

ABSTRAK

Big data menjadi populer bagi penikmat ilmu komputer di seluruh dunia. Begitu banyaknya data dengan jumlah yang besar dan bervariasi akan begitu sulit dalam melakukan analisis data. Dalam pemrosesan *big data mining* tentu akan berbeda dengan *data mining*. *Big data mining* dengan jumlah data sangat besar harus dilakukan secepat mungkin dengan dipecah ke beberapa node (*Parallel computing*) untuk mendapatkan informasi yang bermanfaat. Untuk itu akan dilakukan analisis untuk mengetahui pemrosesan *big data mining*. Analisis dilakukan melalui simulasi dengan konfigurasi *multi-node* yang dimana terdiri dari beberapa komputer yang sudah dikonfigurasi. *Big data mining* pada tugas akhir ini akan menggunakan metode *clustering*. Pada metode *clustering* ini algoritma yang digunakan adalah *K-means*. Hasil dari *k means clustering big data* mendapatkan cluster optimal yaitu $K = 5$ dan waktu pemrosesan bergantung pada *cluster mode*, jumlah *worker* dan spesifikasi *hardware* yang digunakan, jika *hardware* dengan spesifikasi rendah digunakan pada saat *cluster mode* otomatis kemampuan untuk pemrosesan mengikuti *hardware* spesifikasi rendah tersebut.

Kata Kunci : *Big Data, Data Mining, Clustering, K-Means, Cluster Mode*