

Abstrak

Penyakit Diabetes Mellitus (DM) merupakan penyakit penyebab kematian terbesar ke-3 di Indonesia, dan penyakit DM tipe II lebih berbahaya karena kombinasi antara faktor genetik dan gaya hidup. Tingginya tingkat penderita DM tipe II dikarenakan keterlambatan penegakan diagnosis penyakit, oleh karena itu diperlukan pendeteksian awal penyakit untuk pengklasifikasian pasien yang terdeteksi penyakit DM tipe II dan tidak terdeteksi penyakit diabetes tipe II, serta menganalisis variabel utama penentu pendeteksian awal penyakit tersebut. Pada penelitian ini diimplementasikan gabungan metode *Classification and Regression Tree* (CART) dan *Random Forest* (RF) untuk membangun model klasifikasi yang digunakan dalam pendeteksian awal penyakit DM tipe II. Hal ini didasarkan oleh karakteristik *dataset* rekam medik yang digunakan merupakan *dataset* yang memiliki variabel kompleks yang terdiri dari beberapa variabel kategorik dan variabel kontinu, serta kelebihan model CART yaitu mudah diimplementasikan, dan dapat mengeksplorasi struktur data rekam medis yang kompleks, sedangkan metode RF dapat menangani masalah akurasi. Dalam penelitian ini dilakukan pengujian terhadap perbedaan jumlah pohon dan jumlah kandidat atribut pemecah. Dari hasil pengujian didapatkan bahwa penambahan jumlah pohon dan atribut pemecah meningkatkan nilai akurasi dan mengurangi nilai *error rate* yang dihasilkan dengan nilai optimal jumlah pohon 50 dan jumlah atribut pemecah 3 dengan nilai rata-rata akurasi 84%. Serta variabel penentu utama pendeteksian awal penyakit DM tipe II yaitu keturunan, umur, dan IMT.

Kata kunci: Diabetes Mellitus tipe II, CART, *Random Forest*, klasifikasi, pendeteksian awal penyakit, variabel utama.