

## **PENERAPAN DATA MINING PENGELOMPOKKAN PENYAKIT MENULAR SEKSUAL (PMS) MENGGUNAKAN METODE CLUSTERING**

Magdalena Simanjuntak, Ediman Manik, Putri Ratna Sari  
*magdalena.simanjuntak84@gmail.com*, [ediman.kaputama@gmail.com](mailto:ediman.kaputama@gmail.com), *putriratna209@gmail.com*  
Teknik Informatika, Komputerisasi Akuntansi, Sistem Informasi  
STMIK Kaputama

### **ABSTRAK**

Penyakit Kelamin (*veneral disease*) sudah lama dikenal di Indonesia. Sejalan dengan majunya ilmu pengetahuan istilah tersebut sudah tidak digunakan lagi dan dirubah menjadi *Sexually Transmitted Disease* (STD) atau Penyakit Menular Seksual (PMS). Infeksi menular seksual adalah infeksi yang ditularkan dari satu orang ke orang lainnya, penyakit PMS yang banyak diderita oleh beberapa laki-laki, perempuan dan remaja yang mengikuti pergaulan yang begitu luas dan bebas dan adanya fenomena gaya hidup modern, mereka rela melakukan apapun tanpa memikirkan resiko yang tidak tanggung-tanggung hingga terjangkitnya penyakit menular yang dapat merenggut jiwa pasien. Dalam hal ini peneliti juga ingin mengetahui penyakit apa saja yang sering terjadi pada beberapa pasien dan dengan diketahuinya penyakit yang diderita pada pasien, maka pihak rumah sakit dapat memberitahu langsung bagaimana pencegahan serta pengobatan yang harus dilakukan oleh pasien yang akan terjangkit maupun yang sudah terjangkit penyakit menular. jenis penyakit Kondiloma Akuminata dan faktor penyebabnya adalah Virus DNA Golongan Papovavirus. Dan dapat diketahui pada *cluster* 3 dari kriteria usia, jenis penyakit, dan faktor penyebab, kelompok mana yang memiliki himpunan / nilai paling tinggi dan paling banyak pasien yang mengalami sakit, yaitu pada *Cluster* 1 berjumlah 247 data pasien yang sakit usia 18-35 tahun dengan penyakit yang dialami ialah penyakit sifilis dan faktor penyebabnya adalah Bakteri Spiroseta. Dalam pengujian diatas dapat diketahui juga bahwa antara usia, jenis penyakit, dan faktor penyebab memiliki hubungan.

**Kata Kunci :** *Data Mining, Algoritma k-means, Penyakit Menular Seksual*

### **I. PENDAHULUAN**

Rumah sakit merupakan suatu lembaga sosial yang bergerak dibidang kesehatan dan memiliki fungsi dalam menyediakan pelayanan untuk seluruh jenis penyakit yang diderita oleh masyarakat, termasuk juga penyakit PMS yang banyak diderita oleh beberapa laki-laki, perempuan dan remaja yang mengikuti pergaulan yang begitu luas dan bebas dan adanya fenomena gaya hidup modern, mereka rela melakukan apapun

tanpa memikirkan resiko yang tidak tanggung-tanggung hingga terjangkitnya penyakit menular yang dapat merenggut jiwa pasien.

Dalam hal ini peneliti juga ingin mengetahui penyakit apa saja yang sering terjadi pada beberapa pasien dan dengan diketahuinya penyakit yang diderita pada pasien, maka pihak rumah sakit dapat memberitahu langsung bagaimana pencegahan serta pengobatan yang harus dilakukan oleh pasien yang akan terjangkit

maupun yang sudah terjangkit penyakit menular.

*Cluster analysis* adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Objek – objek yang bergabung dalam sebuah kelompok merupakan objek – objek yang mirip (berhubungan) satu sama lain dan berbeda (tidak berhubungan) dengan objek dalam kelompok yang lain. Salah satu metode yang paling banyak digunakan pada metode cluster adalah dengan *algoritma K-Means*.

Penelitian ini diperkuat oleh Jurnal Transformatika, Peneliti Anindya Khrisna Wardhani dengan judul “Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan” yang menyatakan bahwa tujuan dari penelitian ini ialah untuk menentukan konsistensi data kesehatan dapat digunakan teknik data mining yang mampu menggali informasi tersembunyi dari kumpulan data *multidimensi* yang telah diperoleh, selain itu data yang terhubung dengan data lain juga dapat dilakukan oleh teknik data mining.

## II. TINJAUAN PUSTAKA

### 2.1 Pengertian Data Mining

Data Mining menurut Santosa, Budi (2007, h.10) *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan.

### 2.2 Konsep Algoritma K-Means

Dalam statistic dan mesin pembelajaran, menurut Eko Prasetyo (2012, h.178) pengelompokan *K-Means* merupakan metode analisis kelompok yang mengarah pada pemartisipasian N objek pengamatan ke dalam K, Kelompok (*cluster*) dimana setiap objek pengamatan

dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* diaman keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan kedua algoritma.

*K-Means* merupakan salah satu metode pengelompokan data non-hirarki (sekata) yang berusaha mempartisipasi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisipasi data kedalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antara kelompok (Eko Prasetyo, 2012, h.178).

### 2.3 Mengukur Jarak Data Ke Pusat Kelompok

Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, di antaranya Euclidean (Bezdek, 1981), Manhattan/City Block (Miyamoto dan Agusta, 1995), dan Minkowsky (Miyamoto dan Agusta, 1995).

Masing-masing cara mempunyai kelebihan dan kekurangan.

Pengukuran jarak pada ruang jarak (distance space) Euclidean menggunakan formula :

$$D(x_2, x_1) = \|x_2, x_1\|_2 \\ = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2} \dots\dots\dots(1)$$

D adalah jarak antara data  $x_2$  dan  $x_1$ , dan  $|\cdot|$  adalah nilai mutlak.

Pengukuran jarak pada ruang jarak Manhattan menggunakan formula :

$$D(x_2, x_1) = \|x_2, x_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}| \dots\dots\dots (2)$$

Pengukuran jarak pada ruang jarak Minkowsky menggunakan formula :

$$D(x_2, x_1) = \|x_2, x_1\|_\lambda = \sqrt[\lambda]{\sum_{j=1}^p |x_{2j} - x_{1j}|^\lambda} \dots\dots\dots (3)$$

$\lambda$  adalah parameter jarak Minkowsky.

Cara yang paling banyak digunakan adalah Euclidean dan Manhattan. Euclidean menjadi pilihan jika kita ingin memberikan jarak terpendek antara dua titik (jarak lurus), seperti yang ditunjukkan pada rumus Euclidean. Sedangkan Manhattan memberikan jarak terjauh pada dua data.

#### 2.4 Pengertian Penyakit Menular Seksual (PMS)

Menurut Verra Scorviani, SH dan dr. Taufan Nugroho (2012, h.1) Penyakit Menular Seksual (PMS) adalah penyakit yang ditularkan melalui hubungan seksual. Penyakit menular seksual akan lebih beresiko bila melakukan hubungan seksual dengan berganti-ganti pasangan baik melalui vagina, oral maupun anal. Penyakit menular seksual dapat menyebabkan infeksi pada alat reproduksi yang harus dianggap serius. Bila tidak diobati secara tepat, infeksi dapat menjalar dan menyebabkan penderitaan, sakit berkepanjangan, kemandulan, dan bahkan kematian. Wanita lebih beresiko untuk terkena penyakit menular seksual lebih besar dari pada laki-laki sebab wanita mempunyai alat reproduksi yang lebih rentan dari pada laki-laki. Dan seringkali berakibat lebih parah karena gejala awal tidak segera dikenali, sedangkan penyakit melanjut ketahap yang lebih parah, Karena letak dan bentuk kelamin laki-laki yang agak menonjol, gejala penyakit menular seksual pada laki-laki lebih mudah dikenali, dilihat, dan dirasakan sedangkan pada wanita sebagian besar gejala yang timbul hampir tidak dapat dirasakan.

#### 2.5 Bahaya Penyakit Menular

Menurut Verra Scorviani, SH dan dr. Taufan Nugroho (2011, h.2) Penyakit menular seksual menyebabkan infeksi saluran reproduksi yang harus dianggap serius. Bila tidak diobati secara tepat, infeksi dapat menjalar dan menyebabkan penderitaan, sakit berkepanjangan, kemandulan dan kematian. (3)

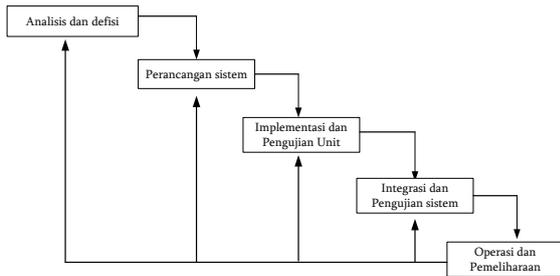
Tanda dan Gejala Penyakit Menular Seksual ;

- a. Rasa sakit atau nyeri pada saat buang air kecil atau berhubungan seksual.
- b. Rasa nyeri pada perut bagian bawah.
- c. Pengeluaran lendir pada vagina / alat kelamin.
- d. Keputihan bewarna putih susu, bergumpal dan disertai rasa gatal dan kemerahan pada alat kelamin atau sekitarnya.
- e. Keputihan yang berbusa, kehijauan, berbau busuk, dan gatal.
- f. Timbul bercak bercak darah setelah berhubungan seks.
- g. Bintil-bintil berisi cairan, lecet atau borok pada alat kelamin, dll.

### III. ANALISA SISTEM

#### 3.1 Analisis Sistem

Dalam membangun sebuah aplikasi untuk mengelompokkan jenis penyakit menular pada pasien dapat dilakukan beberapa tahap analisis, dimana tujuan utama analisis sistem ini adalah menentukan kesalahan dan kelemahan dari sistem yang sedang berjalan, sehingga dapat diusulkan perbaikannya. Dalam penelitian ini juga dilakukan analisis terhadap sistem yang telah diteliti adapun proses pembuatan perangkat lunak Penerapan Data Mining pengelompokkan Jenis Penyakit Menular pada pasien dengan menggunakan model air terjun (*waterfall*). Dimana siklus perangkat lunak didefinisikan pada gambar dibawah ini:



**Gambar III.1 Proses pembuatan rekayasa perangkat lunak**

Tahap-tahap yang akan dilakukan untuk pengembangan dasar yaitu:

1. Analisis dan definisi persyaratan  
Pada tahap ini dilakukan pengumpulan data, analisis kebutuhan, batasan masalah, dan tujuan dari pembuatan sistem dengan melakukan wawancara langsung serta pengambilan data.
2. Perancangan sistem dan perangkat lunak  
Proses ini dilakukan dengan perancangan menu sistem, modul serta arsitektur sistem secara keseluruhan, sehingga dihasilkan rancangan sistem yang digunakan sebagai dasar pembuatan sistem pada tahap implementasi.
3. Implementasi sistem dan perangkat lunak  
Pada tahap ini dilakukan dengan pembuatan pemrograman perangkat lunak (*coding*) dan pengujian perangkat lunak dengan memasukkan beberapa *sample* data.
4. Integrasi dan pengujian sistem  
Perangkat lunak akan terintegrasi dan diuji sebagai sistem yang lengkap untuk menjamin bahwa persyaratan sistem telah dipenuhi dan sesuai dengan kebutuhan.
5. Operasi dan pemeliharaan  
Melakukan koreksi dari berbagai macam kesalahan yang tidak ditemukan pada tahap sebelumnya sehingga dapat dilakukan perbaikan, agar dapat menghasilkan sistem yang lebih baik.

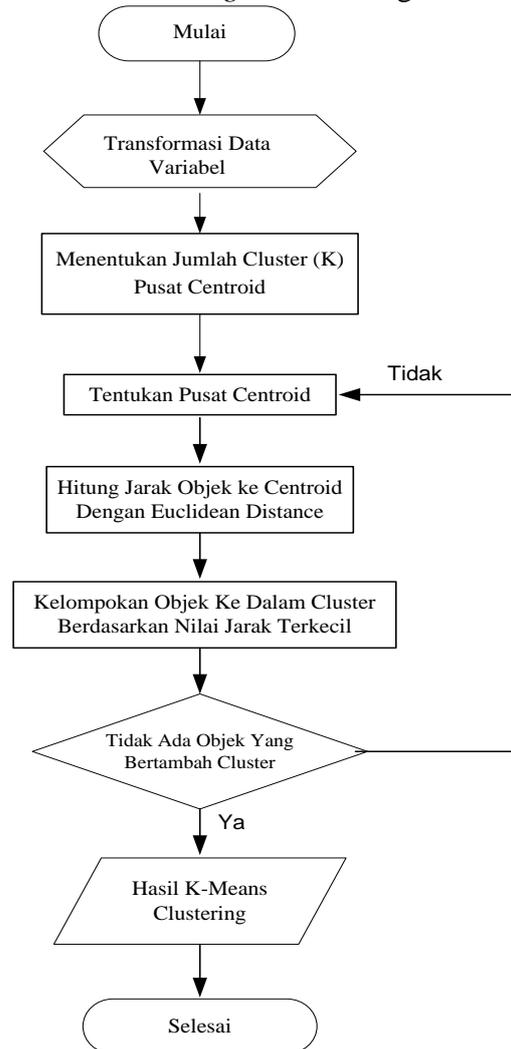
### 3.2 Perancangan Sistem

Berdasarkan analisis yang dilakukan, peneliti akan menggali data

untuk dikelompokkan dengan menggunakan metode *clustering* dengan algoritma *K-Means*, dimana variabel-variabel dari data penyakit menular pada pasien yang akan diambil untuk penelitian tersebut adalah usia, jenis penyakit dan faktor penyebab. Kemudian dari data tersebut di masukan ke dalam *Matlab* sehingga terdapat hasil 3 kelompok.

### 3.3 Flowchart

Adapun *flowchart K-Means Clustering* adalah sebagai berikut:



**Gambar III.2 Flowchart K-Means Clustering**

**IV. PEMBAHASAN**

**4.1 Pembahasan**

Langkah-langkah yang dilakukan untuk perhitungan data penyakit menular seksual pada pasien menggunakan metode *clustering* dengan algoritma *K-means* ini, agar dapat dihasilkan sebuah pengetahuan baru, mengenai berapa banyak kelompok data penyakit menular seksual pada pasien, kelompok jenis penyakit dan kelompok faktor penyebab. Sehingga dapat diketahui hubungan terdekat antara kelompok data penyakit menular seksual pada pasien.

**4.2 Data Input/Import**

Data yang dibutuhkan untuk diinput kedalam pemrograman *matlab* adalah berupa data penyakit menular seksual pada pasien dan yang diambil ialah data usia, jenis penyakit, faktor penyebab. Kemudian data-data penyakit menular seksual pada pasien tersebut diinput kedalam *microsoft excel* sebagai *database* dengan nama “data akhir” dimana data yang akan diproses sebanyak data 536 data. Dapat dilihat data yang akan diolah menggunakan algoritma *K-means* dengan *Matlab* adalah sebagai berikut:

Dari data variabel-variabel diatas, maka data penyakit menular seksual pada pasien dapat di transformasikan sesuai kebutuhan untuk perhitungan dengan metode *Clustering* data pemrograman *Matlab*. Berikut adalah data penyakit menular seksual pada pasien yang sudah di transformasikan.

**Tabel IV.1 Data Penyakit Menular Seksual Pada Pasien Yang Telah Ditransformasi**

N O	X	Y	Z	N O	X	Y	Z
1	8	3	2	5	9	2	7
2	4	3	9	6	1	3	2
3	8	3	2	7	1	3	6
4	1	3	2	8	1	2	6
9	2	2	3	50	8	3	2
10	9	2	7	51	6	4	8

11	2	2	3	52	8	3	2
12	2	2	3	53	8	2	2
13	4	3	9	54	8	3	2
14	8	3	2	55	1	4	4
15	8	3	2	56	6	2	8
16	8	2	2	57	8	3	2
17	9	1	7	58	8	3	2
18	9	2	7	59	8	3	2
19	9	1	7	60	8	2	2
20	8	3	2	61	2	2	3
21	8	2	2	62	1	3	2
22	8	2	2	63	2	3	3
23	8	2	2	64	6	3	8
24	8	2	2	65	1	4	4
25	8	3	2	66	8	3	2
26	8	3	2	67	9	1	7
27	1	0	2	68	9	1	7
28	8	3	2	69	6	2	8
29	9	3	7	70	8	3	2
30	8	3	2	71	9	1	7
31	8	3	2	72	9	3	7
32	1	3	2	73	9	3	7
33	8	2	2	74	6	2	8
34	8	3	2	75	1	0	2
35	8	3	2	76	8	2	2
36	8	2	2	77	1	1	3
37	5	2	5	78	1	1	3
38	8	3	2	79	1	1	2
39	8	3	2	80	1	1	2
40	3	3	2	81	1	1	3

**4.3 Perhitungan Centroid**

Adapun penentuan hasil jumlah *centroid* untuk setiap *group* adalah sebagai berikut:

$Centroid\ 1 = total\ group\ 1 / banyak\ group\ 1$

$C1 = 2761 / 253 = 10.91$

$C2 = 627 / 253 = 2.47$

$$C3 = 1390 / 253 = 5.49$$

Centroid 2 = total group 2/banyak group 2

$$C1 = 1790 / 283 = 6.32$$

$$C2 = 723 / 283 = 2.55$$

$$C3 = 841 / 283 = 2.97$$

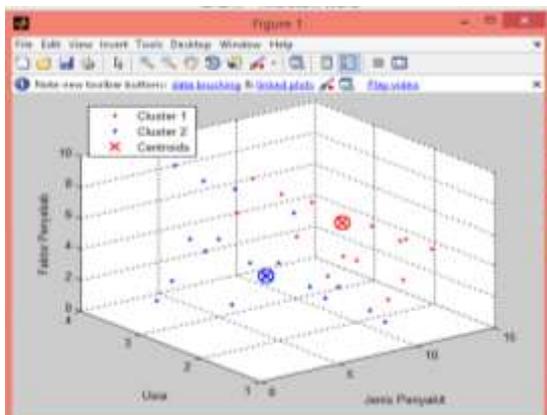
**Tabel IV.3 Perhitungan Centroid**

Usia (X)	Jenis Penyakit (Y)	Faktor Penyebab (Z)	Keterangan
10.91	2.47	5.49	Centroid 1
6.32	2.55	2.97	Centroid 2

Untuk group 1 memiliki nilai yang baik. Karena untuk hasil Jenis Penyakit (X), Usia (Y), Faktor Penyebab (Z) jumlah data yang dimiliki sebanyak 253 data. Dari keterangan data menunjukkan bahwa pasien yang sering terserang penyakit adalah jenis penyakit dengan range 10.91 kemudian usia 2.47 dan faktor penyebab 5.49.

Untuk group 2 memiliki nilai yang baik. Karena untuk hasil Jenis Penyakit (X), Usia (Y), Faktor Penyebab (Z) jumlah data yang dimiliki sebanyak 283 data. Dari keterangan data menunjukkan bahwa pasien yang sering terserang penyakit adalah jenis penyakit dengan range 6.32 kemudian usia 2.55 dan faktor penyebab 2.97.

Hasil cluster 2 dapat dilihat pada grafik dibawah ini :



## V. KESIMPULAN

Dari 536 data penyakit menular pada pasien diperoleh 2 cluster, dimana cluster 1 berjumlah 253 pasien dan cluster 2 berjumlah 283 pasien sebagai berikut adalah

1. **10** (10.91) **2** (2.47) **5** (5.49)

Dapat diketahui bahwasannya pada cluster 1. Kelompok data penyakit menular pada pasien yang sering terserang penyakit pada group jenis penyakit adalah penyakit sifilis dengan usia 18-35 tahun dan faktor penyebabnya adalah Bakteri Spiroseta.

2. **6** (6.32) **2** (2.55) **2** (2.97)

Dapat diketahui bahwasannya pada cluster 2. Kelompok data penyakit menular pada pasien yang sering terserang penyakit pada group jenis penyakit adalah penyakit Kondiloma Akuminata dengan usia 18-35 tahun dan faktor penyebabnya adalah Virus DNA Golongan Papovavirus.

## DAFTAR PUSTAKA

- Astuti Hermawati, Fajar. 2013 Data mining. Andi, Yogyakarta.
- Kusrini, Emha Taufiq Luthfi (2009, h.7) Algoritma Data mining, Yogyakarta, Penerbit CV Andi.
- Prasetyo, Eko. 2012. Data mining konsep dan aplikasi menggunakan matlab, Andi, Yogyakarta.
- Verra Scoviani, SH dan dr. Taufan Nugroho. 2012, Mengungkap tuntas 9 jenis PMS ( Penyakit Menular Seksual), Nuha Medika, Yogyakarta.