Seminar Nasional Royal (SENAR) 2018 ISSN 2622-9986 (cetak)
STMIK Royal – AMIK Royal, hlm. 153 – 158 ISSN 2622-6510 (online)
Kisaran, Asahan, Sumut - 3 September 2018

ANALISA DECISION TREE UNTUK PREDIKSI DIAGNOSA DIABETES MELLITUS

Dewi Anggraeni[1], Ramadhani[2]
[1,2]Sistem Informasi, STMIK ROYAL
email : anggraeni1987@gmail.com[1], rdhani2916@gmail.com[2]

***Abstrak** : tahapan memodelkan suatu informasi mengenai keputusan bahwa seseorang positif diabetes mellitus atau negatif diabetes mellitus . data yang dikumpulkan melalui rekam medis pasien. Data diolah dan dianalisa dari hasil laboratorium , selanjutnya data dianalisa untuk menentukan atribut atau penentu pohon keputusan. Berdasarkan dari klasifikasi diabetes mellitus, maka dapat dilihat karakteristik dari pasien positif diabetes mellitus dan negatif mellitus.*

Kata Kunci : *Algoritma C4.5, Decision Tree, Rapid Miner, Diabetes Mellitus*

PENDAHULUAN

Perkembangan data mining yang sangat pesat tidak dapat lepas dari perkembangan teknologi informasi, yang memungkinkan data dalam jumlah besar terakumulasi.

Data mining decision tree klasifikasi dengan menggunakan algoritma C4.5 telah banyak diterapkan oleh peneliti yang bertujuan untuk mendapatkan pola informasi yang tersimpan pada suatu basis data yang dapat digunakan untuk pengolahan informasi dan sebagai untuk prediksi diagnosa positif dan negatif diabetes pada pasien.

Dataset yang digunakan pada penelitian ini adalah berdasarkan dari data rekap medik pasien. Data yang terdiri dari kumpulan klinis hasil laboratorium dari pasien.

Berdasarkan data dan informasi yang dianalisa, rumusan masalah pada penelitian ini adalah, bagaimana memodelkan dan menguji dataset decision tree diabaetes mellitus, bagaimana model decision tree dapat mempredikasi pasien positif diabetes mellitus dan pasien negatif mellitus.

Pada penelitian ini penulis membatasi batasan terhadap masalah-masalah tersebut. Adapun batasan masalah input data untuk pelatihan (Training), dan pengujian (Testing) menggunakan Metode Algoritma C4.5, metode algoritma C4.5 teknik klasifikasi decision tree untuk melakukan pelatihan dan pengujian, menerapkan teori-teori dan algoritma C4.5 dalam menerapkan decision tree unutk diagnosa pasien diabetes mellitus.

METODOLOGI

Metodologi yang digunakan dalam penelitian ini adalah :

1. Studi Pustaka, penelitian yang dilakukan dengan menggali pustaka yang relevan dan berkenaan dengan penelitian.
2. Jurnal-jurnal yang membahas tentang decision tree dengan menggunakan algoritma C4.5
3. Sumber data yang digunakan, berdasarkan informasi dari rekam medik di Rumah Sakit M.Djamil Padang.

KDD *(Knowledge Discovery In Database)* menurut Fayyad dalam buku (Kusrini, 2009) Istila data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Proses KDD secara garis besar sebagai berikut :

1. Data Selection
   Pemilihan (Seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai.
2. Pre Procesing / Pembersihan Data
   Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD.
3. Transformation
   Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai proses data mining.

ISSN 2622-9986 (cetak)
ISSN 2622-6510 (online)

4. Data Mining
   Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu
5. Evaluasi
   Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

HASIL DAN PEMBAHASAN

Data penelitian ini berdasarkan data pemilihan dari rekam medik pasien rumah sakit M.Djamil Padang.

1. Data Selection
   Data yang menjadi variabel input (Penentu) dalam pembentukan Pohon Keputusan berdasarkan dari hasil laboratorium.
   a. Umur
   b. Glukosa
   c. Pregnant (Tingkat Kehamilan)
   d. DBP (Tekanan Darah)
   e. BMI (Index Masa Tubuh)
2. Pre Procesing

Dari hasil rekam medik dirumah sakit M.Djmail Padang, pasien yang terdeteksi positif diabetes mellitus dan negatif diabetes mellitus rata-rata berusia >>=22 tahun, dilihat dari riwayat hidup pasien, persentasi 75% keturunan dari keluarga yang sudah terdeteksi positif diabetes mellitus.

Tabel 1 Format Data Pasien

No	No.Rekam Medis	Umur	Jenis Kelamin	Asal Kota	SMF
1	01	63	PR	Padang	Penyakit Dalam
2	02	63	PR	Padang	Penyakit Dalam
3	03	64	PR	Padang	Penyakit Dalam
4	04	52	PR	Padang	Penyakit Dalam
5	05	52	PR	Padang	Penyakit Dalam
6	06	53	PR	Padang	Penyakit Dalam
7	07	22	LK	Padang	Penyakit Dalam
8	08	28	LK	Padang	Penyakit Dalam
9	09	28	LK	Padang	Penyakit Dalam
10	010	68	PR	Padang	Penyakit Dalam
11	011	66	LK	Padang	Penyakit Dalam
12	012	66	LK	Padang	Penyakit Dalam
13	013	66	LK	Padang	Penyakit Dalam
14	014	36	PR	Padang	Penyakit Dalam
15	015	36	PR	Pariaman	Penyakit Dalam
16	016	36	PR	Pariaman	Penyakit Dalam
17	017	68	LK	Pariaman	Penyakit Dalam
18	018	44	PR	Pariaman	Penyakit Dalam
19	019	44	PR	Pariaman	Penyakit Dalam
20	020	44	PR	Solok	Penyakit Dalam

ISSN 2622-9986 (cetak)
ISSN 2622-6510 (online)

Glukosa	Pregnant	DBP	BMI	Status
101	1	80	24	negatif
101	1	80	24	negatif
250	1	80	24	positif
250	1	80	24	positif
250	2	70	24	positif
165	2	70	24	negatif
250	1	70	22	positif
210	1	70	23	positif
110	1	70	25	positif
120	2	70	25	positif
300	2	70	25	positif
165	2	90	25	negatif
150	2	90	25	negatif
165	5	90	23	positif
140	2	90	23	negatif
110	2	90	23	negatif
260	3	70	23	positif
110	4	70	25	negatif
329	5	70	25	negatif
309	2	70	25	positif

3. Transformation
   Proses Transformasi yang dilakukan adalah mengklasifikasi hasil laboratorium menjadi variabel sebagai berikut :

Tabel 2 Atribut Variabel

No	Atribut	Ket	Nilai kontinyu	Level
1	Umur	Umur Pasien	<20	young
2	Umur	Umur Pasien	20<=Umur <=40	medium
3	Umur	Umur Pasien	>40	old
4	Glun	Glukosa	<95	low
5	Glun	Glukosa	95-140	medium
6	Glun	Glukosa	>90	high
7	Pregnant	Banyaknya Kehamilan	<1	low
8	Pregnant	Banyaknya kehamilan	<5	medium
9	Pregnant	Banyaknya kehamilan	>6	high
10	DBP	Tekanan Darah	<80	normal
11	DBP	Tekanan Darah	80-90	normal to high
12	DBP	Tekanan Darah	>90	high
13	BMI	Index Masa Tubuh	<24.9	low
14	BMI	Index Masa Tubuh	25-29.9	Normal
15	BMI	Index Masa Tubuh	30-34.9	obesitas
16	DPF	Riwayat Diabetes	<0.5287	low
17	DPF	Riwayat Diabetes	>0.5287	high

ISSN 2622-9986 (cetak)
ISSN 2622-6510 (online)

Tabel 3 Data Transformasi

No	Umur	Glukosa	Pregnant	DBP	BMI	Status
1	Medium	Low	Low	Normal	Normal	negatif
2	Medium	Low	Low	Normal	Normal	negatif
3	Old	High	Low	Normal	Rendah	positif
4	Old	High	Low	Normal	Rendah	positif
5	Old	High	Low	Normal	Rendah	positif
6	Old	High	Medium	Normal	Rendah	negatif
7	Old	High	Medium	Normal	Rendah	positif
8	Old	High	Medium	Normal	Rendah	positif
9	Old	High	Medium	High	Rendah	positif
10	Old	High	Medium	High	Rendah	positif
11	Old	High	Low	Normal	Rendah	positif
12	Medium	Low	Low	Normal	Normal	negatif
13	Medium	Low	Medium	High	Normal	negatif
14	Medium	Normal	Medium	High	Normal	positif
15	Medium	Normal	Medium	High	Normal	negatif
16	Medium	Normal	High	Normal	Rendah	negatif
17	Old	Normal	Medium	Normal	Rendah	positif
18	Old	Normal	Medium	Normal	Rendah	negatif
19	Medium	Normal	Medium	Normal	Rendah	positif
20	Medium	Normal	Medium	Normal	Rendah	positif

4. Analisa Decision Tree

Untuk memilih atribut sebagai akar, didasarkan pada gain tertinggi dari atribut yang ada. Untuk menghitung gain digunakan rumus sebagai berikut :

$$Gain\ (S.A) = Entropy(S) \sum_{i=1}^{n} \frac{|Si|}{|S|} + Entropy(Si)$$

Dimana :

1. S : Himpunan Kasus
2. A : Atribut
3. n : Jumlah Partisi Atribut A
4. $|Si|$: Jumlah Kasus Pada Partisi ke-i
5. $|S|$: Jumlah Kasus dalam S

Sementara perhitungan nilai entropy dapat dilihat pada persamaan berikut : [2]

$$Entropy\ (S) = \sum_{i=1}^{n} -Pi * Log2pi$$

Dimana :

1. S : Himpunan Kasus
2. A : Atribut
3. n : Jumlah Partisi S
5. Pi : Proporsi dari Si terhadap S

Hasil perhitungan menggunakan algoritma C4.5 untuk mencari node pertama.

Tabel 4 Hasil Perhitungan Pencarian Node 1

	Atribut	Nilai	Sum (nilai)	Sum (positif)
Node 1	Umur	Medium	9	3
		Old	11	9
	Pregnant	Low	7	4
		Medium	12	8
		High	1	0
	Glukosa	Low	4	0
		high	9	8
		Normal	7	4
	DBP	Normal	15	9
		high	4	3
	BMI	Rendah	14	11
		Normal	6	1

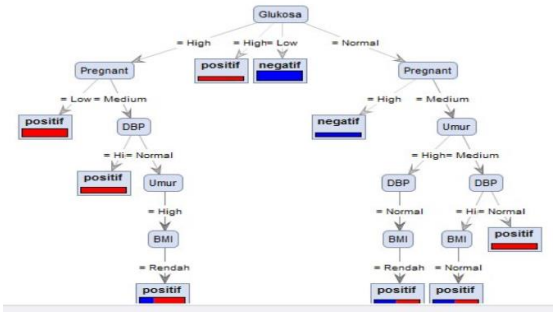
ISSN 2622-9986 (cetak)
ISSN 2622-6510 (online)

Sum (negatif)	Entropi	Gain
6	0.9316345	0.083888
2	0.8505942	
3	0.8749305	0.068831
4	0.9931569	
1	0	
4	0	0.613305
1	0.1142656	
3	0.8749305	
6	1.0394911	0.042785
2	0.7427376	
3	0.8849179	0.136679
5	0.7160964	0.136679

Sesuai hasil perhitungan bahwa atribut dengan gain tertinggi adalah atribut glukosa yaitu sebesar 0.613305. glukosa dapat menjadi node akar. Proses perhitungan decision tree dilanjutkan hingga semua atribut memiliki nilai keputusan.

Uji Coba

Uji coba sistem menggunakan tool Rapid Miner 5. Hasil uji coba dengan menggunakan sistem berupa pohon keputusan. Berikut hasil uji coba :



Gambar 1. Pohon Keputusan Menggunakan Rapid Miner

SIMPULAN

Dari hasil yang diperoleh, dapat diambil kesimpulan sebagai berikut :

1. Dari variabel input dalam pembentukan pohon keputusan berdasarkan hasil laboratorium yang dapat dilihat dari umur, glukosa, pregnant (tingkat kehamilan), DBP (tekanan darah), BMI (index masa tubuh). Dari variabel yang telah ditentukan, kita dapat memodelkan decision tree dengan menggunakan algoritma C4.5, dengan menentukan nilai entropi dan gain.
2. Dari hasil pengujian dengan penerapan decision tree menggunakan algoritma C4.5, kita dapat melihat prediksi kemungkinan dan menggambarkan bahwa pasien positif dan pasien negatif diabetes mellitus.

DAFTAR PUSTAKA

Kuo et all,(2012). “Data Mining with decision tree for diagnosis of breast tumor in medical ultrasonic image.

Olaiya, Folorunsho. (2012). “Appication of Data Mining Techniques in Weather Prediction and Climate Change Studies”. *Department of Computer & Information System, Achievers University, Owa, Negaria*. **1**. 51-59.

Khan Ahmed Naeem Muhammad, at all. (2013). “Gender Classification with Decision Trees”. *International Jurnal of Signal Processing, Image Processing and Pattern Recognition vol 6.*

ISSN 2622-9986 (cetak)
ISSN 2622-6510 (online)

Curtis G. Panayiotis. (2009). "A Decision Tree Application in Tourism Based Regional Economic Development". *International Multidiciplinary Journal Of Tourism, vol 4 Number 2 2009,pp. 169-178.*

Kusrini, Emha Taufiq Luthfi (2009). " Algoritma Data Mining", Andi, STMIK AMIKOM Yogyakarta.

Rekam Medik (2011-2013). "Rumah Sakit Dr.M.Djamil Padang", Padang.

Abidin Zaenal Zezen Aa, (2011) "Implementasi Algortima C4.5 Untuk Menentukan Tingkat Bahaya Tsunami". *Seminar National Informatika, Jurusan Teknik Informatika STMIK Subang, Jawa Barat* 2011 1979-2328.

Haim-Ben Yael (2010). "A Streaming Parallel Decision Tree Algorithm". *Journal of Machine Learning Research 11*. 849-872.

Khoonsari Emami Payam, AhmadReza Motie (2012). "A Comparison of Efficiency and Robustness of ID3 and C4.5 Algorithm Using Dynamic Test and Tranining Data Sets". *International Journal of Machine Learning and Computing vol 2.*

Chauhan Harvinder, Chauhan Anu (2013). "Implementasi of Decision Tree algorithm C4.5". *International Journal of Scientific and Research Publication, vol 3.*

Mazid M Mohammed, Ali Shawkat A B M, Kevin S Tickle (2009). "Improved C4.5 Algorithm for Rule Based Classification". *School of Computing Science Central Queensland university, Australia.*