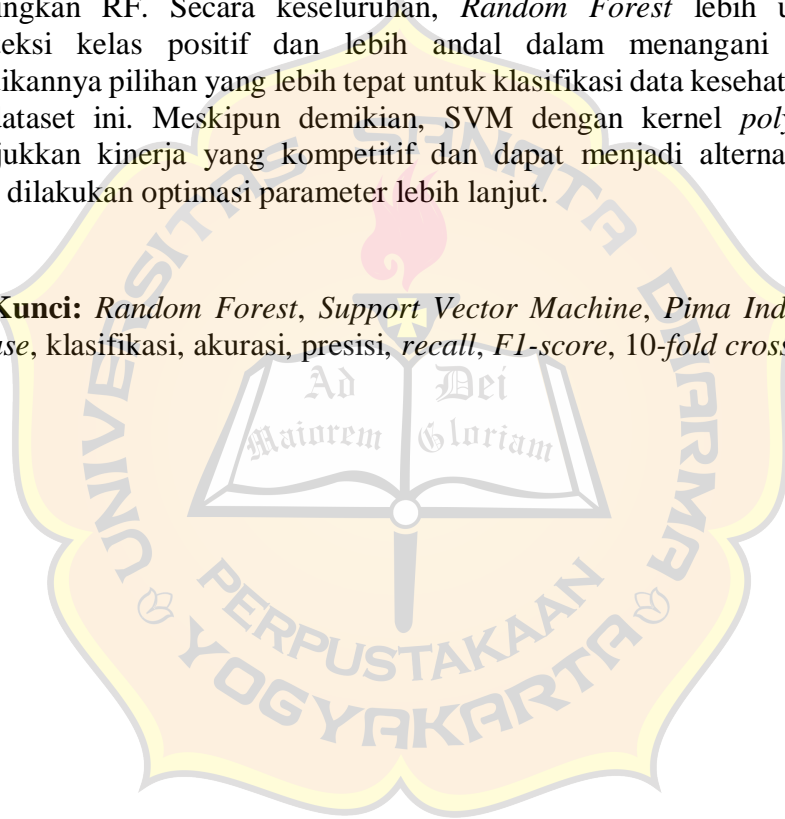


ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM) dalam melakukan klasifikasi pada dataset *Pima Indians Diabetes Database*, yang digunakan untuk memprediksi kemungkinan seseorang menderita diabetes. Kedua algoritma diuji menggunakan *10-fold cross-validation* untuk memperoleh gambaran yang lebih stabil dan akurat mengenai kinerja masing-masing model. Evaluasi dilakukan dengan menggunakan beberapa metrik klasifikasi, yaitu akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *Random Forest* (RF) memiliki keunggulan dalam hal stabilitas kinerja, dengan rata-rata akurasi 76,3% dan performa yang konsisten antar *fold*. Sementara itu, SVM dengan kernel *polynomial* menunjukkan sedikit keunggulan dalam hal presisi (74,57%), namun kalah dalam *recall* dan *F1-score* dibandingkan RF. Secara keseluruhan, *Random Forest* lebih unggul dalam mendeteksi kelas positif dan lebih andal dalam menangani variasi data, menjadikannya pilihan yang lebih tepat untuk klasifikasi data kesehatan, khususnya pada dataset ini. Meskipun demikian, SVM dengan kernel *polynomial* tetap menunjukkan kinerja yang kompetitif dan dapat menjadi alternatif yang baik apabila dilakukan optimasi parameter lebih lanjut.

Kata Kunci: *Random Forest*, *Support Vector Machine*, *Pima Indians Diabetes Database*, klasifikasi, akurasi, presisi, *recall*, *F1-score*, *10-fold cross-validation*.



ABSTRACT

This study aims to compare the performance of the Random Forest (RF) and Support Vector Machine (SVM) algorithms in classifying the Pima Indians Diabetes Database, which is used to predict the likelihood of an individual having diabetes. Both algorithms were tested using 10-fold cross-validation to obtain a more stable and accurate picture of each model's performance. Evaluation was performed using several classification metrics, including accuracy, precision, recall, and F1-score. The results indicate that Random Forest (RF) has the advantage in terms of stability, with an average accuracy of 76.3% and consistent performance across folds. Meanwhile, SVM with the polynomial kernel shows a slight advantage in precision (74.57%), but lags behind RF in recall and F1-score. Overall, Random Forest is superior in detecting the positive class and more reliable in handling data variability, making it a more suitable choice for classifying health data, particularly for this dataset. Nevertheless, SVM with the polynomial kernel still demonstrates competitive performance and could serve as a good alternative if parameter optimization is further conducted.

Keywords: *Random Forest, Support Vector Machine, Pima Indians Diabetes Database, classification, accuracy, precision, recall, F1-score, 10-fold cross-validation.*

