

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cuaca merupakan faktor lingkungan fundamental yang memiliki pengaruh signifikan terhadap berbagai dimensi kehidupan manusia, khususnya dalam sektor pertanian yang merupakan pilar strategis ekonomi Indonesia (Busthanul et al., 2023; Sholihah et al., 2024; Rusmayadi et al., 2023). Sebagai negara dengan populasi keempat terbesar di dunia, Indonesia memiliki tanggung jawab krusial dalam menjaga ketahanan pangan global, namun peran penting ini dihadapkan pada tantangan kompleks yang dipicu oleh dinamika iklim yang terus berubah, berkurangnya lahan pertanian yang produktif, dan ketidakefisienan yang melekat pada praktik pertanian tradisional (Rusmayadi et al., 2023). Perubahan cuaca yang tidak menentu dan ekstrem di Indonesia, termasuk peristiwa El Niño yang menyebabkan penundaan curah hujan dan penurunan penanaman padi di wilayah-wilayah utama penghasil beras, mengancam stabilitas produksi pertanian dan sistem penghidupan petani, terutama petani skala kecil yang memiliki keterbatasan sumber daya (Busthanul et al., 2023 ;Naylor et al., 2007). Oleh karena itu, kebutuhan akan metode prediksi cuaca yang lebih efektif dan responsif menjadi semakin mendesak untuk mendukung pengambilan keputusan pertanian yang lebih presisi dan adaptif terhadap variabilitas iklim yang meningkat.

Secara tradisional, peramalan cuaca dilaksanakan melalui penerapan model numerik yang didasarkan pada persamaan matematis kompleks yang menggambarkan proses-proses fisika atmosfer. Namun, pendekatan konvensional ini menghadapi keterbatasan substansial dalam hal efisiensi pemrosesan komputasi, sensitivitas terhadap ketidakpastian parameter atmosfer, dan akurasi prediksi yang seringkali tidak memuaskan, terutama dalam mengantisipasi peristiwa cuaca ekstrem yang semakin sering terjadi (Meenal et al., 2021). Di Indonesia, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) sebagai lembaga pemerintah yang bertanggung jawab menyediakan informasi perubahan cuaca, seringkali menghadapi tantangan dalam memberikan prediksi yang spesifik dan terdiferensiasi untuk wilayah-wilayah tertentu, sehingga prediksi yang dihasilkan cenderung bersifat umum dan kurang responsif terhadap kebutuhan lokal (Hamami & Dahlan, 2022). Keterbatasan ini menciptakan kesenjangan

informasi yang signifikan antara ketersediaan data cuaca nasional dan kebutuhan spesifik petani lokal yang memerlukan informasi cuaca yang lebih detail dan akurat untuk wilayah mereka masing-masing.

Seiring dengan kemajuan pesat dalam teknologi pengolahan data dan komputasi, pendekatan berbasis machine learning (ML) dan deep learning (DL) mulai diadopsi sebagai alternatif yang lebih responsif, berbasis data, dan mampu menangani kompleksitas non-linear dalam sistem cuaca. Teknik-teknik machine learning, khususnya deep learning, telah menunjukkan kemampuan yang luar biasa dalam mengkuantifikasi data cuaca multidimensi, mengelola hubungan non-linear yang sangat kompleks, dan memanfaatkan kumpulan data besar yang tersedia dari berbagai sumber meteorologi (Meenal et al., 2021). Pendekatan berbasis data ini memiliki keunggulan fundamental dalam menghasilkan prediksi cuaca yang masuk akal dan akurat tanpa memerlukan pengetahuan presisi tentang fisika atmosfer yang rumit, sehingga membuka peluang baru untuk pengembangan sistem prediksi yang lebih efisien dan terjangkau (Meenal et al., 2021). Dengan kemampuan untuk menangani karakteristik cuaca yang multidimensi dan dataset energi terbarukan yang sangat besar, machine learning menawarkan solusi yang lebih fleksibel dan adaptif dibandingkan dengan model numerik konvensional yang memerlukan sumber daya komputasi yang sangat besar.

Random Forest sebagai algoritma ensemble learning telah terbukti secara empiris unggul dalam menangani kompleksitas iklim tropis Indonesia dan menunjukkan performa yang konsisten dalam berbagai aplikasi prediksi cuaca (Hamami & Dahlan, 2022); (Mao & Sorteberg, 2020); (Putra et al., 2024). Keunggulan algoritma ini berasal dari beberapa karakteristik fundamental yang sangat relevan dengan tantangan prediksi cuaca: pertama, Random Forest mampu menangkap pola-pola non-linear yang kompleks tanpa memerlukan asumsi distribusi statistik yang ketat, sehingga sangat sesuai untuk menangani variabilitas cuaca yang tinggi di daerah tropis kedua, algoritma ini menunjukkan robustness yang tinggi terhadap outlier dan data ekstrem yang umum ditemukan dalam dataset cuaca ketiga, Random Forest menyediakan mekanisme interpretasi yang transparan melalui analisis feature importance, memungkinkan identifikasi variabel meteorologi yang paling berpengaruh terhadap prediksi. Penelitian empiris menunjukkan bahwa Random Forest mampu mencapai akurasi tinggi dalam

prediksi curah hujan dengan nilai ROC-AUC mendekati sempurna (1,00), mengindikasikan kemampuan diskriminasi yang sangat baik, dan bahkan dapat mencapai akurasi 71% dalam klasifikasi cuaca harian di Indonesia hanya dengan menggunakan enam variabel meteorologis dasar, mengungguli efisiensi komputasi model numerik konvensional.

Dengan mempertimbangkan karakteristik-karakteristik tersebut, Random Forest dipandang sebagai pilihan metodologi yang paling tepat dan optimal untuk membangun model klasifikasi cuaca harian yang andal dan responsif di wilayah-wilayah spesifik seperti Daerah Istimewa Yogyakarta. Algoritma ini menggabungkan secara sinergis kemampuan menangani data non-linear yang kompleks, robustness terhadap outlier dan data ekstrem, efisiensi komputasi yang superior, dan interpretabilitas yang memungkinkan pemahaman mendalam tentang faktor-faktor yang mempengaruhi cuaca lokal (Hamami & Dahlan, 2022). Pengembangan sistem prediksi cuaca berbasis Random Forest untuk wilayah lokal diharapkan dapat mengatasi keterbatasan prediksi umum yang disediakan oleh BMKG, memberikan informasi cuaca yang lebih spesifik dan akurat untuk mendukung pengambilan keputusan pertanian yang lebih baik, meningkatkan ketahanan petani terhadap variabilitas iklim, dan pada akhirnya berkontribusi pada peningkatan ketahanan pangan dan kesejahteraan komunitas pertanian lokal (Hamami & Dahlan, 2022); Sholihah et al., 2024). Dengan demikian, implementasi Random Forest untuk prediksi cuaca lokal merepresentasikan langkah strategis menuju sistem peringatan dini cuaca yang lebih responsif, akurat, dan bermanfaat bagi pengambil keputusan di tingkat lokal.

Berdasarkan pertimbangan tersebut, penelitian ini difokuskan pada wilayah Daerah Istimewa Yogyakarta (DIY) yang dipilih karena memiliki karakteristik cuaca yang dinamis dan kompleks akibat pengaruh iklim monsun tropis, variasi topografi yang signifikan, serta kedekatannya dengan wilayah pesisir. Selain itu, Yogyakarta merupakan wilayah strategis dengan ketergantungan tinggi terhadap kondisi cuaca pada sektor pertanian, pariwisata, dan mitigasi bencana. Prediksi cuaca harian dalam konteks ini tidak hanya berfungsi sebagai antisipasi sesaat, tetapi menjadi dasar penting dalam perencanaan pertanian berkelanjutan, pengelolaan operasional sektor pariwisata, serta sebagai komponen awal sistem peringatan dini bencana hidrometeorologi. Oleh karena itu, pengembangan model klasifikasi cuaca harian berbasis Random Forest di wilayah

Yogyakarta diharapkan mampu menghasilkan prediksi cuaca yang lebih spesifik, akurat, dan aplikatif untuk mendukung pengambilan keputusan di tingkat lokal, meningkatkan ketahanan terhadap variabilitas iklim, serta berkontribusi pada ketahanan pangan dan kesejahteraan masyarakat.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang diatas, maka penulis dapat merumuskan beberapa permasalahan sebagai berikut:

1. Bagaimana model klasifikasi untuk prediksi cuaca harian (cerah, mendung, hujan) dapat dibangun menggunakan algoritma Random Forest berdasarkan data BMKG DIY 2022–2025?
2. Seberapa tinggi tingkat akurasi dan performa yang dicapai oleh model Random Forest dalam mengklasifikasikan kondisi cuaca tersebut?
3. Faktor-faktor cuaca apa saja yang paling berpengaruh dalam prediksi menurut model Random Forest?

1.3 Tujuan

Penelitian ini bertujuan untuk mencapai beberapa sasaran utama sebagai berikut:

1. Membangun dan mengimplementasikan sebuah model klasifikasi cuaca harian menggunakan algoritma Random Forest.
2. Mengukur dan menganalisis tingkat akurasi serta performa model Random Forest yang telah dibangun.
3. Mengidentifikasi faktor meteorologis yang paling dominan dalam memengaruhi hasil prediksi melalui analisis feature importance.

1.4 Manfaat

Adapun manfaat dari penelitian ini antara lain sebagai berikut:

1. Memberikan kontribusi pada pengembangan model prediksi cuaca yang lebih akurat melalui penerapan dan evaluasi algoritma Random Forest pada data iklim tropis di Indonesia.

2. Meningkatkan pemahaman tentang algoritma random forest, sehingga dapat diketahui kekuatan dan kelemahan dalam klasifikasi kondisi cuaca
3. Menghasilkan model yang dapat menjadi alat bantu dalam pengambilan keputusan dari berbagai sektor untuk DIY ,seperti perencanaan masa tanam dan panen untuk pertanian,dukungan peringatan dini terhadap cuaca ekstrem pada mitigasi bencana dan pemberian informasi terhadap prakiraan yang lebih akurat dalam pariwisata.

1.5 Ruang Lingkup Masalah

Adapun ruang lingkup yang dibahas dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan pendekatan machine learning dengan algoritma Random Forest sebagai metode utama dalam membangun model klasifikasi cuaca harian.
2. Data yang digunakan merupakan data cuaca harian yang bersumber dari BMKG untuk wilayah DIY.
3. Periode data yang dianalisis mencakup rentang waktu tahun 2022 hingga 2025.
4. Variabel meteorologis yang digunakan terbatas pada suhu minimum (TN), suhu maksimum (TX), kelembapan rata-rata (RH_AVG), curah hujan (RR), lama penyinaran matahari (SS), dan kecepatan angin maksimum (FF_X).
5. Output penelitian dibatasi pada klasifikasi cuaca harian ke dalam tiga kelas, yaitu Cerah, Mendung, dan Hujan, sesuai dengan kriteria operasional BMKG.
6. Penelitian ini tidak membahas prediksi numerik cuaca, prakiraan jangka menengah/panjang, maupun penerapan model deep learning atau numerical weather prediction (NWP).

1.6 Kabaharuan

Penelitian ini memiliki kontribusi orisinal dalam bentuk dua aspek utama yaitu:

1. Penerapan validasi berbasis time-series split yang ketat dikombinasikan dengan penanganan ketidakseimbangan kelas melalui `class_weight='balanced'`, menjawab gap metodologis dalam studi terdahulu (Zian Asti Dwiyantri, 2023) yang masih menggunakan random split berisiko data leakage. Kombinasi ini memastikan evaluasi model mencerminkan skenario operasional nyata: memprediksi cuaca masa depan hanya dengan data historis, sekaligus menjaga sensitivitas terhadap kelas minoritas (Cerah) yang kritis untuk peringatan dini kondisi ekstrem.

2. Pengembangan prototipe sistem prediksi berbasis Streamlit yang tidak hanya menghasilkan label kategori, tetapi juga menyediakan interpretasi visual melalui feature importance dan probabilitas kelas, sehingga menjembatani kesenjangan antara model machine learning dan kebutuhan pengambil keputusan di sektor pertanian, pariwisata, dan mitigasi bencana di DIY. Antarmuka interaktif ini memungkinkan pengguna non-teknis memahami faktor meteorologis dominan yang mendasari setiap prediksi suatu fitur yang belum tersedia dalam sistem prediksi cuaca berbasis ML di wilayah tropis Indonesia.