

INTISARI

PENANGANAN DATA TIDAK SEIMBANG MENGGUNAKAN SMOTE PADA MODEL KLASIFIKASI UNTUK DIAGNOSA ANEMIA

Muhamad Fahmi

22/499651/PPA/06339

Anemia ditandai dengan kurangnya sel darah merah atau hemoglobin dalam tubuh yang dapat menyebabkan gejala seperti kelelahan, sesak napas, pusing, dan penurunan stamina. Oleh karena itu, diagnosis anemia yang cepat dan akurat penting untuk memberikan pengobatan yang tepat dan mencegah komplikasi yang lebih serius.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi untuk diagnosa anemia pada spesifik 4 kelas BTT, HbE, DB, dan kombinasi (mix) dengan menggunakan metode tunggal *Random Forest*, *SVM*, *Logistic Regression*, *Ensemble Learning Bagging* dengan teknik *oversampling* SMOTE pada data kelas minoritas dengan meningkatkan nilai *Recall*.

Hasil pengujian model pada penelitian ini mengungkap wawasan penting terkait performa berbagai model klasifikasi dalam mengidentifikasi anemia. *Logistic Regression* mencapai tingkat akurasi tertinggi 74.60%, menunjukkan ketepatan prediksinya. Model *Random Forest* dengan penerapan SMOTE mencapai nilai *Precision* tertinggi 68.84%. *Bagging SVM* dengan SMOTE menunjukkan *Recall* tertinggi 59.32%, efektif dalam mendeteksi sejumlah besar kelas yang sebenarnya. *F1-Score* tertinggi dicapai model SVM sebesar 58.14%, mencerminkan keseimbangan baik antara *Precision* dan *Recall*. Penggunaan SMOTE mampu meningkatkan *Recall* pada model *Bagging* seperti *Bagging RF* dengan SMOTE sebesar 2.73%, *Bagging SVM* dengan SMOTE sebesar 12.34%, begitu juga dengan model *Bagging LR* dengan SMOTE sebesar 3.79%. Temuan ini mengonfirmasi bahwa SMOTE dapat menjadi strategi yang efektif untuk meningkatkan kemampuan model dalam menangani kelas minoritas dan meningkatkan kualitas prediksi pada kondisi data yang tidak seimbang, meskipun tidak selalu berdampak positif pada akurasi keseluruhan model.

Kata Kunci: *Anemia, Imbalance Data, SMOTE, Ensemble Learning, Bagging*

ABSTRACT

HANDLING IMBALANCED DATA USING SMOTE IN CLASSIFICATION MODELS FOR ANEMIA DIAGNOSIS

Muhamad Fahmi

22/499651/PPA/06339

Anemia is characterized by a lack of red blood cells or hemoglobin in the body which can cause symptoms such as fatigue, shortness of breath, dizziness, and decreased stamina. Therefore, a quick and accurate diagnosis of anemia is important to provide appropriate treatment and prevent more serious complications.

This study aims to develop a classification model for anemia diagnosis on specific 4 classes of BTT, HbE, DB, and combination (mix) using a single method *Random Forest*, *SVM*, *Logistic Regression*, *Ensemble Learning Bagging* with SMOTE oversampling technique on minority class data by increasing the *Recall* value.

The results of model testing in this study reveal important insights into the performance of various classification models in identifying anemia. Logistic Regression achieved the highest accuracy rate of 74.60%, demonstrating its prediction accuracy. The Random Forest model with the application of SMOTE achieved the highest Precision value of 68.84%. SVM bagging with SMOTE showed the highest Recall of 59.32%, effective in detecting a large number of true classes. The highest F1-Score was achieved by the SVM model at 58.14%, reflecting a good balance between Precision and Recall. The use of SMOTE was able to increase Recall in Bagging models such as RF Bagging with SMOTE by 2.73%, SVM Bagging with SMOTE by 12.34%, as well as LR Bagging model with SMOTE by 3.79%. These findings confirm that SMOTE can be an effective strategy to improve the model's ability to handle minority classes and improve prediction quality under conditions of imbalanced data, although it does not always have a positive impact on the overall accuracy of the model.

Keyword: *Anemia, Imbalance Data, SMOTE, Ensemble Learning, Bagging*