

Abstrak

Di masa sekarang informasi berita sangat mudah didapatkan. Untuk mendapatkan suatu informasi, sekarang masyarakat tidak perlu lagi membeli media cetak karena bermacam-macam artikel berita banyak berbaran di internet. Berbagai portal berita dapat menyajikan berbagai macam kategori mulai dari kategori kriminal, politik, bisnis, olahraga dan lain-lain. Dengan banyaknya jumlah artikel, klasifikasi kategori penting dilakukan karena minat pembaca terhadap kategori artikel berbeda-beda. Apalagi jika dalam satu artikel terdapat banyak kata yang berelasi terhadap beberapa kategori yang berbeda, ini akan menyulitkan dalam pemilihan kategori. Dengan adanya sistem ini diharapkan dapat menemukan kategori terbaik terhadap suatu artikel. Penelitian ini menggunakan 1125 aritikel dari protal berita Kompas yang dibagi menjadi lima kategori berita. Tahapan klasifikasi yang dilakukan biasanya setelah preprocessing adalah proses klasifikasi. Tetapi sebelum melakukan proses klasifikasi, data yang telah preprocessing bisa di pilah dengan berbagai metode feature selection. Metode tersebut diharapkan dapat membuat akurasi meningkat. Dalam Tugas Akhir ini akan menganalisis beberapa metode feature selection. Metode-metode yang digunakan adalah *Collection Frequency*, *Document Frequency*, *Word Frequency*, dan *Information Gain*. Dengan menggunakan metode klasifikasi *Naive Bayes*, *Document Frequency* dan *Word Frequency* menghasilkan akurasi yang paling baik yaitu sebesar 88.6%. Fitur ini mengalami peningkatan sebesar 8 % dibandingkan dengan tidak menggunakan *feature selection*.

Kata kunci : *feature selection*, *document classification*, klasifikasi berita Indonesia