh dari BMKG.
- Data kejadian banjir dari BNPB untuk wilayah DIY dan Jawa Barat.
- Data *real-time* cuaca yang diambil melalui API *OpenWeatherMap*.
- Data peta dari *OpenStreetMap* untuk kebutuhan visualisasi lokasi.

2. *Preprocessing Data*

- Tahapan praproses yang dilakukan meliputi:
 - Pembersihan data: menghapus duplikasi, menangani *missing values* (0, *NaN*, 8888, 9999, '-'), dan menyamakan format tanggal serta lokasi.
 - Normalisasi fitur: menggunakan metode *StandardScaler* atau *Min-Max Scaling* untuk fitur kontinu.
 - Penggabungan data: menggabungkan data historis dari BMKG, data kejadian banjir dari BNPB.
 - *Labeling*: memberikan label target *biner* (1 = banjir, 0 = tidak banjir) berdasarkan catatan kejadian.
 - Penanganan data tidak seimbang: model dilatih tanpa teknik *oversampling* agar data tetap organik.

3. Evaluasi Model

- Model dievaluasi menggunakan metrik-metrik berikut:
 - Akurasi, *Precision*, *Recall*, dan *F1-Score* untuk prediksi banjir.
 - R^2 (*R-squared*), *MAE*, dan *RMSE* untuk prediksi cuaca (regresi).
 - *Confusion Matrix* dan *ROC-AUC score* untuk analisis lebih mendalam pada klasifikasi banjir.
 - Evaluasi dilakukan secara *cross-validation* dan validasi temporal untuk memastikan performa stabil pada data waktu berbeda.

4. Implementasi Sistem

- Hasil pelatihan model disimpan dalam bentuk *file .pkl* dan diintegrasikan ke dalam aplikasi *web* berbasis *Flask*.
- Aplikasi dapat menerima input lokasi, mengambil data cuaca *real-time* dari API, melakukan prediksi banjir maupun cuaca, dan menampilkan hasil dalam bentuk kartu informasi dan peta interaktif (*Leaflet.js*).

- Pengujian sistem dilakukan untuk memastikan seluruh fitur aplikasi (input, prediksi, visualisasi peta) berjalan dengan baik dan akurat secara *real-time*.

F. Sistematika Penulisan

BAB I : Pendahuluan

Pada BAB I berisi adapun penjelasan masalah yang terjadi sebagai latar belakang penelitian, dari latar belakang ini kemudian dapat ditemukan rumusan permasalahan, batas permasalahan, tujuan penelitian, metode penelitian, dan sistematika penelitian.

BAB II : Tinjauan Pustaka

Pada BAB II ini berisikan penelitian-penelitian terdahulu yang menyangkut dengan penelitian yang dilakukan, yang merupakan perbandingan antara penilitan yang dilakukan dengan penelitian-penelitian terdahulu.

BAB III : Landasan Teori

Pada BAB III ini terdiri dari penjelasan mengenai teori-teori yang berkaitan sebagai landasan penelitian yang dilakukan. Pada bab ini juga dibantu dengan sumber penelitian terdahulu yang mendukung penelitian.

BAB IV : Dataset Dan Pengembangan Model

Pada BAB IV ini akan terdiri dari deskripsi masalah yang akan diselesaikan dengan metode yang dipilih, kemudian terdapat dataset yang mendukung adanya penelitian seperti dengan menganalisis data dan *preprocessing* data, lalu

kemudian model akan dikembangkan menggunakan teknik yang dipilih dengan pelatihan model, evaluasi model, dan pengujian model.

BAB V : Implementasi Model Dan Pengujian Sistem

Pada BAB V ini akan terdiri dari implementasi model yang membahas proses implementasi model ke dalam web, implementasi sistem yang akan membahas bagaimana sistem dapat menggunakan model, dan pengujian sistem yang dilakukan untuk menguji apakah sistem yang dibuat semua fungsinya dapat berjalan dengan baik atau belum.

BAB VI : Kesimpulan dan Saran

Pada BAB VI ini akan menjelaskan kesimpulan dari hasil penelitian yang dilakukan, yang kemudian akan dilanjutkan dengan saran yang terkait dengan bagaimana penelitian lanjutan dapat dilakukan.