

# Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng

Ulil Ma'rifatini

Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: [ulil\\_sempok@yahoo.com](mailto:ulil_sempok@yahoo.com)

**Abstrak** – Dalam menentukan konsistensi data kesehatan, dapat menggunakan teknik penambangan data yang dapat menggali informasi tersembunyi dari data set yang telah diperoleh. Selain itu, data yang terhubung dengan data lain juga dapat dilakukan dengan teknik penambangan data ini. Salah satu teknik penambangan data yang cukup terkenal yaitu clustering. Metode ini cukup populer dalam teknik penambangan data yang disebut metode k-means. Ini digunakan untuk memfasilitasi perekam medis untuk menganalisis situasi kesehatan umum kelompok populasi dalam pengarsipan data perawatan kesehatan. Hasil analisis ini, pengelompokan penyakit berdasarkan usia, jenis kelamin, durasi penyakit dan diagnosis penyakit. Penelitian ini menggunakan alat Rapid Miner 5.3. Berdasarkan data dari puskesmas Warujayeng, hasil pengelompokan ini adalah 376 item akut. dan 624 penyakit tidak akut dari 1000 total data.

**Kata Kunci** — Clustering, Data Mining, K-Means

## 1. PENDAHULUAN

Clustering adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh. Ciri khas dari teknik data mining ini adalah mempunyai sifat tanpa arahan (unsupervised), 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 dibentuk, 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 dipindahkan lagi.

Berdasarkan permasalahan yang telah dipaparkan di atas maka penelitian ini akan menerapkan metode k-means untuk menghasilkan informasi mengenai pengelompokan penyakit “AKUT” dan “TIDAK AKUT” yang banyak diderita oleh pasien pada Puskesmas Warujayeng. Yang kemudian hasil tersebut dapat dijadikan bahan atau dasar penyuluhan kesehatan oleh Dinas Kesehatan setempat.

## 2. METODE PENELITIAN

### 2.1 Metode Pengumpulan Data

Dalam kegiatan pengumpulan data untuk penelitian ini digunakan metode pengumpulan studi pustaka yang mana pada metode ini kegiatan dilakukan adalah mempelajari, mencari dan mengumpulkan data yang berhubungan dengan penelitian ini. Data yang digunakan dalam pengelompokan penyakit pasien ini diperoleh dari data pasien Puskesmas Warujayeng dimana peneliti-an ini dilakukan. Data yang diperoleh kemudian akan di olah menggunakan metode k-means dengan mengambil nilai – nilai dari setiap atribut pada data untuk mengelompokkan data penyakit pasien.

### 2.2 Metode Analisis Data

Analisis data merupakan proses mencari dan menyusun secara sistematis data yang telah diperoleh dari wawancara, dokumentasi, dokumen pribadi, observasi, catatan lapangan, gambar foto dan sebagainya, dengan cara mengorganisasikan data tersebut ke dalam kategori, menjabarkan ke dalam unit- unit, melakukan sintesa, menyusun ke dalam pola, memilih mana yang penting dan mana yang akan dipelajari dan kemudian membuat kesimpulan agar dapat dipahami diri sendiri dan orang lain.

Dalam penulisan penelitian ini menggunakan analisis data yang bersifat kualitatif, penelitian kualitatif adalah analisis yang dilakukan dengan mengelompokkan data untuk mencari suatu pola dari hal yang dipelajari dan membandingkan konsep – konsep yang ada dalam sumber.

### 2.3 Studi Literatur

Dalam penelitian studi literatur adalah kegiatan ilmiah yang dilakukan untuk menemukan jawaban satu permasalahan, dan yang tujuan akhirnya adalah memberikan kontribusi teoritis atau praktis pada pengembangan bidang ilmu yang bersangkutan. Studi literatur yang digunakan disini meliputi pengolahan data penyakit pasien Puskesmas Warujayeng dan pemrograman.

$$(x + a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k} \dots\dots\dots(1)$$

### 2.4 Instrumen Penelitian

Dalam penelitian selain menggunakan data sebagai bahan penelitian juga diperlukan komponen pendukung seperti *software* dan *hardware* sebagai bahan pendukung berlangsungnya penelitian tersebut, komponen pendukung tersebut adalah:

#### 1) Kebutuhan Hardware

Kebutuhan perangkat keras (*hardware*) yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi Processor Intel core I3, Sistem Operasi Windows 7, RAM 2 GB dan HDD 500 GB.

#### 2) Kebutuhan Software

##### a. Sistem Operasi

Dalam penelitian ini sistem operasi yang dipakai adalah sistem operasi windows 7.

##### b. *Microsoft Word*

*Microsoft Word* dalam penelitian disini digunakan untuk menyusun laporan penelitian, *Microsoft Word* yang dipakai adalah versi 2012.

##### c. Xampp

XAMPP merupakan web server open source yang dapat berjalan pada sistem Windows, Linux, MacOS serta dapat dimanfaatkan untuk mengelola website seperti Apache, MySQL/MariaDB, PHP, dan Perl. XAMPP juga digunakan untuk membuat web server lokal di komputer.

##### d. MySQL

MySQL merupakan sebuah *database management system* (manajemen basis data) yang menggunakan perintah dasar SQL (*Structured Query Language*). MySQL disebut juga dengan database server.

#### 3) Sampel Data-set

Data set merupakan kumpulan data tabel dan juga di dalamnya terdapat relasi antar data tabel (*data relation*) atau lebih mudahnya di dalam satu dataset bisa terdapat banyak data tabel yang berelasi

#### 4) Variabel Penelitian

Pada sebuah penelitian data mining terdapat data yang akan diolah dengan metode yang telah ditentukan sebelumnya, pada penelitian ini data yang digunakan adalah data pasien dari Puskesmas

Warujayeng yang akan diolah menggunakan metode k-means untuk mengelompokkan data penyakit pasien tersebut kedalam kelompok penyakit “Akut” atau penyakit “Tidak Akut” berdasarkan beberapa variabel inputan. Variabel inputan yang digunakan dalam pengelompokkan penyakit pasien tersebut adalah nama, nomor ID, jenis kelamin, umur, kode penyakit, dan lama mengidap penyakit tersebut dalam hitungan bulan. Kemudian variabel tersebut akan diolah menggunakan metode k-means yang kemudian menghasilkan output kelompok penyakit berdasarkan hitungan dari metode k-means.

#### 5) Metode yang Diusulkan

K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara partisi yang memisahkan data ke dalam kelompok yang berbeda-beda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses.

### 2.5 Landasan Teori

#### 1) Data Mining

Data mining yang juga dikenal dengan istilah *pattern recognition* merupakan suatu metode yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang diolah. Data yang diolah dengan teknik data mining ini kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan di masa depan.

Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Dalam data mining juga terdapat metode – metode yang dapat digunakan seperti klasifikasi, clustering, regresi, seleksi variabel, dan market basket analisis.

Data mining juga bisa diartikan sebagai rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data – data tersebut dapat disimpan dalam database, data warehouse atau penyimpanan informasi. Ada beberapa ilmu yang mendukung teknik data mining diantaranya adalah data analisis, signal processing, neural network dan pengenalan pola.

#### 2) Clustering

Clustering atau pengklasteran adalah suatu teknik data mining yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokkan data atau lebih tepatnya mempartisi dari dataset ke dalam subset. Pada teknik

clustering targetnya adalah untuk kasus pendistribusian (objek, orang, peristiwa dan lainnya) ke dalam suatu kelompok, hingga derajat tingkat keterhubungan antar anggota cluster yang sama adalah kuat dan lemah antara anggota cluster yang berbeda [1].

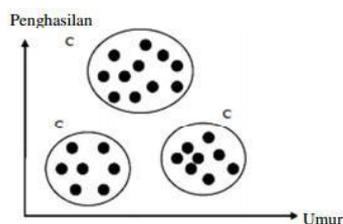
Teknik cluster mempunyai dua metode dalam pengelompokkannya yaitu hierarchical clustering dan non-hierarchical clustering. hierarchical clustering merupakan suatu metode pengelompokkan data yang cara kerjanya dengan mengelompokkan dua data atau lebih yang mempunyai kesamaan atau kemiripan, kemudian proses dilanjutkan ke objek lain yang memiliki kedekatan dua, proses ini terus berlangsung hingga cluster membentuk semacam tree dimana ada hirarki atau tingkatan yang jelas antar objek dari yang paling mirip hingga yang paling tidak mirip. Namun secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster [2].

Sedangkan non-hierarchical clustering pada teknik ini dimulai dengan menentukan jumlah cluster yang diinginkan (dua cluster, tiga cluster, empat cluster atau lebih), setelah jumlah yang cluster yang diinginkan maka proses cluster dimulai tanpa mengikuti proses hirarki, metode ini juga sering disebut sebagai metode K-Means Clustering [2].

Pada proses analisis cluster metode yang digunakan untuk membagi data menjadi subset data berdasarkan kesamaan atau kemiripan yang telah ditentukan sebelumnya. Jadi analisis cluster secara umum dapat dikatakan bahwa:

- Data yang terdapat dalam satu cluster memiliki tingkat kesamaan yang tinggi.
- Dan yang terdapat dalam suatu cluster yang berbeda memiliki tingkat kesamaan yang rendah.

Sebagai contoh dapat dilihat pada gambar 1 dibawah ini :



Gambar 1. Grafik Clustering

Pada gambar 1 dapat dilihat kita misalkan data tersebut merupakan data konsumen sederhana yang terdapat dua atribut didalamnya, yaitu umur dan penghasilan. Pada data yang berdasarkan dua atribut tersebut kemudian dibagi menjadi tiga cluster yaitu cluster C1 yang terdiri dari konsumen usia muda dan berpenghasilan rendah, cluster C2 terdiri dari

konsumen usia muda dan tua berpenghasilan tinggi, dan cluster C3 terdiri dari konsumen usia tua dan berpenghasilan relatif rendah.

### 3) K-Means

K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – berda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses [1].

Sehingga dalam penggunaan algoritma k-means terdapat aturan sebagai berikut [1]:

- Berapa jumlah *cluster* yang perlu dimasukkan.
- Hanya memiliki atribut bertipe *numeric*.

Pada dasarnya algoritma k-means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat *cluster* awal, pada penentuan pusat *cluster* ini dipilih secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing – masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat *cluster*. Selanjutnya posisi pusat *cluster* akan dihitung kembali samapi semua komponen data digolongkan ke dalam tiap – tiap *cluster* dan terakhir akan terbentuk *cluster* baru [1].

Algoritma K-Means pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat *cluster* dan proses pencarian anggota dari tiap-tiap *cluster*. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster,  $X_{ij}$  ( $i=1, \dots, n$ ;  $j=1, \dots, m$ ) dengan  $n$  adalah jumlah data yang akan dikluster dan  $m$  adalah jumlah variabel. Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang),  $C_{kj}$  ( $k=1, \dots, k$ ;  $j=1, \dots, m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean. Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak data tersebut ke pusat *cluster* ke- $k$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lain.

Proses dasar algoritma k-means dapat dilihat di bawah ini :

- Tentukan jumlah *cluster* yang ingin dibentuk dan tetapkan pusat *cluster*  $k$ .
- Menggunakan jarak *euclidean* kemudian hitung setiap data ke pusat *cluster*.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots \dots \dots (1)$$

- c. Kelompokkan data ke dalam cluster dengan jarak yang paling pendek dengan persamaan

$$\text{Min } \sum_k^k = d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots(2)$$

- d. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \dots\dots\dots(3)$$

Dengan :  
 $x_{ij} \in$  Kluster ke - k  
 $p$  = banyaknya anggota kluster ke - k

- e. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

### 2.6 Kerangka Penelitian

Kerangka pemikiran merupakan garis besar dari langkah – langkah penelitian yang sedang dilakukan, kerangka pemikiran dijadikan acuan untuk melakukan tahap – tahap yang sedang dilakukan dalam penelitian

Permasal		
Data penyakit pasien Puskesmas Warujayang hanya menjadi tumpukan statistik yang belum digali informasinya. Mengatahui keakuratan algoritma k-means dalam pengelompokkan data penyakit pasien Puskesmas Warujayang.		
T		
Mengelompokkan data penyakit yang diderita oleh pasien. Menganalisa keakuratan penggunaan algoritma k-means untuk pengelompokkan penyakit pada Puskesmas Warujayang		
Eksper		
Inputan	Metode	Implementasi
Data pasien Puskesmas Warujayang	K – Means	Aplikasi Berbasis Web
H		
Menghasilkan sistem yang mampu mengelompokkan data penyakit pasien pada Puskesmas Warujayang dengan menggunakan algoritma k-means		
Ma		
Membantu pihak Puskesmas dalam mengetahui jenis penyakit apa yang menjadi peringkat pertama yang diderita oleh pasien Puskesmas Warujayang, dan juga bisa menjadi bahan acuan dalam memonitoring dan pemberian penyuluhan Kesehatan kepada masyarakat sekitar		

Gambar 2. Kerangka Pemikiran

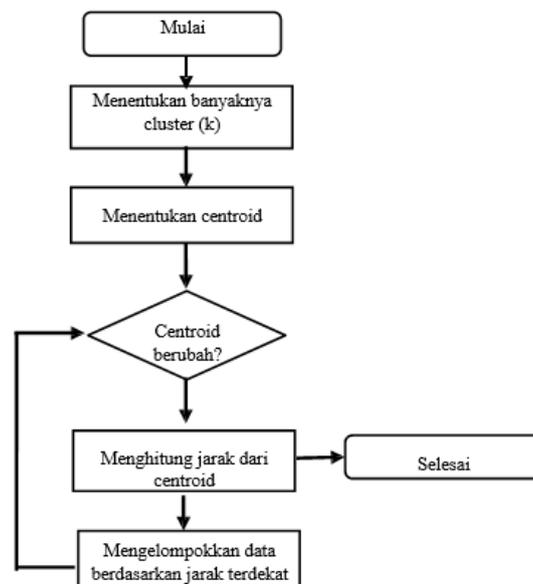
## 3. HASIL DAN PEMBAHASAN

### 3.1. Implementasi Algoritma K-Means

Dapat dilihat pada gambar 3 di bawah merupakan diagram alur dari metode k-means yang digunakan dalam pengelompokkan penyakit di Puskesmas Warujayang, pada umumnya kinerja

metode k-means secara berurutan adalah sebagai berikut :

- 1) Menentukan banyaknya cluster (k)
- 2) Menentukan centroid
- 3) Apakah nilai centroidnya berubah?
  - a. Jika ya, hitung jarak data dari centroid
  - b. Jika tidak, selesai.
- 4) Mengelompokkan data berdasarkan jarak terdekat



Gambar 3. Alur Implementasi Algoritma K-Means

Data penelitian yang sedang dilakukan merupakan data penyakit pasien Puskesmas Warujayang sebanyak 1000 data yang akan dikelompokkan ke dalam penyakit “AKUT(C1)” dan penyakit “TIDAK AKUT(C2)” pengelompokkan tersebut berdasarkan atribut umur, kode penyakit dan lama mengidap penyakit, yang kemudian atribut tersebut akan diolah menggunakan algoritma k-means. Sampel dari data penyakit pasien Puskesmas Warujayang dapat dilihat pada tabel 1 di bawah ini.

Tabel 1. Sampel Data Penyakit Pasien Puskesmas Warujayang

Data ke-	Jenis Kelamin	Umur (th)	Kode Diagnosa	Lama Mengidap (bln)	Diagnosa
1.	Perempuan	25	16	3	Demam tifoid dan paratifoid
2.	Perempuan	25	16	5	Demam tifoid dan paratifoid
3.	Laki-laki	15	18	3	Amoebiasis
4.	Laki-laki	11	18	2	Amoebiasis
5.	Laki-laki	17	21	2	Diare dan Gastroenteritis
6.	Laki-laki	12	21	3	Diare dan Gastroenteritis

7.	Laki-laki	10	21	5	Diare dan Gastroenteritis
8.	Laki-laki	14	21	1	Diare dan Gastroenteritis
9.	Laki-laki	10	21	4	Diare dan Gastroenteritis
10.	Laki-laki	12	21	5	Diare dan Gastroenteritis
11.	Laki-laki	3	21	1	Diare dan Gastroenteritis
12.	Perempuan	21	19	4	TB paru BTA
13.	Laki-laki	10	17	8	TB selain paru
14.	Laki-laki	51	20	8	Kusta
15.	Perempuan	18	20	9	Kusta
16.	Perempuan	18	20	6	Kusta
17.	Perempuan	27	15	6	Tetanus obstetric
18.	Laki-laki	20	22	11	Scabies
19.	Laki-laki	15	24	21	Anemia
20.	Laki-laki	7	23	8	Campak

Iterasi Ke-1

1) Penentuan pusat awal cluster

Tabel 2. Titik Pusat Awal Cluster

Data ke-	Umur	Kode penyakit	Lama mengidap
8	14	21	1
11	3	21	1

2) Perhitungan jarak pusat cluster

Untuk menghitung jarak antara data dengan pusat awal cluster menggunakan persamaan euclidean distace sebagai berikut :

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots (3)$$

Dimana :

$C_{ik}$  : pusat cluster

$C_{kj}$  : data

Maka akan didapatkan nilai matrik jarak sebagai berikut :

Jarak Data ke-1 ke pusat cluster

$$C1 = \sqrt{(14 - 25)^2 + (21 - 16)^2 + (1 - 3)^2}$$

$$C1 = 12,2474487$$

$$C2 = \sqrt{(3 - 25)^2 + (21 - 16)^2 + (1 - 3)^2}$$

$$C2 = 22,6495033$$

Jarak Data ke-2 ke pusat cluster

$$C1 = \sqrt{(14 - 25)^2 + (21 - 16)^2 + (1 - 5)^2}$$

$$C1 = 12,7279221$$

$$C2 = \sqrt{(3 - 25)^2 + (21 - 16)^2 + (1 - 5)^2}$$

$$C2 = 22,9128785$$

Jarak Data ke-3 ke pusat cluster

$$C1 = \sqrt{(14 - 15)^2 + (21 - 18)^2 + (1 - 3)^2}$$

$$C1 = 3,741657387$$

$$C2 = \sqrt{(3 - 15)^2 + (21 - 18)^2 + (1 - 3)^2}$$

$$C2 = 12,52996409$$

Jarak Data ke-4 ke pusat cluster

$$C1 = \sqrt{(14 - 11)^2 + (21 - 18)^2 + (1 - 2)^2}$$

$$C1 = 4,358898944$$

$$C2 = \sqrt{(3 - 11)^2 + (21 - 18)^2 + (1 - 2)^2}$$

$$C2 = 8,602325267$$

Jarak Data ke-5 ke pusat cluster

$$C1 = \sqrt{(14 - 17)^2 + (21 - 21)^2 + (1 - 2)^2}$$

$$C1 = 3,16227766$$

$$C2 = \sqrt{(3 - 17)^2 + (21 - 21)^2 + (1 - 2)^2}$$

$$C2 = 14,03566885$$

Dan seterusnya dilanjutkan menghitung untuk data ke-6.....N terhadap pusat awal cluster hingga didapatkan matrik jarak.

3) Pengelompokkan data

Jarak hasil perhitungan pada point ke-2 akan dilakukan perbandingan dan dipilih jarak yang paling dekat antara data dengan pusat cluster, jarak ini akan menunjukkan bahwa data yang memiliki jarak terdekat berada dalam satu kelompok dengan pusat cluster terdekat, pengelompokkan data tersebut dapat dilihat pada tabel 3 di bawah ini, nilai 1 berarti data tersebut berada dalam kelompok.

Tabel 3 Pengelompokkan Data Berdasarkan Cluster Terdekat

Data Ke-	C1	C2
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	1	
10	1	
11		1
12	1	
13	1	
14	1	
15	1	
16	1	
17	1	
18	1	
19	1	
20		1

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1: data 1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18 dan 19  
C2: data 11 dan data 20

#### 4) Penentuan pusat *cluster* baru

Setelah didapatkan anggota dari setiap *cluster* kemudian pusat *cluster* baru dihitung berdasarkan data anggota tiap – tiap *cluster* yang sudah didapatkan menggunakan rumus yang sesuai dengan pusat anggota *cluster* sebagai berikut :

$$C1 = \left( \frac{25 + 25 + 15 + 11 + 17 + 12 + 10 + 14 + 10 + 12 + 21 + 10 + 51 + 18 + 18 + 27 + 20 + 15}{18}; \frac{16 + 16 + 18 + 18 + 21 + 21 + 21 + 21 + 21 + 19 + 17 + 20 + 20 + 20 + 15 + 22 + 24}{18}; \frac{3 + 5 + 3 + 2 + 2 + 3 + 5 + 1 + 4 + 5 + 4 + 8 + 8 + 9 + 6 + 6 + 11 + 21}{18} \right)$$

$$C1 = (18,4; 19,5; 5,9)$$

$$C2 = \left( \frac{7 + 3}{2}; \frac{21 + 23}{2}; \frac{1 + 8}{2} \right)$$

$$C2 = (5; 22; 4,5)$$

Dari perhitungan di atas maka didapatkan pusat cluster baru dalam matrik tabel sebagai berikut:

Tabel 4. Pusat *Cluster* Baru

C1	18,4	19,5	5,9
C2	5	22	4,5

Iterasi selanjutnya dilakukan dengan cara yang sama hingga tidak ada perubahan data dalam suatu cluster.

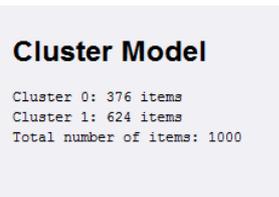
### 3.2. Implementasi Rapid Miner

Berikut adalah pengolahan data dengan menggunakan *k-means* pada *RapidMiner* :



Gambar 4. Pemodelan Clustering K-Means pada Rapid Miner

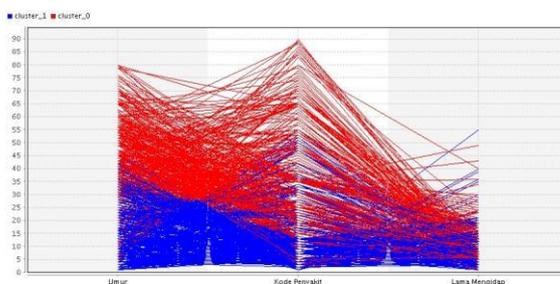
Dengan menggunakan pemodelan *k-means clustering* seperti gambar 4 diatas, dengan inialisasi jumlah *cluster* sebanyak 2 buah, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 2, sesuai dengan pendefinisian nilai k dengan jumlah *cluster\_0* ada 376 item, *cluster\_1* ada 624 item dengan total jumlah data adalah 1000.



Gambar 5. Hasil data cluster K-Means dalam implementasi Rapid Miner

Attribute	cluster_0	cluster_1
Umur	45.691	17.277
Kode Penyakit	34.324	12.199
Lama Mengidap	9.633	7.362

Gambar 6. Hasil perhitungan antara jarak cluster dan centroid



Gambar 7. Grafik data hasil Clustering K-Means

## 4. SIMPULAN

Dalam perangkaan kesehatan pada suatu daerah masih menggunakan cara manual yaitu perhitungan yang masih menggunakan rata – rata seluruh hasil indikator atau didasarkan pada distribusi data pada setiap daerah, pengolahan data juga masih menggunakan teknik statistik dasar, ini menghasilkan output yang kurang maksimal dan memiliki permasalahan pada konsistensi data pada setiap Dinas Kesehatan[2]. Tumpukan data yang berada pada dinas kesehatan, poliklinik maupun rumah sakit dan puskesmas.

Berdasarkan uraian masalah di atas, maka dapat dirumuskan kesimpulan inialisasi jumlah *cluster* sebanyak 2 buah sesuai dengan pendefinisian nilai k dengan jumlah *cluster* akut ada 376 item, *cluster* tidak akut ada 624 item dengan total jumlah data adalah 1000.

Untuk menentukan konsistensi data kesehatan dapat digunakan teknik data mining yang mampu menggali informasi tersembunyi dari kumpulan data multidimensi yang telah diperoleh, selain itu pengekstrakan data yang terhubung dengan data lain juga dapat dilakukan oleh teknik data mining ini. Salah satu teknik data mining yang cukup terkenal yaitu *clustering*

dan metode yang cukup populer dalam teknik data mining ini adalah metode k-means.

## 5. SARAN

Beberapa Saran pengembangan penelitian antara lain:

- 1) Pada penelitian ini dapat dikembangkan menjadi aplikasi maupun dibuat menggunakan bahasa pemrograman.
- 2) Penelitian ini dapat dikembangkan menggunakan perbandingan algoritma lainnya.

## DAFTAR PUSTAKA

- [1] Wallace, V. P., Bamber, J. C. dan Crawford, D. C. 2000. Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminate analysis and artificial neural network. *Journal Physical Medical Biology*. No.45. Vol.3. 2859-2871.
- [2] Yusoff, M. Rahman, S.,A., Mutalib, S., and Mohammed, A. 2006 Diagnosing Application Development for Skin Disease Using Backpropagation Neural Network Technique. *Journal of Information Technology*. vol 18. hal 152-159.