

Abstrak

Masalah yang sering dihadapi dalam masalah *document clustering* adalah menentukan algoritma atau metode yang tepat dengan jumlah dokumen tertentu. Terkadang metode yang dengan hasil yang cukup akurat membutuhkan waktu yang lama untuk pemrosesan. Ada beberapa pendekatan yang dipakai untuk menyelesaikan masalah *clustering* ini, yaitu *clustering* dengan pendekatan partisi dan *clustering* dengan pendekatan hirarki. Masing-masing memiliki kekurangan dan kelebihan. Salah satu dari sekian banyak metode *clustering* yang bisa digunakan adalah *sequential information bottleneck (sIB) method*[5]. Algoritma *sequential information bottleneck* dalam *document clustering* memberikan jaminan ditemukannya solusi yang merupakan lokal maksimum dari fungsi target.

Pada tugas akhir ini dilakukan penerapan Algoritma *sequential information bottleneck* sebagai metode *clustering*. Hasil *clustering* kemudian diukur akurasi *cluster*-nya menggunakan *micro-averaged precision* dan *micro averaged recall* dengan memperhatikan perubahan parameter-parameter masukan yang digunakan seperti *maximum loop (maxL)*, jumlah inisialisasi random cluster, dan nilai *error*. Dari hasil pengujian didapatkan bahwa Algoritma *sequential information bottleneck* sangat baik digunakan sebagai metode *clustering*, hal ini bisa dilihat dari hasil percobaan dimana nilai akurasi yang dicapai rata-rata diatas 70%.

Akurasi *cluster* semakin meningkat seiring dengan peningkatan nilai parameter *maximum loop* hingga sampai pada batas tertentu dimana kondisi perulangan sudah berhenti karena parameter lain (nilai *error*) sudah terpenuhi nilainya. karena hasil dari dokumen yang telah diproses menggunakan *sequential information bottleneck method* memiliki nilai presisi yang sangat tinggi, maka hasil *clustering* tersebut dapat digunakan sebagai *training set* untuk *supervised classification method*[5]. *supervised classification method* yang akan digunakan adalah *naive bayes clasification method*. *Naive bayes clasification method* yang menggunakan hasil *clustering* dengan sIB sebagai *training set* kemudian dibandingkan akurasinya dengan *Naive bayes clasification method yang tidak menggunakan hasil clustering dengan sIB sebagai training*.

Kata Kunci : *Sequential Information Bottleneck Method, Naive bayes clasification method, clustering, klasifikasi.*