

**PENGEMBANGAN PERANGKAT LUNAK DIAGNOSA PENYAKIT  
DIABETES MELLITUS TIPE II BERBASIS TEKNIK KLASIFIKASI DATA**

**Rodiyatul FS<sup>1</sup>, Bayu Adhi Tama<sup>2</sup>, Megah Mulya<sup>3</sup>**

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Sriwijaya  
<sup>1</sup>rodiyatulfs@yahoo.co.id, <sup>2</sup>bayu@unsri.ac.id, <sup>3</sup>megahmulya@unsri.ac.id

**ABSTRACT**

Type-2 of Diabetes Mellitus (Type II DM) is the most common type of diabetes whose patient about 90-95% of all diabetes population. Early Detection of Type II DM from various risk factors is a way to prevent the complication that causes mortality. Developing software for diagnosis Type II DM could be utilized as an alternative and would enhance medical care for the increasing number of patients. In this paper, the classification technique of data mining with C4.5 algorithm classifier is employed to acquire valuable information and extract pattern from medical record data. This pattern is used as knowledge base in medical diagnosis process. It absolutely helps doctors and other clinicians for making decision through early detection of Type II DM.

**Keywords.** data mining, Type II DM, diagnosis, software

**PENDAHULUAN**

Saat ini penyakit Diabetes Mellitus (DM) Tipe II telah menjadi salah satu penyakit kronik yang paling sering diderita di Indonesia. Berdasarkan survei, diperkirakan pada tahun 2020 akan ada 178 juta penduduk berusia diatas 20 tahun memiliki prevelansi terkena DM, suatu jumlah yang besar untuk dapat ditangani sendiri oleh para ahli DM [4]. Tingginya angka-angka statistik diatas, tentunya patut diantisipasi oleh pihak penyedia layanan kesehatan seperti rumah sakit untuk mencegah timbulnya ledakan pasien DM.

Pada zaman modern ini, banyak rumah sakit telah mengimplementasikan teknologi informasi untuk meningkatkan pelayanan medis dan mengatur data warehouse Ketersediaan instrumentasi dan informatika medis modern (telemedis) seperti adanya suatu perangkat lunak untuk menunjang keputusan seorang dokter dalam mendiagnosa suatu penyakit sangat dibutuhkan. Teknologi data mining hadir sebagai solusi nyata bagi para pengambil keputusan seperti dokter dalam

memprediksi pasien yang beresiko terkena terkena penyakit DM Tipe II. Dengan menerapkan teknik klasifikasi data mining, dapat ditemukan informasi yang berharga pada sekumpulan data yang berukuran besar.

### TEKNIK KLASIFIKASI DATA MINING

Klasifikasi adalah proses penemuan pola atau fungsi yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui[3]. Konsep klasifikasi dengan pengawasan (supervised classification) adalah untuk membangun sebuah model dari data yang telah diketahui, atau sering disebut sebagai classifier. Model atau fungsi ini kemudian dapat digunakan untuk memetakan data didalam suatu basis data kepada suatu atribut target, selanjutnya dapat memperkirakan suatu kelas dari data yang baru

Tiap record berisi banyak atribut dimana masing-masing atribut memiliki satu dari beberapa kemungkinan nilai. Di dalam klasifikasi diberikan sejumlah record yang dinamakan sekumpulan data latih yang terdiri dari beberapa atribut, dimana salah satu atribut menunjukkan kelas untuk record.

### ALGORITMA KLASIFIKASI DATA MINING

Pada paper ini digunakan algoritma klasifikasi C4.5 yang melakukan pemilihan atribut terbaik berdasarkan informasi gain. Atribut dengan informasi gain tertinggi akan dipilih untuk membuat keputusan. Informasi gain merupakan selisih antara kebutuhan informasi awal (yang hanya bergantung pada jumlah dan proporsi tiap kelas di dalam D) dan kebutuhan informasi baru (yang diperoleh setelah melakukan partisi terhadap atribut A). Untuk menghitung gain, digunakan rumus :

$$\text{Gain (A)} = \text{Info (D)} - \text{Info}_A (\text{D})$$

Informasi yg dibutuhkan untuk mengklasifikasi sebuah record di D diberikan dengan rumus berikut.

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$

Informasi yg dibutuhkan untuk mengklasifikasi sebuah record di D diberikan dengan rumus berikut berdasarkan hasil partisi di A.

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{Info}(D_j)$$

Pseudocode algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [7] :

1. Cek Sekumpulan Data Latih
2. Pada masing-masing atribut a, hitung informasi gain
3. Ubah a terbaik ( dengan informasi gain tertinggi) menjadi akar
4. Buat simpul keputusan yang berakar dari a terbaik
5. Ulangi proses untuk masing-masing cabang dengan memilih a terbaik berikutnya dan jadikan anak dibawah simpul keputusan terakhir.

## ANALISIS

### 1. Data Preparation

Data utama yang digunakan pada penelitian ini berupa sekumpulan data rekam medis pasien rawat inap RSMH Palembang untuk penyakit Diabetes Mellitus Tipe 2 sepanjang tahun 2008 yang berjumlah 435 instances.

### 2. Pembangkitan dan Pengujian Model Klasifikasi

Sebelum pembangkitkan model klasifikasi berupa 39 rules, dibentuk sebuah pohon keputusan menggunakan algoritma C4.5. Untuk menguji tingkat akurasi model klasifikasi digunakan pengujian kualitatif dengan mewawancarai dua dokter untuk mengetahui seberapa besar user-acceptance terhadap model klasifikasi. Dari hasil penilaian, didapatkan 34 rules yang sesuai dengan standar penegakan diagnosa WHO dan PERKENI (Perkumpulan Endokrinologi Indonesia). Disamping

itu, dilakukan juga pengujian kuantitatif dengan menggunakan metode confusion matrix yang menghasilkan tingkat akurasi model sebesar 95,63%.

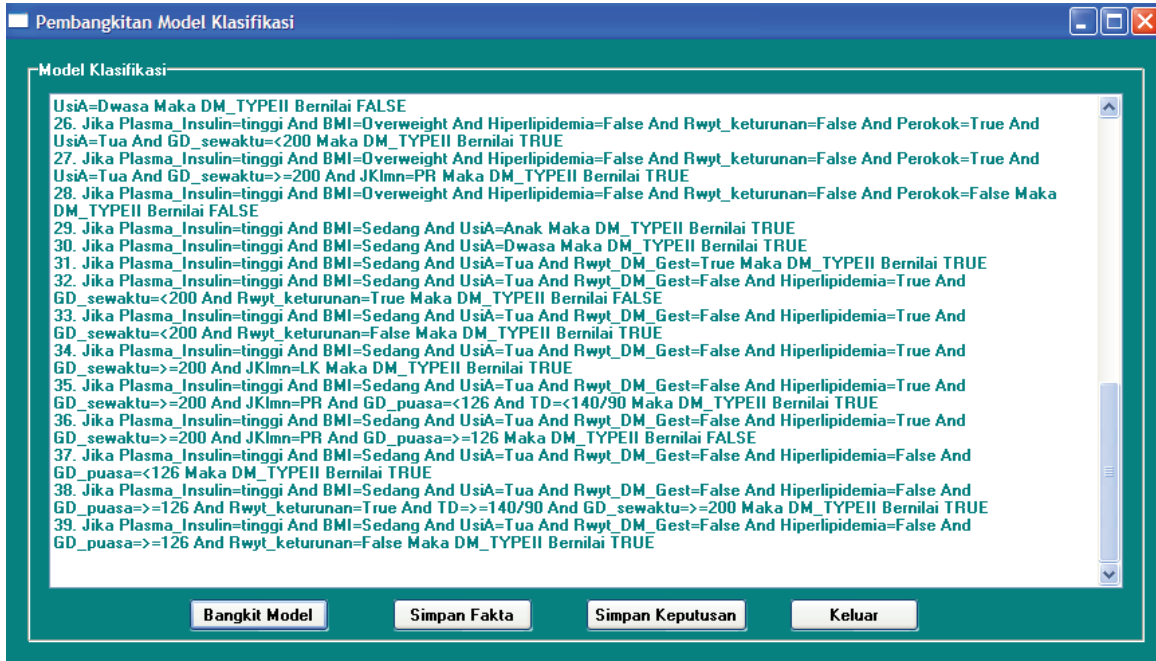
### 3. Hasil dan Pembahasan

Hasil implementasi perangkat lunak dapat dilihat pada beberapa tampilan berikut :

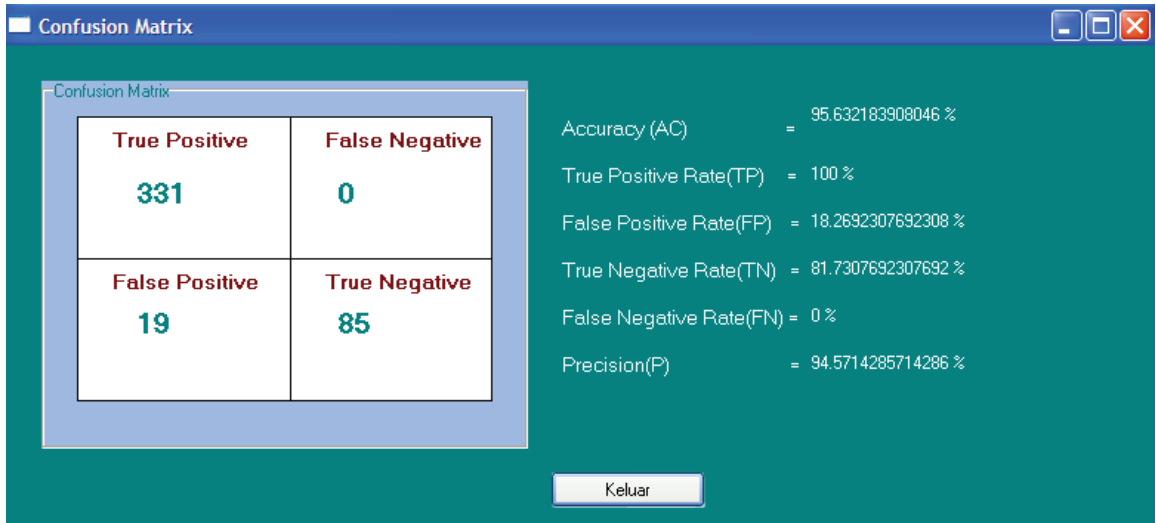
The screenshot displays the 'Pembelajaran Data Latih' application window. On the left, the 'TABEL DATA KASUS' table lists 23 cases with attributes like UsjA, Tua, JkImn, PR, BMI, Dverweight, TD, and TD. Below the table, the 'GAIN TERTINGGI' section shows the highest gain for the 'Plasma\_Insulin' attribute. On the right, a decision tree is visualized, starting with the root node 'Plasma\_Insulin' and branching based on various attributes and their values to classify the data.

**Gambar 4.1 Form Pembangkitan Model Pohon**

Berdasarkan hasil pembangkitan model klasifikasi, maka dapat dilihat bahwa plasmainsulin memiliki gain ratio yang paling tinggi, oleh karena itu atribut ini paling berpengaruh terhadap penyakit diabetes mellitus tipe II



Gambar 4.2 Form Pembangkitan Model Klasifikasi



Gambar 4.3 Form Pengujian Model Klasifikasi

**Gambar 4.4 Form Proses Diagnosa**

## KESIMPULAN DAN SARAN

Melalui teknik klasifikasi data mining yang digunakan, paper ini telah berhasil mengumpulkan dan menganalisa data rekam medis pasien diabetes mellitus tipe II, dan menghasilkan beberapa rules yang dapat digunakan pihak rumah sakit dalam pengambilan keputusan di bidang kesehatan, khususnya dalam mendiagnosa penyakit diabetes mellitus tipe II.

Kuantitas training data yang digunakan sebaiknya ditambah untuk mengekstrak kemungkinan munculnya tambahan informasi bernilai lainnya. Disamping itu, sebaiknya dilakukan penelitian lanjutan dengan membandingkan algoritma klasifikasi yang lain pada proses pembangkitan model klasifikasi sehingga dihasilkan performance yang lebih baik.

**DAFTAR PUSTAKA**

- American Diabetes Association. 2004. Diagnosis and Classification of Diabetes Mellitus. American Journal of Diabetes Care. 27. S5-S10
- Bramer, Max. Principles of Data Mining, Springer-Verlag London Limited, 2007.
- Han, J., et al. Data Mining: Concepts and Techniques 2nd Edition, Morgan Kaufmann Publisher, 2006.
- Indah. 2009. [Online] Tersedia: [www.indahmuhariani.com/index.php /2009/02/01](http://www.indahmuhariani.com/index.php/2009/02/01). [diakses terakhir tanggal 10 Februari 2009]
- Larose, D.T. 2005. Discovering Knowledge in Data : An Introduction to Data Mining. John Wiley & Sons, Inc, New Jersey.
- Palaniappan, S. and R.Awang. 2008. Intelligent Heart Disease Prediction System Using Data Mining Techniques. International Journal of Computer Sciences and Network security. 8 (8), 343-350.
- Quinlan.2009.C4.5 Algorithm. [Online] Tersedia : [http://en.wikipedia.org/wiki /C4.5\\_algorithm](http://en.wikipedia.org/wiki/C4.5_algorithm). [10 Februari 2009]
- Witten, Ian H. And Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques 2nd Edition, Morgan Kaufmann Publisher, 2005.
- World Health Organization. 2002. Laboratory Diagnosis and Monitoring of Diabetes Mellitus. WHO Publication, Switzerland.