

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan :

1. Perbedaan performa model antara data yang mengalami *data leakage* dan yang bebas dari *data leakage* ditunjukkan melalui nilai metrik evaluasi yang dihasilkan. Skenario dengan *data leakage* cenderung menghasilkan nilai ROC-AUC, *accuracy*, *precision*, *recall*, dan F1-score yang lebih tinggi dibandingkan skenario tanpa *data leakage*. Namun, peningkatan performa tersebut bersifat artifisial dan tidak sepenuhnya mencerminkan kemampuan generalisasi model terhadap data baru. Evaluasi menggunakan berbagai kombinasi *random seed* dan *test size* menunjukkan bahwa skenario tanpa *data leakage* menghasilkan performa yang lebih stabil dan konsisten, sehingga lebih valid sebagai dasar penilaian kinerja model.

Berdasarkan hasil tersebut, *dataset* klinis *pediatric appendicitis* dipilih sebagai objek penelitian karena merepresentasikan data klinis nyata yang berpotensi digunakan dalam pengembangan model prediktif diagnosis. Hasil analisis menunjukkan bahwa meskipun *dataset* yang sama dapat menghasilkan performa model yang tinggi, *dataset* tersebut hanya layak digunakan apabila diproses melalui *preprocessing* tanpa *data leakage*. Oleh karena itu, data yang dibutuhkan untuk pengembangan dan evaluasi

model *machine learning* yang valid adalah data klinis yang telah melalui seluruh tahapan *preprocessing* secara bebas *data leakage*.

2. Identifikasi dan pembuktian adanya *data leakage* pada tahap *preprocessing* dilakukan melalui perancangan skenario *preprocessing* yang berbeda, yaitu skenario tanpa *data leakage* dan skenario dengan *data leakage* yang disengaja. *Data leakage* dibuktikan dengan membandingkan hasil evaluasi performa model, analisis kurva ROC, serta perbedaan kestabilan nilai antara *cross-validation* dan data testing. Peningkatan nilai performa yang tidak konsisten dan cenderung lebih optimistis pada skenario dengan *data leakage* menjadi indikator utama terjadinya kebocoran informasi pada tahap *preprocessing*. Temuan ini memperkuat bahwa identifikasi *data leakage* bukan hanya bersifat metodologis, tetapi merupakan prasyarat utama dalam menentukan kelayakan *dataset* klinis untuk digunakan pada pengembangan model *machine learning* yang valid dan dapat diandalkan.

5.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang telah diperoleh, beberapa saran berikut diajukan sebagai bahan pertimbangan bagi penelitian selanjutnya maupun praktik penerapan *machine learning* :

1. Penelitian selanjutnya disarankan untuk menggunakan *dataset* dengan ukuran yang lebih besar dan berasal dari berbagai sumber guna meningkatkan kemampuan generalisasi model.

2. Perlu dilakukan eksplorasi metode *feature selection* multivariat agar interaksi antar fitur dapat dimanfaatkan secara lebih optimal.
3. Penggunaan model lain, seperti *ensemble* lanjutan atau *deep learning*, dapat dipertimbangkan untuk membandingkan dampak *data leakage* pada arsitektur model yang lebih kompleks.
4. Dalam praktik *machine learning*, khususnya pada data klinis, disarankan untuk selalu memisahkan data *training* dan *testing* sebelum seluruh proses *preprocessing* guna mencegah *data leakage*.

