

INTISARI

PENANGANAN DATA TIDAK SEIMBANG MENGGUNAKAN ADAPTIVE NEIGHBOR SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (ANS) UNTUK ANALISIS KLASIFIKASI

oleh

Nafisha Nur Fatima

16/394200/PA/17291

Perkembangan teknologi memicu berkembangnya ketersediaan data dalam jumlah yang besar dan bentuk yang kompleks. Hal tersebut merupakan suatu alasan penting untuk memajukan pengetahuan dan pemahaman mengenai analisis data. Pada kenyataannya sering ditemui data yang tidak seimbang, yaitu data dengan rasio yang tidak seimbang antara satu kelas dengan kelas lainnya. Dalam analisis klasifikasi, data tidak seimbang menimbulkan masalah karena model yang dihasilkan kurang mampu memprediksi data yang berasal dari kelas yang sedikit (kelas minoritas). Metode ANS, yang merupakan pengembangan dari metode SMOTE, dapat diterapkan untuk menangani permasalahan tersebut. Dengan metode ANS dapat dibuat tupel sintesis yang berasal dari kelas minoritas sehingga dapat menghasilkan data yang lebih seimbang. Pada skripsi ini dilakukan analisis pada data *Pima Indians Diabetes* dengan menggunakan model klasifikasi *Random Forest* dan *Naïve Bayes* yang dikombinasikan dengan metode ANS. Dari analisis yang dilakukan diperoleh kesimpulan bahwa model *Random Forest* pada data dengan ANS menghasilkan performa yang paling baik, terutama dalam akurasi secara keseluruhan dan akurasi kelas positif.

Kata kunci: data tidak seimbang, *oversampling*, ANS, SMOTE, klasifikasi.

ABSTRACT

IMBALANCED DATA HANDLING WITH ADAPTIVE NEIGHBOR SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (ANS) FOR CLASSIFICATION ANALYSIS

by

Nafisha Nur Fatima

16/394200/PA/17291

The development in technology initiates the growth of data availability in big quantity and complex form. It is an important reason to improve knowledge and understanding about data analysis. In fact, we usually found imbalanced data, which is data with imbalanced ratio between one class with other classes. In classification analysis, imbalanced data causes problem because the output model does not really predict data from fewer class (minority class) accurately. ANS method, the development from SMOTE method, can be used to solve that problem. With ANS method we can make synthetic instance from minority classes so we can get more balanced data. In this thesis, analysis in Pima Indians Diabetes dataset was done using Random Forest and Naïve Bayes classification method combined with ANS method. From the analysis, we can conclude that Random Forest method in data with ANS produced the best performance, especially in accuracy and sensitivity.

Keyword: imbalanced data, oversampling, ANS, SMOTE, classification.