

ANALISIS *CLUSTERING* TINGKAT KEPARAHAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA *K-MEANS* (STUDI KASUS DI PUSKESMAS BANDAR SEIKIJANG)

Mentari Tri Indah Rahmayani¹

¹ *Jurusan Informatika FST UPTT*

Jln. Tuanku Tambusai No.23 Bangkinang 28412 INDONESIA

¹mentari.tri@gmail.com

Intisari— Puskesmas Bandar Seikijang merupakan salah satu pelayanan pusat kesehatan masyarakat yang terletak di wilayah kecamatan Bandar Seikijang Kabupaten Pelalawan. Jumlah pasien pada puskesmas tersebut terus meningkat setiap tahunnya. Dengan meningkatnya jumlah pasien berbanding lurus dengan peningkatan jenis-jenis penyakit pasien. Untuk memudahkan pelayanan kesehatan masyarakat maka perlu dilakukan pengelompokan tingkat keparahan penyakit pasien. Pengelompokan (*Clustering*) penyakit pasien tersebut terdiri dari 3 *cluster* yaitu penyakit berat, penyakit sedang dan penyakit ringan. Pengelompokan penyakit menggunakan *metode K-Means*. Tujuan dari penelitian ini adalah untuk mengelompokkan tingkat keparahan penyakit pasien, mengetahui tingkat akurasi dari *cluster* tingkat keparahan penyakit pasien, mengetahui penyakit terbanyak yang diderita masyarakat sekitar puskesmas, dan mengetahui persamaan dalam pengolahan data manual dan menggunakan *software Rapid Miner*. Sampel data pada pengolahan secara manual sebanyak 15 pasien dan data secara keseluruhan sebanyak 278 pasien. Hasil penelitian menunjukkan tingkat keparahan penyakit pasien *Cluster* (C0) berada pada penyakit berat dengan pasien sebanyak 47 orang, *cluster* (C1) berada pada penyakit ringan dengan pasien sebanyak 82 orang, dan *cluster* (C2) berada pada penyakit sedang dengan pasien sebanyak 149 orang. Tingkat keparahan penyakit pasien berada pada penderita penyakit sedang dengan persentase 53,59%. Penyakit yang paling sering diderita oleh masyarakat sekitar puskesmas adalah penyakit ISPA, demam berdarah dan malaria. Dan hasil implementasi menggunakan *Software Rapid Miner* bernilai sama dengan pengolahan data secara manual.

Kata Kunci : *Data Mining, Clustering, K-Means, Tingkat Keparahan penyakit pasien, Rapid Miner.*

Abstract— Puskesmas Bandar Seikijang is one of the community health center services located in the Bandar Seikijang sub-district of Pelalawan Regency. The number of patients in the puskesmas continues to increase every year. With the increasing number of patients is directly proportional to the increase in the types of patient's disease. To facilitate public health services it is necessary to classify the severity of the patient's disease. Clustering of the patient's disease consists of 3 clusters namely severe disease, moderate disease and mild disease. Disease grouping using the K-Means method. The purpose of this study was to classify the severity of the patient's disease, determine the level of accuracy of the cluster the severity of the patient's disease, find out the most diseases suffered by the community around the health center, and find out the similarities in manual data processing and using Rapid Miner software. Data samples on manual processing were 15 patients and overall data were 278 patients. The results showed the severity of Cluster patient disease (C0) was in severe disease with 47 patients, cluster (C1) was in mild disease with 82 patients, and cluster (C2) was in moderate disease with 149 patients. The severity of the patient's disease is in patients with moderate disease with a percentage of 53.59%. The diseases most often suffered by the community around the puskesmas are ARI, dengue fever and malaria. And the implementation results using Rapid Miner Software are the same as manual data processing.

Keywords : *Data Mining, Clustering, K-Means, Severity of Patient Disease, Rapid Miner*

1. PENDAHULUAN

Puskesmas Bandar Seikijang merupakan salah satu puskesmas yang terletak di kecamatan Bandar Seikijang Kabupaten Pelalawan Provinsi Riau. Setiap tahunnya jumlah pasien yang berobat semakin meningkat. Saat ini jumlah pasien semakin meningkat peningkatan jumlah pasien berbanding lurus dengan jenis-jenis penyakit pasien. Dari data jumlah pasien terhadap jenis penyakit tersebut, belum ada pendataan terhadap tingkat keparahan dari penyakit yang diderita pasien. Tingkat keparahan penyakit dapat dikelompokkan berdasarkan penderita penyakit ringan, sedang dan berat. Dengan tidak adanya data tingkat keparahan penyakit pasien ini, maka menyulitkan pihak puskesmas dalam melakukan penyuluhan kesehatan, menyediakan obat-obatan dan peralatan kesehatan dalam melakukan pelayanan kesehatan yang tepat sasaran. Untuk itu diperlukan pengelompokan (*cluster*) terhadap penyakit-penyakit berdasarkan tingkat keparahan penyakit. Metode yang digunakan adalah *Clustering*.

Clustering adalah salah satu metode dalam data mining yang digunakan untuk mencari dan mengelompokkan data yang memiliki kesamaan karakteristik dengan lainnya (Wardhani, 2016). Data yang diolah dengan teknik data mining akan menghasilkan suatu pengetahuan (*knowledge*) yang baru bersumber dari informasi data lama (Ong, 2013). Metode *Clustering* yang ada pada saat ini sangat banyak, salah satunya adalah *K-Means*. *K-Means* mempunyai sifat relatif efektif dan efisien (Wardhani, 2016).

1.2 Perumusan Masalah

Berdasarkan dari uraian latar belakang masalah diatas dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana menentukan kelompok tingkat keparahan penyakit pasien berdasarkan kategori penyakit yaitu penyakit ringan, penyakit sedang dan penyakit berat ?
2. Bagaimana mendapatkan suatu pengetahuan baru (*knowledge*) terhadap *clustering* berdasarkan kategori penyakit pasien untuk memudahkan pihak puskesmas dalam meningkatkan pelayanan kesehatan masyarakat dimasa yang akan datang?

1.3 Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah untuk :

1. Mendapatkan *Clustering* kategori penyakit ringan, sedang dan berat pada pasien yang berobat pada Puskesmas Bandar Seikijang.
2. Mendapatkan tingkat akurasi dari *cluster* tingkat keparahan penyakit pasien.
3. Mengetahui penyakit yang lebih banyak

diderita masyarakat disekitar puskesmas Bandar Seikijang.

Mengetahui persamaan dalam pengolahan data secara manual dan menggunakan *software Rapid Miner*.

1.4 Ruang Lingkup Masalah

Ruang lingkup masalah yang digunakan dalam penelitian tesis ini adalah :

1. Data yang diolah adalah data penyakit pasien yang berobat pada Puskesmas Bandar Seikijang Kabupaten Pelalawan bulan januari 2017.
2. Pengelompokan (*Cluster*) yang digunakan adalah pasien berdasarkan kategori penderita penyakit, yaitu penyakit ringan, penyakit sedang, dan penyakit berat.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Memberikan informasi-informasi pelayanan kesehatan masyarakat berdasarkan kategori tingkat keparahan penderita, yaitu penyakit ringan, penyakit sedang dan penyakit berat pada masyarakat yang berobat di Puskesmas Bandar Seikijang.
2. Membantu Puskesmas Bandar Seikijang dalam menyediakan obat-obatan dan peralatan kesehatan yang dibutuhkan dan tepat sasaran sesuai tingkat keparahan berdasarkan *cluster* penyakit.
3. Membantu pemerintah untuk memudahkan dalam melakukan penyuluhan kesehatan masyarakat berdasarkan tingkat keparahan penyakit yang diderita oleh pasien.

1.6 Sistematika Penulisan

Untuk mempermudah dalam penyusunan tesis ini, penulis membagikan pokok-pokok permasalahan ke dalam lima bab sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini diuraikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi tentang tinjauan pustaka yang diperlukan dalam pembahasan masalah, yaitu konsep dasar Algoritma *K-Means* untuk *Clustering*

BAB III METODOLOGI PENELITIAN

Dalam bab ini membahas tentang metodologi penelitian yang diterapkan dalam penulisan thesis ini.

BAB IV ANALISA DAN PEMBAHASAN

Bab ini menjelaskan tentang bagaimana

menerapkan Algoritma *K-Means* Untuk *Clustering* Penyakit pada Pasien

BAB V PENGUJIA

Dalam bab ini berisikan tentang pengujian dengan bantuan *Software Rapid Miner*.

BAB VI KESIMPULAN DAN SARAN

Pada bab yang terakhir ini berisikan kesimpulan dan saran dari hasil penelitian yang telah dilakukan

2. LANDASAN TEORI

2.1 Data Mining

Data mining disebut juga dengan *pattern recognition* yang merupakan metode dalam pengolahan data untuk menemukan pola tersembunyi dari data yang diolah kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan di masa yang akan datang (Wardhani, 2016).

2.2 Pengertian Clustering

Clustering adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak.

2.3 K-Means Clustering

K-means clustering merupakan metode data *clustering non-hirarki* yang mengelompokan data dalam bentuk satu atau lebih *cluster*. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu *cluster* data yang memiliki karakteristik yang berbeda dikelompokan dengan *cluster* yanglain sehingga data yang berada dalam satu

cluster/kelompok memiliki tingkat variasi yang kecil (Ong, 2013).

Adapun langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut :

1. Menentukan jumlah *cluster k* yang akan dibentuk.
2. Inisialisasi *k* pusat *cluster* dapat dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberikan nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster*

4. *r* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam

cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

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

dimana:

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

X *ki* = Data ke *i* pada atribut data ke *k*

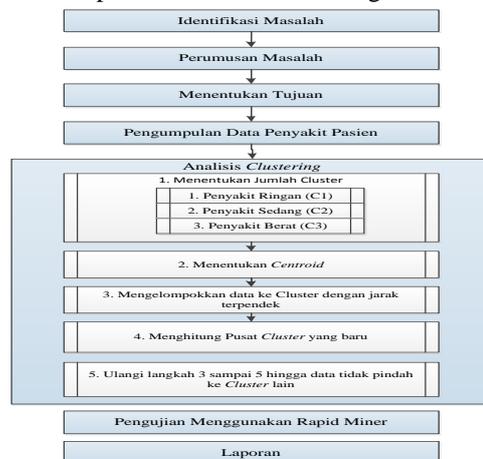
X *kj* = Titik pusat ke *j* pada atribut ke *k*

5. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.

6. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

3. METODOLOGI

Diagram penelitian dilakukan secara sistematis yang dapat digunakan oleh peneliti sebagai pedoman dalam menyelesaikan penelitian ini agar hasil yang dicapai tidak menyimpang dan tujuan yang diinginkan sesuai dengan yang telah ditetapkan. Adapun Diagram penelitian dalam melakukan penelitian ini adalah sebagai berikut :



Gambar 3.1 Metodologi Penelitian

4. ANALISA DAN PEMBAHASAN

Pada proses perhitungan data secara manual data yang dijadikan data sampel adalah 15 pasien dari 278 pasien. Berikut adalah sampel data penyakit puskesmas Bandar Seikijang dapat dilihat pada tabel 4.1 berikut ini :

Tabel 4.1 Sampel Data Pasien

| Data ke | Jenis Kelamin | Umur (th) | Kode Diagnosa | Lama Mengidap (minggu) | Diagnosa |
|---------|---------------|-----------|---------------|------------------------|-----------------------|
| 1 | Perempuan | 3 | 9 | 1 | Diare |
| 2 | Perempuan | 4 | 4 | 1 | Cacar Air |
| 3 | Laki-laki | 16 | 17 | 3 | Mimisan |
| 4 | Perempuan | 8 | 9 | 1 | Diare |
| 5 | Laki-laki | 25 | 77 | 10 | Maag Akut |
| 6 | Perempuan | 10 | 4 | 2 | Cacar Air |
| 7 | Perempuan | 9 | 33 | 1 | Masuk Angin |
| 8 | Laki-laki | 50 | 147 | 24 | Diabetes Militus |
| 9 | Laki-laki | 15 | 72 | 2 | Typus |
| 10 | Perempuan | 4 | 4 | 1 | Cacar Air |
| 11 | Perempuan | 27 | 90 | 10 | Anemia (Kurang Darah) |
| 12 | Perempuan | 2 | 130 | 3 | Polio |
| 13 | Perempuan | 29 | 77 | 8 | Maag Akut |
| 14 | Laki-laki | 16 | 70 | 2 | Malaria |
| 15 | Laki-laki | 16 | 9 | 2 | Diare |

Pada cluster penyakit berat nilai centroid yang diambil pada data ke-8, Pada cluster penyakit sedang nilai centroid yang diambil pada data ke-14, dan Pada cluster penyakit ringan nilai centroid yang diambil pada data ke-7. Tahapan berikutnya adalah menghitung jarak antara data dengan pusat awal cluster menggunakan persamaan euclidean distance sebagai berikut :

$$D_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - x_{kj})^2} \dots\dots\dots(4.1)$$

Maka akan didapatkan nilai matrik jarak sebagai berikut : Jarak pada data ke-1 ke pusat cluster C0 = $\sqrt{(50 - 3)^2 + (147 - 9)^2 + (24 - 1)^2}$ C0 = 147,5872623

Dan seterusnya untuk menghitung data ke-2 sampai data ke-15 hingga didapatkan matrik jarak. Setelah nilai pusat cluster didapat selanjutnya dilihat nilai dengan jarak terpendek.

Proses iterasi terus dilakukan hingga nilai centroid dan posisi atau letak cluster tidak berpindah lagi. Pada penelitian ini iterasi dilakukan sampai iterasi 5.

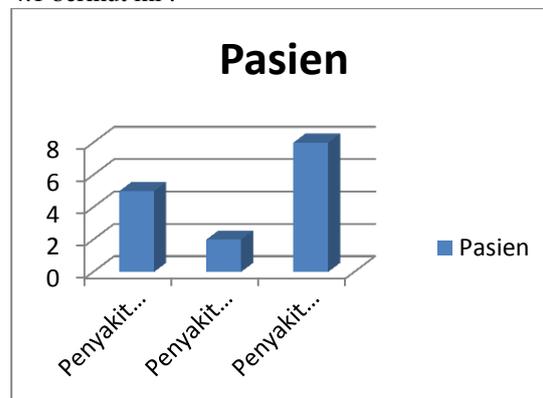
Setelah nilai centroid diperoleh selanjutnya akan diperoleh letak atau posisi cluster. Berikut adalah posisi cluster dari centroid iterasi

ke-5 dapat dilihat pada tabel 4.2 berikut ini :

Tabel 4.2 Pengelompokkan Data Pasien

| No | C0 | C1 | C2 |
|----|----|----|----|
| 1 | * | | |
| 2 | * | | |
| 3 | | * | |
| 4 | * | | |
| 5 | * | | |
| 6 | | * | |
| 7 | | * | |
| 8 | | * | |
| 9 | * | | |
| 10 | | | * |
| 11 | * | | |
| 12 | | * | |
| 13 | * | | |
| 14 | * | | |
| 15 | | | * |

Hasil pengolahan data secara manual dengan sampel 15 pasien dapat dilihat pada gambar 4.1 berikut ini :

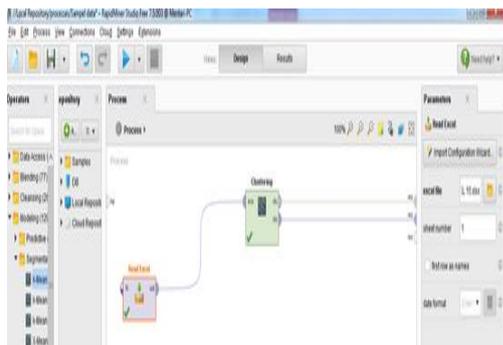


Gambar 4.1 Cluster Tingkat Keparahan Penyakit pasien

Pada grafik tersebut dapat dilihat bahwa untuk cluster 0 pada penyakit sedang yaitu sebanyak 5 pasien dengan data pasien ke-5, ke-9, ke-11, ke-13 dan ke-14. Untuk cluster 1 pada penyakit berat yaitu sebanyak 2 pasien dengan data pasien pada data ke-8 dan data ke-12. Untuk cluster 2 pada penyakit ringan sebanyak 8 orang pada data pasien ke-1, ke-2, ke-3, ke-4, ke-6, ke-7, ke-10 dan data ke-15

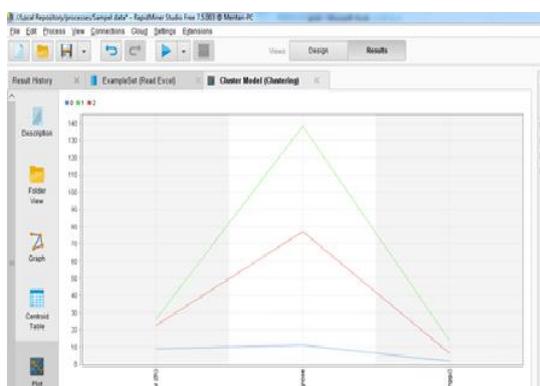
IMPLEMENTASI DAN HASIL

Implementasi menggunakan software Rapid Manir dengan data sampel sebanyak 15 pasien. Berikut adalah pengolahan data menggunakan K-Means untuk sampel data dapat dilihat pada gambar 5.1 berikut ini :



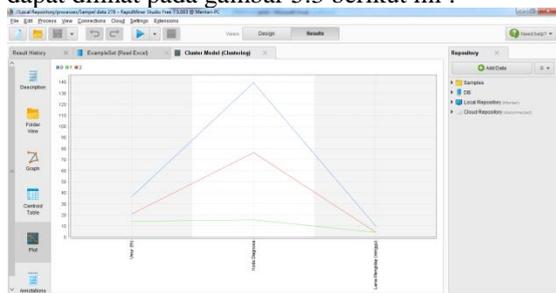
Gambar 5.1 Pemodelan Clustering K-Means

Berikut adalah gambar grafik dari implementasi software Rapid Miner untuk data sampel dapat dilihat pada gambar 5.2 berikut ini :



Gambar 5.2 Grafik hasil Clustering

Berikut adalah grafik hasil clustering menggunakan K-Means untuk keseluruhan data dapat dilihat pada gambar 5.3 berikut ini :



Gambar 5.3 Grafik hasil Clustering

6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan pembahasan dan hasil implementasi software Rapid Miner pada pengelompokan tingkat keparahan penyakit pasien, maka dapat ditarik kesimpulan sebagai berikut

1. Pengolahan data secara manual dengan data sampel 15 pasien, penyakit sedang berada

- cluster (C0) dengan pasien sebanyak 5 orang, penyakit berat berada cluster (C1) dengan pasien sebanyak 2 orang dan penyakit ringan berada pada cluster (C2) dengan pasien sebanyak 8 orang.
2. Pengolahan data secara keseluruhan menggunakan software Rapid Miner dengan jumlah pasien sebanyak 278 orang. Cluster (C0) berada pada penyakit berat dengan pasien sebanyak 47 orang, cluster (C1) berada pada penyakit ringan dengan pasien sebanyak 82 orang, dan cluster (C2) berada pada penyakit sedang dengan pasien sebanyak 149 orang.
3. Dari hasil pengelompokan tingkat keparahan penyakit pasien masyarakat dilingkungan puskesmas Bandar Seikijang lebih banyak menderita penyakit sedang dengan persentase 53,59%.
4. Dari hasil pengolahan data secara keseluruhan masyarakat disekitar puskesmas Bandar Seikijang lebih banyak menderita penyakit sedang (C2) dengan jenis penyakit dominan Infeksi Saluran Pernafasan Akut (ISPA), demam berdarah dan malaria.
5. Pengolahan data secara manual dan menggunakan software Rapid Miner memiliki hasil nilai centroid yang sama.

5.2 Saran

Berdasarkan pada kesimpulan maka disarankan agar diadakan penelitian lebih lanjut untuk clustering tingkat keparahan penyakit pasien dengan menambah atribut pada pengolahan data pasien tersebut.

DAFTAR PUSTAKA

Aprilius, W. (2015) ‘Big Data dan Perawatan Kesehatan Studi Awal Menuju Perawatan Kesehatan Masa Depan’, *ULTIMA Infosys*, VI(1), pp. 64–70.

Dhuhita, W. M. P. (2016) ‘Clustering Menggunakan Metode K-Means Untuk’, *Jurnal Informatika*, 15(2), pp. 160--174.

eko prasetyo (2010) *Data mining*. jogjakarta.

Iswanto, M. H. et al. (2015) ‘Pemanfaatan Teknik Data Mining Untuk Diagnosis Penyakit Tuberculosis (TBC)’, *Seminar Nasional Teknologi Informasi dan Multimedia (STM IK AMIKOM)*, Yogyakarta, pp. 121–126.

Mauladi, K. F. (2013) ‘Information Extraction Data Rekam Medis Pasien Untuk Perolehan Association Rule Sebagai Pendukung Pengambilan Keputusan’, 5(1).Ong, J. O. (2013) ‘Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university’, *Jurnal Ilmiah teknik Industri*, vol.12, no(juni), pp. 10–20.