

## Abstrak

*Big data* menjadi kata yang populer seiring dengan bagaimana dapat menyimpan data dalam jumlah yang besar, melakukan proses serta analisisnya. *Data explosion problem* menjadi bagian penting dalam *big data*, dimana data dengan jumlah besar tidak dapat diolah secara cepat guna memberikan suatu informasi (*Knowledge*). Pemrosesan *big data* diperlukan *data mining* untuk mengolah agar dapat dimanfaatkan salah satunya dalam pencarian informasi. Pada *data mining*, data dapat diproses salah satunya menggunakan metode klasifikasi dengan *naïve bayes*, yaitu suatu metode klasifikasi yang dilakukan berdasarkan probabilitas dari setiap *class variable* inputan. Proses perhitungan klasifikasi *naïve bayes* ini akan menggunakan sistem *parallel computing* dalam melakukan pengolahan data dalam jumlah besar (*big data*) dengan konsep *Resilient Distributed Dataset* (RDD). Hasil perhitungan menggunakan konsep *Resilient Distributed Dataset* membuktikan perhitungan dapat dilakukan secara lebih cepat dibandingkan dengan pengolahan data mining menggunakan *weka* dan dengan akurasi yang lebih baik. Hasil percobaan menunjukkan persentasi akurasi, *recall* dan *precision* terbaik pada kasus *dataset poker hand testing* terletak pada *percentage rasio* 60:40 dengan waktu eksekusi 10,209 detik yang dijalankan pada 3 *worker* dengan spesifikasi *hardware* yang sejenis.

Kata kunci : *data mining, big data, naïve bayes, Resilient Distributed Dataset*