

INTISARI

ADASYN TERMODIFIKASI (*Modified*) UNTUK MULTICLASS PADA KASUS KLASIFIKASI PENYAKIT ANEMIA

Oleh

Velssy

23/525862/PPA06623

Anemia merupakan salah satu penyakit darah paling umum di dunia dan masih menjadi masalah kesehatan yang signifikan, terutama pada remaja perempuan, ibu hamil, dan anak-anak. Dalam beberapa tahun terakhir, *Machine Learning* mulai dimanfaatkan untuk membantu proses klasifikasi jenis anemia, namun proses ini menghadapi tantangan besar berupa ketidakseimbangan data antar kelas diagnosis. Ketidakseimbangan tersebut menyebabkan model lebih mudah mengenali kelas mayoritas, tetapi memberikan *Recall* yang rendah pada keseluruhan prediksi. Penelitian sebelumnya menunjukkan hal ini dengan Akurasi sekitar 74% dan *Recall* 59% meskipun telah menggunakan teknik *oversampling* SMOTE. Namun, SMOTE masih kurang efektif untuk skenario multiclass karena tidak memperhitungkan variasi dan jarak antar kelas. Sementara itu, ADASYN sebagai metode *oversampling* adaptif hanya optimal untuk *binary classification* dan belum mampu menangani kompleksitas hubungan antar kelas pada data multiclass secara menyeluruh.

Untuk mengatasi permasalahan tersebut, diusulkan menggunakan pendekatan *oversampling* ADASYN termodifikasi (*Modified*), yaitu pengembangan dari metode ADASYN standar untuk menangani data tidak seimbang pada kasus multiclass dengan pendekatan *pairwise*. Teknik ini diterapkan pada enam model klasifikasi, yaitu *Random Forest*, *Support Vector Machine*, *Logistic Regression*, *Decision Tree*, *K-Nearest Neighbors*, dan *Naïve Bayes*. Proses evaluasi dilakukan menggunakan *classification report* untuk melihat perubahan nilai Presisi, *Recall*, dan *F1-score* pada kelas mix sebelum dan sesudah penerapan *oversampling*.

Hasil penelitian menunjukkan bahwa metode *Modified* mampu meningkatkan performa klasifikasi secara keseluruhan, terutama pada model yang sensitif terhadap perubahan distribusi data seperti *Random Forest*, *SVM*, dan *Decision Tree*. Pada *Random Forest*, metode ini meningkatkan Akurasi dari 69.09% menjadi 80% dan *Recall* dari 50.28% menjadi 68.11%. *Decision Tree* juga menunjukkan respons yang baik dengan kenaikan *Recall* lebih dari 11 poin, sementara *SVM* mengalami peningkatan moderat pada kedua metrik tersebut. *Modified* berhasil mengatasi masalah rendahnya Akurasi dan *Recall* pada penelitian sebelumnya, sehingga lebih sesuai untuk menangani ketidakseimbangan data anemia yang bersifat *multiclass*.

Kata Kunci: ADASYN, *Machine Learning*, *Imbalanced Class*, *Anemia*

ABSTRACT

Anemia is one of the most common blood disorders worldwide and remains a significant health problem, particularly among adolescent girls, pregnant women, and children. In recent years, Machine Learning has been widely applied to assist anemia classification; however, the process faces a major challenge due to class imbalance in diagnosis data. This imbalance causes classification models to favor majority classes while producing low Recall for minority classes. Previous studies reported an Accuracy of approximately 74% and a Recall of 59% despite using the SMOTE oversampling technique. However, SMOTE is less effective for multiclass scenarios because it does not consider inter-class variation and distance. Meanwhile, ADASYN, as an adaptive oversampling method, is mainly optimized for binary classification and has not fully addressed the complex relationships among classes in multiclass datasets.

To address this issue, this study proposes a modified ADASYN oversampling approach (Modified ADASYN), an extension of the standard ADASYN method designed for multiclass imbalanced data using a pairwise strategy. The method was implemented on six classification models: Random Forest, Support Vector Machine (SVM), Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Naïve Bayes. Model evaluation was conducted using a classification report to analyze changes in Precision, Recall, and F1-score, particularly for the minority Mix class before and after oversampling.

The results demonstrate that Modified ADASYN improved overall classification performance, especially for models sensitive to data distribution changes such as Random Forest, SVM, and Decision Tree. In Random Forest, the proposed method increased Accuracy from 69.09% to 80% and Recall from 50.28% to 68.11%. Decision Tree also showed a significant improvement with Recall increasing by more than 11 points, while SVM achieved moderate improvements in both metrics. These findings indicate that Modified ADASYN successfully addressed the low Accuracy and Recall reported in previous studies and is more suitable for handling multiclass imbalanced anemia datasets.

Keyword: ADASYN, Machine Learning, Imbalanced Class, Anemia