
Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering

Parasian D P Silitonga¹ Irene Sri Morina²
Program Studi Teknik Informatika¹ Instalasi SIRS²
Universitas Katolik St. Thomas RSU Pusat Haji Adam Malik Medan
e-mail : parasianiren@gmail.com¹ morina_ginting@yahoo.com²

Abstrak

Data mining merupakan suatu proses yang dapat digunakan untuk menemukan hubungan yang berarti dari sekumpulan data dengan memeriksa data yang tersimpan dalam media penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Salah satu teknik pada data mining adalah teknik *clustering*. Teknik *clustering* merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Tujuan dari penelitian ini adalah untuk menemukan pola kecenderungan penyakit pasien berdasarkan kelompok usia pasien. Penelitian ini dilakukan dengan menggunakan metode K-Means Clustering yang diimplementasikan pada data pasien pada suatu periode tertentu.

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

1. Latar Belakang

Data mining merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai *database* yang besar (Larose, 2005).

Salah satu fungsi *data mining* adalah *clustering*. *Clustering* merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data (Jiawei Han, Micheline, 2006). Konsep dasar dari *clustering* adalah mengelompokkan sejumlah objek ke dalam *cluster* dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu *cluster* dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* yang lainnya.

Terdapat banyak metode *clustering* yang dalam penggunaannya tergantung pada tipe data yang akan dikelompokkan dan apa tujuan dari pembuatan aplikasinya. Salah satu metode *clustering* adalah *K-Means Clustering*. Metode *K-Means Clustering* membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas (Berkhin., 2002). Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Metode *clustering* hadir dikarenakan kondisi berlimpahnya data (*data overload*) yang dialami oleh berbagai institusi, seperti rumah sakit. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam selama bertahun-tahun.

Saat ini tumpukan data pasien yang ada di rumah sakit pada umumnya hanya dimanfaatkan pada pemenuhan laporan dan grafik pasien rumah sakit, data penyakit serta biaya perawatan pasien. Tumpukan data yang ada belum menyajikan pola khusus terhadap data seperti kecenderungan penyakit pasien dikaitkan dengan usia pasien. Dengan diketahuinya pola kecenderungan tersebut diharapkan pihak rumah sakit dapat melakukan program penyuluhan dengan tepat serta dapat melakukan tindakan antisipasi prioritas layanan.

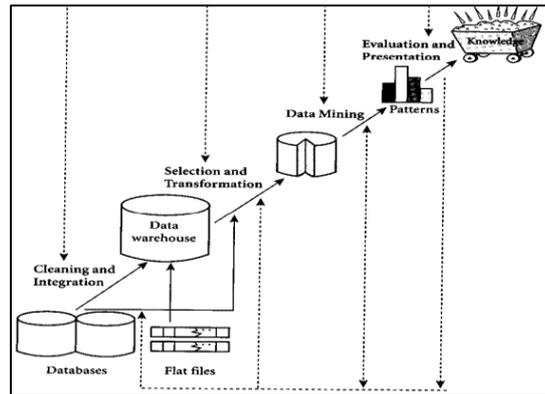
2. Landasan Teori

Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai *database* besar.

Selain itu *data mining* disebut sebagai suatu proses yang digunakan untuk menemukan pola dan kecenderungan data dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005). *Data mining* hadir dikarenakan kondisi berlimpahnya data (*overload data*) yang dialami oleh berbagai institusi, perusahaan atau organisasi. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam bertahun-tahun.

Terdapat 7 (tujuh) tahapan proses data mining, dimana 4 (empat) tahap pertama disebut tahapan *preprocessing* (terdiri dari *data cleaning, data integration, data selection, dan data transformation*), yang dalam implementasinya membutuhkan waktu sekitar 60% dari keseluruhan proses. Kemudian penggalian data (*data mining*), evaluasi pola (*pattern evaluation*) dan penyajian pola (*knowledge presentation*) (Witten and Frank, 2005). Proses data mining tersebut disajikan seperti pada Gambar 1.



Gambar 1. Tahapan Proses *Data Mining*

Clustering

Salah satu fungsi *data mining* adalah *clustering*. *Clustering* merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data (Jiawei Han, Micheline, 2006).

Clustering melakukan proses pengelompokan data tanpa berdasarkan kelas data tertentu. Proses *clustering* dapat dipakai untuk memberikan label pada kelas data yang belum diketahui. *Clustering* termasuk ke dalam *descriptive methods*, dan juga termasuk *unsupervised learning* (Ian H, Eibe., 2005). Dimana tidak ada pendefinisian kelas objek sebelumnya. Sehingga *clustering* dapat digunakan untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya.

Konsep dasar *clustering* adalah pengelompokan sejumlah objek ke dalam *cluster* dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu *cluster* dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* yang lainnya. Analisa *clustering* mengidentifikasi kumpulan objek yang memiliki kemiripan satu dengan yang lainnya. Metode *clustering* yang baik dapat menghasilkan *cluster* yang berkualitas untuk memastikan kesamaan data-data yang ada pada sebuah *cluster*.

Terdapat banyak algoritma *clustering* yang dalam penggunaannya tergantung pada tipe data yang akan dikelompokkan dan apa tujuan dari pembuatan aplikasinya. Algoritma tersebut dapat digunakan untuk mengelompokkan objek ke dalam *cluster-cluster*, kemudian dari hasil *clustering* akan dideteksi keberadaan outlier dalam data tersebut.

Algoritma *clustering* diklasifikasikan ke dalam 5 kategori, yaitu (Ian H, Eibe., 2005) :

1. **Partitioning methods.** Merupakan metode dengan pengelompokan objek dimana tiap objek dimiliki oleh 1 *cluster*. Algoritma partisi dapat meminimalkan mean kuadrat jarak dari setiap titik data ke pusat terdekatnya. Yang termasuk ke dalam metode ini adalah algoritma *K-Means*, *K-Medoid* atau *PAM*, *CLARA*, dan *CLARANS*.
2. **Hierarchical methods.** Merupakan proses *clustering* dimana pengelompokan objek dapat dilakukan dengan 2 cara yaitu *agglomerative* dan *divisive*. *Agglomerative* dimulai dengan menggabungkan beberapa *cluster* hingga menjadi satu, atau *divisive* yang dimulai dengan *cluster* yang sama kemudian dipecah menjadi beberapa *cluster* yang lebih kecil. Yang termasuk ke dalam metode ini adalah algoritma *CURE*, *BIRCH*, dan *Chameleon*.
3. **Density-based methods.** Merupakan pengelompokan objek berdasarkan tingkat kerapatan objek atau densitas. Yang termasuk dalam metode ini adalah algoritma *DBSCAN*, *DENCLUE*, dan *OPTICS*.
4. **Grid-based methods.** Adalah metode pengelompokan objek dengan menggunakan struktur data *grid multiresolusi* yang mampu menangani data berdimensi tinggi. Yang termasuk dalam metode ini adalah algoritma *CLIQUE*, *WaveCluster*, dan *STING*.
5. **Model-based methods.** Adalah pengelompokan objek dengan memodelkan tiap *cluster*, dan mencoba mengoptimasikan kesesuaian data dengan model matematika. Yang termasuk ke dalam metode ini adalah algoritma *COBWEB*.

K-Means Clustering

Metode *K-Means Clustering* merupakan metode yang paling populer dan banyak digunakan pada proses klusterisasi data. Metode *K-Means Clustering* membagi data menjadi beberapa kelompok serta dapat menerima masukan berupa data tanpa label kelas (Berkhin., 2002). Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

K-Means Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan

dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

3. Metode Penelitian

Analisis Masalah

Penelitian dilakukan terhadap data pasien. Proses pengumpulan data dilakukan melalui data rekam medik pasien pengguna BPJS Kesehatan pada bulan Februari tahun 2014.

Kemudian dilakukan proses transformasi data untuk merubah data ke dalam bentuk yang sesuai untuk di klusterisasi. Proses transformasi data dilakukan untuk menemukan jumlah diagnosa data penyakit pasien berdasarkan kelompok usia muda, parobya dan tua.

Implementasi

Implementasi *K-Means Clustering* dilakukan dengan menggunakan metode *Simple K-Means Clustering* pada perangkat lunak *data mining* WEKA.

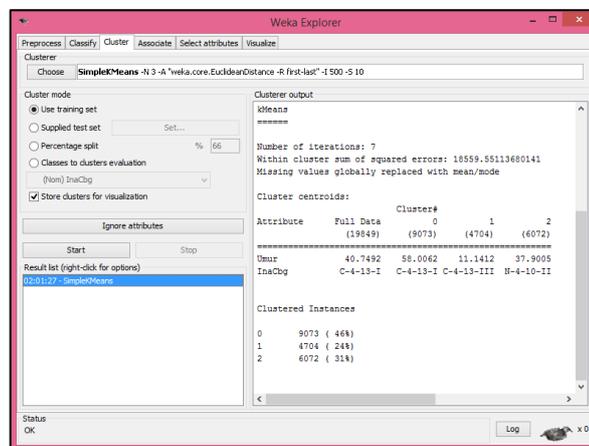
WEKA (*Waikato Environment for Knowledge Analysis*) merupakan perangkat lunak data mining yang dikembangkan oleh Universitas Waikato, New Zealand. Diimplementasikan pertama kali pada tahun 1997 dan mulai menjadi *open source* pada tahun 1999 (Mardiana T, Rudy D, 2015).

WEKA terdiri dari beberapa tools yang dapat digunakan untuk melakukan tugas *pre-processing data*, *classification*, *regression*, klustering, *association rules*, dan visualisasi (N. Sharma, et.al, 2012).

Proses clustering dilakukan dengan menggunakan perangkat Lunak WEKA dengan menggunakan jumlah cluster sebanyak 3.

4. Hasil

Hasil proses clustering dengan K-Means Clustering data penyakit pasien berdasarkan kelompok usia diperoleh seperti pada Gambar 1.



Gambar 2. Hasil Pengolahan Data

Berdasarkan proses transformasi data diperoleh bahwa :

1. Usia Minimum Pasien adalah 9 Tahun, dan Usia Maksimum Pasien adalah 94 Tahun.
2. Rata-rata usia pasien adalah 40 Tahun.

Hasil clustering dengan menggunakan K-Means Clustering dengan jumlah cluster sebanyak 3, diperoleh :

1. Cluster 1, sebesar 46 % dengan kategori usia Tua.
2. Cluster 2, sebesar 24% dengan kategori usia muda.
3. Cluster 3, sebesar 31% dengan kategori usia parobaya.
4. Penyakit dengan jumlah terbanyak adalah Kemoterapi Ringan (C-4-13-I).

5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan disimpulkan bahwa dari sejumlah pasien yang ada, persentasi usia pasien paling tinggi adalah pasien dengan usia tua dan kemudian pasien dengan usia parobaya.

Melalui penelitian ini diharapkan dapat menjadi masukan bagi Dinas Kesehatan berkaitan dengan penyuluhan kesehatan bagi masyarakat. Penelitian ini dapat dikembangkan dengan variasi data yang lebih beragam dan metode clustering yang berbeda.

Daftar Pustaka

- [1] Agusta, Y, 2007, *K-means - Penerapan, Permasalahan dan Metode Terkait*, Jurnal Sistem dan Informatika Vol. 3 (Februari 2007) : 47-60.
- [2] Berkhin Pavel, 2002, *Survey of Clustering Data Mining Techniques*. Accrue Software, Inc.
- [3] Ian H and Eibe Frank, 2005, *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco.
- [4] Jiawei Han and Micheline Kamber, 2006, *Data Mining: Concepts and Techniques Second Edition* , Morgan Kaufmann Publishers, San Francisco.
- [5] Larose, Daniel, *Discovery Knowledge in Data*, A Jhon Wiley & Sons, Inc Publication. Canada: 2005
- [6] Manning, Christopher D., Prabhakar Raghavan, Hinrich Schutze, 2009, *An Introduction to Information Retrieval*, Cambridge: Cambridge University Press.
- [7] Mardiana T, Rudy D, 2015, *Kluster Bag-of-Word Menggunakan Weka*, Jurnal Edukasi dan Penelitian Informatika (JEPIN) Vol. 1, No. 1, (Juni 2015) ISSN 2460-7041.
- [8] N. Sharma, A. Bajpai, and R. Litoriya, *Comparison the Various Clustering Algorithms of WEKA Tools*, Int. J. Emerg. Technol. Adv. Eng., vol. 2, No. 5, May 2012.
- [9] Witten, I. H and Frank, E, *Data Mining : Practical Machine Learning Tools and Techniques Second Edition*, Morgan Kauffman, San Francisco, 2005.