

Kesimpulan

Proyek klasifikasi tingkat obesitas ini berhasil kami laksanakan melalui serangkaian tahapan yang sistematis, komprehensif, dan mendalam, mulai dari preprocessing data hingga deployment model akhir. Pada tahap awal preprocessing, kami melakukan seleksi fitur dengan menghapus atribut yang menunjukkan korelasi sangat rendah terhadap target, yaitu SMOKE dan NCP, guna meningkatkan efisiensi pelatihan tanpa kehilangan informasi penting. Selain itu, kami membersihkan seluruh data dari nilai null, outlier, dan inkonsistensi, lalu melakukan standarisasi pada fitur numerik agar distribusi nilainya lebih seimbang sehingga model dapat belajar secara optimal tanpa bias akibat perbedaan skala. Hasil akhir preprocessing adalah dataset bersih berisi 1.871 rekaman yang siap digunakan untuk pelatihan.

Pada tahap pemodelan dan pelatihan, kami menguji berbagai teknik sampling untuk menangani ketidakseimbangan kelas yang kerap menjadi tantangan pada data multikelas, meliputi No Sampling, Random Oversampling, Random Undersampling, dan SMOTE. Dari eksperimen tersebut, kami menemukan bahwa kombinasi SMOTE dengan algoritma Random Forest secara konsisten menunjukkan performa unggul dengan akurasi mencapai sekitar 97%, serta nilai presisi, recall, dan F1-score yang seimbang dan tinggi di atas 96%. Model ini juga menunjukkan efisiensi komputasi dengan waktu pelatihan yang singkat dan penggunaan CPU yang optimal, menandakan bahwa solusi ini tidak hanya akurat tetapi juga praktis untuk aplikasi nyata.

Untuk memperkuat hasil tersebut, kami melakukan hyperparameter tuning secara ekstensif pada Random Forest dan menghasilkan konfigurasi terbaik dengan parameter `bootstrap=True`, `max_depth=30`, `min_samples_leaf=1`, `min_samples_split=5`, dan `n_estimators=200`. Konfigurasi ini secara signifikan meningkatkan performa model, mengukuhkan keandalan dan stabilitasnya dalam mengklasifikasikan data multikelas dengan distribusi tidak seimbang. Evaluasi mendalam melalui confusion matrix dan classification report menunjukkan kemampuan model kami dalam memprediksi hampir seluruh kelas dengan sangat presisi, meminimalkan kesalahan klasifikasi sekaligus menjaga keseimbangan antara presisi dan recall.

Akhirnya, model terbaik yang telah kami optimasi dan validasi berhasil kami simpan secara efisien dan deploy melalui platform Streamlit dengan integrasi GitHub, membuka akses penggunaan secara real-time dan mudah bagi pengguna akhir. Keseluruhan proses ini membuktikan bahwa penerapan SMOTE sebagai teknik sampling untuk mengatasi imbalance data, dipadukan dengan pengoptimalan hyperparameter Random Forest, mampu menghasilkan model klasifikasi obesitas yang tidak hanya sangat akurat dan stabil, tetapi juga efisien secara komputasi dan siap diimplementasikan dalam konteks dunia nyata.

Link Streamlit: <https://capstonebengkod-obesitas-mafif.streamlit.app/>