

Nama: Glenn Claudio Ivan Petrus

NIM: D121171519

## Essay Tugas Akhir

### Parallelisasi Algoritma K-Means untuk Klasterisasi Artikel Berita

Jika saya menguasai pemrograman parallel maka saya akan membuat tugas akhir dengan judul “Parallelisasi Algoritma K-Means untuk Klasterisasi Artikel Berita”. Pesatnya kemajuan teknologi informasi dan web membuatnya lebih mudah untuk menyimpan sejumlah besar informasi tekstual yang dikumpulkan, misalnya artikel berita. Meningkatnya ukuran *dataset* tekstual menimbulkan tantangan besar sehingga sulit bagi informasi tersebut untuk dikategorikan secara efisien. Menggunakan algoritma klasterisasi K-Means untuk secara otomatis menemukan kluster dari *dataset* yang besar. Selanjutnya, algoritma *iterative* K-Means diparalelkan menggunakan teknik Map-Reduced untuk meningkatkan efisiensi komputasi dan dengan demikian dapat memprediksi kluster terbaik secara efektif.

Tugas akhir ini bertujuan menggunakan algoritma K-Means yang diparalelisasi untuk mengkluster artikel berita. K-means adalah metode klasterisasi yang populer digunakan di bidang *data mining*. Untuk bekerja dengan dataset yang besar saya menggunakan K-Means yang diparalelisasi, yang merupakan K-means parallel pada Map-Reduce. Memiliki tahapan pasca pemrosesan yang cepat untuk menghasilkan hasil akhir.

Fokus dengan bagaimana mengurangi ukuran *dataset*. Ide dasarnya dengan mempartisi seluruh *dataset* menjadi beberapa *subset* yang kecil. Setiap *reducer* menjalankan satu K-Means lengkap untuk setiap *subset* dan menghasilkan satu grup *centroid*, jadi K-Means dapat diselesaikan dalam satu Map-Reduce *Job*. *M Reducer* menghasilkan *M* grup *intermediate centroid* yang akan digabungkan menjadi satu *final grup centroid*.

K-means adalah salah satu metode pengelompokan yang paling populer dan berguna untuk *dataset* yang besar. Untuk menghadapi meningkatnya jumlah data, MapReduce dirancang untuk memproses dan menghasilkan dataset besar dengan algoritma paralel dan terdistribusi pada sebuah kluster. Banyak algoritma *single-machine* telah berhasil diadaptasi dan diimplementasikan menggunakan MapReduce untuk mendapatkan *speed-up* yang signifikan.