

Abstrak

Saat ini jumlah dokumen teks berbentuk digital semakin banyak dan beragam. Untuk memudahkan pengambilan informasi yang dibutuhkan dari dokumen teks tersebut, maka perlu dilakukan pengelompokan dokumen sesuai dengan topiknya. Pengelompokan ini dapat dilakukan dengan menggunakan teknik yang terdapat dalam *data mining* yaitu *clustering*. *Clustering* merupakan teknik yang bersifat *unsupervised learning*, yaitu data tidak diberi label sehingga algoritma *clustering* yang akan mengelompokkan dokumen berdasarkan nilai kesamaan. Untuk melakukan pengelompokan dokumen tersebut, digunakan algoritma *committee clustering* yang bekerja dengan cara membangun pusat *cluster* dengan merata-ratakan nilai *feature vector* dari himpunan bagian dari anggota *cluster* yang disebut *committee*, yang akan bertugas untuk menentukan suatu dokumen untuk masuk ke dalam suatu *cluster*. Dengan memilih anggota *committee* secara hati-hati, *feature* dari pusat *cluster* akan cenderung mengarah kepada *class target*[8].

Dalam tugas akhir ini, jumlah *cluster* yang dibentuk oleh algoritma *committee clustering* disesuaikan dengan jumlah kategori dari data yang digunakan dan didapatkan nilai rata-rata *silhouette coefficient* sebesar 0.2296. Dengan demikian, kualitas *cluster* yang dihasilkan bersifat *no structure*.

Kata kunci: algoritma, *committee clustering*, *clustering*, dokumen, *feature*, *silhouette coefficient*.