

PERBANDINGAN KINERJA METODE NAIVE BAYES DAN C4.5 DALAM PENGKLASIFIKASIAN PENYAKIT DIABETES MELITUS DI RUMAH SAKIT KUMALA SIWI KUDUS

Eska Sarti Kundari¹

^{1,2} Teknik Informatika, Ilmu Komputer, Universitas Dian Nuswantoro
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA
111201106265@mhs.dinus.ac.id¹

Abstrak

Seiring berjalannya waktu data – data yang dimiliki rumah sakit akan bertambah banyak tetapi informasi yang bisa dipetik dari data – data tersebut tidak seimbang dengan data – data yang didapat. Hal tersebut dipengaruhi oleh banyaknya jumlah pasien yang telah melakukan pemeriksaan kesehatan termasuk penderita penyakit diabetes melitus. Dengan memanfaatkan data mining, pengklasifikasian penyakit bisa dilakukan untuk mengetahui apakah pasien positif diabetes melitus atau negatif diabetes melitus. Data mining ini dapat menggali informasi dari gudang data dengan menggunakan metode - metode tertentu untuk mendapat informasi atau pengetahuan baru. Oleh karena itu data mining bisa digunakan untuk pengklasifikasian, algoritma naive bayes dan c4.5 adalah contoh algoritma yang bisa di gunakan untuk pengklasifikasian. Tujuan dari penelitian ini adalah untuk mengetahui akurasi yang lebih baik antara kedua algoritma tersebut. Sehingga dalam studi kasus yakni rumah sakit kumala siwi kudus bisa menggunakan algoritma yang memiliki akurasi lebih baik. Peneliti akan membuat prototipe menggunakan bahasa pemrograman visual basic 6.0 dan menggunakan microsoft access sebagai database. Dengan prototipe ini diharapkan memberi informasi tentang akurasi algoritma naive bayes dan c4.5 dalam dalam pengklasifikasian penyakit diabetes melitus di rumah sakit kumala siwi kudus. **Kata Kunci:** *Klasifikasi, naive bayes, decision tree c4.5, data penyakit diabetes melitus, visual basic 6.0, access 2013.*

Abstract

Over time many data owned by the hospital will multiply but the information that can be gleaned from the data is not balance by the data obtained. It is influenced by the large number of patients who have a medical examination, including patients with diabetes melitus. By utilizing data mining, classification of the disease could be conducted to determine whether the patient is positive or negative diabetes melitus diabetes melitus. Data mining can dig up information from the data warehouse using the method - specific method to obtain new information or knowledge. Therefore, data mining can be used for classification, Naive Bayes and C4.5 algorithms are examples of algorithms that can be used for classification. The purpose of this study was to determine the accuracy is better between the two algorithms are. So in the case study that is kumala siwi kudus hospital can use the algorithm has better accuracy. Researchers will create a prototype using the programming language Visual Basic 6.0 and uses Microsoft Access as the database. With this prototype is expected to provide information about the accuracy of Naive Bayes and C4.5 algorithms in the classification of diabetes melitus in kumala siwi kudus hospital. **Keywords :** *classification, naive bayes, decision tree c4.5, data disease diabetes melitus, visual basic 6.0, access 2013*

I. PENDAHULUAN

Diabetes Militus merupakan penyakit yang terjadi akibat kadar glukosa di dalam darah tinggi karena tubuh tidak dapat melepaskan atau menggunakan insulin secara normal. Kadar glukosa darah sepanjang hari bervariasi, meningkat setelah

makan dan kembali normal dalam waktu dua jam. Glukosa darah normal pada pagi hari setelah malam sebelumnya berpuasa adalah 70-110 mg/dL. Glukosa darah biasanya kurang dari 120-140 mg/dL pada dua jam setelah makan atau minum cairan yang mengandung gula maupun karbohidrat lainnya. Glukosa darah normal cenderung

meningkat secara ringan tetapi progresif setelah usia 50 tahun, terutama pada orang – orang yang tidak aktif beraktifitas. Insulin adalah hormon yang dilepaskan oleh pankreas, merupakan zat utama yang bertanggungjawab dalam mempertahankan kadar glukosa darah yang tepat. Insulin menyebabkan glukosa berpindah ke dalam sel sehingga bisa menghasilkan energi. Diabetes Militus terjadi jika tubuh tidak menghasilkan insulin yang cukup untuk mempertahankan glukosa darah normal atau jika sel tidak memberikan respon yang tepat terhadap insulin [1] .

Rumah Sakit Kumala Siwi adalah salah satu rumah sakit swasta yang berada di Jl. Jeparo Km 6 Desa Mijen RT 10 RW 06, Kec. Kaliwungu Kab, Kudus 59361 yang memiliki data – data mengenai pasien yang melakukan pemeriksaan kesehatan. Hasil dari pemeriksaan tersebut dapat mendiagnosa potensi seorang pasien dalam menderita Diabetes Militus. Seiring berjalannya waktu data – data yang dimiliki rumah sakit akan bertambah banyak tetapi informasi yang bisa dipetik dari data – data tersebut tidak seimbang dengan data – data yang didapat. Oleh karena itu diperlukan suatu teknik dan perangkat yang dapat membantu dalam mendukung data tersebut menjadi suatu informasi yang berguna yaitu dengan penerapan data mining dengan tujuan untuk mentukan apakah pasien yang telah melakukan pemeriksaan kesehatan termasuk penderita penyakit Diabetes Militus atau tidak..

Seseorang dapat dikatakan memiliki resiko penyakit diabetes melitus apabila diketahui pada pemeriksaan kadar gula darah dalam puasannya melebihi angka 126 mg/dl atau dua kali berturut – turut pemeriksaan kadar gula darah setelah 2 jam melebihi 180 mg / dl , tekanan darah tinggi yaitu lebih dari 140 / 85 mm Hg, berat badan obese yaitu untuk laki laki 75 kg dan perempuan 55 kg, usia lebih dari 45 tahun. [11].

II STUDI PUSTAKA

2.1. Penelitian Terkait

Penulis memulai penelitian ini dengan

terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang akan digunakan penulis.

1. F. Gian, “Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes Dan Decision Tree C4.5 Untuk Prediksi Ketetapan Waktu Kelulusan Mahasiswa,” 2014.
2. Y. G. d. Syawli Almira, “Diagnosa Penyakit Diabetes Melitus Dengan Metode Naive Bayes Berbasis Destop Aplication,” 2009.
3. T. A. B. FS Rodiyatul, “Implementasi Teknik Data Mining Didalam Analisis Penyakit Diabetes Melitus Tipe II Menggunakan Decision Tree,” 2009

2. 2 Tinjauan Pustaka

A. 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, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [9].

Untuk pengklasifikasian penyakit ini dapat digunakan metode klasifikasi yaitu suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Ada beberapa algoritma dalam metode klasifikasi ini antara lain algoritma naive bayes dan C4.5

B. Teorema Bayes

Teorema Bayes adalah sebuah pendekatan untuk sebuah ketidak tentuan yang diukur dengan probabilitas. Pendekatan bayes pada saat klasifikasi adalah mencari probabilitas tertinggi (VMAP) dengan masukan atribut (a1, a2, a3, ..., an) seperti tampak pada persamaan 1 berikut [2] : VMAP = arg max

$$P(v_j|a_1,a_2,a_3,\dots,a_n) \quad (1)$$

$v_j \in V$.

Teorema Bayes sendiri berawal dari rumus persamaan 2 berikut :

$$P(A|B) = \frac{P(B \cap A)}{p(B)} \quad (2)$$

Dimana $P(A|B)$ artinya peluang A jika diketahui keadaan B. Kemudian dari persamaan rumus 2 diatas didapat persamaan 3 seperti berikut :

$$P(B \cap A) = P(B | A) P(A) \quad (3)$$

Sehingga didapatkan teorema bayes seperti persamaan 4 berikut :

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (4)$$

Yang mana :

- A adalah hipotesis data A (class spesifik)
- B adalah data dengan class yang belum diketahui
- $P(A|B)$ adalah probabilitas hipotesis A berdasar kondisi B (posterior | probability)
- $P(B|A)$ adalah probabilitas B berdasar kondisi pada hipotesis A
- $P(A)$ adalah probabilitas hipotesis A (prior probability)
- $P(B)$ adalah probabilitas dari B

Menggunakan teorema bayes ini, persamaan diatas dapat ditulis menjadi persamaan 5 berikut ini :

$$V_{\text{map}} = \arg \max_{v_j} \frac{P(a_1, a_2, a_3, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (5)$$

Karena nilai $P(a_1, a_2, a_3, \dots, a_n)$ konstan untuk semua v_j , maka persamaan ini dapat ditulis menjadi persamaan 6 berikut ini :

$$V_{\text{MAP}} = \arg \max_{v_j \in V} P(a_1, a_2, a_3, \dots, a_n | v_j) P(v_j)$$

Untuk menghitung $P(a_1, a_2, a_3, \dots, a_n | v_j)$ bisa jadi semakin sulit karena jumlah term $P(a_1, a_2, a_3, \dots, a_n | v_j)$ bisa jadi sangat besar [2]. Hal ini disebabkan jumlah term tersebut [3].

Secara umum algoritma naive bayes apabila diberikan inputan baru, maka klasifikasi dapat ditentukan sebagai berikut :

1. Menghitung jumlah class / label
2. Menghitung jumlah kasus yang sama dengan class yang sama
3. Kalikan semua hasil variabel,

4. Dari hasil variabel yang dikalikan tersebut cari probabilitas yang tertinggi [8].

C. Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (*kontinyu*) dan *diskret*, dapat menangani nilai atribut yang hilang menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain [4].

D. Pohon Keputusan C4.5

Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan/ internal (*branch/ internal node*) dan simpul daun (*leaf node*) [8].

Dalam algoritma C4.5 untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiap-tiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari

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

atribut-tribut yang ada. Untuk menghitung gain digunakan rumus sebagai berikut [4].

Keterangan:

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

Sehingga akan diperoleh nilai gain dari atribut yang paling tertinggi. Gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute*

tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu *node*. Sementara itu, penghitungan nilai entropy dapat dilihat pada persamaan :

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

Keterangan :

S : Himpunan kasus

A : Atribut

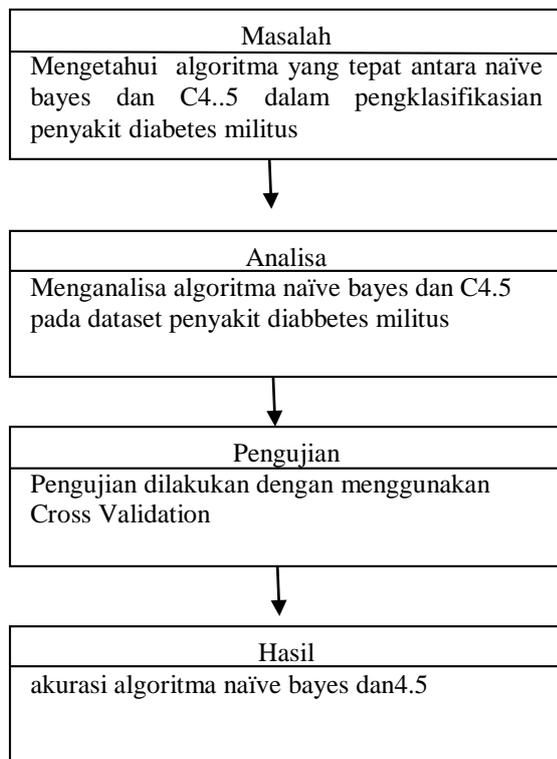
N : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap – tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. [8].

2.3. Kerangka Pikir



III METODOLOGI PENELITIAN

3.1 objek penelitian

Penulis melakukan penelitian di Rumah Sakit Kumala Siwi yang beralamatkan di Jl. Jeparu Km 6 Desa Mijen RT 10 RW 06, Kec. Kaliwungu Kab, Kudus 59361. Penelitian ini dilakukan untuk mengambil data pasien yang digunakan untuk membandingkan kinerja metode naive bayes dan C4.5 dalam pengklasifikasian penyakit diabetes militus.

3.2 Instrumen Penelitian

Berdasarkan permasalahan yang telah diuraikan pada bab sebelumnya, maka bahan dan peralatan yang diperlukan untuk penelitian ini meliputi :

A. Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah data pasien di Rumah Sakit Kumala Siwi Kudus.

B. Peralatan

Peralatan dalam penelitian ini meliputi kebutuhan software dan kebutuhan hardware. Dibawah ini merupakan kebutuhan dari sistem , diantaranya:

Kebutuhan Software :

1. Windows Edition : Windows 7 Professiona (Copyright © 2009 Microsoft Corporation)
2. System Type : 64-bit Operating System

Kebutuhan Hardware :

1. Processor Intel(R) Core(TM) i3-2310M CPU @2.10GHz 2.1 GHz
2. Installed memory (RAM) 2.00 GB

3.3 Pengumpulan Data

Data yang digunakan pada penelitian ini berasal dari data pasien Rumah Sakit Kumala Siwi Kudus, dengan 7 variabel.

3.4 Teknik Analisis Data

Data pasien Rumah Sakit Kumala Siwi Kudus yang di peroleh dari pengumpulan data sebanyak 323 record, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena

harus melalui beberapa tahap pengolahan awal data (preparation data). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut:

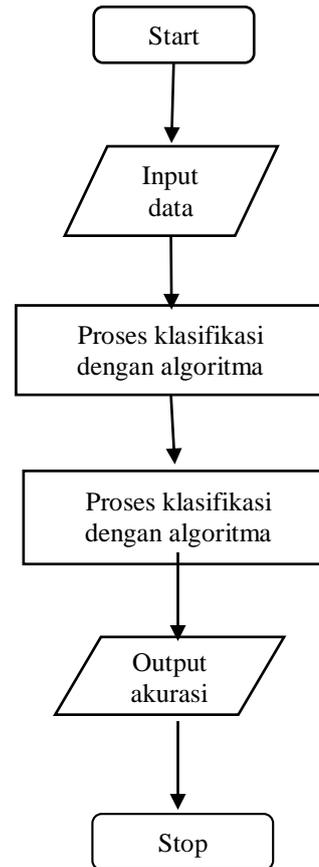
1. Data cleaning.
2. Data integration and transformation
3. Data reduction

Tabel 3.1 atribut

No	Nama Atribut	Penjelasan
1	Kadar glukosa Puasa	Low (<95), medium (95 – 140), high (>140),
2	Kadar glukosa setelah 2 jam	Low (<110), medium (110 – 180), high (>180),
3	Teanan darah	Low (<100), normal (100 – 140), high (> 140)
4	Berat badan	Low (<45), normal (45 – 50), obese (50 – 60), severely-obese(>60)
5	Umur	Young (<45), medium (45-59), old (>60)
6	Jenis kelamin	Perempuan (P), Laki – laki (L)
7	Status (Positif / Negatif)	Positif , atau negatif .

3.4 Eksperimen

Pada tahap akhir pada penelitian ini akan dilakukan pengujian pada eksperimen yang telah dilakukan. Pada tahap eksperimen ini penulis menerapkan model yang telah diusulkan mejadi sebuah prototype yang di bangun menggunakan vb 6.0 dan microsoft access 2013



IV HASIL PENELITIAN DAN PEMBAHASAN

Didalam bab ini akan dibahas mengenai data yang akan digunakan dalam penelitian, data tersebut akan di hitung menggunakan algoritma naïve bayes dan C4.5 kemudian diuji menggunakan cross validation.

4.1 Data yang digunakan

Pada penelitian ini, data yang digunakan adalah data penyakit diabetes di Rumah Sakit Kumala Siwi Kudus dengan jumlah data 152 record.

Tabel 4.1 data asli

No CM	Kadar glukosa puasa (mg /dl)	Kadar glukosa setelah dua jam (mg / dl)	Tekanan darah (mm hg)	Berat badan (kg)	Umur (th)	Jenis kelamin	status
0007687	271	150	150 / 70	75 kg	49 th	P	positif
0010576	251	167	130 / 100	53 kg	56 th	P	positif
0007556	409	400	160 / 90	66 kg	59 th	L	positif
0006876	344	165	150 / 70	47 kg	42 th	P	positif

Data asli dari rumah sakit tersebut kemudian di olah dan dikonversikan menjadi tabel sebagai berikut :

Tabel 4.2 data pasien yang sudah dikelompokkan

Kadar glukosa puasa	Kadar glukosa setelah 2jam	Tekanan darah	Berat badan	umur	Jenis kelamin	status
high	medium	high	severely_obese	medium	P	positif
high	medium	normal	obese	medium	P	positif
high	high	high	severely_obese	medium	L	positif
high	medium	high	obese	young	P	positif

4.2 Perhitungan Algoritma Naïve Bayes

Dibawah ini adalah contoh perhitungan manual penerapan algoritma naïve bayes untuk pengklasifikasian penyakit diabetes melitus menggunakan data training dan data testing sebagai berikut:

Tabel 4.3 kasus naive bayes

Kadar glukosa puasa	Kadar glukosa setelah 2 jam	Tekanan darah	Berat badan	umur	Jenis kelamin	status
high	medium	high	Severely obese	medium	P	?

Dalam perhitungan naive bayes ada beberapa langkah yang harus dilakukan, berikut adalah langkah – langkahnya

- Menghitung jumlah class / label
 $P (Y = \text{Positif}) = 86 / 152$ “ jumlah data positif pada data trining dibagi dengan keseluruhan data “
 $P (Y = \text{Negatif}) = 66 / 152$ “ jumlah data negatif pada data trining dibagi dengan keseluruhan data ”
- Menghitung jumlah kasus yang sama dengan class yang sma
 $P (\text{kadar glukosa puasa} = \text{high} \mid Y = \text{Positif}) = 41 / 86$
 $P (\text{kadar glukosa puasa} = \text{high} \mid Y = \text{negatif}) = 0 / 66$

 $P (\text{kadar glukosa setelah 2 jam} = \text{medium} \mid Y = \text{Positif}) = 22 / 86$
 $P (\text{kadar glukosa setelah 2 jam} = \text{medium} \mid Y = \text{Negatif}) = 37 / 66$

 $P (\text{tekanan darah} = \text{high} \mid Y = \text{Positif}) = 51 / 86$
 $P (\text{tekanan darah} = \text{high} \mid Y = \text{Negatif}) = 11 / 66$

 $P (\text{berat badan} = \text{severely obese} \mid Y = \text{Positif}) = 27 / 86$
 $P (\text{berat badan} = \text{severely obese} \mid Y = \text{Negatif}) = 0 / 66$

$$P (\text{umur} = \text{medium} \mid Y = \text{Positif}) = 49 / 86$$

$$P (\text{umur} = \text{medium} \mid Y = \text{Negatif}) = 28 / 66$$

$$P (\text{jenis kelamin} = P \mid Y = \text{Positif}) = 61 / 86$$

$$P (\text{jenis kelamin} = P \mid Y = \text{Negatif}) = 25 / 66$$

- kalikan semua hasil variabel positif dan negatif

$$P (\text{kadar glukosa puasa} \setminus \text{positif}) * P (\text{kadar glukosa setelah 2 jam} \setminus \text{positif}) * P (\text{tekanan darah} \setminus \text{positif}) * P (\text{berat badan} \setminus \text{positif}) * P (\text{umur} \setminus \text{positif}) * P (\text{jenis kelamin} \setminus \text{positif})$$

$$= \frac{41}{86} \times \frac{22}{86} \times \frac{51}{86} \times \frac{27}{86} \times \frac{49}{86} \times \frac{61}{86}$$

$$= 0.4767 \times 0.2558 \times 0.5930 \times 0.3139 \times 0.5697 \times 0.7093$$

$$= 0.009$$

$$P (\text{kadar glukosa puasa} \setminus \text{negatif}) * P (\text{kadar glukosa setelah 2 jam} \setminus \text{negatif}) * P (\text{tekanan darah} \setminus \text{negatif}) * P (\text{berat badan} \setminus \text{negatif}) * P (\text{umur} \setminus \text{negatif}) * P (\text{jenis kelamin} \setminus \text{negatif})$$

$$= \frac{0}{66} \times \frac{37}{66} \times \frac{11}{66} \times \frac{0}{66} \times \frac{28}{66} \times \frac{25}{66}$$

$$= 0 \times 0.5606 \times 0.1666 \times 0 \times 0.4242 \times 0.3787$$

$$= 0$$

- Dari nilai diatas, terlihat bahwa nilai probabilitas tertinggi pada class (Y | Positif) sehingga dapat disimpulkan bahwa status pasien termasuk dalam klasifikasi “ Positif Diabetes Militus “.

4.2 Perhitungan Algoritma C4.5

Perhitungan menggunakan algoritma C4.5 berbeda dengan menggunakan algoritma naïve bayes. Pada algoritma C4.5 harus menggunakan pohon keputusan yang kemudian akan menjadi rule untuk mengklasifikasikan penyakit Diabetes Militus.

- Menghitung jumlah kasus, jumlah kasus untuk keputusan positif, jumlah kasus untuk keputusan negatif, dan

entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut Kgp (kadar glukosa puasa), Kg2jm (kadar glukosa setelah 2 jam), TD (tekanan darah), BB (berat badan), umur, jenis kelamin. Setelah itu, lakukan perhitungan gain untuk setiap atribut. Hasil perhitungan di tunjukkan oleh tabel berikut :

Tabel 4.4 perhitungan C4.5 node 1

Node 1	Jumlah kasus (S)	Negatif (S1)	Positif (S2)	Entropy	Gain
Total	152	66	86	0.9874750083	
Kgp					0.3096921135
-high	41	0	41	0	
-medium	65	32	33	0.9998292601	
-low	46	34	12	0.8280557253	
Kg2jm					0.5413697451
-high	64	1	63	0.0848650753	
-medium	59	37	22	0.9528588097	
-low	29	28	1	0.2163969324	
TD					0.272646061
-high	62	11	51	0.67439888	
-normal	67	32	35	0.9985532861	
-low	23	23	0	0	
BB					0.4830276399
-severely obese	27	0	27	0	
-obese	65	14	51	0.7516499461	
-normal	37	29	8	0.7531979911	
-low	23	23	0	0	
Umur					0.0266460609
-old	33	20	13	0.9672947789	
-medium	77	28	49	0.9456603045	
-young	42	18	24	0.985228136	
JK					0.080264482
-P	86	25	61	0.8696207741	
-L	66	41	25	0.9571814385	

Berdasarkan tabel diatas, berikut adalah contoh perhitungan mencari entropy dan gain pada field Kgp (kadar glukosa puasa) menggunakan algoritma C4.5

$$\text{Entropy (total)} = (- (66 / 152 * \log_2 (66 / 152)) + (- (86 / 152) * \log_2 (86 / 152))) = 0.5225868685 + 0.4648881398 = 0.9874750083$$

$$\text{Entropy (medium)} = (- (32 / 65) * \log_2 (32 / 65)) + (- (33 / 65) * \log_2 (33 / 65)) = 0.5033195387 + 0.4965097214 = 0.9998292601$$

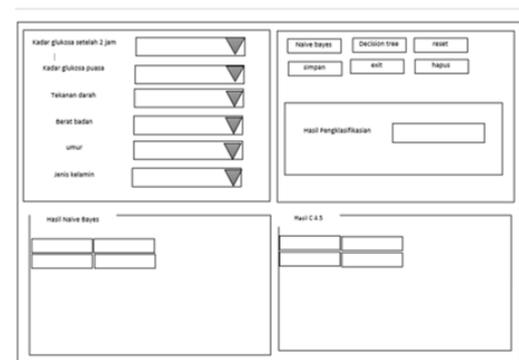
$$\text{Entropy (low)} = (- (34 / 46) * \log_2 (34 / 46)) + (- (12 / 46) * \log_2 (12 / 46)) = 0.3223341283 + 0.505721597 = 0.8280557253$$

$$\text{Gain (Kgp)} = 0.9874750083 - ((41 / 152 * 0) + (65 / 152 * 0.999) + (46 / 152 * 0.828)) = 0.9874750083 - (0 + 0.4272039474 + 0.2505789474) = 0.9874750083 - 0.6777828948 = 0.309697451$$

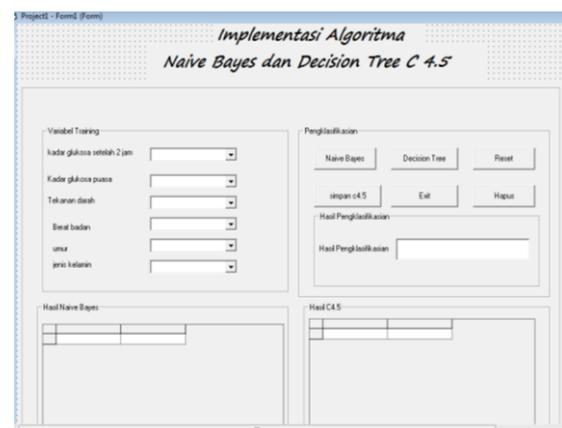
4.3 Pengujian Cross Validation

Dari dua kali percobaan memperbandingkan algoritma naive bayes dan C4.5 untuk **pengklasifikasian** penyakit Diabetes Militus di Rumah Sakit Kumala Siwi Kudus dengan 37 data testing yang berbeda dihasilkan persentase keakurasian yang berbeda pula. Dari dua akurasi diatas masing – masing algoritma memiliki rata – rata yaitu 74.31 % untuk algoritma naive bayes dan 85,13 % untuk algoritma C4.5

4.4 Prototype



Gambar 4.1 tampilan prototype naive bayes dan C4.5



Gambar 4.2 implementasi prototype naive bayes dan C4.5

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa dari hasil percobaan memperbandingkan algoritma naive bayes dan C4.5 untuk mengklasifikasikan penyakit Diabetes Militus di Rumah Sakit Kumala Siwi Kudus menghasilkan akurasi 74,31 % untuk algoritma naive bayes dan 85,13 % untuk algoritma C4.5 dengan menggunakan 37 data testing dan 152 data training. Maka dapat disimpulkan bahwa algoritma yang paling tepat untuk pengklasifikasian penyakit Diabetes Militus di Rumah Sakit Kumala Siwi Kudus adalah algoritma C4.5.

5.2 Saran

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut :

1. Diharapkan dapat membandingkan dengan algoritma klasifikasi yang lain agar mengetahui algoritma yang lebih baik dalam pengklasifikasian penyakit Diabetes Militus di Rumah Sakit Kumala Siwi Kudus.
2. Diharapkan dapat menambah data training yang lebih banyak lagi agar tingkat keakurasiannya lebih tinggi.

DAFTAR PUSTAKA

- [1] L. I. P. Dody, "Perbandingan Kinerja Decision Tree J48 dan ID3 Dalam Pengklasifikasian Diagnosa Penyakit Diabetes Melitus," 2012.
- [2] P. Eko, Data Mining Konsep dan Aplikasi menggunakan Matlab, Yogyakarta: Andi, 2012.
- [3] Y. G. d. Syawli Almira, "Diagnosa Penyakit Diabetes Melitus Dengan Metode Naive Bayes Berbasis Destop Aplication," 2009.
- [4] d. Luthfi, Algoritma Data Mining, Yogyakarta: Andi, 2009.
- [5] I. d. Ahmad, "Perbandingan Algoritma Klasifikasi dalam Pendeteksian Penyakit Kanker," 2013.
- [6] F. Gian, "Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes Dan Decision Tree C4.5 Untuk Prediksi Ketetapan Waktu Kelulusan Mahasiswa," 2014.
- [7] T. A. B. FS Rodiyatul, "Implementasi Teknik Data Mining Didalam Analisis Penyakit Diabetes Melitus Tipe II Menggunakan Decision Tree," 2009.
- [8] A. M.Khoirul, "Penerapan Data Mining Untuk Menentukan Kriterion Calon Nasabah Potensial Pada AJB BUMIPUTRA 1912 PALEMBANG," 2013.
- [9] d. E Turban, Decision Support System and Intelligent Systems, Andi Offset, 2005.
- [10] Pramudiono, "pengertian-data-mining-apa-itu-data-mining.html," 2006. [Online]. Available: <http://apapengertianahli.blogspot.com/2014/09/>.
- [11] d. Trisnawati kurnia shara, "Faktor Resiko Diabetes Militus Tipe II di Puskesmas Kecamatan Cengkaren Jakarta Barat Tahun 2012," 2013.
- [12] K. Indah, "Penggunaan Pohon Keputusan Untuk Data Mining," 2008.