

ABSTRAK

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

ABSTRACT

RENY RAHAYU.S. *Unbalance Analysis with Undersampling and Oversampling Techniques for Naïve Bayes Accuracy (supervised by Titin Wahyuni S.Pd., M.T. and Fahrin Irhamna Rachman S.Kom., M.T.).*

Dataset imbalance is a condition where the distribution of classes in a dataset is unbalanced, with the majority class having a much larger amount of data compared to the minority class, which can cause errors in classification results where minority class data is often classified as the majority class. The aim of this research is to understand the impact of data imbalance on the accuracy of the Naïve Bayes model. The Naïve Bayes model referred to in this research is a simple but effective probability-based classification algorithm, which is based on Bayes' Theorem. This algorithm assumes that the features in the dataset are independent of each other, an assumption that is rarely true in reality but often produces quite good results in practice. The research results show that balance in the dataset greatly influences the accuracy of the Naïve Bayes model. The undersampling technique produces the highest accuracy of 96.6%, while the oversampling technique achieves a maximum accuracy of 92%, and the use of SMOTE (Synthetic Minority Over-sampling Technique) provides good results with the highest accuracy of 91.8%. The conclusion of this research is that undersampling techniques tend to reduce accuracy by removing important information from the majority class, while oversampling techniques risk causing overfitting. Therefore, the selection of balancing methods must be done carefully, adapting to the nature of the data and analysis objectives, and carefully evaluated to achieve optimal model performance.

Keywords : *Unbalance Analysis, Naïve Bayes, Oversampling, SMOTE, Undersampling.*