

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan pemaparan pada bab II, metodologi penelitian pada bab III dan hasil dari studi kasus pada bab IV, diperoleh kesimpulan bahwa konsep CLARA (*Clustering Large Application*) hampir sama dengan algoritma PAM yaitu:

1. Menentukan jumlah  $k$  klasternya terlebih dahulu dari  $n$  objek
2. Memilih sampel terbaik sebanyak  $40+2k$  sebanyak  $m$  sampel
3. Memilih jarak terkecil dari  $m$  sampel misalnya  $m_j$
4. Pilih medoid sembarang sebanyak  $k$  pada  $m_j$
5. Menghitung jarak dari masing-masing objek ke medoidnya, objek yang berdekatan akan membentuk satu klaster
6. Pilih objek baru sebagai kandidat medoid baru, lalu hitung kembali jaraknya, jika kandidat medoid baru jaraknya lebih kecil maka kandidat medoid baru tersebut menjadi medoid baru
7. Ulangi sampai tidak ada perubahan medoid
8. Hitung jarak setiap objek dalam  $n$  objek ke medoid, jarak yang berdekatan dengan medoid akan membentuk suatu klaster.

Dibandingkan dengan menemukan medoids untuk keseluruhan kumpulan data, CLARA mempertimbangkan sampel kecil dari data dengan ukuran tetap (*sample size*) dan menerapkan algoritma PAM (*Partitioning Around Medoid*) untuk menghasilkan nilai medoid yang optimal untuk masing-masing sampel. Kualitas medoid yang dihasilkan diukur dengan perbedaan jarak rata-rata antara setiap objek di seluruh himpunan data dengan medoid klasternya. Dari sejumlah data yang besar tersebut, CLARA akan memartiskannya ke dalam beberapa sampel acak, lalu dari sampel acak tersebut akan diolah untuk menemukan medoid terbaik menggunakan PAM. Dari beberapa sampel acak yang dilakukan diperoleh nilai *Silhouette Coefficient*. Setelah itu, dari nilai *Silhouette Coefficient* terbesar tersebutlah yang menunjukkan bahwa sampel acak yang diambil tersebut adalah sampel acak terbaik atau teroptimal.

Grachella Yowanda Br Ginting, 2018

CLARA (CLUSTERING LARGE APPLICATION) PADA DATA SIMULASI TRIVARIAT 1000 OBJEK

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Hasil penerapan metode CLARA pada data simulasi trivariat 1000 objek untuk jumlah kluster = 3, diperoleh nilai rata-rata *silhouette width* atau nilai *silhouette coefficient*nya pada ketiga kluster di atas sebesar 0,79. Yang artinya berdasarkan kriteria subjektif pengukuran pengelompokan oleh Kaufmann dan Rousseeuw pengujian penelitian ini berada pada kriteria pengukuran pengelompokan struktur kluster kuat. Sehingga, pengelompokan data objek dalam suatu kluster tersebut sudah baik dengan jarak rata-rata antar objek dalam satu kluster sangat dekat atau sangat kecil dan jarak objek antar kluster sudah saling berjauhan atau sangat besar. Seluruh hasil dalam skripsi ini diolah menggunakan program R.

## 5.2 Saran

Selain menggunakan data simulasi trivariat, para peneliti yang tertarik dapat mengkaji analisis kluster non-hirarki dengan metode CLARA (*Clustering Large Application*) dapat menggunakan data asli atau primer dan data sekunder.