

Abstrak

Seringkali, bila diberikan suatu koleksi dokumen, akan muncul kebutuhan untuk mengelompokkan dokumen tersebut ke dalam kluster-kluster berdasarkan tingkat kemiripan isi dokumen yang ada. Untuk jumlah dokumen yang sedikit, proses pengelompokan secara manual masih mungkin untuk dilakukan. Namun, untuk mengelompokkan dokumen dengan jumlah banyak, proses pengelompokan secara manual akan memakan waktu yang lama. Proses otomatisasi menjadi satu solusi yang bisa digunakan untuk mengurangi waktu yang diperlukan untuk melakukan pengelompokan. Oleh karena itu, dibutuhkan suatu metode pengelompokan (*clustering*) yang menghasilkan kelompok (*cluster*) dokumen yang sesuai dengan topik yang ada.

Pada Tugas Akhir ini diimplementasikan metode Non-negative Matrix Factorization (NMF) untuk melakukan proses *clustering* dokumen. Metode ini melakukan faktorisasi pada *term-document matrix* yang merupakan hasil dari proses *preprocessing* dokumen menjadi matriks dua dimensi yang setiap baris mewakili vektor term dan setiap kolom mewakili bobot yang berhubungan dengan term pada baris. Pengujian pada proses dalam proses *clustering* ini dilakukan untuk mengetahui akurasi kualitas cluster yang dihasilkan oleh metode Non-negative Matrix Factorization (NMF) bila dibandingkan dengan cluster awal dari dataset. Selain itu juga diuji pengaruh setiap parameter input terhadap hasil kualitas cluster. Hasil kualitas cluster ditunjukkan dengan nilai *Accuracy*.

Hasil *clustering* dengan metode Non-negative Matrix Factorization (NMF) ini menunjukkan bahwa nilai *Accuracy* cluster setelah proses *clustering* menggunakan metode Non-negative Matrix Factorization (NMF), dipengaruhi oleh jumlah kluster yang diinginkan dan jumlah dokumen yang digunakan. Dimana nilai akurasi tidak akan selalu naik pada jumlah kluster yang berbeda dengan jumlah dokumen yang sama.

Kata Kunci : Klasterisasi dokumen, *Non-negative Matrix Factorization (NMF)*, *Accuracy*