

## INTISARI

### ANALISIS PERBANDINGAN RANDOM FOREST, ADABOOST DAN GRADIENT TREE BOOSTING PADA IMBALANCED DATASET

Oleh

Rian Junianto Suryo Prayoga

18/433792/PPA/05607

Pengelolaan *imbalanced dataset* memiliki permasalahan ketika *minority class* menjadi penting untuk diprediksi dengan benar. Keadaan *imbalanced* menyebabkan *minority class* sulit diprediksi. *Imbalanced dataset* dapat ditemukan pada beberapa penelitian yang mengembangkan model klasifikasi. SMOTE merupakan salah satu metode *oversampling*, membangkitkan data sintetik pada *minority class* sehingga data seimbang. Penelitian ini fokus untuk mengkaji signifikansi metode SMOTE pada model klasifikasi dalam meningkatkan kinerja model yang dihasilkan oleh *ensemble learning*.

*Ensemble learning* adalah metode *learning* yang menghasilkan beberapa model kemudian model-model yang dihasilkan digunakan untuk menyelesaikan masalah, pada penelitian ini adalah masalah klasifikasi pada *imbalanced dataset*. *Ensemble learning* menggabungkan beberapa prediksi untuk menentukan suatu kelas. Metode *ensemble learning* yang digunakan adalah *Random Forest*, *AdaBoost* dan *Gradient Tree Boosting*. Pada beberapa penelitian *ensemble learning* memiliki prediksi lebih baik daripada *single learning*. Kemampuan model dievaluasi menggunakan *precision*, *recall*, *TNR* dan *F-measure* dan dataset yang digunakan diambil dari *UCI Machine Learning*.

Berdasarkan hasil percobaan tanpa SMOTE, model yang dihasilkan *AdaBoost* menunjukkan nilai *recall* dan *F-measure* yang lebih baik diantara *Random Forest* dan *Gradient Tree Boosting* pada dataset kedua dan ketiga. Secara berurutan untuk nilai *recall* 0.315, 0.1401 dan untuk nilai *F-measure* 0.2958, 0.104. Pada percobaan menggunakan SMOTE, *AdaBoost* masih menunjukkan nilai *recall* dan *F-measure* lebih baik antara *Random Forest* dan *Gradient Tree Boosting* pada dataset ke dua dan ke tiga, secara berurutan untuk nilai *recall* 0.638, 0.613 dan untuk nilai *F-measure* 0.4432, 0.2112. Dengan demikian, secara keseluruhan metode *oversampling* SMOTE dapat meningkatkan nilai *recall* pada masing-masing model yang dihasilkan.

**Kata Kunci:** *Imbalanced Dataset, Ensemble Classifier, Oversampling, SMOTE*

## **ABSTRACT**

### **COMPARATIVE ANALYSIS RANDOM FOREST, ADABOOST AND GRADIENT TREE BOOSTING ON IMBALANCED DATASET**

by

Rian Junianto Suryo Prayoga

18/433792/PPA/05607

Handling Imbalanced Dataset has a problem when the minority class is important to be classified correctly. Imbalanced Dataset causes the minority class hard to be classified. Imbalanced Dataset often to be found in research that focus on developing classification model. SMOTE is one of oversampling methods, SMOTE generates data synthetic for minority class. This research focus on studying the effect of SMOTE to increase the performance of ensemble learning.

Ensemble learning trains multiple classifiers to classify the data, in this research is Imbalanced Dataset. Ensemble learning combine some predictions to decide which class a data belongs to. In this research we use Random Forest, AdaBoost and Gradient Tree Boosting. We compare each other using precision, recall, F1 and TNR. We took the dataset we use from UCI Machine Learning.

Based on experiment without using SMOTE, AdaBoost performs better in two dataset, second dataset and third dataset with recall score 0.315 and 0.141 and F1 score 0.2958 and 0.104. Experiment with SMOTE also shows that AdaBoost performs better in two datasets, second dataset and third dataset with recall score 0.638 and 0.613 and F1 score 0.4432 and 0.2112. This research also shows that SMOTE can increase recall score from ensemble learning.

**Key Words:** *Imbalanced Dataset, Ensemble Classifier, Oversampling, SMOTE*