

INTISARI

IMPLEMENTASI *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (SMOTE) UNTUK KLASIFIKASI DATA TIDAK SEIMBANG

oleh

I Dewa Gede Natih Bernan

16/398656/PA/17617

Analisis klasifikasi merupakan salah satu metode *supervised learning* yang dapat digunakan untuk memprediksi label dari suatu data berdasarkan data latih yang diberikan. Dalam klasifikasi seringkali dijumpai data yang tidak seimbang, yaitu salah satu label memiliki jumlah yang jauh lebih banyak dari label yang lain. Hal ini dapat menimbulkan masalah yaitu kurangnya kemampuan model untuk memprediksi data yang berasal dari label yang sedikit tersebut (kelas minoritas), sehingga walaupun akurasi yang dihasilkan tinggi namun sebagian besar prediksi tepat hanya berasal dari kelas mayoritas saja. Metode SMOTE dapat diterapkan untuk mengatasi permasalahan tersebut, di mana dengan metode ini dibuat *instance* sintetis yang berasal dari kelas minoritas sehingga menghasilkan data yang lebih seimbang. Pada skripsi ini dilakukan analisis pada data *Bank Marketing* dengan menggunakan model klasifikasi *Naïve Bayes* dan *Random Forest* yang dikombinasikan dengan metode SMOTE serta menggunakan rasio yang berbeda-beda. Dari analisis yang dilakukan diperoleh kesimpulan bahwa metode *Naïve Bayes* pada data SMOTE1.0 menghasilkan performa yang paling baik, terutama dilihat dari akurasi kelas positif.

Kata kunci: klasifikasi, data tidak seimbang, SMOTE, *Naïve Bayes*, *Random Forest*

ABSTRACT

IMPLEMENTATION OF SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) FOR CLASSIFICATION OF IMBALANCED DATA

by

I Dewa Gede Natih Bernan

16/398656/PA/17617

Classification analysis is one of the supervised learning methods that can be used to predict the label of a data based on the training data provided. In classification, it is often found imbalanced data, that is, one label has far more amount than the other label. This can cause the lack of the ability of the model to predict data from the fewer label (minority class), so that even though the accuracy produced is high, most of the correct predictions only come from the majority class. The SMOTE method can be applied to overcome these problems, where from this method synthetic instances are made that come from minority class to produce more balanced data. This undergraduate thesis analyzes the Bank Marketing data using the Naïve Bayes and Random Forest classification models combined with the SMOTE method and using different ratios. From the analysis conducted it was concluded that the Random Forest method on SMOTE1.0 data produced the best performance, especially viewed from the accuracy of the positive class.

Keyword: classification, imbalanced data, SMOTE, Naïve Bayes, Random Forest