

## INTISARI

### IMPLEMENTASI *RANDOM OVER-SAMPLING EXAMPLES* (ROSE) UNTUK MENANGANI DATA TIDAK SEIMBANG PADA ANALISIS KLASIFIKASI BINER

oleh

Sindhi Mery Handayani

19/442609/PA/19358

Kegagalan metode klasifikasi ketika estimasi model didasarkan pada data yang tidak seimbang merupakan permasalahan yang sering ditemui. Banyak jurnal yang menunjukkan bahwa data yang tidak seimbang sangat mengganggu proses analisis, karena model akan cenderung berfokus pada kelas mayoritas dan mengabaikan kelas minoritas. Mempertimbangkan pentingnya masalah ini, banyak metode telah dikembangkan untuk mencoba mengatasi masalah ini. *Random Over-Sampling Examples* (ROSE) merupakan salah satu metode *resampling* yang dapat digunakan untuk mengatasi masalah ini. ROSE akan membangkitkan data sintesis baru yang seimbang dengan menggunakan pendekatan *smoothed bootstrap*. Pada skripsi ini, data *Abalone* diklasifikasikan dengan metode klasifikasi *Naïve Bayes* dan *Random Forest*. Berdasarkan analisis yang telah dilakukan, disimpulkan bahwa penanganan data tidak seimbang dengan metode ROSE dapat meningkatkan performa klasifikasi, baik pada klasifikasi dengan metode *Naïve Bayes* maupun *Random Forest*.

Kata kunci: Klasifikasi, Data Tidak Seimbang, *Random Over-Sampling Examples*, *Naïve Bayes*, *Random Forest*.

## **ABSTRACT**

### **IMPLEMENTATION OF RANDOM OVER-SAMPLING EXAMPLES (ROSE) FOR HANDLING IMBALANCED DATA IN BINARY CLASSIFICATION ANALYSIS**

by

Sindhi Mery Handayani

19/442609/PA/19358

Failure of classification methods when the model estimation is based on an imbalanced dataset is a very well-known problem. It has been widely reported that the imbalanced dataset heavily compromises the analysis process because the model tends to focus on the majority class and ignore the minority class. Considering the importance of this issue, a large amount of method has been developed trying to address this problem. Random Over-Sampling Examples (ROSE) is one of the resampling methods that can be used to solve this problem. ROSE builds on the generation of new synthetic balanced data according to a smoothed bootstrap approach. In this thesis, the Abalone dataset is classified using the Naïve Bayes and Random Forest classifier. From the analysis conducted, it was concluded that handling imbalanced data with the ROSE method can improve the classification performance, both in classification using the Naïve Bayes method and the Random Forest method.

**Keywords:** Classification, Imbalanced Data, Random Over-Sampling Examples, Naïve Bayes, Random Forest.