

ABSTRAK

Penelitian ini berfokus pada penerapan algoritma *Naïve Bayes* untuk memprediksi churn pelanggan di industri telekomunikasi menggunakan dataset *Telecom Customer Churn*. Penelitian ini merujuk pada proses ketika pelanggan meninggalkan atau berhenti menggunakan produk atau layanan dari suatu perusahaan. Proses penelitian dimulai dengan analisis data awal, preprocessing untuk membersihkan dan normalisasi data, serta transformasi data menggunakan *Principal Component Analysis* (PCA). Seleksi fitur dilakukan dengan metode *Recursive Feature Elimination*, dengan variasi nilai 'k' yang dipertimbangkan. Data diimbangi menggunakan *Systematic Missingness Over-sampling* (SMOTE), dan variasi pembagian training-testing diterapkan, termasuk *Stratified* dan *Random*.

Hasil eksperimen menunjukkan bahwa metode *Stratified* tanpa *SMOTE*, menggunakan 10 fitur, memberikan akurasi tertinggi sebesar 82,82%. Pada *Random* tanpa *SMOTE*, hasil yang baik diperoleh, sedangkan penggunaan *SMOTE* cenderung menurunkan akurasi. Peningkatan jumlah fitur dari 5 menjadi 10 menghasilkan kompleksitas model yang meningkat. Kesimpulannya, metode *Stratified* tanpa *SMOTE* pada 10 fitur memberikan akurasi optimal. Saran untuk penelitian berikutnya mencakup eksperimen dengan variasi parameter model, seperti *Hyperparameter* dan evaluasi pada dataset yang lebih besar dan beragam.

Kata Kunci : *SMOTE*, Klasifikasi, *Naïve Bayes*, *Stratified*, *Random*

