

Abstrak

Perkembangan teknologi yang semakin pesat menyebabkan jumlah data dan dokumen, seperti dokumen berita pun semakin besar jumlahnya. Jumlah yang besar tersebut menyebabkan pencari data kesulitan untuk menemukan data yang sesuai dengan kebutuhan mereka. Salah satu kegunaan *Text Mining* yaitu *Text Categorization* adalah salah satu solusi untuk permasalahan tersebut.

Text Categorization adalah sebuah proses klasifikasi dokumen yang tidak terstruktur menjadi kategori-kategori tertentu secara otomatis. Pada *text categorization*, pembobotan *term* adalah langkah penting untuk meningkatkan efektivitas *text categorization*, karena *term* yang berbeda mempunyai tingkat kepentingan yang berbeda pula dalam sebuah dokumen. Sistem yang dibangun bertujuan untuk membandingkan dua metode pembobotan *term* yaitu TF • IDF dan TF • Chi-Square. Setelah mendapatkan hasil pembobotan, maka sistem melakukan klasifikasi terhadap dokumen yang dimasukkan.

Performansi sistem diukur dengan metode SVM melalui akurasi hasil klasifikasi yang dipengaruhi oleh hasil pembobotan dua metode tersebut. Parameter evaluasi yang digunakan adalah akurasi, *recall*, *precision*, dan *F-measure*. Hasil percobaan membuktikan bahwa pembobotan *term* menggunakan TF • Chi-Square memberikan performansi yang lebih baik daripada TF • IDF. Performansi ini juga dipengaruhi oleh besarnya *threshold* dan jumlah dokumen yang digunakan.

Kata kunci : *Text Categorization, Term Weighting, TF • IDF, TF • Chi-Square, recall, precision, F-measure*