

ABSTRAK

Penanganan ketidakseimbangan kelas merupakan tantangan utama dalam klasifikasi data medis, seperti pada dataset Diabetes dan Heart Disease. Penelitian ini membandingkan efektivitas metode oversampling SMOTE, SMOTE-Tomek, dan SMOTE-ENN yang dikombinasikan dengan dua classifier klasik, yaitu K-Nearest Neighbors ($K=3$) dan Support Vector Machine (SVM). Evaluasi dilakukan menggunakan stratified 5-fold cross-validation dengan metrik utama accuracy, precision, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa metode hybrid SMOTE-Tomek dan SMOTE-ENN secara signifikan meningkatkan performa klasifikasi, khususnya recall dan F1-score pada kelas minoritas, dengan SVM sebagai classifier terbaik untuk kedua dataset. Temuan ini memberikan rekomendasi praktis dalam pemilihan metode oversampling dan classifier guna meningkatkan deteksi pasien berisiko, serta meminimalkan kesalahan diagnosis pada data medis yang tidak seimbang.

Kata kunci — Data tidak seimbang, *k-Nearest Neighbors* (K-NN), *Support Vector Machine* (SVM), SMOTE, SMOTE-Tomek, SMOTE-ENN, Oversampling.