

INTISARI

PENANGANAN *IMBALANCED DATA MENGGUNAKAN SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE AND BOOSTING (SMOTEBOOST) PADA ANALISIS KLASIFIKASI KLASIFIKASI*

Oleh

Hanifah Nur Widianingtyas

19/442594/PA/19343

Data menjadi salah satu komponen yang sangat penting dalam era digital saat ini. Untuk menggali dan menggunakan informasi yang terkandung dalam data diperlukan analisis data yang tepat. Analisis klasifikasi bekerja dengan mengidentifikasi pola dan mengelompokkan data dalam kelas-kelas yang memiliki pola yang sama. Dalam analisis klasifikasi sering ditemukan masalah ketidakseimbangan data atau *imbalanced data*. *Imbalanced data* adalah kondisi saat kelas-kelas dalam analisis klasifikasi memiliki jumlah observasi yang berbeda secara signifikan. *Imbalanced data* menyebabkan terjadinya bias pada model yang dihasilkan dalam analisis klasifikasi, sehingga model tersebut lebih representatif terhadap kelas mayoritas. SMOTEBoost merupakan salah satu metode *hybrid-level* pada penanganan masalah *imbalanced data*. Pada tugas akhir ini dilakukan analisis pada data *Iranian Churn* menggunakan metode SMOTEBoost (*Synthetic Minority Oversampling Technique and Boosting*) dengan penerapan beberapa algoritma *boosting* serta dibandingkan performanya dengan metode *ensemble learning* lain, yaitu AdaBoost, *Gradient Boosting*, dan *LightGBM* dan dengan model dasarnya, yaitu *Decision Tree*. Berdasarkan analisis yang dilakukan didapatkan kesimpulan bahwa SMOTEBoost secara umum mampu meningkatkan performa klasifikasi dibanding model dasar dan *ensemble learning* lain. Selain itu, SMOTEBoost yang menerapkan *LightGBM* memberikan performa klasifikasi paling baik.

Kata Kunci: data tidak seimbang, analisis klasifikasi, SMOTEBoost, AdaBoost, *Gradient Boosting*, *LightGBM*

ABSTRACT

IMBALANCED DATA HANDLING USING SYNTHETIC MINORITY OVERSAMPLING AND BOOSTING (SMOTEBOOST) IN CLASSIFICATION ANALYSIS

by

Hanifah Nur Widianingtyas

19/442594/PA/19343

The world has entered the digital era, in which data plays a critical role. Proper data analysis is required to investigate and utilize the information found in the data. Classification analysis works by recognizing patterns in data and categorizing them into classes with the same pattern. Imbalanced data is a common issue in classification analysis. Imbalanced data is a condition where the classes in the classification analysis have a significantly different number of observations. In a classification analysis, imbalanced data results in a biased model that is more representative of the majority class. SMOTEBoost is a hybrid-level approach for dealing with data imbalances. This final project will analyze Iranian Churn data using the SMOTEBoost (Synthetic Minority Oversampling Technique and Boosting) method with the application of several boosting algorithms and compare its performance to other ensemble learning methods, such as AdaBoost, Gradient Boosting, and LightGBM, as well as the base model, DecisionTree. Based on the results of the research, it is possible to conclude that SMOTEBoost can generally increase classification performance compared to other base models and ensemble learning. Furthermore, SMOTEBoost, which uses LightGBM, provides the best classification performance.

Keyword: imbalanced data, classification analysis, SMOTEBoost, AdaBoost, Gradient Boosting, LightGBM