SISTEM PREDIKSI PENYAKIT DIABETES BERBASIS *DECISION TREE*

Anik Andriani
Manajemen Informatika
AMIK BSI Jakarta
Jl. R.S. Fatmawati No.24, Pondok Labu, Jakarta Selatan
Email: anik.aai@bsi.ac.id

ABSTRAK

Data penderita diabetes bertambah dari tahun ketahun. Tingkat diagnosa diabetes memberikan kontribusi yang signifikan terhadap komorbiditas dan tingkat komplikasi diabetes. Berdasarkan data histori penderita diabetes dapat dibuat rekomendasi prediksi penyakit diabetes yang membantu tenaga kesehatan. Klasifikasi merupakan salah satu teknik dari data mining yang dapat digunakan untuk membuat prediksi. Klasifikasi dapat dilakukan dengan decision tree salah satunya dengan algoritma C4.5. Penelitian ini bertujuan membuat klasifikasi data diabetes dan menerapkannya dalam pembangunan sistem prediksi penyakit diabetes. Hasil klasifikasi data diabetes dievaluasi dengan confusion matrix dan kurva ROC(Receiver Operating Characteristic) untuk mengetahui tingkat akurasi hasil klasifikasi. Evaluasi yang dilakukan menunjukkan hasil yang termasuk Excellent Classification. Rule hasil klasifikasi diimplementasikan untuk pembuatan sistem prediksi penyakit diabetes. Sistem yang dibangun menggunakan Microsoft Visual Basic 6.0 dan database MySQL.

Kata kunci: Diabetes, Decision tree, Microsoft Visual Basic 6.0

ABSTRACT

Data diabetics increased from year to year. Diabetes diagnosis rate contributes significantly to the level of comorbidities and complications of diabetes. Based on historical data diabetics can made recommendations that help predict diabetes health professionals. Classification is one of data mining techniques that can be used to make predictions. Classification can be done with a decision tree, one of them with the C4.5 algorithm. This study to create a data classification of diabetes and apply them in the development of diabetes prediction system. Diabetes data classification results are evaluated by confusion matrix and ROC(Receiver Operating Characteristic) curves to determine the accuracy of the classification results. Evaluation results have shown that including Excellent Classification. Rule classification results are implemented for prediction system of diabetes. The system is built using Microsoft Visual Basic 6.0 and MySQL database.

Keyword: Diabetics, Decision tree, Microsoft Visual Basic 6.0

I. Pendahuluan

Menurut American Diabetes Association (ADA, 2005), diabetes mellitus merupakan suatu penyakit yang disebabkan karena kelainan sekresi insulin, kerja insulin atau bisa karena kedua-duanya yang juga merupakan penyakit metabolik dengan karakteristik hiperglikemia. Berdasarkan data pada laporan World Health Organization (WHO, 2010) menyebutkan dari 57 juta kematian global di tahun 2008, 36 juta atau 63% disebabkan karena penyakit tidak menular seperti jantung, diabetes kanker, dan penyakit pernafasan kronis. Dan angka tersebut diprediksikan akan terus meningkat dari tahun-ketahun.

Diabetes adalah penyakit yang kompleks dan rumit. Tingkat diagnosa diabetes memberikan kontribusi yang signifikan terhadap komorbiditas dan tingkat komplikasi diabetes (CDA, 2008). Berdasarkan data histori penderita penyakit diabetes dapat dibuat rekomendasi prediksi penyakit diabetes yang membantu tenaga kesehatan yaitu menggunakan klasifikasi data dengan *decision tree*.

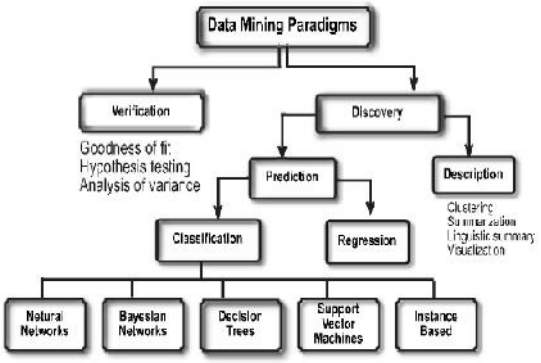
Tujuan penelitian ini adalah membuat sistem untuk prediksi penyakit diabetes dengan cara membuat klasifikasi data penderita diabetes menggunakan *decision tree* yang dapat diterapkan

dengan penggunaan algoritma C4.5. Hasil klasifikasi selanjutnya diaplikasikan pada sistem prediksi penyakit diabetes yang dibangun dengan Microsoft Visual Basic 6.0 dan *database* MySQL.

II. Tinjauan Pustaka

Data mining merupakan sebuah proses terpadu dari analisis data yang terdiri dari serangkaian kegiatan yang berjalan berdasarkan pada pendefinisian tujuan dari apa yang akan dianalisis sampai pada interpretasi dan evaluasi hasil (Giudici & Figini, 2009). Data mining juga dapat diartikan sebagai proses dalam menemukan pola dari sebuah set data dimana proses tersebut harus otomatis atau biasanya semi-otomatis dan pola yang dihasilkan harus berarti bahwa pola tersebut memberikan beberapa keuntungan (Witten, Frank, & Hall, 2011).

Klasifikasi atau taksonomi adalah proses menempatkan suatu objek atau konsep kedalam satu set kategori berdasarkan objek atau konsep yang bersangkutan (Gorunescu, 2011). Ada banyak metode data mining yang digunakan untuk tujuan yang berbeda-beda. Metode klasifikasi digunakan untuk membantu dalam memahami pengelompokkan data. Klasifikasi sendiri merupakan cabang dari *discovery data mining* seperti yang ditunjukkan pada gambar 1 (Maimon & Rokach, 2010) .



Gambar 1. Data mining taksonomi
Sumber: (Maimon & Rokach, 2010)

Pada gambar 1 menunjukkan data mining terdiri dari dua jenis yaitu *data mining verification oriented* yang berorientasi untuk memverifikasi data pengguna dan *data mining discovery oriented* yang digunakan untuk menemukan aturan baru atau pola mandiri dari sebuah set data. Jenis *data mining discovery* terdiri dari dua cabang yaitu *prediction* dan *description*. Metode *description* berorientasi pada interpretasi data yang berfokus pada pemahaman dengan visualisasi. Sedangkan metode *prediction* berorientasi bertujuan secara otomatis membangun model perilaku yang memperoleh sampel baru dan tak terlihat dan mampu memprediksi nilai-nilai dari satu atau lebih dari variabel yang terkait dengan sampel. Selain itu prediksi dikembangkan untuk memperoleh pola yang membentuk pengetahuan dengan cara yang dapat dimengerti dan mudah dioperasikan. *Data mining prediction* terdiri dari teknik klasifikasi dan regressi. Algoritma yang dapat digunakan untuk klasifikasi antara lain *Neural Network*, *Bayesian Network*, *Decision Trees*, *Support Vector Machine*, dan *Instance Based* (Maimon & Rokach, 2010).

Algoritma C4.5 dan *decision tree* merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah *decision tree*, dibutuhan algoritma C4.5. Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (*Iterative Dichotomiser*), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis *supervised learning* (Han & Kamber, 2006).

Menurut (Witten, Frank, & Hall, 2011) serangkaian perbaikan yang dilakukan pada ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan berpengaruh untuk *decision tree* yaitu C4.5. Perbaikan ini meliputi metode untuk menangani *numeric attributes, missing values, noisy* data, dan aturan yang menghasilkan *rules* dari *trees*.

Ada beberapa tahapan dalam membuat sebuah *decision tree* dalam algoritma C4.5 (Larose, 2005) yaitu :

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :

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

Keterangan :
S= Himpunan kasus
n = jumlah partisi S
P_i = proporsi S_i terhadap S

Kemudian hitung nilai gain menggunakan rumus :

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

Keterangan :
S = Himpunan Kasus
A = Fitur
n = jumlah partisi atribut A
|Si| = Proporsi Si terhadap S
|S| = jumlah kasus dalam S

3. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpatisi
4. Proses partisi *decision tree* akan berhenti saat :
   a. semua *record* dalam simpul N mendapat kelas yang sama.
   b. Tidak ada atribut didalam *record* yang dipartisi lagi
   c. Tidak ada *record* didalam cabang yang kosong

Sebuah *rule* hasil klasifikasi dengan *decision tree* jika diterapkan untuk prediksi perlu dilakukan evaluasi dan validasi hasil sehingga diketahui seberapa akurat hasil prediksi. Untuk evaluasi dan validasi hasil *rule* klasifikasi dapat menggunakan *confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*).

1. *Confusion Matrix*
   Metode ini hanya menggunakan tabel matriks seperti pada Tabel 1, jika dataset hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007).
   Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precison*, dan *recall*. *Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Han & Kamber, 2006). Sedangkan *precision* atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar (Powers, 2011).

Tabel 1 Model Confusion Matrix

Correct Classification	*Classified as*	
	+	-
+	*True positives*	*False negatives*
-	*False positives*	*True negatives*

Sumber: (Han & Kamber, 2006)

True Positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai positif, *false negative* adalah jumlah *record* positif yang diklasifikasikan sebagai negative, *true negative* adalah jumlah *record negative* yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall), Specifity, precision,* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *t_pos* terhadap jumlah *record* yang positif sedangkan *Specifity, precision* adalah perbandingan jumlah *t_neg* terhadap jumlah *record* yang negative. Untuk menghitung digunakan persamaan dibawah ini (Han & Kamber, 2006).

$$\text{Sensitifity} = \frac{t\_pos}{pos}$$

$$\text{Specifity} = \frac{t\_neg}{neg}$$

$$\text{Precision} = \frac{t\_pos}{t\_pos + f\_pos}$$

$$\text{accuracy} = \text{Sensitivity}\frac{pos}{(pos+neg)} + \text{Specifity}\frac{neg}{(pos+neg)}$$

Keterangan :
t_pos = Jumlah *true positives*
t_neg = Jumlah *true negative*
p = Jumlah *record positives*
n = Jumlah *tupel negatives*
f_pos = Jumlah *false positives*

2. Kurva ROC(*Receiver Operating Characteristic*)
   Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertical (Vercellis, 2009). *The area under curve* (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC digunakan dengan menggunakan rumus (Liao, 2007):

$$\theta^r = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} \psi\ (xt^r, xj^r)$$

Dimana :

$$\boldsymbol{\psi}(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

Keterangan :
X = Output positif
Y = Output negatif

ROC memiliki tingkat nilai diagnosa yaitu(Gorunescu, 2011):

Akurasi bernilai 0.90-1.00 = *excellent classification*
Akurasi bernilai 0.80-0.90 = *good classification*
Akurasi bernilai 0.70-0.80 = *fair classification*
Akurasi bernilai 0.60-0.70 = *poor classification*
Akurasi bernilai 0.50-0.60 = *failure*

Pembangunan sistem prediksi penyakit diabetes berdasarkan *rule* hasil klasifikasi dengan *decision tree* yaitu dengan algoritma C4.5 dibangun dengan Microsoft Visual Basic atau sering disebut dengan VB saja. VB merupakan sebuah bahasa pemrograman yang menawarkan *Integrated Development Environment* (IDE) visual untuk membuat program perangkat lunak berbasis sistem operasi Microsoft Windows dengan menggunakan model pemrograman (COM) (Jones, 2001).

Microsoft Visual Basic juga dapat diartikan sebagai sebuah aplikasi yang digunakan untuk pengembangan dengan memanfaatkan keistimewaan konsep-konsep antar muka grafis dalam Microsoft Windows. Aplikasi yang dihasilkan Visual Basic berkaitan erat dengan windows itu sendiri sehingga dibutuhkan pengetahuan bagaimana cara kerja windows (Suryana, 2009).

Visual Basic merupakan pengembangan dari bahasa komputer BASIC (*Beginner's All-purpose Symbolic Instruction Code*) yang diciptakan oleh Professor John Kemeny dan Thomas Eugene Kurtz pada pertengahan tahun 1960-an. Popularitas dan pemakaian BASIC yang luas dalam berbagai jenis komputer turut berperan dalam mengambangkan dan memperbaiki bahasa BASIC hingga akhirnya lahir Visual Basic yang berbasis GUI (*Graphic User Interface*) bersamaan dengan Microsoft Windows (Liberty, 2005).

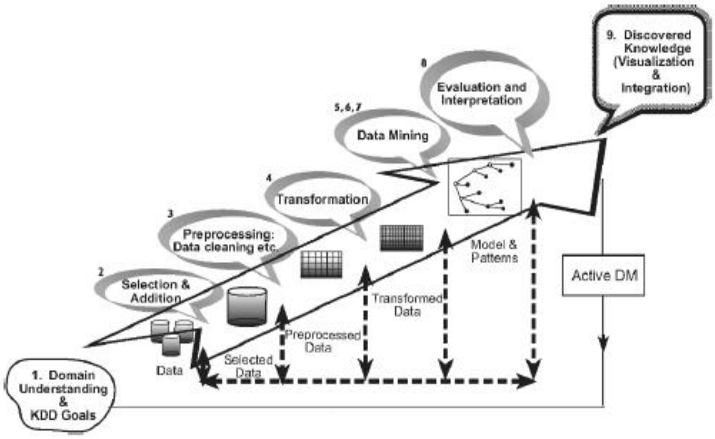
Microsoft Visual Basic 6.0 muncul pada pertengahan tahun 1998. Versi ini telah diimprovisasi di beberapa bagian termasuk kemampuan membuat aplikasi web meskipun kini VB6 sudah tidak didukung lagi, tetapi *file runtime*-nya masih didukung hingga Windows 7.

Sedangkan MySQL adalah sistem manajemen database yang bersifat bebas atau *open source*. MySQL mengolah *database* menggunakan bahasa SQL (Cabral & Murphy, 2009)

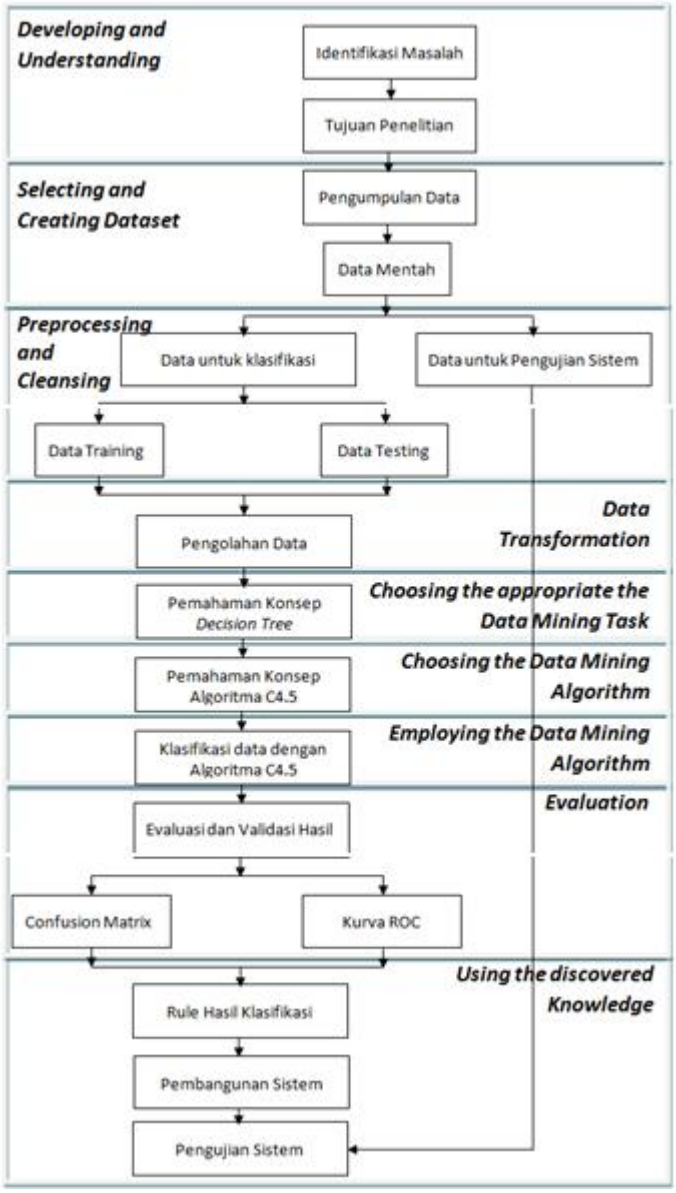
III.Metode Penelitian

Penelitian ini termasuk jenis penelitian eksperimen yang menggunakan data dalam penelitiannya dan menghasilkan kesimpulan yang mampu dibuktikan oleh pengamatan atau percobaan. Dalam penelitian eksperimen, peneliti harus memberikan hipotesis kerja atau membuat dugaan pada kemungkinan hasil. Setelah itu peneliti bekerja mendapatkan data untuk membuktikan atau menyangkal hipotesis. Kemudian peneliti membangun desain eksperimental yang menurutnya akan dapat memanipulasi orang atau data yang bersangkutan sehingga memperoleh informasi yang dikehendaki (Kothari, 2004).

Analisis data pada penelitian ini menggunakan *Knowledge Discovery in Databases* (KDD) yang terdiri dari 9 tahap seperti pada gambar 2.



Gambar 2. *Knowledge Discovery in Databases* (Maimon & Rokach, 2010)



Gambar 3. Kerangka Pemikiran

Gambar 3 menjelaskan kerangka pemikiran dari penelitian yang akan dilakukan dimana terdiri dari 9 tahap yang diadopsi dari 9 langkah KDD. Tahap pertama yaitu tahap *discovering and understanding* yang bertujuan memahami apa yang akan dilakukan dalam penelitian yaitu dilakukan dengan mengidentifikasi masalah dan menetapkan tujuan penelitian. Tahap kedua yaitu tahap *selecting and creating dataset* yang dilakukan dengan mengumpulkan data yang akan digunakan dalam penelitian yaitu data penderita diabetes dari repository uci.edu. Tahap ketiga selanjutnya yaitu *preprocessing and cleansing* yaitu memproses data dengan cara memilih data yang akan digunakan dan membuang data yang *missing value* (bernilai kosong/tidak lengkap) dan *noise* (tidak tepat). Hasilnya diperoleh data sejumlah 750 *record*. Setelah itu data dibagi menjadi dua yaitu data untuk klasifikasi sebanyak 700 *record* dan data untuk pengujian sistem sebanyak 50 *record*. Data klasifikasi dibagi menjadi dua untuk *testing* dan *training* dengan komposisi 80% dan 20% sehingga diperoleh data *training* sebanyak 560 *record* dan data *testing* sebanyak 140 *record*. Tahap keempat yaitu *data transformation* yang merupakan tahap untuk mengembangkan data menjadi lebih baik sesuai

dengan pola informasi yang dibutuhkan. Dalam penelitian ini dilakukan transformasi data kedalam kategori-kategori. Tahap kelima yaitu *choosing the appropriate data mining task* yaitu memilih teknik data mining yang akan digunakan. Untuk membuat suatu prediksi penyakit diabetes dipilih data mining klasifikasi. Tahap keenam yaitu *choosing data mining algorithm* yaitu memilih algoritma klasifikasi yang akan digunakan yaitu algoritma C4.5. Tahap ketujuh yaitu *employing data mining algorithm* yaitu membuat klasifikasi pada dataset dengan algoritma yang C4.5. Tahap kedelapan yaitu *evaluation* yang dilakukan dengan mengevaluasi hasil klasifikasi untuk mengukur tingkat akurasi dalam membuat prediksi penyakit diabetes. Evaluasi dilakukan dengan menggunakan *confusion matrix* dan kurva ROC. Tahap kesembilan yaitu using *the discovered knowledge* yaitu mengimplementasikan *rule* hasil klasifikasi kedalam pembangunan sistem prediksi penyakit diabetes dengan menggunakan Microsoft Visual Basic 6.0 dan *database* MySQL.

IV. Hasil dan Pembahasan

Klasifikasi data diabetes berbasis *decision tree* dengan menggunakan algoritma C4.5 dilakukan menggunakan *software* Rapidminer 5. Hasilnya dievaluasi untuk mengetahui tingkat akurasi dalam memprediksi penyakit diabetes sebagai berikut:

1. Evaluasi dengan *confusion matrix*
   Hasil evaluasi dengan *confusion matrix* dilakukan dua kali yaitu evaluasi hasil klasifikasi terhadap data *training* dan evaluasi terhadap data *testing*. Evaluasi terhadap data training ditampilkan dalam tabel 2 yang menunjukkan tingkat akurasi hasil klasifikasi jika diuji dengan data *training* sebesar 87,86%.

Tabel 2. Hasil evaluasi nilai *accuracy* terhadap data *training* dengan *Confusion Matrix*

accuracy: 87.86%			
	true Diabetes	true Tidak	class precision
pred. Diabetes	151	24	86.29%
pred. Tidak	44	341	88.57%
class recall	77.44%	93.42%	

Evaluasi hasil klasifikasi terhadap data *testing* ditampilkan dalam tabel 3. Dari tabel tersebut dapat diketahui tingkat *accuracy* terhadap data testing sebesar 84,29%.

Tabel 3. Hasil evaluasi nilai *accuracy* terhadap data *testing* dengan *Confusion Matrix*

accuracy: 84.29%			
	true Tidak	true Diabetes	class precision
pred. Tidak	36	14	86.00%
pred. Diabetes	3	32	80.00%
class recall	91.43%	69.57%	

2. Evaluasi dengan Kurva ROC
   Hasil evaluasi dengan kurva ROC dilakukan dua kali yaitu evaluasi hasil klasifikasi terhadap data *training* dan terhadap data *testing*. Evaluasi dengan kurva ROC terhadap data training dapat dilihat pada gambar 4. Dari gambar 4 dapat dilihat nilai diagnosa sebesar 0,963 sehingga termasuk *Excellent Classification*.

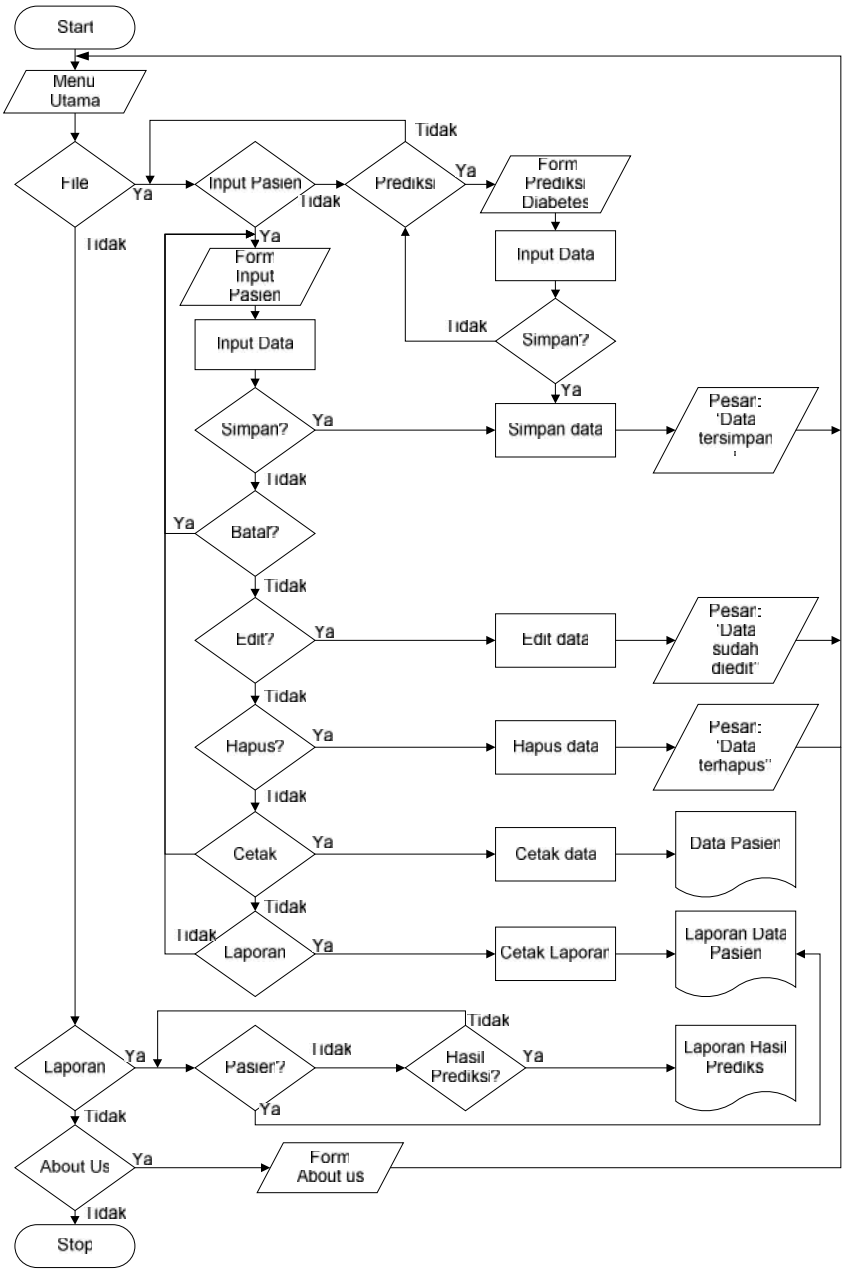
Gambar 4. Kurva ROC Data *Training*

Evaluasi hasil klasifikasi terhadap data testing dengan kurva ROC dapat dilihat pada gambar 5. Pada gambar 5 dapat dilihat hasil diagnosa sebesar 0,941 yang termasuk Excellent Classification.

Gambar 5. Kurva ROC Data *Testing*

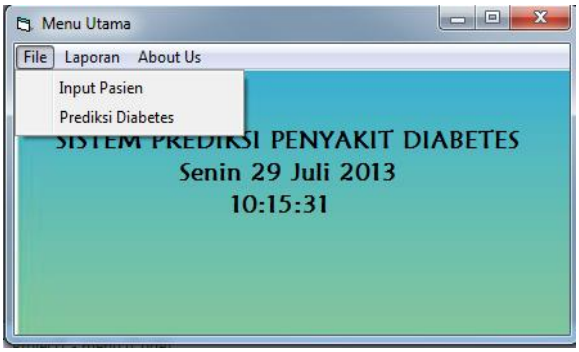
Berdasarkan hasil evaluasi dengan *Confusion Matrix* dan Kurva ROC menunjukkan *rule* hasil klasifikasi untuk memprediksi penyakit diabetes termasuk *Excellent Classification* sehingga dapat diimplementasikan kedalam sistem prediksi penyakit diabetes. Sistem dibangun dengan Microsoft Visual Basic 6.0 dan *database* MySQL. Database yang dibuat terdiri dari 2 tabel yaitu tabel pasien dan tabel prediksi.

Tahap setelah evaluasi hasil klasifikasi adalah tahap pembangunan sistem. Alur penggunaan sistem dapat dilihat pada gambar 6.



Gambar 6. *Flowchart* Sistem Prediksi Penyakit Diabetes

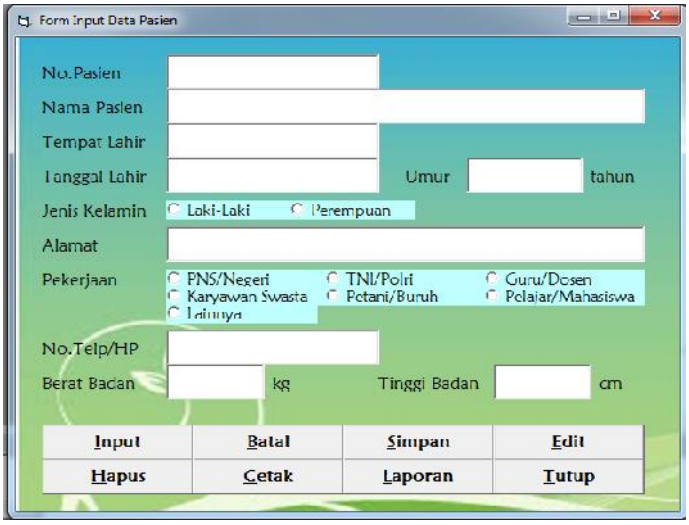
Gambar 6 menunjukkan menu utama sistem prediksi penyakit diabetes terdapat 3 sub menu yaitu File, Laporan, dan About Us. Sub menu file terdiri dari input data pasien dan form prediksi. Sub menu Laporan terdiri dari Laporan data pasien dan Laporan hasil prediksi. Sub menu About Us berisi informasi tentang sistem dan tata cara penggunaan. Bentuk dari form utama dari sistem ini dapat dilihat pada gambar 7.



Gambar 7. Menu Utama Sistem Prediksi Diabetes

Pada gambar 7 dapat dilihat menu utama terdiri dari menu File yang terdiri dari sub menu Input Pasien dan Prediksi Diabetes. Menu Laporan terdiri dari sub menu Laporan Data Pasien dan Laporan Hasil Prediksi.

Form input data pasien dapat dilihat pada gambar 8. Untuk menginput data pasien klik tombol Input maka otomatis tampil No.Pasien yang terdiri dari format tahun, bulan, tanggal, dan nomor urut pendaftaran pasien. Kemudian input data lainnya secara lengkap. Tombol Simpan digunakan untuk menyimpan inputan data kedalam tabel pasien. Batal untuk membatalkan inputan. Edit untuk mengubah data pasien yang dapat dilakukan dengan mencari dahulu data pasien dengan menginputkan nomor pasien. Tombol Hapus digunakan untuk menghapus data pasien dari tabel pasien yang dapat dilakukan dengan mencari dahulu data pasien dengan menginputkan nomor pasien. Cetak digunakan untuk mencetak data pasien, sedangkan laporan digunakan untuk mencetak laporan data pasien.



Gambar 8. Form Input Data Pasien

Form prediksi penyakit diabetes dapat dilihat pada gambar 9. Pada form tersebut kita input dulu no pasien kemudian tekan enter, otomatis tampil data pasien sesuai yang tersimpan di *database*. Setelah itu input data hasil Lab maka akan tampil Hasil Prediksi Diabetes atau Tidak.

Gambar 9. Form Prediksi Penyakit Diabetes

Sistem diuji menggunakan data baru sebanyak 50 record data. Hasil pengujian sistem menunjukkan tingkat akurasi pada confusion matrix sebesar 73.33% dan 0,815 pada kurva ROC sehingga termasuk *Good Classification*.

V. Kesimpulan dan Saran

Kesimpulan yang diperoleh dari penelitian ini adalah klasifikasi data penderita diabetes dengan teknik data mining klasifikasi yang menggunakan algoritma C4.5 menghasilkan *rule* yang dapat digunakan untuk prediksi penyakit diabetes. Pengujian hasil klasifikasi menggunakan *confusion matrix* dan kurva ROC menghasilkan klasifikasi yang termasuk *Excellent Classification*. *Rule* hasil klasifikasi diterapkan dalam pembangunan sistem prediksi diabetes dengan menggunakan Microsoft Visual Basic 6.0 dan database MySQL. Hasil pengujian sistem prediksi diabetes pada data baru menunjukkan hasilnya termasuk *Good Classification*.

Saran untuk penelitian selanjutnya adalah komparasi dengan beberapa algoritma untuk prediksi penyakit diabetes sehingga diperoleh algoritma dengan tingkat akurasi tertinggi. Selain itu penelitian dilakukan dengan data diabetes yang lain dengan jumlah yang lebih banyak.

Daftar Pustaka

Bramer, M. (2007). Principles of Data Mining. United Kingdom: Springer.

Cabral, S. K., & Murphy, K. (2009). *MySQL Administrator's Bible*. Indianapolis: Wiley Publishing, Inc.

CDA. (2008). Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Canadian Journal of Diabetes* .

Giudici, P., & Figini, S. (2009). *Applied Data Mining for Business and Industry Second Edition.* United Kingdom: John Willey and Sons.

Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques.* Berlin: Springer.

Han, J., & Kamber, M. (2006). *Data Mining Concepts And Techniques 2nd Edition.* San Fransisco: Elsevier.

Jones, P. (2001). *Visual Basic: A Complete Course.* London: Continuum.

Kothari, C. R. (2004). *Research Methodology Methods and Techniques, Second Revised Edition.* New Delhi: New Age International Publishers.

Larose, D. T. (2005). *Discovering Knowledge in Data An Introduction to Data Mining.* New Jersey: John Willey and Sons.

Liao, T. W. (2007). Enterprise Data Mining: A Review and Research Directions. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications* , 1-109.

Liberty, J. (2005). *Programming Visual Basic 2005.* USA: O'Reilly Media.

Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition.* London: Springer.

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure To ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies* , 37-63.

Suryana, T. (2009). *Visual Basic.* Yogyakarta: Graha Ilmu.

Vercellis, C. (2009). *Business Intelligence.* United Kingdom: John Wiley and Sons.

WHO. (2010). *Global Status Report on noncommunicable Disease.* Switzerland: WHO Press.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition.* Burlington: Elsevie