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

**Dini Marlina1, Nurelina Fauzer Putri2, Andri Fernando3, Aditya Ramadhan4**

123Program Studi Sistem Informasi, Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. HR. Soebrantas Km. 15 Panam Pekanbaru-Riau

e-mail: 1dinimarlinacosco@gmail.com, 2fauz3rlina@gmail.com, 3andrifernando61@gmail.com, 4adhitkha@gmail.com

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT**  |
| ***Article history:***Received , 2018Revised , 2018Accepted , 2018 |  | 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 *Silhoutte 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*. |
| ***Keyword:***Cacat*Clustering**Data Mining**K-Medoids**Validasi* |
| *Copyright © 2017Predatech* |
| ***Corresponding Author:***First Author, Departement of Sceince and Technology,State Islamic University of Sultan Syarif Kasim Riau ,HR Soebrantas Road Km. 15, Panam, Pekanbaru. |

1. **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 pengelompokkan dapat dilakukan dengan mengimplementasikan metode *clustering*.

Ada beberapa teknik untuk melakukan klasterisasi, diantaranya *K-Means*, *K-Medoids* dan lainnya. Teknik *K-Means* diperkenalkan oleh MacQueenm (1967), sedangkan teknik *K-Medoids* (*Partitioning Around Medoids*) diperkenalkan oleh Kaufman dan Rousseeuw (1990)[4].Berdasarkan penelitian yang dilakukan oleh (Aishwaya) didapati kesimpulan bahwaalgoritma*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 Pengelompokkan Wilayah Sebaran Cacat pada Anak”. Penelitian ini dapat menjadi sebuah solusi agar memudahkan dalam mengetahui wilayah-wilayah yang memiliki tingkat anak cacat yang tinggi nsehingga dapat dilakukan penanggulangan secara dini.

1. **LANDASAN TEORI**
	1. **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]:

1. Penyandang cacat fisik.
2. Penyandanga ccat mental.
3. Penyandang cacat fisik dan mental.
	1. ***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 dataoperasional perlu dilakukan sebelum tahappenggalian informasi dalam KDD dimulai. Data hasilseleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah daribasis data operasional.
2. *Pre-processing / Cleaning,* sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yangmenjadi fokus KDD. Proses pembersihan mencakupantara lain membuang *duplikasi data,* memeriksa data yang *inkosisten,* dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).
3. Transformation *Coding* adalah transformasi pada data yang telahdipilih, sehingga data tersebut sesuai untuk proses*data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis ataupola informasi yang akan dicari dalam basis data.
4. *Data mining, data mining* adalah proses mencari pola atauinformasi menarik dalam data terpilih denganmenggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangatbervariasi. Pemilihan metode atau algoritma yangtepat sangat bergantung pada tujuan dan proses KDDsecara keseluruhan.
5. *Interpretation / Evaluation,* pola informasi yang dihasilkan dari proses *datamining*perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahapini merupakan bagian dari proses KDD yang disebut*interpretation*. Tahap ini mencakup pemeriksaanapakah pola atau informasi yang ditemukanbertentangan dengan fakta atau hipotesis yang ada sebelumnya



Gambar 1. Proses *Data Mining* Dalam Penemuan Pengetahuan Dalam Database [8]

* 1. ***Clustering***

*Clustering* merupakan suatu prosespengelompokan data, observasi, atau mengelompokan 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 klastering yang telah digunakan oleh peneliti sebelumnya seperti *K-Means, Improved K-Means, K-Medoids* (PAM), *Fuzzy C-Means, DBSCAN, CLARANS dan Fuzzy Substractive.*

*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 mengelompokan data sesuaidengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok walaupun setiap algortima*clustering* memiliki kelebihan dan kekurangan masing-masing.

* 1. **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\_minA, new\_MaxA] dengan persamaan 1.

$v^{'}=\frac{v-min\_{A}}{max\_{A}-min\_{A}}\left(new max\_{A}-new min\_{A}\right)+ new\_{minA}$……………………(1)

* 1. ***K-Medoids***

*K-Medoid*sadalah 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]

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:

………………...(2)

$$d\left(x,y\right)=\sqrt{\sum\_{i=1}^{n}(x\_{i}-y\_{i})^{2}}$$

1. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
2. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
3. 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*.
4. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster*masing-masing.
	1. ***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].

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
5. 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
	1. ***Silhoutte Coefficient Index* (SI)**

*Silhoutte 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 dimasukan ke dalam cluster A atau cluster B. Akan tetapi, jika *si* = -1 artinya struktur *cluster* yang dihasilkan *overlapping*, sehingga objek i lebih tepat dimasukan ke dalam *cluster* yang lain[15].

Adapun rumus dari *Silhoutte Coefficient Index* adalah:

1. **METODOLOGI 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

Adapun alur algoritma dari K-Medoid adalah sebagai berikut:



Gambar 2. Alur algoritma *K-Medoids*

Adapun alur algoritma dari K-Medoid adalah sebagai berikut:



Gambar 2. Alur algoritma *K-Means*[16]

1. **HASIL DAN ANALISIS**

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-Medoidss*yaitu mengetahui pengelompokan desa berdasarkan jumlah penyandang cacat. Adapun atribut yang digunakan yaitu:

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 adalah sebagai berikut:

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 |

Sebelum dilakukan pengolahan data menggunakan Algoritma *K-Medoids,* lakukan proses normalisasi. Adapun hasil normalisasi data sebaran cacat adalah sebagai berikut:

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 | … | … |

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).

* 1. **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 terpiih akan merepresentasikan pusat dari cluster sebagai *medoids*, untuk kasus ini akan dipilih tiga buah objek sebagai *medoids* awal.

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 |

Langkah selanjutnya adalah menghitung jarak setiap seluruh objek non-*medoidss*. Objek dengan jarak terdekat dengan masing-masing *medoidss* akan bergabung dengan *medoidss* tersebut. Berikut adalah objek yang paling dekatdengan *medoidss* 1 (C1):

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 |

Berikut adalah objek yang paling dekat dengan *medoids* 2 (C2):

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 |

Berikut adalah objek yang paling dekat dengan *medoids* 3 (C3):

Tabel 6. 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 |

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

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

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

$$\sum\_{}^{}S\left(C1,Oi\right)=d\left(C1,O2\right)+d\left(C1,O4\right)+…+d(C1,O1636)$$

$$\sum\_{}^{}S\left(C1,Oi\right)=0.0492+0.0857+…+0.0651=32.1567$$

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

$$\sum\_{}^{}S\left(C2,Oi\right)=0.1047+0.1390+…+0.0774=26.5401$$

$$\sum\_{}^{}S\left(C2,Oi\right)=0.0751+0.1431+…+0.4091=119.6672$$

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

$$\sum\_{}^{}S\left(Ci,Oi\right)=32.1567+26.5401+119.6672=178.3640$$

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-*medoidss*. 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:

$$\sum\_{}^{}S\left(Ci,Oi\right)=49.0599+7.9079+185.6273=185.6273$$

Perhitungan yang dihasilkan dari dua kali percobaan adalah sebagai berikut:

Tabel 7. Hasil *cost* pada Kedua Percobaan

|  |  |
| --- | --- |
| Cost | Hasil |
| $$\sum\_{}^{}S\left(Ci,Oi\right)pertama$$ | $$178.3640$$ |
| $$\sum\_{}^{}S\left(Ci,Oi\right)kedua$$ | $$185.6273$$ |

Tabel diatas menjelaskan bahwa terdapat perubahan jumlah *cost* dari $178.3640 $menjadi $185.6273$ maka nilai ∑S(Ci,Oi)pertama < ∑S(Ci,Oi)kedua, oleh karena itu algoritma berhenti pada tahap ini dan tidak ada perubahan susunan *cluster* pada percobaan pertama.

Berikut adalah data pada klaster 1:

Tabel 8. 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 |

Jumlah data yang terdapat pada Klaster 1 adalah 601 data. Pola yang terbentuk pada klaster 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.

Berikut adalah data yang terdapat pada klaster 2:

Tabel 9. 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 |

Jumlah data yang terdapat pada klaster kedua adalah 352 data. Pola yang terbentuk pada klaster 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.

Berikut adalah data yang terdapat pada klaster 3:

Tabel 10. 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 |

Jumlah data yang terdapat pada klaster kedua adalah 698 data. Pola yang terbentuk pada klaster 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.

Berikut adalah grafik hasil klaster pada data sebaran cacat pada anak:

* 1. **Perhitungan K-Means**

 Perhitungan juga dilakukan pada Algoritma K-Means untuk mengetahui perbedaan pola yang terentuk. Dalam algoritma *K-Means*, *cluster*1 yang terbentuk adalah sebagai berikut:

Tabel . *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 |

*Cluster* 2 yang terbentuk antara lain:

Tabel . *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 |

*Cluster* 3 yang terbentuk antara lain:

Tabel . *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 |

* 1. **Validasi *K-Medoids*dan *K-Means***

 Perhitungan *Silhoutte Coefficient Index* (SI) digunakan untuk mengetahui kinerja tiap algoritma *clustering*. Adapun hasil perhitungan validasi *Silhoutte* pada metode *K-Medoids* adalah sebagai berikut:

Tabel . Hasil Validasi Silhoutte algoritma *K-Medoids*

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

 Rata-rata Silhoutte 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 *Silhoutte Coefficient* adalah sebagai berikut:

Tabel . Hasil Validasi Silhoutte algoritma *K-Means*

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

 Rata-rata *Silhoutte 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

1. **KESIMPULAN**

 *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. Klaster 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 klaster pertama memiliki data dengan nilai yang paling rendah dibandingkan klaster lainnya. Pada klaster 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 klaster kedua mempunyai data tingkat sebaran anak cacat yang lebih tinggi dibandingkan pada klaster yang pertama. Sedangkan pada klaster yang ketiga, Pola yang terbentuk pada klaster 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 klaster ketiga mempunya data dengan nilai cacat yang paling tinggi dibanding klaster 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 *Silhoutte 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.Kom., M.T selaku dosen pembimbing matakuliah *Data Mining,* sehingga penelitian ini dapat diselesaikan dengan baik.

**REFERENSI**

**If references are Journal:**

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 MiningAlgorithms. 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

**BIBLIOGRAPHY OF AUTHORS (10 PT)**

|  |  |
| --- | --- |
| IMG-20180111-WA0260.jpg | Dini MarlinaDuri, 27th Agustus 1996dinimarlinacosco@gmail.comStudents of State University of Sultan Syarif Kasim Riau |
|  |  |
| NURELINA_FAUZER_PUTRI_l.png | Nurelina Fauzer PutriBukittinggi, 20th Agustus 1995fauz3rlina@gmail.comStudents of State University of Sultan Syarif Kasim Riau |
|  |  |
| IMG-20180111-WA0261.jpg | Andri FernandoPekanbaru, 21 Mei 1997andrifernando61@gmail.com Students of State University of Sultan Syarif Kasim Riau |
| IMG-20180111-WA0259.jpg | Aditya RamadhanPekanbaru, 14th February 1996adhitkha@gmail.comStudents of State University of Sultan Syarif Kasim Riau |