

BAB V KESIMPULAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Perbandingan Metode Ensemble dan Single Classifier

Hasil evaluasi menunjukkan bahwa metode ensemble learning secara umum memberikan performa yang lebih baik dibandingkan metode single classifier. Model Stacking mencatat skor tertinggi baik dari segi akurasi maupun macro average F1-score, mengungguli model-model individu seperti Logistic Regression (F1 0.6772), K-Nearest Neighbors (0.6579), dan Decision Tree (0.5948). Ini membuktikan bahwa pendekatan ensemble lebih mampu menangkap kompleksitas data cuaca.

2. Implementasi Model Stacking Classifier

Model Stacking Classifier berhasil diimplementasikan untuk memprediksi kondisi cuaca berbasis klasifikasi biner dengan dua kategori utama, yaitu hujan dan tidak hujan. Model ini memanfaatkan data cuaca historis dari London dengan variabel-variabel penting seperti suhu maksimum, suhu minimum, suhu rata-rata, durasi sinar matahari, radiasi global, tekanan permukaan laut, kelembapan relatif, dan tutupan awan. Tahapan preprocessing data dilakukan secara sistematis melalui pemilihan fitur, normalisasi menggunakan StandardScaler, encoding label target, dan pembagian data menjadi set pelatihan dan pengujian.

3. Evaluasi Model Sebelum dan Sesudah Tuning

Model Stacking Classifier dievaluasi sebelum dan sesudah dilakukan proses hyperparameter tuning. Sebelum tuning, model menghasilkan akurasi pelatihan sebesar 69.79% dan akurasi pengujian sebesar 70.00%, dengan F1-score pada data uji sebesar 0.7004. Setelah dilakukan tuning dengan pendekatan terbaik menggunakan sequential tuning, performa model pada data latih justru mengalami sedikit penurunan, di mana akurasi turun menjadi 69.41%, F1-score menjadi 0.6941, precision sebesar 0.6943, dan recall sebesar 0.6941.

Sementara itu, pada data pengujian terlihat adanya peningkatan yang sangat tipis. Akurasi test meningkat tidak signifikan sebesar +0.0013 menjadi 70.16%, namun F1-score test mengalami peningkatan kecil sebesar +0.0013

menjadi 0.7017. Selain itu, precision test meningkat +0.0015 menjadi 0.7019, dan recall test naik +0.0013 menjadi 0.7016.

Secara keseluruhan, tuning tidak memberikan lonjakan performa yang berarti. Performa model pada data latih sedikit menurun, namun pada data uji terjadi peningkatan yang sangat kecil dan relatif stabil. Hal ini menunjukkan bahwa proses tuning membuat model lebih seimbang dalam melakukan generalisasi, meskipun tanpa adanya peningkatan performa yang signifikan.

4. Kemampuan Model dalam Mengenali Pola Cuaca

Berdasarkan confusion matrix dan classification report pada data pengujian, model mampu mengklasifikasikan kategori "tidak hujan" (label 0) dengan cukup baik (precision 0.72, recall 0.72, f1-score 0.72), namun performanya sedikit lebih rendah dalam mengenali "hujan" (label 1), dengan recall sebesar 0.69 dan f1-score 0.68. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kejadian hujan secara konsisten, kemungkinan besar disebabkan oleh distribusi data yang tidak seimbang antar kelas dan keterbatasan informasi dari variabel yang digunakan.

5.2. Saran

Untuk meningkatkan akurasi dan generalisasi model prediksi cuaca di masa mendatang, penelitian selanjutnya disarankan untuk memperhatikan beberapa hal berikut sebagai bagian dari strategi pengembangan model:

1. Peningkatan Kualitas dan Cakupan Data

Dalam rangka meningkatkan akurasi model, penelitian selanjutnya disarankan untuk menggunakan dataset dengan resolusi temporal yang lebih tinggi misalnya data per jam, serta mencakup lebih banyak wilayah agar model tidak terbatas pada karakteristik cuaca di London saja.

2. Penambahan Variabel Pendukung

Menambahkan fitur eksternal seperti ketinggian wilayah, indeks UV, tingkat polusi udara, atau pergerakan angin dapat membantu model mengenali pola cuaca secara lebih komprehensif. Fitur-fitur ini memberikan informasi tambahan yang relevan terhadap dinamika atmosfer sehingga model dapat memahami konteks data dengan lebih baik. Selain itu, untuk meningkatkan performa prediksi, kolom target yang digunakan sebaiknya lebih spesifik dan sesuai dengan fokus

penelitian, agar model dapat dilatih untuk mengenali pola yang lebih jelas dan tidak terlalu umum.

3. Pengujian Kombinasi dan Alternatif Model

Eksplorasi model alternatif dapat dilakukan dengan menguji penggabungan Stacking Classifier bersama algoritma lain seperti Gradient Boosting, LightGBM, maupun pendekatan berbasis deep learning seperti kombinasi CNN-LSTM. Pendekatan ini bertujuan untuk mengevaluasi sejauh mana metode-metode tersebut mampu meningkatkan akurasi dan stabilitas model prediksi jika dibandingkan dengan metode sebelumnya.

4. Penerapan pada Sistem *Real-Time*

Model yang telah dikembangkan dapat dikemas menjadi sistem prediksi cuaca real-time yang terintegrasi dengan API cuaca. Hal ini akan meningkatkan manfaat praktis model dalam dunia nyata, seperti perencanaan pertanian, logistik, atau mitigasi bencana.

5. Eksplorasi Metode Optimasi Hyperparameter yang Lebih Efisien

Selain Random Search dan konfigurasi manual, disarankan untuk mengeksplorasi teknik optimasi hyperparameter yang lebih canggih seperti Grid Search, Bayesian Optimization, atau bahkan AutoML. Pendekatan ini dapat membantu menemukan kombinasi parameter yang lebih optimal.

6. Evaluasi dengan Skenario dan Dataset yang Berbeda

Pengujian model dengan dataset dari wilayah atau iklim yang berbeda, serta dengan skenario ekstrem seperti hujan deras, badai, atau gelombang panas, dapat membantu mengevaluasi daya tahan dan generalisasi model terhadap kondisi yang tidak umum.

Dengan mempertimbangkan saran-saran tersebut, diharapkan pengembangan model prediksi cuaca berbasis Stacking Classifier dapat mencapai performa yang lebih optimal dan memberikan manfaat nyata dalam berbagai sektor yang membutuhkan prakiraan cuaca yang akurat dan presisi.