

Abstrak

Saat ini teknologi informasi yang semakin berkembang pesat, menyebabkan jumlah dokumen berita *online* maupun *offline* juga bertambah banyak. Karena jumlah dokumen berita yang ada banyak, pencarian terhadap informasi mengenai topik-topik tertentu dalam kumpulan dokumen berita cenderung lebih sulit. Untuk itu diperlukan pengelompokan artikel berita dengan menerapkan salah satu teknik yang terkait dengan *Text Mining*, yaitu *Clustering*.

Clustering merupakan salah satu metode data mining yang bersifat *unsupervised learning* untuk mengelompokkan dokumen berdasarkan kemiripannya. Untuk melakukan pengelompokan tersebut, dalam Tugas Akhir ini digunakan salah satu algoritma hierarki *clustering* yaitu *Divisive Analysis* (DIANA). Algoritma *Divisive Analysis* (DIANA) bersifat *top-down clustering* dimana pada awalnya seluruh objek dikelompokkan pada *cluster* yang sama, objek yang memiliki *dissimilarity* terbesar dengan objek lainnya akan dipisahkan ke dalam suatu *cluster* yang baru, pada objek lainnya akan dilihat bila memiliki kesamaan dengan *cluster* yang baru, maka objek tersebut akan dikelompokkan ke dalam *cluster* tersebut, bila objek tersebut tidak memiliki kesamaan terhadap *cluster* yang baru dan dengan *cluster* semula maka objek tersebut akan dikategorikan ke dalam *cluster* yang baru lagi, begitu seterusnya hingga akhirnya minimal terdapat satu objek tunggal dalam suatu *cluster*.

Dalam Tugas Akhir ini dilakukan pengujian bagaimana *cluster* yang dihasilkan serta evaluasi *cluster* dengan melakukan pengujian terhadap *Divisive Coeficient* (DC), hubungan antara *minimum distance* dengan *F-measure* serta hubungan antara *minimum distance* dengan jumlah *cluster* yang dihasilkan. Dari hasil pengujian yang telah dilakukan dimana jumlah dokumen yang berbeda dan jumlah *term* yang paling kecil memiliki kondisi paling maksimal pada saat nilai *minimum distance* sebesar 7.5 yang memiliki nilai DC sebesar 0.6812 dan nilai *F-measure* total sebesar 0.4154. Sementara pada inputan *dataset* yang memiliki jumlah dokumen sama dengan jumlah *term* sama kondisi paling maksimal pada saat nilai *minimum distance* sebesar 22.5 yang memiliki nilai DC sebesar 0.6323 dan nilai *F-measure* total sebesar 0.4071.

Kata Kunci : *Clustering*, *Divisive Analysis* (DIANA), *Divisive Coeficient* (DC), *F-measure*, *minimum distance*.