

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan oleh penulis mengenai analisis penggunaan teknik *oversampling* pada model XGBoost untuk menangani ketidakseimbangan kelas, dapat disimpulkan sebagai berikut:

1. Model XGBoost tanpa teknik *oversampling* menunjukkan kinerja akurasi yang tinggi, yaitu di atas 0,90. Namun, nilai metrik *g-mean* yang diperoleh cukup rendah, yaitu di bawah 0,50. Hal ini mengindikasikan bahwa model cenderung lebih baik dalam mengklasifikasikan individu yang sehat (kelas mayoritas), dibandingkan dengan individu yang menderita penyakit jantung (kelas minoritas). Dengan kata lain, model mengalami kesulitan dalam mengidentifikasi kelas minoritas.
2. Model XGBoost dengan teknik *oversampling*, baik ROS, SMOTE, maupun ADASYN, menunjukkan nilai metrik *g-mean* yang jauh lebih tinggi, yaitu di atas 0,75. Hasil ini mengindikasikan bahwa model-model tersebut mampu mengklasifikasikan kelas mayoritas dan minoritas lebih baik dibandingkan dengan model XGBoost tanpa *oversampling*.
3. Terdapat *trade-off* antara waktu komputasi dan kinerja model ketika menerapkan teknik *oversampling*. Peningkatan rasio *sampling* secara langsung meningkatkan waktu komputasi, namun tidak selalu berbanding lurus dengan peningkatan kinerja model. Di antara ketiga teknik *oversampling* yang digunakan, ROS merupakan teknik yang paling efisien secara komputasi, sedangkan ADASYN adalah yang paling lambat. Penggunaan rasio *sampling* melebihi nilai 0,1 memiliki potensi terjadinya *overfitting* pada model, yaitu model menjadi terlalu cocok dengan data latih sehingga performanya menurun pada data uji ketika terlalu banyak sampel sintetis ditambahkan.

#### 5.2 Saran

Berdasarkan temuan penelitian ini, terdapat beberapa saran untuk penelitian selanjutnya sebagai berikut:

1. Melakukan perbandingan efektivitas berbagai teknik *oversampling* pada dataset dengan tingkat ketidakseimbangan yang lebih rendah untuk menguji apakah penggunaan *oversampling* tetap efektif pada kondisi data yang lebih seimbang.
2. Menerapkan teknik *resampling* lain, seperti *undersampling* atau kombinasi *undersampling* dan *oversampling*, untuk mengidentifikasi metode *resampling* yang paling efektif dalam mengatasi ketidakseimbangan kelas pada dataset yang diteliti.
3. Menggunakan teknik *hyperparameter tuning* lain, seperti *Bayesian Optimization*, *Tree-structured Parzen Estimator (TPE)*, atau teknik lainnya untuk membandingkan efektivitasnya dengan metode *Grid Search*. Selanjutnya, perluasan ruang pencarian untuk setiap parameter dapat dilakukan untuk meningkatkan kemungkinan menemukan kombinasi *hyperparameter* yang optimal.