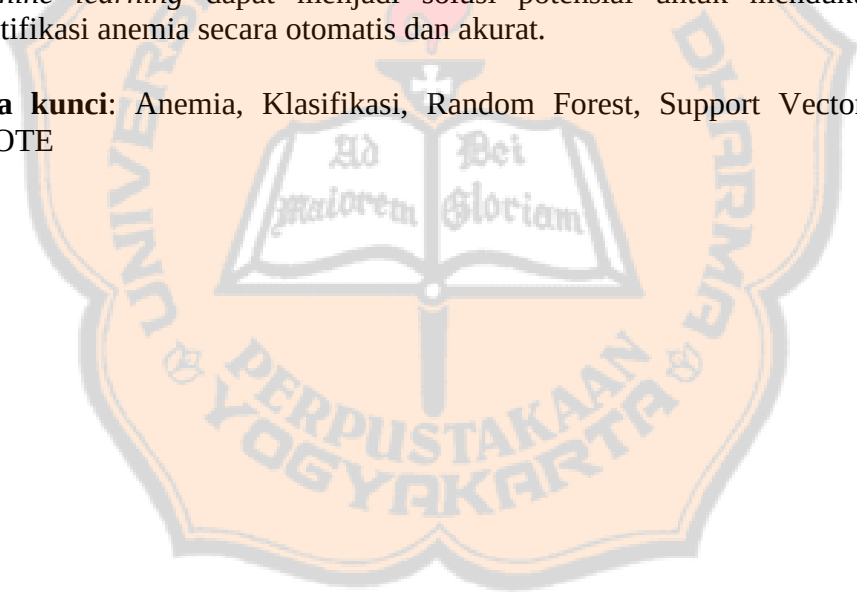


ABSTRAK

Anemia merupakan salah satu masalah kesehatan yang umum terjadi, terutama pada kelompok usia remaja dan dewasa muda. Untuk membantu proses diagnosis secara otomatis, penelitian ini membandingkan dua algoritma klasifikasi, yaitu *Random Forest* dan *Support Vector Machine* (SVM), dalam mengklasifikasikan penyakit anemia. Dataset yang digunakan terdiri dari 1001 data pasien dan 13 fitur kesehatan yang telah melalui tahap preprocessing serta penyeimbangan kelas menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Pemodelan dilakukan dengan melakukan pencarian hyperparameter terbaik menggunakan *GridSearchCV* dengan *10-fold cross-validation*. Evaluasi performa model dilakukan menggunakan metrik akurasi, *F1-score*, *confusion matrix*. Hasil menunjukkan bahwa model *Random Forest* menghasilkan akurasi tertinggi sebesar 98,47%, dengan *F1-score* sebesar 0,99 untuk kelas mayoritas dan 0,98 untuk kelas minoritas. Sementara itu, model *SVM* dengan *kernel RBF* dan parameter terbaik menghasilkan akurasi sebesar 98,06%, dengan *F1-score* 0,99 pada kelas mayoritas dan 0,97 pada kelas *minoritas*. Berdasarkan evaluasi secara keseluruhan, *Random Forest* menunjukkan performa yang lebih unggul dalam klasifikasi penyakit anemia pada dataset yang digunakan. Penelitian ini menunjukkan bahwa *algoritma machine learning* dapat menjadi solusi potensial untuk mendukung proses identifikasi anemia secara otomatis dan akurat.

Kata kunci: Anemia, Klasifikasi, Random Forest, Support Vector Machine, SMOTE



ABSTRACT

Anemia is one of the most common health issues, particularly among adolescents and young adults. To assist the diagnostic process automatically, this study compares two classification algorithms, namely Random Forest and Support Vector Machine (SVM), in classifying anemia cases. The dataset used consists of 1001 patient records and 13 health-related features, which have undergone preprocessing and class balancing using the Synthetic Minority Oversampling Technique (SMOTE). Model training was performed by tuning hyperparameters using GridSearchCV with 10-fold cross-validation. The performance of each model was evaluated using accuracy, precision, recall, F1-score, confusion matrix, ROC curve, and precision-recall curve. The results show that the Random Forest model achieved the highest accuracy of 98.47%, with an F1-score of 0.99 for the majority class and 0.98 for the minority class. Meanwhile, the SVM model with an RBF kernel and optimal parameters achieved an accuracy of 98.06%, with an F1-score of 0.99 for the majority class and 0.97 for the minority class. Overall, Random Forest outperformed SVM in classifying anemia cases in the given dataset. This study demonstrates that machine learning algorithms can serve as a potential solution to support automatic and accurate identification of anemia.

Keywords: *Anemia, Classification, Random Forest, Support Vector Machine, SMOTE*

