

IMPLEMENTASI ALGORITMA K-MEANS UNTUK PENGELOMPOKKAN PENYAKIT PASIEN PADA PUSKESMAS KAJEN PEKALONGAN

(K-MEANS ALGORITHM IMPLEMENTATION FOR CLUSTERING OF PATIENTS DISEASE IN KAJEN CLINIC OF PEKALONGAN)

Anindya Khrisna Wardhani
Magister Sistem Informasi
Universitas Diponegoro
nindywardhani77@gmail.com

Abstract - In determining the consistency of health data, can use data mining techniques that can dig the hidden information from multidimensional data sets that have been obtained. In addition, data wick connected with other data can also be done by these data mining techniques. One of the data mining techniques is quite well known namely clustering. The methods are quite popular in data mining techniques that called k-means method. It is used to facilitate medical recorder for analyzing the general health situation of population groups in archiving health care data. The results of this analysis, the clustering of the disease based on age, sex, duration of disease and disease diagnosis. This research used tool Rapid Miner 5.3. Based on the data from clinic centers Kajen Pekalongan, the result of clustering is 376 items of acute and 624 unacute diseases from 1000 total of data.

Keywords : Data Mining, K-Means, Clustering

I. PENDAHULUAN

Clustering adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengolompokkan 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*[4].

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[5].

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

II. LANDASAN TEORI

A. 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[4].

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[4].

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[10].

B. Clustering

Clustering atau pengklasteran adalah suatu teknik data mining yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokan 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[8].

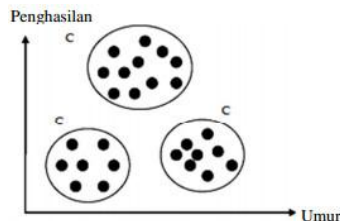
Teknik *cluster* mempunyai dua metode dalam pengelompokannya yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *hierarchical clustering* merupakan suatu metode pengelompokan data yang cara kerjanya dengan mengelompokkan dua data atau lebih yang mempunyai kesamaan atau kemiripan, kemudian 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*[4].

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 diinginkan maka proses *cluster* dimulai tanpa mengikuti proses hirarki, metode ini juga sering disebut sebagai metode K-Means *Clustering*[4].

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 [8]:

- Data yang terdapat dalam satu *cluster* memiliki tingkat kesamaan yang tinggi, dan
- 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.

C. K-Means

K-Means merupakan suatu algoritma yang digunakan dalam pengelompokan 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[8].

Sehingga dalam penggunaan algoritma k-means terdapat aturan sebagai berikut [8] :

- 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[8].

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 [8].

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

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

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots \dots \dots (2.1)$$

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

$$\text{Min } \sum_k^k = d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots \dots \dots (2.2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \dots \dots \dots (2.3)$$

Dengan :

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

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

D. Kerangka Pemikiran

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

| Permasalahan | | |
|---|-----------|---------------|
| Data penyakit pasien Puskesmas Kejen hanya menjadi tumpukan statistik yang belum digali informasinya. Mengetahui keakuratan algoritma k-means dalam pengelompokan data penyakit pasien Puskesmas Kajen Pekalongan. | | |
| Tujuan | | |
| Mengelompokkan data penyakit yang diderita oleh pasien. Menganalisa keakuratan penggunaan algoritma k-means untuk pengelompokan penyakit pada Puskesmas Kajen Pekalongan | | |
| Eksperimen | | |
| Inputan | Metode | Implementasi |
| Data pasien Puskesmas Kajen Pekalongan | K – Means | Java Netbeans |
| Hasil | | |
| Menghasilkan sistem yang mampu mengelompokkan data penyakit pasien pada Puskesmas Kajen Pekalongan dengan menggunakan algoritma k-means | | |
| Manfaat | | |
| Membantu pihak Puskesmas dalam mengetahui jenis penyakit apa yang menjadi peringkat pertama yang diderita oleh pasien Puskesmas Kajen Pekalongan, dan juga bisa menjadi bahan acuan dalam memonitoring dan pemberian penyuluhan kesehatan kepada masyarakat sekitar | | |

dilakukan dalam penelitian.

Gambar 2 Kerangka Pemikiran

III. METODE PENELITIAN

A. 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 Kajen Pekalongan dimana penelitian 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.

B. 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.

C. 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 Kajen Pekalongan dan pemrograman.

D. 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*

Kebutuhan perangkat lunak (*software*) merupakan salah satu faktor yang sangat penting dalam penyusunan penelitian ini, perangkat lunak yang dipakai adalah :

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. Java Netbeans

Netbeans merupakan aplikasi berbasis desktop yang dalam penelitian ini digunakan untuk memproses data – data inputan.

d. MySQL

Perangkat lunak basis data yang dalam penelitian ini digunakan untuk menampung data – data dan nilai dari setiap atribut.

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 Kajen Pekalongan 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 pengelompokan 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

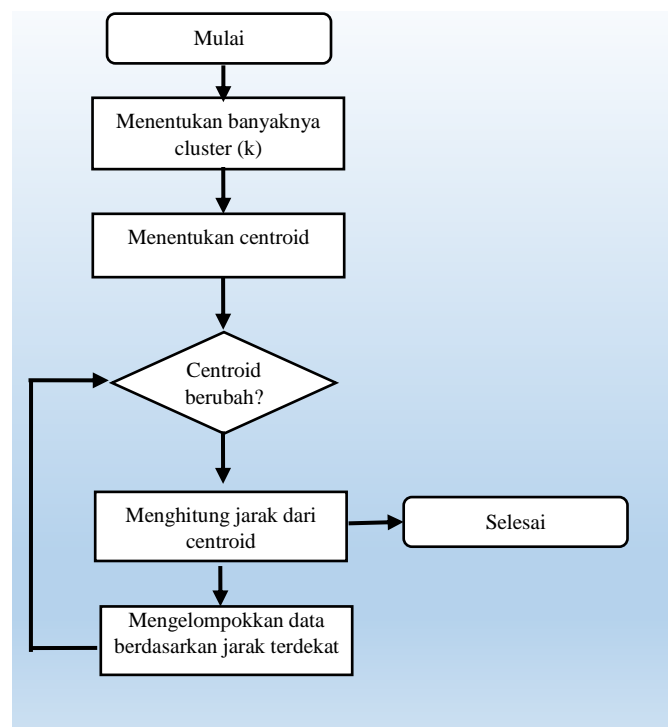
Means merupakan suatu algoritma yang digunakan dalam pengelompokan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – berda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang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

IV. HASIL PENELITIAN DAN PEMBAHASAN

A. Implementasi Algoritma K-Means

Dapat dilihat pada gambar 3 di bawah merupakan diagram alur dari metode k-means yang digunakan dalam pengelompokan penyakit di Puskesmas Kajen Pekalongan, 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 Kajian Pekalongan 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 Kajian Pekalongan dapat dilihat pada tabel 1 di bawah ini.

Tabel 1 Sample Data Penyakit Pasien Puskesmas Kajian Pekalongan

| Data ke- | Jenis Kelamin | Umur (th) | Kode Diagnose | Lama Mengidap (bln) | Diagnose |
|----------|---------------|-----------|---------------|---------------------|-----------------------------|
| 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 *eucledian distace* sebagai berikut :

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

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. Pengelompokan 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, pengelompokan data tersebut dapat dilihat pada tabel 3 di bawah ini, nilai 1 berarti data tersebut berada dalam kelompok.

Tabel 3 Pengelompokan Data Berdasarkan *Cluster* Terdekat

G1 =

| 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 :

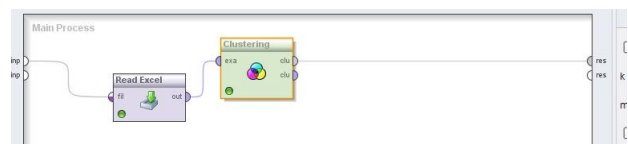
Tabel 4 Pusat *Cluster* Baru

| 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

B. 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.

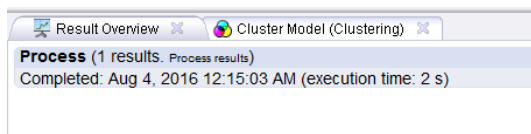
Cluster Model

Cluster 0: 376 items
 Cluster 1: 624 items
 Total number of items: 1000

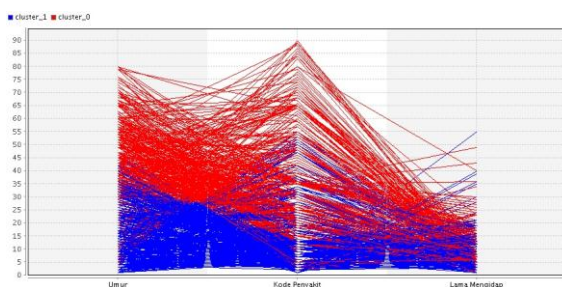
Gambar 5 Hasil data cluster K-Means dalam implementasi Rapid Miner

| 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 Result Overview



Gambar 8 Grafik data hasil Clustering K-Means

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam perangkan 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 indikator 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 inisialisasi jumlah *cluster* sebanyak 2 buah sesuai dengan pendefinisian nilai *k* dengan jumlah *cluster* akut ada 376 item, *cluster* tidak 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 pengestrakan 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.

5.2 Saran

- Pada penelitian ini dapat dikembangkan menjadi aplikasi maupun dibuat menggunakan bahasa pemrograman.
- Untuk pengembangan penelitian selanjutnya dapat dikembangkan menggunakan perbandingan algoritma lainnya.

DAFTAR PUSTAKA

- [1] M. M. Putri dan K. Fithriasari, "Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Metode Kohonen SOM dan K-Means," *Jurnal Sains dan Seni ITS*, vol. 4, no. 1, pp. D13 - D18, 2015.
- [2] N. Atthina dan L. Iswari, "Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat kesehatan Daerah dengan Metode K-Means," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Vol. %1 dari %2ISSN 1907 - 5022, pp. B52 - B59, 2014.
- [3] V. Handayani, A. dan A. P. kurniati, "Analisa Clustering Menggunakan Algoritma K-Modes," *Telkom University*, pp. 1-8, 2010.
- [4] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10 - 20, 2013.
- [5] Y. Ardilla, H. Tjandra dan I. Ariesanti, "Deteksi Penyakit Epilepsi dengan Menggunakan Entropi Permutasi, K-Means Clustering, dan Multilayer Perceptron," *Jurnal Teknik POMITS*, vol. 3, no. 1, pp. A70 - A74, 2014.
- [6] M. E. Putra, "Implementasi Algoritma K-Means Pada Pendeteksian Warna Untuk Membantu Penderita Buta Warna," *Jurnal Ilmiah komputer dan Informatika (KOMPUTA)*, Vol. %1 dari %2ISSN 2089 - 9033, pp. 1-8, 2012.
- [7] S. P. Tulus dan Hendry, "Supprot Vektor Machines Yang Didukung K-Means Clustering Dalam Klasifikasi Dokumen," *Jurnal Teknologi Informasi-Aiti*, vol. 11, no. 2, pp. 101-202, 2014.

- [8] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana dan F. Khusnu, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," Universitas Brawijaya Malang, Malang, 2012.
- [9] F. E. Agustin, A. Fitria dan A. H. S, "Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus : SMP Negeri 101 Jakarta)," *Jurnal Teknik Informatika*, vol. 8, no. 1, pp. 73 - 78, 2015.
- [10] K. R. Prilianti dan H. Wijaya, "Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering," *Jurnal Cybermatika*, vol. 2, no. 1, pp. 1-6, 2014.
- [11] A. Yusuf dan T. Priambadha, "Supprot Vektor Machines Yang Didukung K-Means Clustering Dalam Klasifikasi Dokumen," *JUTI*, vol. 11, no. 1, pp. 13-16, 2013.
- [12] N. A, S. B dan P. U, "Implementasi Naive Bayes Classifier Pada Program Bantu Penentuan Buku Referensi Mata Kuliah," *Jurnal Informatika*, vol. 3, no. 2, pp. 32-36, 2007.
- [13] D. Septiari, "Implementasi Metode Naive Bayes Classifier Dalam Klasifikasi Kelayakan Calon Pendorong Darah (Studi Kasus : PMI Kabupaten Demak)," Universitas Dian Nuswantoro, Semarang, 2016.
- [14] Budiono, a. Fahmi dan Pujiono, "Penerapan Metode Association Rule Discovery Menggunakan Algoritma Apriori Untuk Mengidentifikasi Pola Penyakit Radang Sendi," *Techno.COM*, vol. 13, no. 2, pp. 115-124, 2014.
- [15] Supardi, D. E. Ratnawati dan W. F. Mahmudy, "Pengenalan Pola Transaksi Sirkulasi Buku Pada Database Perpustakaan Menggunakan Algoritma Generalized Sequential Pattern," *Jurnal Mahasiswa PTIIK Universitas Brawijaya*, vol. 4, no. 11, pp. 1-8, 2014.
- [16] R. F. Jannah, "Rancang Bangun Sistem Hasil Produksi Dengan Memanfaatkan Metode Least Square Regression Line (Studi Kasus : Toko Keeava Salad & Puding)," Universitas Jember, Jember, 2015.
- [17] N. P. E. Merliana, Ernawati dan J. Santoso, "Analisis Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering," dalam *Prosiding Seminar Nasional Multi Disiplin Ilmu dan Call For Papers Unisbank (SENDI_U)*, Yogyakarta, 2014.