

Abstrak

Dengan berkembangnya ilmu pengetahuan dan teknologi dalam ilmu kedokteran, gen yang dimiliki oleh manusia dapat direpresentasikan dalam bentuk data dengan jumlah attribute yang banyak. Data tersebut dapat dikatakan sebagai data berdimensi tinggi. Data tersebut perlu diolah untuk menggali informasi yang tersimpan, banyak algoritma yang dapat mengolah data tersebut. Salah satunya ialah algoritma CS4 (*Cascading and Sharing for Ensemble of Decision Trees*) yang merupakan varian dari *ensemble method*.

Attribute selection berfungsi untuk memilih attribute terbaik berdasarkan nilai *information gain*. Hal yang membedakan algoritma CS4 dengan algoritma bagging dan boosting adalah penggunaan data dalam pembangunan model. Algoritma CS4 menggunakan data asli dalam setiap pembangunan model, sedangkan algoritma bagging dan boosting menggunakan data bootstrap.

Dari pengujian penggunaan algoritma CS4 yang telah dilakukan, didapatkan bahwa tingkat keakurasian sangat ditentukan oleh penentuan jumlah attribute yang digunakan pada saat pembangunan model.

Kata kunci: CS4, *ensemble method*, *attribute selection*, *bagging*, *boosting*.