

ABSTRAK

Diabetes merupakan penyakit serius yang disebabkan oleh ketidakmampuan pankreas dalam memproduksi insulin secara optimal. Jika tidak dideteksi dan ditangani dengan baik, diabetes dapat menyebabkan komplikasi berat seperti kebutaan, gagal jantung, bahkan kematian. Faktor-faktor penyebab diabetes antara lain keturunan, berat badan, usia, tekanan darah serta gaya hidup. Namun, gejala awal sering kali tidak terlihat sehingga diperlukan metode deteksi dini untuk mengurangi risiko komplikasi. Penelitian ini bertujuan untuk menerapkan metode klasifikasi dalam data mining menggunakan algoritma C4.5 dalam membentuk pohon Keputusan (*Decision Tree*) untuk mengklasifikasikan penderita penyakit diabetes. Penelitian ini menggunakan dataset dari *Kaggle* yang berisi 2.000 data dengan label target 2 kelas yaitu positif diabetes sebanyak 684 data dan negatif diabetes sebanyak 1316 data. Data dilakukan transformasi sebelum diklasifikasikan menggunakan *algoritma Decision Tree C4.5*. Pengujian dilakukan menggunakan *software rapidminer* untuk mengevaluasi performa model berdasarkan akurasi, presisi, *recall* dan *f1score* serta menggunakan 3 kali skenario uji. Dari hasil pengujian skenario uji pertama, diperoleh akurasi sebesar 72%, skenario uji kedua diperoleh akurasi sebesar 76,5%, skenario uji ketiga, diperoleh akurasi sebesar 74,8%. Hasil ini menunjukkan bahwa model klasifikasi yang diterapkan memiliki kinerja yang cukup baik dalam mengidentifikasi penderita diabetes.

Kata kunci : Diabetes, *Decision Tree*, Algoritma C4.5, Klasifikasi, *Confusion Matrix*

ABSTRACT

Diabetes is a serious disease caused by the inability of the pancreas to produce insulin optimally. If not detected and treated properly, diabetes can cause serious complications such as blindness, heart failure, and even death. Factors that cause diabetes include heredity, weight, age, blood pressure, and lifestyle. However, early symptoms are often not visible so early detection methods are needed to reduce the risk of complications. This study aims to apply the classification method in data mining using the C4.5 algorithm in forming a Decision Tree to classify people with diabetes. This study uses a dataset from Kaggle containing 2,000 data with 2 class target labels, namely 684 positive diabetes data, and 1316 negative diabetes data. The data was transformed before being classified using the Decision Tree C4.5 algorithm. Testing was carried out using rapid miner software to evaluate model performance based on accuracy, precision, recall, and f1-score and using 3 test scenarios. From the results of the first test scenario, an accuracy of 72% was obtained, the second test scenario obtained an accuracy of 76.5%, and the third test scenario obtained an accuracy of 74.8%. These results indicate that the applied classification model has performed quite well in identifying diabetes sufferers.

Keywords: Diabetes, Decision Tree, C4.5 Algorithm, Classification, Confusion Matrix