

**IMPLEMENTASI ALGORITMA *CLUSTERING*
(*COBWEB* DAN *ITERATE*) SEBAGAI BAGIAN DARI DATA
MINING ALGORITHMS COLLECTION**

SKRIPSI

**ADOLF PANDAPOTAN
1204000041**



**UNIVERSITAS INDONESIA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
DEPOK
JULI 2008**

**IMPLEMENTASI ALGORITMA *CLUSTERING*
(*COBWEB* DAN *ITERATE*) SEBAGAI BAGIAN DARI DATA
MINING ALGORITHMS COLLECTION**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Ilmu Komputer**

**ADOLF PANDAPOTAN
1204000041**



**UNIVERSITAS INDONESIA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
DEPOK
JULI 2008**

Universitas Indonesia

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh

Nama : Adolf Pandapotan

NPM : 1204000041

Program Studi : Ilmu Komputer

Judul Tugas Akhir : Implementasi Algoritma *Clustering (Cobweb dan Iterate)*
Sebagai Bagian Dari *Data Mining Algorithms Collection*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Ilmu Komputer pada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia.

DEWAN PENGUJI

Pembimbing : Yova Ruldeviyani, M.Kom. (.....)

Penguji : Dr. Indra Budi (.....)

Penguji : Siti Aminah, M.Kom (.....)

Ditetapkan di : Depok

Tanggal : 21 Juli 2008

KATA PENGANTAR

Berkat rahmat dan karunia Tuhan, akhirnya Laporan Tugas Akhir berjudul “IMPLEMENTASI ALGORITMA *CLUSTERING (COBWEB DAN ITERATE)* SEBAGAI BAGIAN DARI *DATA MINING ALGORITHMS COLLECTION*” ini dapat diselesaikan. Pada kesempatan ini penulis juga tidak lupa mengucapkan terima kasih kepada semua pihak yang telah memberikan bimbingan, bantuan, dan dukungan kepada penulis, baik selama melakukan penelitian Tugas Akhir atau pun pada saat penulisan Tugas Akhir. Ucapan terima kasih penulis sampaikan kepada:

1. Papa, Mama, Abang Teddy, Kakak Ita, dan Adik Evi yang selalu memberikan semangat serta mendoakan sehingga dapat menyelesaikan Tugas Akhir ini.
2. Ibu Yova Ruldeviyani selaku Pembimbing Tugas Akhir yang telah memberikan saran dan masukan selama implementasi hingga penulisan Laporan Tugas Akhir ini.
3. Bapak Ade Azurat selaku Pembimbing Akademik yang telah memberikan dorongan semangat.
4. Anita Megasari selaku teman dekat penulis yang telah memberikan semangat dan doa agar dapat menyelesaikan Tugas Akhir dengan baik.
5. Hugo, Ishak, dan Henri selaku teman Kelompok Kecil (KK) PO Fasilkom yang telah memberikan semangat dan doa.
6. Sadar Baskoro, Yemima (Mimi), Jeremia, Jani, Daniel Albert, serta seluruh teman-teman PO Fasilkom yang telah memberikan dukungan dan doa.
7. Franova (Smile), Martin, Mulki, Hendra, Agung (AW), Andre, Rangga, Arudea (2005), Michael, selaku teman Lab Tugas Akhir.
8. Teman-teman Fasilkom Angkatan 2004 yang telah bersama selama empat tahun di Fasilkom. Terima kasih atas persahabatan selama ini.

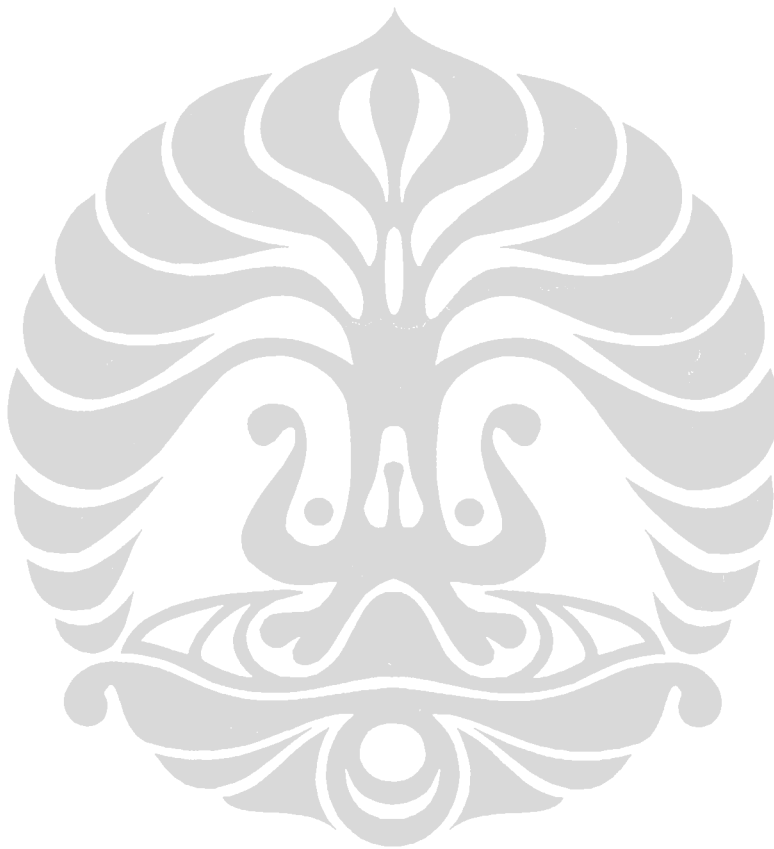
Universitas Indonesia

9. Segenap civitas akademika Fasilkom UI atas dukungannya dan bantuannya.

Penulis menyadari masih banyak kekurangan dalam Laporan Tugas Akhir ini. Untuk itu, penulis mengharapkan saran dan kritik dari berbagai pihak. Akhir kata, penulis berharap agar Laporan Tugas Akhir ini dapat bermanfaat bagi yang membaca.

Depok, Mei 2008

Penulis



Universitas Indonesia

DAFTAR ISI

KATA PENGANTAR	iv
ABSTRAK	vi
<i>ABSTRACT</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
1. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Tujuan penelitian.....	3
1.3 Ruang Lingkup Tugas Akhir	4
1.4 Metode Penelitian	4
1.5 Sistematika Penulisan Laporan.....	4
2. LANDASAN TEORI.....	7
2.1 <i>Data Mining</i>	7
2.2 <i>Clustering</i>	9
2.4 Algoritma <i>Cobweb</i>	13
2.4.1 Fungsi evaluasi heuristik.....	14
2.4.2 Representasi <i>state</i> (representasi <i>concept</i>).....	16
2.4.3 Operator.....	17
2.4.4 Strategi kontrol.....	21
2.5 Algoritma <i>Iterate</i>	29
2.5.1 Pembentukan klasifikasi <i>tree</i> (<i>classified</i>)	33
2.5.2 Pengumpulan partisi awal yang “baik” dari klasifikasi <i>tree</i> untuk membentuk <i>cluster-cluster</i> yang diinginkan (<i>extract</i>)	35
2.5.3 <i>Iterative redistribution</i>	37

2.6 WEKA	45
3. ANALISA DAN PERENCANAAN.....	47
3.1 Analisa Kebutuhan	47
3.1.1 Analisa <i>Input</i>	47
3.1.2 Analisa <i>Output</i>	48
3.2 Analisa dan Perancangan Algoritma	50
3.3 Permasalahan Analisa dan Perencanaan.....	60
4. IMPLEMENTASI ALGORITMA	61
4.1 Membaca data.....	61
4.2 Algoritma <i>Cobweb</i>	65
4.2.1 Representasi <i>State</i>	66
4.2.2 Fungsi evaluasi heuristik.....	69
4.2.3 Operator.....	70
4.2.4 Strategi Kontrol.....	72
4.3 Algoritma <i>Iterate</i>	74
4.3.1 Langkah pertama: Pembentukan klasifikasi <i>tree(classified)</i>	77
4.3.2 Langkah kedua: Pengumpulan partisi awal yang “baik” dari klasifikasi <i>tree (extract)</i>	80
4.3.3 Langkah ketiga: <i>Iterative redistribution</i>	81
4.4 Tampilan sistem.....	82
5. HASIL UJI COBA IMPLEMENTASI	84
5.1 Penjelasan Umum Hasil Pengujian	84
5.2 Hasil Pengujian <i>Dataset</i> Kecil.....	87
5.2.1 Algoritma <i>Cobweb</i>	87
5.2.2 Algoritma <i>Iterate</i>	89
5.3 Hasil Pengujian <i>Dataset</i> Besar	90

5.3.1 Algoritma <i>Cobweb</i>	90
5.3.2 Algoritma <i>Iterate</i>	93
6. KESIMPULAN DAN SARAN	95
6.1 Kesimpulan	95
6.2 Saran	96
DAFTAR PUSTAKA	97
Lampiran A	99
Lampiran B	100
Lampiran C	101
Lampiran D	102
Lampiran E	103
Lampiran F	104
Lampiran G	105
Lampiran H	106
Lampiran I	107
Lampiran J	108
Lampiran K	109
Lampiran L	113
Lampiran M	116

DAFTAR TABEL

Tabel 2.1 Contoh pasangan <i>attribute-value</i>	11
Tabel 2.2 Tabel acuan <i>attribute-value</i> contoh <i>clustering</i>	12
Tabel 2.3 Tabel acuan <i>attribute-value</i>	22
Tabel 4.1 Hasil akhir <i>array pointer</i>	65
Tabel 4.2 <i>Node</i> yang dimasukan 2 <i>instance</i> yaitu [3][1][2][5] dan [3][5][6][2]... ..	68
Tabel 5.1 Isi <i>file</i> data_kecil. <i>arff</i>	85
Tabel 5.2 Tabel acuan untuk “data_kecil. <i>txt</i> ” dan “data_kecil. <i>arff</i> ”	87
Tabel 5.3 Isi <i>file</i> data_kecil. <i>txt</i>	87
Tabel 5.4 Rata-rata <i>PU dataset</i> A_1, A_2, A_3, A_4, A_5	88
Tabel 5.5 Rata-rata <i>PU dataset</i> $A_6, A_7, A_8, A_9, A_{10}$	88
Tabel 5.6 Perbandingan nilai-nilai <i>PU cluster</i> pada uji coba <i>Cobweb</i> dan <i>Iterate</i>	89
Tabel 5.7 Hasil ujicoba WEKA dan <i>Cobweb</i> dengan jumlah data 1000.	91
Table 5.8 Hasil ujicoba WEKA dengan <i>Cobweb</i> dengan nilai <i>cutoff</i> 0.	92
Tabel 5.9 Hasil uji coba <i>dataset Cobweb</i> dan <i>Iterate</i>	93

DAFTAR GAMBAR

Gambar 2.1. Konsep pada algoritma <i>Cobweb</i> berupa <i>node</i>	12
Gambar 2.2. Probabilitas <i>node A</i> yang bukan <i>node</i> daun (<i>leaf</i>) dengan yang <i>node</i> daun (<i>leaf</i>).	15
Gambar 2.3 Proses memasukan obyek sementara untuk memilih <i>node</i> mana yang terbaik.	18
Gambar 2.4 Proses memasukan obyek ke dalam <i>node</i> baru yaitu <i>node n+1</i>	19
Gambar 2.5 Proses menggabungkan 2 buah <i>node</i> menjadi 1 <i>node</i> (<i>node z</i>).	20
Gambar 2.6 Proses <i>split node</i> terbaik, <i>node z</i> , menjadi beberapa <i>node</i>	20
Gambar 2.7 <i>Root</i> mengalami proses <i>expanded leaf</i>	23
Gambar 2.8 Hasil pemasukan obyek ikan paus [2,1,2].	25
Gambar 2.9 Proses pemasukan obyek burung hantu [2,3,2].	27
Gambar 2.10 Hasil pemasukan obyek ikan piranha [1,1,2].	29
Gambar 2.11 Manhattan distance antara dua buah titik x dan y adalah 7.	32
Gambar 2.12 Grafik nilai <i>CU</i> sepanjang penelusuran <i>node</i> di tiap levelnya.	35
Gambar 2.13 <i>PU root</i> jika obyek manusia [2,2,3] dimasukan pada <i>node A</i> dan <i>node</i> baru.	39
Gambar 2.14 Hasil akhir iterasi <i>node root</i>	39
Gambar 2.15 Hasil akhir langkah pertama.	42
Gambar 2.16 Ilustrasi langkah kedua dan ketiga <i>Iterate</i>	43
Gambar 2.17 Contoh <i>file</i> format <i>arff</i>	46
Gambar 3.1 Contoh penyimpanan <i>instance</i>	51
Gambar 3.2 Contoh <i>instance-instance</i> yang disimpan <i>node</i>	52
Gambar 3.3 <i>Node-node</i> yang memiliki anak dan membentuk klasifikasi <i>tree</i>	53
Gambar 3.4. Proses <i>loop</i> yang mungkin terjadi pada operasi <i>merge node</i>	56
Gambar 4.1 Struktur data <i>Cobweb</i>	65
Gambar 4.2 Struktur data <i>NodeCobweb</i>	67
Gambar 4.3 Struktur data <i>AttributeCobweb</i>	67
Gambar 4.4 Struktur data <i>Iterate</i>	74
Gambar 4.5 Struktur data <i>NodeIterate</i>	76
Gambar 4.6 Tampilan sistem.	83
Gambar 5.1 Grafik perbandingan nilai <i>cutoff</i> dengan waktu eksekusi.	92
Gambar 5.2 Grafik perbandingan jumlah data dengan waktu eksekusi dengan nilai <i>cutoff</i> 0.	93

Gambar 5.3 Grafik perbandingan *Cobweb* dengan *Iterate*. 94

