

Prediksi Cuaca Harian Berdasarkan Data BMKG Daerah Istimewa Yogyakarta dengan Menggunakan Algoritma Random Forest

Rafi Dio Adibta, 202231031

Dibawah Bimbingan Bapak Rakhmadi Irfansyah Putra, S.Kom, MMSI

ABSTRAK

Cuaca merupakan variabel lingkungan krusial yang berpengaruh signifikan terhadap sektor-sektor strategis, termasuk pertanian, transportasi, pariwisata, serta mitigasi bencana. Di Daerah Istimewa Yogyakarta (DIY), wilayah yang memiliki intensitas aktivitas agrikultur dan pariwisata yang tinggi, ketepatan informasi cuaca harian menjadi faktor penentu dalam perencanaan operasional dan pengambilan keputusan. Mengingat dinamika iklim tropis Indonesia yang ditandai oleh ketidakstasioneran pola dan fluktuasi ekstrem, pendekatan prediksi berbasis model numerik konvensional sering kali kurang adaptif. Penelitian ini bertujuan mengembangkan model klasifikasi cuaca harian menggunakan algoritma Random Forest berdasarkan data historis dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Klimatologi Mlati, Sleman, periode 2022–2025. Dataset terdiri atas 1.230 observasi harian dengan enam variabel meteorologis: suhu minimum (TN), suhu maksimum (TX), kelembapan rata-rata (RH_AVG), curah hujan (RR), lama penyinaran matahari (SS), dan kecepatan angin maksimum (FF_X).

Proses pra-pemrosesan mencakup penanganan missing value, pembentukan lag features ($t-1$) untuk menangkap ketergantungan temporal, pelabelan kelas berdasarkan kriteria operasional BMKG (Cerah: $RR = 0 \text{ mm}$ & $RH_AVG < 80\%$; Mendung: $RR = 0 \text{ mm}$ & $RH_AVG \geq 80\%$; Hujan: $RR > 0 \text{ mm}$), serta pembagian data secara kronologis (time-series split) dengan rasio 80% untuk pelatihan (Januari 2022–Agustus 2024) dan 20% untuk pengujian (September 2024–Desember 2025). Evaluasi model dilakukan menggunakan metrik yang robust terhadap ketidakseimbangan kelas, yaitu Macro-average F1-score, Accuracy, Precision, Recall, dan ROC-AUC (One-vs-Rest). Hasil menunjukkan bahwa model Random Forest mencapai F1-score (Macro) sebesar 0,4486, ROC-AUC 0,7248, dan akurasi 50,00% pada data uji.

Analisis feature importance mengidentifikasi kelembapan rata-rata hari sebelumnya (RH_AVG_lag1) dan curah hujan hari sebelumnya (RR_lag1) sebagai prediktor paling dominan, yang konsisten dengan prinsip fisika atmosfer tropis. Temuan ini mengonfirmasi bahwa Random Forest mampu menangkap pola temporal cuaca harian di wilayah tropis, meskipun performanya masih dibatasi oleh kesederhanaan fitur input dan kompleksitas dinamika iklim lokal. Penelitian ini memberikan fondasi metodologis bagi pengembangan sistem prediksi cuaca berbasis machine learning yang lebih adaptif, interpretable, dan spesifik wilayah untuk mendukung sektor strategis di Indonesia.

Keywords: *Random Forest, Prediksi Cuaca Harian, Machine Learning, Data BMKG, Iklim Tropis, Feature Importance, Prediksi Berbasis Lag, Time-Series Split*

Prediksi Cuaca Harian Berdasarkan Data BMKG Daerah Istimewa Yogyakarta dengan Menggunakan Algoritma Random Forest

Rafi Dio Adibta, 202231031

Dibawah Bimbingan Bapak Rakhmadi Irfansyah Putra, S.Kom, MMSI

ABSTRACT

Weather is a critical environmental variable that significantly influences strategic sectors, including agriculture, transportation, tourism, and disaster mitigation. In the Special Region of Yogyakarta (DIY) a region characterized by high agricultural and tourism activity accurate daily weather information is a decisive factor in operational planning and decision-making. Given the dynamic nature of Indonesia's tropical climate, marked by non-stationary patterns and extreme fluctuations, conventional numerical weather prediction models are often insufficiently adaptive. This study aims to develop a daily weather classification model using the Random Forest algorithm based on historical data from the Meteorology, Climatology, and Geophysics Agency (BMKG), Mlati Climatological Station, Sleman, covering the period 2022–2025. The dataset comprises 1,230 daily observations of six meteorological variables: minimum temperature (TN), maximum temperature (TX), average relative humidity (RH_AVG), rainfall (RR), sunshine duration (SS), and maximum wind speed (FF_X).

The preprocessing pipeline includes handling missing values, engineering lag features ($t-1$) to capture temporal dependencies, class labeling according to BMKG operational criteria (Clear: $RR = 0 \text{ mm}$ & $RH_AVG < 80\%$; Cloudy: $RR = 0 \text{ mm}$ & $RH_AVG \geq 80\%$; Rainy: $RR > 0 \text{ mm}$), and chronological data splitting (time-series split) with an 80% training set (January 2022–August 2024) and a 20% test set (September 2024–December 2025). Model evaluation employs metrics robust to class imbalance: Macro-average F1-score, Accuracy, Precision, Recall, and ROC-AUC (One-vs-Rest).

Results show that the Random Forest model achieves a Macro F1-score of 0.4486, ROC-AUC of 0.7248, and accuracy of 50.00% on the test set. Feature importance analysis identifies average relative humidity from the previous day (RH_AVG_lag1) and rainfall from the previous day (RR_lag1) as the most dominant predictors findings consistent with the physical principles of tropical atmospheric dynamics. These results confirm that Random Forest can effectively capture temporal weather patterns in tropical regions, although its performance remains constrained by the simplicity of input features and the complexity of local climatic dynamics. This research provides a methodological foundation for developing more adaptive, interpretable, and region-specific machine learning-based weather forecasting systems to support strategic sectors in Indonesia.

Keywords: *Random Forest, Daily Weather Prediction, Machine Learning, BMKG Data, Tropical Climate, Feature Importance, Time-Series Split*