

Abstrak

Saat ini jumlah informasi seperti artikel berita yang ada didalam web terus berkembang dengan cepat. Banyaknya jumlah informasi tersebut bisa menyebabkan pengguna mengalami kesulitan dalam mencari informasi yang penting. Salah satu *task Data Mining* yaitu *Text Categorization* yang merupakan suatu cara untuk mengelompokan dokumen kedalam kelas atau kategori tertentu bisa dijadikan sebagai solusi untuk mengorganisasikan dokumen berita.

Salah satu metode *Text Categorization* adalah Metode *Centroid Based Classifier*. Metode *Centroid Based* merepresentasikan dokumen kedalam bentuk vektor. Metode ini membentuk *vector centroid* pada sekumpulan dokumen yang termasuk pada suatu kelas tertentu. *Vector centroid* tersebut yang akan digunakan sebagai model untuk mengelompokan suatu dokumen dengan menggunakan kesamaan kosinus.

Pada tugas akhir ini dilakukan perbandingan performansi antara metode *Centroid Based Classifier* dengan k-NN dan Naïve Bayes. Parameter performansi yang digunakan adalah akurasi dan *f-measure*. Selain itu, tugas akhir ini juga menganalisis deteksi *outlier* untuk meningkatkan akurasi dari metode *Centroid Based*. Hasil percobaan menunjukan bahwa metode *Centroid Based* memberikan performansi yang lebih baik dari pada k-NN dan Naïve Bayes sedangkan percobaan deteksi *outlier* menunjukan bahwa deteksi *outlier* bisa meningkatkan akurasi metode *Centroid Based*.

Kata Kunci : *text categorization, centroid based classifier, centroid vector, akurasi, f-measure.*