

BAB I

PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan salah satu unsur iklim yang mempengaruhi kehidupan dan aktivitas manusia di berbagai sektor, antara lain pertanian, transportasi, dan pengelolaan sistem drainase di suatu wilayah (Anwar *et al.*, 2021). Kota Surabaya, sebagai salah satu kota metropolitan di Indonesia, memiliki pola hujan yang fluktuatif, yang menunjukkan adanya variasi curah hujan dari waktu ke waktu.

Data Stasiun Meteorologi Perak I mencatat bahwa curah hujan tertinggi pada periode Januari 2018 – September 2025 terjadi pada Februari 2024 sebesar 166 mm (Badan Meteorologi, Klimatologi, dan Geofisika, 2024). Selain nilai maksimum tersebut, data curah hujan juga menunjukkan perubahan atau fluktuasi dari waktu ke waktu. Variabilitas curah hujan tersebut menunjukkan dinamika kondisi cuaca yang dapat mempengaruhi aktivitas sehari-hari. Oleh karena itu, informasi mengenai curah hujan menjadi penting sebagai dasar dalam mendukung upaya mitigasi banjir, perencanaan tata ruang, serta peningkatan ketahanan iklim perkotaan (Putra *et al.*, 2024). Dengan demikian, diperlukan sistem klasifikasi curah hujan yang akurat untuk mengidentifikasi curah hujan secara tepat.

Penerapan *machine learning* dalam bidang sains dan rekayasa lingkungan semakin meluas seiring berkembangnya teknologi, termasuk untuk prediksi fenomena meteorologis. Salah satu algoritma yang banyak digunakan adalah *Extreme Gradient Boosting (XGBoost)*, yaitu sistem *boosting* berbasis pohon yang efisien untuk dataset besar dengan keunggulan pada kecepatan pelatihan, akurasi

tinggi, kemampuan menangani *missing value*, serta regularisasi untuk mencegah *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan dengan data pelatihan sehingga kurang mampu melakukan generalisasi terhadap data uji (Chen & Guestrin, 2016). Dalam konteks meteorologi, algoritma ini mampu menangkap hubungan nonlinier antarvariabel cuaca lebih baik dibandingkan model statistik konvensional seperti SARIMA (Putra *et al.*, 2024). Namun, sifatnya yang “*black box*” menyulitkan interpretasi hasil, sehingga diperlukan metode *SHapley Additive exPlanations* (SHAP) untuk menjelaskan kontribusi setiap fitur terhadap keputusan model.

Berbagai penelitian menunjukkan bahwa variabel meteorologis memiliki pengaruh besar terhadap pembentukan dan prediksi curah hujan. Secara umum, fitur seperti suhu udara serta kecepatan dan arah angin saling berinteraksi dalam menentukan intensitas curah hujan (Pathan *et al.*, 2021). Kombinasi variabel-variabel tersebut dapat dimodelkan secara efektif menggunakan algoritma XGBoost, sementara *SHapley Additive exPlanations* (SHAP) berperan dalam menjelaskan kontribusi tiap fitur terhadap hasil klasifikasi (Shao *et al.*, 2025). Pendekatan ini relevan untuk menghasilkan model klasifikasi curah hujan yang akurat dalam menentukan faktor-faktor meteorologis yang berperan di wilayah perkotaan seperti Surabaya.

Penggunaan algoritma XGBoost dalam pemodelan dan klasifikasi curah hujan telah banyak diteliti pada berbagai wilayah. XGBoost terbukti efektif untuk klasifikasi curah hujan di Kota Bandung menggunakan data observasi Stasiun Geofisika Bandung dengan akurasi 81% (Hapsari & Wisesa, 2025), tetapi belum mengintegrasikan SHAP. Pendekatan yang mengombinasikan XGBoost dan SHAP

telah digunakan untuk memprediksi curah longsor yang dipicu hujan dengan memanfaatkan data hujan harian, kelembapan tanah, serta faktor lingkungan (Shao *et al.*, 2025), tetapi fokusnya bukan pada klasifikasi curah hujan harian berdasarkan variabel meteorologi BMKG.

Penelitian sebelumnya memperlihatkan adanya kesenjangan penelitian dalam penerapan gabungan XGBoost dan SHAP untuk klasifikasi curah hujan dengan variabel meteorologis yang lebih komprehensif seperti arah angin maksimum dan lama penyinaran matahari. Oleh karena itu, penelitian ini berfokus pada pengembangan model klasifikasi curah hujan berbasis XGBoost dengan interpretasi SHAP menggunakan data meteorologis harian Kota Surabaya. Pendekatan ini diharapkan mampu memberikan prediksi yang akurat sekaligus penjelasan mengenai faktor-faktor yang memengaruhi curah hujan, guna mendukung mitigasi banjir, perencanaan tata ruang adaptif, dan peningkatan ketahanan iklim di Kota Surabaya.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian ini adalah:

1. Bagaimana cara menerapkan metode klasifikasi curah hujan menggunakan algoritma XGBoost berdasarkan variabel-variabel meteorologis yang memengaruhi curah hujan di Kota Surabaya?
2. Bagaimana kinerja model XGBoost dalam mengklasifikasikan curah hujan berdasarkan hasil evaluasi performa model?
3. Bagaimana menginterpretasikan kontribusi setiap variabel meteorologis terhadap hasil klasifikasi menggunakan metode *SHapley Additive exPlanations* (SHAP)?

1.3 Batasan Masalah

Permasalahan dalam penelitian ini dibatasi dengan:

1. Penelitian ini menggunakan data harian dari *website* BMKG Stasiun Meteorologi Perak I Surabaya, periode Januari 2018 hingga September 2025. Variabel yang digunakan meliputi suhu minimum, suhu maksimum, suhu rata-rata, kelembapan rata-rata, lama penyinaran matahari, arah angin saat kecepatan maksimum, arah angin terbanyak, kecepatan angin maksimum dan kecepatan angin rata-rata.
2. Penelitian ini hanya menggunakan algoritma XGBoost tanpa membandingkannya dengan metode lain seperti *Random Forest*, *Decision Tree*, atau *Support Vector Machine* (SVM).
3. Proses *hyperparameter tuning* dilakukan menggunakan metode *grid search* dengan pengujian terhadap beberapa parameter utama, yaitu *n_estimators*, *max_depth*, *learning_rate*, *subsample*, *colsample_bytree*, dan *min_child_weight*. Ruang pencarian parameter dibatasi agar sesuai dengan kapasitas komputasi dan efisiensi proses pelatihan model.
4. Evaluasi kinerja model difokuskan pada metrik akurasi, presisi, *recall*, *F1-score*, serta ditampilkan dalam bentuk *confusion matrix* untuk menilai kemampuan model dalam mengklasifikasikan curah hujan.
5. Interpretasi hasil klasifikasi hanya dilakukan menggunakan metode *SHapley Additive exPlanations* (SHAP) untuk menjelaskan kontribusi setiap variabel terhadap hasil prediksi. Metode interpretasi lain tidak digunakan dalam penelitian ini.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Menerapkan model klasifikasi curah hujan menggunakan algoritma XGBoost guna menghasilkan model prediksi curah hujan yang akurat.
2. Mengevaluasi performa model XGBoost dalam mengklasifikasikan curah hujan menggunakan metrik evaluasi yang telah ditentukan, yaitu metrik evaluasi dan *confusion matrix* untuk menilai efektivitas model dalam melakukan klasifikasi.
3. Menganalisis kontribusi setiap variabel independen terhadap hasil prediksi model XGBoost SHAP, dengan tujuan untuk mengidentifikasi variabel yang paling berpengaruh terhadap hasil prediksi model.