

ABSTRAK

Model data mining bisa digunakan untuk melakukan klasifikasi penyakit yang dapat mengidap seseorang. Pada klasifikasi terdapat metode Algoritma C4.5, ID3, Naïve bayes, dan lain-lain. Metode Algoritma C4.5 lebih baik pada metode lain dikarenakan tingkat akurasi yang tinggi dan waktu komputasi yang cepat menjadikan Algoritma C4.5 sangat cocok digunakan. Pada penelitian ini data yang digunakan adalah Stroke Prediction Datasets dari Kaggle UCI repository. Dalam dataset ini masih terdapat missing value maka dari itu digunakanlah metode imputasi menggunakan Mean untuk menghilangkan data abnormal yang masih kosong pada dataset tersebut. Adapun karena dataset ini memiliki nilai kontinyu maka dilakukan diskritisasi untuk mengubah atribut menjadi diskrit agar mudah dipahami. Dari metode metode yaitu Algoritma C4.5, Imputasi mean, dan diskritisasi berbasis equal width interval yang diterapkan pada dataset Kaggle ini. Penggunaan *SMOTE* untuk menyeimbangkan data yang dimana peneliti gunakan adalah dataset stroke yang tidak berimbang untuk nilai positif dan negatif. Sedangkan untuk imputasi mean digunakan dengan tujuan untuk menghilangkan *missing value* dengan hasil meningkatkan akurasi. Hasil fold tertinggi tanpa menggunakan *SMOTE* adalah 95,05%

Kata kunci : Penyakit Stroke, Algoritma C4.5, Kecerdasan Komputasional, Imputasi Mean, Diskritisasi, Smote.