

## **ABSTRAK**

### **ANALISIS PENGARUH PENERAPAN SMOTE PADA AKURASI RANDOM FOREST DAN C4.5 DALAM KLASIFIKASI DATA TIDAK SEIMBANG**

Oleh

Fanisa Nimastiti 20110004

Penanganan distribusi data yang tidak seimbang dalam sebuah dataset dapat menyulitkan proses klasifikasi, sehingga menjadi tantangan serius dalam analisis data. Ketidakseimbangan antara kelas mayoritas dan minoritas dapat menyebabkan model pembelajaran mesin tidak dapat belajar dengan baik dari kelas minoritas. Pendekatan pada tingkat data, seperti oversampling dan undersampling, dapat mengatasi masalah ketidakseimbangan dataset, tetapi masing-masing pendekatan memiliki kelebihan dan kekurangan. Penelitian ini berfokus pada pendekatan oversampling, terutama menerapkan Synthetic Minority Over-sampling Technique (SMOTE). Tujuannya adalah menganalisis dampak penerapan SMOTE dalam mengatasi masalah ketidakseimbangan dataset, baik dari segi akurasi model maupun distribusi data. Selain itu, kinerja algoritma Random Forest dan C4.5 setelah penerapan SMOTE akan dibandingkan. Hasil penelitian menunjukkan bahwa penggunaan teknik oversampling SMOTE terbukti efektif dalam mengatasi masalah ketidakseimbangan dataset. SMOTE menciptakan data baru yang mirip dengan kelas minoritas, sehingga distribusi data menjadi lebih seimbang. Hal ini menghasilkan peningkatan performa model mencapai 24% hingga 28,7% pada data 1, peningkatan 17,1% hingga 19,1% pada data 2, peningkatan 3,5% hingga 5,1% pada data 3, dan terakhir pada data 4 peningkatannya hanya sebesar sekitar 1% saja saat diterapkan SMOTE. Selain itu juga ditemukan bahwa performa akurasi Random Forest lebih unggul dibandingkan dengan algoritma C4.5, baik dengan menggunakan SMOTE maupun tanpa SMOTE. Hal ini disebabkan karena Random Forest lebih tahan terhadap overfitting, sehingga dapat menghasilkan model yang lebih akurat pada dataset yang tidak seimbang.

**Kata Kunci:** Ketidakseimbangan data, *oversampling*, *Synthetic Minority Over-sampling Technique (SMOTE)*, *Random Forest*, *C4.5*.