

## Implementasi Algoritma *K-Medoids* dan *K-Means* untuk Pengelompokan Wilayah Sebaran Cacat pada Anak

Dini Marlina<sup>1</sup>, Nurelina Fauzer Putri<sup>2</sup>, Andri Fernando<sup>3</sup>, Aditya Ramadhan<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi

Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru 28293

<sup>1</sup>dinimarlinacosco@gmail.com, <sup>2</sup>fauz3rlina@gmail.com, <sup>3</sup>andrifernando61@gmail.com, <sup>4</sup>adhikha@gmail.com

**Abstrak** – Usia dibawah 18 tahun merupakan usia yang baik dalam pertumbuhan dan perkembangan fisik dan mental pada seseorang. Pertumbuhan dan perkembangan yang baik akan menjadi modal bagi kelangsungan anak sebagai generasi penerus yang baik. Namun, seorang anak yang dilahirkan dalam keadaan cacat fisik yang berat beresiko untuk mengalami stress dan hambatan penyesuaian. Dinas Sosial Provinsi Riau mengaku fasilitas yang diberikan kepada penyandang cacat masih rendah. Selain itu, angka penyandang cacat di Provinsi Riau lebih dari 11 ribu tersebar di seluruh Kabupaten/ Kota di Provinsi Riau. *K-Medoids* mampu melakukan pengelompokan pada data sebaran anak cacat yang ada pada Provinsi Riau. Klaster yang dihasilkan pada penelitian ini adalah berjumlah tiga klaster. Validitas yang digunakan pada penelitian ini adalah validitas Silhouette Coefficient Adapun nilai validitas yang dihasilkan pada algoritma *K-Medoids* adalah sebesar 0.5009. Sedangkan nilai validitas yang dihasilkan pada algoritma *K-Means* adalah 0.1443. Hal ini menunjukkan bahwa algoritma *K-Medoids* lebih baik dalam melakukan pengelompokan pada data sebaran Anak Cacat dibandingkan dengan algoritma *K-Means*.

**Kata Kunci** – *Cacat, Clustering, Data Mining, K-Medoids, Validasi*

### PENDAHULUAN

Usia dibawah 18 tahun merupakan usia yang baik dalam pertumbuhan dan perkembangan fisik dan mental pada seseorang. Pertumbuhan dan perkembangan yang baik akan menjadi modal bagi kelangsungan anak sebagai generasi penerus yang baik. Namun, seorang anak yang dilahirkan dalam keadaan cacat fisik yang berat beresiko untuk mengalami stress dan hambatan penyesuaian. Kecacatan sering mengakibatkan masalah-masalah sosial, seperti penolakan oleh lingkungan sosialnya, kesulitan dalam membina hubungan sosial, dan sikap *over*-proteksi dari orang lain[1].

Penyandang cacat adalah setiap orang yang mempunyai kelainan fisik dan/ atau mental, yang dapat mengganggu atau merupakan rintangan dan

hambatan baginya untuk melakukan secara selayaknya[2].Dinas Sosial Provinsi Riau mengaku fasilitas yang diberikan kepada penyandang cacat masih rendah. Selain itu, angka penyandang cacat di Provinsi Riau lebih dari 11 ribu tersebar di seluruh Kabupaten/ Kota di Provinsi Riau [3]

Dampak penolakan lingkungan dan masalah-masalah sosial lainnya yang menimpa penyandang cacat mendorong berbagai pihak untuk melakukan pencegahan secara dini. Mengingat pertumbuhan dan perkembangan anak dapat terganggu saat usia dibawah 18 tahun. Salah satu langkah awal yang dapat dilakukan yaitu dengan memprakirakan wilayah-wilayah penyandang cacat yang tinggi. Pemanfaatan data sebaran anak cacat dapat dilakukan untuk proses pengelompokan sesuai dengan informasi yang dimiliki oleh data, sehingga dapat diketahui wilayah mana yang memiliki sebaran cacat yang tinggi. Proses pengelompokan dapat dilakukan dengan mengimplementasikan metode *clustering*.

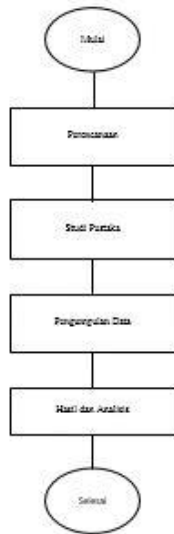
Ada beberapa teknik untuk melakukan klusterisasi, diantaranya *K-Means*, *K-Medoids* dan lainnya. Teknik *K-Means* diperkenalkan oleh MacQueen (1967), sedangkan teknik *K-Medoids* (*Partitioning Around Medoids*) diperkenalkan oleh Kaufman dan Rousseeuw (1990)[4].Berdasarkan penelitian yang dilakukan oleh (Aishwaya) didapati kesimpulan bahwa algoritma *K-Means* efisien untuk kumpulan data yang lebih kecil dan *K-Medoids* nampaknya berperforma lebih baik untuk dataset besar[5].

Berdasarkan permasalahan dan paparan penelitian sebelumnya, penulis membuat sebuah penelitian berjudul “Implementasi Algoritma *K-Medoids* untuk Pengelompokan Wilayah Sebaran Cacat pada Anak”. Penelitian ini dapat menjadi sebuah solusi agar memudahkan dalam mengetahui wilayah-wilayah yang memiliki tingkat anak cacat yang tinggi sehingga dapat dilakukan penanggulangan secara dini.

### METODE PENELITIAN

Metodologi penelitian menjelaskan langkah-langkah yang akan digunakan serta perancangan dalam melakukan implementasi metode *K-Medoids*

*clustering* untuk pengelompokan data wilayah sebaran cacar berdasarkan jumlah penderita dibawah 18 tahun. Berikut adalah metodologi penelitian pada penelitian ini:



Gambar 1. Metodologi penelitian

### Cacat

Cacat adalah setiap orang yang memiliki kelainan fisik dan/atau mental, yang dapat mengganggu atau merupakan rintangan dan hambatan baginya untuk melakukan secara selayaknya, yang terdiri dari [6]:

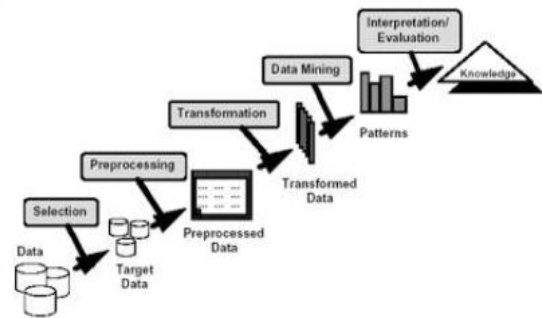
- a. Penyandang cacat fisik.
- b. Penyandang cacat mental.
- c. Penyandang cacat fisik dan mental.

### Knowledge Discovery in Database (KDD)

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut: [7]

1. Data Selection, pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing / Cleaning*, sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang *duplikasi data*, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*typografi*).

3. *Transformation Coding* adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis ataupun informasi yang akan dicari dalam basis data.
4. *Data mining*, *data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation / Evaluation*, pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya



Gambar 2. Proses Data Mining Dalam Penemuan Pengetahuan Dalam Database [8]

### Clustering

*Clustering* merupakan suatu proses pengelompokan data, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek [9]. Berbeda dengan proses klasifikasi, *clustering* tidak mempunyai target variable dalam melakukan. *Clustering* sering dilakukan sebagai langkah awal dalam proses data mining. Terdapat banyak algoritma klustering yang telah digunakan oleh peneliti sebelumnya seperti *K-Means*, *Improved K-Means*, *K-Medoids (PAM)*, *Fuzzy C-Means*, *DBSCAN*, *CLARANS* dan *Fuzzy Subtractive*.

*Clustering* telah digunakan secara luas dan pentingnya pengelompokan tumbuh dengan cepat dikarenakan jumlah data yang berhubungan dengan eksponen aljabar dalam kecepatan pengolahan komputer sangat banyak [10]. Algoritma *clustering* berfungsi untuk mengelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok walaupun setiap

algoritma *clustering* memiliki kelebihan dan kekurangan masing-masing.

**Normalisasi data**

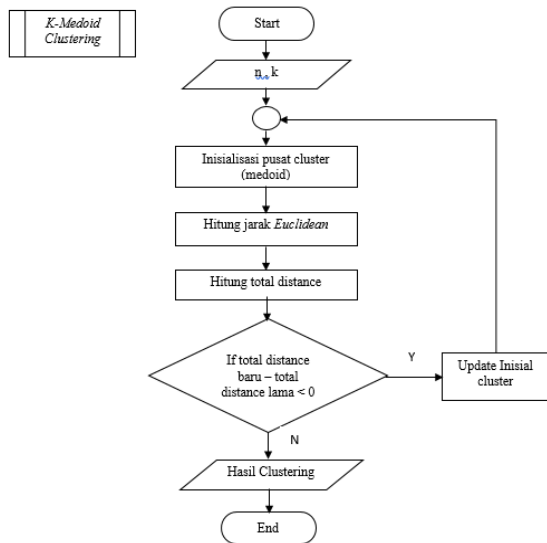
Normalisasi adalah proses transformasi untuk merubah nilai data. Normalisasi digunakan untuk menyamakan skala atribut data kedalam *range* yang spesifik yang lebih kecil seperti -1 sampai 1 atau 0 sampai 1. *min-Max Normalization* merupakan teknik normalisasi dengan melakukan transformasi linear pada atribut data asli untuk menghasilkan *range* nilai yang sama [11]. *Min-Max Normalization* memetakan sebuah value  $v$  dari atribut A menjadi  $v'$  kedalam *range*  $[new\_min_A, new\_Max_A]$  dengan persamaan 1.

$$v' = \frac{v - min_A}{max_A - min_A} (new\ max_A - new\ min_A) + new\_min_A \quad (1)$$

**K-Medoids**

*K-Medoids* adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (*medoids*) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster*. Metode *K-medoids* lebih cocok untuk mengelompokkan data dibandingkan metode *K-Means* [12].

Adapun alur algoritma dari *K-Medoid* adalah sebagai berikut:



Gambar 3. Alur algoritma *K-Medoids*

Langkah-langkah algoritma *K-Medoids*:

1. Inisialisasi pusat *cluster* sebanyak  $k$  (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:  

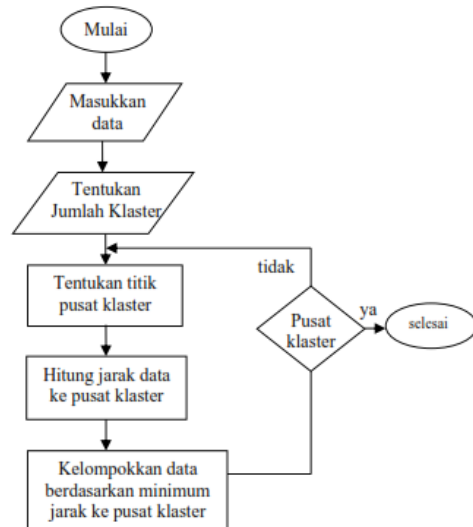
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$
3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.

4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
5. Hitung total simpangan ( $S$ ) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan  $k$  objek baru sebagai *medoid*.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

**K-Means**

*K-means* adalah algoritma untuk pelatihan *unsupervise*, pertama kali dipublikasikan oleh Stuart Loyd pada tahun 1984 dan merupakan algoritma *clustering* yang banyak digunakan. Algoritmanya cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Prinsip utama dari teknik ini adalah menyusun kbuah partisi/pusat (*centroid*)/ rata-rata (*mean*) dari sekumpulan data. Algoritma *K-means* dimulai dengan pembentukan partisi *klaster* di awal kemudian secara iteratif partisi *klaster* ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi *klaster* [13].

Adapun alur algoritma dari *K-Means* adalah sebagai berikut:



Gambar 4. Alur algoritma *K-Means*[16]

Data *clustering* menggunakan algoritma *K-Means* secara umum dilakukan dengan algoritma dasar sebagai berikut: [14]

1. Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara *random*
3. Hitung *centroid*/ rata-rata dari data yang ada di masing-masing *cluster*
4. Alokasikan masing-masing data ke *centroid*/ rata-rata terdekat

- Kembali ke Step 3, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan

### Silhouette Coefficient Index (SI)

*Silhouette Coefficient Index* adalah salah satu metode analisa untuk mendapatkan nilai validasi pada sebuah metode clustering. Hasil perhitungan nilai *silhouette coefficient* dapat bervariasi antara -1 hingga 1. Jika  $si = 1$  berarti objek  $i$  sudah berada dalam cluster yang tepat. Jika nilai  $si = 0$  maka objek  $i$  berada di antara dua cluster sehingga objek tersebut tidak jelas harus dimasukkan ke dalam cluster A atau cluster B. Akan tetapi, jika  $si = -1$  artinya struktur *cluster* yang dihasilkan *overlapping*, sehingga objek  $i$  lebih tepat dimasukkan ke dalam *cluster* yang lain [15].

## HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan bagaimana mengimplementasikan algoritma K-Medoidss pada data Sebaran Anak Cacat yang terdapat pada Provinsi Riau. Sebelum dilakukan perhitungan, Hasil dari metode K-Medoids yaitu mengetahui pengelompokan desa berdasarkan jumlah penyandang cacat. Adapun atribut yang digunakan terdapat pada Tabel 1.

Tabel 1. Atribut yang Digunakan

Atribut	Inisialisasi
Desa	Desa
Tidak cacat <5%	X1
Tidak cacat 6%-10%	X2
Tidak cacat 11%-20%	X3
Tidak cacat 21%-30%	X4
Tidak cacat 31%-40%	X5
Cacat (Semua Fisik)	X6
Tuna Daksa	X7
Tuna Netra	X8
Tuna Rungu	X9
Tuna Wicara	X10
Cacat Mental	X11
Mantan Penderita Gangguan Jiwa	X12

Data yang digunakan dalam penelitian ini adalah data Sebaran Anak Cacat yang diunduh dari data.go.id. Adapun data yang digunakan adalah data desa yang berada pada Provinsi Riau. Data yang diolah dapat dilihat pada Tabel 2.

Sebelum dilakukan pengolahan data menggunakan Algoritma K-Medoids, lakukan proses normalisasi. Adapun hasil normalisasi data sebaran cacat dapat dilihat pada Tabel 3.

Percobaan perhitungan merujuk pada lima atribut jumlah penyandang cacat setiap desa, yaitu Tidak cacat <5%, Tidak cacat 6%-10%, Tidak cacat

11%-20%, Tidak cacat 21%-30%, Tidak cacat 31%-40% dan cacat (Semua Fisik).

### Perhitungan K-Medoids

Algoritma *K-Medoids* dimulai pada penentuan awal pusat klaster dengan memilih secara acak di antara objek yang ada pada dataset. Jumlah objek yang dipilih sesuai dengan jumlah cluster yang diinginkan, hal ini dikarenakan objek yang terpilih akan merepresentasikan pusat dari cluster sebagai *medoids*, untuk kasus ini akan dipilih tiga buah objek sebagai *medoids* awal. *Medoids* awal dapat dilihat pada Tabel 4.

Langkah selanjutnya adalah menghitung jarak setiap seluruh objek *non-medoids*. Objek dengan jarak terdekat dengan masing-masing *medoids* akan bergabung dengan *medoids* tersebut. Objek yang paling dekat dengan *medoids* 1 (C1) dapat dilihat pada Tabel 5. Objek yang paling dekat dengan *medoids* 2 (C2) dapat dilihat pada Tabel 6. Objek yang paling dekat dengan *medoids* 3 (C3) dapat dilihat pada Tabel 7.

Selanjutnya mencari nilai *cost* dengan melakukan perhitungan jarak antara objek dengan setiap *medoids* dengan menggunakan metoda *Eucleadean Distance*. Adapun rumus *Eucleadian Distance* adalah sebagai berikut:

$$d(C1, O2) = \sqrt{(0.0018 - 0.0108)^2 + (0.0051 - 0.0255)^2 + \dots + (0.0581 - 0.0000)^2} = 0.0492$$

Lakukan perhitungan jarak pada seluruh data yang mendekati *medoids* 1 dan sehingga didapatkan jumlah *cost medoids* 1 ke objek data yang mendekati *medoids* 1.

$$\sum S(C1, Oi) = d(C1, O2) + d(C1, O4) + \dots + d(C1, O1636)$$

$$\sum S(C1, Oi) = 0.0492 + 0.0857 + \dots + 0.0651 = 32.1567$$

Lakukan perhitungan jarak pada seluruh data yang mendekati *medoids* lainnya. Adapun jumlah *cost* pada *medoids* 2 dan 3 yaitu:

$$\sum S(C2, Oi) = 0.1047 + 0.1390 + \dots + 0.0774 = 26.5401$$

$$\sum S(C3, Oi) = 0.0751 + 0.1431 + \dots + 0.4091 = 119.6672$$

Setelah dihitung jumlah *cost* pada data masing-masing *medoids*, maka didapatkan jumlah seluruh *cost* yang ada pada percobaan pertama.

$$\sum S(Ci, Oi) = 32.1567 + 26.5401 + 119.6672 = 178.3640$$

Setelah nilai *cost* pada percobaan pertama didapatkan, lakukan hal yang sama pada percobaan kedua dengan melakukan pemilihan *medoids*

kembali secara acak, lakukan perubahan *medoids* yang ada pada data. Setelah menentukan *medoids*, hitunglah jarak setiap seluruh objek non-*medoids*. Objek dengan jarak terdekat dengan masing-masing *medoids* akan bergabung dengan *medoids* tersebut. Kemudian mencari nilai *cost* dengan melakukan perhitungan jarak antara objek dengan setiap *medoids* dengan menggunakan metoda *Eucleadean Distance*. Sehingga pada percobaan kedua didapatkan nilai jumlah *cost* sebagai berikut:

$$\sum S(C_i, O_i) = 49.0599 + 7.9079 + 185.6273 = 185.6273$$

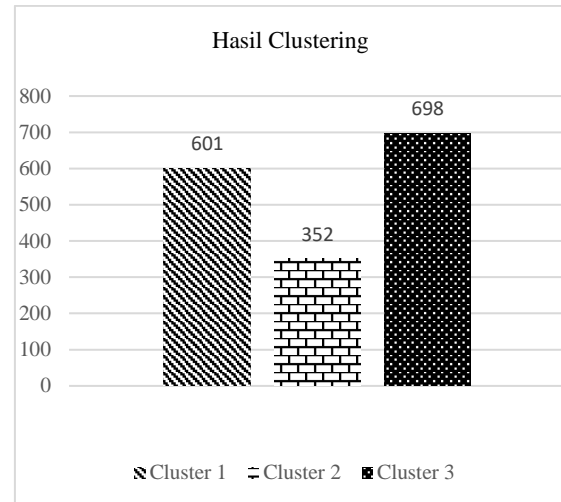
Perhitungan yang dihasilkan dari dua kali percobaan dapat dilihat pada Tabel 8. Tabel 8 menjelaskan bahwa terdapat perubahan jumlah *cost* dari 178.3640 menjadi 185.6273 maka nilai  $\sum S(C_i, O_i)$  pertama <  $\sum S(C_i, O_i)$  kedua, oleh karena itu algoritma berhenti pada tahap ini dan tidak ada perubahan susunan cluster pada percobaan pertama.

Data pada kluster ke 1 dapat dilihat pada Tabel 9. Jumlah data yang terdapat pada Kluster 1 adalah 601 data. Pola yang terbentuk pada kluster 1 adalah jumlah tidak cacat <5% antara 0 hingga 64. Sedangkan untuk jumlah cacat (semua fisik) antara 0-5. Sementara jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 3, 2, 1, 2, dan 3.

Data pada kluster kedua dapat dilihat pada Tabel 10. Jumlah data yang terdapat pada kluster kedua adalah 352 data. Pola yang terbentuk pada kluster 2 adalah jumlah tidak cacat <5% antara 0 hingga 164. Sedangkan untuk jumlah cacat (semua fisik) antara 0-8. Sementara jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 1, 3, 2, dan 3.

Data pada kluster ketiga dapat dilihat pada Tabel 11. Jumlah data yang terdapat pada kluster ketiga adalah 698 data. Pola yang terbentuk pada kluster 3 adalah jumlah tidak cacat <5% antara 0 hingga 1116. Sedangkan untuk jumlah cacat (semua fisik) antara 0-15. Sementara jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 5, 3, 4, dan 5.

Grafik hasil kluster pada data sebaran cacat pada anak dapat dilihat pada Gambar 5.



Gambar 5. hasil kluster data sebaran cacat pada anak

### Perhitungan K-Means

Perhitungan juga dilakukan pada Algoritma *K-Means* untuk mengetahui perbedaan pola yang terbentuk. Dalam algoritma *K-Means*, *cluster 1* yang terbentuk dapat dilihat pada Tabel 12, *cluster 2* yang terbentuk dapat dilihat pada Tabel 13, *cluster 3* yang terbentuk dapat dilihat pada Tabel 14

### Validasi K-Medoids dan K-Means

Perhitungan *Silhouette Coefficient Index* (SI) digunakan untuk mengetahui kinerja tiap algoritma *clustering*. Adapun hasil perhitungan validasi *Silhouette* pada metode *K-Medoids* dapat dilihat pada Tabel 15.

Rata-rata *Silhouette Coefficient* pada *K-Medoids* adalah 0.5009, ini menunjukkan bahwa metode *K-Medoids* cukup baik dalam melakukan Pengelompokan pada data sebaran cacat pada Anak di wilayah Provinsi Riau.

Sedangkan pada Algoritma *K-Means*, nilai yang dihasilkan pada validasi *Silhouette Coefficient* dapat dilihat pada Tabel 16.

Rata-rata *Silhouette Coefficient* pada *K-Means* adalah 0.1443, ini menunjukkan bahwa metode *K-Means* cukup baik. Namun jika dibandingkan dengan algoritma *K-Medoids*, algoritma *K-Means* mempunyai nilai validitas yang lebih rendah dalam melakukan pengelompokan pada data sebaran cacat pada anak di wilayah Provinsi Riau.

Tabel 2. Data yang Digunakan

No	Desa	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	Sungai Besar	51	86	94	70	17	3	1	0	0	1	0	0
2	Ibul	12	20	34	34	5	0	0	0	0	0	0	0
3	Pangkalan	49	56	68	47	2	2	0	0	1	0	0	0
4	Muara Petai	35	17	35	21	4	1	0	0	0	0	1	0
5	Pantai	22	30	39	14	7	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1656	Bintan	25	14	148	238	156	4	1	0	0	0	3	0

Tabel 3. *Medoids* secara acak

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
1	Sungai Besar	0.0457	0.1096	0.0421	0.0390	0.0494	0.2000	0.1429	...	0.0000
2	Ibul	0.0108	0.0255	0.0152	0.0189	0.0145	0.0000	0.0000	...	0.0000
3	Pangkalan	0.0439	0.0713	0.0305	0.0262	0.0058	0.1333	0.0000	...	0.0000
4	Muara Petai	0.0314	0.0217	0.0157	0.0117	0.0116	0.0667	0.0000	...	0.0000
5	Pantai	0.0197	0.0382	0.0175	0.0078	0.0203	0.0000	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1656	Bintan	0.0224	0.0178	0.0663	0.1324	0.4535	0.2267	0.1429	...	...

Tabel 4. *Medoids* secara acak

No	Nama Desa	X1	X2	X3	X4	X5	X6	X7	...	X12
411	Kuala Keritang	0.0018	0.0051	0.0166	0.0234	0.0581	0.0000	0.0000	...	0.0000
828	Sungai Tengah	0.0054	0.0204	0.0143	0.0256	0.0872	0.0000	0.0000	...	0.0000
1242	Harapan Baru	0.0573	0.0471	0.0708	0.0668	0.0523	0.0667	0.0000	...	0.0000

Tabel 5. Objek yang mendekati *medoids* 1

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
2	Ibul	0.0108	0.0255	0.0152	0.0189	0.0145	0.0000	0.0000	...	0.0000
4	Muara Petai	0.0314	0.0217	0.0157	0.0117	0.0116	0.0667	0.0000	...	0.0000
5	Pantai	0.0197	0.0382	0.0175	0.0078	0.0203	0.0000	0.0000	...	0.0000
8	Koto Cengar	0.0251	0.0229	0.0076	0.0061	0.0000	0.0000	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1636	Batu Teritip	0.0152	0.0038	0.0282	0.0000	0.0000	0.0000	0.0000	...	0.0000

Tabel 6. Objek yang mendekati *medoids* 2

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
3	Pangkalan	0.0439	0.0713	0.0305	0.0262	0.0058	0.1333	0.0000	...	0.0000
6	Air Buluh	0.0968	0.0777	0.0125	0.0017	0.0029	0.3333	0.2857	...	0.0000
7	Lubuk Ramo	0.0439	0.0331	0.0296	0.0195	0.0174	0.0000	0.0000	...	0.0000
9	Seberang Cengar	0.0367	0.0522	0.0219	0.0128	0.0116	0.0667	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1623	Tebing Tinggi Okura	0.0421	0.0433	0.0511	0.0306	0.0349	0.2667	0.1429	...	0.0000

Tabel 7. Objek yang mendekati *medoids* 3

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	x8	x12
1	Sungai Besar	0.0457	0.1096	0.0421	0.0390	0.0494	0.2000	0.1429	0.0000	0.0000
57	Muara Lembu	0.1308	0.1516	0.0287	0.0273	0.0233	0.2667	0.4286	0.0000	0.0000
66	Sungai Kuning	0.1263	0.1006	0.0143	0.0161	0.0000	0.0000	0.0000	0.0000	0.0000
69	Sungai Buluh	0.1102	0.1210	0.0578	0.0356	0.0436	0.1333	0.1429	0.0000	0.0000
...	...	...	...	...	...	...	...	...	...	...
1656	Bintan	0.0224	0.0178	0.0663	0.1324	0.4535	0.2667	0.1429	...	0.0000

Tabel 8. Hasil *cost* pada Kedua Percobaan

Cost	Hasil
$\sum S(C_i, O_i)_{\text{pertama}}$	178.3640
$\sum S(C_i, O_i)_{\text{kedua}}$	185.6273

Tabel 9. Data pada Klaster 1

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
2	Ibul	0.0108	0.0255	0.0152	0.0189	0.0145	0.0000	0.0000	...	0.0000
4	Muara Petai	0.0314	0.0217	0.0157	0.0117	0.0116	0.0667	0.0000	...	0.0000
5	Pantai	0.0197	0.0382	0.0175	0.0078	0.0203	0.0000	0.0000	...	0.0000
8	Koto Cengar	0.0251	0.0229	0.0076	0.0061	0.0000	0.0000	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1636	Batu Teritip	0.0152	0.0038	0.0282	0.0000	0.0000	0.0000	0.0000	...	0.0000

Tabel 10. Data pada Klaster 2

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
3	Pangkalan	0.0439	0.0713	0.0305	0.0262	0.0058	0.1333	0.0000	...	0.0000
6	Air Buluh	0.0968	0.0777	0.0125	0.0017	0.0029	0.3333	0.2857	...	0.0000
7	Lubuk Ramo	0.0439	0.0331	0.0296	0.0195	0.0174	0.0000	0.0000	...	0.0000
9	Seberang Cengar	0.0367	0.0522	0.0219	0.0128	0.0116	0.0667	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1623	Tebing Tinggi Okura	0.0421	0.0433	0.0511	0.0306	0.0349	0.2667	0.1429	...	0.0000

Tabel 11. Objek yang mendekati medoids 3

No	Nama Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
1	Sungai Besar	0.0457	0.1096	0.0421	0.0390	0.0494	0.2000	0.1429	...	0.0000
57	Muara Lembu	0.1308	0.1516	0.0287	0.0273	0.0233	0.2667	0.4286	...	0.0000
66	Sungai Kuning	0.1263	0.1006	0.0143	0.0161	0.0000	0.0000	0.0000	...	0.0000
69	Sungai Buluh	0.1102	0.1210	0.0578	0.0356	0.0436	0.1333	0.1429	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1656	Bintan	0.0224	0.0178	0.0663	0.1324	0.4535	0.2667	0.1429	...	0.0000

Tabel 12. Cluster 1 pada Algoritma K-Means

No	Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
232	Bukit Meranti	0.0260	0.0484	0.0363	0.0384	0.1860	0.2667	0.0000	...	0.0000
252	Aur Cina	0.1057	0.1134	0.0305	0.0095	0.0233	0.4000	0.1429	...	0.0000
430	Pulau Kijang	0.4337	0.2981	0.2360	0.2065	0.2820	0.3333	0.1429	...	0.0000
481	Sungai Perak	0.1953	0.1618	0.1594	0.0729	0.0465	0.4667	1.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1656	Bintan	0.0224	0.0178	0.0663	0.1324	0.4535	0.2667	0.1429	...	0.0000

Tabel 13. Cluster 2 pada Algoritma K-Means

No	Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
2	IBUL	0.0108	0.0255	0.0152	0.0189	0.0145	0.0000	0.0000	...	0.0000
3	PANGKALAN	0.0439	0.0713	0.0305	0.0262	0.0058	0.1333	0.0000	...	0.0000
4	MUARA PETAI	0.0314	0.0217	0.0157	0.0117	0.0116	0.0667	0.0000	...	0.0000
5	PANTAI	0.0197	0.0382	0.0175	0.0078	0.0203	0.0000	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1653	BULUH KASAP	0.0099	0.0306	0.0394	0.1041	0.2849	0.0000	0.0000	...	0.0000

Tabel 14. Cluster 3 pada Algoritma K-Means

No	Desa	x1	x2	x3	x4	x5	x6	x7	...	x12
1	Sungai Besar	0.0457	0.1096	0.0421	0.0390	0.0494	0.2000	0.1429	...	0.0000
28	Perhentian Sungkai	0.0206	0.0611	0.0309	0.0150	0.0087	0.1333	0.0000	...	0.0000
6	Air Buluh	0.0968	0.0777	0.0125	0.0017	0.0029	0.3333	0.2857	...	0.0000
31	Sungai Kelelawar	0.0242	0.0229	0.0090	0.0033	0.0087	0.0667	0.0000	...	0.0000
...	...	...	...	...	...	...	...	...	...	...
1651	Sukajadi	0.0502	0.0357	0.0596	0.1269	0.5610	0.3333	0.0000	...	0.0000

Tabel 15. Hasil Validasi Silhouette algoritma K-Medoids

Cluster	Rata-rata Jarak pada Cluster	Rata-rata Jarak pada Luar Cluster	Silhouette
1	0.053059	0.30076	0.823583
2	0.075398	0.323903	0.76722
3	0.171443	0.156363	-0.08796

Tabel 16. Hasil Validasi Silhouette algoritma K-Means

Cluster	Rata-rata Jarak pada Cluster	Rata-rata Jarak pada Luar Cluster	Silhouette
1	0.6671	0.2245	-0.6635
2	0.1219	0.5279	0.7691
3	0.3113	0.4628	0.3274

## KESIMPULAN DAN SARAN

*K-Medoids* mampu melakukan pengelompokan pada data sebaran anak cacat yang ada pada Provinsi Riau. Kluster yang dihasilkan pada penelitian ini adalah berjumlah tiga kluster. Kluster pertama mempunyai pola dengan jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 3, 2, 1, 2, dan 3. Hal ini menunjukkan bahwa pada kluster pertama memiliki data dengan nilai yang paling rendah dibandingkan kluster lainnya. Pada kluster kedua jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 1, 3, 2, dan 3. Hal ini menunjukkan kluster kedua mempunyai data

tingkat sebaran anak cacat yang lebih tinggi dibandingkan pada kluster yang pertama. Sedangkan pada kluster yang ketiga, Pola yang terbentuk pada kluster 3 adalah jumlah cacat (semua fisik) antara 0-15. Sementara jumlah penyandang Tuna Daksa, Tuna Netra, Tuna Rungu, Tuna Wicara, dan Cacat Mental masing-masing adalah 7, 5, 3, 4, dan 5. Hal ini menunjukkan kluster ketiga mempunyai data dengan nilai cacat yang paling tinggi dibanding kluster lainnya. Adapun hasil validasi yang dihasilkan adalah 0.5009, ini menunjukkan bahwa Algoritma *K-Medoids* cukup baik dalam melakukan Pengelompokan pada data sebaran cacat pada Anak di wilayah Provinsi Riau.

Algoritma *K-Medoids* dan *K-Means* dibandingkan dengan melihat nilai validitasnya. Validitas yang digunakan pada penelitian ini adalah

validitas *Silhouette Coefficient* Adapun nilai validitas yang dihasilkan pada algoritma *K-Medoids* adalah sebesar 0.5009. Sedangkan nilai validitas yang dihasilkan pada algoritma *K-Means* adalah 0.1443. Hal ini menunjukkan bahwa algoritma *K-Medoids* lebih baik dalam melakukan pengelompokan pada data sebaran Anak Cacat dibandingkan dengan algoritma *K-Means*.

#### UCAPAN TERIMA KASIH

Terima kasih yang sebesar-besarnya kepada Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau yang telah memberikan dukungan dan fasilitas dari pimpinan. Demikian juga ucapan terima kasih kepada Bapak Mustakim, S.T., M.Kom selaku dosen pembimbing matakuliah Data Mining, sehingga penelitian ini dapat diselesaikan dengan baik.

#### REFERENSI

- [1] Fizdalifar, M. Gengki. Tingkat Kecemasan Sosial pada Anak yang Mengalami Cacat Fisik di YPAC. ISBN 978-979-796-324-8. 2015
- [2] UU 4/1997, Penyandang cacat pada Bab 1 Pasal 1 ayat 1
- [3] Republika.co.id,jum,at, 6 Desember 2017
- [4] Dr. T. Velmurugan. Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points. ISSN: 2229-6093. Vol 3. 2012
- [5] Chrisnanto, Yulison Herry Dan Abdillah, Gunawan. Penerapan Algoritma *Partitioning Around Medoids* (PAM) Clustering untuk Melihat Gambaran Umum Kemampuan Akademik Mahasiswa. Seminar Nasional Teknologi Informasi dan Komunikasi 2015 (SENTIKA 2015). 28 Maret 2015
- [6] UU 4/1997, Penyandang cacat pada Bab 1 Pasal 1 ayat 1
- [7] Batra. Aishwarya. Analysis and Approach: K-Means and K-Medoids Data Mining Algorithms. 5th IEEE International Conference on Advanced Computing & Communication Technologies. 2011
- [8] Prasmesti, Dyang Falila, Furqon, M. Tanzil, Dewi, Candra. Implementasi Metode *K-Medoids Clustering* untuk Pengelompokan Data Potensi Kebakaran Hutan/ Lahan Berdasarkan Persebaran Titik Panas (*Hotspot*). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 1, No. 9, Juni 2017, hlm. 723-732
- [9] Junaedi, Hartanto., Herman, Budianto, 2011. Data Transformation Pada Data Mining. Prosiding Konferensi nasional "Inovasi dalam Desain dan Teknologi". IDEaTech 2011
- [10] Kalpit G. Soni and Dr. Atul Patel. "Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data". International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 899-906
- [11] Prasmesti, Dyang Falila, Furqon, M. Tanzil, Dewi, Candra. Implementasi Metode *K-Medoids Clustering* untuk Pengelompokan Data Potensi Kebakaran Hutan/ Lahan Berdasarkan Persebaran Titik Panas (*Hotspot*). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 1, No. 9, Juni 2017, hlm. 723-732
- [12] Mustakim. Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan *Knowledge Discovery in Databases (KDD)* dengan Teknik *K-Means Mining*. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI). Pekanbaru. 3 Oktober 2012.
- [13] Mustakim. Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan *Knowledge Discovery in Databases (KDD)* dengan Teknik *K-Means Mining*. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI). Pekanbaru. 3 Oktober 2012.
- [14] Aditya, Zuliar, Mustakim .Perbandingan Algoritma *Clustering K-Means* dan *Fuzzy C Means* pada Data *Knowledge User Modelling*. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI). Pekanbaru 19 Mei 2017
- [15] Aditya, Zuliar, Mustakim .Perbandingan Algoritma *Clustering K-Means* dan *Fuzzy C Means* pada Data *Knowledge User Modelling*. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI). Pekanbaru 19 Mei 2017