

IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENGETAHUI WILAYAH RENTAN PENYAKIT MENGGUNAKAN DATA RL RAWAT INAP RSUP DR.KARIADI SEMARANG

Nurfaizin¹, Heru Lestiawan, M.Kom.²

¹Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
email : 1111201106442@mhs.dinus.ac.id

²Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
email : herul.at.dosen.dinus.ac.id

Abstrak

Implementasi data mining dapat memberikan informasi tentang daerah endemik penyakit dengan jelas, cepat, dan akurat baik dalam bentuk aplikasi. Melalui data penyakit rawat inap dapat dikelola untuk mencari pola daerah endemi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Algoritma K-Means adalah algoritma pengelompokan yang dipilih untuk pengelolaan data sehingga informasi yang dibutuhkan dapat terpenuhi. Pada tahap clustering dengan menggunakan K-Means ini dimulai dengan pembentukan cluster, pembagian cluster yang akan dipilih secara random. Dalam tugas akhir ini, penulis membuat implementasi persebaran pasien rawat inap dalam distribusi di wilayah Kecamatan Semarang, informasi yang ditampilkan termasuk data rawat inap di Kecamatan distribusi di RSUP Dr.Kariadi Semarang, informasi Penyakit, dan informasi fasilitas kesehatan. Dengan adanya aplikasi ini dapat membantu pemantauan daerah yang terkena penyakit dan memberikan informasi yang lengkap dan aktual.

Kata Kunci: Clustering, K-means, Rawat inap, Data mining

Abstract

Implementation of data mining can provide information about the area endemic disease clearly, quickly, and accurately both in the application form. Through the data inpatient disease can be managed to find patterns of endemic areas. Data mining is a mining or the discovery of new information by looking for patterns or particular rules of a number of large amounts of data are expected to treat the condition. K-Means algorithm is a clustering algorithm selected for the management of data so that the information needed can be met. At this stage of clustering using K-Means begins with the formation of clusters, cluster division which will be chosen at random. In this thesis, the author makes the implementation of the distribution of inpatients in the distribution in the District of Semarang, the information displayed includes the data of hospitalization in the District RSUP Dr.Kariadi distribution in Semarang, disease information, and information about health facilities. With this application can help to monitor the areas affected by the disease and provide information that is complete and actual.

Keywords: Clustering, K-means, inpatient, Data mining

1. PENDAHULUAN

Rumah sakit adalah salah satu sarana kesehatan tempat menyelenggarakan upaya kesehatan dengan memberdayakan berbagai kesatuan

personel terlatih dan terdidik dalam menghadapi dan menangani masalah medik untuk pemulihan dan pemeliharaan kesehatan yang baik (Siregar, 2004).[1]

Upaya kesehatan adalah setiap kegiatan

untuk memelihara dan meningkatkan kesehatan yang bertujuan untuk mewujudkan derajat kesehatan yang optimal bagi masyarakat dan tempat yang digunakan untuk menyelenggarakannya disebut sarana kesehatan. Sarana kesehatan berfungsi melakukan upaya kesehatan dasar, kesehatan rujukan dan atau upaya kesehatan penunjang. Upaya kesehatan diselenggarakan dengan pendekatan pemeliharaan, peningkatan kesehatan (promotif), pencegahan penyakit (preventif), penyembuhan penyakit (kuratif) dan pemulihan kesehatan (rehabilitatif) yang diselenggarakan secara menyeluruh, terpadu dan berkesinambungan (Siregar, 2004).[1]

Perkembangan teknologi informasi komunikasi dewasa ini telah dianggap telah membuka wawasan dan jawaban alternatif bagi masyarakat dalam mendapatkan suatu kesimpulan yang membantu menentukan keputusan. Secara faktual dibutuhkan sebuah tool yang dapat bekerja memodelkan pola penularan penyakit rawat inap supaya dapat memberikan informasi untuk menanggulangi atau menindak lanjuti penyakit endemi yang paling dominan di daerah tertentu dengan cara menambahkan puskesmas atau apotek di daerah tersebut. K-means adalah algoritma yang dapat digunakan untuk membantu menganalisa kondisi suatu daerah terhadap penyakit untuk menentukan tindakan yang harus dilakukan untuk menangani penyakit. Dengan adanya sistem informasi ini diharapkan pihak RSUP khususnya bagian rawat inap bisa mengatasi masalah yang terjadi saat ini dan mampu memberikan informasi pengambilan keputusan terhadap daerah endemi. Sehingga kegiatan karyawan menjadi efektif dan efisien. Hal ini berperan sebagai suatu media untuk mengetahui, mengawasi, dan menyiapkan strategi preventive penyakit endemi di kota tersebut.

Kemudian hasilnya dapat digunakan sebagai suatu acuan ilmiah untuk tindakan berikutnya dalam upaya pencegahan kasus yang hampir sama. Berdasarkan permasalahan yang terjadi, dengan memanfaatkan teknik data mining metode K-Means, penulis ingin mengolah dataset yang akan digunakan untuk mengolah data tersebut sehingga menjadi informasi untuk memberikan pelayanan ke depannya.

2. METODE

Dari beberapa jurnal yang telah dibaca, akan digunakan sebagai acuan literature yang memiliki kesamaan konsep data mining untuk mengatasi suatu permasalahan. Maka penelitian ini akan menggunakan teknik klastering metode k-means untuk mengolah data RL rawat inap RSUP Dr.Kariadi Semarang. Penelitian ini juga menganalisa kemungkinan untuk membuat cluster otomatis (*automatic clustering*).

2.1 Metode Pengumpulan Data

A. Jenis Data

Jenis data yang akan digunakan untuk penelitian ini adalah data kuantitatif. Data kuantitatif adalah data yang berbentuk angka atau bilangan. Sesuai dengan bentuknya, data kuantitatif dapat diolah atau dianalisis menggunakan teknik perhitungan matematika atau statistika. Data kuantitatif berfungsi untuk mengetahui jumlah atau besaran dari sebuah objek yang akan diteliti. Data yang digunakan dalam penelitian ini berupa data persebaran penyakit menular pada bulan Januari sampai dengan bulan Februari 2015. Dari data tersebut hanya diambil data yang berkaitan dengan atribut yang dibutuhkan.

B. Sumber Data

Sumber data yang digunakan dalam penelitian ini ada dua, yaitu:

1. Data Primer

Data Primer adalah data yang berasal dari sumber asli atau pertama. Pada penelitian ini, data primer adalah data pasien rawat inap RSUP Dr.Kariadi Semarang pada bulan Januari sampai dengan Februari tahun 2015.

2. Data Sekunder

Data sekunder merupakan hasil pengolahan lanjut dari data primer yang ditampilkan dalam bentuk lain dan digunakan untuk mendukung data primer. Dalam penelitian ini data sekunder diperoleh dari sebaran pasien. Peneliti menggunakan data sekunder ini untuk memperkuat data primer dan melengkapi informasi yang telah dikumpulkan melalui observasi langsung.

C. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini merupakan teknik pengumpulan observasi dan dokumentasi. Data pasien rawat inap pada bulan Januari sampai dengan Februari 2015 akan dianalisa untuk memperoleh sebuah informasi baru yang akan digunakan untuk mengelompokkan alamat berdasarkan dataset dengan teknik clustering menggunakan algoritma k-means.

D. Metode Prototype

Dalam pelaksanaannya, sistem informasi yang diperoleh pada data RSUP Dr.Kariadi kota Semarang dirasa belum optimal, hal ini dikarenakan sistem yang digunakan masih bersifat manual. Dengan permasalahan tersebut maka muncul berbagai permasalahan terutama pada daerah yang paling rentan terkena penyakit pada daerah endemi. Untuk itu, diperlukan suatu sistem informasi yang mampu mendukung pengambilan keputusan dalam memperoleh informasi kegiatan medis. Pembuatan Sistem Informasi medis RSUP Dr.Kariadi Semarang

menggunakan pendekatan terstruktur, sedangkan metode pengembangan menggunakan prototype dengan teknik pengumpulan data observasi dan dokumentasi, sedangkan alat yang digunakan dalam merancang sistem berupa pengembangan aplikasi berbasis desktop. Sistem yang dibangun diharapkan dapat mengatasi sebagian besar permasalahan yang ada seperti mengimplementasikan pola pada daerah endemi.

2.2 Analisa Data Mining

Dalam penulisan tugas akhir ini akan dicari atribut yang memiliki hubungan antara daerah endemi yang dibutuhkan untuk mengelompokkan data pasien rawat inap Semarang. Tidak semua data pasien rawat inap dicari hubungannya dengan daerah endemi yang dibutuhkan untuk mengelompokkan alamat, hanya beberapa atribut yang berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses *mining* memakan waktu lama dan tingkat hubungannya pun rendah. Data pasien yang akan dicari hubungannya meliputi hubungan atribut alamat pasien dengan daerah endemi yang dibutuhkan untuk mengimplementasikannya.

Proses lengkap analisa data mining pada penelitian ini menggunakan tahapan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dengan enam fase berikut ini :

1. Fase Pemahaman Bisnis

Tujuan dilakukannya penelitian ini adalah untuk mengetahui pola yang dibutuhkan untuk mengimplementasikan data mining berdasarkan data pasien rawat inap Semarang. *Data set* pada bulan Januari sampai dengan Februari tahun 2015 merupakan data yang akan dianalisa untuk memperoleh informasi yang dibutuhkan. *Data set* akan diolah dengan cara data mining menggunakan teknik *clustering*

metode k-means untuk memudahkan dalam pengelompokkan daerah endemi yang dibutuhkan dalam membuat sistem informasi.

2. Fase Pemahaman Data

Fase pemahaman data adalah menentukan data apa yang akan diambil dan diolah untuk mencapai tujuan yang telah ditentukan. Data yang menjadi training pada metode clustering dengan algoritma k-means ini adalah data pasien rawat inap ditahun 2015 RSUP Dr.Kariadi Semarang. Dari survey yang dilakukan peneliti, didapatkan data bulan Januari sampai Februari adalah 295 data, jadi total data mentah yang didapat peneliti untuk training ini sebanyak 295 data.

Gambar 2.1 Data Mentah

3. Fase Pengolahan Data

Pada fase ini, data mentah yang telah didapat tidak semuanya dapat digunakan karena masih ada data yang mengandung missing value (keterangan tidak lengkap). Oleh karena itu, harus dilakukan preprocessing yaitu cleaning data dan selection data. Cleaning data (pembersihan data) dan selection data merupakan tahap awal dari preprocessing data mining. Pembersihan ini dilakukan dengan membuang data yang mempunyai informasi tidak lengkap.

No	Jenis Kelamin	UmurTh	Kecamatan
1	P	52	Genuk
2	L	35	Rowosari
3	L	23	Semarang Selatan
4	L	62	Genuk
5	P	15	Semarang Barat
6	P	36	Gayamsari
7	P	9	Getasan
8	L	11	Banyumanik
9	P	9	Semarang Barat
10	L	23	Candisari
11	L	79	Semarang Selatan
12	L	6	Semarang Selatan
13	P	28	Gunung Pati
14	P	36	Genuk
15	L	78	Semarang Barat
16	P	25	Semarang Barat
17	L	2	Ngaliyan
18	L	4	Semarang Selatan
19	L	2	Gajah Mungkur
20	P	26	Semarang Timur
21	L	1	Semarang Barat
22	P	1	Gayamsari
23	P	18	Semarang Selatan
24	P	63	Semarang Barat
25	P	14	Tembalang
26	P	66	Semarang Selatan

Gambar 2.2 Data Selection

Setelah dilakukan proses cleaning didapat 295 data, 37 data dihilangkan karena atribut wilayah/daerahnya tidak ada value (nilai), atau bukan merupakan wilayah kecamatan di kota Semarang:

232	L	10	Semarang Barat
233	P	12	Semarang Barat
234	L	1	Mranggen
235	P	12	Candisari
236	L	44	Semarang Utara
237	L	5	Semarang Utara
238	L	7	Ngaliyan
239	P	30	Tugu
240	P	15	Gajah Mungkur
241	L	10	Banyumanik
242	L	9	Pedurungan
243	L	2	Jambu
244	L	52	Mejobo
245	P	4	Ngaliyan
246	L	36	Banyumanik
247	L	5	Genuk
248	P	1	Semarang Timur
249	L	2	Ngaliyan
250	P	43	Semarang Utara
251	L	1	Candisari
252	P	7	Tugu
253	P	41	Ambarawa
254	L	56	Semarang Barat
255	L	50	Tembalang
256	L	3	Tembalang
257	P	15	Kebonagung
258	P	17	Jati

Gambar 2.3 Data Cleaning

4. Fase Permodelan

Pada fase ini, memilih dan mengaplikasikan teknik pemodelan yang sesuai dengan kalibrasi aturan model untuk mengoptimalkan hasil. Perlu diperhatikan bahwa beberapa teknik mungkin digunakan pada permasalahan data mining yang sama. Jika diperlukan proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu. Pada permasalahan ini, teknik *clustering* data mining menggunakan algoritma k-means dianggap paling cocok untuk diterapkan di kasus ini.

5. Fase Evaluasi

Pada fase ini akan mengevaluasi model yang digunakan dalam fase permodelan serta menetapkan apakah model tadi sudah sesuai dengan tujuan pada fase awal. Pada penelitian ini akan dilakukan 2 proses evaluasi, yakni pengujian dan perbandingan untuk mengetahui hasil setelah pengimplementasian dengan data awal apakah sesuai dengan tujuan yang diinginkan.

6. Fase Implementasi

Fase implementasi adalah fase akhir dari enam fase CRISP-DM. Dengan terbentuknya model tidak menandakan telah terselesaikannya proyek. Contoh sederhana implementasi adalah pembuatan laporan. Contoh kompleks dari implementasi adalah penerapan proses data mining secara paralel pada departemen lain.

3. Implementasi

Pada aplikasi *data mining* ini terdiri dari beberapa interface yaitu *Menu Login*, *Halaman Utama*, *Menu About*, dan *Menu Input Data*.

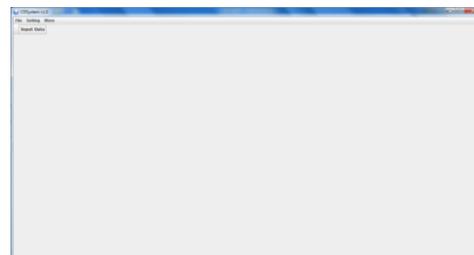
a) Login



Gambar 3.1 *User Interface Login*

Pada halaman ini *user / admin* memasukkan akun yang telah terdaftar agar bisa masuk ke halaman utama.

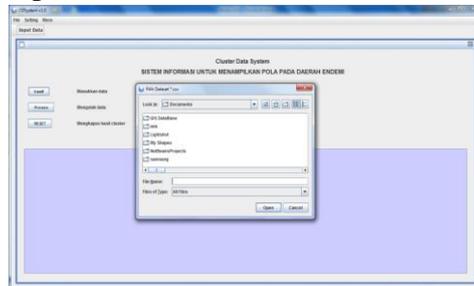
b) Halaman Utama



Gambar 3.2 *User Interface Home*

Halaman utama merupakan *interface* yang menampilkan menu fungsional berupa load data, reset data, proses data.

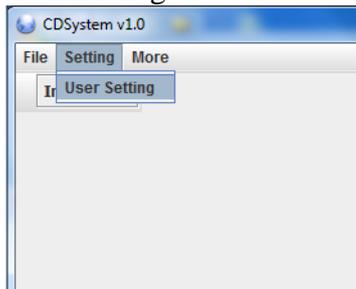
c) Input Data



Gambar 3.3 *User Interface Input Data*

Pada halaman Input Data *user* dapat memasukkan dataset yang telah di seleksi dan *cleaning data*.

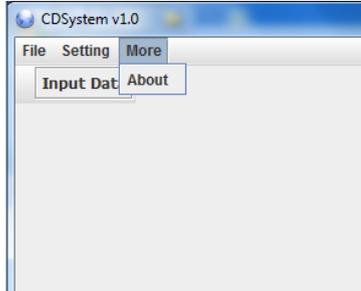
d) User Setting



Gambar 3.4 *User Interface UserSetting*

Halaman *User setting* merupakan *interface* yang mengatur akun dari pengguna, untuk penambahan, merubah dan menghapus user yang telah terdaftar

e) About



Gambar 3.5 User Interface About Pada halaman about dijelaskan mengenai versi dari *developer* perangkat lunak data mining.

4. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Pada hasil penelitian ini, terciptanya sistem untuk mengelompokkan data untuk memperoleh informasi pola. Sistem terdiri dari beberapa *interface* yaitu: *Login*, *Halaman Utama*, *Input Data*, *User Setting* dan *About*. Setiap antar muka mempunyai fungsi yang berbeda. Pada implementasi program dan *interface* pengguna dapat memperoleh informasi berupa data output daerah endemi, sehingga ada penanganan khusus untuk menangani kasus tersebut dengan menggunakan algoritma k-means.

B. Pembahasan

Tahap *clustering* dengan menggunakan K-Means ini dimulai dengan pembentukan *cluster*, pembagian *cluster* ini dipilih secara random, penulis membentuk 5 *cluster* dari 20 data yang diambil dari dataset sebagai contoh.

Proses penghitungan *centroid* awal dimulai dengan pemberian nama awal cluster (dari *cluster* pertama sampai dengan cluster kelima) secara random

pada data hasil *cleaning* (data 1-20).

Tabel 4.1 Membuat *cluster* pada masing-masing data:

No	Nama	Anak	Remaja	Dewasa	Tua	P	K.smgel	K.smgut	K.smgbar	K.smgang	Master ka
1	IRTEJAL	0	0	0	1	1	0	0	0	0	1
2	AHMAD SODIKIN	0	0	1	0	0	1	0	0	0	1
3	KOKO SAPUTRO	0	0	1	0	0	1	1	0	0	1
4	SUBARHAN TN	0	0	0	1	0	1	0	0	0	1
5	HELIA ADEL YANI RUSMELAN	0	1	0	0	1	0	0	1	0	1
6	PRABU CEMKARATI	0	0	1	0	1	0	0	0	0	1
7	SADRA SALMA TISTANA	1	0	0	0	1	0	0	0	0	1
8	GENAWAN GHANI	0	1	0	0	0	1	0	0	0	1
9	ANITA DPT TERAPARE	1	0	0	0	1	0	0	1	0	1
10	GALFY BISTANG VILLA	0	0	1	0	0	1	0	0	0	1
11	RAYOSO	0	0	0	1	0	1	1	0	0	1
12	ALDY FAJAR SETIAWAN PUTRA	1	0	0	0	0	1	1	0	0	1
13	NINIK WIHAYANTI	0	0	1	0	1	0	0	0	0	1
14	LISA NOVI PESHTASARI DR	0	0	1	0	1	0	0	0	0	1
15	PRAPTONO SALEH PN	0	0	0	1	0	1	0	0	1	0
16	NURLE INANWILKAB	0	0	1	0	1	0	0	1	0	1
17	IBRAHIM SYR GHAZANI	1	0	0	0	0	1	0	0	0	1
18	RADIAN FABRIK RAMADHAN	1	0	0	0	0	1	1	0	0	1
19	BROHIM	1	0	0	0	0	1	0	0	0	1
20	TULAHNTI	1	0	0	0	1	0	0	0	0	1

Setelah pemberian nama *cluster* untuk masing-masing data, maka selanjutnya akan dilakukan penghitungan untuk mendapatkan nilai *centroid* awal dengan menghitung rata-rata pada masing-masing cluster dengan membagi jumlah data yang didapatkan untuk setiap cluster-nya. Penghitungan *centroid* awal dengan menggunakan mean (rata-rata) ini ditujukan agar setiap cluster memiliki anggota data pada iterasi pertama. Adapun penghitungan nilai *centroid* awal adalah sebagai berikut:

- Untuk nilai *centroid* awal pada *cluster* pertama :

$$\text{Centroid awal } C_0 = \frac{\text{data1} + \text{data6} + \text{data11} + \text{data16}}{4}$$

- Untuk nilai *centroid* awal pada *cluster* kedua :

$$\text{Centroid awal } C_1 = \frac{\text{data2} + \text{data7} + \text{data12} + \text{data17}}{4}$$

- Untuk nilai *centroid* awal pada *cluster* ketiga :

$$\text{Centroid awal } C_2 = \frac{\text{data3} + \text{data8} + \text{data13} + \text{data18}}{4}$$

- Untuk nilai *centroid* awal pada *cluster* keempat :

$$\text{Centroid awal } C_3 = \frac{\text{data4} + \text{data9} + \text{data14} + \text{data19}}{4}$$

- Untuk nilai *centroid* awal pada *cluster* kelima :

$$\text{Centroid awal } C_4 = \frac{\text{data5} + \text{data10} + \text{data15} + \text{data20}}{4}$$

Adapun hasil dari *centroid* awal dari masing-masing *cluster* adalah sebagai berikut :

1. *Cluster* pertama
(C₀) = [0; 0; 0.5; 0.5; ...; 0]
2. *Cluster* kedua
(C₁) = [0.75; 0; 0.25; 0; ...; 0]
3. *Cluster* ketiga
(C₂) = [0.25; 0.25; 0.5; 0; ...; 0]
4. *Cluster* keempat
(C₃) = [0.5; 0; 0.25; 0.25; ...; 0]
5. *Cluster* kelima
(C₄) = [0.25; 0.25; 0.25; 0.25; ...; 0]

Kemudian melakukan penghitungan untuk menentukan jarak setiap data dengan *centroid* awal yang telah dibentuk dengan menggunakan rumus *euclidance distance*. Hasil dari perhitungan jarak dengan rumus *euclidance distance* ini akan berpengaruh pada penempatan setiap data ke *cluster* tertentu.

Berikut ini disajikan perhitungan data pertama terhadap nilai *centroid* masing-masing *cluster*.

- Jarak antara data pertama dengan *centroid* pertama (C₀)
 $d_{1,0} = \sqrt{(1-0)^2 + (0-0)^2 + (0-0.5)^2 + \dots + (0-0)^2} = 1.17260394$
- Jarak antara data pertama dengan *centroid* kedua (C₁)
 $d_{1,1} = \sqrt{(1-0.75)^2 + (0-0)^2 + (0-0.25)^2 + \dots + (0-0)^2} = 2$
- Jarak antara data pertama dengan *centroid* ketiga (C₂)
 $d_{1,2} = \sqrt{(1-0.25)^2 + (0-0.25)^2 + (0-0)^2 + \dots + (0-0)^2} = 1.968501969$
- Jarak antara data pertama dengan *centroid* keempat (C₃)
 $d_{1,3} = \sqrt{(1-0.5)^2 + (0-0.25)^2 + (0-0)^2 + \dots + (0-0)^2} = 1.322875656$
- Jarak antara data pertama dengan *centroid* kelima (C₄)
 $d_{1,4} = \sqrt{(1-0)^2 + (0-0.25)^2 + (0-0.5)^2 + \dots + (0-0)^2} = 1.620185175$

Penghitungan jarak data dengan *centroid* tiap *cluster*, pada 20 record data, selanjutnya akan disajikan dalam bentuk tabel dibawah ini.

Tabel 4.2 Penghitungan jarak

No	C0	C1	C2	C3	C4		
1	1.17260394		2	1.968501969	1.322875656	1.620185175	
2	1.695582496	1.414213562	1.369306394	1.658312395	1.620185175		
3	1.541103501	1.414213562	0.935414347	1.658312395	1.620185175		
4	1.541103501	1.732050808	1.695582496	1.322875656	1.620185175		
5	1.541103501		2	1.837117307	1.658312395	1.274754878	
6	1.17260394	1.870828693	1.695582496	1.658312395	1.620185175		
7	1.695582496	1.414213562	1.837117307		1.5	1.620185175	
8	1.968501969	1.732050808	1.369306394	1.802775638	1.620185175		
9	1.541103501	1.58113883	1.837117307	1.322875656	1.274754878		
10	1.695582496	1.58113883	1.369306394	1.658312395	1.457737974		
11	1.541103501	1.58113883	1.369306394	1.658312395	1.620185175		
12	1.837117307		1	1.17260394		1.5	1.620185175
13	1.369306394	1.870828693	1.541103501	1.658312395	1.620185175		
14	1.17260394	1.870828693	1.695582496	1.322875656	1.620185175		
15	1.541103501	1.732050808	1.695582496		1.5	1.274754878	
16	1.17260394	1.870828693	1.695582496		1.5	1.274754878	
17	1.968501969		1	1.541103501		1.5	1.620185175
18	1.837117307		1	1.17260394		1.5	1.620185175
19	1.968501969	1.224744871	1.541103501	1.322875656	1.620185175		
20	1.695582496	1.58113883	1.837117307		1.5	1.457737974	

Dari tabel diatas dapat dilihat bahwa untuk data pertama memiliki jarak terkecil yaitu 1.17260394 (C₀), maka data pertama akan menjadi anggota dari *cluster* ke-1. Untuk lebih jelasnya anggota data untuk masing-masing *cluster* akan disajikan dalam bentuk tabel dibawah ini.

Tabel 4.3 Anggota data tiap *cluster*

No	C0	C1	C2	C3	C4	Cluster ke		
1	1.172604		2	1.968502	1.322876	1.620185	1	
2	1.695582	1.414214	1.369306	1.658312	1.620185		3	
3	1.541104	1.414214	0.935414	1.658312	1.620185		3	
4	1.541104	1.732051	1.695582	1.322876	1.620185		4	
5	1.541104		2	1.837117	1.658312	1.274755	5	
6	1.172604	1.870829	1.695582	1.658312	1.620185		1	
7	1.695582	1.414214	1.837117		1.5	1.620185	2	
8	1.968502	1.732051	1.369306	1.802776	1.620185		3	
9	1.541104	1.581139	1.837117	1.322876	1.274755		5	
10	1.695582	1.581139	1.369306	1.658312	1.457738		3	
11	1.541104	1.581139	1.369306	1.658312	1.620185		3	
12	1.837117		1	1.172604		1.5	1.620185	2
13	1.369306	1.870829	1.541104	1.658312	1.620185		1	
14	1.172604	1.870829	1.695582	1.322876	1.620185		1	
15	1.541104	1.732051	1.695582		1.5	1.274755	5	
16	1.172604	1.870829	1.695582		1.5	1.274755	5	
17	1.968502		1	1.541104		1.5	1.620185	2
18	1.837117		1	1.172604		1.5	1.620185	2
19	1.968502	1.224745	1.541104	1.322876	1.620185		2	
20	1.695582	1.581139	1.837117		1.5	1.457738	5	

Pada iterasi pertama ini jumlah anggota yang didapatkan oleh masing-masing *cluster* adalah sebagai berikut :

1. Pada *cluster* pertama (C₀) memiliki jumlah anggota 4 data, yaitu pada data ke 1, 6, 13, 14.
2. Pada *cluster* kedua (C₁) memiliki jumlah anggota 5 data, yaitu pada data ke 7, 12, 17, 18, 19.

3. Pada *cluster* ketiga (C₂) memiliki jumlah anggota 5 data, yaitu pada data ke 2, 3, 8, 10, 11.
4. Pada *cluster* keempat (C₃) memiliki jumlah anggota 1 data, yaitu pada data ke 4.
5. Pada *cluster* kelima (C₄) memiliki jumlah anggota 5 data, yaitu pada data ke 5, 9, 15, 16, 20.

Iterasi pada *clustering* ini akan berhenti, jika anggota data *cluster* pada iterasi sebelumnya sama dengan anggota data *cluster* pada iterasi selanjutnya atau nilai centroid pada iterasi awal sama dengan nilai centroid pada iterasi selanjutnya.

Dari 20 dataset tersebut, untuk memperoleh nilai centroid yang sama terbentuklah 3 iterasi. Dari iterasi ke-3, didapatkan hasil bahwa nilai minimum centroid ada di data ke-12 (C₁) dan nilai maksimum centroid ada di data ke-8 (C₂).

Tabel 4.4 Hasil *Cluster* terakhir pada iterasi ke-3:

No	C0	C1	C2	C3	C4	Cluster ke
1	1.224744871	2.318404624	2.207940217	1.632993162	1.697056275	1
2	1.870828693	1.837117307	0.935414347	1.825741858	2.116601049	3
3	1.870828693	1.541103501	0.935414347	1.632993162	2.116601049	3
4	1.870828693	1.837117307	1.695582496	0.816496581	2.209072203	4
5	1.732050808	2.318404624	2.091650066	2.160246899	1.13137085	5
6	1	2.318404624	1.837117307	2.309401077	1.574801575	1
7	1.732050808	1.837117307	2.207940217	2.309401077	1.13137085	5
8	2.236067977	1.837117307	1.369306394	1.825741858	2.116601049	3
9	1.732050808	1.837117307	2.207940217	2.160246899	0.692820323	5
10	1.870828693	1.837117307	0.935414347	1.825741858	2.116601049	3
11	2.121320344	1.541103501	1.541103501	0.816496581	2.209072203	4
12	2.236067977	0.612372436	1.541103501	1.632993162	1.918332609	2
13	1	2.318404624	1.837117307	2.309401077	1.574801575	1
14	0.707106781	2.318404624	1.837117307	2.160246899	1.574801575	1
15	2.121320344	1.837117307	1.695582496	0.816496581	1.918332609	4
16	1.224744871	2.318404624	1.837117307	2.160246899	1.13137085	5
17	2.236067977	0.935414347	1.695582496	1.825741858	1.918332609	2
18	2.236067977	0.612372436	1.541103501	1.632993162	1.918332609	2
19	2.236067977	0.935414347	1.695582496	1.825741858	1.918332609	2
20	1.732050808	1.837117307	2.207940217	2.309401077	1.13137085	5

Dari hasil *clustering*, diperoleh karakteristik masing-masing *cluster*. Karakteristik yang diperoleh dari 20 dataset di atas adalah :

1. *Cluster* pertama (C₀) memiliki karakteristik tipe penyakit *Dengue Fever (Classical Dengue)*, banyak menyerang usia dewasa, jenis kelamin perempuan, pada kecamatan genuk.
2. *Cluster* kedua (C₁) memiliki karakteristik tipe penyakit *Diarrhoea And Gastroenteritis*, banyak menyerang usia anak-anak,

jenis kelamin laki-laki, pada kecamatan Semarang Selatan.

3. *Cluster* ketiga (C₂) memiliki karakteristik tipe penyakit *Dengue Fever (Classical Dengue)*, banyak menyerang usia dewasa, jenis kelamin laki-laki, pada kecamatan Semarang Selatan.
4. *Cluster* keempat (C₃) memiliki karakteristik tipe penyakit *Diarrhoea And Gastroenteritis*, banyak menyerang usia tua, jenis kelamin laki-laki, pada kecamatan Semarang Selatan.
5. *Cluster* kelima (C₄) memiliki karakteristik tipe penyakit *HIV Disease Resulting in Multiple*, banyak menyerang usia anak-anak, jenis kelamin perempuan, pada kecamatan Semarang Barat.

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Pada penelitian yang telah dilakukan diatas, dapat ditarik kesimpulan sebagai berikut :

Aplikasi ini dapat digunakan untuk menampilkan informasi pola daerah endemi penyakit. Informasi yang ditampilkan berupa *cluster* daerah endemi, endemi penyakitnya, jenis kelamin dan usia yang rentan terserang. Data RL rekam medis Rumah Sakit Kota Semarang yang diproses mining meliputi data alamat, kode penyakit, usia dan jenis kelamin.

Dari hasil proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang sikap apa yang harus segera diambil untuk menangani wabah penyakit sesuai daerah endemi.

B. Saran

Saran untuk penelitian selanjutnya adalah :

Dari hasil proses mining ini bisa dikembangkan untuk dijadikan aplikasi

dengan gambaran peta kota Semarang dengan memanfaatkan teknologi SIG. Penelitian ini dapat dikembangkan menjadi sebuah peramalan daerah endemi menggunakan algoritma untuk prediksi maupun algoritma untuk peramalan.

DAFTAR PUSTAKA

- [1] Siregar, Charles. 2003. Farmasi Rumah Sakit: Teori dan Penerapan. Jakarta: Penerbit Buku Kedokteran EGC.
- [2] S. Ray, R.H. Turi, 1999, "Determination of Number of Clusters in K-means Clustering and Application in Colour Image Segmentation", 4th ICAPRDT Proc., pp.137-143.
- [3] Ali Ridho Barakhbah, Kohei Arai, 2004, "Determining Constraints of Moving Variance to Find Global Optimum and Make Automatic Clustering" IES Proc. Pp 409-413.
- [4] Britton, Carol; Jill Doake (2001). Object-Oriented Systems Development. McGraw-Hill. hlm. 28-29, 269. ISBN 0-07-709544-8.
- [5] Connolly, T., Begg C. (2010). Database System: A Practical Approach to Design, Implementation, and Management. (5th Edition). New York: Addison Wesley.
- [6] B. Santosa, Data Mining. Teknik Pemanfaatan Data untuk Keperluan Bisnis, 1 ed., Yogyakarta: Graha Ilmu, 2007.
- [7] L. T. Daniel, Data Mining Methods and Models, John Wiley & Sons, Inc Publication, 2006.
- [8] David M. Kroenke, Database Processing Jilid 1 edisi 9, halaman 60.
- [9] Erlangga. Ir. M. FARID AZIS, M. Kom, Object Oriented Programming Php 5, halaman 118. Elex Media Komputindo.
- [10] Muhammad, F. & Ida, W. (2015). Jurnal Manajemen Informasi