

## ABSTRAK

Diabetes Mellitus atau juga disebut kencing manis adalah penyakit yang terjadi ketika kadar glukosa darah terlalu tinggi. Sebagian besar kematian terjadi karena peningkatan kadar glukosa, yang merusak pembuluh darah, saraf, dan struktur internal lainnya. Di bidang medis, banyak sekali catatan penyakit pasien, salah satunya adalah data diabetes. Namun banyaknya data tersebut belum dapat digunakan dengan baik jika belum ada informasi dan kesimpulan dari data tersebut. Seperti kesulitan dalam memprediksi apakah pasien tersebut menderita diabetes atau tidak. Dengan begitu diperlukan proses ekstraksi untuk menemukan informasi dalam data yang sebelumnya tidak diketahui disebut *data mining* dengan menggunakan klasifikasi. Algoritma *Decision Tree C4.5* adalah salah satu metode untuk klasifikasi menggunakan pohon keputusan. Penelitian ini menggunakan dataset *Pima Indian Diabetes Database* (PIDD) dari *UCI Machine Learning* menghasilkan akurasi yang berbeda menggunakan *Confusion Matrix*. Hasil akurasi yang dilakukan pada algoritma C4.5 sebesar 75%, sedangkan hasil akurasi pada algoritma C4.5 dengan seleksi fitur *Particle Swarm Optimization* sebesar 81%. Terjadi peningkatan akurasi pada saat menambahkan *Particle Swarm Optimization* sebanyak 6%. Dengan menggunakan *Particle Swarm Optimization* terjadi seleksi fitur pengurangan jumlah atribut dari 9 atribut menjadi 4 atribut yaitu *Glucose*, *SkinThickness*, *BMI* dan *DiabetesPedigreeFunction*.

Kata Kunci : Klasifikasi, *Decision Tree C4.5*, Seleksi Fitur, *Particle Swarm Optimization*

## ABSTRACT

Diabetes Mellitus, also known as diabetes, is a disease that occurs when blood glucose levels are too high. Most deaths occur due to elevated glucose levels, which damage blood vessels, nerves and other internal structures. In the medical field, there are lots of patient records, one of which is diabetes data. However, the amount of data cannot be used properly if there is no information and conclusions from the data. Such as the difficulty in predicting whether the patient has diabetes or not. Thus, an extraction process is needed to find information in previously unknown data, which is called data mining using classification. Decision Tree C4.5 algorithm is one method for classification using decision trees. This study uses the Pima Indian Diabetes Database (PIDD) dataset from UCI Machine Learning to produce different accuracy using the Confusion Matrix. The accuracy results for the C4.5 algorithm are 75%, while the accuracy results for the C4.5 algorithm with the Particle Swarm Optimization feature selection are 81%. There is an increase in accuracy when adding Particle Swarm Optimization by 6%. By using Particle Swarm Optimization, there is a feature selection that reduces the number of attributes from 9 attributes to 4 attributes, namely Glucose, SkinThickness, BMI and DiabetesPedigreeFunction.

Keyword : Classification, Decision Tree C4.5, Feature Selection, Particle Swarm Optimization