

Abstrak

Big data memproses data dalam jumlah besar secara *parallel computing* (*multi-node*) agar data dalam jumlah besar mampu diproses secepat mungkin. Pemrosesan *big data* diperlukan *data mining*. Sedangkan pada *data mining*, data diproses dengan metode salah satunya yaitu *association* dengan *FP-Growth*. Untuk itu pada penelitian ini dilakukan analisis bagaimana metode pada *data mining* dapat digunakan pada pemrosesan *big data*. Analisis akan dilakukan dengan *tools Apache Spark* dan beberapa komputer yang telah dikonfigurasi dengan menggunakan *library* pada *Apache Spark*, yaitu *FP-Growth*. Pada penelitian ini dilakukan pengujian dengan data dosen, mahasiswa, dan pegawai. Untuk data dosen nilai *minimum support* yang digunakan yaitu 0,2; 0,3; dan 0,4 dengan nilai *minimum confidence* 0,8. Sedangkan untuk data mahasiswa dan pegawai nilai *minimum support* yang digunakan yaitu 0,1; 0,2; dan 0,3 dengan nilai *minimum confidence* 0,5. Dari pengujian tersebut dan pengujian perbandingan perhitungan sistem dengan perhitungan manual, sistem mampu memproses perhitungan *data mining* baik secara *single-node* maupun *multi-node*. Proses yang dihasilkan dari pemrosesan yang dilakukan terdiri dari empat *job*. Pengujian selanjutnya dilakukan dengan *single-node* dan *multi-node* untuk memperoleh perbandingan waktu pemrosesan. Dari hasil pengujian didapatkan waktu pemrosesan dipengaruhi oleh oleh spesifikasi *hardware* dan jumlah *worker*.

Kata kunci : *Data Mining, Big Data, FP-Growth, Apache Spark, Minimum Support, Minimum Confidence*