

Abstrak

Meningkatnya penggunaan internet telah memicu pertumbuhan dan pertukaran informasi menjadi jauh lebih pesat dibandingkan era sebelumnya. Volume berita elektronik berbahasa Indonesia semakin bertambah besar dan menyimpan informasi yang berharga di dalamnya. Pengelompokan berita berbahasa Indonesia merupakan salah satu solusi yang dapat digunakan untuk mempermudah mencerna informasi penting yang ada di dalamnya. *Clustering* dapat digunakan untuk membantu menganalisis berita dengan mengelompokkan secara otomatis berita yang memiliki kesamaan.

Pada *text clustering* terdapat suatu permasalahan yaitu adanya fitur – fitur yang berdimensi tinggi. Diperlukan metode *Feature selection* untuk mengurangi dimensi fitur ini. *Feature selection* memiliki kemampuan mengurangi dimensionalitas suatu data sehingga dapat meningkatkan performansi *clustering*. Ada beberapa pendekatan sebagai teknik dari implementasi *feature selection*, salah satunya adalah *filter based feature selection*.

Pada tugas akhir ini, dilakukan analisis perbandingan metode *feature selection* antara *Term contribution* dan *Document Frequency*. Metode-metode *feature selection* tersebut diterapkan secara *filter feature selection*. Pada akhir pengujian, dapat dibuktikan bahwa metode *Term contribution* lebih baik daripada *Document Frequency* karena memperhitungkan frekuensi kemunculan *term* pada suatu dokumen dan jumlah dokumen yang dimiliki *term* tersebut, sehingga *term* yang terpilih adalah *term* yang khas atau bersifat diskriminator. Hal ini dapat meningkatkan performansi *clustering* dokumen berdasarkan *precision* dan *entropy*.

Kata Kunci: *clustering, filter feature selection, Term contribution, Document Frequency.*