

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rangkaian analisis dan pemodelan yang telah dilaksanakan, penelitian ini menyimpulkan hasil yang menjawab keempat rumusan masalah sebagai berikut.

1. Pearson *Correlation* dan *TabNet* berhasil diterapkan untuk melakukan seleksi variabel pada data radiosonde dalam prediksi intensitas curah hujan harian di Kota Surabaya. Tahapan seleksi dilakukan dengan menggunakan Pearson *Correlation* sebagai penyaringan awal dengan kriteria $|r| \geq 0,07$, yang menghasilkan 18 variabel dari 23 variabel awal. Variabel dengan korelasi tertinggi terhadap curah hujan adalah Humi0 sebesar 0,365735, LCL sebesar -0,322427, 700 sebesar 0,296928, 500 sebesar 0,291545, dan TPW sebesar 0,291313. Selanjutnya, *TabNet* digunakan untuk menentukan tingkat kepentingan variabel dan menghasilkan 15 variabel teratas, yaitu Humi0, LCL, WS, 700, *Height*, KI, LI, CCL, Press0, TPW, 500, 850, LFC, SWEAT, dan CAPE. Pengujian jumlah variabel menggunakan 5, 10, 15, 20, dan seluruh variabel menunjukkan bahwa penggunaan 15 variabel memberikan kinerja terbaik sehingga digunakan sebagai dasar pembentukan variabel masukan pada tahap pemodelan.
2. Ketiga skenario model berhasil diterapkan untuk memprediksi intensitas curah hujan harian berdasarkan data radiosonde. Tahapan pemodelan diawali dengan seleksi variabel menggunakan Pearson *Correlation* dan *TabNet*, kemudian dilanjutkan dengan penentuan *lag* menggunakan *Cross-Correlation Function* (CCF). Hasil CCF menunjukkan bahwa sebagian besar variabel memiliki lag optimal 1 hari, sedangkan variabel 850 memiliki lag 4 hari serta Press0 dan Temp0 memiliki lag 7 hari. Humi0 memiliki hubungan CCF paling kuat dengan curah hujan pada lag 1 dengan nilai korelasi sebesar 0,322576. Hasil penentuan lag tersebut digunakan dalam pembentukan variabel masukan sebelum pemodelan. Berdasarkan tahapan tersebut, tiga skenario yang

diterapkan terdiri atas *Hybrid TabNet-XGBoost*, *Hybrid TabNet-XGBoost* dengan penanganan data curah hujan ekstrem, dan *Hybrid TabNet-SVR*.

3. Hasil evaluasi menggunakan RMSE, MAE, dan MAPE menunjukkan bahwa tidak terdapat satu skenario yang memberikan nilai terbaik pada seluruh metrik. *Hybrid TabNet-XGBoost* tanpa penanganan data ekstrem menghasilkan RMSE sebesar 19,1404 mm, MAE sebesar 11,8515 mm, dan MAPE sebesar 18,53%. Setelah data curah hujan dengan nilai lebih dari 80 mm tidak disertakan dalam proses pelatihan, *Hybrid TabNet-XGBoost* menghasilkan peningkatan berdasarkan RMSE dan MAE, masing-masing menjadi 14,4482 mm dan 10,1127 mm, tetapi MAPE meningkat menjadi 19,03%. Sementara itu, *Hybrid TabNet-SVR* menghasilkan RMSE sebesar 20,8319 mm, MAE sebesar 10,6938 mm, dan MAPE sebesar 12,36%. Dengan demikian, *Hybrid TabNet-XGBoost* dengan penanganan data ekstrem memberikan kinerja terbaik berdasarkan RMSE dan MAE, sedangkan *Hybrid TabNet-SVR* memberikan nilai MAPE terendah. Apabila RMSE dan MAE digunakan sebagai pertimbangan utama dalam mengukur besarnya kesalahan prediksi, maka *Hybrid TabNet-XGBoost* dengan penanganan data ekstrem merupakan skenario dengan kinerja paling baik secara keseluruhan.
4. Hasil prediksi dari ketiga skenario berhasil diimplementasikan ke dalam aplikasi berbasis *Streamlit* sebagai keluaran program. Aplikasi menyajikan hasil prediksi dalam bentuk tabel dan visualisasi sehingga pengguna dapat melihat keluaran masing-masing skenario model. Prediksi tahun 2026 disusun berdasarkan skenario masukan yang tersedia dan menghasilkan keluaran prediksi untuk periode tahun 2026, dengan jumlah data menyesuaikan ketersediaan data masukan pada masing-masing skenario. Hasil prediksi tahun 2026 tidak digunakan untuk mengevaluasi akurasi model karena data aktual curah hujan tahun 2026 belum tersedia sebagai data pembanding. Dengan demikian, implementasi *Streamlit* berfungsi sebagai media penyajian hasil prediksi dari model yang telah dibangun, sedangkan penilaian kinerja model tetap didasarkan pada hasil pengujian menggunakan data tahun 2025..

5.2 Saran Pengembangan

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan selama proses analisis dan pemodelan, beberapa saran yang dapat digunakan sebagai bahan pengembangan penelitian selanjutnya adalah sebagai berikut.

1. Penambahan cakupan dan periode data. Penelitian selanjutnya dapat menggunakan data radiosonde dengan periode pengamatan yang lebih panjang dan cakupan wilayah yang lebih luas untuk meningkatkan jumlah data pelatihan serta memungkinkan model mempelajari pola atmosfer yang lebih beragam. Penambahan periode pengamatan juga dapat digunakan untuk menguji konsistensi kinerja model pada kondisi cuaca dan musim yang berbeda. Penambahan variabel meteorologi dan sumber data. Penelitian selanjutnya dapat mempertimbangkan penambahan variabel atmosfer lainnya, baik dari hasil pengamatan udara atas maupun data pengamatan permukaan dan penginderaan jauh. Penambahan variabel tersebut diharapkan dapat memberikan informasi atmosfer yang lebih lengkap dalam menjelaskan perubahan intensitas curah hujan.
2. Pengembangan metode penanganan curah hujan ekstrem. Penelitian selanjutnya dapat mengeksplorasi metode lain dalam menangani ketidakseimbangan karakteristik data curah hujan, seperti pemberian bobot pada data ekstrem, transformasi target, *oversampling*, atau pendekatan khusus untuk *extreme rainfall*. Hal tersebut diperlukan agar model tidak hanya memiliki kinerja baik pada kondisi curah hujan normal, tetapi juga mampu mempertahankan kemampuan dalam mendeteksi kejadian curah hujan dengan intensitas tinggi. Optimasi dan perbandingan metode *lag*. Penelitian selanjutnya dapat menguji rentang *lag* yang lebih luas serta membandingkan CCF dengan metode penentuan *lag* lainnya. Pengujian tersebut dapat dilakukan secara sistematis untuk memperoleh kombinasi lag yang paling sesuai untuk setiap variabel meteorologi dan mengevaluasi pengaruhnya terhadap kinerja model. Pengembangan model untuk prediksi jangka pendek. Penelitian selanjutnya dapat mengembangkan model untuk menghasilkan prediksi beberapa hari ke depan, misalnya prediksi 1-7 hari, dengan menggunakan data

meteorologi yang tersedia secara aktual atau forecast. Pendekatan tersebut lebih sesuai untuk mendukung kebutuhan informasi cuaca operasional karena hasil prediksi dapat dibandingkan dengan observasi aktual setelah periode prediksi berlangsung.

3. Pengembangan metode *hybrid* dan optimasi model. Penelitian selanjutnya dapat mengeksplorasi kombinasi metode *machine learning* lainnya serta melakukan optimasi *hyperparameter* secara lebih luas. Perbandingan dengan metode seperti *Random Forest*, *LightGBM*, *LSTM*, atau pendekatan *ensemble* dapat dilakukan untuk mengetahui metode yang lebih sesuai dalam menangkap hubungan non linier antara kondisi atmosfer dan intensitas curah hujan.
4. Pengembangan evaluasi model berdasarkan kategori intensitas hujan. Selain menggunakan *RMSE*, *MAE*, dan *MAPE*, penelitian selanjutnya dapat mengevaluasi kemampuan model berdasarkan kategori intensitas curah hujan, seperti tidak hujan, hujan ringan, sedang, lebat, dan ekstrem. Evaluasi tersebut dapat memberikan gambaran yang lebih spesifik mengenai kemampuan model dalam mendeteksi kondisi hujan yang berbeda, khususnya kejadian hujan dengan intensitas tinggi.