

Abstrak

Clustering merupakan proses pengelompokan objek ke dalam suatu *cluster*, sehingga objek dalam sebuah *cluster* memiliki kemiripan yang sangat besar dengan objek lain pada *cluster* yang sama tetapi sangat tidak mirip dengan objek pada *cluster* yang lain. *Clustering* telah secara luas diimplementasikan diberbagai bidang seperti *market research*, *pattern recognition*, *data analysis*, *image processing* dan sebagainya. *Clustering* data bertipe *categorical* mendapat perhatian khusus karena tipe data ini tidak bisa dihitung jarak kedekatan antar objeknya. Selain itu banyak algoritma *clustering* membutuhkan waktu proses yang lama sehingga tidak cocok digunakan untuk data berukuran besar. Algoritma NabSqueezer merupakan suatu metode *clustering* data *categorical* yang terlebih dahulu melakukan pembobotan pada setiap *attribute value* dimana untuk *attribute value* yang jarang muncul diberi bobot yang lebih besar. Kemudian pada proses *clusteringnya* NabSqueezer hanya perlu membaca dataset satu kali, dan menentukan suatu data akan bergabung dengan *cluster* yang sudah ada berdasarkan nilai *threshold similarity*. Dari hasil analisa didapatkan bahwa nilai *threshold* yang diinputkan user dapat mempengaruhi akurasi sistem berdasarkan *purity measure*, *cohesion*, dan *separation*. Selain itu didapatkan algoritma NabSqueezer memiliki skalabilitas yang baik terhadap penambahan jumlah data.

Kata Kunci : *clustering, data categorical, nabsqueezer*