

Implementasi Data Mining Pola Penyakit Pasien Menerapkan Decision Tree J48 (Studi Kasus : RS.Estomihi)

Rohani Situmeang

Program Studi Teknik Informatika, STMIK Budi Darma, Medan, Indonesia
Jalan Sisingamangaraja No. 338 Medan, Indonesia

Abstrak

Berdasarkan hal tersebut peneliti tertarik Suatu pencarian pola dari data yang berskala besar dan yang mengasosiasikan data yang satu dengan yang lainnya dengan menggunakan algoritma Decision tree j48 .Dengan demikian proses penanganan awal dari pasien akan kurang tidak efektif jika dalam menentukan jenis penyakit pasien dibutuhkan waktu yang lama. Selain dari masalah tersebut, hasil analisis juga sangat diperlukan untuk menghindari kesalahan pengobatan, karena jika salah dalam melakukan pengobatan maka bisa saja penyakit yang diderita pasien bisa lebih parah.

Kata Kunci: Algoritma Decision tree j48 ,Data mining, Pola penyakit.

Abstract

Based on this the researcher is interested in A search pattern of large-scale data and that associate data with one another using Decision tree j48 algorithm. Thus the initial handling process of the patient will be less effective if in determining the type of patient's disease takes a long time . Apart from the problem, the results of analysis is also very necessary to avoid treatment errors, because if one in the treatment then it could be the illness suffered by patients can be more severe.

Keywords: Decision tree Algorithm j48, Data mining, Pattern of disease.

1. PENDAHULUAN

Rumah Sakit sebagai salah satu fasilitas pelayanan kesehatan, merupakan bagian dari sumber daya kesehatan yang sangat di perlukan dalam mendukung penyelenggaraan upaya kesehatan. Termasuk Rumah Sakit Umum Estomihi merupakan salah satu instansi kesehatan yang ada di kota Medan. Banyak pasien yang menderita penyakit datang ke rumah sakit tersebut untuk melakukan perobatan. Pola penyakit merupakan suatu bentuk atau pola yang menyebabkan terjadinya suatu penyakit, pola tersebut dapat terbentuk berkat adanya penyebab-penyebab yang mempengaruhi penyakit seperti jenis pekerjaan yang dapat menyebabkan adanya penyakit, usia, jenis kelamin dan sebagainya. Pada rumah sakit Estomihi Medan pola penyakit telah mengalami perubahan dengan adanya transisi epidemiologi. Proses transisi epidemiologi adalah terjadinya perubahan pola penyakit dan kematian yang ditandai dengan beralihnya penyebab kematian yang dahulu didominasi oleh penyakit infeksi bergeser ke penyakit non-infeksi atau penyakit tidak menular.

Berdasarkan masalah yang diuraikan diatas maka penulis menggunakan Algoritma Decision Treej48. Algoritma Decision Treej48 merupakan salah satu algoritma klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node dipresentasikan attribute, cabangnya merepresentasikan nilai dari attribute dan daun merepresentasikan kelas. Dengan demikian maka penyakit yang diderita pasien dapat klasifikasikan berdasarkan data rekam medis dari pasien tersebut, sehingga pola dari penyakit pasien tersebut dapat diketahui. Penggunaan aplikasi weka dalam pembuatan data mining dikarenakan weka memiliki algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling. Kelebihan weka terletak pada algoritma yang semakin lengkap dan canggih, tetapi keberhasilan data mining tetap terletak pada faktor pengetahuan manusia yang melakukan pengumpulan data yang berkualitas dan pengetahuan yang diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

2. LANDASAN TEORI

2.1 Data mining

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam *database* besar yang merupakan bagian dari proses *Knowledge Discovery in Database* (KKD) yang terdiri dari beberapa tahapan seperti pemilihan data, prapengolahan, transformasi, *data mining* dan evaluasi hasil[4].

2.2 Decision Tree J48

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) dimana setiap node merepresentasikan nilai dari atribut, cabangnya merepresentasikan nilai dari atribut dan daun merepresentasikan kelas. Node yang paling atas dari *decision tree* disebut sebagai *root*[6]. *Decision tree* merupakan metode klasifikasi yang paling popular digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami.

Decision tree J48 merupakan implementasi dari algoritma C4.5 yang memproduksi *decision tree*. Ini merupakan standar algoritma yang digunakan dalam *machine learning*. Dalam *k-fold cross validation*, data pengujian dipisah secara acak ke dalam k himpunan bagian yang *mutually exclusive* atau *folds* (lipatan), D_1, D_2, \dots, D_k , yang masing-masing kurang lebih berukuran sama. Pelatihan dan pengujian dilakukan sebanyak k kali. Pada iterasi ke-I, partisi D_i , digunakan sebagai data ter dan partisi sisanya digunakan bersama untuk melatih model. Dalam iterasi pertama, yaitu himpunan bagian D_2, \dots, D_k

secara bersama bertindak sebagai data pelatihan untuk memperoleh model pertama yang diuji pada D_1 , iterasi kedua dilatih pada himpunan bagian D_1, D_2, \dots, D_k dan diuji pada D_2 dan seterusnya[6].

Data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*.

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2 p_i \tag{1}$$

Merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya:

- S : Himpunan kasus
- n : Jumlah partial S
- p_i : Jumlah kasus pada partisi ke-i

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{2}$$

Merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya:

- S : Himpunan kasus
- n : Jumlah partisi atribut A
- | S_i | : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam S

2.3 Pola Penyakit Pasien

Frekuensi dan penyebaran masalah kesehatan (khususnya penyakit) pada umumnya bervariasi. Selain itu dalam kegiatan analisis *epidemiologi* membutuhkan kesadaran adanya interaksi dalam menimbulkan penyakit. Adapun faktor-faktor yang dapat menganalisa suatu pola penyakit[8] adalah sebagai berikut:

1. Tempat
2. Orang
3. Waktu.

3. ANALISA DAN PEMBAHASAN

Untuk mengetahui pola penyakit yang terjadi di rumah sakit Estomihi Medan, maka diambil data pasien dan penyakit yang terjadi pada tahun 2012 sampai 2017 dengan 5 kasus penyakit terbanyak yang diderita oleh pasien sebagai berikut:

Tabel 1. 5 Penyakit Terbanyak di Rumah Sakit Estomihi Medan dari 2012-2017

No.	Daftar Penyakit
1	Diabetes
2	Hypertensi
3	TBC
4	Gastroenteritis
5	Dispepsia

Langkah-langkah algoritma *decision tree* j48 dalam melakukan proses data mining adalah sebagai berikut:

Mengelompokkan setiap variabel untuk mendapatkan attribute yang akan digunakan untuk menghitung *entropy* dari kasus agar mendapatkan akar dan cabang dari pola penyakit. Dalam tahap ini juga tahap klasifikasi pada data mining dilakukan, dimana data tersebut dikelompokkan dan diklasifikasikan berdasarkan variabel-variabel yang telah ditentukan memilih jawaban:

Tabel 2. Klasifikasi Data

Variabel	Attribut	Jumlah	Meninggal	Tidak Meninggal
Jenis Kelamin	Lk	11	6	5
	Pr	19	7	12
Jumlah		30	13	17
Penyakit	Diabetes	8	4	4
	Hypertensi	5	2	3
	TBC	7	4	3
	Dispepsia	4	1	3
	Gastroenteritis	6	2	4
Jumlah		30	13	17
Usia	≤ 15	6	3	3
	16 s/d 50	16	6	10

	≥ 51	8	3	5
Jumlah		30	12	18
Pekerjaan	IRT	11	5	6
	Tidak Bekerja	1	0	1
	Karyawan Swasta	9	4	5
	Mahasiswa	3	1	2
Jumlah	Pelajar	6	3	3
Wilayah Tinggal	Medan Kota	7	3	4
	Medan Amplas	12	5	7
	Medan Denai	3	1	2
	Tembung	2	1	1
	Medan Sunggal	1	1	0
	Medan Area	5	2	3
Jumlah		30	13	17
Golongan darah	A	3	1	2
	B	5	2	3
	AB	7	2	5
	O	15	8	7
Jumlah		30	13	17

1. Menghitung nilai *entropy* total dari semua kasus:

$$\begin{aligned}
 \text{Entropy (total)} &= \left(-\frac{13}{30} * \log_2 \left(\frac{13}{30}\right)\right) + \left(-\frac{17}{30} * \log_2 \left(\frac{17}{30}\right)\right) \\
 &= (-0,43 * -0,06) + (-0,57 * 0,05) \\
 &= 0,026 + 0,03 \\
 &= 0,06
 \end{aligned}$$

2. Menghitung nilai *entropy* tiap nilai atribut dan *gain* tiap atribut. Berikut perhitungan nilai *entropy* nilai jenis kelamin:

Jenis Kelamin

$$\begin{aligned}
 \text{Entropy (Lk)} &= \left(-\frac{6}{11} * \log_2 \left(\frac{6}{11}\right)\right) + \left(-\frac{5}{11} * \log_2 \left(\frac{5}{11}\right)\right) \\
 &= (-0,55 * -0,04) + (-0,45 * -0,04) \\
 &= 0,022 + 0,018 \\
 &= 0,04
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy (Pr)} &= \left(-\frac{7}{19} * \log_2 \left(\frac{7}{19}\right)\right) + \left(-\frac{12}{19} * \log_2 \left(\frac{12}{19}\right)\right) \\
 &= (-0,368 * -0,133) + (-0,632 * -0,101) \\
 &= 0,05 + 0,06 \\
 &= 0,11
 \end{aligned}$$

3. Menghitung nilai *gain* dari atribut. Berikut perhitungannya:

$$\begin{aligned}
 \text{Gain (total, Jenis Kelamin)} &= 0,06 - \left(\left(\frac{11}{30}\right) * 0,04\right) + \left(\left(\frac{19}{30}\right) * 0,11\right) \\
 &= 0,06 - (0,015 + 0,069) \\
 &= 0,06 - 0,084 \\
 &= 0,024
 \end{aligned}$$

4. Dari hasil perhitungan *gain* tersebut, dapat diketahui bahwa variabel dengan *gain* tertinggi adalah variabel usia, sehingga usia dapat menjadi node akar. Ada tiga nilai *attribute* dari usia yaitu ≤ 15 , $16 \text{ s/d } 50$, ≥ 51

4. IMPLEMENTASI

Pengujian dilakukan untuk pada perangkat lunak Weka versi 3.8.1. Algoritma *Decision Tree j48* telah terdapat pada *Weka Software Package*. Pengujian dengan *sampling* yang digunakan yakni *10-fold cross-validation* menggunakan 7 *attribut* dari *dataset*. Adapun percobaan yang dilakukan adalah sebagai berikut:

```

Classifier output
Size of the tree : 1

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances 17 56.6667 %
Incorrectly Classified Instances 13 43.3333 %
  
```

Gambar 1. Hasil *Training Set* Weka

Sedangkan hasil prediksi Weka menggunakan algoritma *decision tree* dapat dilihat dari gambar berikut:

penyakit.arff *									
Relation: Hani-weka.filters.unsupervised.attribute.Remove-R8_predicted									
No.	1: Jenis Kelamin	2: Penyakit	3: Usia	4: Pekerjaan	5: Wilayah	6: Golongan Darah	7: prediction margin	8: predicted Kondisi	9: Kondisi
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	Perempuan	Hyperte...	Tua	IRT	Amplas	A	-0.133333	Tidak Meninggal	Mening...
2	Perempuan	Dispep...	Tua	IRT	Tembu...	O	-0.133333	Tidak Meninggal	Mening...
3	Perempuan	Hyperte...	Tua	Karyawan...	Amplas	O	-0.133333	Tidak Meninggal	Mening...
4	Perempuan	Diabetes	Dew...	IRT	Kota	AB	-0.133333	Tidak Meninggal	Mening...
5	Perempuan	Gastroe...	Muda	Pelajar	Kota	O	-0.133333	Tidak Meninggal	Mening...
6	Perempuan	Diabetes	Tua	IRT	Amplas	O	-0.133333	Tidak Meninggal	Mening...
7	Perempuan	Diabetes	Dew...	IRT	Area	AB	-0.133333	Tidak Meninggal	Mening...
8	Laki-Laki	TBC	Muda	Pelajar	Denai	B	-0.133333	Tidak Meninggal	Mening...
9	Laki-Laki	Diabetes	Dew...	Karyawan...	Amplas	B	-0.133333	Tidak Meninggal	Mening...
10	Laki-Laki	TBC	Dew...	Mahasiswa	Amplas	O	-0.133333	Tidak Meninggal	Mening...
11	Laki-Laki	TBC	Dew...	Karyawan...	Sunggal	O	-0.133333	Tidak Meninggal	Mening...
12	Laki-Laki	Gastroe...	Muda	Pelajar	Area	O	-0.133333	Tidak Meninggal	Mening...
13	Laki-Laki	TBC	Dew...	Karyawan...	Kota	O	-0.133333	Tidak Meninggal	Mening...
14	Perempuan	Diabetes	Dew...	IRT	Amplas	O	0.133333	Tidak Meninggal	Tidak ...
15	Perempuan	Dispep...	Dew...	Karyawan...	Denai	O	0.133333	Tidak Meninggal	Tidak ...
16	Perempuan	Diabetes	Dew...	Karyawan...	Area	A	0.133333	Tidak Meninggal	Tidak ...
17	Perempuan	Dispep...	Tua	IRT	Amplas	AB	0.133333	Tidak Meninggal	Tidak ...
18	Perempuan	Dispep...	Tua	IRT	Amplas	A	0.133333	Tidak Meninggal	Tidak ...
19	Perempuan	Diabetes	Tua	IRT	Kota	B	0.133333	Tidak Meninggal	Tidak ...
20	Perempuan	Gastroe...	Muda	Pelajar	Kota	O	0.133333	Tidak Meninggal	Tidak ...
21	Perempuan	Hyperte...	Dew...	IRT	Tembu...	O	0.133333	Tidak Meninggal	Tidak ...
22	Perempuan	Hyperte...	Tua	IRT	Area	AB	0.133333	Tidak Meninggal	Tidak ...
23	Perempuan	Gastroe...	Muda	Tidak Bek...	Kota	AB	0.133333	Tidak Meninggal	Tidak ...
24	Perempuan	Gastroe...	Muda	Pelajar	Amplas	AB	0.133333	Tidak Meninggal	Tidak ...
25	Perempuan	Hyperte...	Dew...	Karyawan...	Area	O	0.133333	Tidak Meninggal	Tidak ...
26	Laki-Laki	TBC	Dew...	Mahasiswa	Denai	O	0.133333	Tidak Meninggal	Tidak ...

Gambar 2. Hasil Prediksi Algoritma *Decision Tree* J48

Berdasarkan gambar 2. didapatkan hasil bahwa pasien penderita *hypertensi* dengan jenis kelamin perempuan dan berada dalam rentang usia yang tua memiliki pekerjaan sebagai ibu rumah tangga memiliki golongan darah A dan tinggal di wilayah Amplas memiliki resiko kematian yang paling tinggi dengan rasio prediksi 0,133333, kemudian yang kedua adalah perempuan penderita *dispepsia* yang memiliki rentang usia tua memiliki golongan darah O dan tinggal di wilayah Tembung adalah pasien yang memiliki resiko kematian terbesar kedua

5. KESIMPULAN

Berdasarkan hasil-hasil analisis dan percobaan yang dilakukan pada bab sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Pola penyakit pasien dapat diketahui dengan menggunakan data rekam medis dari rumah sakit dan mengambil data-data yang dapat mempengaruhi penyakit pasien tersebut.
2. Algoritma *decision tree* dapat dengan mudah memproses data karena mudah beradaptasi pada interaksi kompleks antar variabel. Setiap cabang pohon dapat terdiri atas kombinasi dari variabel dan variabel yang sama dapat muncul lebih dari satu di berbagai cabang pohon. Hal ini dapat berguna dalam mengetahui bagaimana sebuah variabel bergantung kepada lainnya dan bagaimana spesifik variabelnya.

3. Penghitungan data mining dengan weka sangat mudah dilakukan dengan mengi-input data ke dalam sebuah file yang berformat .csv kemudian menyimpan file tersebut dengan format .arff yang terdapat di dalam weka untuk mendapatkan hasil prediksi dari aplikasi tersebut..

REFERENCES

- [1] A.Aneta, “Implementasi Kebijakan Program Penanggulangan Kemiskinan Perkotaan (P2KP) di Kota Gorontalo”, Jurnal Administrasi Publik, vol.I, pp. 54-65. 2010
- [2] M.H. Iswanto, A.E.Permanasari, et.al, “Pemanfaatan Teknik Data Mining Untuk Diagnosis Penyakit Tuberculosis (TBC)”,vol.II, pp. 121-126.2015
- [3] Sartikha, Maria, et.al, “Analisis Profil Mahasiswa Politeknik Negeri Batam dengan Teknik Data Mining Asosiasi dan Clustering”, vol.VIII, pp.16-21.2016
- [4] G.Abdillah, et.al, “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma K-Means”, vol.I, SENTIKA, pp.498-506.2016
- [5] A.B.Cahyono, “Analisis Pemanfaatan Small Disjunct Pada Decision Tree dengan Algoritma Genetika”, pp.6-9.2010
- [6] H.Susanto, “ Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu”, vol.II, pp.222-231.2014
- [7] F.F.Harranto, S.Hansun, “Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT.WISE”, Vol.III, pp. 95-102.2017
- [8] P.Meilina, “Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi”, Universitas Muhammadiyah, Vol.I, pp.1-9.2016
- [9] H.Kurniawan, Fujiati, et.al, “Analisa Pola Data Penyakit Rumah Sakit dengan Menerapkan Metode Association Rule Menggunakan Algoritma Apriori”, pp.195-201.2014
- [10] Y.Wobisono, Tutorial WEKA, Bandung: Universitas Pendidikan Indonesia.2017