

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengujian klasifikasi penyakit diabetes mellitus mempergunakan metode SMOTE Random Forest, beberapa kesimpulan dapat ditarik sebagai berikut:

- a. Proses penyusunan model klasifikasi *Random Forest* melibatkan dua skenario pelatihan, yaitu menggunakan data yang tidak dengan tahap SMOTE dan menggunakan data yang telah melalui proses SMOTE, dengan tujuan untuk membandingkan performa model dalam kedua kondisi tersebut. Model dibangun dan dilatih mempergunakan *library Scikit-Learn*, dengan parameter *n_estimators* diatur sebanyak 100 untuk menentukan jumlah pohon keputusan yang dibangun dalam model, serta *random_state* diatur menjadi 42 untuk memastikan hasil yang dapat direproduksi.
- b. Faktor-faktor utama yang berpengaruh terhadap penyakit diabetes *mellitus* pada hasil klasifikasi menggunakan algoritma *Random Forest* menunjukkan perbedaan antara tahapan tanpa SMOTE dan dengan SMOTE. pada hasil klasifikasi Random Forest tanpa SMOTE, faktor-faktor yang berpengaruh adalah gula darah puasa, gula darah acak, gula darah 2 jam PP, HbA1c, sistolik. Sedangkan pada hasil klasifikasi Random Forest dengan SMOTE, faktor-faktor utama yang berpengaruh adalah gula darah 2 jam PP, gula darah puasa, HbA1c, gula darah acak, sistolik.
- c. Performa yang dihasilkan model dengan data latih yang melewati tahapan SMOTE memperlihatkan performa yang lebih baik dibanding model yang dilatih pada data yang tidak melewati tahapan SMOTE. Model yang dilatih menggunakan data yang melewati tahapan SMOTE menunjukkan kesalahan prediksi lebih kecil dibandingkan data yang tidak melewati tahapan SMOTE. Model yang dilatih dengan SMOTE pada rasio 80:20 dan 90:10 menghasilkan performa 98%. Sedangkan model yang dilatih tanpa SMOTE memiliki performa 93%. Rasio 70:30 juga menghasilkan performa yang lebih tinggi pada tahapan SMOTE dengan akurasi 97% banding 92%.

5.2 Saran

Berdasarkan hasil penelitian ini, berikut adalah beberapa saran yang dapat diberikan:

- a. Data yang digunakan bisa ditambah sehingga model memiliki lebih banyak input selama proses *training* yang dapat meningkatkan akurasi model.
- b. Dapat menggunakan metode klasifikasi lainnya seperti SVM, Xgboost, dan *Artificial Neural Network* untuk mengetahui bagaimana perbandingan performa yang dihasilkan.