

ABSTRAK

Adelina Dwi Wardani

Perbandingan Akurasi Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dengan Teknik SMOTE untuk Prediksi Penyakit Diabetes

11 lampiran depan + 61 halaman + 11 tabel + 19 gambar + 3 lampiran akhir

Latar Belakang: Diabetes melitus merupakan salah satu penyakit kronis yang banyak diderita masyarakat dan berpotensi menimbulkan komplikasi serius apabila tidak terdeteksi secara dini. Seiring berkembangnya teknologi, machine learning menjadi pendekatan yang efektif dalam memprediksi risiko diabetes secara cepat dan akurat.

Metode: Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk memprediksi risiko diabetes berdasarkan dataset Pima Indian Diabetes. Data dibagi menggunakan metode split 70:30 dan diseimbangkan menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan kelas. Evaluasi performa model dilakukan dengan metrik akurasi, *precision*, *recall*, *F1-score*, dan *ROC AUC*. Proses tuning hyperparameter dilakukan menggunakan *GridSearchCV* dengan *8-fold cross-validation*.

Hasil Penelitian: Model *Random Forest* dengan penerapan SMOTE menunjukkan performa terbaik dengan akurasi sebesar 78,35%, *precision* 84%, *recall* 72%, dan *F1-score* sebesar 70%. Hasil ini menunjukkan bahwa penggunaan teknik *balancing data* dan optimasi parameter dapat meningkatkan kinerja model secara signifikan dalam mendeteksi risiko diabetes.

Kesimpulan: Penerapan algoritma *machine learning* seperti *Random Forest* dan SVM mampu membantu prediksi risiko diabetes secara efektif. *Random Forest* dengan teknik SMOTE memberikan hasil yang lebih optimal dibandingkan metode lainnya. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem deteksi dini diabetes untuk mendukung pengambilan keputusan dalam bidang kesehatan.

Kata Kunci : Diabetes, *Random Forest*, *Support Vector Machine*, SMOTE, Prediksi, *Machine Learning*

Referensi : 21 (2017-2025)

ABSTRACT

Adelina Dwi Wardani

Comparison of the Accuracy of Support Vector Machine (SVM) and Random Forest (RF) Algorithms with the SMOTE Technique for Diabetes Prediction

11 Front Attachments + 61 Pages + 11 Tables + 19 Images + 3 Back Attachments

Background: Diabetes mellitus is a chronic disease that affect many people and has the potential to cause serious complications if not detected early. With the advancement of technology, machine learning has become an effective approach for predicting diabetes risk quickly and accurately.

Methods: This study used the Support Vector Machine (SVM) and Random Forest (RF) algorithms to predict diabetes risk based on the Pima Indian Diabetes dataset. The data was divided using a 70 : 30 split method and balanced using the SMOTE technique to address class imbalance. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC AUC metics. Hyperparameter tuning was performed using GridSearchCV with 8-fold cross-validation.

Research Result: The Random Forest model with SMOTE implementation demonstrated the best performance with an accuracy of 78.35%, a precision of 84%, a recall of 72%, and an F1-score of 70%. These result indicate that the use of data balancing techniques and parameter optimization can significantly improve model performance in detecting diabetes risk.

Conclusion: The application of machine learning algorithms such as Random Forest and SVM can help effectively predict diabetes risk. Random Forest with SMOTE techniques provided more optimal result than other methods. This research is expected to form the basis for developing an early diabetes detection system to support decision-making in the healthcare sector.

Keywords: Diabetes, Random Forest, Support Vector Machine, SMOTE, Prediction, Machine Learning.

References: 21 (2017-2025)