

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan keseluruhan rangkaian penelitian, tahapan eksperimen, dan hasil evaluasi performa yang telah diuraikan pada bab-bab sebelumnya mengenai komparasi algoritma *machine learning* pada klasifikasi penyakit diabetes, dapat ditarik beberapa kesimpulan utama. Pertama, penerapan metode *Synthetic Minority Over-sampling Technique* (SMOTE) terbukti sangat krusial dan efektif dalam mengatasi masalah ketidakseimbangan kelas (*imbalanced data*) yang melekat pada Pima Indians Diabetes Dataset. Dengan menyintesis data pada kelas minoritas hingga mencapai distribusi yang setara dengan kelas mayoritas, model *machine learning* yang dilatih tidak lagi memiliki kecenderungan bias terhadap prediksi pasien sehat. Keberhasilan intervensi data ini dibuktikan secara empiris oleh tingginya metrik *Recall* (sensitivitas) pada seluruh algoritma yang diuji, yakni berada di rentang 88,88% hingga 92,59%. Hal ini menandakan bahwa model memiliki tingkat kepekaan yang sangat tinggi dalam mendeteksi pasien yang benar-benar berisiko diabetes dan berhasil meminimalkan tingkat *false negative* secara signifikan.

Selanjutnya, dari hasil komparasi ketiga arsitektur *ensemble tree* yang diuji, algoritma XGBoost terbukti memberikan performa prediktif yang paling superior dan seimbang secara keseluruhan. XGBoost memimpin dominasi metrik evaluasi dengan tingkat akurasi mencapai 85,71%, *Precision* sebesar 75,00%, *F1-Score* di angka 81,35%, dan tingkat separabilitas *ROC-AUC* sebesar 87,85%. Kemampuan XGBoost dalam mengendalikan *overfitting* melalui mekanisme regularisasinya, yang dipadukan dengan optimisasi *threshold* ke angka 0,433, memungkinkannya untuk menekan angka peringatan palsu (*false positive*) jauh lebih baik dibandingkan model lainnya. Arsitektur ini membuktikan kemampuannya dalam memberikan diagnosis medis yang andal tanpa banyak mengorbankan ketepatan prediksi pada pasien yang sesungguhnya sehat.

Meskipun XGBoost unggul secara global, algoritma CatBoost (dengan akurasi 84,41%) dan LightGBM (dengan akurasi 83,11%) juga mencatatkan sensitivitas deteksi (*Recall*) yang lebih ekstrem, yakni menyentuh angka 92,59%. CatBoost tampil sebagai alternatif model yang sangat tangguh karena mampu menjaga rasio alarm palsu agar tetap rendah di tengah sensitivitasnya yang sangat tinggi, sehingga menjadikannya pilihan yang ideal apabila protokol medis menerapkan toleransi nol (*zero tolerance*) terhadap pasien berisiko yang tidak terdeteksi. Pada akhirnya, seluruh *pipeline* prapemrosesan yang diusulkan dalam penelitian ini mulai dari imputasi nilai kosong menggunakan *median*, penyeimbangan kelas melalui SMOTE, pembagian rasio data latih-uji sebesar 90:10, hingga kalibrasi *threshold* probabilitas terbukti secara empiris mampu mengungguli berbagai performa penelitian terdahulu. Hal ini mengukuhkan bahwa rekayasa kebersihan dan keseimbangan fitur data sama krusialnya dengan pemilihan algoritma *machine learning* itu sendiri.

5.2 Saran

Penelitian ini telah berhasil menghasilkan model klasifikasi yang andal untuk deteksi dini penyakit diabetes, namun masih terdapat berbagai ruang untuk pengembangan lebih lanjut. Bagi peneliti selanjutnya yang ingin mengembangkan atau melanjutkan penelitian di bidang prediksi medis ini, disarankan untuk mengeksplorasi teknik penanganan *imbalanced data* yang lebih mutakhir. Alih-alih hanya menggunakan SMOTE murni, peneliti selanjutnya dapat membandingkan efektivitasnya dengan teknik *resampling* hibrida seperti SMOTE-ENN atau SMOTE-Tomek Links. Pendekatan hibrida ini berpotensi untuk tidak hanya menambah sampel pada kelas minoritas secara sintesis, tetapi juga membersihkan *noise* atau data ambigu di perbatasan antar kelas, yang secara teoretis diharapkan mampu mendongkrak metrik *Precision* pada model agar peringatan palsu dapat ditekan ke tingkat yang lebih minimal.

Selain perbaikan di sisi penanganan data, penelitian di masa depan juga sangat disarankan untuk menggunakan cakupan dataset yang lebih luas, modern, dan beragam. Pima Indians Diabetes Dataset yang digunakan dalam studi ini

memiliki keterbatasan dalam hal jumlah populasi sampel (768 data) dan hanya merepresentasikan spesifikasi demografis yang sangat spesifik (perempuan keturunan Pima Indian). Untuk meningkatkan kemampuan generalisasi (*robustness*) dari model *machine learning* yang dibangun, pengujian sebaiknya dilakukan pada dataset rekam medis klinis multisentris berskala besar yang mencakup demografi pasien yang jauh lebih heterogen. Di samping itu, optimalisasi model dapat lebih disempurnakan dengan beralih dari penggunaan GridSearchCV menuju algoritma optimisasi *hyperparameter* yang lebih efisien dan canggih, seperti *Bayesian Optimization* atau kerangka kerja *Optuna*, untuk menjelajahi ruang parameter yang lebih presisi dengan waktu komputasi yang lebih efisien.

Untuk mendorong tingkat adopsi dan kepercayaan terhadap model kecerdasan buatan di dunia medis, implementasi *Explainable AI* (XAI) menjadi sebuah keharusan bagi penelitian selanjutnya. Dalam praktiknya, tenaga medis tidak hanya membutuhkan prediksi akhir, tetapi juga membutuhkan interpretasi mengenai alasan di balik keputusan yang diambil oleh model (*black-box*). Mengintegrasikan metode penjelasan cerdas seperti SHAP (*SHapley Additive exPlanations*) atau LIME akan sangat membantu dokter dalam mengetahui secara pasti fitur klinis mana, misalnya apakah tingkat glukosa yang tinggi atau metrik BMI, yang memberikan kontribusi paling dominan terhadap vonis risiko diabetes yang dikeluarkan oleh model XGBoost maupun CatBoost.