

BAB V

PENUTUP

5.1. Kesimpulan

Adapun kesimpulan pada penelitian implementasi algoritma SMOTE dan klasifikasi *Decision Tree* untuk mendeteksi kecurangan transaksi online adalah sebagai berikut:

1. Dataset yang digunakan, yaitu "*Online Payments Fraud Detection Dataset*" yang diambil dari Kaggle. Dataset memiliki kelas transaksi curang, dimana nilai 0 ditandai sebagai kelas tidak curang yang terdiri sebanyak 6.354.407 dan nilai 1 sebagai kelas curang sebanyak 8.213. Dari data yang digunakan terdapat 5 jenis pembayaran transaksi diantaranya *cash_out*, *payment*, *cash_in*, *transfer*, dan *debit*. Dari kelima jenis pembayaran transaksi curang hanya dilakukan pada *cash_out* dan *transfer*. Sehingga untuk langkah awal menanggulangi ketidakseimbangan data dilakukan penghapusan pada transaksi yang bertipe pembayaran *payment*, *cash_in*, dan *debit*. Maka kelas 0 berkurang menjadi 2.762.196 dan kelas 1 sebanyak 8.213. Kemudian dengan adanya ketidakseimbangan kedua kelas dilakukan penerapan SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan kelas. Metode penerapan SMOTE dilakukan dengan kerja sama *k Nearest Neighbor*, hal ini dilakukan untuk menghasilkan data sintesis yang menjaga interpolasi dari pola dataset asli. SMOTE dilakukan dengan menentukan nilai alpha 0,9 untuk menjaga konsistensi data yang dihasilkan. Kemudian memilih salah satu sampel minoritas dan tetangga terdekat secara acak dari hasil *k Nearest Neighbor*. Terakhir, menghitung selisih sampel minoritas dan tetangga terdekat. Data sintesis yang dihasilkan dalam penelitian dilakukan dengan penambahan kelas minoritas menjadi setengah dan seimbang dari kelas mayoritas.
2. Dalam penerapan proses klasifikasi *decision tree* membangun pohon keputusan dengan menghitung gini index dan gini split pada setiap kemungkinan pemisahan dalam semua fitur. Nilai gini split yang terendah akan menjadi *root node*. Kemudian membagi *root node* menjadi 2 cabang

sampai kondisi *leaf node* terpenuhi. Adapun parameter yang digunakan dalam membangun pohon keputusan dengan membatasi nilai kedalaman pohon = 3, minimum jumlah sampel yang digunakan untuk pemisahan 2000 sampel dan sampel *leaf node* minimum memiliki 1000 sampel. Kemudian dari nilai gini menunjukkan banyak nilai gini yang rendah di banyak node menunjukkan bahwa pemisahan cukup baik dalam mengisolasi kelas-kelas tertentu, terutama di node yang mengarah ke kelas 0. Sedangkan gini yang tinggi seperti 0.5 pada beberapa node menunjukkan ketidakpastian atau pencampuran antara kelas 0 dan kelas 1, yang bisa menjadi indikasi bahwa split tersebut tidak seefektif yang lain dalam memisahkan kelas.

3. Dari hasil beberapa pengujian penerapan klasifikasi *decision tree* dengan berbagai split data pada dataset yang tidak seimbang atau tanpa SMOTE menunjukkan model mampu melakukan generalisasi yang baik dalam kedua kelas. Hal ini terlihat dari tingginya *precision*, *recall*, dan *F1-score* pada kedua kelas, terutama pada kelas minoritas (kelas 1), menunjukkan bahwa model ini efektif meskipun menghadapi ketidakseimbangan data yang signifikan. Namun pada pengujian menggunakan split data 60% data latih dan 40% data uji nilai *precision*, *recall*, dan *F1-score* menurun, hal ini menunjukkan bahwa proses klasifikasi cenderung hanya melatih kelas mayoritas sehingga salah dalam memprediksi kelas minoritas. Sedangkan hasil pengujian proses klasifikasi Decision Tree yang dilakukan dengan penambahan data sintesis dari hasil penggunaan metode SMOTE memberikan semua nilai metrik evaluasi yang stabil tinggi dengan nilai 99% dan dari hasil klasifikasi report menunjukkan bahwa proses klasifikasi dalam memprediksi kelas 0 dan 1 sempurna. Sedangkan dari hasil nilai gini skenario pengujian tanpa SMOTE lebih murni pada kelas y0, sedangkan skenario pengujian dengan SMOTE nilai gini menjadi tinggi yang menunjukkan adanya campur aduk kelas y0 dan y1. Selain itu didapatkan hasil *confusion matrix* menunjukkan skenario pengujian yang dilakukan dengan menggunakan metode SMOTE mengalami peningkatan pada jumlah *false negatives* yang sangat signifikan. Hal ini terjadi karena

pembentukan data sintesis SMOTE kurang mempertahankan pola dataset asli sehingga semakin memperbanyak jumlah data sintesis yang dihasilkan maka akan semakin bertambah jumlah *false negatives*.

5.2. Saran

Adapun saran yang dapat penulis berikan dari hasil penelitian ini, antara lain :

- a. Mempertimbangkan metode oversampling lainnya seperti ADASYN atau teknik kombinasi dari teknik oversampling dan undersampling yang bisa lebih sesuai untuk dataset tertentu.
- b. Menemukan kombinasi terbaik atau penambahan data sintesis yang lebih representatif dengan data asli sehingga dapat mengurangi *false negatives* secara efektif.
- c. Pengembangan dengan menggunakan metode klasifikasi yang berbeda dan terbaru.
- d. Dapat dilanjutkan pada model pengembangan dalam bentuk prototype program.