

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan serangkaian proses penelitian yang telah dilakukan dapat disimpulkan bahwa penelitian ini berhasil mengimplementasikan dan membandingkan tiga algoritma *machine learning* yaitu Random Forest, XGBoost, dan *Support Vector Machine* (SVM) untuk prediksi dini PCOS menggunakan dataset klinis yang terdiri dari 541 sampel pasien dari 10 rumah sakit di Kerala, India. Ketiga algoritma dilatih menggunakan skema *Stratified 5-Fold Cross-Validation* yang terintegrasi dengan SMOTE *pipeline* untuk mengatasi ketidakseimbangan kelas, dengan konfigurasi parameter yang seragam untuk menjaga objektivitas komparasi.

Berdasarkan hierarki kriteria seleksi tiga tahap yang telah ditetapkan yang mana Recall $\geq 0,80$ sebagai ambang batas klinis utama, F1-Score sebagai penyeimbang, dan AUC-ROC sebagai penentu akhir, Random Forest ditetapkan sebagai model paling optimal. Random forest mencatat nilai Recall 0,8078, F1-Score 0,8454, dan AUC-ROC tertinggi sebesar 0,9576 dengan standar deviasi yang secara konsisten lebih kecil dibandingkan XGBoost pada seluruh metrik evaluasi AUC-ROC ($\pm 0,0098$) sebagaimana terlihat pada distribusi Box Plot antar-lipatan yang mengindikasikan konsistensi kinerja terbaik di antara ketiga algoritma. SVM menunjukkan kinerja terendah dengan nilai Recall 0,7800 yang tidak memenuhi ambang batas minimum klinis meskipun SMOTE telah diterapkan. Sementara XGBoost mencatat Recall tertinggi (0,8471) namun sedikit tertinggal pada metrik AUC-ROC dibandingkan Random Forest. Evaluasi final pada data uji independen mengonfirmasi konsistensi kinerja Random Forest dengan Recall 0,8611 dan AUC-ROC 0,9427, membuktikan model tidak mengalami *overfitting* yang signifikan.

Penerapan SMOTE secara eksklusif di dalam setiap lipatan *Stratified K-Fold* melalui *imblearn.pipeline.pipeline* terbukti merupakan pendekatan metodologis yang valid untuk mengatasi ketidakseimbangan kelas. Desain ini memastikan data uji senantiasa mempertahankan distribusi kelas aslinya ($\approx 2:1$), sehingga seluruh metrik evaluasi yang dilaporkan merupakan cerminan kemampuan generalisasi model yang objektif dan bebas dari bias kebocoran data. Perbandingan antara skenario tanpa SMOTE dan dengan SMOTE mengonfirmasi efektivitas teknik ini dengan adanya peningkatan konsisten pada metrik Recall di ketiga algoritma.

Analisis Permutation Feature Importance mengidentifikasi *Follicle No. (R)*

sebagai fitur klinis paling berpengaruh dengan nilai *importance* sebesar 0,0598 jauh melampaui fitur kedua. Temuan ini sangat konsisten dengan kriteria diagnostik Rotterdam yang menjadikan morfologi ovarium *polikistik* dan *hiperandrogenisme* klinis sebagai pilar utama diagnosis PCOS, membuktikan bahwa model Random Forest tidak hanya akurat secara statistik tetapi juga logis secara medis.

Model Random Forest terpilih berhasil diimplementasikan ke dalam prototipe sistem skrining berbasis web menggunakan kerangka kerja Streamlit. Prototipe yang dihasilkan mampu menerima 10 parameter klinis dari pengguna dan menghasilkan prediksi risiko PCOS beserta nilai probabilitasnya secara *real-time*, sebagaimana dibuktikan melalui pengujian dengan data klinis contoh yang menghasilkan prediksi positif PCOS dengan probabilitas 98,5%. Keberhasilan implementasi ini membuktikan kelayakan teknis penerapan model komputasional sebagai alat bantu skrining awal PCOS dalam skenario klinis sederhana.

5.2 Saran

Berdasarkan temuan dan keterbatasan yang diidentifikasi selama penelitian, beberapa saran dikemukakan untuk penelitian dan pengembangan selanjutnya. Pertama, disarankan untuk menerapkan teknik *hyperparameter tuning* pada ketiga algoritma untuk mengeksplorasi apakah optimasi parameter dapat lebih jauh meningkatkan nilai Recall terutama pada SVM yang dalam penelitian ini belum *parameter default*. Kedua, mengeksplorasi perbandingan berbagai varian SMOTE untuk evaluasi apakah teknik yang lebih canggih dapat meningkatkan sensitivitas model terhadap kelas minoritas secara lebih signifikan. Ketiga, disarankan untuk melibatkan dokter spesialis obstetri dan ginekologi dalam proses evaluasi kegunaan sistem, sehingga dapat disesuaikan dengan kebutuhan praktis di lingkungan klinis yang nyata. Terakhir, penelitian selanjutnya dapat mengintegrasikan metode *Explainable AI (XAI)* yang lebih komprehensif seperti SHAP (*SHapley Additive exPlanations*) sehingga pengguna mendapatkan penjelasan visual mengenai faktor klinis mana yang paling berkontribusi terhadap prediksi pada setiap kasus pasien secara individual.