

Abstrak

Clustering adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang digunakan. Dalam sebuah dataset berupa kumpulan dokumen berita, *clustering* dapat bermanfaat untuk mengelompokkan berita-berita yang memiliki tingkat kesamaan tinggi sehingga dapat memudahkan user untuk mencari berita-berita sejenis yang diinginkan. Namun, dengan semakin tinggi tingkat kompleksitas data yang digunakan, algoritma *clustering* mulai menunjukkan kesulitan dalam menemukan *cluster* yang baik. Masalah yang umum muncul adalah adanya dokumen *noise* atau *outliers* yang tidak terdeteksi. Untuk mengatasi masalah tersebut sebuah metode berbasis *shared nearest neighbor* akan coba diterapkan dalam buku ini. Metode ini akan membentuk *cluster* secara otomatis dengan melihat pada ketetanggaan antar dokumen dengan dipengaruhi tiga parameter input (yaitu: K, EPS, dan MinPts). Dengan menggunakan dataset pengujian yang diperoleh dari www.reuters.com dan SMART didapatkan hasil pengujian dengan perbandingan parameter *cohesion* dan *separation* terbaik pada saat nilai ketiga parameternya sama atau hampir sama. Namun, secara umum, nilai optimal untuk ketiga parameter ini tidak dapat ditentukan secara pasti.

Kata kunci: *clustering, shared nearest neighbor, noise, parameter, cohesion, separation.*