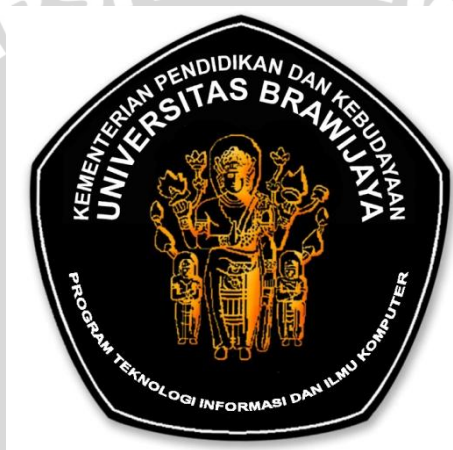


**IMPLEMENTASI ALGORITMA FUZZY C-MEANS CLUSTERING
UNTUK PEMBANGKITAN ATURAN FUZZY PADA PENGELOMPOKAN
TINGKAT RISIKO PENYAKIT KANKER PAYUDARA**

SKRIPSI

Untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer



Disusun oleh :
ELY RATNA SAYEKTI
0910960007

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
MALANG
2014**

LEMBAR PERSETUJUAN

IMPLEMENTASI ALGORITMA *FUZZY C-MEANS CLUSTERING* UNTUK PEMBANGKITAN ATURAN *FUZZY* PADA PENGELOMPOKAN TINGKAT RISIKO PENYAKIT KANKER PAYUDARA

SKRIPSI

Laboratorium Komputasi Cerdas dan Visualisasi

Untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer



Disusun oleh :

ELY RATNA SAYEKTI

NIM. 0910960007

Skrripsi ini telah disetujui oleh dosen pembimbing
pada tanggal 17 Desember 2013

Dosen Pembimbing I,

Dosen Pembimbing II,

Nurul Hidayat, S.Pd., Msc
NIP. 19680430 200212 1 001

Arief Andy Soebroto, ST., M.Kom.
NIP. 19720425 199903 1 002

LEMBAR PENGESAHAN

**IMPLEMENTASI ALGORITMA *FUZZY C-MEANS CLUSTERING*
UNTUK PEMBANGKITAN ATURAN *FUZZY* PADA PENGELOMPOKAN
TINGKAT RISIKO PENYAKIT KANKER PAYUDARA**

SKRIPSI

Laboratorium Komputasi Cerdas dan Visualisasi

Untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer

Disusun oleh :

ELY RATNA SAYEKTI
NIM. 0910960007

Skripsi ini telah diuji dan dinyatakan lulus
pada tanggal 17 Januari 2014

Penguji I

Drs. Marji, MT.
NIP. 19670801 199203 1 001

Penguji II

Rekryan Regasari MP, ST., MT.
NIK. 770414 06 1 2 0253

Penguji III

Imam Cholissodin, S.Si., M.Kom.
NIK. 850719 16 1 1 0422

Mengetahui,
Ketua Program Studi Informatika/ Ilmu Komputer

Drs. Marji, MT.
NIP. 19670801 199203 1 001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Ely Ratna Sayekti
 NIM : 0910960007
 Program Studi : Informatika/ Ilmu Komputer
 Fakultas : Program Teknologi Informasi dan Ilmu Komputer
 Penulis skripsi berjudul : Implementasi Algoritma *Fuzzy C-Means*
Clustering untuk Pembangkitan Aturan *Fuzzy* pada
 Pengelompokan Tingkat Risiko Penyakit Kanker
 Payudara

Dengan ini menyatakan bahwa:

1. Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam skripsi ini.
 2. Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya bersedia menanggung segala resiko yang akan saya terima.
- Demikian pernyataan ini dibuat dengan segala kesadaran dan penuh tanggung jawab dan digunakan sebagaimana mestinya.

Malang, 24 Januari 2014

Penulis,

Ely Ratna Sayekti
 NIM. 0910960007

ABSTRAK

Ely Ratna Sayekti. 2013. Implementasi Algoritma *Fuzzy C-Means Clustering* untuk Pembangkitan Aturan *Fuzzy* pada Pengelompokan Tingkat Risiko Penyakit Kanker Payudara. Skripsi Program Studi Ilmu Komputer, Program Teknologi Informasi Dan Ilmu Komputer Universitas Brawijaya. Pembimbing : Nurul Hidayat, S.Pd., M.Sc. dan Arief Andy Soebroto, ST., M.Kom.

Kanker payudara menduduki posisi pertama sebagai penyebab kematian tertinggi pada wanita akibat kanker di Indonesia. Sebagian besar kanker payudara baru didiagnosis setelah melihat hasil mammogram. Tim dokter menggunakan *computer-aided-diagnosis* untuk mendiagnosa tingkat keganasan kanker payudara guna membantu melakukan keputusan operasi dengan melakukan interpretasi suatu gambar medis. Dari hasil interpretasi tersebut akan diperoleh hasil yang menyatakan apakah kanker yang diderita pasien termasuk jinak atau ganas. Dalam interpretasi sebuah mamografi terkadang dokter mengalami kesalahan dalam menentukan keparahan (jinak dan ganas) sehingga dapat menyebabkan sekitar 70% biopsi yang tidak perlu dilakukan karena kanker itu bersifat jinak. Oleh karena itu diperlukan aturan yang dapat digunakan untuk pengelompokan tingkat risiko kanker payudara. Salah satu teknik yang dapat diterapkan untuk pembangkitan aturan secara otomatis adalah *fuzzy c-means clustering*.

Penelitian ini menggunakan *Mammographic Mass Data Set* yang terdiri dari 500 data latih dan 100 data uji. Proses pelatihan diawali dengan proses *clustering*, dari hasil *clustering* kemudian dilakukan analisa varian. Hasil *cluster* dengan nilai varian terkecil akan dijadikan bahan untuk proses ekstraksi aturan *fuzzy*. Setelah aturan terbentuk barulah dilakukan proses pengujian menggunakan sistem inferensi fuzzy model sugeno orde-satu. Dari pengujian jumlah *cluster* dihasilkan jumlah *cluster* ideal yang terpilih konvergen, yaitu 2 cluster. Sedangkan untuk pengujian akurasi didapatkan akurasi sistem sebesar 78% dengan akurasi tertinggi 84 % pada jumlah aturan 2.

Kata kunci : kanker payudara, mamografi, *fuzzy c-means*, aturan *fuzzy*, fuzzy inference system sugeno, *clustering*.

ABSTRACT

Ely Ratna Sayekti. 2013. *Implementation of Fuzzy C-Means Clustering Algorithm for Fuzzy Rules Extraction in Clustering Breast Cancer Malignancy*. Skripsi Program Studi Ilmu Komputer, Program Teknologi Informasi Dan Ilmu Komputer Universitas Brawijaya. Pembimbing : Nurul Hidayat, S.Pd., MSc dan Arief Andy Soebroto, ST., M.Kom.

Breast cancer is the number one death cause in women by cancer in Indonesia. Most of breast cancer cases can only diagnosed after seeing mammography test. Team of doctors usually use computer-aided-diagnosis to diagnose the malignancy level on breast cancer in order to help in making decisions to operate by interpreting medical graphic. From the result of the interpretation, the team of doctors will then decide whether the patient's cancer is benign or malignant. Sometimes, even doctor can make mistakes in interpreting a mammography test result on the cancer's malignancy, which resulted in 70% of unnecessary biopsy since the cancer itself is benign. Therefore, it is needed to have rules that can be applied to cluster the level of risk on breast cancer. One of many techniques that can be used in automatic fuzzy rule extraction is fuzzy c-means clustering.

This study used Mammographic Mass Data Set that consists of 500 train data and 100 test data. The process of training begin with clustering process, and then variants analysis will begin from then on. The result of the clustering process with the smallest variants value will be used as fuzzy rules extraction material. Testing process using Sugeno first order fuzzy inference system then applied after the system formed fuzzy rules. The ideal number of convergent clusters obtained is 2 clusters. Meanwhile accuracy test resulted in 78% of accuracy and the highest number of accuracy is 84% with 2 rules extracted.

Keyword: breast cancer, mammography, fuzzy c-means, fuzzy rule, sugeno fuzzy inference system, clustering.

KATA PENGANTAR

Syukur *alhamdulillah* penulis panjatkan kehadiran Allah SWT karena berkat rahmat serta hidayah-Nya penulis dapat menyelesaikan skripsi dengan mengambil judul **“IMPLEMENTASI ALGORITMA FUZZY C-MEANS CLUSTERING UNTUK PEMBANGKITAN ATURAN FUZZY PADA PENGELOMPOKAN TINGKAT RISIKO PENYAKIT KANKER PAYUDARA”** sebagai salah satu syarat memperoleh gelar Sarjana Komputer pada program studi Ilmu Komputer/Teknik Informatika PTIIK Universitas Brawijaya - Malang .

Tidak dapat dipungkiri bahwa tidak mungkin penulis dapat menyelesaikan skripsi ini tanpa dukungan serta bantuan dari berbagai pihak. Penulis tidak lupa untuk menyampaikan penghargaan lewat rasa hormat dan terimakasih kepada semua pihak yang telah ikut andil dan berkontribusi, baik dalam bentuk bantuan tenaga, pikiran maupun dukungan moral selama penulisan skripsi sehingga dapat terselesaikannya skripsi ini, yang diantaranya kepada:

1. Ibu Juwanti, Bapak Arifin, Mbah Kakung, Mbah Putri, Kakak dan juga seluruh keluarga tercinta, terima kasih atas semua doa, kasih sayang dan perhatian yang tulus serta dukungan yang telah diberikan.
2. Nurul Hidayat, S.Pd., MSc., selaku pembimbing utama yang telah meluangkan waktu, tenaga, serta pikiran dalam mengarahkan dan membimbing kepada penulis.
3. Arief Andy Soebroto, S.T., M.Kom., selaku pembimbing pendamping yang telah meluangkan waktu, tenaga, serta pikiran dalam mengarahkan dan membimbing kepada penulis.
4. Drs. Marji, MT., selaku Ketua Program Studi Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
5. Bayu Rahayudi, ST., MM., selaku Dosen Penasehat Akademik.
6. Segenap Bapak dan Ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada Penulis selama menempuh pendidikan di Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.

7. Segenap staf dan karyawan di Program Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya yang telah banyak membantu Penulis dalam pelaksanaan penyusunan skripsi ini.
8. Wahyu Aji yang telah mendukung penulis selama masa kuliah.
9. Sahabat-sahabat dan rekan seperjuangan penulis selama kuliah, Deby, Dedy, Inthi, Dila, Arief, Agung, Udin, Zakiya dan Ryno atas segala dukungan dan semua perjuangan selama ini.
10. Rekan-rekan Program Studi Ilmu Komputer A 2009 yang telah berjuang bersama-sama dalam menempuh pendidikan di Universitas Brawijaya.
11. Semua pihak yang telah terlibat baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu

Sebagai manusia biasa, penulis sadar bahwa begitu banyak kekurangan entah itu disengaja ataupun dari kebelummampuan penulis dalam menulis skripsi ini, oleh karena itu saran dan kritik dengan nuansa positif yang bertujuan membangunlah yang sangat dibutuhkan oleh penulis untuk memperbaiki skripsi ini agar kedepannya lebih baik. Penulis berharap skripsi ini dapat bermanfaat bagi penulis sendiri dan bagi semua pihak. Akhirnya penulis berharap agar skripsi ini dapat memberikan sumbangan dan manfaat bagi semua pihak yang berkepentingan.

Malang, Desember 2013

Penulis

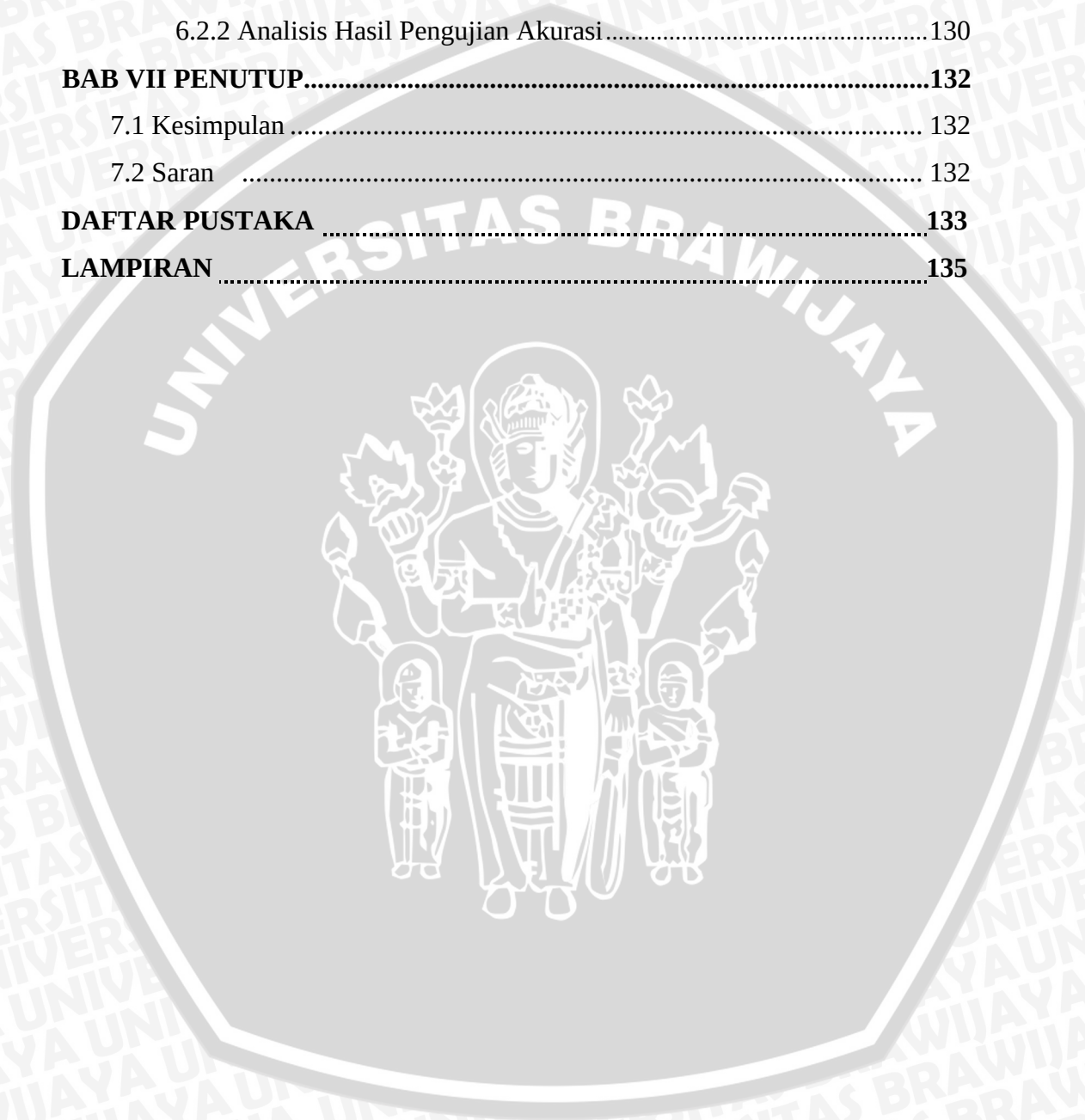
DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Kajian Pustaka	6
2.2 Kanker.....	7
2.2.1 Kanker Payudara.....	8
2.2.2 Mamografi	8
2.3 <i>Mammographic mass data set</i>	10
2.4 <i>Clustering</i>	11
2.4.1 <i>Tipe Clustering</i>	12
2.5 <i>Fuzzy C-Means Clustering (FCM)</i>	12
2.5.1 <i>Algoritma Fuzzy C-Means Clustering (FCM)</i>	13
2.6 <i>Analisa Cluster</i>	14

2.6.1 Analisa Varian	15
2.7 Logika <i>Fuzzy</i>	16
2.7.1 Himpunan <i>Fuzzy</i>	17
2.7.2 Fungsi Keanggotaan	18
2.7.3 Operator Dasar Zadeh.....	19
2.7.4 Fungsi Implikasi	19
2.7.5 <i>Fuzzy Inference System</i>	20
2.7.6 Teknik Pembangkitan Aturan <i>Fuzzy</i>	21
2.7.7 Ekstraksi Aturan <i>Fuzzy</i>	22
2.8 <i>Least Square Estimator</i> (LSE).....	24
2.9 Representasi Linear	25
2.10 Akurasi	27
BAB III METODOLOGI	28
3.1 Studi Literatur	29
3.2 Data Penelitian	29
3.3 Perancangan	29
3.3.1 Deskripsi Sistem Umum.....	29
3.4 Implementasi	31
3.5 Pengujian.....	31
3.6 Pengambilan Kesimpulan	33
3.7 Penulisan Laporan.....	33
BAB IV PERANCANGAN.....	34
4.1 Perancangan Sistem	34
4.1.1 Perancangan Diagram Alir Algoritma.....	36
4.1.1.1 Proses <i>Clustering</i> dengan <i>Fuzzy C-Means Clustering</i>	36
4.1.1.2 Proses Perhitungan Varian.....	45
4.1.1.3 Proses Pemilihan Jumlah <i>Cluster</i> Varian Terkecil	46
4.1.1.4 Proses Ekstraksi Aturan <i>Fuzzy</i> dari <i>Cluster</i>	47
4.1.1.5 Proses Sistem Inferensi <i>Fuzzy</i>	53
4.1.2 Perancangan Kelas Program.....	55
4.1.3 Perancangan Penentuan Kelas Tingkat Risiko Kanker Payudara	57
4.1.4 Perhitungan Manual.....	58

4.1.4.1	Proses Pelatihan	58
4.1.4.2	Proses Pengujian	81
4.2	Perancangan <i>Database</i>	84
4.3	Perancangan Antarmuka	85
4.3.1	Antarmuka Pelatihan 1	85
4.3.2	Antarmuka Pelatihan 2	87
4.3.3	Antarmuka Pengujian	88
4.3.4	Antarmuka Data Latih	89
4.3.5	Antarmuka Data Uji	90
BAB V	IMPLEMENTASI	91
5.1	Spesifikasi Sistem	91
5.1.1	Spesifikasi Perangkat Keras	91
5.1.2	Spesifikasi Perangkat Lunak	92
5.2	Implementasi Algoritma	92
5.2.1	Implementasi Kelas FCM	92
5.2.2	Implementasi Kelas Varian	99
5.2.3	Implementasi Kelas Ekstraksi Aturan <i>Fuzzy</i>	104
5.2.4	Implementasi Kelas FIS Sugeno	107
5.3	Implementasi Antarmuka	111
5.3.1	Implementasi Antarmuka Pelatihan 1	111
5.3.2	Implementasi Antarmuka Pelatihan 2	112
5.3.3	Implementasi Antarmuka Pengujian	114
5.3.4	Implementasi Antarmuka Data Latih	114
5.3.5	Implementasi Antarmuka Data Uji	115
BAB VI	PENGUJIAN DAN ANALISIS	116
6.1	Pengujian	116
6.1.1	Pengujian Jumlah <i>Cluster</i>	117
6.1.1.1	Tujuan Pengujian Jumlah <i>Cluster</i>	117
6.1.1.2	Skenario Pengujian Jumlah <i>Cluster</i>	117
6.1.1.3	Hasil Pengujian Jumlah <i>Cluster</i>	117
6.1.2	Pengujian Akurasi	125
6.1.2.1	Tujuan Pengujian Akurasi	126

6.1.2.2 Skenario Pengujian Akurasi	126
6.1.2.3 Hasil Pengujian Akurasi.....	126
6.2 Analisis	128
6.2.1 Analisis Hasil Pengujian Jumlah <i>Cluster</i>	128
6.2.2 Analisis Hasil Pengujian Akurasi.....	130
BAB VII PENUTUP.....	132
7.1 Kesimpulan	132
7.2 Saran	132
DAFTAR PUSTAKA	133
LAMPIRAN	135



DAFTAR GAMBAR

Gambar 2.1 Karakteristik fungsi gauss	18
Gambar 2.2 Fungsi Implikasi MIN	20
Gambar 2.3 Fungsi Implikasi DOT	20
Gambar 2.4 Representasi Linear Naik	26
Gambar 2.5 Representasi Linear Turun	26
Gambar 3.1 Langkah-Langkah Penelitian	28
Gambar 3.2 Desain Sistem.....	30
Gambar 3.3 Desain Pengujian Jumlah Cluster	32
Gambar 3.4 Desain pengujian Akurasi	32
Gambar 4.1 Diagram Perancangan	34
Gambar 4.2 Gambaran Umum Perancangan Sistem.....	35
Gambar 4.3 Alur proses <i>Fuzzy C-Means Clustering</i>	37
Gambar 4.4 Alur proses pembentukan matriks partisi awal U	39
Gambar 4.5 Alur proses perhitungan pusat cluster	40
Gambar 4.6 Alur proses perhitungan fungsi objektif.....	41
Gambar 4.7 Alur proses perubahan matriks partisi U.....	43
Gambar 4.8 Alur proses pengelompokan data	44
Gambar 4.9 Alur proses perhitungan varian	45
Gambar 4.10 Alur proses pemilihan jumlah <i>cluster</i> dengan varian terkecil..	47
Gambar 4.11 Alur proses ekstraksi aturan <i>fuzzy</i>	48
Gambar 4.12 Alur proses perhitungan perhitungan standar deviasi	49
Gambar 4.13 Alur proses perhitungan koefisien <i>output</i>	50
Gambar 4.14 Alur proses normalisasi.....	51
Gambar 4.15 Alur proses pembentukan matriks U	52
Gambar 4.16 Alur proses perhitungan LSE	53
Gambar 4.17 Alur proses sistem inferensi <i>fuzzy</i>	54
Gambar 4.18 Derajat Keanggotaan terhadap Kelas Jinak dan Ganas.....	57
Gambar 4.19 <i>Physical data model</i>	85
Gambar 4.20 Rancangan antarmuka pelatihan 1	86
Gambar 4.21 Rancangan antarmuka pelatihan 2	87

Gambar 4.22 Rancangan antarmuka pengujian	88
Gambar 4.23 Rancangan antarmuka data latih	89
Gambar 4.24 Rancangan antarmuka data uji	90
Gambar 5.1 Diagram Implementasi	91
Gambar 5.2 Implementasi baca data latih	95
Gambar 5.3 Implementasi pembentukan matriks partisi awal	96
Gambar 5.4 Implementasi penentuan pusat <i>cluster</i>	96
Gambar 5.5 Implementasi perhitungan fungsi objektif	97
Gambar 5.6 Implementasi perhitungan perubahan matriks partisi	98
Gambar 5.7 Implementasi pengecekan kondisi berhenti	98
Gambar 5.8 Implementasi pengelompokan data	99
Gambar 5.9 Implementasi perhitungan varian tiap <i>cluster</i>	102
Gambar 5.10 Implementasi perhitungan varian <i>within cluster</i>	102
Gambar 5.11 Implementasi perhitungan varian <i>between cluster</i>	103
Gambar 5.12 Implementasi perhitungan batasan varian	103
Gambar 5.13 Implementasi perhitungan nilai derajat keanggotaan	104
Gambar 5.14 Implementasi perhitungan koefisien <i>output</i>	107
Gambar 5.15 Implementasi Baca Data Uji	108
Gambar 5.16 Implementasi inisialisasi pusat <i>cluster</i>	109
Gambar 5.17 Implementasi perhitungan derajat keanggotaan	109
Gambar 5.18 Implementasi perhitungan <i>alpha predikat</i>	110
Gambar 5.19 Implementasi perhitungan nilai Z	110
Gambar 5.20 Implementasi proses <i>defuzzy</i>	111
Gambar 5.21 Implementasi antarmuka pelatihan 1	112
Gambar 5.22 Implementasi antarmuka pelatihan 2	113
Gambar 5.23 Implementasi antarmuka pengujian	114
Gambar 5.24 Implementasi antarmuka data latih	115
Gambar 5.25 Implementasi antarmuka data uji	115
Gambar 6.1 Diagram pengujian dan analisis	116
Gambar 6.2 Grafik jumlah <i>cluster</i> ideal yang terpilih	129
Gambar 6.3 Grafik pengaruh jumlah <i>cluster</i> terhadap akurasi	130
Gambar 6.4 Grafik Akurasi Aturan	131

DAFTAR TABEL

Tabel 3.1 Atribut dan kelas <i>Mammography Mass Data Set</i>	29
Tabel 4.1 Perancangan kelas pembentuk aplikasi	55
Tabel 4.2 Data latih	58
Tabel 4.3 Bilangan random	60
Tabel 4.4 Nilai Qi	60
Tabel 4.5 Pusat cluster iterasi 1	62
Tabel 4.6 Fungsi objektif	63
Tabel 4.7 Perhitungan matriks partisi U	64
Tabel 4.8 Pusat <i>cluster</i> iterasi 10	65
Tabel 4.9 Fungsi objektif iterasi 10	66
Tabel 4.10 Kecenderungan data terhadap <i>cluster</i>	67
Tabel 4.11 Data anggota <i>cluster</i> 1	67
Tabel 4.12 Data anggota <i>cluster</i> 2	68
Tabel 4.13 Rata-rata	69
Tabel 4.14 Standar Deviasi	71
Tabel 4.15 Derajat keanggotaan	71
Tabel 4.16 Koefisien <i>output</i>	80
Tabel 4.17 Derajat keanggotaan aturan <i>fuzzy</i>	82
Tabel 5.1 Method-method pembentuk kelas FCM.java	92
Tabel 5.2 Method-method pembentuk kelas varian.java	99
Tabel 6.1 Nilai batasan varian uji coba 1	118
Tabel 6.2 Nilai batasan varian uji coba 2	118
Tabel 6.3 Nilai batasan varian uji coba 3	119
Tabel 6.4 Nilai batasan varian uji coba 4	119
Tabel 6.5 Nilai batasan varian uji coba 5	120
Tabel 6.6 Nilai batasan varian uji coba 6	121
Tabel 6.7 Nilai batasan varian uji coba 7	121
Tabel 6.8 Nilai batasan varian uji coba 8	122
Tabel 6.9 Nilai batasan varian uji coba 9	122
Tabel 6.10 Hasil pengujian akurasi jumlah <i>cluster</i> terhadap akurasi	123

Tabel 6.11 Hasil pengujian akurasi berdasarkan aturan *fuzzy* 126



DAFTAR LAMPIRAN

Lampiran 1 Data Latih	135
Lampiran 2 Data Uji	148

