ABSTRAK

Enos, NIM 4213550026 (2025). Implementasi Algoritma *Random Forest* Untuk Prediksi Risiko Diabetes Berbasis Data Klinis

Diabetes merupakan penyakit metabolik kronis yang memerlukan deteksi dini untuk mencegah komplikasi serius. Penelitian ini bertujuan membangun model prediksi risiko diabetes menggunakan algoritma *Random Forest* berdasarkan data klinis pasien RSUD dr. M. Thomsen Nias. Data penelitian berjumlah 140 sampel pasien dengan variabel usia, tekanan darah, glukosa darah, kolesterol, HbA1c, pola makan, dan jenis kelamin. Tahapan penelitian meliputi praproses data, pembangunan *pipeline Random Forest*, optimasi *hyperparameter* (*RandomizedSearchCV*), serta evaluasi kinerja model menggunakan 5-*Fold Cross Validation*. Hasil menunjukkan kinerja model dengan akurasi 93,6%, presisi 94,2%, *recall* 99,2%, dan *F1-Score* 9,67%. Faktor paling berpengaruh dalam prediksi adalah glukosa darah sewaktu, glukosa darah puasa, HbA1c, dan usia. Model kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan *Streamlit*, yang memudahkan deteksi dini risiko diabetes. Penelitian ini membuktikan bahwa algoritma *Random Forest* efektif dan andal untuk prediksi risiko diabetes berbasis data klinis serta dapat diintegrasikan ke sistem informasi medis.

Kata kunci: *Random Forest*, Prediksi Diabetes, Data Klinis, *Machine Learning*, *Streamlit*

ABSTRACT

Enos Christean Oimolala Telaumbanua, NIM 4213550026 (2025). ***Implementation of the Random Forest Algorithm for Diabetes Risk Prediction Based on Clinical Data.***

Diabetes Mellitus is a chronic metabolic disease that requires early detection to prevent severe complications. This study aims to develop a diabetes risk prediction model using the Random Forest algorithm based on clinical data from patients at RSUD dr. M. Thomsen Nias. The dataset consists of 140 samples containing variables such as age, blood pressure, blood glucose, cholesterol, HbA1c, dietary patterns, and gender. The research stages include data preprocessing, Random Forest model pipeline construction, hyperparameter optimization (RandomizedSearchCV), and model evaluation using 5-Fold Cross Validation. The results show strong model performance with accuracy of 93.6%, precision of 94.2%, recall of 99.2%, and F1-score of 96.7%. The most influential features were random blood glucose, fasting glucose, HbA1c, and age. The model was then implemented into a web-*based application using Streamlit, allowing users to easily perform early detection of diabetes risk. This study demonstrates that the Random Forest algorithm is effective and reliable for predicting diabetes risk based on clinical data and can be integrated into medical information systems.*

Keywords: *Random Forest, Diabetes Prediction, Clinical Data, Machine Learning, Streamlit*