



Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan *International Classification Diseases* (ICD-10)

Yudha Aditya Fiandra^a, Sarjon Defit^b, Yuhandri^c

^aPasca Sarjana, Magister Komputer, Universitas Putra Indonesia, yudhaaditya1994@gmail.com

^bPasca Sarjana, Magister Komputer, Universitas Putra Indonesia, sarjonde@yahoo.co.uk

^cPasca Sarjana, Magister Komputer, Universitas Putra Indonesia, yuhandri.yunus@gmail.com

Abstract

The medical record data is the patient's current record of medical records, the medical record data only being data stacked and not traced to generate useful knowledge for the hospital. This study can process the medical record data to classify the disease that occurs in sleeping sickness based on ICD-10. The method used in this research is C4.5 algorithm method by using attribute of international disease code as attribute of destination label as many as 21 international disease group, that is: A00-B99 up to Z00-Z99. This study yields a decision of the value code, C4.5 code can represent as many as 14 attribute values of disease code objectives and data percentage that read more than 66%. The conclusion of this research is C4.5 algorithm help classify international disease code based on ICD-10 and decision tree making which can give information of any disease that often happened at hospital

Keywords: data mining, classification, C4.5, medical records, ICD-10

Abstrak

Data rekam medis adalah catatan khusus pasien yang sedang, seringkali data rekam medis hanya menjadi data yang menumpuk dan tidak dilakukan penelusuran untuk menghasilkan pengetahuan yang berguna bagi rumah sakit. Penelitian ini bertujuan mengolah tumpukan data rekam medis untuk mengklasifikasikan jenis penyakit yang terjadi pada rumah sakit berdasarkan pada kode penyakit internasional (*International Classification of Disease*) ICD-10. Metode yang digunakan dalam penelitian ini adalah metode algoritma C4.5 dengan menggunakan atribut kode penyakit internasional sebagai atribut label tujuan sebanyak 21 kelompok penyakit internasional, yaitu: A00-B99 sampai dengan Z00-Z99. Penelitian ini menghasilkan pohon keputusan dari 21 berupa value atribut tujuan kode penyakit, algoritma C4.5 dapat merepresentasikan sebanyak 14 value atribut tujuan kode penyakit dan persentasi data yang terbaca lebih dari 66%. Kesimpulan dari penelitian ini yaitu algoritma C4.5 membantu mengklasifikasikan kode penyakit internasional berdasarkan ICD-10 dan membentuk pohon keputusan yang dapat memberikan informasi penyakit apa saja yang sering terjadi pada rumah sakit

Kata kunci: data mining, klasifikasi, C4.5, rekam medis, ICD-10

© 2017Jurnal RESTI

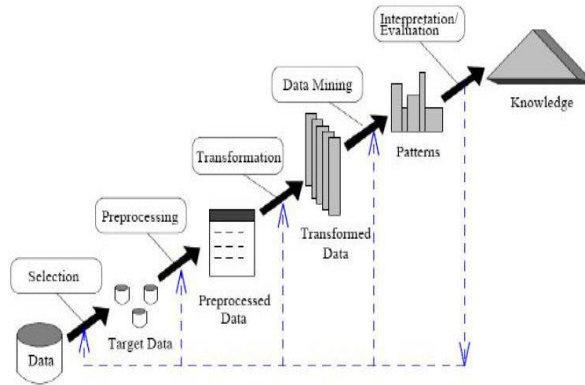
1. Pendahuluan

Knowledge discovery in database (KDD) merupakan proses untuk menemukan informasi yang berguna dalam database. Seluruh proses KDD biasanya terdiri dari langkah-langkah, yaitu memahami bidang aplikasi, membuat data target yang ditetapkan dari data mentah yang tersimpan dalam *database*, pembersihan data dan *preprocessing* data [1].

Istilah *knowledge discovery in database* atau mencari pengetahuan dalam *database* atau KDD singkatnya, mengacu pada proses pencarian pengetahuan dalam data yang luas dan menekankan pada penerapan

metode tingkat tinggi atau metode penambangan data tertentu. Ini menarik minat para peneliti dalam melakukan pengembang penelitian baik dalam bidang *machine learning* atau pembelajaran mesin [2].

Pendekatan data mining menjadi sangat penting dalam industri kesehatan dalam mengambil keputusan berdasarkan analisis dari data klinis yang besar. *Data mining* berperan dalam proses mengekstrak informasi tersembunyi dari dataset yang besar dan teknik-teknik seperti klasifikasi, *clustering*, regresi dan asosiasi telah digunakan oleh bidang medis [3].



Gambar 1. Proses Knowledge discovery in database[2]

Salah satu teknik klasifikasi yang sering digunakan oleh peneliti adalah penerapan algoritma C4.5. Algoritma C4.5 adalah algoritma yang menghasilkan pohon keputusan. Ini memiliki masukan berupa sampel klasifikasi [4]. Penerapan fungsi algoritma C4.5 untuk menghasilkan tingkat akurasi data sebagai dataset yang mengandung data dalam jumlah besar [5].

2. Tinjauan Pustaka/Penelitian Sebelumnya

Khusus untuk fungsionalitas *data mining* menggunakan klasifikasi, beberapa penelitian telah dilakukan misalnya oleh Wenefrida pada tahun 2013 [6], mengenai klasifikasi data rekam medis berdasarkan kode penyakit internasional dimana dalam penelitian ini terfokus pada bagaimana melakukan klasifikasi data rekam medis pasien agar diklasifikasikan berdasarkan kode penyakit menurut standar ICD-10. Klasifikasi penyakit berdasarkan tumpukan data rekam medis, menggunakan salah satu metode dalam *data mining*, yaitu algoritma C4.5, untuk mencapai tujuan penelitian, maka dipilih 4 atribut sesuai data rekam medis. Data rekam medis yang dimaksud terdiri dari atribut diagnosa penyakit berdasarkan *International Classification of Diseases-10th* (ICD-10), jenis kelamin, umur pasien, bulan masuk pasien ke rumah sakit. Hasil penelitian menunjukkan bahwa ada 5 jenis klasifikasi penyakit, yaitu A00-B99, I00-I99, J00-J99, O00-O99 dan Z00-Z99.

Penelitian mengenai klasifikasi *data mining* juga dilakukan oleh Yulia pada tahun 2016 [7] dalam bidang medis mengenai metode klasifikasi penyakit kanker serviks, dalam penelitian ini, data yang digunakan diperoleh dari rekam medis hasil uji pap smear. Ada 38 gejala dan 7 kelas, *Naïve Bayes*, *Support Vector Machines* (SVM), dan *Random Forest Tree* digunakan untuk mengevaluasi kinerja pengklasifikasi. Matrik kinerja yang digunakan dalam penelitian ini adalah akurasi, daya ingat, presisi, dan kurva ROC. Berdasarkan kinerja matrik, *Random Forest Tree* adalah *classifier* terbaik di antara pengklasifikasi lainnya untuk mengklasifikasikan hasil Pap smear

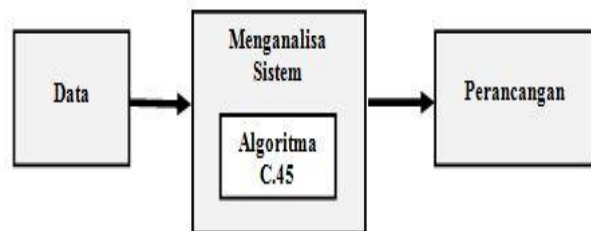
Penelitian lain juga dilakukan oleh Sucipto tahun 2016 [8], mengenai klasifikasi penyakit pada ikan, untuk memperoleh kualitas data dengan baik sehingga bisa untuk mengidentifikasi penyakit ikan. Kualitas hubungan data antara gejala penyakit ikan harus dipahami untuk diketahui akurasi klasifikasi yang diperoleh. Metode ini diperlukan untuk mengekstrak informasi dari data yang diperoleh. Ada banyak algoritma klasifikasi data mining tapi, algoritma C4.5 sesuai untuk penelitian ini digunakan untuk membentuk pohon keputusan.

3. Metodologi Penelitian

Penelitian fokus pada proses menganalisis data rekam medis dengan algoritma C4.5 menggunakan program *RapidMiner Studio 7.5 (Tools Data Mining)* untuk memperoleh hasil klasifikasi. Ada 4 atribut yang dipakai dalam penelitian, yaitu: (1) umur yang dikelompokkan dalam kategori bayi & anak (umur <15 tahun), muda & dewasa (umur 15- 50 tahun), tua (umur >50 tahun); (2) jenis kelamin yang terdiri dari perempuan (P) dan laki-laki (L); (3) bulan yang terdiri dari Januari, Februari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, Nopember, Desember; (4) diagnosa (ICD-10) merupakan atribut tujuan yang terdiri dari kelompok penyakit sesuai kode penyakit internasional (ICD-10), yaitu: A00-B99, C00-D48, D50-D89, E00-E90, F00-F99, G00-G99, H00-H59, H60-H95, I00-I99, J00-J99, K00-K93, L00-L99, M00-M99, N00-N99, O00-O99, P00-P96, Q00-Q99, R00-R99, S00-T98, V01-Y98, Z00-Z99 [9].

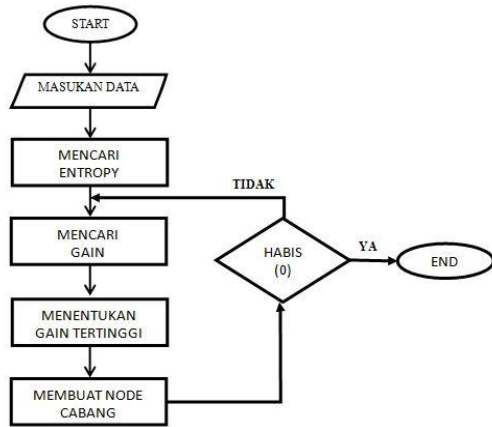
Algoritma C4.5 dimulai dari proses memilih atribut dengan gain tertinggi sebagai akar pohon, kemudian membuat cabang untuk tiap-tiap nilai, lalu membagi kasus dalam cabang, setelah itu mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memudahkan dalam penerapan metodologi dan perancangan sistem maka dibuat bagan alir analisa dan perancangan seperti pada gambar 2 dibawah ini.



Gambar 2. Bagan Alir Analisa

Bentuk diagram alir (*flowchart*) dapat menggambarkan dengan jelas mengenai proses tahapan maupun langkah dalam klasifikasi menggunakan algoritma C4.5. Dapat dilihat dalam bentuk gambar 4.2 berupa *flowchart* sebagai berikut:



Gambar 3. Flowchart Proses pada Algoritma C4.5

Teknik Klasifikasi Algoritma C4.5 dimulai dengan processing dan transformasi data agar data mentah yang digunakan untuk analisa adalah data dengan atribut yang lengkap dan dapat menghasilkan pohon keputusan, berikut urutan kerja untuk menentukan pohon keputusan

3.1 Processing dan Transformasi Data

Tidak semua atribut dalam database rekam medis pasien digunakan dalam penelitian, atribut seperti: tanggal lahir, nama dokter penanggung jawab dan nomor rekam medis pasien tidak dibutuhkan dalam penelitian ini.

Pengolahan hanya memerlukan empat buah atribut, seperti: Jenis Kelamin, Kategori Umur, Bulan Berobat dan Kode Penyakit sebagai atribut tujuan dari penelitian ini, sehingga diperlukan proses transformasi data yang akan diolah, setelah dilakukan transformasi data maka nantinya akan dilakukan pengolahan menggunakan algoritma C4.5, Sampel data yang telah dilakukan transformasi dapat dilihat pada tabel 1 dibawah ini

Tabel 1. Data Transformasi Rekam Medis Januari - Juni 2017

No	Jenis Kelamin	Kategori Umur	Bulan	Group ICD
1	Laki-Laki	Tua	Jan	J00-J99
2	Perempuan	Muda	Feb	E00-E90
3	Laki-Laki	Bayi & Anak	Mar	H00-H59
4	Laki-Laki	Bayi & Anak	Apr	A00-B99
5	Perempuan	Muda	Mei	I00-I99
6	Perempuan	Muda	Mei	I00-I99
7	Laki-Laki	Tua	Mei	I00-I99
8	Laki-Laki	Tua	Mei	I00-I99
9	Perempuan	Muda	Jun	I00-I99
10	Perempuan	Muda	Jun	I00-I99

3.2 Pengolahan Data

Pengolahan data dimulai dengan mencari entropi total dari seluruh atribut dan kemudian menentukan gain tertinggi. Untuk mendapatkan nilai gain dalam

pembentukan pohon keputusan, perlu menghitung dulu nilai informasi dalam satuan bits dari kumpulan objek. Bentuk perhitungan untuk entropi adalah sebagai berikut :

$$Entropy(X) = \sum_{j=1}^k p_j * \log_2 \frac{1}{p_j} = - \sum_{j=1}^k p_j * \log_2 p_j$$

dimana,

X : Himpunan Kasus

k : jumlah partisi X

p_j : Proporsi X_j terhadap X

Besar nilai $Entropy(X)$ menunjukkan bahwa X adalah atribut yang acak. Nilai entropi mencapai nilai minimum 0, ketika semua p_j lain = 0 atau berada pada kelas yang sama. Pada kontruksi pohon C4.5, di setiap simpul pohon diisi oleh atribut dengan nilai $gain ratio$ tertinggi, dengan rumus sebagai berikut:

$$Gain(a) = Entropy(X) - \sum_{j=1}^k \frac{|X_j|}{|X|} * Entropy(X_j)$$

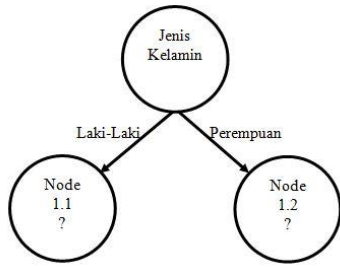
3.2.1 Mencari Entropy Total dan Gain (Root)

Proses pencarian entropi total dan gain dilakukan dengan mengelompokan data dengan benar, kemudian menghitung data serta menggunakan rumus pencarian $entropy$ dan $gain$ pada masing-masing atribut data.

Tabel 2. Tabel Perhitungan Gain Tertinggi (Root)

Perhitungan Node 1					
Jumlah Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
Total	1000	21	152	3,455	
Jenis Kelamin					1,346
L	325	11	101	0	
P	675	10	51	3,124	
Umur					0,555
BA	81	6	17	0	
MD	400	10	58	3,324	
TUA	519	5	77	3,025	
Bulan					0,406
JAN	141	1	42	3,096	
FEB	148	10	44	2,883	
MAR	105	3	1	2,883	
APR	353	1	63	3,128	
MEI	173	5	1	3,141	
JUN	80	1	1	2,931	

Dari hasil perhitungan pada tabel diatas, dapat diketahui bahwa nilai $gain$ terbesar yaitu pada atribut "Jenis Kelamin" sebesar 1,346. Sehingga atribut "Jenis Kelamin" menjadi node akar. Pada atribut "Jenis Kelamin" terdapat 2 nilai atribut, yaitu Laki-laki dan Perempuan, maka perlu dilakukan perhitungan lanjut. Dari proses tersebut maka dapat dihasilkan pohon sementara seperti berikut ini:



Gambar 4. Pohon Keputusan Sementara (Root)

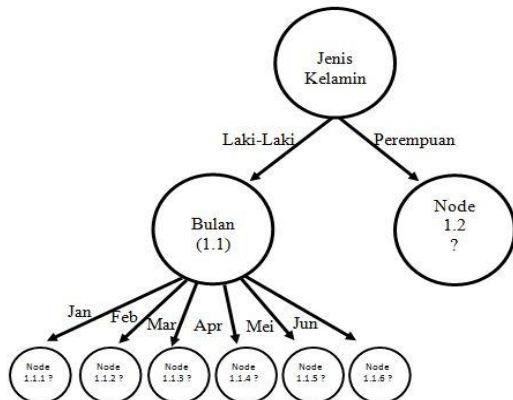
3.2.2 Mencari Node 1.1

Perhitungan dilakukan untuk mencari node cabang atribut "Jenis Kelamin Laki-Laki" kemudian lakukan perhitungan mencari nilai *gain* terbesar untuk menjadi *node* cabang.

Tabel 3. Tabel Perhitungan *Gain* Tertinggi (Node 1.1)

Node 1.1 (Jenis Kelamin_Laki-Laki)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
325	11	..	101	3,525	
Umur					3,525
BA	42	1	..	9	0
MD	72	7	..	17	0
TUA	211	3	..	75	0
Bulan					3,525
JAN	60	0	..	42	0
FEB	62	2	..	44	0
MAR	28	3	..	0	0
APR	102	1	..	15	0
MEI	68	5	..	0	0
JUN	5	0	..	0	0

Dari hasil perhitungan pada tabel diatas, dapat diketahui bahwa nilai *gain* terbesar yaitu sama besar pada atribut "Umur" dan atribut "Bulan" yaitu sebesar 3,525. Sehingga dapat dipilih diantara kedua atribut tersebut untuk dijadikan node cabang, dalam penelitian ini peneliti memilih atribut "Bulan" untuk dijadikan node cabang. Pada atribut "Bulan" terdapat 6 nilai atribut, yaitu Januari, Februari, Maret, April, Mei dan Juni, maka perlu dilakukan perhitungan lebih lanjut. Dari proses tersebut maka dapat dihasilkan pohon sementara seperti berikut ini:



Gambar 5. Pohon Keputusan Sementara (Node 1.1)

3.2.3 Mencari Node 1.1.1 - Node 1.1.6

Perhitungan dilakukan untuk mencari node cabang atribut "Jenis Kelamin Laki-Laki" pada bulan Januari sampai Juni, kemudian lakukan perhitungan mencari nilai *gain* terbesar untuk menjadi *node* cabang.

Tabel 4. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.1)

Node 1.1.1 (Jenis Kelamin_Laki-Laki) (Bulan_Januari)					
Banyak Kasus	C00-D48	..	Z00-Z99	Ent.	Gain
60	5	..	42	1,567	
Umur					1,567
BA	3	0	..	0	0,0000
MD	18	5	..	13	0,0000
TUA	39	0	..	29	0,0000

Tabel 5. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.2)

Node 1.1.2 (Jenis Kelamin_Laki-Laki) (Bulan_Februari)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
62	2	..	44	1,644	
Umur					1,644
BA	8	0	..	0	0,0000
MD	4	0	..	0	0,0000
TUA	50	2	..	44	0,0000

Tabel 6. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.3)

Node 1.1.3 (Jenis Kelamin_Laki-Laki) (Bulan_Maret)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
28	3	..	1	2,263	
Umur					2,263
BA	2	0	..	0	0,0000
MD	6	3	..	0	0,0000
TUA	20	0	..	1	0,0000

Tabel 7. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.4)

Node 1.1.4 (Jenis Kelamin_Laki-Laki) (Bulan_April)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
102	1	..	15	3,2306	
Umur					3,2306
BA	25	1	..	9	0,0000
MD	32	0	..	4	0,0000
TUA	45	0	..	2	0,0000

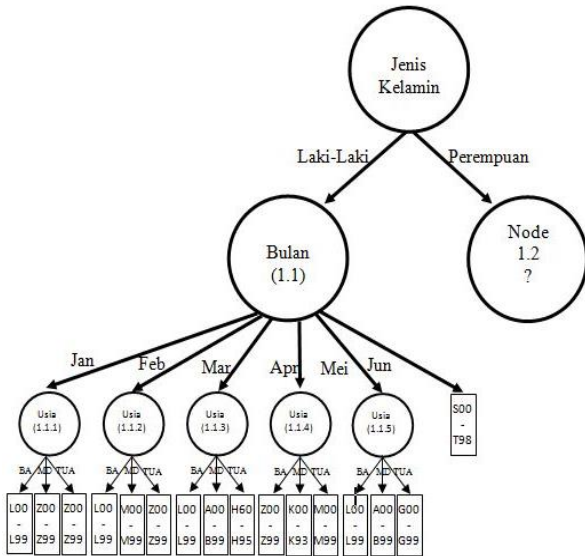
Tabel 8. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.5)

Node 1.1.5 (Jenis Kelamin_Laki-Laki) (Bulan_Mei)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
68	5	..	5	3,1469	
Umur					3,1469
BA	3	0	..	0	0,0000
MD	11	4	..	0	0,0000
TUA	54	1	..	5	0,0000

Tabel 9. Tabel Perhitungan *Gain* Tertinggi (Node 1.1.6)

Node 1.1.6 (Jenis Kelamin_Laki-Laki) (Bulan_Juni)					
Banyak Kasus	A00-B99	..	Z00-Z99	Ent.	Gain
5	2	..	2	1,5219	
Umur					0,3219
BA	1	0	..	1	0,0000
MD	0	0	..	0	0,0000
TUA	4	2	..	1	1,5000

Dari perhitungan tabel diatas, maka dapat dihasilkan pohon sementara seperti berikut ini:



Gambar 6. Pohon Keputusan Sementara (Node 1.1.2 - Node 1.1.6)

3.2.4 Mencari Node 1.2

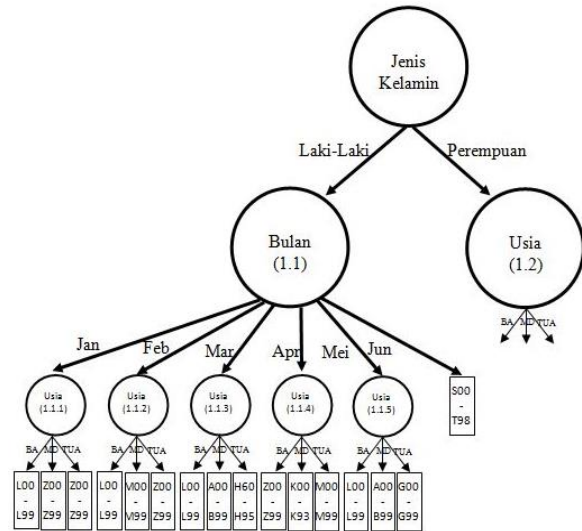
Perhitungan dilakukan untuk mencari *node* cabang atribut "Jenis Kelamin Perempuan", perhitungan dilakukan dengan mencari nilai dari atribut selain yang menjadi *node* akar (Jenis Kelamin), dan kemudian lakukan perhitungan untuk mencari nilai *Gain* dan atribut dengan nilai *Gain* terbesar, maka akan menjadi *node* cabang.

Tabel 10. Tabel Perhitungan *Gain* Tertinggi (Node 1.2)

Node 1.1 (Jenis Kelamin_Perempuan)					
Banyak Kasus	A00- B99	..	Z00- Z99	Ent.	Gain
325	10	..	51	3,395	
Umur					3,349
BA	39	5	..	8	0
MD	328	3	..	41	0
TUA	308	2	..	2	0
Bulan					3,021
JAN	81	1	..	0	0
FEB	86	8	..	0	0
MAR	77	1	..	1	0
APR	251	0	..	48	0
MEI	105	0	..	1	0
JUN	75	1	..	1	0

Dari hasil perhitungan pada tabel diatas, dapat diketahui bahwa nilai *gain* terbesar pada atribut "Umur" yaitu sebesar 3,349. Pada atribut "Umur" terdapat 3 nilai atribut, yaitu Bayi dan Anak, Muda dan Dewasa serta Tua, maka perlu dilakukan perhitungan lebih lanjut.

Dari proses tersebut maka dapat dihasilkan pohon sementara seperti berikut ini:



Gambar 7. Pohon Keputusan Sementara (Node 1.2)

3.2.5 Mencari Node 1.2.1 - Node 1.2.3

Perhitungan dilakukan untuk mencari *node* cabang atribut "Jenis Kelamin Perempuan" pada kategori umur Balita, Muda dan Tua, lakukan perhitungan nilai *gain* terbesar untuk *node* cabang

Tabel 11. Tabel Perhitungan *Gain* Tertinggi (Node 1.2.1)

Node 1.2.1 (Jenis Kelamin_Perempuan (Usia_Bayi&Anak))					
Banyak Kasus	A00- B99	Z00- Z99	Ent.	Gain
39	5	8	3,2818	
Bulan					3,2818
JAN	4	0	0	0,0000
FEB	7	5	0	0,0000
MAR	4	0	0	0,0000
APR	12	0	8	0,0000
MEI	11	0	0	0,0000
JUN	1	0	0	0,0000

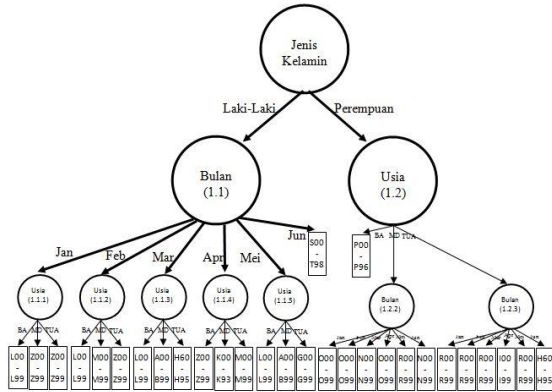
Tabel 12. Tabel Perhitungan *Gain* Tertinggi (Node 1.2.2)

Node 1.2.2 (Jenis Kelamin_Perempuan (Usia_Muda&Dewasa))					
Banyak Kasus	A00- B99	Z00- Z99	Ent.	Gain
39	3	41	3,094	
Bulan					3,094
JAN	42	0	0	0,0000
FEB	40	3	0	0,0000
MAR	30	0	0	0,0000
APR	159	0	40	0,0000
MEI	30	0	1	0,0000
JUN	27	0	0	0,0000

Tabel 13. Tabel Perhitungan *Gain* Tertinggi (Node 1.2.3)

Node 1.2.3 (Jenis Kelamin_Perempuan (Usia_Tua))					
Banyak Kasus	A00- B99	Z00- Z99	Ent.	Gain
308	2	2	2,777	
Bulan					2,777
JAN	35	1	0	0,0000
FEB	39	0	0	0,0000
MAR	43	0	1	0,0000
APR	80	0	0	0,0000
MEI	64	0	0	0,0000
JUN	47	1	1	0,0000

Dari perhitungan tabel diatas, maka dapat dihasilkan pohon keputusan yang telah digabungkan dengan pohon keputusan sebelumnya, seperti berikut ini:



Gambar 8. Pohon Keputusan Akhir

Untuk lebih jelasnya, hasil dalam pohon keputusan dapat menghasilkan rule seperti dapat dilihat pada tabel 14 seperti dibawah ini:

Tabel 14. Rule Knowledge

Rule 1	IF	Jenis Kelamin =Laki-Laki and Bulan = Januari and Usia = Balita dan Anak
	THEN	Kode Penyakit = L00 - L99
Rule 2	IF	Jenis Kelamin =Laki-Laki and Bulan = Januari and Usia = Muda dan Dewasa
	THEN	Kode Penyakit = Z00 - Z99
Rule 3	IF	Jenis Kelamin =Laki-Laki and Bulan = Januari and Usia = Tua
	THEN	Kode Penyakit = Z00 - Z99
Rule 4	IF	Jenis Kelamin =Laki-Laki and Bulan = Februari and Usia = Balita dan Anak
	THEN	Kode Penyakit = L00 - L99
Rule 5	IF	Jenis Kelamin =Laki-Laki and Bulan = Februari and Usia = Muda dan Dewasa
	THEN	Kode Penyakit = M00 - M99
Rule 6	IF	Jenis Kelamin =Laki-Laki and Bulan = Februari and Usia = Tua
	THEN	Kode Penyakit = Z00 - Z99
Rule 7	IF	Jenis Kelamin =Laki-Laki and Bulan = Maret and Usia = Balita dan Anak
	THEN	Kode Penyakit = L00 - L99
Rule 8	IF	Jenis Kelamin =Laki-Laki and Bulan = Maret and Usia = Muda dan Dewasa
	THEN	Kode Penyakit = A00 - B99
Rule 9	IF	Jenis Kelamin =Laki-Laki and Bulan = Maret and Usia = Tua
	THEN	Kode Penyakit = H60 - H95
Rule 10	IF	Jenis Kelamin =Laki-Laki and Bulan = April and Usia = Balita dan Anak
	THEN	Kode Penyakit = Z00 - Z99
Rule 11	IF	Jenis Kelamin =Laki-Laki and Bulan = April and Usia = Muda dan Dewasa

	THEN	Kode Penyakit = K00 - K93
Rule 12	IF	Jenis Kelamin =Laki-Laki and Bulan = April and Usia = Tua
	THEN	Kode Penyakit = M00 - M99
Rule 13	IF	Jenis Kelamin =Laki-Laki and Bulan = Mei and Usia = Balita dan Anak
	THEN	Kode Penyakit = L00 - L99
Rule 14	IF	Jenis Kelamin =Laki-Laki and Bulan = Mei and Usia = Muda dan Dewasa
	THEN	Kode Penyakit = A00 - B99
Rule 15	IF	Jenis Kelamin =Laki-Laki and Bulan = Mei and Usia = Tua
	THEN	Kode Penyakit = G00 - G99
Rule 16	IF	Jenis Kelamin =Laki-Laki and Bulan = Juni
	THEN	Kode Penyakit = S00 - T98
Rule 17	IF	Jenis Kelamin = Perempuan and Usia = Balita dan Anak
	THEN	Kode Penyakit = P00 - P96
Rule 18	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = Januari
	THEN	Kode Penyakit = O00 - O99
Rule 19	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = Februari
	THEN	Kode Penyakit = O00 - O99
Rule 20	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = Maret
	THEN	Kode Penyakit = N00 - N99
Rule 21	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = April
	THEN	Kode Penyakit = O00 - O99
Rule 22	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = Mei
	THEN	Kode Penyakit = R00 - R99
Rule 23	IF	Jenis Kelamin = Perempuan and Usia = Muda dan Dewasa and Bulan = Juni
	THEN	Kode Penyakit = N00 - N99
Rule 24	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = Januari
	THEN	Kode Penyakit = R00 - R99
Rule 25	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = Februari
	THEN	Kode Penyakit = R00 - R99
Rule 26	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = Maret
	THEN	Kode Penyakit = R00 - R99
Rule 27	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = April
	THEN	Kode Penyakit = I00 - I99
Rule 28	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = Mei
	THEN	Kode Penyakit = R00 - R99
Rule 29	IF	Jenis Kelamin = Perempuan and Usia = Tua and Bulan = Juni
	THEN	Kode Penyakit = H60 -H95

4. Hasil dan Pembahasan

Hasil analisa menunjukan bahwa algoritma C4.5 berhasil mengelompokan penyakit ke dalam **13 jenis kategori** dari **21 jenis kategori** yang menjadi label tujuan berdasarkan ICD (*International Code Diseases*) atau kode penyakit internasional, sehingga dapat dikatakan bahwa Algoritma C4.5 berhasil mendefinisikan **61,9%** dari kategori label tujuan yang ada.

Hasil pengujian terhadap data yang telah dilakukan, dirangkum pada tabel 15 seperti dibawah ini:

Tabel 15. Hasil Pengujian Data

No	Group ICD	Klasifikasi	Keterangan
1	A00-B99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
4	E00-E90	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "E00-E90" sangat sedikit, hanya terjadi 14 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 1,4%
5	F00-F99	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "F00-F99" sangat sedikit, hanya terjadi 18 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 1,8%
6	G00-G99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
7	H00-H59	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "H00-H59" sangat sedikit, hanya terjadi 12 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 1,2%
8	H60-H95	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
9	I00-I99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
10	J00-J99	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "J00-J99" sangat sedikit, hanya terjadi 17 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 1,7%
11	K00-K93	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
12	L00-L99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
13	M00-M99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
14	N00-N99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
15	O00-O99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
16	P00-P96	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)

17	Q00-Q99	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "F00-F99" sangat sedikit, hanya terjadi 8 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 0,8%
18	R00-R99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
19	S00-T98	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)
2	C00-D48	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "C00-D48" sangat sedikit, hanya terjadi 13 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 1,3%
3	D50-D89	TIDAK	Kasus yang terjadi untuk grup penyakit pada ICD "D50-D89" sangat sedikit, hanya terjadi 9 kali selama bulan Januari sampai dengan bulan Juni 2017, dengan presentasi data sebesar 0,9%
20	V01-Y98	TIDAK	Tidak ada kasus yang terjadi untuk grup penyakit pada ICD "V01-Y98", dengan presentasi data sebesar 0%
21	Z00-Z99	ADA	Terklasifikasikan sebagai <i>leaf node</i> (label)

5. Kesimpulan

Hasil pengujian algoritma C4.5 yang telah dilakukan, terdapat beberapa kesimpulan serta saran yang berhubungan dengan pembahasan sebelumnya.

5.1. Simpulan

1. Teknik klasifikasi menggunakan algoritma C4.5 membantu mengklasifikasikan data rekam medis pasien Rumah Sakit Tamar Medical Centre (TMC) Kota Pariaman sesuai dengan kode penyakit internasional.
2. Algoritma C4.5 dengan *output* pohon keputusan memberikan informasi rule diagnosa kode penyakit yang sering terjadi dengan mengklasifikasikan diagnosa penyakit berdasarkan jenis kelamin, umur dan tanggal pasien dirawat.
3. Algoritma C4.5 mengelompokan penyakit ke dalam 13 jenis kategori dari 21 jenis kategori yang menjadi label tujuan berdasarkan ICD (*International Code Diseases*) atau kode penyakit internasional, sehingga Algoritma C4.5 berhasil mendefinisikan label tujuan sebesar 61,90%

5.2.Saran

Beberapa saran yang dapat diperhatikan untuk masa yang akan datang maupun untuk pengembangan lebih lanjut terhadap pengklasifikasian data rekam medis pasien adalah sebagai berikut :

1. Penelitian ini diharapkan dapat dikembangkan lagi dengan menambah lebih banyak data, sehingga tingkat akurasi atribut label tujuan semakin mewakili data yang ada.
2. Penambahan variabel atau atribut lainnya sangat diperlukan atau dalam kata lain perlu penelitian lebih lanjut baik di lapangan maupun berdasarkan data yang ada guna menghasilkan rule yang tepat.
3. Belum adanya perbandingan software yang dilakukan penulis, misalnya terhadap Software Data Mining lainnya seperti *Weka*, *Orange*, *Tanagra*, dan lainnya. Perbandingan ini juga perlu dilakukan guna melihat hasil dari pengolahan yang dilakukan terhadap *software* yang digunakan.

6. Daftar Rujukan

- [1] Pupezescu, V., 2016. The Influence of Data Replication in the Knowledge Discovery in Distributed Databases Process. *ECAI 2016 International Conference Electronics (Computers and Artificial Intelligence)*, 8th Edition, pp.17
- [2] Priyadharsini, C., 2014. An Overview of Knowledge Discovery Database and Data mining Techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2 (Special Issue 1), pp.1571-1572
- [3] Veenita K., 2016. Chronic Kidney Disease Analysis Using Data Mining Classification Techniques. *IEEE*, pp.300-305
- [4] Harvinder C. and Anu C., 2013. Implementation of decision tree C4.5 algorithm. *International Journal of Scientific and Research Publications*, Vol. 3(10)
- [5] Chauhan H. and Chauhan A., 2013. Implementation of decision tree C4.5 algorithm. *International Journal of Scientific and Research Publications*, Vol. 3(Issue 10)
- [6] Wenefrida, T.I., 2013. Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5. *Jurnal Media Elektro*, Vol.1(3), pp.105
- [7] Yulia, E.K., Adhistya E.P., Silmi F., 2016. Comparative Study on Data Mining Classification Methods for Cervical Cancer Prediction Using Pap Smear Results. *1st International Conference on Biomedical Engineering (BIOMED)*, pp.1
- [8] Sucipto, Kusri, Emha L.T., 2016. Classification Method of Multi-class on C4.5 Algorithm for Fish Diseases. *2016 2nd International Conference on Science in Information Technology (ICSITech)*, pp.5
- [9] WHO. 2010. *International Statistical Classification of Diseases and Related Health Problems (ICD-10)*.