

INTISARI

ANALISIS KLASIFIKASI JAWABAN ESAI DENGAN SMOTE RANDOM FOREST DAN SMOTE ADABOOST

Oleh :

Wilia Satria

19/448736/PPA/05819

Pada penelitian sebelumnya telah banyak dilakukan metode machine learning dalam penerapan penilaian dalam bentuk uraian singkat, digunakan dalam mengevaluasi dan menilai esai siswa yang ditulis berdasarkan soal yang diberikan. Namun, terdapat ketidakseimbangan data kelas jawaban salah dan benar yang mengakibatkan pada performa ketepatan prediksi jawaban benar dan salahnya.

Berdasarkan analisis dataset yang telah dilakukan, terdapat ketidakseimbangan labeling dataset antar kelas jawaban benar dan salah, sehingga diperlukan teknik untuk mengatasi ketidakseimbangan data. Berdasarkan literatur, untuk mengatasi permasalahan tersebut dapat digunakan algoritma klasifikasi Random Forest dan Adaboost untuk meningkatkan konsistensi keakuratan klasifikasi dan metode SMOTE untuk mengatasi ketidakseimbangan data. Maka penerapan metode-metode tersebut diharapkan dapat meningkatkan performa klasifikasi sistem.

Hasil penelitian menunjukkan penggunaan SMOTE pada Random Forest dengan data yang digunakan pada penelitian ini menghasilkan peningkatan 12% pada parameter *recall*, 8% pada parameter *f1-score* dan 16% pada parameter *g-mean*. Penggunaan SMOTE pada Adaboost menghasilkan peningkatan 1,7% untuk *precision*, 12% pada *recall*, 17% pada parameter *g-mean* dan pada *f1-score* menghasilkan peningkatan 7%.

Kata Kunci: SMOTE, Data imbalance, Random Forest, AdaBoost.

ABSTRACT

ANALYSIS OF ESSAY ANSWER CLASSIFICATION WITH SMOTE RANDOM FOREST AND SMOTE ADABOOST

By

WILIA SATRIA

19/448736/PPA/05819

Previous study, machine learning methods have been applied to the application of assessments in the form of short descriptions used in evaluating and assessing student essays written based on the given subject. However, there is an imbalance in the data class of the correct and wrong answers resulting in the accuracy performance of the prediction of the right and incorrect answers.

Based on the analysis of data sets that has been performed, there is an imbalance in labelling data sets between classes, so techniques are needed to overcome the data imbalances. Based on the literature, to overcome these problems, the Random Forest and AdaBoost classification algorithms can be used to improve the consistency of classification accuracy and the SMOTE method to overcome data imbalances. So the implementation of these methods is expected to improve the performance of the system classification.

The results showed that the use of SMOTE in Random Forest with the data used in this study resulted in a 12% increase in the recall parameter, 16% in the g-mean parameter, and 8% in the f1-score parameter. Using SMOTE on Adaboost resulted in a 1,7% increase in precision, 12% in recall, 17% increase in g-mean and an f1-score resulted in a 7% increase.

Keywords: SMOTE, Data imbalance, Random Forest, AdaBoost.