

## Abstrak

Dengan banyaknya jumlah kategori yang dapat dimiliki oleh sebuah data teks, maka penggolongan secara manual tentu saja akan menjadi suatu masalah baru untuk user. Dengan data teks yang telah terklasifikasi, maka akan memudahkan pencari informasi dalam menemukan data teks yang dibutuhkan karena data teks telah dikelompokkan berdasarkan kategori yang mencerminkan isi data teks tersebut. Dewasa ini, kebanyakan data teks sudah bersifat banyak kategori atau *multi-label*.

*Decision tree* merupakan salah satu metoda klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. *Tree* yang dihasilkan dari pemodelan suatu data kemudian membentuk suatu *rules* atau aturan "jika - maka".

Salah satu algoritma *tree* yang paling terkenal adalah algoritma C4.5. Algoritma C4.5 sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Data yang *multi-label* merupakan sebuah problem pada proses klasifikasi dengan algoritma C4.5, oleh sebab itu akan dilakukan modifikasi *classifier* pada algoritma C4.5 yang sejatinya digunakan untuk menangani proses klasifikasi *single label* sehingga dapat melakukan proses klasifikasi data teks *multi-label*. Modifikasi yang dilakukan adalah melakukan modifikasi rumus untuk menghitung nilai *entropy*.

Hasil yang didapat menunjukkan implementasi modifikasi *entropy* pada algoritma C4.5 untuk data teks *multi-label* dapat dilakukan. Selain itu, performansi yang dihasilkan baik dilihat dari evaluasi hasil akurasi dan waktu.

**Kata kunci:** modifikasi *classifier*, *entropy*, data teks, *multi-label*, performansi