

INTISARI

PENERAPAN *SAFE-LEVEL SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (*SAFE-LEVEL SMOTE*) PADA ANALISIS KLASIFIKASI DENGAN DATA TIDAK SEIMBANG

Oleh

Oktavia Marwatin Hudanti

19/439214/PA/19037

Analisis klasifikasi merupakan salah satu metode yang cukup populer dalam *machine learning*. Namun pada penerapannya, masih sering ditemukan analisis klasifikasi dengan data tidak seimbang. Data tidak seimbang merupakan kondisi di mana distribusi kelas data tidak seimbang, yakni jumlah pengamatan pada salah satu kelas lebih banyak dibandingkan dengan kelas lainnya. Data tidak seimbang dapat menyebabkan kesalahan klasifikasi, dimana kelas minoritas salah diklasifikasikan menjadi kelas mayoritas. Sehingga proses klasifikasi dilakukan hanya berdasarkan kelas mayoritas saja tanpa memperhatikan kelas minoritas. Oleh karena itu diperlukan penanganan data tidak seimbang, diantaranya dengan melakukan *undersampling* atau *oversampling*. *Safe-level SMOTE* merupakan salah satu teknik *oversampling* hasil pengembangan dari SMOTE. *Safe-level SMOTE* membangkitkan data sintesis berdasarkan kriteria yang disebut dengan *safe level ratio*. Pada studi kasus ini dilakukan penerapan *Safe-level SMOTE* pada dua jenis data yakni data simulasi dengan proporsi kelas minoritas berbeda-beda dan data *Indian Liver Patient Dataset*.

Kata kunci : klasifikasi, data tidak seimbang, SMOTE, *Safe-level SMOTE*

ABSTRACT

IMPLEMENTATION OF SAFE-LEVEL SYNTETIC MINORITY OVERSAMPLING TECHNIQUE (SAFE-LEVEL SMOTE) IN IMBALANCED CLASSIFICATION ANALYSIS

By

Oktavia Marwatin Hudanti

19/439214/PA/19037

Classification analysis is one of popular method in machine learning. However, in its applications classification analysis is often found with imbalanced data. Imbalanced data is a condition which the distribution of classes data is unbalance, the number of observations in one class is more than the other class. Imbalanced data can cause misclassification, where minority class is incorrectly classified as majority class. So the classification process is only based on majority class without notice minority class. There are several methods for handling this imbalanced data problem, including undersampling and oversampling. Safe-level SMOTE is one of the oversampling techniques developed by SMOTE. Safe-level SMOTE generate the synthetic data based on criterion called safe-level ratio. In this case study, Safe-level SMOTE applied with two types of data, simulation data with different minority class proportion and Indian Liver Patient Dataset.

Keywords : classification, imbalanced data, SMOTE, Safe-level SMOTE