

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada bab hasil dan pembahasan telah dibuat model *Random Forest* untuk klasifikasi pasien positif dan negatif tuberkulosis. Berdasarkan hasil dan pembahasan tersebut, diperoleh kesimpulan:

1. Model klasifikasi *Random Forest* yang dibagi menjadi tiga rasio data *split*, yaitu 70:30, 80:20, dan 90:10, dibentuk dengan menerapkan kombinasi parameter terbaik yang telah diperoleh sebelumnya. Penerapan ketiga model *Random Forest* tersebut dalam mengklasifikasikan pasien tuberkulosis dalam kelas positif dan negatif menghasilkan performa sebagai berikut.
 - a. Rasio data *split* 70:30, dari total 8.450 data *testing*, model mampu memprediksi secara benar 3.152 pasien positif tuberkulosis dan 2.778 pasien negatif tuberkulosis. Namun, masih ditemukan kesalahan klasifikasi, yaitu 1.073 pasien negatif tuberkulosis yang salah diklasifikasikan sebagai pasien positif tuberkulosis dan 1.447 pasien positif tuberkulosis yang salah diklasifikasikan sebagai pasien negatif tuberkulosis.
 - b. Rasio data *split* 80:20, dari total 5.663 data *testing*, model mampu memprediksi secara benar 2.127 pasien positif tuberkulosis dan 1.850 pasien negatif tuberkulosis. Namun, masih ditemukan kesalahan klasifikasi, yaitu 690 pasien negatif tuberkulosis yang salah diklasifikasikan sebagai pasien positif tuberkulosis dan 966 pasien positif tuberkulosis yang salah diklasifikasikan sebagai pasien negatif tuberkulosis.
 - c. Rasio data *split* 90:10, dari total 2.817 data *testing*, model mampu memprediksi secara benar 1.079 pasien positif tuberkulosis dan 928 pasien negatif tuberkulosis. Namun, masih ditemukan kesalahan klasifikasi, yaitu 330 pasien negatif tuberkulosis yang salah diklasifikasikan sebagai pasien positif tuberkulosis dan 480 pasien positif tuberkulosis yang salah diklasifikasikan sebagai pasien negatif tuberkulosis.
2. Berdasarkan hasil evaluasi klasifikasi, diperoleh akurasi dari masing-masing model *Random Forest*. Model dengan rasio data *split* 70:30, 80:20, dan 90:10

menghasilkan akurasi berturut-turut sebesar 70,1%, 70,6%, dan 71,2%. Selain itu, nilai metrik evaluasi lainnya menunjukkan performa yang relatif sebanding, dengan nilai *precision*, *recall*, dan *F1-score* yang tidak berbeda secara signifikan pada setiap rasio data *split*. Model *Random Forest* dengan rasio data *split* 90:10 menghasilkan nilai akurasi tertinggi, diikuti oleh model dengan rasio 80:20 dan 70:30. Nilai performa model yang dihasilkan relatif serupa antar rasio data *split*, yang menunjukkan bahwa kinerja model lebih lebih dipengaruhi fitur pada *dataset* dibandingkan oleh pemilihan rasio pembagian data. Dengan demikian, algoritma *Random Forest* memiliki stabilitas yang baik dalam melakukan klasifikasi pada variasi pembagian data.

5.2 Saran

Berdasarkan hasil analisis dan pembahasan, berikut saran untuk penelitian selanjutnya:

1. melakukan pengembangan dan seleksi fitur untuk meningkatkan relevansi antara variabel input dengan variabel target,
2. melakukan eksplorasi metode penanganan ketidakseimbangan data selain SMOTE, seperti *hybrid sampling*, untuk mengetahui pengaruhnya terhadap peningkatan performa model klasifikasi, dan
3. melakukan perbandingan performa dengan metode klasifikasi lain, seperti *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), atau algoritma klasifikasi lainnya untuk mengetahui metode yang paling optimal.