

ABSTRAK

Wahyu Sukma Erlangga, Anak Agung Gde (2024), *Kombinasi oversampling dan undersampling dalam menangani class unbalanced dan overlapping pada klasifikasi data bank marketing.* Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Kata-kata kunci: bank marketing, class imbalance, class overlapping, oversampling, undersampling,

Ketidakseimbang kelas dapat terjadi pada berbagai jenis *datasets*. Salah satunya adalah *datasets bank marketing* yang diambil dari web UCI Machine Learning. Masalah itu dapat menyebabkan model *machine learning* yang dihasilkan lebih baik dalam memprediksi kelas mayoritas dibanding kelas minoritas. Oleh karena itu, pada penelitian ini akan dilakukan penanganan pada masalah tersebut dengan menggunakan metode *oversampling* yaitu SMOTE akan tetapi, penerapan SMOTE juga dapat menimbulkan masalah lainnya yaitu potensi *class overlapping* atau tumpang tindih kelas yang juga dapat mengganggu performa model. Penelitian ini memiliki tujuan untuk menanganani kedua masalah tersebut dengan mengkombinasikan metode SMOTE dengan metode *undersampling*, dimana metode *undersampling* yang akan digunakan terdiri dari ENN, NCL, dan TomekLink. Untuk algoritma *machine learning* yang digunakan adalah *Logistic Regression* dan evaluasi performa dilakukan dengan *confusion matrix*, sehingga dapat dihitung *sensitivity*, *specificity* dan *g-means* dari model. Hasil penelitian ini menunjukkan bahwa kombinasi SMOTE-ENN menghasilkan hasil paling optimal dengan *sensitivity*, *specificity* dan *g-means* berturut-turut 94,05%, 83,22% dan 88,47% pada *datasets bank marketing*.



ABSTRACT

Wahyu Sukma Erlangga, Anak Agung Gde (2024), *Combination of oversampling and undersampling in handling class unbalanced and overlapping in bank marketing data classification. Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.*

Keywords: bank marketing, class imbalance, class overlapping, oversampling, undersampling.

Class imbalance can occur in various types of datasets. One of them is bank marketing datasets taken from UCI Machine Learning web. The problem can cause the machine learning model to predict the majority class better than the minority class. Therefore, this research will handle the problem by using an oversampling method, namely SMOTE. However, the application of SMOTE can also cause other problems, namely class overlapping which can also interfere with model performance. This research aims to handle both problems by combining the SMOTE method with undersampling methods, where the undersampling methods to be used consist of ENN, NCL, and TomekLink. The machine learning algorithm used is Logistic Regression and the performance evaluation is done with confusion matrix, so that the sensitivity, specificity, and g-means of the model can be obtained. The results of this study show that the SMOTE-ENN combination produces the most optimal results with sensitivity, specificity, and g-means of 94.05%, 83.22% and 88.47% respectively on bank marketing datasets.

