

## INTISARI

### **Metode Optimasi Ambang Batas pada Analisis Klasifikasi Data Tidak Seimbang**

Oleh

Daud Muhamad Azhari

21/474059/PA/20462

Klasifikasi merupakan bagian dari *supervised learning* yang bertujuan untuk memprediksi variabel respon dengan tipe kategorik. Dalam persoalan klasifikasi, tantangan yang sering dijumpai adalah ketidakseimbangan pada data. Tantangan tersebut muncul ketika label pada persoalan klasifikasi memiliki frekuensi dengan perbandingan cukup ekstrem. Pada kondisi tersebut, model klasifikasi cenderung bias dan berfokus pada kelas mayoritas dan cenderung mengabaikan kelas minoritas. Optimasi ambang batas merupakan salah satu metode untuk menangani permasalahan data tidak seimbang dengan menyesuaikan nilai ambang batas dalam proses klasifikasi. Tugas akhir ini berfokus pada analisis klasifikasi data tidak seimbang pada data *Drug Consumption (Quantified)* menggunakan metode optimasi ambang batas yang dikombinasikan dengan model dasar *Logistic Regression* dan SVM, serta metode *ensemble learning* berupa *Random Forest* dan *Gradient Boosting*. Berdasarkan analisis yang telah dilakukan, penggunaan metode optimasi ambang batas secara umum dapat meningkatkan performa klasifikasi, khususnya pada kelas minoritas, dengan meningkatkan metrik evaluasi *recall*, *F-measure*, dan *G-mean*, meski menurunkan metrik evaluasi *accuracy*, *precision*, dan *specificity*.

**Kata Kunci:** Optimasi Ambang Batas, *Threshold Moving*, Data Tidak Seimbang, *Random Forest*, *Gradient Boosting*

## ABSTRACT

### Threshold Optimization for Classification on Imbalanced Data

By

Daud Muhamad Azhari

21/474059/PA/20462

Classification is a part of supervised learning that aims to predict categorical response variables. In classification problems, imbalanced data is a common challenge. This challenge occurs when the labels in the classification problem have extreme proportions. In such conditions, the classification model tends to be biased, focusing on the majority class and often ignoring the minority class. One method of handling the issue of imbalanced data is threshold optimization, which works by adjusting the threshold value while classifying instances. This final project focuses on the classification of imbalanced data in the Drug Consumption (Quantified) dataset using threshold optimization, combined with basic model Logistic Regression, SVM, and also ensemble learning methods such as Random Forest and Gradient Boosting. Based on the results of the research, threshold optimization generally improves classification performance, especially in the minority class, by improving evaluation metrics such as recall, F-measure, and G-mean, although it lowers evaluation metrics like accuracy, precision, and specificity.

**Keyword:** Threshold Optimization, Threshold Moving, Imbalanced Data, Random Forest, Gradient Boosting