

Implementasi Data Mining K-Means Clustering Dalam Pengelompokan Data Penyakit Pasien Pada Puskesmas Bestari Petisah

Fahrur Rozy¹, Ahmad Fitri Boy², Egi Affandi³

¹Sistem Informasi, STMIK Triguna Dharma

^{2,3}Sistem Informasi, STMIK Triguna Dharma

Article Info

Article history:

Received May 9th, 2020

Revised May 11th, 2020

Accepted May 30th, 2020

Keyword:

Penyakit,
Pengelompokan,
Data Mining,
K-Means

ABSTRACT

Pusat kesehatan masyarakat (puskesmas) merupakan kesatuan organisasi fungsional yang menyelenggarakan upaya kesehatan yang bersifat menyeluruh, terpadu, merata dapat di terima dan terjangkau oleh masyarakat, salah satu puskesmas yang ada di Kota Medan yaitu puskesmas Bestari. Puskesmas Bestari ingin berupaya terus meningkatkan pelayanan dan juga sistem-sistem yang belum mendukung kecerdasan teknologi di era global pada sekarang ini, yang bertujuan untuk memberikan yang terbaik dalam bidang pelayanan kesehatan untuk masyarakat dan juga mempermudah para medis dalam melayani pasien di puskesmas Bestari.

Namun pada saat ini pihak puskesmas masih menggunakan pencatatan manual yaitu dengan microsoft excel dan word untuk pengelompokan data penyakit pada pasien, sehingga masih sulit menemukan informasi mengenai pengelompokan penyakit "Akut dan "Tidak Akut" yang banyak di derita oleh pasien pada puskesmas Bestari. Dengan alasan itulah maka diangkat penelitian dengan judul "Implementasi Data Mining K-Means Clustering Dalam Pengelompokan Data Penyakit Pasien Pada Puskesmas Bestari Petisah".

Penelitian tersebut bertujuan untuk menciptakan suatu sistem berbasis komputerisasi, kemudian dengan diterapkannya sistem tersebut maka hasil yang didapatkan akan benar-benar akurat dan cepat. Dapat dikatakan bahwa dengan pengujian sistem berdasarkan atribut-atribut yang ada akan memberikan jawaban pasti dalam pengelompokan data penyakit pasien. Hal ini karena penerapan metode yang di masukkan ke dalam coding program sehingga sistem ini dapat membantu Puskesmas Bestari Petisah.

Copyright © 2020 STMIK Triguna Dharma.
All rights reserved.

Corresponding Author: *First Author

Nama : Fahrur Rozy

Program Studi

STMIK Triguna Dharma

Email: rozyozy1659@yahoo.com

1. PENDAHULUAN

Pusat kesehatan masyarakat (puskesmas) merupakan kesatuan organisasi fungsional yang menyelenggarakan upaya kesehatan yang bersifat menyeluruh, terpadu, merata dapat di terima dan terjangkau oleh masyarakat dengan peran aktif untuk masyarakat dan menggunakan hasil pengembangan ilmu pengetahuan dan teknologi tepat guna dengan biaya yang dapat dipikul oleh pemerintah dan masyarakat luas guna mencapai derajat kesehatan yang optimal tanpa mengabaikan mutu pelayanan kepada perorangan. Salah satu puskesmas yang ada di Kota Medan yaitu puskesmas Bestari.

Puskesmas Bestari terletak di Jl. Rotan No.1, Petisah Tengah, Kec. Medan Petisah, Kota Medan, Merupakan suatu layanan kesehatan yg disediakan pemerintah untuk masyarakat. Puskesmas Bestari

memberikan pelayanan kedokteran berupa kesehatan, pengobatan dan penyuluhan kepada masyarakat. Pasien yang khususnya berdomisili di daerah puskesmas Bestari Medan lebih diutamakan untuk mendapatkan pelayanan penanganan kesehatan seperti layanan kesehatan dari gula darah, tensi dll. Oleh sebab itu puskesmas Bestari ingin berupaya terus meningkatkan pelayanan dan juga sistem-sistem yang belum mendukung kecerdasan teknologi di era global pada sekarang ini, yang bertujuan untuk memberikan yang terbaik dalam bidang pelayanan kesehatan untuk masyarakat dan juga mempermudah para medis dalam melayani pasien di puskesmas Bestari. Algoritma clustering di gunakan untuk mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah cluster, hitung jarak terdekat dengan pusat cluster. Data dengan jarak terdekat menyatakan anggota dari cluster tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada cluster lain, untuk meminimalkan fungsi objektif. *Clustering* adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan menggolongkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh. Ciri khas dari teknik data mining adalah mempunyai sifat tanpa arahan, yang dimaksud adalah teknik ini diterapkan tanpa perlunya data training dan tanpa ada teacher serta tidak memerlukan target output.

Metode *clustering* yang mempunyai sifat efisien dan cepat yang dapat digunakan salah satunya adalah metode k-means, metode ini bertujuan untuk membuat cluster objek berdasarkan atribut menjadi k partisi. Cara kerja metode ini adalah mula-mula ditentukan cluster yang akan di bentuk, pada elemen pertama dalam tiap cluster dapat dipilih untuk dijadikan sebagai titik tengah (centroid), selanjutnya akan dilakukan pengulangan langkah-langkah hingga tidak ada objek yang dapat di pindahkan lagi. Data Mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data Mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah Data Mining kadang disebut juga *knowledge Discovery* [1].

Data mining merupakan proses *iterative* dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambil keputusan diwaktu yang akan datang, pola-pola ini dikenali perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lain [2].

Istilah data mining sudah muncul sejak tahun 1990-an dan mulai populer sejak awal tahun 2000-an. Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah *pattern recognition*[3].

Makna lain yang mengenai *clustering* adalah sebuah mekanisme klasifikasi dan banyak digunakan untuk dirancang dalam mengkategorikan data. Penggunaan teknik seperti ini sering digunakan dalam mode eksplorasi data, dan *clustering* dapat dilakukan dengan menggunakan berbagai algoritma yang berbeda salah satunya adalah *K-Means*[4].

K-Means merupakan salah satu metode pengelompokan data *nonhierarki* (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama di masukan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain.

Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok [5]

Berikut ini adalah algoritma atau langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut:

1. Pilih jumlah *cluster k*.
2. Inisialisasi *k* pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random,
3. Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

2. METODE PENELITIAN

Dalam melakukan suatu penelitian memerlukan langkah-langkah atau cara tertentu yang menjadi pedoman selama proses penelitian, agar hasil penelitian sesuai dengan tujuan yang telah ditetapkan. Jika metodologi penelitian yang dilakukan baik, maka semakin baik pula hasil penelitian yang didapatkan. Adapun metode penelitian yang akan dilakukan dalam penulisan skripsi ini adalah sebagai berikut :

Teknik Pengumpulan Data

1. Penelitian Kepustakaan (Library Research)

Penelitian kepustakaan merupakan salah satu elemen yang mendukung sebagai landasan teoritis peneliti dalam mengkaji masalah yang dibahas. Dalam hal ini, peneliti menggunakan beberapa sumber kepustakaan diantaranya: Buku, Jurnal Nasional, Jurnal Internasional dan Sumber-sumber lainnya.

2. Penelitian Lapangan (Field Research)

Setelah observasi, yang dilakukan selanjutnya adalah teknik wawancara kepada staff Kantor wali Nagari Sitombol yang disetujui oleh Bapak Busri Wali Nagari Sitombol. Teknik wawancara ini dilakukan untuk menggali informasi mengenai prosedur dalam menentukan masyarakat yang layak menerima bantuan bedah rumah keluarga miskin melalui Kantor Wali Nagari Sitombol, dan apa saja yang menjadi kendala selama ini serta mencari solusi untuk kendala yang dihadapi. Berdasarkan hasil wawancara, adapun hasil dari wawancara ini telah dirumuskan pada latar belakang yang mendukung bahwasanya penelitian ini perlu dilakukan.

Tabel 1. Data Pasien Puskesmas Bestari Petisah Periode Mei-Juni 2021

| Kode | Nama Pasien | Usia (Tahun) | Jenis Kelamin | Diagnosa |
|------|---------------------|--------------|---------------|----------------------|
| P01 | Dahlia Anggraini | 30 | Perempuan | Diare, Mencret |
| P02 | M. Ramsel | 22 | Laki-Laki | Gatal-Gatal |
| P03 | Krista yudha | 20 | Laki-Laki | Batuk |
| P04 | Fikri Aulia | 15 | Laki-Laki | Gatal-Gatal |
| P05 | Cut Maryam | 60 | Perempuan | Asam Lambung |
| P06 | Rosma Rolinda Silau | 56 | Perempuan | Sakit Jantung |
| P07 | Nurazizah S | 60 | Perempuan | Sakit Maag |
| P08 | Emrizal | 57 | Laki-Laki | Gatal-Gatal |
| P09 | Suyatman | 49 | Laki-Laki | TB Paru |
| P10 | Petty | 44 | Perempuan | Flu, Batuk |
| P11 | Teguh Susan | 47 | Laki-Laki | Diare |
| P12 | Wagimun Wisno | 60 | Laki-Laki | TB Paru |
| P13 | Arini Magdalena S | 31 | Perempuan | Kolestrol |
| P14 | Ronal Parlindungan | 57 | Laki-Laki | Kejang-Kejang |
| P15 | Jeni | 50 | Perempuan | Cuci Darah |
| P16 | Tjut Murri Trm | 63 | Perempuan | Sakit Gigi |
| P17 | Magdalena | 78 | Perempuan | Ginjal |
| P18 | Rian | 1.2 | Laki-Laki | Gatal-Gatal Berdarah |
| P19 | Pahala S | 61 | Laki-Laki | Batuk |
| P20 | Susanty | 29 | Perempuan | Asam Lambung |

Berikut ini adalah data nilai atribut pada data mining untuk pengelompokan data penyakit pada pasien Puskesmas Bestari Medan menggunakan metode *K-Means*.

Tabel 2. Data Atribut

| No | Kode | Atribut |
|----|------|-------------------------------|
| 1 | A1 | Jenis Penyakit |
| 2 | A2 | Tingkatan Keseriusan Penyakit |
| 3 | A3 | Waktu Penyakit |

Tabel 3. Konversi Nilai Atribut Jenis Penyakit

| Atribut | Keterangan | Nilai |
|----------------|---------------|-------|
| Jenis Penyakit | Tidak Menular | 0 |
| | Menular | 1 |

Tabel 4. Konversi Nilai Atribut Tingkat Keseriusan Penyakit

| Atribut | Keterangan | Nilai |
|-----------------------------|------------------|-------|
| Tingkat Keseriusan Penyakit | Tidak Berbahaya | 1 |
| | Kurang Berbahaya | 2 |
| | Cukup Serius | 3 |
| | Serius | 4 |
| | Sangat Serius | 5 |

Tabel 5. Konversi Nilai Atribut Jenis Penyakit

| Atribut | Keterangan | Nilai |
|----------------|-----------------------|-------|
| Waktu Penyakit | Penyakit Baru Dialami | 0 |
| | Penyakit Berulang | 1 |

Tabel 6. Data Alternatif Untuk Setiap Atribut

| Kode | Penyakit | Jenis Penyakit | Tingkatan Keseriusan Penyakit | Waktu Penyakit |
|------|----------------|----------------|-------------------------------|-----------------------|
| P01 | Diare, Mencret | Tidak Menular | Cukup Serius | Penyakit Baru Dialami |
| P02 | Gatal-Gatal | Menular | Cukup Serius | Penyakit Baru Dialami |
| P03 | Batuk | Menular | Kurang Berbahaya | Penyakit Baru Dialami |
| P04 | Gatal-Gatal | Menular | Cukup Serius | Penyakit Baru Dialami |
| P05 | Asam Lambung | Tidak Menular | Serius | Penyakit Berulang |
| P06 | Sakit Jantung | Tidak Menular | Sangat Serius | Penyakit Berulang |
| P07 | Sakit Maag | Tidak Menular | Serius | Penyakit Berulang |

| | | | | |
|-----|----------------------|---------------|------------------|-----------------------|
| P08 | Gatal-Gatal | Menular | Cukup Serius | Penyakit Baru Dialami |
| P09 | TB Paru | Menular | Sangat Serius | Penyakit Baru Dialami |
| P10 | Flu, Batuk | Menular | Cukup Serius | Penyakit Berulang |
| P11 | Diare | Tidak Menular | Kurang Berbahaya | Penyakit Baru Dialami |
| P12 | TB Paru | Menular | Sangat Serius | Penyakit Baru Dialami |
| P13 | Kolestrol | Tidak Menular | Kurang Berbahaya | Penyakit Berulang |
| P14 | Kejang-Kejang | Tidak Menular | Sangat Serius | Penyakit Baru Dialami |
| P15 | Cuci Darah | Tidak Menular | Sangat Serius | Penyakit Berulang |
| P16 | Sakit Gigi | Tidak Menular | Kurang Berbahaya | Penyakit Berulang |
| P17 | Ginjal | Tidak Menular | Serius | Penyakit Baru Dialami |
| P18 | Gatal-Gatal Berdarah | Menular | Serius | Penyakit Baru Dialami |
| P19 | Batuk | Menular | Kurang Berbahaya | Penyakit Baru Dialami |
| P20 | Asam Lambung | Tidak Menular | Serius | Penyakit Berulang |

Tabel 7. Data Konversi Nilai Alternatif Untuk Setiap Atribut

| Alternatif | A1 | A2 | A3 |
|------------|----|----|----|
| P01 | 0 | 3 | 0 |
| P02 | 1 | 3 | 0 |
| P03 | 1 | 2 | 0 |
| P04 | 1 | 3 | 0 |
| P05 | 0 | 4 | 1 |
| P06 | 0 | 5 | 1 |
| P07 | 0 | 4 | 1 |
| P08 | 1 | 3 | 0 |
| P09 | 1 | 5 | 0 |
| P10 | 1 | 3 | 1 |
| P11 | 0 | 2 | 0 |
| P12 | 1 | 5 | 0 |
| P13 | 0 | 2 | 1 |
| P14 | 0 | 5 | 0 |
| P15 | 0 | 5 | 1 |
| P16 | 0 | 2 | 1 |
| P17 | 0 | 4 | 0 |
| P18 | 1 | 4 | 0 |
| P19 | 1 | 2 | 0 |
| P20 | 0 | 4 | 1 |

Langkah-langkah penyelesaian pada metode *K-Means* adalah sebagai berikut :

1. Tentukan berapa banyak *cluster* k dari dataset yang akan dibagi.
2. Tetapkan secara acak data k menjadi pusat awal lokasi klaster.. Dalam penelitian ini, titik pusat awal ditentukan secara manual oleh pemakai.

3. Hitung pusat *cluster* (*centroid*) menggunakan *mean* untuk masing-masing kelompok. Oleh karena itu, telah terbentuk cluster k: C1, C2, C3, ..., Ck
4. Alokasikan masing-masing data ke *centroid* terdekat.
5. Tentukan Rasio BCV dan WCV untuk tolak ukur perulangan ke Iterasi
6. Ulangi langkah ke-3 dan ke-5 hingga data-data pada tiap *cluster* menjadi terpusat atau selesai.

Untuk mengukur jarak antara data dan *centroid* di ukur dengan cara *Euclidean*. Berikut ini adalah penyelesaian masalah dengan metode *K-Means*:

1. Iterasi ke – 1
 - a. Penentuan pusat (*centroid*) awal *cluster*
 Untuk menentukan pusat (*centroid*) awal ditentukan dengan mengacak (*random*) dari data nilai yang sudah ada. Pada kasus ini pusat *centroid* awal adalah data ke-2 dan data ke-6 :

Tabel 8. Titik Pusat (*Centorid*) Awal *Cluster*

| Titik Pusat <i>Centorid</i> | Nilai Awal <i>Cluster</i> | | |
|-----------------------------|---------------------------|---|---|
| Data ke-2 (C1) | 1 | 3 | 0 |
| Data ke-6 (C2) | 0 | 5 | 1 |

- b. Perhitungan jarak dengan pusat *cluster*
 Perhitungan jarak dari data ke-1 terhadap pusat *cluster* adalah :

$$D(1,1) = \sqrt{(0 - 1)^2 + (3 - 3)^2 + (0 - 0)^2}$$

$$= 1$$

$$D(1,2) = \sqrt{(0 - 0)^2 + (3 - 5)^2 + (0 - 1)^2}$$

$$= 2$$

Perhitungan jarak dari data ke-2 terhadap pusat *cluster* adalah :

$$D(2,1) = \sqrt{(1 - 1)^2 + (3 - 3)^2 + (0 - 0)^2}$$

$$= 0$$

$$D(2,2) = \sqrt{(1 - 0)^2 + (3 - 5)^2 + (0 - 1)^2}$$

$$= 2$$

Dan seterusnya di lakukan perhitungan jarak untuk data ke-3 sampai data ke-20 kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat *cluster* baru, kemudian dilakukan pengelompokan dengan melihat jarak terdekat antara data dengan pusat *cluster*.

Tabel 9. Hasil Perhitungan Jarak (*Distance Score*) untuk Masing-masing *Cluster* dan Pengelompokan Pusat *Cluster*

| Alternatif | Nilai Jarak | | Jarak Terdekat |
|------------|-------------|----|----------------|
| | C1 | C2 | |
| P01 | 1 | 2 | C1 |
| P02 | 0 | 2 | C1 |
| P03 | 1 | 3 | C1 |
| P04 | 0 | 2 | C1 |
| P05 | 2 | 1 | C2 |
| P06 | 2 | 0 | C2 |
| P07 | 2 | 1 | C2 |
| P08 | 0 | 2 | C1 |
| P09 | 2 | 1 | C2 |
| P10 | 1 | 2 | C1 |
| P11 | 1 | 3 | C1 |

| | | | |
|-----|---|---|----|
| P12 | 2 | 1 | C2 |
| P13 | 2 | 3 | C1 |
| P14 | 2 | 1 | C2 |
| P15 | 2 | 0 | C2 |
| P16 | 2 | 3 | C1 |
| P17 | 1 | 1 | C2 |
| P18 | 1 | 2 | C1 |
| P19 | 1 | 3 | C1 |
| P20 | 2 | 1 | C2 |

Hasil Cluster :

C1 : {P01, P02, P03, P04, P08, P10, P11, P13, P16, P18, P19}

C2 : {P05, P06, P07, P09, P12, P14, P15, P17, P20}

- c. Menghitung Nilai BCV (*Between Cluster Variation*)

Karena *centroid* m1 (1;3;0), m2 (0;5;1)

$$D(m1,m2) = \sqrt{(1-0)^2 + (3-5)^2 + (0-1)^2}$$

$$= 2$$

$$BCV = 2$$

- d. Menghitung Nilai WCV (*Within Cluster Variation*)

Memilih jarak terdekat antara data dengan *centorid*. Berikut ini adalah jarak nilai terdekat yaitu :

Tabel 10. Jarak Nilai Terdekat Antara Data dan *Centorid*

| Alternatif | Jarak Terdekat |
|------------|----------------|
| P01 | 1 |
| P02 | 0 |
| P03 | 1 |
| P04 | 0 |
| P05 | 1 |
| P06 | 0 |
| P07 | 1 |
| P08 | 0 |
| P09 | 1 |
| P10 | 1 |
| P11 | 1 |
| P12 | 1 |
| P13 | 2 |
| P14 | 1 |
| P15 | 0 |
| P16 | 2 |
| P17 | 1 |
| P18 | 1 |
| P19 | 1 |
| P20 | 1 |

$$WCV = 1^2 + 0^2 + 1^2 + 0^2 + 1^2 + 0^2 + 1 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 0^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2$$

$$WCV = 21$$

e. Menghitung Besar Rasio

Rasio besaran antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*) :

$$\begin{aligned} \text{Besar Rasio} &= BCV/WCV \\ &= 2/21 \\ &= 0,116642369 \end{aligned}$$

Karena ini merupakan Iterasi-1, maka dilanjutkan ke Iterasi selanjutnya.

Pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster*.

Untuk *cluster* C1, ada 11 anggota didalamnya yaitu:

C1 : {P01, P02, P03, P04, P08, P10, P11, P13, P16, P18, P19}

Tabel 11. Hasil *Cluster* C1

| Alternatif | A1 | A2 | A3 |
|------------|------|------|----|
| P01 | 0 | 3 | 0 |
| P02 | 1 | 3 | 0 |
| P03 | 1 | 2 | 0 |
| P04 | 1 | 3 | 0 |
| P08 | 1 | 3 | 0 |
| P10 | 1 | 3 | 1 |
| P11 | 0 | 2 | 0 |
| P13 | 0 | 2 | 1 |
| P16 | 0 | 2 | 1 |
| P18 | 1 | 4 | 0 |
| P19 | 1 | 2 | 0 |
| Mean | 0,58 | 2,42 | 0 |

Untuk *cluster* C2, ada 9 anggota didalamnya yaitu:

C2 : {P05, P06, P07, P09, P12, P14, P15, P17, P20}

Tabel 12. Hasil *Cluster* C2

| Alternatif | A1 | A2 | A3 |
|------------|------|------|----|
| P05 | 0 | 4 | 1 |
| P06 | 0 | 5 | 1 |
| P07 | 0 | 4 | 1 |
| P09 | 1 | 5 | 0 |
| P12 | 1 | 5 | 0 |
| P14 | 0 | 5 | 0 |
| P15 | 0 | 5 | 1 |
| P17 | 0 | 4 | 0 |
| P20 | 0 | 4 | 1 |
| Mean | 0,25 | 4,61 | 1 |

Setelah perhitungan diatas, maka akan diperoleh *centroid* baru dengan nilai seperti terlihat pada Tabel 3.13 berikut ini :

Tabel 13. Titik Pusat Awal *Cluster* Baru

| Titik Pusat <i>Centorid</i> | Nilai <i>Cluster</i> Baru | | |
|-----------------------------|---------------------------|------|---|
| <i>Centroid</i> 1 | 0,58 | 2,42 | 0 |
| <i>Centroid</i> 2 | 0,25 | 4,63 | 1 |

2. Iterasi ke – 2

Setelah dilakukan perhitungan pusat *centroid* maka selanjutnya melakukan perhitungan jarak sebagai berikut:

a. Perhitungan jarak dari data ke-1 terhadap pusat *cluster* baru adalah :

$$D(1,1) = \sqrt{(0 - 0,58)^2 + (3 - 2,42)^2 + (0 - 0)^2}$$

$$= 1$$

$$D(1,2) = \sqrt{(0 - 0,25 + (3 - 4,63))^2 + (0 - 1)^2}$$

$$= 2$$

b. Perhitungan jarak dari data ke-2 terhadap pusat *cluster* baru adalah :

$$D(2,1) = \sqrt{(1 - 0,58)^2 + (3 - 2,42)^2 + (0 - 0)^2}$$

$$= 1$$

$$D(2,2) = \sqrt{(1 - 0,25 + (3 - 4,63))^2 + (0 - 1)^2}$$

$$= 2$$

Dan seterusnya di lakukan perhitungan jarak untuk data ke-3 sampai data ke-20. Berikut ini adalah hasil perhitungan jaraknya, yaitu :

Tabel 14. Hasil Perhitungan *Distance Score* dan Pengelompokan *Cluster*

| Alternatif | Nilai Jarak | | Jarak Terdekat |
|------------|-------------|----|----------------|
| | C1 | C2 | |
| P01 | 1 | 2 | C1 |
| P02 | 1 | 2 | C1 |
| P03 | 1 | 3 | C1 |
| P04 | 1 | 2 | C1 |
| P05 | 2 | 1 | C2 |
| P06 | 3 | 0 | C2 |
| P07 | 2 | 1 | C2 |
| P08 | 1 | 2 | C1 |
| P09 | 3 | 1 | C2 |
| P10 | 1 | 2 | C1 |
| P11 | 1 | 3 | C1 |
| P12 | 3 | 1 | C2 |
| P13 | 1 | 3 | C1 |
| P14 | 3 | 1 | C2 |
| P15 | 3 | 0 | C2 |
| P16 | 1 | 3 | C1 |
| P17 | 2 | 1 | C2 |
| P18 | 2 | 1 | C2 |
| P19 | 1 | 3 | C1 |
| P20 | 2 | 1 | C2 |

Hasil *Cluster* :

C1 : {P01, P02, P03, P04, P08, P10, P11, P13, P16, P19}

C2 : {P05, P06, P07, P09, P12, P14, P15, P17, P18, P20}

c. Menghitung Nilai BCV (*Between Cluster Variation*)

Karena *centroid* m1 (0,58; 2,42; 0), m2 (0,25; 4,63; 1)

$$D(m1,m2) = \sqrt{(0,58 - 0,25)^2 + (2,42 - 4,63)^2 + (0 - 1)^2}$$

$$= 2$$

$$BCV = 2$$

d. Menghitung Nilai WCV (*Within Cluster Variation*)

Memilih jarak terdekat antara data dengan *centorid*. Berikut ini adalah jarak nilai terdekat yaitu :

Tabel 15. Jarak Nilai Terdekat Antara Data dan Centorid

| Alternatif | Jarak Terdekat |
|------------|----------------|
| P01 | 1 |
| P02 | 1 |
| P03 | 1 |
| P04 | 1 |
| P05 | 1 |
| P06 | 0 |
| P07 | 1 |
| P08 | 1 |
| P09 | 1 |
| P10 | 1 |
| P11 | 1 |
| P12 | 1 |
| P13 | 1 |
| P14 | 1 |
| P15 | 0 |
| P16 | 1 |
| P17 | 1 |
| P18 | 1 |
| P19 | 1 |
| P20 | 1 |

$$WCV = 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2$$

$$WCV = 18$$

e. Menghitung Besar Rasio

Rasio besaran antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*) :

$$\begin{aligned} \text{Besar Rasio} &= BCV/WCV \\ &= 2/18 \\ &= 0,116642369 \end{aligned}$$

Karena rasio yang baru (0,116642369) tidak lebih besar dari rasio sebelumnya (0,116642369) maka proses perhitungan iterasi dihentikan.

3. Kesimpulan

Karena iterasi telah mencapai akhir atau selesai maka dari perhitungan tersebut diperoleh kesimpulan sebagai berikut:

Tabel 16. Hasil Akhir Pengelompokkan Penyakit Dengan Metode K-Means

| Alternatif | Cluster | | Keterangan | |
|------------|---------|----|---------------------|---------------|
| | C1 | C2 | Penyakit Tidak Akut | Penyakit Akut |
| P01 | C1 | | Tidak Akut | |
| P02 | C1 | | Tidak Akut | |
| P03 | C1 | | Tidak Akut | |
| P04 | C1 | | Tidak Akut | |
| P05 | | C2 | | Akut |
| P06 | | C2 | | Akut |
| P07 | | C2 | | Akut |
| P08 | C1 | | Tidak Akut | |
| P09 | | C2 | | Akut |
| P10 | C1 | | Tidak Akut | |
| P11 | C1 | | Tidak Akut | |
| P12 | | C2 | | Akut |
| P13 | C1 | | Tidak Akut | |
| P14 | | C2 | | Akut |

Tabel 17. Hasil Akhir Pengelompokan Penyakit Dengan Metode *K-Means* (lanjutan)

| Alternatif | Cluster | | Keterangan | |
|------------|---------|----|-------------|---------------|
| | C1 | C2 | Rawat Jalan | Harus Dirujuk |
| P15 | | C2 | | Akut |
| P16 | C1 | | Tidak Akut | |
| P17 | | C2 | | Akut |
| P18 | | C2 | | Akut |
| P19 | C1 | | Tidak Akut | |
| P20 | | C2 | | Akut |

Dari tabel 3.16 di atas di dapat kesimpulan yaitu mana saja penyakit yang rawat jalan dan penyakit yang harus di rujuk, berikut ini adalah kesimpulan yang di dapat:

1. Penyakit Yang Tidak Akut
 - a. Diare, Mencret
 - b. Gatal-Gatal
 - c. Batuk
 - d. Flu, Batuk
 - e. Diare
 - f. Kolestrol
 - g. Sakit Gigi
2. Penyakit Yang Akut
 - a. Asam Lambung
 - b. Sakit Jantung
 - c. Sakit Maag
 - d. TB Paru
 - e. Kejang-Kejang
 - f. Cuci Darah
 - g. Ginjal
 - h. Gatal-Gatal Berdarah

Gambar 1. *Form Login*

Form Data Pasien

Input Data

Kode :

Nama Pasien :

Jenis Kelamin :

Diagnosa Penyakit :

Aksi

| Kode | Nama Pasien | Jenis Kelamin | Diagnosa Penyakit |
|------|--------------------|---------------|----------------------|
| P13 | Arini Magdalena S | Perempuan | Kolestrol |
| P14 | Ronal Parlindungan | Laki-Laki | Kejang-Kejang |
| P15 | Jeni | Perempuan | Cuci Darah |
| P16 | Tjut Murri Trm | Perempuan | Sakit Gigi |
| P17 | Magdalena | Perempuan | Ginjal |
| P18 | Rian | Laki-Laki | Gatal-Gatal Berdarah |

Gambar 2. Form Masukan Data Alternatif

Form Data Atribut

Input Data

Kode Atribut :

Atribut :

Input Data

| Kode Atribut | Atribut |
|--------------|-------------------------------|
| A1 | Jenis Penyakit |
| A2 | Tingkatan Keseriusan Penyakit |
| A3 | Waktu Penyakit |

Gambar 3. Form Masukan Data Atribut

Form Penilaian

Penilaian Alternatif Setiap Atribut

Kode :

Diagnosa Penyakit :

Atribut :

Jenis Penyakit (A1) : 0

Tingkatan Keseriusan Penyakit (A2) : 4

Waktu Penyakit (A3) : 1

Aksi

| Kode | Diagnosa Penyakit | K1 | K2 | K3 |
|------|----------------------|----|----|----|
| P11 | Diare | 0 | 2 | 0 |
| P12 | TB Paru | 1 | 5 | 0 |
| P13 | Kolestrol | 0 | 2 | 1 |
| P14 | Kejang-Kejang | 0 | 5 | 0 |
| P15 | Cuci Darah | 0 | 5 | 1 |
| P16 | Sakit Gigi | 0 | 2 | 1 |
| P17 | Ginjal | 0 | 4 | 0 |
| P18 | Gatal-Gatal Berdarah | 1 | 4 | 0 |

Gambar 4. Form Proses Penilaian

Gambar 5. Form Proses Perhitungan

LAPORAN HASIL PERHITUNGAN DENGAN METODE K-MEANS

| No | Diagnosa Penyakit | Tidak Akut | Akut |
|----|----------------------|------------|------|
| 1 | Diare, Mencret | Tidak Akut | |
| 2 | Gatal-Gatal | Tidak Akut | |
| 3 | Batuk | Tidak Akut | |
| 4 | Gatal-Gatal | Tidak Akut | |
| 5 | Asam Lambung | | Akut |
| 6 | Sakit Jantung | | Akut |
| 7 | Sakit Maag | | Akut |
| 8 | Gatal-Gatal | Tidak Akut | |
| 9 | TB Paru | | Akut |
| 10 | Flu, Batuk | Tidak Akut | |
| 11 | Diare | Tidak Akut | |
| 12 | TB Paru | | Akut |
| 13 | Kolestrol | Tidak Akut | |
| 14 | Kejang-Kejang | | Akut |
| 15 | Cuci Darah | | Akut |
| 16 | Sakit Gigi | Tidak Akut | |
| 17 | Ginjal | | Akut |
| 18 | Gatal-Gatal Berdarah | | Akut |
| 19 | Batuk | Tidak Akut | |
| 20 | Asam Lambung | | Akut |

Gambar 6. Tampilan Form Laporan Hasil

Kesimpulan

Berdasarkan perumusan dan pembahasan bab-bab sebelumnya dapat diambil kesimpulan sebagai berikut

1. Dalam menganalisa suatu masalah dalam mengelompokkan data penyakit pasien menggunakan data mining metode *k-means* yaitu dengan menentukan atribut yang mempengaruhi pengelompokkan data penyakit, kemudian setiap atribut diberikan nilai berdasarkan ketentuan metode *K-Means*, selanjutnya dilakukan proses pengelompokkan.
2. Dalam merancang aplikasi yang mengadopsi metode *k-means* dalam pemecahan masalah untuk mengelompokkan data penyakit pasien di Puskesmas Bestari Medan yaitu dengan merancang aplikasi berbasis *Desktop Programming* kemudian membuat *form-form* yang berkaitan dan mendukung untuk proses pengelompokkan seperti membuat *form* data alternatif, *form* data atribut, *form* proses penilaian dan membuat *form* proses perhitungan. Setelah semua *form* dibuat selanjutnya melakukan perhitungan pengelompokan data penyakit pasien.

3. Sistem yang telah dirancang selanjutnya diimplementasikan dengan memasukkan data-data sesuai dengan yang ada pada bab-bab sebelumnya, kemudahan jika hasil *output*nya sesuai dengan data manual maka dalam pengujian ini sistem berjalan dengan baik, menambahkan data ke *database*, perintah *update* untuk merubah data di *database*, perintah *delete* untuk menghapus data di *database*.

Saran

Untuk lebih mengembangkan dan meningkatkan sistem dalam pengelompokan data penyakit pasien dengan menggunakan metode *K-Means* ada beberapa saran yang dapat dijadikan pertimbangan, yaitu :

1. Sistem dirasa belum sempurna, untuk itu perlu dilakukan pengembangan bagi penelitian selanjutnya seperti penambahan atribut baru secara otomatis.
2. Perangkat lunak sebaiknya menggunakan bahasa pemrograman berbasis web sehingga dapat dengan mudah diakses oleh pihak Puskesmas Bestari Petisah dengan media apa saja.
3. Sistem dapat dikembangkan dengan mengembangkan metode yang digunakan dengan metode lainnya seperti metode *C4.5*.

UCAPAN TERIMA KASIH

Alhamdulillah rabbil alamin kita ucapkan kepada Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat dan karunia-Nya sehingga dapat menyelesaikan Skripsi ini dengan baik. Skripsi ini berjudul "Implementasi Data Mining K-Means Clustering Dalam Pengelompokan Data Penyakit Pasien Pada Puskesmas Bestari Petisah. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Pendidikan Strata 1 program studi Sistem Informasi di STMIK Triguna Dharma Medan. Teristimewa saya ucapkan kepada Ayahanda Ir. Hasnaldi dan Ibunda Hasrah Martha ATLM tercinta, serta kakak atas segala doa, cinta dan kasih serta semangat dan dukungan semangat maupun materi yang tak terhingga. Peneliti menyadari sepenuhnya bahwa apa yang peneliti sampaikan dalam skripsi ini masih banyak kekurangan dan jauh dari sempurna. Hal ini disebabkan keterbatasan pengetahuan dan kemampuan peneliti, sehingga dengan kerendahan hati peneliti mengharapakan saran dan kritik yang bersifat membangun untuk kesempurnaan skripsi ini. Dalam kesempatan ini, peneliti mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Ahmad Fitri Boy, S.Kom, M.Kom sebagai Dosen Pembimbing I. Bapak Egi Affandi, S.Kom., M.Kom selaku Dosen Pembimbing II. Bapak/ Ibu Dosen dan staf STMIK Triguna Dharma Medan yang banyak membantu. Kepada seluruh Karyawan /Staff/Kapus Puskesmas Bestari Petisah Medan Teman-teman Kelas 8SIA8 STMIK Triguna Dharma Medan yang telah banyak membantu saya dalam menyelesaikan skripsi ini. Seluruh anggota Team Pejuang Sidang Sejati yaitu : Nurin Fadhillah Adani, Sallimatul Mardiah, Artsen Dwi Cahya, Musnar Gayo, Yessi Ruhama asirva Tanjung, Ferry Syahrizal, Ramadhan Yandra.

REFERENSI

- [1] R. Gunawan, "Implementasi Data Mining Menggunakan Regresi Linier Berganda dalam Memprediksi Jumlah Nasabah Kredit Macet Pada BPR Tanjung Morawa," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 87, 2019, doi: 10.53513/jis.v18i1.117.
- [2] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *J. Tek. Komput. AMIK BSI*, vol. Vol 4, No., no. September, pp. 1–4, 2018.
- [3] R. Astuti, "Data Mining untuk Klasifikasi dengan Algoritma Cart (Classification and Regression Trees)," *Media Inform.*, vol. 17, no. 3, pp. 114–124, 2018, doi: 10.37595/mediainfo.v17i3.15.
- [4] A. Fitri, Y. Syahra, and R. Kustini, "Penerapan Data Mining Dalam Mengklusterisasi Location Best Pb Tambahan Pada Regional IV PT Indomarco Prismatama Cab. Medan Dengan Menggunakan Metode K-Means," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 19, no. 2, p. 11, 2020, doi: 10.53513/jis.v19i2.2330.
- [5] S. Yakub, "Penerapan Data Mining dalam pengelompokan Bibit Padi Unggul Berdasarkan Minat Beli Konsumen Pada PT. Sang Hyang Seri Regional IV Deli Serdang dengan Menggunakan Metode Clustering Algoritma K-Means," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 17, no. 2, p. 192, 2018, doi: 10.53513/jis.v17i2.43.

| | | | |
|---|---------------|---|---|
|  | Nama | : | Fahrur Rozy |
| | TTL | : | Medan, 16 Oktober 1996 |
| | Jenis Kelamin | : | Laki-Laki |
| | Program Studi | : | Sistem Informasi STMIK Triguna Dharma |
| | Deskripsi | : | Sedang Menempuh jenjang Strata Satu (S1) dengan program studi sistem informasi di STMIK Triguna Dharma. |
| | | | |
|  | Nama | : | Ahmad Fitri Boy, S.Kom., M.Kom |
| | Jenis Kelamin | : | Laki-Laki |
| | Program Studi | : | Sistem Informasi |
| | Deskripsi | : | Dosen tetap Stmik triguna dharma yang aktif mengajar dan meneliti yang berfokus pada bidang keilmuan pemrograman, multimedia dan desain grafis. |
| | | | |
|  | Nama | : | EGI AFFANDI, S.Kom, M.Kom |
| | Program Studi | : | Sistem Informasi |
| | Alamat Email | : | egi.afandi46@gmail.com |
| | Kemampuan | : | Database System, Analisis dan Perancangan Sistem Informasi |
| | | | |