

BAB 5

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan, berikut merupakan kesimpulan yang didapatkan dari hasil penelitian.

1. Penelitian ini berhasil mengimplementasikan algoritma *Random Forest* untuk deteksi penyakit jantung koroner menggunakan data *Framingham*. Kombinasi teknik *hyperparameter tuning* dan SMOTE terbukti mampu meningkatkan performa model, khususnya dalam mendeteksi kelas minoritas. Hasil ini memberikan wawasan baru terkait kesuksesan *Random Forest* dalam permasalahan klasifikasi medis.
2. Hasil terbaik dalam mendeteksi penyakit jantung ditemukan pada skenario keempat, yaitu ketika model *Random Forest* dioptimalkan menggunakan teknik SMOTE dan pengaturan *hyperparameter tuning*. Pada skenario ini, hasil yang dicapai pada rasio data *train* dan *test* 70:30 yaitu akurasi sebesar 73.89%, dengan nilai *precision* 25.89%, *recall* 38.67%, dan F1-score 31.02%. Hasil tersebut diperoleh dari hasil setelah pengaturan *hyperparameter tuning* terbaik, seperti jumlah pohon sebanyak 125, kedalaman maksimum pohon hingga 75, dan metode penghitungan pembagian data menggunakan *log loss*. Kemudian, jumlah maksimum cabang keputusan dalam pohon adalah 410 dan akan membuat keputusan akhir jika terdapat minimal 6 data di cabang tersebut, serta hanya membagi suatu *node* jika terdapat setidaknya 2 data di dalamnya. Selain itu, berdasarkan hasil evaluasi terhadap metrik performa model, akurasi dan *precision* terbaik diperoleh dari skenario ketiga (*hyperparameter tuning*), masing-masing sebesar 85% dan 58.33%, sedangkan *recall* dan F1-score terbaik berasal dari skenario keempat (*hyperparameter tuning* dan SMOTE), yaitu sebesar 38.67% dan 31.02%.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, berikut merupakan saran-saran yang dapat dilakukan untuk penelitian selanjutnya.

1. Mengingat data yang digunakan pada *dataset framingham.csv* tergolong tidak seimbang pada kelas klasifikasi, disarankan untuk menerapkan teknik penanganan data *imbalanced* yang lain. Selain SMOTE, dapat menggunakan teknik seperti ADASYN (*Adaptive Synthetic Sampling Approach*) untuk meningkatkan performa klasifikasi terutama peningkatan metrik evaluasi.
2. Melakukan *feature selection* seperti *Feature Importance*, *Recursive Feature Elimination* (RFE), *Boruta* untuk mengeliminasi fitur-fitur yang kurang relevan atau tidak berkontribusi signifikan terhadap kinerja model, sehingga dapat meningkatkan efisiensi dan akurasi prediksi.
3. Disarankan untuk menambah fitur-fitur yang memiliki relevansi medis dalam proses pelatihan model. Dengan mempertimbangkan fitur yang secara klinis penting, seperti kadar gula darah HbA1c, riwayat keluarga penyakit jantung, dislipidemia, dan obesitas agar model diharapkan dapat mengenali pola risiko yang lebih kompleks dan meningkatkan performa prediksi terhadap penyakit jantung.

