

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dari rangkaian penelitian yang telah dilakukan mengenai klasifikasi penyakit Polycystic Ovary Syndrome (PCOS) menggunakan algoritma Random Forest serta penerapan metode SMOTE, dapat disimpulkan sebagai berikut:

1. Penelitian ini menerapkan serangkaian tahapan pengolahan data yang meliputi data understanding, preprocessing, exploratory data analysis (EDA), pemilihan fitur, pemodelan, hingga evaluasi. Pada tahap preprocessing dilakukan penanganan nilai kosong, pembersihan data, dan pengkodean variabel kategorikal agar data dapat diproses oleh algoritma machine learning. Tahap EDA dilakukan untuk memahami distribusi data, mengidentifikasi adanya ketidakseimbangan kelas, serta melihat hubungan antar fitur dengan variabel target. Pemilihan fitur dilakukan berdasarkan analisis pentingnya fitur (feature importance) pada model Random Forest. Selanjutnya, pemodelan dilakukan dengan membangun dua skenario model, yaitu Random Forest tanpa SMOTE dan Random Forest dengan SMOTE. Tahap akhir berupa evaluasi menggunakan classification report, confusion matrix, kurva ROC, dan nilai AUC untuk menilai kemampuan model dalam membedakan kelas PCOS dan tidak PCOS.
2. Model Random Forest tanpa SMOTE menghasilkan akurasi sebesar 87% pada data uji, dengan nilai AUC sebesar 0.948, sedangkan Random Forest dengan SMOTE juga menghasilkan akurasi yang sama yaitu 89%, namun memiliki peningkatan pada recall kelas PCOS dan nilai AUC menjadi 0.950.
3. Penerapan SMOTE pada data latih memberikan pengaruh yang positif terhadap kemampuan model dalam memprediksi kelas PCOS yang merupakan kelas minoritas. Pada kelas PCOS, nilai recall meningkat dari 0,75 menjadi 0,78 dan F1-score meningkat dari 0,80 menjadi 0,81 setelah

SMOTE diterapkan. Nilai AUC juga meningkat dari 0,945 menjadi 0,950, yang menunjukkan bahwa model dengan SMOTE memiliki kemampuan pemisahan kelas yang lebih baik. Meskipun precision untuk kelas PCOS sedikit menurun (dari 0,86 menjadi 0,83) akibat meningkatnya jumlah false positive karena model menjadi lebih sensitif dalam mengenali kasus PCOS dan mengurangi risiko pasien yang terlewat (false negative). Dengan demikian, SMOTE terbukti efektif membantu model menangani ketidakseimbangan kelas pada dataset klinis PCOS.

4. Model klasifikasi PCOS berhasil diimplementasikan dalam aplikasi berbasis web menggunakan Streamlit. Model yang telah dilatih dan disimpan dalam format .sav diunggah ke GitHub dan terintegrasi dengan Streamlit Cloud. Aplikasi ini mampu menerima input data klinis dari pengguna, menjalankan proses prediksi menggunakan model Random Forest terbaik, serta menampilkan hasil klasifikasi beserta probabilitas tiap kelas secara interaktif.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, adapun beberapa saran yang dapat penulis berikan untuk perbaikan pada penelitian selanjutnya adalah sebagai berikut.

1. Penelitian selanjutnya disarankan untuk menggunakan dataset dengan jumlah sampel yang lebih besar.
2. Peneliti berikutnya dapat mempertimbangkan untuk menggunakan atau membandingkan algoritma lain, seperti XGBoost, Gradient Boosting, atau metode berbasis deep learning, sehingga dapat diketahui algoritma yang memberikan kinerja lebih baik untuk kasus klasifikasi PCOS.
3. Aplikasi web yang dikembangkan masih bersifat prototipe dan belum melalui uji coba langsung dengan tenaga medis atau pengguna akhir. Penelitian selanjutnya dapat melakukan pengujian kegunaan (usability testing), menambahkan fitur manajemen data dan autentikasi pengguna, serta mengintegrasikan aplikasi dengan sistem informasi klinis yang sudah ada. Dengan demikian, aplikasi dapat dikembangkan menjadi alat bantu

pendukung keputusan yang lebih siap digunakan dalam konteks praktik klinis.

