

Abstrak

World Wide Web atau WWW merupakan salah satu fenomena teknologi yang berkembang sangat pesat saat ini. WWW menyediakan berbagai layanan informasi mengenai berita, iklan, pendidikan, e-commerce dan sebagainya. Informasi yang tersedia dalam WWW tersebut memiliki ukuran yang sangat besar dan terdistribusi secara global di seluruh dunia. Web juga mengandung kekayaan informasi dilihat dari struktur dan penggunaannya (*web usage*). Web merupakan kumpulan data dan informasi yang sangat berpotensi untuk dilakukan penggalian (*mining*) agar menghasilkan pengetahuan (*knowledge*) yang dapat berguna bagi masyarakat maupun pihak-pihak tertentu.

Algoritma C5.0 merupakan algoritma untuk mengklasifikasikan dengan menghasilkan *decision tree*. Pemilihan atribut yang akan diproses menggunakan ukuran *information gain*. Ukuran *information gain* digunakan untuk memilih atribut uji pada setiap *node* di dalam *tree*. Atribut dengan nilai *information gain* tertinggi akan terpilih sebagai *parent* bagi *node* selanjutnya. Algoritma ini membentuk pohon keputusan dengan cara pembagian dan menguasai sampel secara rekursif dari atas ke bawah. Untuk memudahkan pengguna informasi dalam menafsirkan terhadap hasil klasifikasi C5.0 disajikan dalam dua bentuk, menggunakan pohon keputusan dan sekumpulan aturan IF-T HEN yang lebih mudah untuk dimengerti.

Berdasarkan hasil analisa yang telah dilakukan dapat diketahui bahwa akurasi untuk *tree* yang digenerate dari data training menghasilkan akurasi kurang baik terhadap data testing, oleh karena itu dilakukan proses *pruning*. Rule yang dihasilkan setelah proses *pruning* memiliki akurasi lebih baik terhadap data testing dan memiliki simplisitas aturan yang rendah, sehingga dihasilkan aturan yang lebih sederhana dibandingkan *tree* sebelumnya. Pola akses dari user yang telah terklasifikasi kurang memberikan perbedaan yang signifikan hal ini disebabkan user pengakses web tersebut memiliki kepentingan terhadap informasi yang hampir sama. Dari hasil klasifikasi ini dapat diperoleh feedback terhadap admin web untuk peningkatan performansi web dalam hal navigasi.

Kata kunci : *web usage, C5.0, decision tree, information gain, pruning*