KLASIFIKASI DATA MEDIS DENGAN MENGGUNAKAN ALGORITMA *CART*

Tugas Akhir disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer (S.Kom.)
Di
Universitas Trunojoyo Madura

Oleh :
Nama : Mega Septi Yolanda

NRP : 08.04.111.00047

Disetujui oleh Tim Penguji Tugas Akhir:		**Tanggal Sidang: 2013**
Bain Khusnul K, S.T.,M.Kom NIP. 198003252003122002	_____	(Pembimbing I)
Yeni K.,S.Kom.,M.Kom. NIP. 19770921 2008122002	_____	(Pembimbing II)
Mula'ab,S.Si.,M.Kom. NIP. 197305262003122001	_____	(Penguji I)
Andharini Dwi C,S.Kom.,M.Kom. NIP. 198005032003122002	_____	(Penguji II)
Arik Kurniawati.S.Kom.,M.T NIP. 197803092003122009	_____	(Penguji III)

Bangkalan, 2013
Mengetahui,
Ketua Program Studi

Firdaus Solihin,S.Kom.,M.Kom
NIP. 19760627 200801 1 008

MOTTO

Sukses itu
Belajar, berdo'a dan berusaha

PERSEMBAHAN

Kupersembahkan Tugas Akhir ini
untuk ibunda dan ayahanda tercinta,
Atas segala doa dan motifasi yang diberikan tiada henti-hentinya
kepada penulis,
Serta untuk orang tersayang dan temen-temen yang selalu
memberikan motifasi sehingga tugas akhir ini dapat terselesaikan
dengan baik.

Universitas Trunojoyo

KLASIFIKASI DATA MEDIS DENGAN MENGGUNAKAN *ALGORITMA CART*

Mega Septi Yolanda
(08.04.112.00047)

Pembimbing 1 : Bain Khusnul Hotimah,S.T.,M.Kom
Pembimbing 2 : Yeni Kustiyahningsih,S.Kom.,M.Kom

ABSTRAK

Konsep pohon merupakan salah satu konsep teori graf yang paling penting. Pemanfaatan pohon dalam kehidupan sehari-hari adalah untuk menggambarkan hierarki dan pemodelan persoalan, contohnya algoritma pohon keputusan CART.

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Pohon keputusan merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk dipahami. CART merupakan algoritma pohon keputusan yang sering digunakan untuk membuat suatu pohon keputusan karena memiliki tingkat akurasi yang tinggi dalam menentukan keputusan.

Pada penelitian ini dibangun aplikasi untuk mengklasifikasi data medis. Aplikasi data medis ini menggunakan algoritma CART yang merupakan algoritma pohon keputusan yang sering digunakan karena memiliki tingkat akurasi yang tinggi dalam menentukan keputusan.

Dari hasil ujicoba pengukuran kinerja algoritma menggunkanan 2 skenario yang telah dilakukan, dapat disimpulkan bahwa pada skenario 1 Breast Cancer merupakan ujicoba paling efektif karena akurasi yang dihasilkan sama-sama mencapai 92% pada algoritma CART 150:50 dan 100:100 sebesar 92%. Sedangkan algoritma CART memiliki kinerja (*precision, recall, dan accuracy*) yang lebih tinggi. dibandingkan algoritma CART Diabetes. Ini terlihat dari nilai akurasi CART 150:50 sebesar 70% sedangkan untuk CART Diabetes 100:100 sebesar 68%.

Kata kunci : Pohon Keputusan, *CART*, *Klasifikasi* Data Medis, *Breast Canser, Diabetes.*

KATA PENGANTAR

Puji syukur penulis ucapkan kehadiran Allah SWT yang telah memberika rahmat dan hidayahnya kepada kita semua, serta sholawat dan salam semoga tercurah dan terlimpahkan kepada junjungan kita Muhammad SAW, sehingga penulis dapat menyelesaikan laporan Tugas Akhir dengan judul :

Klasifikasi Data Medis Dengan Menggunakan *Algoritma CART*

Adapun tujuan penyusunan laporan Tugas Akhir ini guna memenuhi persyaratan untuk mengakhiri studi dan memperoleh gelar sarjana Teknik Informatika pada program studi Strata-1 Jurusan Teknik Informatika Universitas Trunojoyo.

Keberhasilan dalam menyelesaikan Tugas Akhir ini tidak terlepas dari bantuan, dukungan saran dan bimbingan dari berbagai pihak. Dan perkenankanlah pada kesempatan ini penulis menyampaikan rasa terima kasih sedalam-dalamnya kepada :

1. Allah SWT atas tuntunan dan bimbingannya.
2. Bapak dan Ibu tercinta dirumah yang telah memberikan doa, dorongan dan dukungannya baik moril maupun materil sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir ini.
3. Ibu Bain Khusnul K, S.T.,M.Kom dan Ibu Yeni K., S.Kom.,M.Kom. selaku Dosen Pembimbing yang telah memberikan petunjuk, arahan dan bimbingan sejak dari perencanaan tugas akhir sampai pada penulisan laporan tugas akhir ini.
4. Semua Teman-temanku yang tidak bisa penulis sebutkan satu per satu, terima kasih untuk dukungannya.
5. Seluruh pihak baik langsung maupun tidak langsung yang telah membantu dalam penyusunan tugas akhir ini.

Penulis menyadari sepenuhnya keterbatasan kemampuan yang penulis miliki, sehingga laporan Tugas Akhir ini masih jauh dari kesempurnaan dan masih banyak kekurangan dan kesalahannya. Akhir kata penulis berharap semoga semoga segala sesuatu yang telah dihasilkan dalam Tugas Akhir ini dapat bermanfat bagi masyarakat, dan kemajuan Ilmu pengetahuan dan teknologi untuk dapat dikembangkan lebih baik lagi.

DAFTAR ISI

Hal.

Halaman Judul i
Lembar Pengesahan ii
Motto iii
Abstrak iv
Kata Pengantar v
Daftar Isi vi
Daftar Gambar viii
Daftar Tabel ix

BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 2
1.3 Tujuan dan Manfaat 2
1.4 Batasan Masalah 2
1.5 Metodologi Pengerjaan Tugas Akhir 2
1.6 Sistematika Penulisan 3

BAB II TINJAUAN PUSTAKA 5
2.1 Penelitian Sebelumnya 5
2.2 Data Mining 5
2.3 Pohon Keputusan 9
2.3.1 Pohon Keputusan CART 14
2.3.2 Data Medis 16
2.3.2.1 Diabetes 17
2.3.2.2 Breast Cancer 18

BAB III PERANCANGAN SISTEM 20
3.1 Flowchart 20
3.1.1 Flowchart Sistem 20
3.1.2 Flowchart Pembentukan Pohon Keputusan CART ... 21
3.2 Use Case Diagram 22
3.3 Activity Diagram 23
3.3.1 Activity Diagram Manage Data Training……..… 24

3.3.2 Activity Diagram Proses Mining..... 25
3.3.3 Activity Diagram Manage Penentu Keputusan..... 26
3.3.4 Activity Diagram View Keputusan..... 27
3.4 Pengukuran Kinerja..... 28

BAB IV IMPLEMENTASI..... 30
4.1 Batasan Implementasi..... 30
4.1.1 Spesifikasi Perangkat Keras..... 30
4.1.2 Spesifikasi Perangkat Lunak..... 30
4.2 Proses CART..... 30
4.2.1 Proses pembagian candidate Splite..... 30
4.2.2 Menghitung PL (Purity Left)..... 36
4.2.3 Menghitung dan PR (Purity Right)..... 36
4.2.4 Menghitung P(j/tL) Malignant..... 36
4.2.5 Menghitung P(j/tL) Benint..... 36
4.2.6 Mengitung P (j/tR) Malignant..... 37
4.2.7 Mengitung P (j/tR) Benint..... 37
4.2.8 Mengitung 2PLPR..... 37
4.2.9 Mengitung Q (s/t)..... 38
4.2.10 Mengitung Φ(s/t)..... 39
4.3 Data yang Digunakan..... 45
4.4 Data Breast Cancer..... 45
4.4.1 Skenario 1..... 45
4.5 Pohon Keputusan Breast Cancer Skenario 1..... 46
4.5.1 Breast Cancer Data 150:50..... 46
4.5.2 Breast Cancer Data 100:100..... 47
4.6 Penentu Keputusan Breast Cancer..... 48
4.7 Kinerja Algoritma Breast Cancer..... 49
4.7.1 Skenario 1..... 49
4.8 Data Diabetes..... 50
4.8.1 Skenario 2..... 50
4.9 Pohon Keputusan Diabetes Skenario 2..... 51
4.9.1 Diabetes Data 150:50..... 51
4.9.2 Diabetes Data 100:100..... 53
4.10 Penentu Keputusan Diabetes..... 54

4.11 Kinerja Algoritma Diabetes 55
4.11.1 Skenario 2 55
4.12 Analisa Algoritma 56
BAB V PENUTUP 60
5.1 Kesimpulan 60
5.2 Saran 60
DAFTAR PUSTAKA 61
LAMPIRAN 64

Universitas Trunojoyo

DAFTAR GAMBAR

		Hal.
Gambar 2.1	Tahap-Tahap Data Mining	8
Gambar 2.2	Konsep Dasar Dalam Pohon Keputusan	10
Gambar 2.3	Pohon Keputusan Pada Pembelian Komputer	12
Gambar 2.4	Pohon Keputusan pada Penyakit *Diabetes*	16
Gambar 2.5	Pohon Keputusan pada Penyakit *Breast Cancer*	17
Gambar 3.1	Flowchart Sistem	21
Gambar 3.2	Flowchart Pohon Keputusan CART	22
Gambar 3.3	Use Case Diagram	23
Gambar 3.4	Activity Diagram Manage Data Training	25
Gambar 3.5	Activity Diagram Proses Mining	26
Gambar 3.5	Activity Diagram View Pohon Keputusan	26
Gambar 3.5	Activity Diagram Manage Penentu Keputusan	27
Gambar 3.5	Activity Diagram View Keputusan	27
Gambar 4.1	Pohon Keputusan *Tree*	46
Gambar 4.1	Pohon Keputusan *Breast Cancer* 150:50	47
Gambar 4.2	Pohon Keputusan *Breast Cancer* 100:100	48
Gambar 4.3	Pohon Keputusan *Diabetes* 150:50	52
Gambar 4.4	Pohon Keputusan *Diabetes* 100:100	54
Gambar 4.5	Grafik Skenario 1	58
Gambar 4.6	Grafik Skenario 2	59

DAFTAR TABEL

Hal.

Tabel 2.1 Frekuensi Penggunaan Algoritma Pohon Keputusan …12
Tabel 2.2 Atribut Diabetes 18
Tabel 2.3 Atribut Breas Cancer 19
Tabel 3.1 Pengukuran Kinerja 28
Tabel 4.1 *Training Set* Untuk *Klasifikas*i (*Numerik*) 31
Table 4.2 Data *Transformasi*.....31
Table 4.3 *Training Set* untuk klasifikasi (*kategorikal*).....32
Table 4.4 *Candidate Split*.....33
Table 4.5 Data *Frekuensi* kemunculan.....35
Table 4.6 Hasil Perhitungan Φ(s/t) dari setiap *candidate split*……40
Table 4.7 Data *Frekuensi* kemunculan kedua 42
Table 4.8 Hasil perhitungan (s/t) dari setiap *Candidate splite* ke2.44
Table 4.9 Data Skenario 1 45
Table 4.10 Penentu Keputusan *Breast Cancer CART*……49
Table 4.11 Penilaian CART Breas Canser150:50 ..... .49
Table 4.12 Penilaian CART Breas Cancer 100:100…..... .50
Table 4.13 Data Skenario 2 ..... .50
Table 4.14 Penentu Keputusan *Diabetes CART*……55
Table 4.15 Penilaian CART Diabetes 150:50 ..... .55
Table 4.16 Penilaian CART Diabetes 100:100 ..... .56
Table 4.17 Kinerja Algoritma ..... .56