

## INTISARI

### IMPLEMENTASI SMOTE-ENN XGBOOST PADA ANALISIS KLASIFIKASI DATA TIDAK SEIMBANG

oleh

Rahayu Isnaini

19/440080/PA/19069

Data tidak seimbang menjadi masalah yang penting dalam klasifikasi biner. Model klasifikasi pada data tidak seimbang akan mengabaikan kelas minoritas dan cenderung lebih fokus mempelajari karakteristik data pada kelas mayoritas. Penanganan masalah data tidak seimbang pada analisis klasifikasi berfokus pada dua aspek utama yaitu pendekatan pada level data dan level algoritma. Pada level data digunakan metode SMOTE-ENN yang merupakan gabungan dari metode SMOTE dan ENN. Metode SMOTE menambah jumlah sampel pada kelas minoritas dengan cara membuat sampel sintetis sedangkan metode ENN digunakan untuk menghapus sampel yang dianggap *noise* dari data. Pada level algoritma digunakan XGBoost yang merupakan sistem *boosting* berbasis *decision tree* yang sangat *scalable* dan efektif dalam menyelesaikan masalah klasifikasi pada *machine learning*. Pada tugas akhir ini dilakukan analisis pada data *heart failure* dan *diabetes* dengan menggunakan SMOTE-ENN XGBoost, kemudian dibandingkan dengan XGBoost tanpa penanganan data tidak seimbang. Dari analisis yang dilakukan diperoleh kesimpulan bahwa metode SMOTE-ENN XGBoost menghasilkan performa yang paling baik.

Kata kunci: klasifikasi, data tidak seimbang, SMOTE-ENN, XGBoost.

***ABSTRACT***

***IMPLEMENTATION OF SMOTE-ENN XGBOOST FOR  
CLASSIFICATION ANALYSIS ON IMBALANCED DATA***

*by*

Rahayu Isnaini

19/440080/PA/19069

Imbalanced data becomes an important issue in binary classification. Classification models trained on imbalanced data tend to ignore the minority class and focus more on learning the characteristics of the majority class. Handling imbalanced data in classification analysis focuses on two main aspects: data-level approaches and algorithm-level approaches. At the data level, the SMOTE-ENN method is utilized, which is a combination of the SMOTE and ENN methods. SMOTE generates synthetic samples to increase the number of samples in the minority class, while ENN is used to remove noisy samples from the data. At the algorithm level, XGBoost is employed, which is a boosting system based on decision trees that is highly scalable and effective in solving classification problems in machine learning. In this final project, an analysis is conducted on heart failure and diabetes data using SMOTE-ENN XGBoost and it is compared with XGBoost without handling imbalanced data. From the analysis performed, it can be concluded that SMOTE-ENN XGBoost method has the best performance.

Keywords: classification, imbalanced data, SMOTE-ENN, XGBoost.