

Abstrak

Mengatasi ketidakseimbangan data dan pencilan merupakan tantangan penting dalam klasifikasi data. Tujuan dari penelitian ini adalah untuk meningkatkan kinerja klasifikasi dengan mengurangi efek ketidakseimbangan kelas dan keberadaan pencilan dalam dataset. SMOTE-LOF menggabungkan metode oversampling SMOTE dengan Local Outlier Factor (LOF) untuk membuat sampel sintesis yang juga memperhitungkan outlier potensial. Sementara itu, Borderline-SMOTE mengidentifikasi sampel "batas" di kelas minoritas dan kemudian membuat sampel sintesis di sepanjang perbatasan antara kelas mayoritas dan minoritas. Pada penelitian ini, eksperimen dilakukan dengan menggunakan algoritma klasifikasi seperti Naïve Bayes, dan Support Vector Machine pada dataset yang tidak seimbang dan mengandung outlier. Dataset yang digunakan dalam penelitian ini antara lain Pima Indians, Haberman, Glass, dan Rainfall. Skenario penelitian ini meliputi perbandingan dengan penelitian sebelumnya yang telah dilakukan mengenai SMOTE-LOF dan Borderline-SMOTE pada dataset Curah Hujan. Hasil penelitian menunjukkan bahwa pada ketiga dataset tersebut, Borderline-SMOTE mengungguli SMOTE-LOF pada ketiga pengklasifikasi dengan rata-rata akurasi 4-6%, presisi 2-4%, recall 5-10%, dan F1 score 5-6%. Ketika teknik ini diterapkan pada dataset Curah Hujan, hasilnya menunjukkan peningkatan akurasi sebesar 10-25%. Hasilnya secara konsisten menunjukkan bahwa, ketika diterapkan pada dataset Pima Indians, Haberman, dan Glass, Borderline-SMOTE meningkatkan kinerja beberapa algoritme klasifikasi. Akurasi, presisi, recall, dan skor F1 yang lebih baik merupakan bukti dari hal ini jika dibandingkan dengan penerapan teknik SMOTE-LOF.

Kata kunci : *Imbalanced data, smote-lof, borderline-smote, naïve bayes, support vector machine*