

INTISARI

PENANGANAN DATA TIDAK SEIMBANG MENGGUNAKAN *BORDERLINE SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE* (*BORDERLINE-SMOTE*) PADA ANALISIS KLASIFIKASI

oleh

Ika Rahmatunnisa

17/409519/PA/17826

Data dan informasi terus mengalami pertambahan seiring dengan berkembangnya teknologi digital. Ketersediaan data menjadi semakin banyak dan kompleks. Data tidak seimbang merupakan masalah yang sering muncul dalam analisis klasifikasi. Data tidak seimbang merupakan kondisi di mana distribusi kelas data yang tidak seimbang. Saat melakukan analisis klasifikasi pada data tidak seimbang, model klasifikasi yang dihasilkan cenderung memprediksi data yang berasal dari kelas mayoritas mengakibatkan performa klasifikasi menjadi kurang baik. Metode *Borderline-SMOTE* dapat diterapkan untuk menangani permasalahan data tidak seimbang. Metode *Borderline-SMOTE* merupakan metode yang dikembangkan dari metode *SMOTE*. *Borderline-SMOTE* merupakan metode *oversampling* yang menyeimbangkan data dengan membentuk *instance* sintetis dari kelas minoritas di area *borderline*. Pada skripsi ini dilakukan penerapan metode *Borderline-SMOTE* pada data yang tidak seimbang dengan menggunakan metode klasifikasi *Random Forest* dan *Naïve Bayes* pada dua *dataset*. Dari analisis yang dilakukan penanganan data tidak seimbang dengan *Borderline-SMOTE* secara umum dapat meningkatkan performa klasifikasi metode *Random Forest* dan *Naïve Bayes*.

Kata kunci: data tidak seimbang, klasifikasi, *Borderline-SMOTE*, *oversampling*.

ABSTRACT

IMBALANCED DATA HANDLING WITH BORDERLINE SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (BORDERLINE-SMOTE) FOR CLASSIFICATION ANALYSIS

By

Ika Rahmatunnisa

17/409519/PA/17826

Data and information continue to increase along with the development of digital technology. The availability of the data is becoming more and more complex. Imbalanced data is a problem that is often found in classification analysis. Imbalanced data is a condition where the distribution of data classes is not balanced. When classifying imbalanced data, the classification model tends to predict data that comes from the majority class resulting in poor classification performance. Borderline-SMOTE can be used to solve that imbalanced data problem. Borderline-SMOTE method is a method developed from the SMOTE method. Borderline-SMOTE is an oversampling method that creates a synthetic instance of the minority class in the borderline area. In this thesis, the Borderline-SMOTE method is applied to imbalanced data using the Random Forest and Naïve Bayes classification methods on two datasets. From the analysis, we can conclude that handling imbalanced data with Borderline-SMOTE method can improve the classification performance of Random Forest and Naïve Bayes methods.

Keywords: imbalanced data, classification, Borderline-SMOTE, oversampling