

ABSTRAK

RENY RAHAYU.S. Analisis *Unbalance* dengan Teknik *Undersampling* dan *Oversampling* untuk Akurasi *Naïve Bayes* (dibimbing Titin Wahyuni S.Pd.,M.T dan Fahrin Irhamna Rachman S.Kom.,M.T.).

Ketidakseimbangan dataset adalah kondisi dimana distribusi kelas dalam dataset tidak seimbang, dengan kelas mayoritas memiliki jumlah data yang jauh lebih banyak dibandingkan dengan kelas minoritas, yang dapat menyebabkan kesalahan dalam hasil klasifikasi dimana data kelas minoritas sering diklasifikasikan sebagai kelas mayoritas. Tujuan dari penelitian ini adalah untuk memahami dampak ketidakseimbangan data terhadap akurasi *model Naïve Bayes*. *Model Naïve Bayes* yang dimaksud dalam penelitian ini adalah salah satu algoritma klasifikasi berbasis probabilitas yang sederhana namun efektif, yang didasarkan pada Teorema Bayes. Algoritma ini mengasumsikan bahwa fitur-fitur dalam dataset bersifat independen satu sama lain, sebuah asumsi yang jarang benar dalam kenyataan tetapi sering kali menghasilkan hasil yang cukup baik dalam praktik. Hasil penelitian menunjukkan bahwa keseimbangan pada dataset sangat mempengaruhi akurasi model *Naïve Bayes*. Teknik *undersampling* menghasilkan akurasi tertinggi sebesar 96,6%, sementara teknik *oversampling* mencapai akurasi maksimal 92%, dan penggunaan *SMOTE* (*Synthetic Minority Over-sampling Technique*) memberikan hasil yang baik dengan akurasi tertinggi 91,8%. Kesimpulan dari penelitian ini adalah bahwa teknik *undersampling* cenderung mengurangi akurasi dengan menghilangkan informasi penting dari kelas mayoritas, sedangkan teknik *oversampling* berisiko menyebabkan *overfitting*. Oleh karena itu, pemilihan metode penyeimbangan harus dilakukan dengan hati-hati, menyesuaikan sifat data dan tujuan analisis, serta dievaluasi dengan cermat untuk mencapai performa model yang optimal.

Kata Kunci : Analisis *Unbalance*, *Naïve Bayes*, *Oversampling*, *SMOTE*, *Undersampling*.