

ABSTRAK

Diabetes atau dapat disebut dengan penyakit gula darah adalah salah satu jenis penyakit kronis yang mempunyai tanda awal berupa meningkatnya kadar gula darah akibat adanya gangguan sistem metabolisme di dalam tubuh. Pada masa pandemi ini pengidap penyakit diabetes bukanlah masalah yang dapat dianggap remeh. Pada kasus pasien Covid-19 Diabetes merupakan salah satu faktor risiko kematian. Teknik *data mining* dapat dimanfaatkan untuk memprediksi suatu penyakit dari data rekam medis pasien. Dengan metode klasifikasi pada *data mining*, berupa data seperti umur, jenis kelamin, tekanan darah dan atribut lainnya, dapat digunakan menjadi faktor pendukung untuk mengetahui kemungkinan pasien terkena suatu penyakit. Salah satu algoritma yang dapat digunakan adalah Algoritma C4.5. Dataset yang digunakan pada penelitian ini adalah *Pima Indians Diabetes Database* yang didapat dari situs Kaggle. Dataset berjumlah 768 data yang semuanya berasal dari jenis kelamin wanita dengan umur sekurang-kurangnya 21 tahun yang terdiri dari 8 variabel independen dan 1 variabel dependen (target). Untuk mengatasi *missing value* dan atribut kontinu pada dataset tersebut peneliti menggunakan imputasi K-NN dan diskritisasi berbasis *Entropy*. Untuk meningkatkan kinerja algoritma C4.5 peneliti menggunakan *ensemble Adaboost*. Penggunaan algoritma *ensemble adaboost* dapat meningkatkan akurasi klasifikasi pada algoritma C4.5. Hal ini dapat dibuktikan dengan perolehan nilai rata-rata akurasi tertinggi pada keseluruhan fold terdapat pada algoritma *adaboost* dengan nilai 77,86%. Sedangkan pengaruh penggunaan imputasi K-NN untuk mengatasi *missing value* dapat meningkatkan nilai akurasi. Hal ini dapat dibuktikan dengan perolehan nilai rata-rata akurasi pada keseluruhan *fold* dengan menggunakan imputasi K-NN mendapat nilai lebih tinggi jika dibandingkan dengan tanpa imputasi K-NN pada data *missing value*.

Kata kunci :Adaboost, C4.5, Diabetes, Ensemble