

## BAB 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai klasifikasi penyakit tumor payudara menggunakan metode *Random Forest*, maka diperoleh kesimpulan sebagai berikut:

1. Metode *Random Forest* dapat digunakan secara efektif untuk mengklasifikasikan tumor payudara menjadi dua kelas, yaitu tumor jinak dan tumor ganas. Proses klasifikasi dilakukan melalui beberapa tahapan, yaitu pengumpulan data rekam medis, analisis deskriptif variabel numerik dan kategorik, data *preprocessing* yang meliputi *encoding* variabel kategorik serta pembagian data menjadi data latih dan data uji, *hyperparameter tuning*, serta evaluasi model menggunakan *confusion matrix* dan parameter evaluasi.
2. Tingkat akurasi metode *Random Forest* dalam mengklasifikasikan tumor payudara menunjukkan hasil yang sangat baik pada seluruh skenario pembagian data. Pada rasio data 80:20, model menghasilkan akurasi sebesar 97,2 %, pada rasio 75:25 sebesar 97%, dan pada rasio 70:30 sebesar 96%. Selain itu, nilai presisi dan *recall* yang tinggi pada ketiga rasio pembagian data juga menunjukkan bahwa model tidak hanya mampu memberikan prediksi yang tepat secara keseluruhan, tetapi juga efektif dalam mendeteksi kasus tumor payudara dengan tingkat kesalahan yang rendah.

Secara keseluruhan, hasil penelitian ini membuktikan bahwa model *Random Forest* memiliki kinerja yang baik dan stabil dalam mengklasifikasikan tumor payudara meskipun proporsi data latih dan data uji yang berbeda. Hasil penelitian ini menunjukkan bahwa metode *Random Forest* layak digunakan sebagai pendekatan dalam mendukung proses klasifikasi tumor payudara berbasis data rekam medis.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya, yaitu sebagai berikut:

1. Menggunakan dataset dengan jumlah data yang lebih besar dan lebih beragam, baik dari sisi karakteristik pasien seperti riwayat keluarga, status hormonal, dan gaya hidup, maupun dari sisi sumber data yang lebih bervariasi, agar kemampuan model dapat diuji secara lebih luas.
2. Penggunaan metode klasifikasi lain seperti *Support Vector Machine* (SVM) atau *K-Nearest Neighbor* (KNN) dapat dipertimbangkan untuk membandingkan kinerja dengan metode *Random Forest* sehingga diperoleh model terbaik, karena setiap algoritma memiliki karakteristik dan mekanisme pembelajaran yang berbeda dalam menangani pola data, sehingga memungkinkan ditemukannya performa yang lebih optimal pada *dataset* yang digunakan.

