

## Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering

Edy Kurniawan

e-mail : [edy@umpo.ac.id](mailto:edy@umpo.ac.id)

### *Abstract*

*Patient information data located in medical centers and hospitals are in the form of ICD codes. Medical Record data is not just a pile of files, but can be used to extract the information contained therein. Some information that can be utilized is the information group of diseases that often suffered by patients in the hospital. Information group of this disease can be utilized by the manager in developing services to patients. From the side of the patient can take advantage of treatment goals are more appropriate target in accordance with the disease he suffered. To determine the analysis of this stack of medical records using groupings, including grouping of regions, gender groupings, age groupings, and ICD code groupings. In this study ICD code used is ICD 10. Preprocessing medical record data is divided into two, first data training and second is test data. From the results of data processing using cross validation 10 fold obtained the most optimal results with an accuracy of 53.8%*

**Keywords:** *Clustering Algorithm, Data Mining, Medical Record*

### 1. PENDAHULUAN

Salah satu layanan kepada masyarakat oleh dinas kesehatan adalah memberikan layanan dan penyuluhan tentang kesehatan masyarakat. Penyuluhan tersebut sering dilakukan apabila di suatu daerah telah mengalami wabah penyakit (contohnya: demam berdarah) atau biasanya penyakit musiman lainnya. Tindakan pencegahan-pencegahan supaya tidak terjangkit penyakit di masyarakat masih rendah, dan seringkali usaha pencegahan tersebut tidak pada sasarannya. Hal ini sebenarnya masih kurangnya informasi kepada dinas kesehatan terkait penyakit-penyakit yang berpotensi terjangkit di kelompok masyarakat.

Hubungan antara Rumah sakit atau balai pengobatan dengan dinas kesehatan sangatlah kuat. Hal ini dibuktikan dengan adanya pelaporan hasil rekam medis dari rumah sakit/balai pengobatan ke dinas kesehatan. Laporan rekam medis ke dinas kesehatan tersebut dilakukan setiap triwulan. Data rekam medis yang diberikan oleh pihak rumah sakit merupakan data mentah yang pada dasarnya untuk mengetahui kandungan informasi di dalamnya diperlukan pengolahan secara tersendiri. Laporan rekam medis tersebut bisa memberikan informasi penting kepada dinas kesehatan.

Rumah sakit atau balai pengobatan secara resmi sesuai aturan dari dinas kesehatan dimana setiap pendirian rumah sakit atau balai pengobatan harus mempunyai tenaga rekam medis (sesuai dengan pendidikannya), dimana perekam medis ini adalah bertugas mencatat segala sesuatu yang terjadi pada pasien yang berobat setiap saat ataupun kejadian-kejadian tertentu apabila pasien tersebut mengalami rawat inap. Data tersebut akan digunakan oleh dokter untuk acuan analisa penyakit dan landasan untuk memberikan pengobatan berdasar pada data diri serta riwayat penyakit yang di derita oleh pasien. Setelah pasien tersebut pulang dengan alasan sembuh, pulang paksa, atau meninggal, maka data tersebut akan menjadi tumpukan berkas di rumah sakit/balai pengobatan.

Tujuan dan kegunaan rekam medis adalah menunjang tercapainya tertib administrasi dalam rangka upaya peningkatan pelayanan kesehatan. Tanpa di dukung suatu

sistem pengelolaan rekam medis yang baik dan benar, maka tertib administrasi tidak akan berhasil. Kegunaan rekam medis antara lain adalah :

**Aspek Administrasi**, rekam medis mempunyai nilai administrasi, karena isinya menyangkut tindakan berdasarkan wewenang dan tanggung jawab sebagai tenaga medis dan perawat dalam mencapai tujuan pelayanan kesehatan,

**Aspek Medis**, Catatan tersebut dipergunakan sebagai dasar untuk merencanakan pengobatan/perawatan yang harus diberikan kepada pasien. Contoh adalah, identitas pasien, umur pasien, alamat pasien, status, dan masih ada beberapa yang menyangkut informasi identitas pasien.

**Aspek Hukum**, menyangkut masalah adalah jaminan kepastian hukum atas dasar keadilan, dalam rangka usaha menegakkan hukum serta penyediaan bahan tanda bukti untuk menegakkan keadilan.

**Aspek penelitian**, berkas rekam medis mempunyai nilai penelitian, karena isinya menyangkut data/informasi yang dapat digunakan sebagai aspek penelitian.

Informasi di dalam rekam medis mengandung paling tidak ada 23 variabel yang menjelaskan identitas dari pasien tersebut. Beberapa attribute yang penting tersebut diantaranya adalah :

- a) Jenis Kelamin
- b) Alamat
- c) Umur
- d) Diagnosa
- e) Informasi Penyakit (ICD)

ICD (International Classification Of Diseases), standart pengelompokan penyakit yang dilakukan oleh WHO (World Health Organization). Kode ICD ini berkembang terus, setiap waktu ada perubahan mendasar pada perkembangan penyakit baru yang ada di dunia. Sampai saat ini teknik pengkodean ICD tersebut mencapai ICD-10. ICD-10 adalah acuan seluruh penyelenggara layanan kesehatan (Rumah sakit, balai pengobatan, puskesmas) untuk dijadikan pedoman dalam melakukan arsip. [1]

Di dalam pedoman ICD-10 dilakukan pengelompokan dari penyakit menjadi 20 kelompok penyakit. Dalam ICD-10 pengelompokan ini didasarkan pada karakter dan jenis penyakit tersebut. Hal ini difungsikan untuk memudahkan dalam administrasi perekam medis sebab keterangan detail dari seorang pasien akan menjadi kesulitan tersendiri dalam pengarsipan oleh penyelenggara layanan kesehatan apabila tidak dikelompokkan.

## 2. KAJIAN LITERATUR

*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata *mining* sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basisdata. Beberapa teknik yang sering disebut-sebut dalam literatur *data mining* antara lain yaitu *association rule mining, clustering, klasifikasi, neural network*, dan lain-lain.[2]

### Preprocess

Pada proses ini data dipersiapkan agar dapat digunakan untuk digali informasinya. Hal ini dilakukan untuk mendapatkan hasil analisis yang lebih akurat dalam pemakaian teknik-teknik *machine learning* maupun *data mining*. Komponen yang terdapat dalam data itu sendiri terdiri dari : Obyek (record, point, case, sampel, entitas, instan) dan Atribut / variabel / field yaitu karakteristik dari obyek (status pernikahan, umur, dll)

*Data cleaning* terdapat proses pembersihan data yang kosong atau data-data lain yang dapat mengakibatkan *noise/error* dengan memperkecil adanya data *outlier*. *Data Cleaning* menjadi proses yang sangat penting dikarenakan data perlu dibersihkan agar analisa menjadi lebih akurat.[3]

Ada beberapa cara transformasi data yang dilakukan sebelum kita menerapkan suatu metode tergantung pada metode apa yang kita lakukan untuk proses *data mining*. Sebelum kita mempelajari teknik/metode yang digunakan dalam *data mining*, ada baiknya kita bedakan dulu metode belajar (*learning*) secara garis besar ke dalam dua pendekatan : *supervised* dan *unsupervised*. Dalam pendekatan pertama, *unsupervised learning*, metode kita terapkan tanpa adanya latihan (training) dan tanpa ada label dari data. Misalkan kita punya sekelompok pengamatan atau data tanpa ada label (output) tertentu yang menandai kemana data dikelompokkan.

### Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah di definisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas-kelas tambahan. Salah satu contoh yang mudah dan populer adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi.[4]

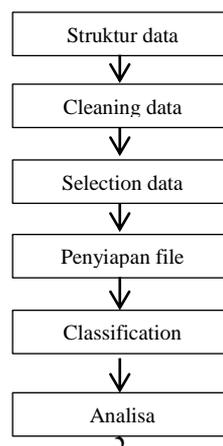
### Clustering

*Clustering* adalah proses pengelompokan sejumlah data/obyek kedalam kelompok-kelompok data (klaster) sehingga setiap klaster akan berisi data yang saling mirip. *Clustering* adalah salah satu teknik *unsupervised learning* dimana kita tidak perlu melatih metode tersebut atau dengan kata lain, tidak ada *fase learning*. Tujuan dari metode clustering adalah untuk mengelompokkan sejumlah data atau objek kedalam klaster sehingga setiap klaster akan terisi data yang semirip mungkin.[5]

Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data *clustering*. Ada dua macam teknik klaster yang cukup sering dipakai. Yang pertama adalah *k-means* (termasuk partitioning cluster) dan yang berikutnya adalah *hierarcichal clustering*. [6]

## 3. METODE PENELITIAN

Untuk melakukan penelitian terhadap tumpukan data rekam medis di Rumah Sakit ini adalah dengan membagi data menjadi dua. Data pertama digunakan untuk melakukan training pada aplikasi, kemudian data berikutnya digunakan sebagai data uji dari metode yang telah digunakan. Metode clustering akan digunakan dalam penelitian ini.



Gambar 1. Sistematika Penelitian [7]

### **Cleaning Data**

Tahap awal dari preprocessing data mining ini adalah *cleaning data*, hal ini dilakukan untuk membuang data yang mempunyai informasi tidak lengkap. Dimungkinkan banyak data yang mempunyai informasi tidak lengkap, seperti tidak masuknya informasi umur, informasi wilayah, informasi jenis kelamin, informasi pengelompokan penyakit.

### **Seleksi Data**

Data-data yang sudah mempunyai informasi lengkap dalam setiap attribute selanjutnya data tersebut dilakukan seleksi. Seleksi ini dilakukan untuk mengelompokkan attribute sesuai dengan informasi yang dibutuhkan. Dari attribute yang di lakukan seleksi adalah pada attribute alamat, attribute umur dan attribute ICD-10. Sedangkan attribute gender merupakan target attribute.

Pada penelitian ini seleksi data dilakukan pada data yang mempunyai informasi wilayah diluar kabupaten ponorogo. Sebab Rumah sakit yang bersangkutan juga menerima pasien dari luar ponorogo, data yang mengandung informasi wilayah diluar kabupaten ponorogo akan dibuang.[8]

### **Pengelompokan Umur**

Proses pengelompokan umur mempunyai dua metode (1) menggunakan discrete (2) pengelompokan umur berdasar pada aturan WHO. Pengelompokan umur secara discrete tidak relevan apabila diterapkan pada penelitian ini, sebab penentuan discrete pada preprocessing yang ada pada aplikasi weka 3.6.6 menggunakan sistem interval. Jadi pengelompokan umurnya tidak mendasar pada jenjang.[9]

Pengelompokan umur untuk kesehatan mendasar pada metode yang dilakukan oleh WHO yaitu umur dalam tahun dikelompokkan menjadi tiga kelompok jenjang umur (kelompok bayi dan balita, kelompok muda dan dewasa, kelompok tua).

Tabel 1. Pengelompokan umur berdasarkan jenjang menurut WHO

<b>Umur</b>	<b>Jenjang</b>
15 ≤	Bayi dan Anak-anak
15 s/d 50	Muda dan Dewasa
≥ 50	Tua

### **Pengelompokan wilayah**

Isi pada attribute alamat pada rekam medis merupakan keterangan alamat secara lengkap, dimana alamat yang tercantum di dalamnya memuat informasi RT/RW, Desa/kelurahan, jalan dan nomernya, Kecamatan. Pada proses pengelompokan wilayah ini penulis menggunakan *MySql* dan *php* untuk melakukan pengelompokan secara otomatis. Teknik pengelompokan ini di dasar pada informasi alamat yang mengandung kata kecamatan yang ada di ponorogo.

Ada dua tahap proses pengelompokan wilayah, (1) conversi alamat menjadi nama kecamatan, (2) pengelompokan kecamatan menjadi wilayah seperti dalam pembagian WHO (Wilayah perkotaan, wilayah pedesaan, wilayah pesisir / pegunungan).

## **4. HASIL PENELITIAN**

Tahapan preprocessing pada penelitian ini telah diolah sebanyak 1921 data mentah. Pembuangan data *missing value* mendapatkan hasil data bersih sebanyak 1809 data rekam medis.

Tahapan selanjutnya adalah melakukan seleksi data yang tidak masuk dalam kategori untuk analisa. Sebagai attribute yang dilakukan seleksi ini adalah atributte kelompok wilayah. 1809 data bersih tersebut setelah dilakukan seleksi untuk mencari data yang sebenarnya akan digunakan untuk analisa penelitian ini, maka peneliti mendapatkan 1628 data bersih yang sudah terseleksi.

### Hasil uji data menggunakan *cross validation*

Untuk melakukan uji data dengan metode *cross validation* ini peneliti menggunakan 2 pengujian, dimana pengujian menggunakan *cross validation* dengan 10 *fold*. Dari hasil pengujian di dapatkan hasil sebagai berikut :

Tabel 2. Hasil percobaan dengan menggunakan 10 fold

	Fold										Rata-rata
	1	2	3	4	5	6	7	8	9	10	
Data Uji	53,2	33	67,9	52,3	51,9	49,1	76,9	51,9	44,4	57,4	53,8%

Dengan melihat dari hasil percobaan yang menggunakan metode *cross validation* di ketahui *Correctly Classified Instances* adalah menggunakan 10 fold.

Dari pohon keputusan yang dihasilkan dari data uji menggunakan data training sebanyak 542 di dapat rule untuk menentukan kecenderungan penyakit yang di derita oleh masyarakat di dasarkan pada jenis kelamin, umur, serta wilayah tempat tinggalnya. Berikut rule yang didapat untuk mengetahui dari hasil informasi ICD berdasarkan dari pohon keputusan yang ada.

```
#Rule dari Hasil Pohon Keputusan Analisa Rekam Medis
if kelompok umur = anak-anak dan bayi then
  if Gender = P then
    if (wilayah = perkotaan) then
      A00-B99
    else if wilayah = pedesaan then
      Z00-Z99
    else P00-P96
    end if
  else A00-B99
  end if
else if kelompok umur = muda dan dewasa then
  if Gender = L then
    if wilayah = perkotaan then
      S00-T98
    else if wilayah = pedesaan then
      A00-B99
    else S00-T98
    end if
  else O00-O99
  end if
else I00-I99
end if
```

Gambar 2. Pseudocode hasil pohon keputusan analisa rekam medis

## 5. SIMPULAN

Data rekam medis mempunyai lebih dari 15 attribute, namun dalam penelitian ini untuk menentukan kelompok penyakit yang diderita oleh sekelompok masyarakat dalam wilayah tertentu hanya di butuhkan 4 atributte saja. Atributte yang mempengaruhi pengelompokan penderita penyakit dalam sekelompok wilayah adalah (1) jenis kelamin, (2) Wilayah, (3) Umur, (3) kelompok ICD. Penelitian ini menggunakan metode algoritma clustering.

Dari 1921 data kotor setelah dilakukan pembersihan data tidak lengkap dan melalui proses seleksi sesuai yang dibutuhkan dalam analisa rekam medis maka didapat data 1628 data bersih. Dari data yang ada dilakukan split untuk digunakan 1/3 data training dan 2/3 data test. Pengujian dilakukan menggunakan metode cross validation dengan *fold* 10 yang mempunyai hasil paling optimal. Dengan melakukan cross validation secara random dari empat attribute yang diujikan ini di dapat informasi gain tertinggi pada kelompok ICD.

Pada data rekam medis ini dalam pengujian yang menggunakan pruning maupun unpruning mempunyai hasil yang sama. Dari hasil pengujian keseluruhan data dengan memakai cross calidation 10 fold di dapat akurasi pengujian sebesar 53,8 %.

## 6. REFERENSI

- ICD-10 (*International Classification of Diseases*)  
[<http://www.who.int/classifications/icd/en/>] diakses Mei 2017
- Duen-Yian Yeh, Ching-Hsue Cheng, Yen-Wen Chen, *A predictive model for cerebrovascular disease using data mining*, *Expert Systems with Applications* 38 (2011) 8970–8977
- Jiawei Han and Micheline Kamber, *“Data Mining Concepts and Techniques”*, Second Edition, 2006
- Ian H. Witten and Eibe Frank, ” *Data Mining Practical Machine Learning Tools and Techniques*”, Morgan Kaufmann Publishers is an imprint of Elsevier., San Francisco, (2005)
- Berkhin, Pavel. *Survey of Clustering Data Mining Techniques*. Accrue Software, Inc. 2002.
- Krzysztof J. Cios, Witod Pedrycz, *“Data Mining A Knowledge Discovery Approach”*, 2007
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. *Discovering datamining: From concept to implementation*. New Jersey: Prentice Hall. (1997).
- BPPS Ponorogo, “Ponorogo dalam Angka” 2015

Remko R.B, Eibe Frank, Mark Hall., “Weka Manual for Version 3.6.6”, Oktober 2011.