

BAB V

KESIMPULAN DAN SARAN

Setelah melalui rangkaian tahapan mulai dari perumusan masalah, pengumpulan data, analisis, hingga pembahasan yang mendalam, maka pada bagian ini akan disajikan kesimpulan sebagai rangkuman dari hasil yang telah dicapai selama proses penelitian/proyek berlangsung. Kesimpulan ini dirumuskan berdasarkan temuan-temuan penting yang diperoleh dan menjadi jawaban atas tujuan yang telah ditetapkan sebelumnya. Selain itu, bagian ini juga memuat saran yang diharapkan dapat memberikan masukan konstruktif bagi pihak-pihak terkait, serta menjadi acuan untuk pengembangan atau penelitian selanjutnya agar hasil yang diperoleh dapat lebih optimal dan bermanfaat.

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi teknik *SMOTETomek* untuk penyeimbangan data dalam klasifikasi penyakit diabetes menggunakan algoritma *XGBoost*, maka pada subbab ini disampaikan kesimpulan yang merupakan jawaban dari rumusan masalah yang telah diajukan. Kesimpulan ini disusun berdasarkan analisis terhadap hasil eksperimen dan evaluasi model yang dilakukan selama proses penelitian.

1. Teknik *SMOTETomek* diimplementasikan dengan menggabungkan metode oversampling (*SMOTE*) dan undersampling (*Tomek Links*) secara bersamaan untuk mengatasi ketidakseimbangan data pada dataset klasifikasi penyakit diabetes. Setelah proses *balancing*, data tersebut dilatih menggunakan algoritma *XGBoost* dengan parameter *n_estimators* sebanyak 500 dan pembagian data *train-test split* sebesar 80:20. Proses ini menghasilkan model klasifikasi yang lebih stabil dan efektif, dengan distribusi kelas yang lebih seimbang sebelum dilakukan pelatihan.
2. Hasil penelitian menunjukkan bahwa penggunaan teknik *SMOTETomek* menghasilkan performa model yang lebih baik dibandingkan dengan penggunaan data asli maupun teknik *SMOTE* saja. Hal ini ditunjukkan oleh nilai akurasi tertinggi sebesar 95,25% dan keseimbangan antara akurasi pelatihan dan pengujian, yang menunjukkan bahwa model tidak mengalami *overfitting*.

secara signifikan. Sebaliknya, model yang dibangun dengan data asli dan *SMOTE* menunjukkan perbedaan akurasi yang besar, mengindikasikan potensi *overfitting*. Oleh karena itu, *SMOTETomek* terbukti menjadi teknik balancing yang lebih efektif dalam konteks ini.

5.2 Saran

Berdasarkan kesimpulan yang telah diperoleh, maka pada subbab ini disampaikan beberapa saran yang diharapkan dapat menjadi masukan konstruktif bagi penelitian selanjutnya. Saran ini bertujuan untuk meningkatkan kualitas penelitian di masa depan, baik dari segi penggunaan data, metode balancing, maupun pengembangan model klasifikasi yang lebih akurat dan dapat diandalkan dalam konteks aplikasi medis.

1. Penelitian ini menggunakan dataset dari *Kaggle* yang mungkin mengandung bias tertentu. Oleh karena itu, disarankan agar penelitian selanjutnya menggunakan dataset yang diperoleh langsung dari lembaga kesehatan resmi agar representasi data lebih akurat dan realistis.
2. Disarankan untuk melakukan eksplorasi lebih lanjut terhadap parameter lain dalam algoritma *XGBoost*, serta membandingkan hasilnya dengan algoritma klasifikasi lain seperti *Random Forest*, SVM, atau *Neural Networks* guna memperoleh performa model yang lebih optimal.
3. Untuk mendapatkan hasil evaluasi model yang lebih andal, penggunaan teknik validasi seperti *K-Fold Cross Validation* sangat direkomendasikan dalam penelitian lanjutan.
4. Mengingat aplikasi dari model ini berhubungan dengan kesehatan, maka penting untuk mempertimbangkan interpretabilitas model dan kolaborasi dengan ahli medis agar hasil klasifikasi dapat dipahami dan digunakan secara tepat dalam pengambilan keputusan klinis.