

## INTISARI

### **PENERAPAN KOMBINASI SMOTE DAN TOMEK LINKS UNTUK KLASIFIKASI DATA TIDAK SEIMBANG DENGAN METODE *RANDOM FOREST***

oleh

Rona Agustika

17/412748/PA/18067

Perkembangan pesat di bidang teknologi berbanding lurus dengan ketersediaan data yang menjadi lebih banyak dan kompleks. Namun seringkali ditemui data yang memiliki distribusi kelas yang tidak seimbang. Melakukan klasifikasi pada data tidak seimbang mengakibatkan model klasifikasi yang dihasilkan cenderung memprediksi kelas mayoritas dan mengabaikan kelas minoritas. Terdapat beberapa metode untuk mengatasi data tidak seimbang antara lain *oversampling* dan *undersampling*. SMOTE merupakan metode *oversampling* yang menyeimbangkan data dengan membuat *instance* sintetis untuk kelas minoritas. Sedangkan Tomek Links merupakan metode *undersampling* yang menghapus data dari kelas mayoritas yang memiliki karakteristik yang serupa. Namun Tomek Links hanya menghapus *instance* yang didefinisikan sebagai "Tomek Links" sehingga data yang dianalisis tidak dapat seimbang dan dalam penerapannya metode tersebut dikombinasikan dengan metode lain. Pada skripsi ini dilakukan penerapan metode kombinasi SMOTE dan Tomek Links pada data yang tidak seimbang dengan menggunakan metode klasifikasi *Random Forest* pada tiga *dataset*. Dari analisis yang dilakukan, diperoleh kesimpulan bahwa penerapan metode kombinasi SMOTE dan Tomek Links menghasilkan performa yang lebih baik dari pada metode SMOTE dan metode Tomek Links untuk analisis klasifikasi *Random Forest*.

Kata kunci : klasifikasi, data tidak seimbang, SMOTE, Tomek Links, Random Forest

***ABSTRACT***

***IMPLEMENTATION OF COMBINATION OF SMOTE AND TOMEK LINKS  
FOR CLASSIFICATION OF IMBALANCED DATA USING RANDOM  
FOREST METHOD***

by

Rona Agustika

17/412748/PA/18067

The rapid development in technology is directly proportional to the availability of data which is becoming more numerous and complex. However, we often encounter data that has an unbalanced class distribution. Classifying the unbalanced data results in the resulting classification model that tends to predict the majority class and ignore the minority class. There are several methods to deal with unbalanced data, including oversampling and undersampling. SMOTE is an oversampling method that balances data by creating synthetic instances for minority classes. Meanwhile, Tomek Links is an undersampling method that removes data from majority classes that have similar characteristics. However, Tomek Links only removes instances defined as "Tomek Links" so that the analyzed data cannot be balanced. In practice, this method is combined with other methods. In this thesis, the application of the combination of SMOTE and Tomek Links method is carried out on unbalanced data using the Random Forest classification method on three dataset. From the analysis, we can conclude that the combination of SMOTE and Tomek Links method results in better performance than SMOTE and Tomek Links for Random Forest classification analysis.

**Keywords :** classification, imbalanced data, SMOTE, Tomek Links, Random Forest