BAB V
PENUTUP

5.1 Kesimpulan

Penelitian ini berfokus pada tantangan dalam menangani ketidakseimbangan data pada berbagai dataset menggunakan algoritma Random Forest dengan pendekatan balancing SMOTE dan NearMiss. Berdasarkan hasil implementasi dan evaluasi terhadap delapan dataset yang berasal dari berbagai domain seperti kesehatan, keuangan, dan industri, diperoleh beberapa kesimpulan utama berikut:

1. Pengaruh ketidakseimbangan kelas terbukti menurunkan performa model klasifikasi, khususnya dalam mendeteksi kelas minoritas. Model cenderung bias terhadap kelas mayoritas jika tidak dilakukan penanganan ketidakseimbangan data.

2. Metode SMOTE dan NearMiss berhasil diimplementasikan dalam sistem klasifikasi universal berbasis web. Sistem ini mampu melakukan balancing secara otomatis dan fleksibel sesuai karakteristik dataset yang diunggah oleh pengguna.

3. Hasil evaluasi menunjukkan bahwa SMOTE lebih efektif dibandingkan NearMiss dalam meningkatkan performa klasifikasi, khususnya pada metrik precision, recall, dan F1-score. Ini menandakan bahwa teknik oversampling seperti SMOTE mampu memperbaiki representasi kelas minoritas secara lebih optimal dibandingkan teknik undersampling.

Dengan demikian, penelitian ini berhasil menjawab rumusan masalah yang diajukan dan membuktikan bahwa penerapan metode balancing yang tepat dapat meningkatkan kinerja model klasifikasi terhadap dataset tidak seimbang secara signifikan.

5.2 Saran

Meskipun sistem berhasil dibangun dan diuji, terdapat sejumlah keterbatasan dan peluang perbaikan yang dapat dijadikan dasar untuk pengembangan sistem di masa mendatang, antara lain:

1. Belum mendukung visualisasi distribusi kelas secara grafik (histogram atau pie chart), sehingga analisis awal terhadap kondisi dataset masih terbatas pada tampilan tabular. Penambahan grafik dapat meningkatkan pemahaman pengguna terhadap ketidakseimbangan data secara visual.
2. Sistem saat ini hanya mendukung satu algoritma klasifikasi, yaitu Random Forest. Untuk ke depan, perlu ditambahkan opsi algoritma lain seperti SVM, KNN, atau Gradient Boosting agar pengguna dapat membandingkan performa berbagai model terhadap dataset mereka.
3. Beberapa dataset dengan fitur kategorikal yang kompleks memerlukan penyesuaian encoding khusus. Sistem belum sepenuhnya menangani kasus di mana fitur kategorikal memiliki label yang tidak konsisten atau sangat banyak kelas.
4. Belum mendukung fitur upload dataset dalam format selain CSV dan belum tersedia deteksi otomatis terhadap target label jika tidak ditentukan oleh pengguna.
5. Fitur prediksi manual saat ini masih terbatas pada satu baris input. Pengembangan lebih lanjut dapat mencakup kemampuan untuk melakukan prediksi batch atau multi-input sekaligus.
6. Sistem belum memiliki validasi input yang ketat dan tidak menampilkan proses preprocessing seperti encoding dan imputasi secara eksplisit ke antarmuka. Ini menyulitkan pengguna untuk memahami transformasi data secara mendalam jika ingin belajar dari sistem.
7. Penggunaan metode balancing belum sepenuhnya otomatis berdasarkan kondisi distribusi kelas. Sistem sebaiknya mampu mendeteksi secara mandiri apakah balancing perlu dilakukan dan memilih metode yang paling optimal berdasarkan rasio kelas.

Dengan memperhatikan dan mengatasi keterbatasan-keterbatasan tersebut, diharapkan sistem klasifikasi universal yang dikembangkan dalam penelitian ini dapat ditingkatkan fungsionalitasnya, digunakan pada lebih banyak skenario, serta memberikan hasil klasifikasi yang semakin akurat dan interpretatif..