

Abstrak

Terdapat banyak sekali dokumen teks yang tersedia dalam bentuk elektronik. Dan semakin banyak dari hari ke hari. Dokumen-dokumen tersebut mewakili sejumlah informasi yang sangat banyak yang dapat diakses dengan sangat mudah. Mencari nilai dalam koleksi yang sangat banyak ini memerlukan semacam organisasi; sebagian besar dari kegiatan mengorganisasi dokumen dapat diotomasi melalui klasifikasi teks. Akurasi dan tingkat pemahaman atas sistem yang digunakan dapat meningkatkan kegunaannya secara signifikan.

Dalam Tugas Akhir ini dilakukan penelitian tentang peningkatan kinerja *Naïve Bayes Classifier* pada klasifikasi *multiclass* dengan menggunakan *error-correcting output coding (ECOC)*, dalam implementasinya untuk mengklasifikasikan dokumen berita ke dalam sebuah kategori kelas berita. *Dataset* yang digunakan adalah berita berbahasa Indonesia dari beberapa website berita.

Hasil penelitian menunjukkan bahwa penggunaan *ECOC* dapat meningkatkan akurasi pengklasifikasian dengan menggunakan *Naïve Bayes Classifier* sampai dengan tingkat akurasi 85% dan dapat dijelaskan mengapa peningkatan ini dapat terjadi. Peningkatan dalam akurasi ini harus dibayar dengan adanya peningkatan dalam lama waktu *training*, yang berbanding lurus dengan panjang *code*. Dalam Tugas Akhir ini, penelitian terhadap pengklasifikasian *ECOC* dilakukan dengan menggunakan panjang *code* yang berbeda-beda, dan dicari yang menghasilkan akurasi paling baik untuk dibandingkan dengan metode *Naïve Bayes* murni.

Kata kunci: *text mining*, klasifikasi teks, *multiclass*, *naïve bayes*, *error-correcting output coding (ECOC)*, berita berbahasa indonesia.