

Abstrak

Classification adalah salah satu task dalam data mining yaitu predictive task yang digunakan untuk memprediksi nilai dari variable target yang diskrit. *Classification* telah secara luas diimplementasikan diberbagai bidang seperti *Direct Marketing*, *Spam detection*, *Functional Classification of Proteins*, *Galaxy Classficaiton* dan sebagainya. Metode dalam *Data Mining Classification* antara lain adalah *Decision tree*, *Neural Network*, *k-Nearest Neighbor*, *Case Based Reasoning*, *Algoritma Genetika* dan sebagainya. *Decision Tree* sendiri merupakan salah satu fungsional yang menggunakan representasi tree untuk menentukan aturan-aturan klasifikasi. Salah satu metode dalam *Decision Tree* ialah *ID3 (iterative Dichotomiser three)*. Algoritma memiliki fungsi yaitu dengan melakukan pembobotan pada setiap atribut, kemudian dengan melihat bobot tersebut dibangun tree berdasarkan bobot terbesar ke bobot terkecil dan pohon ini akan rekursif dibangun dengan memanggil fungsi itu sendiri. Walaupun pohon yang dibangun cukup baik, namun terdapat cara lebih baik untuk meningkatkan efisiensi waktu dan ukuran pohon, dengan tidak mengabaikan akurasi. Hasil analisa didapatkan bahwa penggunaan nilai ambang batas atau *threshold* cenderung memberikan waktu pembangunan tree yang lebih efisiesn, ukuran pohon yang lebih kecil, dan akurasi yang cukup baik, bergantung terhadap karakteristik dari setiap dataset.

Kata Kunci : data mining, classification, ID3