

## BAB 5

### SIMPULAN DAN SARAN

#### 5.1 Simpulan

Kesimpulan yang dapat ditarik sesuai hasil penelitian yang sudah dilaksanakan pada skripsi ini adalah :

1. Data ini dapat dikerjakan melalui langkah pre-processing seperti pembersihan data. Kemudian dilakukan pencarian k optimal melalui metode elbow. Selanjutnya diclusterkan menggunakan algoritma k-means. Tahapan terakhir yaitu dilakukan evaluasi menggunakan teknik *davies bouldin index* (DBI) dan *calinski harabasz index* (CHI).
2. K yang optimal pada dataset penelitian ini bernilai 3 yang didapatkan dari hasil metode elbow. Sehingga dataset diclusterkan menjadi 3 kelas, yaitu sangat terampil, cukup terampil, dan kurang terampil. Berdasarkan hasil pelabelan data, kelas sangat terampil dilihat dari tingginya presentase terampil teknologi informasi dan computer pada provinsi tersebut, kelas cukup terampil dilihat dari standarnya presentase terampil teknologi informasi dan computer pada provinsi tersebut, dan kelas kurang terampil dilihat dari kecilnya presentase terampil teknologi informasi dan computer pada provinsi tersebut. Hasil performa terbaik yang didapatkan melalui teknik *davies bouldin index* bernilai 0.4681510344138588 dan *calinski harabasz index*(CHI) bernilai 99.68654406812013 dengan  $k=7$ .
3. Dari hasil klasterisasi didapatkan tiga klaster. Klaster nol berisi provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi

Tenggara, Gorontalo, Maluku, Papua Barat. Klaster satu merupakan klaster sangat terampil berisi Kepulauan Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur, Kalimantan Utara. Klaster dua merupakan klaster kurang terampil berisi Provinsi Kalimantan Barat, Sulawesi Barat, Maluku Utara, Papua.

## 5.2 Saran

Berikut merupakan beberapa saran yang diberikan oleh penelitian ini sehingga dapat peningkatan hasil dari penelitian berikutnya :

1. Memakai algoritma lain untuk komparasi performa dengan algoritma k-means yang dikerjakan pada penelitian ini
2. Menggunakan dataset penelitian dalam skala yang lebih besar.
3. Menggunakan beberapa teknik evaluasi clustering yang belum digunakan pada penelitian ini.