

BAB V PENUTUP

5.1 Kesimpulan

Penelitian ini dilakukan untuk mengetahui kinerja dari model machine learning dalam mengklasifikasi setelah dilakukan teknik resampling kepada beberapa tipe dataset, dengan tujuan untuk mengetahui. Terdapat 12 teknik resampling yang digunakan, yaitu 6 teknik oversampling "SMOTE", "ADASYN", "Borderline SMOTE", "SVM SMOTE", "SMOTENC", "Random Over Sampler (ROS)", dan 6 teknik undersampling "Random Under Sampler (RUS)", "Edited Nearest Neighbours (ENN)", "Cluster Centroids", "Near Miss", "Neighbourhood Cleaning Rule (NCL)", "One-Sided Selection". Algoritma klasifikasi yang digunakan yaitu decision tree karena decision tree adalah algoritma populer yang digunakan untuk memecahkan permasalahan klasifikasi.

Berdasarkan eksperimen yang dilakukan :

1. Performa tertinggi yang diraih sebesar 100%, yaitu pada dataset 4 dan dataset 8, dari 13 percobaan dataset 4 dan 8 mendapat kan performa 100% pada 11 percobaan mendapat kinerja 100% di pada Akurasi, Presisi, Recall, dan F1-score, dan dimungkinkan klasifikasi yang dilakukan pada dataset 4 dan 8 mengalami overfitting.
2. Peningkatan kinerja paling terlihat pada dataset 6 yang memiliki 4 kelas, peningkatan kinerja sangat signifikan, kinerja sebelum resampling didapat Akurasi 99.5%, Presisi 67.2%, Recall 66.6% dan F1-score 66.5% dan mengalami peningkatan pada Presisi, Recall dan F1-score, dengan rata rata peningkatan sebesar 32.3% setelah di lakukan teknik oversampling, dan rata rata peningkatan sebesar 12.1% setelah dilakukan teknik undersampling, kecuali pada teknik Random Under Sampling yang mengalami penurunan kinerja.
3. Tidak semua teknik resampling bisa diterapkan ke semua dataset, dalam penelitian ini ditemukan bahwa teknik SMOTENC tidak bisa di terapkan pada dataset 3, dataset 4, dataset 5 dan dataset 8, dikarenakan dataset

tersebut hanya memiliki 1 jenis atribut yaitu atribut kontinu (numeric) atau kategorikal, menjadikan SMOTENC hanya bisa diterapkan dengan dataset yang memiliki 2 jenis atribut kategorikal dan kontinu secara bersamaan.

5.2 Saran

Berdasarkan hasil penelitian ini, ada beberapa saran untuk peneliti yang ingin meneliti lebih lanjut dengan topik yang sama:

1. Menerapkan *Feature Selection* pada dataset untuk mengurangi jumlah fitur yang memiliki korelasi rendah.
2. Mengganti algoritma *Decision Tree* dengan algoritma lain yang memiliki kemungkinan *overfitting* rendah seperti *Random Forest*.

