

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan analisis terhadap performa empat algoritma machine learning—Regresi Linier, Regresi Logistik, Decision Tree, dan Random Forest—dalam mengklasifikasikan cuaca di Kota Semarang, maka diperoleh kesimpulan sebagai berikut:

1. Algoritma Regresi Linier, meskipun bukan model klasifikasi murni, digunakan sebagai baseline. Model ini menghasilkan nilai AUC sebesar 0,8272, akurasi 75,42%, presisi 75,44%, recall 75,42%, dan F1-score 75,41%. Nilai-nilai ini menunjukkan bahwa Regresi Linier kurang optimal dalam klasifikasi cuaca dibandingkan model lainnya, terutama karena membutuhkan proses threshold untuk dapat digunakan sebagai model klasifikasi biner.
2. Algoritma Regresi Logistik menunjukkan performa yang lebih baik dibanding Regresi Linier. Sebelum dilakukan tuning, model ini memiliki nilai evaluasi yang identik dengan Regresi Linier. Namun, setelah dilakukan hyperparameter tuning, performanya meningkat menjadi AUC sebesar 0,8251, akurasi 74,79%, presisi 74,81%, recall 74,79%, dan F1-score 74,79%.
3. Algoritma Decision Tree mampu menangkap pola klasifikasi cuaca dengan hasil yang cukup baik. Model ini menghasilkan AUC sebesar 0,7534, akurasi 72,69%, presisi 72,85%, recall 72,69%, dan F1-score 72,64%. Hasil ini tidak mengalami peningkatan setelah dilakukan tuning, menandakan keterbatasan fleksibilitas model terhadap data ini.
4. Algoritma Random Forest terbukti memberikan performa terbaik dibandingkan seluruh model lainnya. Sebelum tuning, model ini mencatat AUC sebesar 0,8326, akurasi 74,16%, presisi 74,69%, recall 74,16%, dan F1-score 74,02%. Setelah dilakukan tuning, performanya meningkat

signifikan dengan AUC sebesar 0,8450, akurasi 77,31%, presisi 77,41%, recall 77,31%, dan F1-score 77,29%.

5. Berdasarkan evaluasi metrik klasifikasi, Random Forest merupakan model terbaik dalam mengklasifikasikan cuaca di Kota Semarang. Model ini mengungguli algoritma lainnya dalam seluruh metrik evaluasi yang digunakan, dan direkomendasikan sebagai metode utama untuk implementasi sistem prediksi cuaca berbasis machine learning.

5.2 Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat disampaikan untuk pengembangan penelitian selanjutnya maupun implementasi praktis:

1. Pendalaman Hyperparameter Tuning:

Penelitian ini menunjukkan bahwa tuning hyperparameter memiliki dampak signifikan terhadap peningkatan performa model, khususnya pada algoritma Random Forest dan Logistic Regression. Untuk penelitian selanjutnya, disarankan agar teknik optimasi hyperparameter dikaji lebih mendalam dengan menggunakan metode yang lebih efisien, seperti Random Search, Bayesian Optimization, atau Genetic Algorithm, yang mampu menjelajahi ruang parameter secara lebih efektif dibandingkan grid search tradisional.

2. Eksplorasi Model Lain:

Selain model yang digunakan dalam penelitian ini, peneliti selanjutnya dapat mengeksplorasi algoritma lain seperti Gradient Boosting, XGBoost, atau Neural Networks yang sering kali memberikan performa lebih baik dalam klasifikasi dengan kompleksitas data tinggi.

3. Pemilihan algoritma klasifikasi yang tepat sangat berpengaruh terhadap akurasi dan efektivitas sistem prediksi cuaca. Berdasarkan hasil penelitian, algoritma Random Forest memberikan hasil terbaik dan direkomendasikan untuk digunakan pada sistem prediksi cuaca ke depannya.

4. Penggunaan Linear Regression untuk klasifikasi biner sebaiknya dibatasi hanya sebagai model baseline. Meskipun model ini mampu memberikan gambaran awal performa klasifikasi dengan nilai AUC sebesar 0,8272 dan akurasi 75,42%, Linear Regression bukanlah algoritma yang dirancang untuk tugas klasifikasi. Model ini membutuhkan proses tambahan berupa threshold untuk mengubah output kontinu menjadi kelas biner, dan tidak mampu menangani probabilitas klasifikasi secara langsung. Oleh karena itu, untuk tujuan klasifikasi yang lebih kompleks atau ketika dibutuhkan interpretasi probabilistik, disarankan menggunakan algoritma klasifikasi murni seperti Logistic Regression, Decision Tree, atau Random Forest.

Dengan memperhatikan hal-hal tersebut, diharapkan hasil prediksi yang diperoleh akan lebih optimal dan akurat, serta dapat berkontribusi dalam pengambilan keputusan yang lebih tepat di berbagai bidang yang bergantung pada kondisi cuaca.

