Abstrak

Sebuah website di Internet memiliki banyak konten informasi dalam tiap-tiap blok halaman yang ditampilkan. Kemudian tidak seperti kebanyakan data atau teks konvensional lainnya, suatu halaman web selain mengandung konten utama juga mengandung banyak blok informasi yang tidak berhubungan dengan konten utama misalnya, panel navigasi, *copyright*, *user guide*, *links*, sinopsis suatu berita, berbagai macam iklan dan lain-lain. Dalam hal ini blok-blok informasi yang tidak relevan dengan konten utama pada suatu halaman web disebut sebagai *web pages noise.*

Dalam tugas akhir ini akan digunakan teknik *feature weighting* untuk meningkatkan performansi hasil klasifikasi dengan mendeteksi noise yang ada pada halaman website. Dengan teknik *feature weighting* ini suatu halaman web pertama kali akan dimodelkan dengan pohon struktur *Dokumen Object Model (DOM)* tree dan *Compressed structure tree(CST)* untuk memperoleh struktur umum dan membandingkan blok-blok informasi dalam suatu website. Informasi yang didapatkan digunakan untuk melakukan pengukuran dan mengevaluasi tingkat kepentingan dari masing-masing node yang terbentuk dari *compress struktur tree(CST)*.

Berdasarkan tree yang terbentuk dan tingkat kepentingan dari nilai node yang didapatkan, metoda ini memberikan bobot pada masing-masing individual word *(feature)* pada masing-masing blok kontent. Hasil pembobotan *(weight)* akan digunakan dalam proses web mining.

Kata Kunci : *CST, DOM, deteksi noise, eliminasi noise, web mining.*