

**IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK  
PEMBANGKITAN ATURAN FUZZY PADA KLASIFIKASI  
KETAHANAN HIDUP PENDERITA KANKER PAYUDARA**

**SKRIPSI**

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh :

**Khoirul Sholeh**

**0910960045**

**PROGRAM STUDI ILMU KOMPUTER  
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER  
UNIVERSITAS BRAWIJAYA**

**MALANG**

**2013**

LEMBAR PERSETUJUAN

**IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK  
PEMBANGKITAN ATURAN FUZZY PADA KLASIFIKASI  
KETAHANAN HIDUP PENDERITA KANKER PAYUDARA**

**SKRIPSI**

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun oleh :

**Khoirul Sholeh**

**0910960045**

Skripsi ini telah disetujui oleh dosen pembimbing  
pada tanggal 19 Maret 2013

**Dosen Pembimbing I,**

**Dosen Pembimbing II,**

**Budi Darma Setiawan, S.Kom., M.Cs.**  
NIK. 841015 06 1 1 0090

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

LEMBAR PENGESAHAN

**IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK  
PEMBANGKITAN ATURAN FUZZY PADA KLASIFIKASI  
KETAHANAN HIDUP PENDERITA KANKER PAYUDARA**

**SKRIPSI**

**KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI**

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun oleh :

**Khoirul Sholeh**  
**0910960045**

Skripsi ini telah diuji dan dinyatakan lulus  
pada tanggal 19 April 2013

**Penguji I**

**Suprpto, ST., MT.**  
**NIP. 19710727 196603 1 001**

**Penguji II**

**Indriati, ST., M.Kom.**  
**NIK. 831013 06 1 2 0035**

**Penguji III**

**Lailil Muflikhah, S.Kom., M.Sc.**  
**NIP.19741113 200501 2 001**

**Mengetahui**  
**Ketua Program Studi Ilmu Komputer**

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

## KATA PENGANTAR

Assalamualaikum wr. wb. Alhamdulillah, puji syukur kehadirat Allah SWT yang telah memberikan rahmat serta hidayah-Nya penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk menyelesaikan studi di Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya secara tepat waktu.

Dalam penulisan skripsi ini, tentunya banyak pihak yang telah memberikan bantuan baik moril maupun materil, petunjuk, dan bimbingan. Oleh karena itu, penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada :

1. Kedua orang tua penulis Bapak Mokhammad Akhsanun dan Ibu Romelah Idhayanti yang tidak pernah lelah untuk mendidik, mengajarkan, membantu, menasehati, merawat, mendorong penulis untuk bekerja keras dan yang selalu mendoakan agar menjadi orang yang baik, sholeh, sukses dan menjadi orang yang bermanfaat untuk Negara.
2. Bapak Budi Darma Setiawan, S.Kom., M.Cs selaku Dosen Pembimbing I yang telah memberikan bimbingan, arahan, dan nasehat untuk kesempurnaan tugas akhir ini.
3. Bapak Imam Cholissodin, S.Si., M.Kom selaku Dosen Pembimbing II yang telah memberikan bimbingan, arahan, dan nasehat untuk kesempurnaan tugas akhir ini.
4. Semua teman-teman di Ilmu Komputer angkatan 2009, teman-teman di Jurusan Matematika, teman-teman di PTIIK khususnya teman-teman Ilkom A'09 terimakasih atas kerjasamanya selama ini.
5. Dan semua pihak dan teman-teman yang tidak bisa penulis sebutkan satu-persatu terimakasih atas bantuan, dorongan, masukan dan do'a sehingga tugas akhir ini bisa diselesaikan oleh penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka saran dan kritik yang konstruktif dari semua pihak sangat diharapkan demi

penyempurnaan selanjutnya dan semoga skripsi ini bisa memberi manfaat untuk penelitian selanjutnya.

Akhirnya hanya kepada Allah SWT kita kembalikan semua urusan dan semoga skripsi ini dapat bermanfaat bagi semua pihak, khususnya bagi penulis dan para pembaca pada umumnya, semoga Allah SWT meridhloi semua usaha penulis dan dicatat sebagai ibadah disisi-Nya, amin.

Malang, 1 April 2013

Penulis



**LEMBAR PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : Khoirul Sholeh  
NIM : 0910960045  
Program Studi : Ilmu Komputer  
Jurusan : Ilmu Komputer  
Fakultas : Program Teknologi Informasi dan Ilmu Komputer  
Penulis skripsi berjudul : IMPLEMENTASI METODE K-MEANS

CLUSTERING UNTUK PEMBANGKITAN  
ATURAN FUZZY PADA KLASIFIKASI  
KETAHANAN HIDUP PENDERITA  
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, 1 April 2013

Yang menyatakan,

Khoirul Sholeh

NIM.0910960045

## IMPLEMENTASI METODE K-MEANS *CLUSTERING* UNTUK PEMBANGKITAN ATURAN *FUZZY* PADA KLASIFIKASI KETAHANAN HIDUP PENDERITA KANKER PAYUDARA

### ABSTRAK

K-Means *clustering* merupakan salah satu metode pengelompokan yang paling sering digunakan diantara algoritma *clustering* lainnya karena kesederhanaan dan efisiensinya. Metode ini digunakan untuk mengelompokkan data ketahanan hidup penderita kanker payudara. Data ketahanan hidup penderita kanker payudara yang digunakan yaitu data Haberman's Survival. Proses *clustering* data juga bertujuan untuk membangkitkan aturan sebagai rule pengganti yang didefinisikan oleh seorang pakar. Hasil dari pembangkitan aturan, nantinya akan digunakan untuk proses pengujian klasifikasi data dari data ketahanan hidup penderita kanker payudara. Dari proses pengujian bertujuan untuk mengetahui seberapa besar nilai akurasi hasil pengujian dengan menggunakan aturan hasil pembangkitan dengan metode K-Means *clustering*.

Pada penelitian ini secara umum terdapat dua proses utama yaitu proses pelatihan dan proses pengujian. Proses pelatihan sendiri dilakukan dimulai dari proses *clustering*, perhitungan mean dan standar deviasi, analisa varian, sampai proses ekstraksi aturan *fuzzy*. Sedangkan untuk proses pengujian dilakukan proses pengujian klasifikasi data menggunakan *Fuzzy Inference System Sugeno* ordo 1. Pada tahap proses pengujian, dilakukan pengujian data sebanyak 30 data untuk setiap skenario pengujian di masing-masing jumlah *cluster*. Dari hasil pengujian, dilakukan perhitungan akurasi untuk setiap skenario uji untuk setiap jumlah *cluster*. Dari seluruh skenario uji, dicari rata-rata akurasi dari seluruh skenario percobaan untuk setiap jumlah *cluster*. Dari skenario global ini dihasilkan seberapa besar nilai akurasi yang diperoleh dari masing-masing jumlah aturan yang dibangkitkan. Dari hasil penelitian menunjukkan bahwa, dari jumlah *cluster* yang diuji yaitu *cluster* 2, 3, 4, 6, 7, 8, dan 9. Hasil akurasi tertinggi diperoleh pada pengujian dengan jumlah *cluster* 7 sebesar 83%. Dan menghasilkan nilai akurasi terendah pada pengujian dengan jumlah *cluster* 3 sebesar 66%. Bisa disimpulkan untuk penggunaan algoritma K-Means *clustering* sudah memiliki hasil yang maksimal dan bisa dikatakan metode ini berhasil dalam membangkitkan aturan *fuzzy*.

**Kata kunci** : penderita kanker payudara, aturan *fuzzy*, *fuzzy inference system Sugeno*, K-Means, *clustering*

repository.ub.ac.id

## IMPLEMENTATION METHOD K-MEANS CLUSTERING FOR FUZZY RULES GENERATION SECURITY CLASSIFICATION OF LIVING IN BREAST CANCER PATIENTS

### ABSTRACT

K-Means clustering is a method of grouping the most commonly used among other clustering algorithms because they simplicity and efficiency. This method is used to classify the data survival of breast cancer patients. Data survival of breast cancer patients who used the Haberman's Survival Data. Data clustering process also aims to generate rules as a replacement rule defined by an expert. The results from the generation of the rules, will be used to process the data from the data classification testing survival of breast cancer patients. The testing process aims to find out how much value the accuracy of the test results by using a rule generation results with K-Means clustering method.

In this study, in general, there are two main processes, namely the process of training and testing process. The training process itself is done starting from the clustering process, the calculation of the mean and standard deviation, variance analysis, to the extraction of fuzzy rules. As for the testing process performed the testing process of data classification using Fuzzy Inference System Sugeno order 1. At this stage of the testing process, testing the data as much as 30 data points for each test scenario in each cluster number. From the test results, the calculation accuracy for each test scenario for each number of clusters. From all the test scenarios, sought an average accuracy of all experimental scenarios for each number of clusters. From all the test scenarios, sought an average accuracy of all experimental scenarios for each number of clusters. From this global scenario resulting accuracy how much the value obtained from each of the number of rules generated. The results showed that, of the number of clusters that were tested, namely cluster 2, 3, 4, 5, 6, 7, 8, and 9. Highest accuracy results obtained in testing the number of cluster 7 by 83%. And produced the lowest accuracy on the test values by the number of cluster 3 by 66%. Can be inferred for the use of K-Means clustering algorithm already has the maximum results and can say this method is successful in generating fuzzy.

**Keywords:** breast cancer, fuzzy rule, fuzzy inference system of Sugeno, K-Means, clustering



DAFTAR ISI

	halaman
<b>LEMBAR PERSETUJUAN</b> .....	ii
<b>LEMBAR PENGESAHAN</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>LEMBAR PERNYATAAN</b> .....	vi
<b>ABSTRAK</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xv
<b>KODE PROGRAM</b> .....	xvii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan .....	4
1.5 Manfaat .....	5
1.6 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 Kanker .....	7
2.2 Kanker Payudara .....	10
2.3 Ketahanan Hidup Penderita Kanker Payudara .....	12
2.4 Data Haberman's Survival .....	15
2.5 Data Mining .....	15
2.5.1 Definisi Data Mining .....	15
2.5.2 Tahapan Data Mining .....	17
2.5.3 Teknik Data Mining .....	19
2.5.3.1 <i>Clustering</i> .....	21
2.5.3.2 Tipe <i>Clustering</i> .....	22
2.5.4 K-Means <i>Clustering</i> .....	23



2.5.4.1	Pengertian K-Means <i>Clustering</i> .....	23
2.5.4.2	Algoritma K-Means <i>Clustering</i> .....	24
2.5.4.3	Analisa <i>Cluster</i> .....	26
2.5.4.4	Analisa Varian.....	27
2.5.4.5	<i>Least Square Estimator</i> (LSE).....	28
2.6	Logika <i>Fuzzy</i> .....	31
2.7	Himpunan <i>Fuzzy</i> .....	33
2.8	Fungsi Keanggotaan.....	34
2.9	Penalaran Monoton.....	37
2.10	Fungsi Implikasi.....	38
2.11	<i>Fuzzy Inference System</i> .....	39
2.12	Aturan Dasar <i>Fuzzy</i> .....	41
2.13	Teknik Pembangkitan Aturan <i>Fuzzy</i> .....	42
2.14	Ekstraksi Aturan <i>Fuzzy</i> dari <i>Cluster</i> .....	43
2.15	Perhitungan Akurasi.....	45
<b>BAB III METODOLOGI DAN PERANCANGAN SISTEM</b> .....		46
3.1	Studi Literatur.....	47
3.2	Data Penelitian.....	48
3.3	Analisa dan Perancangan Sistem.....	48
3.3.1	Deskripsi Umum Sistem.....	48
3.3.2	Perancangan Proses.....	50
3.3.2.1	Proses Pelatihan.....	52
a)	Proses <i>Clustering</i> .....	52
b)	Analisis Varian.....	54
c)	Memilih Jumlah <i>Cluster</i> dengan Varian Terkecil.....	55
d)	Pembangkitan Aturan.....	56
e)	Menghitung Mean dan Standar Deviasi.....	57
f)	Menghitung Derajat Keanggotaan.....	59
g)	Perhitungan Koefisien Output.....	60
h)	Proses Membentuk Matriks U.....	61
i)	Proses Normalisasi.....	62
j)	Proses Perhitungan LSE.....	63



3.3.2.2	Proses Pengujian .....	64
a)	Sistem Inferensi <i>Fuzzy Model Sugeno</i> .....	64
b)	Proses Fuzzifikasi .....	65
c)	Proses Perhitungan Nilai Z.....	66
d)	Proses Defuzzifikasi.....	67
3.3.3	Perancangan Tabel .....	68
3.3.4	Perancangan Antarmuka .....	71
3.3.5	Perhitungan Manual .....	77
3.3.5.1	Proses Perhitungan K-Means <i>Clustering</i> .....	77
3.3.5.2	Analisis Varian.....	85
3.3.5.3	Ekstraksi Aturan <i>Fuzzy</i> .....	86
3.3.5.4	Perhitungan Fungsi Keanggotaan .....	87
3.3.5.5	Menghitung Koefisien Output .....	88
3.3.5.6	Proses Uji Data.....	91
a)	Proses Fuzzifikasi.....	91
b)	Proses Perhitungan <i>Fire Strengh</i> ( $\alpha$ -predikat) .....	92
c)	Proses Perhitungan Nilai Z Setiap Aturan .....	93
d)	Proses Defuzzifikasi.....	93
3.3.6	Sistematika Pengujian .....	93
3.3.6.1	Uji Jumlah <i>Cluster</i> Ideal .....	93
3.3.6.2	Uji Tingkat Akurasi.....	94
<b>BAB IV</b>	<b>IMPLEMENTASI</b> .....	<b>96</b>
4.1	Lingkungan Implementasi .....	96
4.1.1	Lingkungan Implementasi Perangkat Keras.....	96
4.1.2	Lingkungan Implementasi Perangkat Lunak.....	96
4.2	Implementasi Program.....	97
4.2.1	Tahapan Proses Pelatihan .....	97
4.2.1.1	Proses Input Data .....	97
4.2.1.2	Proses K-Means <i>Clustering</i> .....	98
4.2.1.3	Proses Perhitungan Mean dan Standar Deviasi.....	102
4.2.1.4	Proses Analisis Varian .....	104
4.2.1.5	Proses Perhitungan Varian <i>Cluster</i> .....	105



4.2.1.6	Proses Perhitungan Derajat Keanggotaan .....	106
4.2.1.7	Proses Pembentukan Matriks U .....	107
4.2.2	Tahapan Proses Pengujian .....	109
4.2.2.1	Proses Fuzzifikasi .....	109
4.2.2.2	Proses Perhitungan Z Untuk Setiap Aturan .....	110
4.2.2.3	Proses Defuzzifikasi .....	111
4.3	Implementasi Antarmuka .....	111
4.3.1	Form Utama .....	112
4.3.2	Form Data Haberman's Survival .....	116
4.3.3	Form Data Uji .....	116
4.4	Implementasi Uji Coba .....	117
4.4.1	Skenario Uji Coba .....	117
4.4.2	Hasil Uji Coba .....	119
4.4.2.1	Uji Coba 1 .....	119
4.4.2.2	Uji Coba 2 .....	121
4.4.2.3	Uji Coba 3 .....	124
4.4.2.4	Uji Coba 4 .....	126
4.4.2.5	Uji Coba 5 .....	128
<b>BAB V ANALISA DAN PEMBAHASAN .....</b>		<b>131</b>
<b>BAB VI PENUTUP .....</b>		<b>138</b>
6.1	Kesimpulan .....	138
6.2	Saran .....	138
<b>DAFTAR PUSTAKA .....</b>		<b>140</b>

## DAFTAR GAMBAR

	halaman
Gambar 2.1 Tahap-Tahap Data Mining atau KDD	17
Gambar 2.2 Contoh <i>Data Matrix Object by Variable Structure</i>	27
Gambar 2.3 Contoh <i>Data Matrix Object by Object Structure</i>	27
Gambar 2.4 Pemetaan Suatu Input-Output	31
Gambar 2.5 Kurva Segitiga	35
Gambar 2.6 Kurva Trapesium	36
Gambar 2.7 Kurva Gaussian	36
Gambar 2.8 Kurva Lonceng	37
Gambar 2.9 Kurva Sigmoidal	37
Gambar 2.10 Fungsi Implikasi Min	39
Gambar 2.11 Fungsi Implikasi Dot	39
Gambar 2.12 Inferensi dengan Menggunakan Metode Tsukamoto	40
Gambar 3.1 Alur Penelitian	47
Gambar 3.2 Alur Perancangan Sistem	50
Gambar 3.3 Alur Proses Pengelompokan Data dengan <i>K-Means Clustering</i>	53
Gambar 3.4 Alur Proses Batasan Varian	54
Gambar 3.5 Alur Proses Pemilihan <i>Cluster</i> dengan Varian Terkecil	55
Gambar 3.6 Alur Proses Pembangkitan Aturan <i>Fuzzy</i>	57
Gambar 3.7 Alur Proses Menghitung Mean dan Standar Deviasi	58
Gambar 3.8 Alur Proses Menghitung Derajat Keanggotaan ( $\mu$ )	59
Gambar 3.9 Alur Proses Menghitung Koefisien Output	60
Gambar 3.10 Alur Proses Pembentukan Matriks U	61
Gambar 3.11 Alur Proses Pembentukan Matriks U Ternormalisasi	62
Gambar 3.12 Alur Proses Menghitung LSE	63
Gambar 3.13 Alur Uji Data Menggunakan FIS Sugeno	64
Gambar 3.14 Alur Proses Fuzzifikasi	65
Gambar 3.15 Alur Proses Perhitungan Z Masing-Masing Cluster	66
Gambar 3.16 Alur Proses Menghitung Nilai Kategori Ketahanan Penderita Kanker Payudara	67

Gambar 3.17 Conceptual Data Model	68
Gambar 3.18 Physical Data Model	68
Gambar 3.19 Tampilan Antarmuka Sistem	72
Gambar 3.20 Tampilan Bagian-Bagian Proses Sistem	73
Gambar 3.21 Tampilan Bagian-Bagian Proses Pelatihan	74
Gambar 3.22 Tampilan Bagian-Bagian Proses Pengujian Data	75
Gambar 3.23 Tampilan Form Data Haberman	76
Gambar 3.24 Tampilan Form Data Uji	76
Gambar 4.1 Form Utama	112
Gambar 4.2 Tampilan Form Utama Setelah Proses Pelatihan Dilakukan	113
Gambar 4.3 Form Data Uji Haberman's Survival Setelah Tombol Load Ditekan	114
Gambar 4.4 Hasil Proses Inputan Menggunakan Form Data Haberman's	115
Gambar 4.5 Hasil Proses Pengujian Data Haberman's Survival	115
Gambar 4.6 Form Data Haberman's Survival	116
Gambar 4.7 Form Data Uji	117
Gambar 4.8 Grafik Nilai Varian Pada Proses Pengelompokkan	119
Gambar 4.9 Grafik Nilai Perhitungan Akurasi Percobaan ke-1	121
Gambar 4.10 Grafik Nilai Perhitungan Akurasi Percobaan ke-2	123
Gambar 4.11 Grafik Nilai Perhitungan Akurasi Percobaan ke-3	126
Gambar 4.12 Grafik Nilai Perhitungan Akurasi Percobaan ke-4	128
Gambar 4.13 Grafik Nilai Perhitungan Akurasi Percobaan ke-5	130
Gambar 5.1 Grafik Hasil Pehitungan Akurasi Untuk Setiap Cluster	136

## DAFTAR TABEL

	halaman
Tabel 3.1 Perancangan Tabel tb_datahaberman	69
Tabel 3.2 Perancangan Tabel tb_kategori	70
Tabel 3.3 Perancangan Tabel tb_datatraining	70
Tabel 3.4 Perancangan Tabel tb_centroidawal	71
Tabel 3.5 Contoh Data Penderita Kanker Payudara Pada Data Haberman's Survival	78
Tabel 3.6 Data Hasil Perhitungan Jarak Data Training dengan Centroid Awal	79
Tabel 3.7 Data Hasil Perhitungan Jarak Pada Iterasi Ke-2	80
Tabel 3.8 Data Hasil Perhitungan Jarak Pada Iterasi Ke-3	82
Tabel 3.9 Data Hasil Perhitungan Jarak Pada Iterasi Ke-4	83
Tabel 3.10 Data Hasil Perhitungan Jarak Pada Iterasi Ke-5	84
Tabel 3.11 Nilai Rata-Rata Setiap Atribut Pada Centroid Akhir Data Training	85
Tabel 3.12 Hasil Proses Perhitungan Fungsi Keanggotaan	88
Tabel 3.13 Hasil Perhitungan Derajat Keanggotaan Data Uji	92
Tabel 3.14 Tabel Data Hasil Perhitungan Batasan Varian	94
Tabel 3.15 Tabel Data Hasil Proses Perhitungan Akurasi Pengujian Data	95
Tabel 4.1 Hasil Perhitungan Varian Masing-Masing <i>Cluster</i>	118
Tabel 4.2 Hasil Pengujian 30 Data Pertama	119
Tabel 4.3 Hasil Perhitungan Kategori Percobaan ke-1	121
Tabel 4.4 Hasil Pengujian 30 Data Kedua	122
Tabel 4.5 Hasil Perhitungan Kategori Percobaan ke-2	123
Tabel 4.6 Hasil Pengujian 30 Data Ketiga	124
Tabel 4.7 Hasil Perhitungan Kategori Percobaan ke-3	125
Tabel 4.8 Hasil Pengujian 30 Data Keempat	126
Tabel 4.9 Hasil Perhitungan Kategori Percobaan ke-4	127
Tabel 4.10 Hasil Pengujian 30 Data Kelima	128
Tabel 4.11 Hasil Perhitungan Kategori Percobaan ke-5	130
Tabel 5.1 Hasil Perhitungan Varian Masing-Masing <i>Cluster</i>	131
Tabel 5.2 Hasil Perhitungan Kategori Percobaan Untuk Seluruh Skenario	132





## Kode Program

	halaman
Kode Program 4.1 Method Input Data	98
Kode Program 4.2 Proses K-Means <i>Clustering</i>	100
Kode Program 4.3 Proses Perhitungan Jarak (Euclidean)	101
Kode Program 4.4 Proses Perhitungan Centroid Baru	102
Kode Program 4.5 Proses Perhitungan Mean dan Standar Deviasi	103
Kode Program 4.6 Analisis Varian	105
Kode Program 4.7 Proses Perhitungan Varian	106
Kode Program 4.8 Proses Perhitungan Derajat Keanggotaan	106
Kode Program 4.9 Proses Perhitungan Matriks U	107
Kode Program 4.10 Proses <i>Least Square Estimator</i>	109
Kode Program 4.11 Proses Fuzzifikasi	109
Kode Program 4.12 Proses Perhitungan $\alpha$ -Predikat	110
Kode Program 4.13 Proses Perhitungan Z Untuk Setiap Aturan	110
Kode Program 4.14 Proses Defuzzifikasi	111

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Di Negara-negara yang telah maju pola penyakit telah beralih dari penyakit infeksi dan penyakit menular ke arah penyakit yang sifatnya menahun, seperti penyakit jiwa, penyakit akibat kecelakaan, penyakit kardio-vaskuler dan penyakit degeneratif seperti arthrosis deformans, dan penyakit kanker. Kanker merupakan kematian no. 2 setelah penyakit jantung [KOS-82].

Kanker merupakan suatu penyakit dimana sel telah kehilangan pengendalian dan mekanisme normalnya, sehingga mengalami pertumbuhan yang tidak normal, cepat dan tidak terkendali. Sel-sel kanker akan terus membelah diri, dan tidak mengindahkan kaidah hukum-hukum pembiakan. Kanker bisa terjadi dari berbagai jaringan dalam berbagai organ, seperti sel kulit, sel hati, sel darah, sel otak, sel lambung, sel usus, sel paru, sel saluran kencing, dan berbagai macam sel tubuh lainnya. Kanker merupakan jenis penyakit yang paling membunuh di dunia. Salah satu jenis penyakit kanker yang berbahaya yaitu kanker payudara. Kanker ini merupakan momok bagi wanita-wanita diseluruh dunia.

Kanker Payudara merupakan jenis kanker yang menjadi salah satu masalah kesehatan pada wanita. Kanker payudara merupakan salah satu penyakit keganasan kedua terbanyak yang sering menyebabkan kematian. Berdasarkan data dari IARC (*International Agency for Research on Cancer*) yang dikutip oleh Pusat Komunikasi Publik Sekjen Depkes tahun 2008, bahwa pada tahun 2002 kanker payudara menempati urutan pertama dari seluruh kanker pada perempuan (*incident rate* 38 per 100.000 perempuan) dengan kasus baru sebesar 22,7% dan jumlah kematian 14% per tahun dari seluruh kanker pada perempuan di dunia [TIO-09].

Dihitung dari sepuluh tahun yang lalu, penderita kanker payudara terus meningkat. Tidak hanya di Indonesia namun di Amerika dan Eropa kanker merupakan pembunuh nomor dua di dunia. Bahkan WHO memprediksi pada tahun 2030 nanti lonjakan penderita kanker di Indonesia sampai tujuh kali lipat. Jumlah penderita kanker yang meninggal juga kian memprihatinkan.

Dengan adanya banyak kasus yang terjadi dan semakin meningkatnya kasus mengenai kanker payudara. Diperkirakan dalam dua puluh tahun terakhir ini, telah terjadi kemajuan yang sangat menyolok dalam penemuan dini, pengobatan dan penelitian yang berhasil pada beberapa tipe keganasan yang di masa lampau selalu dianggap fatal [RUP-02]. Salah satunya yaitu penelitian ketahanan hidup penderita kanker payudara yang dilakukan oleh Tjen-Sien Lim. Data penelitian ketahanan hidup penderita kanker payudara ini disebut dengan data Haberman's Survival, yang merupakan data hasil dari penelitian mengenai ketahanan pasien yang telah menjalani operasi kanker payudara di University of Chicago's Billings Hospital. Data ini disumbangkan oleh Tjen-Sien Lim. Data Haberman's Survival terdiri dari 306 kasus pasien penderita kanker payudara.

Pengobatan penyakit kanker dapat dikatakan berhasil bila penyembuhan yang diperoleh penderita kanker dapat dibuktikan mempunyai harapan hidup (*life expectancy*) yang sama dengan penduduk yang tidak menderita penyakit kanker di mana penderita kanker itu hidup [RUP-02]. Selain itu tolak ukur keberhasilan pengobatan kanker adalah angka ketahanan hidup (*year survival rate*). Harapan hidup pasien kanker payudara dalam lima tahun digambarkan dalam *five-year survival rate*.

Ketahanan hidup seorang penderita kanker payudara pasti berbeda antara penderita yang satu dengan penderita yang lain. Masalah perbedaan ini tentunya ada beberapa faktor yang mempengaruhi daya tahan penderita kanker payudara tersebut. Faktor-faktor tersebut bisa dijadikan sebagai karakteristik penderita kanker payudara. Cara paling mudah dalam mengetahui karakteristik penderita kanker payudara ini yaitu dengan mengelompokkan data penderita berdasarkan klasifikasi atau kelompok tertentu. Namun karena data daya tahan penderita itu tidak pasti, dan tiap kasus itu berbeda-beda, bahkan daya tahan tetap berbeda walaupun dengan kasus yang sama. Maka dari itu sulit untuk mengelompokkan data tersebut secara pasti.

Melihat dari permasalahan tersebut, *fuzzy classification system* memungkinkan untuk mengelompokkan data-data ke dalam *cluster* tertentu sehingga dapat dikelompokkan data-data dengan data yang memiliki kemiripan

data yang sama. Data *clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan atau ketidakmiripan untuk mengelompokkan objek.

Salah satu metode *clustering* ini yaitu metode *K-Means*. Metode ini digunakan untuk mengolah data untuk membentuk aturan-aturan yang nantinya digunakan untuk mengelompokkan dan menguji data daya tahan hidup penderita kanker payudara apakah tergolong dari kategori (bertahan hidup lebih dari lima tahun atau tidak) dengan menggunakan metode *Fuzzy Inference Sugeno*. Dengan metode *Fuzzy Inference System Sugeno* data diuji dengan aturan yang dihasilkan dari pembangkitan aturan dengan metode *K-Means*. Dari pengujian tersebut bisa dilihat keakuratan uji data dengan hasil penelitian pada data *Haberman's Survival*.

Metode algoritma *K-Means Clustering* digunakan sebagai solusi proses pengelompokkan data Haberman's Survival (daya tahan hidup penderita). *K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok. Metode *K-Means clustering* ini mempartisi data ke dalam *cluster-cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama, sebaliknya data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain. Adapun tujuan peng-*cluster*-an ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering* yang pada umumnya digunakan untuk usaha meminimalisasikan variasi di dalam suatu *clustering* dan memaksimalkan variasi antar *cluster*.

Dalam skripsi ini, akan diterapkan metode algoritma *K-Means clustering* yang mana digunakan untuk menyelesaikan permasalahan dalam proses pengelompokkan data Haberman's Survival dan nantinya akan digunakan untuk pembangkitan aturan yang digunakan untuk proses uji data menggunakan *Fuzzy Inference System Sugeno*. Dari proses *clustering*, pembangkitan aturan dan perhitungan data-data yang diuji, diharapkan dapat diketahui keakuratan dari proses *clustering* data-data uji serta seberapa akurat hasil uji data perhitungan menggunakan *Fuzzy Inference System Sugeno* dengan data hasil penelitian pada data *Haberman's Survival*.

## 1.2 Rumusan Masalah

Dari uraian latar belakang yang diuraikan, dapat dibuat suatu rumusan masalah yaitu :

- 1) Bagaimana cara mengimplementasikan metode *K-Means Clustering* untuk membangkitkan aturan *fuzzy* pada proses *clustering* data ketahanan hidup penderita kanker payudara.
- 2) Bagaimana perhitungan akurasi hasil pembangkitan aturan *fuzzy* menggunakan metode *K-Means Clustering* untuk pengujian data ketahanan hidup penderita kanker payudara menggunakan *Fuzzy Inferensi Sistem Sugeno* pada data Haberman's Survival.

## 1.3 Batasan Masalah

Untuk membatasi masalah yang dikaji, maka penulis membatasi permasalahan dan menitik beratkan penelitian terhadap masalah sebagai berikut :

- 1) Data yang dijadikan data latih dan data uji pada skripsi ini merupakan dataset Haberman's Survival yang merupakan hasil penelitian di University Chicago's Billings Hospital oleh Tjen-Sien Lim mengenai Ketahanan Hidup Penderita Kanker Payudara, diambil dari :  
<http://archive.ics.uci.edu/ml/machine-learning-databases/haberman/haberman.data>
- 2) Rule hasil proses pembangkitan aturan *fuzzy* hanya berlaku pada data Haberman's Survival.
- 3) Jika nantinya ingin menguji dengan data kasus yang berbeda, maka faktor-faktor lain (internal maupun eksternal) yang mempengaruhi kondisi penderita diabaikan.
- 4) Tidak ada perbandingan metode lain di skripsi ini.

## 1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian skripsi ini adalah :

- 1) Menerapkan algoritma *K-Means Clustering* untuk pembangkitan aturan *fuzzy* pada proses *clustering* data ketahanan hidup penderita kanker payudara berdasarkan perhitungan kemiripan data pada data Haberman's Survival.
- 2) Untuk mengetahui hasil akurasi uji data dan klasifikasi yang dilakukan dengan menggunakan metode *Fuzzy Inferensi Sistem Sugeno* dengan

pembangkitan aturan *fuzzy* menggunakan metode *K-Means Clustering* terhadap kasus ketahanan hidup penderita kanker payudara menggunakan data Haberman's Survival.

### 1.5 Manfaat

Hasil penulisan dan penelitian skripsi ini diharapkan dapat menghasilkan aplikasi yang dapat digunakan dalam mengklasifikasi suatu kasus ketahanan penderita kanker payudara, serta dapat mengimplementasikan metode *K-Means clustering* sebagai alat untuk membangkitkan aturan secara otomatis sebagai ganti rule yang didefinisikan oleh seorang pakar. Dan harapannya dapat memberikan manfaat bagi para praktisi ilmu komputer dan pendidikan yang mempunyai tujuan dalam pengembangan penelitian mengenai keakuratan dalam perhitungan klasifikasi ketahanan hidup penderita kanker payudara pada data *Haberman's Survival* dengan pembangkitan aturan menggunakan metode *K-Means*. Serta harapan yang diinginkan oleh penulis, skripsi ini dapat bermanfaat sebagai referensi serta bisa dikembangkan menjadi penelitian yang lebih kompleks dan menghasilkan aplikasi yang lebih sempurna. Misalnya perhitungan dan penelitian keakuratan dalam proses *clustering* dengan memperhitungkan faktor ketahanan hidup penderita kanker payudara dari faktor-faktor internal maupun eksternal.

### 1.6 Sistematika Penulisan

Skripsi ini disusun berdasarkan sistematika penulisan sebagai berikut :

#### 1. BAB I PENDAHULUAN

Bab I ini berisi latar belakang masalah, rumusan masalah, tujuan, dan manfaat penelitian, serta sistematika penulisan.

#### 2. BAB II TINJAUAN PUSTAKA

Bab II berisi mengenai uraian-uraian teori-teori yang berhubungan dengan kanker (kanker payudara), ketahanan hidup penderita kanker payudara, *Data Mining*, *Clustering*, metode *K-Means Clustering*, Logika Fuzzy, dan *Fuzzy Inference System Sugeno*.

#### 3. BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab III ini membahas tentang metode penelitian yang dilakukan dalam skripsi serta membahas spesifikasi perangkat lunak, perancangan sistem, perancangan basis data, dan perancangan antar muka.

#### 4. BAB IV IMPLEMENTASI

Bab IV ini berisi tentang implementasi algoritma *K-Means* ke dalam sistem dan *code*.

#### 5. BAB V ANALISA DAN PEMBAHASAN

Bab V ini membahas tentang perhitungan dalam penggunaan metode yang digunakan serta menjelaskan mengenai tabel data uji dan menjelaskan grafik pengujian perangkat lunak.

#### 6. BAB VI PENUTUP

Bab VI membahas tentang kesimpulan dari keseluruhan rangkaian penelitian serta saran untuk memungkinkan adanya pengembangan lebih lanjut.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kanker

Kanker adalah suatu penyakit, dimana terdapat pertumbuhan sel dan jaringan yang tidak normal, cepat, dan tidak terkendali. Bila pertumbuhan itu tidak segera dihentikan, maka ia akan berkembang terus sehingga akan mengganggu dan menghancurkan jaringan dan organ yang terkena, serta dapat menyebar ke bagian tubuh yang lainnya, yang akhirnya dapat menyebabkan kematian penderitanya [BUD-02].

Sel-sel kanker akan berkembang dengan cepat, tidak terkendali, dan terus membelah diri, selanjutnya menyusup ke jaringan di sekitarnya (*invasive*) dan terus menyebar melalui jaringan ikat, darah, dan menyerang organ-organ penting serta saraf tulang belakang. Dalam keadaan normal, sel hanya akan membelah diri jika ada penggantian sel-sel yang telah mati dan rusak. Sebaliknya, sel kanker akan membelah terus meskipun tubuh tidak memerlukannya, sehingga akan terjadi penumpukan sel baru. Penumpukan sel tersebut akan mendesak dan merusak jaringan normal, sehingga mengganggu organ yang ditempatinya [MAN-09].

Penyebab kanker bervariasi dan tidak dapat diketahui dengan pasti. Kanker terjadi karena kerusakan struktur genetik yang menyebabkan pertumbuhan sel menjadi tidak terkontrol. Pola insiden kanker bervariasi sesuai jenis kelamin, ras, dan letak geografik. Beberapa kanker dapat dipengaruhi faktor genetik keluarga, namun yang paling sering terjadi karena faktor lingkungan dan gaya hidup. Promotor kanker, yang disebut karsinogen seperti bahan kimia, virus serta faktor lingkungan dan gaya hidup [MEN-00 dan DUY-06].

Penyebab kanker biasanya akan dinetralisir langsung oleh tubuh, tapi kadang bisa langsung menyerang DNA, tetapi hal ini memakan waktu lama untuk tumbuh dan berubah menjadi kanker [BUD-02]. Umumnya, sebelum kanker meluas atau merusak jaringan di sekitarnya, penderita tidak merasakan adanya keluhan atau pun gejala, bila sudah ada keluhan atau gejala biasanya penyakit berada pada taraf stadium lanjut. Awalnya kanker tidak menimbulkan keluhan



karena hanya melibatkan beberapa sel. Bila sel kanker bertambah, maka keadaan bergantung kepada orang yang terkena. Misalnya, pada usus berongga besar, tumor harus mencapai ukuran besar sebelum memicu keluhan [HAS-09].

Stadium kanker merupakan keputusan klinis yang berkaitan dengan ukuran tumor, derajat invasi lokal yang telah terjadi dan derajat penyebarannya ke tempat-tempat jauh pada seseorang [COR-01]. Sebuah prediktor secara kuat mengenai kekambuhan penyakit dan lamanya paparan penyakit pasien yang menyerang di tempat itu serta penyebaran ke organ terdekat, seperti tulang, hati, paru-paru dan sistem saraf pusat. Secara klinis untuk menentukan diagnosa perluasan penyakit, maka akan digunakan hasil pemeriksaan klinis pasien, hasil observasi selama intervensi pembedahan, dan hasil laporan patologis [NAP-08].

Penyakit kanker dapat didefinisikan berdasarkan empat karakteristik, yang dapat menjelaskan bagaimana sel kanker berlaku berbeda dengan sel normal sebagai berikut [MEN-00]:

1. Klonalitas : Kanker berasal dari perubahan genetik yang terjadi pada sebuah sel, yang kemudian berproliferasi membentuk sel ganas.
2. Autonomi : Pertumbuhan tidak teratur dengan benar oleh pengaruh biokimia dan fisik normal dalam lingkungan.
3. Anaplasia : Tidak terdapat diferensiasi sel yang normal dan terkoordinasi
4. Metastasis : Sel kanker memiliki kemampuan tumbuh secara tidak kontinyu dan menyebar ke bagian tubuh lain.

Kanker dapat menyebabkan banyak gejala yang berbeda, bergantung pada lokasinya dan karakter dari keganasan dan apakah ada metastasis. Sebuah diagnosis biasanya membutuhkan pemeriksaan mikroskopik jaringan yang diperoleh dengan biopsi. Setelah didiagnosis, pasien kanker biasanya dirawat dengan operasi, kemoterapi dan/atau radiasi. Kebanyakan pasien kanker dapat dirawat dan banyak disembuhkan, terutama bila perawatan dimulai sejak awal. Bila tidak terawat, kebanyakan kanker menyebabkan kematian pada pasien [HAS-09].

Kondisi dan penanganan penyakit kanker dapat menimbulkan stres, sehingga tidak saja mempengaruhi kondisi fisik tetapi juga kondisi psikologis pasien. Meskipun reaksi psikologis terhadap diagnosis penyakit dan penanganan

kanker sangat beragam dan keadaan serta kemampuan masing-masing penderita tergantung pada banyak faktor, namun ada enam reaksi psikologis yang utama yaitu kecemasan, depresi, perasaan kehilangan kontrol, gangguan kognitif atau status mental (*impairment*), gangguan seksual serta penolakan terhadap kenyataan (*denial*). Jay, Elliot & Varni (1986) menyatakan bahwa profil psikologis pasien yang datang pada pemeriksaan medis menunjukkan tingginya tingkat kecemasan, rasa marah, dan keterasingan [HAS-09].

Pada penderita kanker, gejala fisik yang dikeluhkan mungkin tidak hanya akibat kankernya tetapi bisa juga disebabkan karena penderita mengalami depresi. Kemungkinan ini harus diwaspadai terutama bila keluhannya jelas menyimpang atau tidak sebanding dengan penyakit fisiknya, atau saat keluhan muncul penderita juga menghadapi masalah sosial dan pekerjaan. Walaupun demikian, umumnya gejala somatic kurang berguna sebagai patokan diagnosis depresi untuk penderita kanker stadium lanjut, karena gejala tersebut mungkin akibat dari penyakitnya sendiri. Yang lebih sesuai adalah adanya afek/emosi depresi dan perilaku menarik diri dari teman dan keluarga, namun yang paling akurat untuk mengetahui depresi pada penderita kanker stadium lanjut yaitu [KON-08] :

- perasaan tidak ada harapan lagi
- rasa bersalah
- pikiran bunuh diri

Penderita dengan gejala depresi yang normal terjadi pada krisis akut biasanya hanya memerlukan perawatan suportif. Bila gejalanya berat maka dapat digunakan obat yang sifatnya hypnotic pada malam hari untuk jangka waktu pendek atau obat anti cemas. Penderita dengan gangguan penyesuaian dan depresi major, memerlukan terapi yang intensif berupa psikoterapi dan antidepresan, karena dampaknya cukup luas yang akan mempengaruhi kualitas hidup penderita. Psikoterapi bertujuan meredakan konflik atau membimbing penderita kearah acceptance (menerima) [KON-08].

Berikut adalah panduan dalam menangani penderita kanker yang menunjukkan gejala depresi [KON-08] :

- Menegakkan diagnosis dan pengobatan gangguan mental organic bila ada.

- Mengobati faktor penyebab atau pencetusnya: nyeri, gangguan fisik lain, sosial dan spiritual.
- Dukungan menyeluruh : tanggap dan empati, yakinkan lagi tentang perawatan selanjutnya, berikan informasi tentang penyakitnya, eksplorasi pengertian dan rasa takut terhadap penyakitnya dan prognosis, perkuat dukungan keluarga dan social.
- Perlu dirujuk ke psikiater bila gejalanya lebih intense, lebih lama atau kambuh setelah mengalami perbaikan, karena bagaimanapun gejala yang terus meningkat memerlukan pengobatan yang adekuat.
- Antidepresan : Golongan tricyclic antidepresan dimulai dengan dosis kecil (25 -50mg) dosis maksimum 75 sampai 150mg/hari, untuk penderita kanker stadium lanjut dosisnya lebih rendah dibandingkan yang tidak menderita kanker. Antidepresan lain: Fluoxetine, Sertralin, dll .

## 2.2 Kanker Payudara

Ada berbagai macam jenis kanker yang telah teridentifikasi, salah satunya adalah kanker payudara. Kanker payudara adalah momok menakutkan yang mengintai para wanita. Payudara merupakan salah satu organ yang menjadi identitas kesempurnaan seorang wanita. Jika organ tersebut terserang kanker maka kesempurnaan seorang wanita menjadi berkurang. Sehingga, seseorang yang terserang kanker payudara akan berusaha mencari pengobatan yang bisa menyembuhkan penyakitnya [MAH-12].

Kanker payudara adalah tumor ganas yang menyerang jaringan payudara. Jaringan payudara tersebut terdiri dari kelenjar susu (kelenjar pembuat air susu), saluran kelenjar (saluran air susu), dan jaringan penunjang payudara. Kanker payudara tidak menyerang kulit payudara yang berfungsi sebagai pembungkus [MAR-04].

Penderita kanker payudara terus meningkat tiap tahunnya. Menurut Organisasi Kesehatan Dunia (WHO) pada tahun 2004 penyakit kanker payudara menempati peringkat nomor duadiantara penyakit kanker penyebab kematian di negara-negara berkembang [NAN-09]. Penyakit ini oleh *World Health Organization (WHO)* dimasukkan dalam *International of Diseases (ICD)* dengan

nomor kode 174. Di Indonesia, kanker payudara merupakan kanker terbanyak kedua pada wanita sesudah kanker leher Rahim [WAH-06].

Kanker payudara merupakan penyakit yang paling ditakuti oleh kaum wanita. Meskipun demikian, berdasarkan penemuan terakhir, kaum pria pun bisa terkena kanker payudara. Dari hasil penelitian para ahli menunjukkan bahwa satu dari delapan wanita terkena kanker payudara. Di Indonesia, kanker payudara merupakan kanker kedua terbanyak diderita kaum wanita setelah kanker mulut/leher rahim. Kanker payudara umumnya menyerang wanita yang telah berumur lebih dari 40 tahun. Namun demikian, wanita muda pun bisa terserang kanker ini [MAR-04].

Beberapa faktor risiko pada kanker payudara yang sudah diterima secara luas oleh kalangan “*oncologist*” di dunia adalah sebagai berikut [RUP-02]:

1. Umur lebih dari 30 tahun mempunyai kemungkinan yang lebih besar untuk mendapat kanker payudara dan risiko ini akan bertambah sampai umur 50 tahun dan setelah menopause.
2. Tidak kawin/Nulipara risikonya 2-4 kali lebih tinggi daripada wanita yang kawin dan punya anak.
3. Anak pertama lahir setelah 35 tahun risikonya 2 kali lebih besar.
4. “*Menarche*” kurang dari 12 tahun risikonya 1,7-3,4 kali lebih tinggi daripada wanita dengan “*menarche*” yang datang pada usia normal atau lebih dari 12 tahun.
5. Menopause datang terlambat lebih dari 55 tahun, risikonya 2,5-5 kali lebih tinggi.
6. Pernah mengalami infeksi, trauma atau operasi tumor jinak payudara, risikonya 3-9 kali lebih besar.
7. Adanya kanker pada payudara kontralateral, risikonya 3-9 kali lebih besar.
8. Pernah mengalami operasi ginekologis-tumor ovarium, risikonya 3-4 kali lebih tinggi.
9. Yang mengalami radiasi di dinding dada risikonya 2-3 kali lebih tinggi.
10. Riwayat keluarga ada yang menderita kanker payudara pada ibu, saudara perempuan ibu, saudara perempuan, adik/kakak, risikonya 2-3 kali lebih tinggi.

11. Kontrasepsi oral pada penderita tumor payudara jinak seperti kelainan fibrokistik yang ganas akan meningkatkan risiko untuk mendapatkan kanker payudara 11 kali lebih tinggi.

Seiring dengan berkembangnya teknologi di dunia medis, maka ditemukan beberapa cara pengobatan kanker payudara. Setiap jenis pengobatan terhadap penyakit ini dapat menimbulkan masalah fisiologis, psikologis dan sosial bagi pasien. Salah satu jenis pengobatan tersebut adalah dengan cara mastektomi. Mastektomi adalah pengobatan kanker payudara dengan cara mengangkat seluruh jaringan payudara. Efek jangka panjang dari mastektomi berpengaruh sangat besar terhadap kualitas hidup karena rasa sakit dan ketidaknyamanan berikutnya [MAH-12].

Perempuan yang telah berjuang melawan kanker payudara dan selamat melalui mastektomi memiliki kekuatan dan semangat untuk bertahan. Kekuatan baru mereka yang mereka dapatkan kemudian diterapkan ke area lain dari kehidupan mereka dan mengakibatkan pengambilan risiko serta kepercayaan diri ketika berhadapan dengan tantangan dan kesulitan meningkat [MAH-12].

Hawari (2004) menyatakan bahwa wanita yang menjalani operasi mastektomi menunjukkan ekspresi yang mencerminkan kecemasan dan depresi serta sikap penolakan. Arroyo dan Lopez (2011) yang menemukan bahwa wanita pasca mastektomi akan merasa dirinya tidak menarik, takut akan ditinggalkan dan juga khawatir dengan kesehatannya selanjutnya [MAH-12].

### **2.3 Ketahanan Hidup Penderita Kanker Payudara**

Ketahanan hidup (*survival rate*) seorang penderita kanker payudara bisa dipengaruhi dari adanya program diagnosis dini pada penderita kanker payudara. Tujuan dari suatu program seperti ini bukan saja hanya memperbaiki angka ketahanan hidup (*survival rate*), tetapi juga perbaikan angka penyembuhan (*free of disease*) sebab kanker yang diobati pada stadium yang dini dengan sendirinya menaikkan angka "*survival*" biarpun penyembuhannya belum tentu tercapai [RUP-02].

Tolak ukur keberhasilan pengobatan penyakit kanker lazim digambarkan dengan angka ketahanan hidup (*year survival rate*). Pada setiap kasus, ketahanan

hidup lebih tinggi pada kanker yang masih terlokalisir. Bila terjadi penyebaran, ketahanan hidup menjadi rendah [PRE-94].

Menurut Hack (1994) dalam Pane (2002), ketahanan hidup tergantung dari adanya metastase ke kelenjar getah bening, besar lesi, kedalaman infiltrasi, adanya metastase ke parametrium, serta adanya metastase ke pembuluh darah. Menurut Hawari (2004), angka-angka statistik menunjukkan bahwa para penderita kanker usianya tidak lebih dari lima tahun untuk bertahan (*five years survival rate*). Karnadihardja (1987) menyatakan bahwa jika kanker payudara tidak diobati maka ketahanan hidup lima tahun sebesar 16%-22% dan 1%-5% dalam 10 tahun [TIO-09].

Data dari National Cancer Institute menunjukkan ketahanan hidup lima tahun penderita kanker payudara (kulit putih) umumnya lebih rendah dibandingkan dengan kanker kulit (Melanoma), kanker Rahim (Corpus Uteri), dan kanker tyroid yaitu 78% untuk kanker payudara, 82% untuk Melanoma, 84% untuk kanker Rahim, dan 94% untuk kanker tyroid. Di India, probabilitas ketahanan hidup penderita kanker payudara lebih rendah yaitu 60%. Hal ini dikarenakan 50% penderita berobat setelah stadium lanjut [VOR-80].

Penentuan ketahanan hidup tergantung dari tingkat keganasan serta probabilitas kematian pada penyakit kanker tersebut. Ketahanan hidup penyakit dengan keganasan tinggi, seperti kanker hati dan paru-paru adalah ketahanan hidup satu tahun dan ketahanan hidup tiga tahun. Pada penyakit kanker payudara adalah ketahanan hidup lima tahun (*five year survival*), ketahanan hidup sepuluh tahun, dan seterusnya [WAH-06].

Seiring dengan berkembangnya teknologi di dunia medis, maka ditemukan beberapa cara pengobatan kanker payudara. Setiap jenis pengobatan terhadap penyakit ini dapat menimbulkan masalah fisiologis, psikologis dan sosial bagi pasien. Salah satu jenis pengobatan tersebut adalah dengan cara mastektomi [MAH-12]. Salah satu tindakan operasi pada kanker payudara ini juga mempunyai pengaruh terhadap ketahanan hidup penderita kanker payudara. Macam-macam tindakan operasi pada kanker payudara yang masih operable-kurabel yaitu mastektomi simpleks dan mastektomi radikal [RUP-02].

Mastektomi simpleks dikembangkan oleh McWhirther, dimana proses operasi ini melakukan tindakan pengangkatan seluruh jaringan payudara dalam bentuk elips atau oval dengan seluruh kulit dan puting susu, dengan garis tengah minimal 10-15 cm, tanpa mengangkat kelenjar getah bening regional tetapi dilakukan pengangkatan fascia pektoralis. Pada proses pascabedah, penderita harus segera diikuti dengan penyinaran, yang sebaiknya segera dimulai 7 s/d 10 hari sesudah operasi. Penanggulangan radiasi berminggu-minggu sesudah operasi akan mempengaruhi harapan hidup si penderita sampai 10-15%. *Five years survival rate* yang dapat dicapai adalah 70-80% [RUP-02].

Setelah evaluasi dan perkembangan mastektomi simplek, maka timbullah perbaikan dan penyempurnaan mastektomi yang dinamakan mastektomi radikal yang dipelopori oleh William Stuard Halsted pada tahun 1882. Tetapi bersamaan dengan itu, seorang ahli bedah lainnya yang juga terkenal dalam bidang KPD yaitu Willy Meyer pada tahun 1894 menciptakan metode sama seperti Halsted tetapi berbeda cara pembedahannya. Bila Halsted lebih banyak mengangkat kulit lemak sehingga memerlukan transplantasi kulit sedangkan aksila dibersihkan setelah pektoralis mayor dan minor dibebaskan maka Willy Mayer mengangkat lebih sedikit kulit, lemak tinggalnya lebih banyak dan pembersihan kelenjar getah bening aksila dilakukan pada permulaan operasi. Dari kedua penelitian tersebut, kedua-duanya mencatat "*5 years survival rate*" yang cukup tinggi tanpa penyinaran yaitu 70 s/d 95% [RUP-02].

Beberapa penulis, antara lain Haagensen dkk., tidak melihat perbedaan "*5 years survival rate*" yang menyolok antara "*classical radical mastectomy*" dan "*extended radical mastectomy*" ("*5 years survival rate*" 90 s/d 97%), malah kadang-kadang bila "*extended*" radikal mastektomi dikerjakan oleh tangan yang kurang terampil disertai fasilitas dan kondisi rumah sakit yang tidak memadai maka "*5 years survival rate*" itu lebih buruk dari "*simple mastectomy*". Karena itu "*extended radical mastectomy*" sedikit demi sedikit ditinggalkan di senter-senter kedokteran negara barat. Mereka kembali pada prosedur "*classical radical mastectomy*" [RUP-02].

## 2.4 Data Haberman's Survival

Data Haberman's Survival merupakan data hasil dari penelitian mengenai ketahanan pasien yang telah menjalani operasi kanker payudara di University of Chicago's Billings Hospital yang dilakukan pada tahun 1958-1970. Data ini disumbangkan oleh Tjen-Sien Lim. Data *Haberman's Survival* digunakan untuk memprediksi apakah seorang pasien yang telah menjalani operasi kanker payudara akan bertahan hidup atau tidak [DHS-12]. Data Haberman's Survival terdiri dari 306 kasus pasien penderita kanker payudara. Masing-masing data terdiri dari empat atribut yaitu : umur pasien ketika menjalani operasi, tahun pasien menjalani operasi, jumlah node positif yang terdeteksi, dan status ketahanan hidup. Status ketahanan hidup terdiri dari dua nilai yaitu kelas 0 menunjukkan penderita (pasien) dapat bertahan hidup lima tahun atau lebih, sedangkan kelas 1 menunjukkan bahwa penderita (pasien) meninggal dalam kurun waktu tidak lebih dari lima tahun [LIM-99].

## 2.5 Data Mining

### 2.5.1 Definisi Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan dalam database besar [TUR-05]. Secara sederhana, data mining dapat diartikan sebagai proses atau “menggali” knowledge yang ada pada sekumpulan (sejumlah) data [HAK-06].

Alasan utama mengapa data mining diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan knowledge yang berguna. Informasi dan knowledge yang diperoleh tersebut dapat digunakan untuk aplikasi mulai dari analisis pasar, deteksi penipuan, retensi pelanggan, untuk mengontrol produksi dan eksplorasi ilmu pengetahuan [HAK-06].

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain [LAR-06] :



1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam database yang baik.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Kemajuan dalam pengumpulan data dan teknologi penyimpanan yang cepat memungkinkan organisasi menghimpun jumlah data yang sangat luas. Alat dan teknik analisis data yang tradisional tidak dapat digunakan untuk mengekstrak informasi dari data yang sangat besar. Untuk itu diperlukan suatu metoda baru yang dapat menjawab kebutuhan tersebut. *Data mining* merupakan teknologi yang menggabungkan metoda analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar [KUS-07].

Ada beberapa definisi dari data mining yang dikenal diantaranya adalah [KUS-07] :

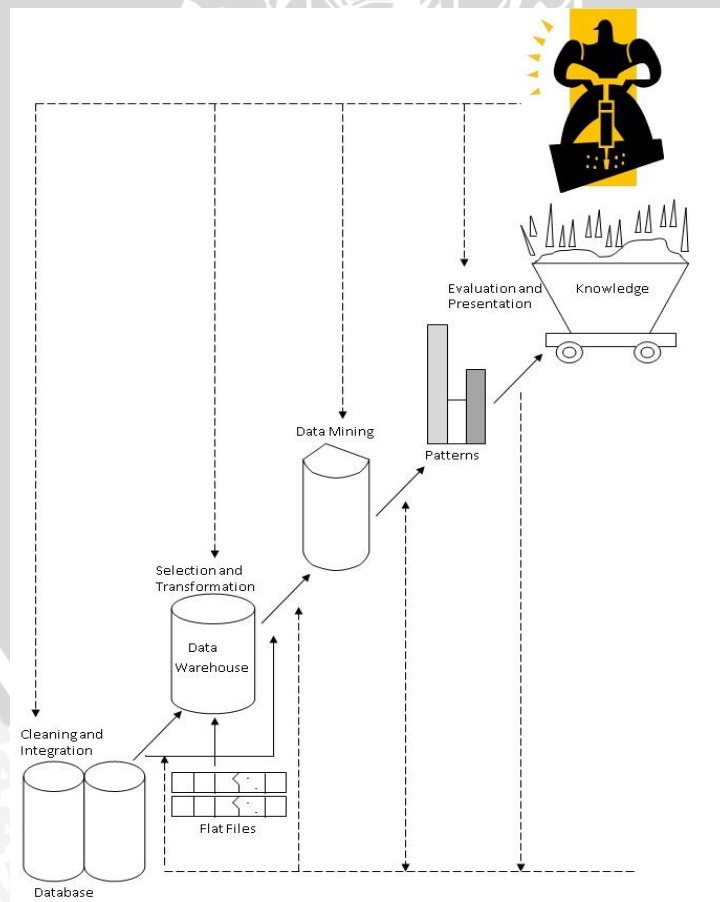
1. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.
2. Data mining adalah analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya
3. *Data mining* atau *Knowledge Discovery in Databases* (KDD) adalah pengambilan informasi yang tersembunyi, dimana informasi tersebut sebelumnya tidak dikenal dan berpotensi bermanfaat. Proses ini meliputi sejumlah pendekatan teknis yang berbeda, seperti *clustering*, *data summarization*, *learning classification rules*.

Secara umum, data mining dapat melakukan dua hal yaitu memberikan kesempatan untuk menemukan informasi menarik yang tidak terduga, dan juga

bisa menangani data berskala besar. Dalam menemukan informasi yang menarik ini, ciri khas data mining adalah kemampuan pencarian secara hampir otomatis, karena dalam banyak teknik data mining ada beberapa parameter yang masih harus ditentukan secara manual atau semi manual. Data mining juga dapat memanfaatkan pengalaman atau bahkan kesalahan di masa lalu untuk meningkatkan kualitas dari model maupun hasil analisisnya, salah satunya dengan kemampuan pembelajaran yang dimiliki beberapa teknik data mining seperti klasifikasi [KUS-07].

### 2.5.2 Tahapan Data Mining

Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge Discovery in Database*, atau yang biasa disebut KDD. Di sisi lain, orang melihat data mining hanya sebagai langkah penting dalam proses KDD. Suatu proses KDD bisa dilihat pada Gambar 2.1 sebagai berikut [HAK-06].



Gambar 2.1 Tahap-Tahapan Data Mining atau KDD

Sumber : Data Mining Concepts and Techniques [HAK-06]

Istilah *data mining* dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut [FPS-96] :

1. *Data Selection* Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing/Cleaning* Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
3. *Transformation Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
4. *Data mining* *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode dan algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation/Evaluation* Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 2.5.3 Teknik Data Mining

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, yaitu [LAR-06] :

#### 1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

#### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:

- 1) Prediksi harga beras dalam tiga bulan yang akan datang.
- 2) Prediksi presentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

#### 4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori apa.

#### 5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal. Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a) Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b) Untuk tujuan audit akutansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.
- c) Melakukan pengklusteran terhadap ekspresi dari gen, dalam jumlah besar.

## 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a) Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b) Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli bersamaan.

Untuk mendukung penelitian ini penulis menggunakan *clustering* untuk membangkitkan aturan.

### 2.5.3.1 Clustering

*Clustering* adalah studi formal metode dan algoritma untuk partisi atau pengelompokan. Analisis *clustering* tidak menggunakan pelabelan kategori sebelumnya. *Clustering* bersifat *unsupervised learning* atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi [JAI-09].

*Clustering* adalah alat penemuan mengungkapkan hubungan dan struktur di dalam data yang sebelumnya tidak jelas menjadi pengetahuan yang bermanfaat ketika ditemukan. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data / obyek ke dalam *cluster* (*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin [BUD-12].

Masalah *clustering* terkait dengan partisi data yang diberikan dan diatur dalam kelompok-kelompok (*cluster*) sehingga titik data dalam *cluster* lebih mirip satu dengan yang lain yang menunjukkan perbedaan kelompok [Halkidi, Batistakis, Vazirgiannis-2001]. *Clustering* adalah metode yang berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster* dan membuat jarak antar *cluster* sejauh mungkin. Ini berarti obyek dalam satu *cluster* sangat mirip satu sama lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain [BUD-12].

*Clustering* digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya *tool data mining*. *Clustering* membagi

item menjadi kelompok-kelompok berdasarkan yang ditemukan *tool data mining*. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar *cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi [KUS-07].

Analisis *cluster* adalah teknik untuk mengelompokkan individu atau objek menjadi beberapa kelompok tertentu dimana setiap objek yang berada dalam *cluster* yang sama mempunyai kemiripan satu dengan yang lain di bandingkan dengan anggota *cluster* yang lain. Secara khusus, tujuan analisis *cluster* adalah untuk mengklasifikasikan entitas sampel (individu atau objek) menjadi sejumlah kecil kelompok khusus yang didasarkan pada kemiripan antar entitas. Analisis *cluster* umumnya membutuhkan dua tahapan. Pertama, mengukur kemiripan atau asosiasi di antara entitas untuk menentukan berapa banyak kelompok yang akan dipakai pada sampel. Kedua, memprofilkan orang atau variabel untuk menentukan posisinya [ABH-95].

### 2.5.3.2 Tipe *Clustering*

Tan, dkk. (2006) membagi *clustering* dalam dua kelompok, yaitu *hierarchical and partitional clustering*. *Partitional Clustering* disebutkan sebagai pembagian obyek-obyek data ke dalam kelompok yang tidak saling overlap sehingga setiap data berada tepat di satu *cluster*. *Hierarchical clustering* adalah sekelompok *cluster* yang bersarang seperti sebuah pohon berjenjang (hirarki) [AND-07].

William (2005), membagi algoritma *clustering* ke dalam kelompok besar seperti berikut [AND-07] :

1. *Partitioning algorithms* : algoritma dalam kelompok ini membentuk bermacam partisi dan kemudian mengevaluasinya dengan berdasarkan beberapa kriteria.
2. *Hierarchy algorithms*: pembentukan dekomposisi hirarki dari sekumpulan data menggunakan beberapa kriteria.
3. *Density-based*: pembentukan *cluster* berdasarkan pada koneksi dan fungsi densitas.

4. *Grid-based*: pembentukan *cluster* berdasarkan pada struktur *multiple-level granularity*.
5. *Model-based*: sebuah model dianggap sebagai hipotesa untuk masing-masing *cluster* dan model yang baik dipilih diantara model hipotesa tersebut.

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. *K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [AGU-07]. Metode ini merupakan metode pengelompokan yang berujuan mengelompokkan objek sedemikian sehingga jarak tiap-tiap objek ke pusat *cluster* di dalam satu *cluster* adalah minimum [BAD-05].

#### 2.5.4 *K-Means Clustering*

##### 2.5.4.1 Pengertian *K-Means Clustering*

Algoritma *K-means* adalah algoritma yang terbaik dalam algoritma *partitional clustering* dan yang paling sering digunakan diantara algoritma *clustering* lainnya karena kesederhanaan dan efisiensinya [Liu-2007]. Algoritma *K-Means* dinilai cukup efisien, yang ditunjukkan dengan kompleksitasnya  $O(nKt)$ , dengan catatan  $n$  adalah banyaknya obyek data,  $k$  adalah jumlah *cluster* yang dibentuk, dan  $t$  banyaknya iterasi. Biasanya, nilai  $k$  dan  $t$  jauh lebih kecil daripada nilai  $n$ . selain itu, dalam iterasinya, algoritma ini akan berhenti dalam kondisi optimum [AND-07].

Pada artikel yang diterbitkan Elsevier “*Data clustering: 50 years beyond K-means*” [JAI-09] diungkapkan bahwa pengorganisasian data ke dalam suatu kluster merupakan suatu model yang paling mendasar untuk pemahaman dan pembelajaran. Analisis kluster adalah studi formal untuk



mengelompokkan, atau *clustering* benda-benda sesuai dengan karakteristik yang diukur berdasarkan kemiripan satu sama lain. *Clustering* adalah pengelompokan menggunakan teknik *unsupervised learning* dimana tidak diperlukan pelatihan pada metode tersebut atau dengan kata lain, tidak ada fase *learning* serta tidak menggunakan pelabelan pada setiap kelompok [BUD-12].

Pada jurnal yang sama diungkapkan pula perihal penggunaan K-means sebagai algoritma yang sejak dulu bahkan sampai saat ini cukup handal digunakan untuk proses *clustering*. Dalam perkembangannya *K-means* juga telah melalui beberapa modifikasi antara lain *Fuzzy C-means*. Namun jurnal “*Performance Evaluation of K-means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points*” [VEL-10] mengungkapkan bahwa melalui serangkaian percobaan didapatkan kesimpulan bahwa performansi algoritma K-means lebih baik daripada algoritma *fuzzy c-means* [BUD-12].

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* [AGU-07].

#### 2.5.4.2 Algoritma *K-Means Clustering*

Beberapa teknik *clustering* yang paling sederhana dan umum adalah *clustering K-Means*. Secara detail teknik ini menggunakan ukuran ketidakmiripan untuk mengelompokkan obyek. Ketidakmiripan dapat diterjemahkan dalam konsep jarak. Dua obyek dikatakan mirip jika jarak dua objek tersebut dekat. Semakin tinggi nilai jarak, semakin tinggi nilai ketidakmiripannya. Algoritma *clustering K-means* dapat diringkas sebagai berikut [SAN-07] :

1. Pilih jumlah *cluster*
2. Inisialisasi *k* pusat *cluster* ( diberi nilai-nilai random )
3. Tempatkan setiap data/obyek ke *cluster* terdekat. Kedekatan dua obyek ditentukan bersarkan jarak kedua obyek tersebut. Jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana.

4. Hitung kembali pusat *cluster* dengan anggota *cluster* yang sekarang. Pusat *cluster* adalah rata-rata semua data/obyek dalam *cluster*
5. Tugaskan lagi setiap obyek memakai pusat *cluster* yang baru. Jika pusat *cluster* sudah tidak berubah lagi, maka proses pengclusteran selesai.
6. Kembali ke langkah 3 sampai pusat *cluster* tidak berubah lagi.

Ada beberapa rumus jarak dua titik  $x$  dan  $y$ , dalam tulisan ini memakai jarak Euclidean adapun rumusnya adalah sebagai berikut [SAN-07] :

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Mean dan standar deviasi dari masing-masing *cluster* dapat dihitung dengan rumus [Bahari-2005] :

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2.2)$$

$$\sigma_{ki} = \sqrt{\frac{\sum_{j=1}^n (x_j - \mu_{ki})^2}{n-1}} \quad (2.3)$$

Dimana :  $k$  = jumlah *cluster* 1... $k$

$i$  = atribut 1... $i$

$n$  = jumlah data

$x_j$  = data ke- $j$

$\mu$  = rata-rata *cluster*

Selain masalah optimasi pengelompokan data ke masing-masing *cluster*, data *clustering* juga diasosiasikan dengan permasalahan penentuan jumlah *cluster* yang paling tepat untuk data yang dianalisa. Penentuan jumlah *cluster* untuk *dataset* yang dianalisa umumnya dilakukan secara *supervised* atau ditentukan dari awal oleh pengguna, walaupun dalam penerapannya ada beberapa metode yang sering dipasangkan dengan metode *K-Means*. Karena secara teori metode penentuan jumlah *cluster* ini tidak sama dengan metode pengelompokan yang dilakukan oleh *K-Means*, kevalidan jumlah *cluster* yang dihasilkan umumnya masih dipertanyakan [AGU-07].

Prosedur pengclusteran pada algoritma *K-Means* pasti bisa mengalami terminate, tetapi hasil yang diberikan dari proses *clustering* menggunakan

algoritma *K-Means* tidak selalu menghasilkan konfigurasi yang optimal. Keakuratan algoritma *K-Means* sangat dipengaruhi oleh penentuan jumlah *k* (banyaknya *cluster*) dan dipengaruhi oleh penentuan centroid awal data uji secara random dengan tepat. Masalah-masalah tersebut bisa diatasi dengan menjalankan proses algoritma *K-Means* secara berulang-ulang [HAK-06].

#### 2.5.4.3 Analisa Cluster

Analisis *cluster* adalah proses pengelompokan objek berdasarkan pada perilaku dan karakteristik yang serupa [SAM-10]. Tujuannya adalah untuk menghasilkan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu *cluster* dan semakin besar perbedaan tiap *cluster* maka kualitas analisis *cluster* semakin baik [KAR-08].

Analisis *cluster* adalah suatu teknik analisa multivariable (banyak variabel) untuk mencari dan mengorganisir informasi tentang variabel tersebut sehingga secara relatif dapat dikelompokkan dalam bentuk yang homogen dalam sebuah *cluster*. Secara umum, bisa dikatakan sebagai proses menganalisa baik tidaknya suatu proses pembentukan *cluster*. Analisa *cluster* bisa diperoleh dari kepadatan *cluster* yang dibentuk (*cluster density*) [NOH-09].

Dalam analisis *cluster* kumpulan objek harus dimodelkan dan diterjemahkan menjadi suatu bentuk representasi formal. Terdapat dua model struktur data untuk mempresentasikan objek, yaitu *data matrix* dan *dissimilarity matrix* [HAK-06].

*Data matrix* (*object-by-variable structure*), model ini mempresentasikan *n* objek, seperti suatu titik dengan variabel *p*, atau disebut juga pengukuran atribut seperti umur, tinggi, lebar, dan sebagainya. Data berasal dari hasil pengukuran atau pengamatan suatu objek direpresentasikan sebagai titik atau vektor pada ruang multidimensi. Struktur dari bentuk hubungan antara objek dan variabel (*n* objek x *p* variabel) [HAK-06] :

$$\begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

Gambar 2.2 Contoh *Data Matrix Object by Variable Structure*

Sumber : Data Mining Concepts and Techniques [HAK-06]

*Dissimilarity matrix (object-by-object structure)*, model ini menyimpan koleksi kedekatan antara sebuah objek dengan objek lain. Model ini direpresentasikan dalam bentuk matrix  $n \times n$ , dimana  $n$  adalah banyaknya objek [HAK-06].

$$\begin{bmatrix} 0 & & & & & \\ d(2,1) & 0 & & & & \\ d(3,1) & d(3,2) & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \\ d(n,1) & d(n,2) & \dots & \dots & 0 & \end{bmatrix}$$

Gambar 2.3 Contoh *Data Matrix Object by Object Structure*

Sumber : Data Mining Concepts and Techniques [HAK-06]

Gambar 2.3 menggambarkan *Dissimilarity matrix* dimana  $d(i,j)$  adalah ukuran perbedaan jarak antara objek  $i$  dan  $j$ . Pada umumnya  $d(i,j)$  bukan bilangan negatif [HAK-06].

**2.5.4.4 Analisa Varian**

Analisa *cluster* bisa diperoleh dari kepadatan *cluster* yang dibentuk (*cluster density*). Kepadatan suatu *cluster* dapat ditentukan dengan *variance within cluster* ( $V_w$ ) dan *variance between cluster* ( $V_b$ ). Varian tiap tahap pembentukan *cluster* dihitung dengan persamaan [MRA-10]:

$$V_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^n (y_i - \bar{y}_c)^2 \tag{2.4}$$

dengan :  $V_c^2$  = varian pada *cluster c*



$c = 1..k$ , dimana  $k =$  jumlah *cluster*

$n_c =$  jumlah data pada *cluster*  $c$

$y_i =$  data ke- $i$  pada suatu *cluster*

$\bar{y}_c =$  rata-rata dari data pada suatu *cluster*

Selanjutnya dari nilai varian diatas, kita bisa menghitung nilai *variance within cluster* ( $V_w$ ) dengan rumus :

$$V_w = \frac{1}{N - c} \sum_{i=1}^c (n_i - 1)v_i^2 \quad (2.5)$$

dimana :  $N =$  jumlah semua data

$c = 1..k$ , dimana  $k =$  jumlah *cluster*

$n_i =$  jumlah data pada *cluster*

$v_i =$  varian pada *cluster*

Dan nilai *variance between cluster* ( $V_b$ ) dengan rumus:

$$V_b = \frac{1}{c - 1} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 \quad (2.6)$$

dimana  $\bar{y} =$  rata-rata dari  $\bar{y}_i$

Salah satu metode yang digunakan untuk menentukan *cluster* yang ideal adalah nilai varian, yaitu dengan menghitung kepadatan *cluster* berupa *variance within cluster* ( $V_w$ ) dan *variance between cluster* ( $V_b$ ). *Cluster* yang ideal mempunyai  $V_w$  minimum yang merepresentasikan internal homogeneity dan maksimum  $V_b$  yang menyatakan external homogeneity.

$$V = \frac{V_w}{V_b} \quad (2.7)$$

Meskipun minimum  $V_w$  menunjukkan nilai *cluster* yang ideal, tetapi pada beberapa kasus kita tidak bisa menggunakannya secara langsung untuk mencapai global optimum. Jika kita paksaan, maka solusi yang dihasilkan akan jatuh pada local optima [NOH-09].

#### 2.5.4.5 Least Square Estimator

Jang (1997), mengemukakan bahwa pembelajaran yang digunakan adalah *steepest descent* (SD) dan *least square estimator* (LSE) dapat digunakan untuk mengidentifikasi parameter-parameter linier [JSM-97].

Fungsi keanggotaan yang diciptakan dari persamaan Gauss berguna untuk fuzzifikasi nilai *input* pada sistem *fuzzy*. Fungsi keanggotaan Gauss dapat dibuat apabila pusat *cluster* dan standar deviasi untuk setiap atribut diketahui. Pusat *cluster* dapat dicari melalui proses *clustering* sehingga diperoleh matriks pusat *cluster*. Sedangkan standar deviasi dicari dengan rumus pada persamaan (2.3). Nilai standar deviasi setiap atribut pada setiap *cluster* dicari sehingga didapatkan matriks standar deviasi berukuran *cluster* x atribut [KUP-10].

Mulai proses berikut, derajat keanggotaan hanya melibatkan variabel-variabel *input* saja. Nilai  $j = 1, 2, \dots, m$  digunakan untuk menyatakan variabel *input* saja ( $m =$  jumlah variabel input). Nilai *output* dicari melalui langkah-langkah sebagai berikut [KUP-10]:

- Derajat keanggotaan setiap data  $i$  dalam *cluster*  $k$  dikalikan dengan setiap atribut  $j$  dari data  $i$ , yang dapat dimisalkan dengan  $d_{ij}^k$  dan dihitung berdasarkan persamaan (2.8).

$$d_{ij}^k = x_{ij} * \mu_{ki} \text{ dan } d_{i(m+1)}^k = \mu_{ki} \quad (2.8)$$

- Proses normalisasi dilakukan dengan cara membagi  $d_{ij}^k$  dan  $d_{i(m+1)}^k$  dengan jumlah derajat keanggotaan setiap titik data  $i$  pada *cluster*  $k$  menggunakan persamaan (2.9) untuk  $d_{ij}^k$  dan persamaan (2.10) untuk  $d_{i(m+1)}^k$ .

$$d_{ij}^k = \frac{d_{ij}^k}{\sum_{k=1}^c \mu_{ki}} \quad (2.9)$$

$$d_{i(m+1)}^k = \frac{d_{i(m+1)}^k}{\sum_{k=1}^c \mu_{ki}} \quad (2.10)$$

- Langkah selanjutnya adalah membentuk matriks  $U$  yang berukuran  $n \times (c*(m+1))$  dengan :

- |                               |                                   |
|-------------------------------|-----------------------------------|
| - $U_{i1} = d_{i1}^1$         | - $U_{i(2m+1)} = d_{i(m+1)}^2$    |
| - $U_{i2} = d_{i2}^1$         | - $U_{i(c*(m+1)-m)} = d_{i1}^c$   |
| - $U_{im} = d_{im}^1$         | - $U_{i(c*(m+1)-m+1)} = d_{i2}^c$ |
| - $U_{i(m+1)} = d_{i(m+1)}^1$ | - $U_{i(c*(m+1)-1)} = d_{im}^c$   |
| - $U_{i(m+2)} = d_{i1}^2$     | - $dst$                           |

Sehingga untuk n titik data akan diperoleh matriks U sebagai berikut :

$$U = \begin{pmatrix} u_{11} & u_{12} & u_{1m} & u_{1(m+1)} & u_{1(m+2)} & \dots & u_{1(c^*(m+1))} \\ u_{21} & u_{22} & u_{2m} & u_{2(m+1)} & u_{2(m+2)} & \dots & u_{2(c^*(m+1))} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & u_{nm} & u_{n(m+1)} & u_{n(m+2)} & \dots & u_{n(c^*(m+1))} \end{pmatrix}$$

Vektor z, sebagai vector *output* berbentuk :

$$z = [z_1 + z_2 + \dots + z_n]^T$$

Dari vektor k, matriks U, dan vektor z dapat dibentuk suatu sistem persamaan linier seperti pada persamaan (2.11).

$$U * k = z \quad (2.11)$$

untuk mencari nilai koefisien *output* tiap-tiap aturan pada setiap variabel ( $k_{ij}$ ,  $i=1,2,\dots,n$ ; dan  $j=1,2,\dots,m+1$ ). Karena matriks U bukan matriks bujursangkar, maka digunakan metode kuadrat terkecil (*least square*).

- Selanjutnya, dilakukan perhitungan kuadrat terkecil untuk membentuk matriks K dengan persamaan (2.12) [FAR-08].

$$[k] = ([U]^T \cdot [U])^{-1} \cdot [U]^T \cdot [Y] \quad (2.12)$$

Keterangan :

k = koefisien *output*

Y = nilai target *output*

U = matriks U

- Untuk mempermudah komputasi, matriks K yang berukuran  $c \times (m+1)$  :

$$K = \begin{pmatrix} k_{11} & k_{12} & k_{13} & \dots & k_{1m} & k_{10} \\ k_{21} & k_{22} & k_{23} & \dots & k_{2m} & k_{20} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{c1} & k_{c2} & k_{c3} & \dots & k_{cm} & k_{c0} \end{pmatrix}$$

disusun menjadi satu vektor k :

$$k = [k_{11} \ k_{12} \ \dots \ k_{1m} \ k_{10} \ k_{21} \ k_{22} \ \dots \ k_{2m} \ k_{20} \ \dots \ k_{c1} \ k_{c2} \ \dots \ k_{cm} \ k_{c0}]^T$$

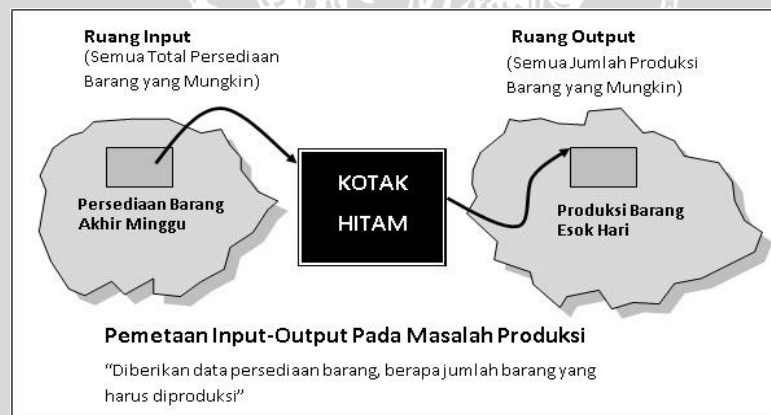
kemudian digabung dengan matriks  $U$  dan vektor  $Z$ , membentuk persamaan (2.11).

Ketika nilai parameter-parameter bagian *premise* telah ditentukan, maka total keluaran dapat dinyatakan sebagai kombinasi linier dari parameter-parameter *consequent* [JSM-97].

## 2.6 Logika Fuzzy

Logika *fuzzy* pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dasar logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Nilai keanggotaan atau derajat keanggotaan atau *membership function* menjadi ciri utama dari penalaran dengan logika *fuzzy* tersebut [KUP-10].

Logika *fuzzy* dapat dianggap sebagai kotak hitam yang menghubungkan antara ruang *input* dengan ruang *output*. Kotak hitam tersebut berisi cara atau metode yang dapat digunakan untuk mengolah data *input* menjadi *output* dalam bentuk informasi yang baik [KUP-10].



Pada gambar 2.4 Pemetaan suatu *input-output*

Kata *Fuzzy* merupakan kata sifat yang berarti kabur, tidak jelas. *Fuzziness* atau kekaburan atau ketidakjelasan atau ketidakpastian selalu meliputi keseharian manusia. Orang yang belum pernah mengenal logika *fuzzy* pasti akan mengira bahwa logika *fuzzy* adalah sesuatu yang rumit dan tidak menyenangkan. Namun, sekali seseorang mulai mengenalnya, pasti akan tertarik untuk ikut mempelajari logika *fuzzy*. Logika *fuzzy* dikatakan sebagai logika baru yang lama, sebab ilmu tentang logika *fuzzy* modern dan metodis baru ditemukan beberapa tahun yang



lalu, padahal sebenarnya konsep tentang logika *fuzzy* itu sendiri sudah ada sejak lama [KUP-10].

Menurut Cox (1994), ada beberapa alasan mengapa orang menggunakan *Fuzzy Logic*, antara lain [KUP-10] :

1. Konsep *Fuzzy Logic* mudah dimengerti. Konsep matematis yang mendasari penalaran *fuzzy* sangat sederhana dan mudah dimengerti.
2. *Fuzzy Logic* sangat fleksibel.
3. *Fuzzy Logic* memiliki toleransi terhadap data-data yang tidak tepat.
4. *Fuzzy Logic* mampu memodelkan fungsi-fungsi non linear yang sangat kompleks.
5. *Fuzzy Logic* dapat membangun dan mengaplikasikan pengalaman-pengalaman para pakar secara langsung tanpa harus melalui proses pelatihan.
6. *Fuzzy Logic* dapat bekerjasama dengan teknik-teknik kendali secara konvensional.
7. *Fuzzy Logic* didasarkan pada bahasa alami.

Menurut Dr. Lotfi A. Zadeh (1988) yang dikutip dari buku *Fuzzy Set, Uncertainty and Information*, Logika *fuzzy* adalah : suatu sistem yang digunakan untuk menangani konsep kebenaran parsial yaitu kebenaran yang berada diantara sepenuhnya benar dan sepenuhnya salah. Logika *fuzzy* merupakan salah satu sistem cerdas, dimana ciri dari sistem cerdas adalah sebagai berikut [KUP-10] :

1. Kemampuan belajar dan penalaran nampak nyata.
2. Kemampuan mengolah data/fakta yang tidak jelas/kabur (*fuzzy*) dengan mudah.
3. Kemudahan pengembangannya.

Misalkan jika terjadi perubahan kegunaan pada konfigurasi sistem logika *fuzzy*, maka tidak perlu merombak seluruh sistem tetapi cukup dengan mengganti bagian yang dapat membuat sistem tersebut bermanfaat kembali [KUP-10)].

Secara umum dalam sistem logika *fuzzy* terdapat empat buah elemen dasar, yaitu [JSM-97] :

1. Basis kaidah (*rule base*), yang berisi aturan-aturan secara linguistik yang bersumber dari pakar;

2. Suatu mekanisme pengambilan keputusan (*inference engine*), yang memperagakan bagaimana para pakar mengambil suatu keputusan dengan menerapkan pengetahuan (*knowledge*);
3. Proses fuzzifikasi (*fuzzification*), yang mengubah besaran tegas (*crisp*) ke besaran *fuzzy*,
4. Proses defuzzifikasi (*defuzzification*), yang mengubah besaran *fuzzy* hasil dari *inference engine*, menjadi besaran tegas (*crisp*).

### 2.7 Himpunan Fuzzy

Himpunan tegas (*crisp*)  $A$  didefinisikan oleh item-item yang ada pada himpunan itu. Pada himpunan tegas (*crisp*), nilai keanggotaan suatu item  $x$  dalam suatu himpunan  $A$  ( $\mu_A(x)$ ) memiliki dua kemungkinan, yaitu [KUP-10] :

1. Satu (1), yang berarti bahwa suatu item menjadi anggota dalam suatu himpunan.
2. Nol (0), yang berarti bahwa suatu item tidak menjadi anggota dalam suatu himpunan.

Himpunan *Fuzzy* didasarkan pada gagasan untuk memperbesar jangkauan fungsi karakteristik sedemikian hingga fungsi tersebut akan mencakup bilangan real pada interval  $[0,1]$ . Nilai keanggotaannya menunjukkan bahwa suatu item dalam semesta pembicaraan tidak hanya berada pada 0 atau 1, namun juga nilai yang terletak diantaranya. Dengan kata lain, nilai kebenaran suatu item tidak hanya benar atau salah. Nilai 0 menunjukkan salah, nilai 1 menunjukkan benar, dan masih ada nilai-nilai yang terletak antara benar atau salah. Himpunan *fuzzy* memiliki 2 atribut, yaitu [KUP-10] :

- a. Linguistik, yaitu penamaan suatu grup yang mewakili suatu keadaan atau kondisi tertentu dengan menggunakan bahasa alami.
- b. Numeris, yaitu suatu nilai (angka) yang menunjukkan ukuran dari suatu variabel.

Ada beberapa hal yang perlu diketahui dalam memahami sistem *fuzzy*, yaitu [KUP-10] :

- a. Variabel *Fuzzy*

Variabel *fuzzy* merupakan variabel yang hendak dibahas dalam suatu sistem *fuzzy*.

b. Himpunan *Fuzzy*

Himpunan *fuzzy* merupakan suatu grup yang mewakili suatu kondisi atau keadaan tertentu dalam suatu variabel.

c. Semesta Pembicaraan

Semesta pembicaraan adalah keseluruhan nilai yang diperbolehkan untuk dioperasikan dalam suatu variabel *fuzzy*. Semesta pembicaraan merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai semesta pembicaraan dapat berupa bilangan positif maupun negatif. Adakalanya nilai semesta pembicaraan ini tidak dibatasi batas atasnya.

d. Domain

Domain himpunan *fuzzy* adalah keseluruhan nilai yang diijinkan dalam semesta pembicaraan dan boleh dioperasikan dalam suatu himpunan *fuzzy*. Seperti halnya semesta pembicaraan, domain merupakan himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai domain dapat berupa bilangan positif maupun negatif.

## 2.8 Fungsi Keanggotaan

Fungsi keanggotaan (*membership function*) adalah suatu kurva yang menunjukkan pemetaan titik-titik input data kedalam nilai keanggotaannya (sering juga disebut dengan derajat keanggotaan) yang memiliki interval antara 0 sampai 1. Salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi [KUP-10].

Himpunan Samar didefinisikan oleh fungsi keanggotaannya. Fungsi keanggotaan merupakan suatu kurva yang menunjukkan pemetaan titik-titik data masukan ke dalam nilai keanggotaannya (JSM-97).

Fungsi-fungsi keanggotaan *fuzzy* terparameterisasi satu dimensi yang umum digunakan diantaranya adalah [JSM-97]:

1. Fungsi keanggotaan segitiga (*Triangular membership function*)

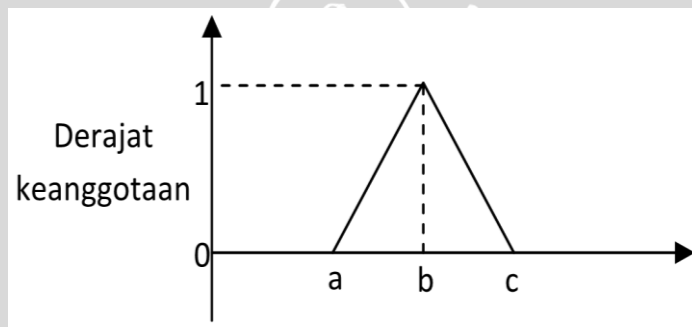
Fungsi keanggotaan segitiga ditentukan oleh 3 parameter yaitu  $\{a, b, c\}$  dengan mengikuti aturan dalam persamaandisifati oleh parameter  $\{a,b,c\}$  yang didefinisikan sebagai berikut :

$$\text{segitiga}(x, a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (2.13)$$

Atau dengan menggunakan min dan max, dapat didefinisikan dengan persamaan :

$$\text{segitiga}(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (2.14)$$

Parameter {a, b, c} dengan  $a < b < c$  menentukan koordinat x dari 3 sudut fungsi keanggotaan segitiga. Fungsi keanggotaan segitiga dapat digambarkan seperti dalam gambar 2.5.



Gambar 2.5 Kurva Segitiga

2. Fungsi keanggotaan trapesium (*Trapezoidal membership function*)

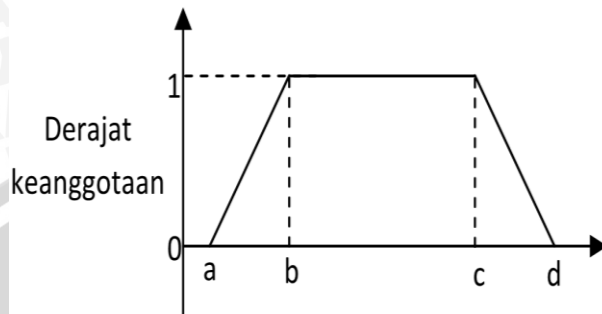
Fungsi keanggotaan trapesium ditentukan 4 parameter {a, b, c, d} yang mengikuti aturan dalam persamaan :

$$\text{trapesium}(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad (2.15)$$

Dan sebagai alternatif dapat digunakan min dan max dalam persamaan :

$$\text{trapesium}(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \quad (2.16)$$

Dalam persamaan (2.10) parameter {a, b, c, d} dengan  $a < b < c < d$  menentukan koordinat x dari 3 sudut fungsi keanggotaan trapesium. Fungsi keanggotaan trapesium dapat digambarkan seperti pada gambar 2.6.



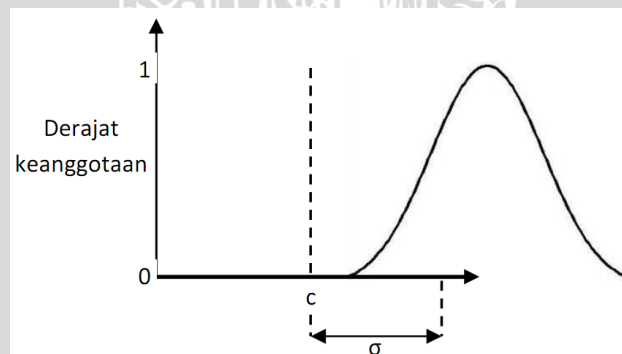
Gambar 2.6 Kurva Trapesium

3. Fungsi keanggotaan Gaussian (*Gaussian membership function*)

Fungsi keanggotaan Gaussian ditentukan dengan 2 parameter {c,  $\sigma$ } dengan mengikuti persamaan :

$$gaussian(x; c; \sigma) = e^{-\frac{1}{2} \left( \frac{x-c}{\sigma} \right)^2} \quad (2.17)$$

Fungsi keanggotaan gaussian ditentukan oleh c dan  $\sigma$ . c merepresentasikan titik tengah (center) dan  $\sigma$  merepresentasikan lebar dari fungsi keanggotaan. Fungsi keanggotaan Gaussian dapat diwujudkan seperti pada gambar 2.7.



Gambar 2.7 Kurva Gaussian

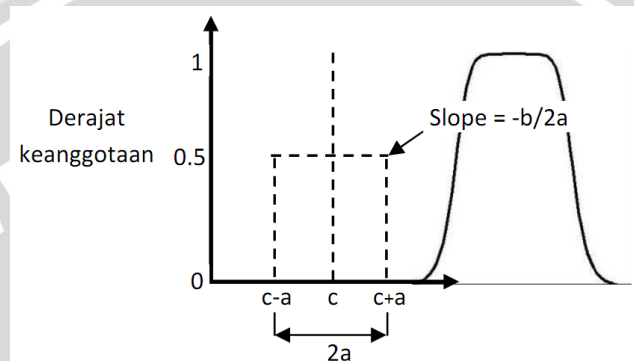
4. Fungsi keanggotaan lonceng (*Bell membership function*)

Fungsi keanggotaan lonceng ditentukan oleh 3 parameter {a, b, c} dengan mengikuti persamaan :



$$bell(x; a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}} \quad (2.18)$$

c mendefinisikan titik tengah, a mendefinisikan lebar kurva dan b digunakan untuk mengendalikan nilai *slope* dan *crossover*. Parameter b biasanya bernilai positif. Fungsi keanggotaan lonceng dapat diilustrasikan seperti gambar 2.8.



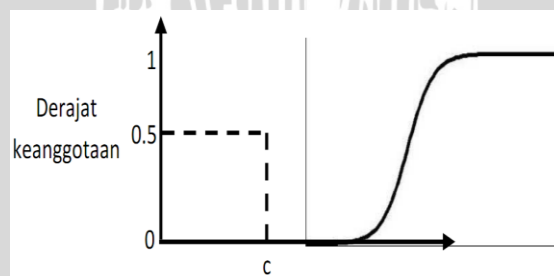
Gambar 2.8 Kurva Lonceng

5. Fungsi keanggotaan sigmoidal (*Sigmoidal membership function*)

Fungsi keanggotaan sigmoidal didefinisikan dengan persamaan :

$$sig(x; a, c) = \frac{1}{1 + \exp[-a(x - c)]} \quad (2.19)$$

Nilai parameter a mengendalikan slope pada nilai crossover  $x = c$ . Fungsi keanggotaan sigmoidal dapat dilihat pada gambar 2.9.



Gambar 2.9 Kurva Sigmoidal

### 2.9 Penalaran Monoton

Metode penalaran secara monoton digunakan sebagai dasar untuk teknik implikasi *fuzzy*. Meskipun penalaran ini sudah jarang sekali digunakan, namun terkadang masih digunakan untuk penskalaan *fuzzy*. Jika 2 daerah *fuzzy* direlasikan dengan implikasi sederhana sebagai berikut [COX-05]:

IF  $x$  is  $A$  THEN  $y$  is  $B$

transfer fungsi :

$$y = f((x, A), B)$$

maka sistem *fuzzy* dapat berjalan tanpa harus melalui komposisi dan dekomposisi *fuzzy*. Nilai *output* dapat diestimasi secara langsung dari nilai keanggotaan yang berhubungan dengan antesedennya [KUP-10].

## 2.10 Fungsi Implikasi

Fungsi implikasi berkaitan dengan bagaimana cara menginterpretasikan proposisi samar menjadi suatu relasi samar. Dimana kaidah *fuzzy* IF-THEN (dikenal juga sebagai kaidah *fuzzy*, implikasi *fuzzy* atau pernyataan kondisi *fuzzy*) [JSM-97].

Tiap-tiap aturan (proposisi) pada basis pengetahuan *fuzzy* akan berhubungan dengan suatu relasi *fuzzy*. Bentuk umum dari aturan yang digunakan dalam fungsi implikasi adalah [KUP-10]:

IF  $x$  is  $A$  THEN  $y$  is  $B$

dengan  $x$  dan  $y$  adalah skalar, dan  $A$  dan  $B$  adalah himpunan *fuzzy*. Proporsi yang mengikuti IF disebut sebagai anteseden, sedangkan proposisi yang mengikuti THEN disebut sebagai konsekuen. Proposisi ini dapat diperluas dengan menggunakan operator *fuzzy*, seperti :

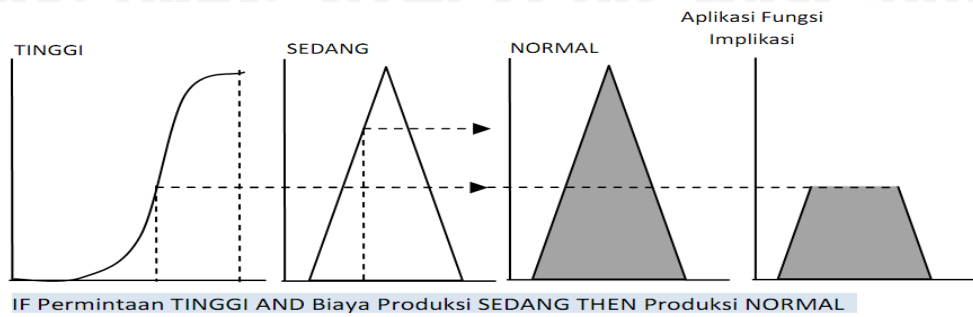
IF ( $x_1$  is  $A_1$ ) o ( $x_2$  is  $A_2$ ) o ( $x_3$  is  $A_3$ ) o ... o ( $x_N$  is  $A_N$ ) THEN  $y$  is  $B$

dengan o adalah operator (missal : OR atau AND).

Secara umum, ada 2 fungsi implikasi yang dapat digunakan, yaitu [KUP-10] :

- a. Min (*minimum*).

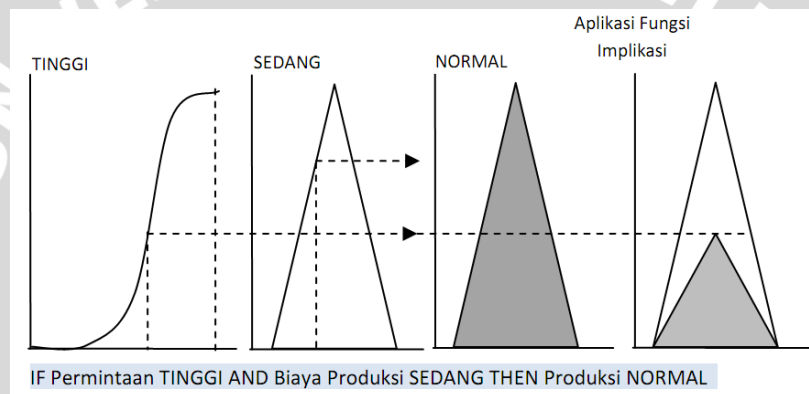
Fungsi ini akan memotong *output* himpunan *fuzzy*. Gambar 2.10 menunjukkan salah satu contoh penggunaan fungsi min.



Gambar 2.10 Fungsi Implikasi : MIN

b. Dot (*product*).

Fungsi ini akan menskala *output* himpunan *fuzzy*. Gambar 2.11 menunjukkan salah satu contoh penggunaan fungsi dot.



Gambar 2.11 Fungsi Implikasi : DOT

### 2.11 Fuzzy Inference System

Sistem inferensi *fuzzy* merupakan proses pengolahan data dalam bentuk *crisp* input yang melalui beberapa tahapan dalam sistem *fuzzy* untuk menghasilkan data dalam bentuk *crisp output* [WID-07].

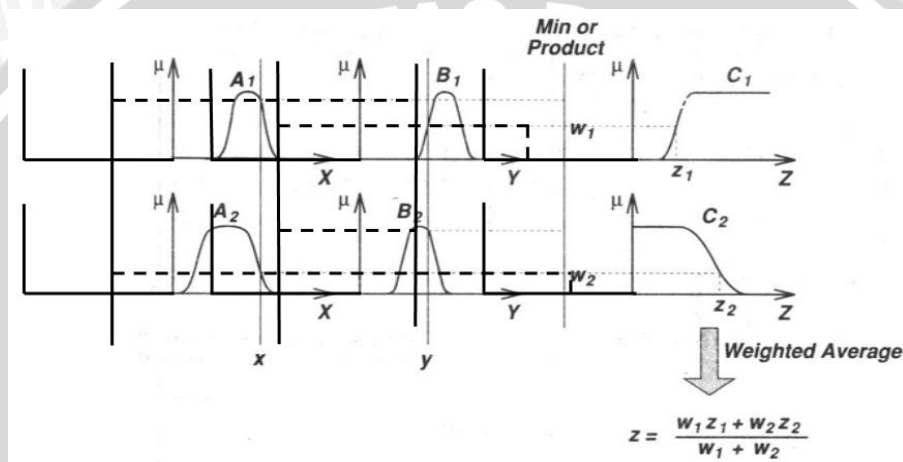
Sistem inferensi *fuzzy* menerima input *crisp*. Input ini kemudian dikirim ke basis pengetahuan yang berisi  $n$  aturan *fuzzy* dalam bentuk If-Then. Fire strength akan dicari pada setiap aturan. Apabila jumlah aturan lebih dari satu, maka akan dilakukan agregasi dari semua aturan. Selanjutnya, hasil agregasi akan dilakukan *defuzzy* untuk mendapatkan nilai *crisp* sebagai keluaran sistem. Terdapat beberapa model Sistem Inferensi Samar [JSM-97], antara lain :

- Model *Fuzzy* Mamdani
- Model *Fuzzy* Sugeno (TSK)
- Model *Fuzzy* Tsukamoto



Perbedaan antara ketiga sistem inferensi *fuzzy* terdapat pada konsekuen dari aturan *fuzzy*, agregasi dan prosedur defuzzifikasi.

Metode Tsukamoto merupakan perluasan dari penalaran monoton. Pada Metode Tsukamoto, setiap konsekuen pada aturan yang berbentuk IF-THEN harus direpresentasikan dengan suatu himpunan *fuzzy* dengan fungsi keanggotaan yang monoton (Gambar 2.12). Sebagai hasilnya, *output* hasil inferensi dari tiap-tiap aturan diberikan secara tegas (*crisp*) berdasarkan  $\alpha$ -predikat (*fire strength*). Hasil akhirnya diperoleh dengan menggunakan rata-rata terbobot [KUP-10].



Gambar 2.12 Inferensi dengan menggunakan Metode Tsukamoto [JSM-97]

Model Mamdani sering dikenal sebagai Metode Max-Min. metode ini diperkenalkan oleh Ebrahim Mamdani pada tahun 1975. Untuk mendapatkan *output*, diperlukan 4 tahapan [KUP-10] :

1. Pembentukan Himpunan *Fuzzy*

Pada Metode Mamdani, baik variabel *input* maupun variabel *output* dibagi menjadi data atau lebih himpunan *fuzzy*.

2. Aplikasi Fungsi Implikasi

Pada Metode Mamdani, fungsi implikasi yang digunakan adalah Min.

3. Komposisi Aturan

Tidak seperti penalaran monoton, apabila sistem terdiri dari beberapa aturan, maka inferensi diperoleh dari kumpulan dan korelasi antar aturan.

Ada 3 metode yang digunakan dalam inferensi sistem *fuzzy*, yaitu : max, additive, dan probabilistic OR (probor).

#### 4. Penegasan (*Defuzzy*)

*Input* dari proses defuzzifikasi adalah suatu himpunan *fuzzy* yang diperoleh dari komposisi aturan-aturan *fuzzy*, sedangkan *output* yang dihasilkan merupakan suatu bilangan pada domain himpunan *fuzzy* tersebut. Sehingga jika diberikan suatu himpunan *fuzzy* dalam range tertentu, maka harus dapat diambil suatu nilai *crisp* tertentu sebagai *output*.

Dalam penelitian skripsi ini penulis akan menggunakan sistem inferensi *fuzzy* menggunakan Metode Sugeno. Model sugeno sering digunakan karena memiliki proses perhitungan yang sederhana sehingga membutuhkan waktu relatif cepat dan efisien dan metode ini tidak terlalu membebani sistem dalam melakukan komputasi [PRI-07].

Penalaran dengan Metode Sugeno hampir sama dengan penalaran Metode Mamdani, hanya saja *output* (konsekuen) sistem tidak berupa himpunan *fuzzy*, melainkan berupa konstanta atau persamaan linear. Metode ini diperkenalkan oleh Takagi-Sugeno Kang pada tahun 1985, sehingga metode ini sering juga dinamakan dengan Metode TSK [KUP-10].

Menurut Cox (1994), metode TSK terdiri dari 2 jenis, yaitu [KUP-10] :

##### a. Model *Fuzzy* Sugeno Orde-Nol

Secara umum bentuk model *fuzzy* Sugeno Orde-Nol adalah :

$$\text{IF } (x_1 \text{ is } A_1) \text{ o } (x_2 \text{ is } A_2) \text{ o } (x_3 \text{ is } A_3) \text{ o } \dots \text{ o } (x_N \text{ is } A_N) \text{ THEN } z = k$$

dengan  $A_i$  adalah himpunan *fuzzy* ke- $i$  sebagai anteseden, dan  $k$  adalah suatu konstanta (tegas) sebagai konsekuen.

##### b. Model *Fuzzy* Sugeno Orde-Satu

Secara umum bentuk model *fuzzy* Sugeno Orde-Satu adalah :

$$\text{IF } (x_1 \text{ is } A_1) \text{ o } \dots \text{ o } (x_N \text{ is } A_N) \text{ THEN } z = p_1 * x_1 + \dots + p_N * x_N + q$$

dengan  $A_i$  adalah himpunan *fuzzy* ke- $i$  sebagai anteseden, dan  $p_i$  adalah suatu konstanta (tegas) ke- $i$  dan  $q$  juga merupakan konstanta dalam konsekuen.

Apabila komposisi aturan menggunakan metode Sugeno, maka defuzzifikasi dilakukan dengan cara mencari nilai rata-ratanya.

#### 2.12 Aturan Dasar *Fuzzy*

Aturan dasar (*rule based*) pada kontrol logika *fuzzy* merupakan suatu bentuk aturan relasi/implikasi “Jika-Maka” atau “IF-THEN”. Sebagaimana telah

dibahas sebelumnya prosedur inferensi pada controller *fuzzy* mengikuti pendekatan berbasis aturan. Ada dua model utama untuk inferensi pada *fuzzy* sistem yaitu Model Mamdani dan Model Takagi-Sugeno-Kang [AKH-10].

Pada penelitian skripsi ini akan digunakan bentuk Model Sugeno. Bentuk Model Sugeno adalah sebagai berikut :

$$\text{IF } (x_1 \text{ is } A_1) \text{ o } \dots \text{ o } (x_N \text{ is } A_N) \text{ THEN } z = p_1 * x_1 + \dots + p_N * x_N + q$$

dengan  $A_i$  adalah himpunan *fuzzy* ke- $i$  sebagai anteseden, dan  $p_i$  adalah suatu konstanta (tegas) ke- $i$  dan  $q$  juga merupakan konstanta dalam konsekuen [KUP-10]. Dalam bentuk ini memiliki anteseden dan konsekuennya dalam bentuk persamaan linear dari *input*. Model ini memungkinkan untuk lebih mempermudah pada teknik pembelajaran untuk pengidentifikasian data. Pada penelitian skripsi ini digunakan *K-Means clustering* untuk otomatisasi aturan pada model Takagi-Sugeno seperti pembangkitan aturan melalui data yang diberikan para pakar. Dengan pendekatan ini tidak diperlukan keahlian khusus dan pengalaman seorang pakar untuk menentukan aturan.

### 2.13 Teknik Pembangkitan Aturan *Fuzzy*

Algoritma pembelajaran untuk pembangkitan aturan *fuzzy* secara otomatis merupakan salah satu algoritma yang sudah banyak dikenal. Algoritma ini menghasilkan pembangkitan aturan *fuzzy* yang di dasarkan pada pembelajaran vektor-vektor dan pengoptimalan perilaku I/O sistem. Sebuah struktur aturan yang sangat sederhana dan bisa membatasi fungsi keanggotaan yang mengarah ke solusi jauh sudah digunakan secara optimum oleh Cheong et al (1996). Aturan *fuzzy* memiliki sebuah geometri yang sederhana dalam ruang input output, biasanya memiliki bentuk ellipsoid. Aturan *fuzzy* menyebabkan terjadinya pembagian *fuzzy* dalam ruang produk rule variabel input output. Selain itu *clustering* adalah teknik yang sangat cocok untuk mendeteksi pembagian partisi dalam *fuzzy* terutama dalam pembangkitan aturan *fuzzy* secara otomatis [PRI-07].

Mesin pembelajaran adalah bagian dari kecerdasan buatan. Termasuk algoritma untuk membuat keputusan berdasarkan data. Arapoglou, et all (2010) meneliti dan menekankan pada pembelajaran tanpa pengawasan, dimana mereka berusaha mengatur sekumpulan data tanpa pengetahuan terstruktur sebelumnya. Algoritma yang digunakan untuk ekstraksi aturan *fuzzy* secara otomatis adalah

dengan menggunakan metode *clustering*. Dalam penelitiannya, Arapoglou menerangkan bahwa banyak algoritma lain yang sudah banyak dikenal yang digunakan dalam diskusi menyelesaikan kasus tentang ekstraksi aturan. Seperti *K-means*, *Fuzzy C-means*, *Subtractive*, dan *Nearest Neighborhood Clustering* [AKH-10].

Dalam penelitian skripsi ini, digunakan *K-means clustering*. Yang mana proses *clustering* ini digunakan untuk menentukan jumlah aturan yang nantinya akan dibangkitkan. Dimana jumlah aturan yang akan dibangkitkan nantinya dipengaruhi oleh jumlah banyaknya *cluster* yang ditentukan. Sehingga jumlah *cluster* yang dibentuk sama dengan jumlah banyaknya aturan yang dibangkitkan.

#### 2.14 Ekstraksi Aturan *Fuzzy* dari *Cluster*

Aturan *Fuzzy* kadang-kadang dapat diperoleh dari ahli manusia. Akuisisi pengetahuan, bagaimanapun juga adalah tugas yang rumit, dan beberapa bagian sistem tidak diketahui, sistem pakar belum tersedia. Oleh karena itu, data pembangun aturan *fuzzy* dari pengukuran input/output data lebih banyak diawasi. Dengan menginisialisasi *clustering* secara berlebihan dalam menentukan jumlah *cluster*, ada kemungkinan peningkatan semua daerah yang penting dalam data yang tersembunyi [MOR-07].

Data mining dapat didefinisikan sebagai proses otomatis mencari ukuran pola-pola data yang besar. Data mining menggunakan proses pencarian melalui ukuran data yang besar menggunakan teknik *clustering* (*K-means*, *Fuzzy K-means*, *Subtractive*) untuk memperoleh data yang relevan dan signifikan dalam pengenalan pola, dan logika *fuzzy* dari *fuzzy* inferensi sistem (Mamdani dan TSK) merupakan teknik berbasis untuk mengesktraksi pembuatan aturan-aturan (IF-THEN). Dalam penelitian tersebut, Moreno et al (2007) meneliti mengenai *clustering* dan ekstraksi aturan dengan dua metode. Yang pertama menangani subjek *clustering* dan ekstraksi menggunakan aturan Mamdani *Fuzzy Inference System*. Dan bagian yang kedua menyajikan *clustering* dan ekstraksi menggunakan aturan Takagi-Sugeno-Kang (TSK) [MOR-07].

Dalam skripsi ini, akan digunakan *clustering* dan esktraksi menggunakan aturan Takagi-Sugeno-Kang dengan beberapa langkah yaitu [MOR-07]:

- Langkah yang pertama yaitu mengelompokkan inputan nilai (*data training*) menggunakan teknik *K-Means clustering*, dimana nantinya digunakan untuk menghitung dan menentukan pusat *cluster* ( $c$ ) dan standar deviasi ( $\sigma$ ).
- Langkah yang kedua yaitu menghitung *consequent* dengan menggunakan inferensi *fuzzy* Sugeno dimana nantinya hasil dari pencarian *consequent* ini berupa derajat keanggotaan dari data uji. Pada metode *K-means clustering*, data ujinya berupa data berbentuk crisp set, sehingga untuk mengujinya diperlukan perubahan data dengan menggunakan fungsi Gaussian (persamaan 2.17).

$$gaussian(x; c; \sigma) = e^{-\frac{1}{2} \left( \frac{x-c}{\sigma} \right)^2}$$

dimana :  $x$  = data input uji  
 $c$  = pusat *cluster*  
 $\sigma$  = standar deviasi

Yang nantinya perhitungan menggunakan fungsi Gaussian tersebut akan menghasilkan derajat keanggotaan masing-masing untuk setiap data uji. Sedangkan untuk menghitung koefisien-koefisien *output* dicari dengan menggunakan metode *least square* yang sudah dijelaskan pada bagian 2.4.4.5 dan hasilnya akan terbentuk matrik koefisien *output*.

- Langkah yang ketiga yaitu mengekstrak aturan *fuzzy* berdasarkan anteseden nilai fungsi keanggotaannya. Misal, if  $\mu^i (\mu_1^i, \mu_3^i, \dots, \mu_n^i)$  adalah pusat *cluster*, maka aturan *fuzzy* yang terbentuk adalah :

$$R = \text{IF } x_1 \text{ is } \mu_1^i \text{ and } x_3 \text{ is } \mu_3^i \text{ and } \dots x_{n-i} \text{ is } \mu_{n-i}^i \text{ THEN } x_n \text{ is } \mu_n^i$$

Bisa dilihat susunan aturan rule yang terbentuk bisa diartikan bahwa setiap dimensi merupakan variabel *fuzzy*. Kemudian dicari  $\alpha$ -predikatnya untuk setiap aturan yang terbentuk menggunakan product sebagai operator pada antesedennya.

$$\alpha_i = \mu_1^i(\hat{x}_1) * \dots * \mu_j^i(\hat{x}_j) * \dots * \mu_n^i(\hat{x}_n) \forall i = 1, 2, 3, \dots, dst \quad (2.20)$$

- Langkah yang terakhir yaitu menghitung nilai  $Z$  dengan cara sebagai berikut:

$$z = p_1 * x_1 + \dots + p_N * x_N + q \quad (2.21)$$

setelah  $z$  ditemukan, dilanjutkan proses *defuzzyfikasi* dengan menggunakan rata-rata terbobot (*weighted average*)

$$Z = \frac{\alpha_1 z_1 + \alpha_2 z_2 + \dots + \alpha_n z_n}{\alpha_1 + \alpha_2 + \dots + \alpha_n} \quad (2.22)$$

### 2.15 Perhitungan Akurasi

Hasil yang diamati pada penelitian ini adalah seberapa tingkat akurasi metode *K-means clustering* dalam membangkitkan aturan *fuzzy* dalam pengujian data ketahanan hidup penderita kanker payudara. Dalam penelitian ini akurasi dihitung dari jumlah uji ketahanan hidup yang tepat dibagi dengan jumlah data. Akurasi merupakan seberapa dekat suatu angka hasil pengukuran terhadap angka sebenarnya (*true value* atau *reference value*). Tingkat akurasi diperoleh dengan perhitungan sesuai dengan persamaan sebagai berikut [RAM-06].

$$\text{Tingkat akurasi} = \frac{\sum \text{dat uji benar}}{\sum \text{total data uji}} \quad (2.23)$$

$$\text{Akurasi} = \frac{\sum \text{data uji benar}}{\sum \text{total data uji}} \times 100\% \quad (2.24)$$

### BAB III

#### METODOLOGI DAN PERANCANGAN SISTEM

Pada Bab III ini akan dijelaskan mengenai metode langkah-langkah yang nantinya dilakukan dalam penelitian dengan menggunakan metode *K-Means clustering* dan *Fuzzy Inference System Sugeno* untuk menguji data ketahanan hidup penderita kanker payudara serta menjelaskan rancangan sistem yang akan digunakan. Adapun tahap-tahapan yang dilakukan dalam penelitian sebagai berikut :

1. Studi Literatur

Mencari dan mempelajari literatur mengenai pembangkitan aturan menggunakan metode *K-Means clustering* dan *Fuzzy Inference System Sugeno* serta mempelajari literatur-literatur yang terkait dengan ketahanan hidup penderita kanker payudara. Studi literatur dilakukan untuk mendukung penelitian serta meningkatkan pemahaman terkait permasalahan yang diangkat dan ingin dicari penyelesaian masalah tersebut.

2. Pengumpulan Data

Mengumpulkan data penderita kanker payudara yang diambil dari data *Haberman's Survival* yang bersumber pada:

(<http://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival>)

yang merupakan data dari penelitian di University of Chicago's Billings Hospital.

3. Analisa dan Perancangan Sistem

Menganalisis dan merancang sistem yang meliputi perancangan proses algoritma yang digunakan, tabel data, perancangan antar muka (GUI), dan uji coba.

4. Implementasi Sistem

Mengimplementasikan hasil analisa dan perancangan sistem dengan membangun perangkat lunak sebagai media input pengujian data.

5. Pengujian Sistem

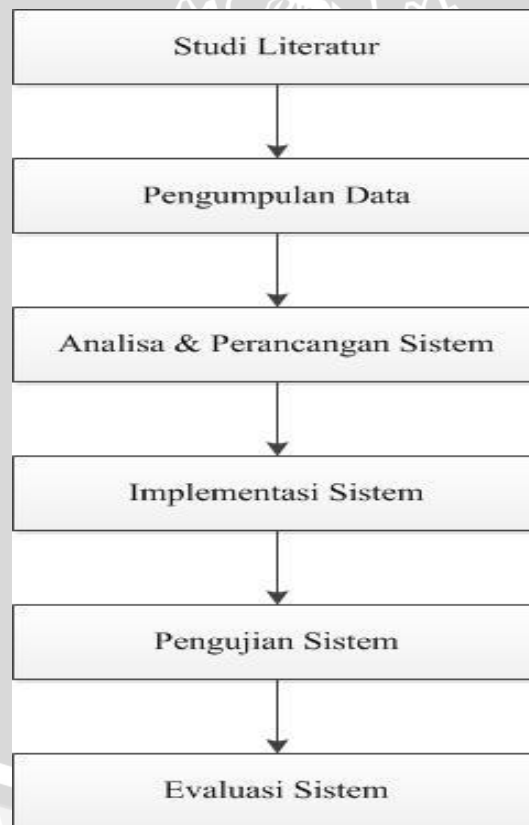
Pada proses ini, akan dilakukan pengujian data terhadap sistem perangkat lunak yang dibangun dengan menggunakan metode *K-Means clustering* dan

menginputkan data sebagai uji coba data dalam menguji keakuratan rule dalam menfklasifikasi (dengan menggunakan *Fuzzy Inferensi Sugeno*) yang dihasilkan dari proses pembangkitan aturan menggunakan metode *K-Means clustering*.

#### 6. Evaluasi Sistem

Output hasil dari pengujian data menggunakan perangkat lunak selanjutnya dianalisis dan dievaluasi keakuratan hasil uji data menggunakan pembangkitan aturan hasil dari metode *K-Means* yang telah diimplementasikan ke dalam sistem.

Berikut ini bagan alur yang digunakan dalam penelitian penggunaan metode *K-Means clustering* untuk pembangkitan aturan pada kasus ketahanan hidup penderita kanker payudara :



Gambar 3.1 Alur Penelitian

### 3.1 Studi Literatur

Pada proses studi literatur ini, dilakukan proses mempelajari literatur-literatur yang digunakan sebagai acuan terutama yang berkaitan dengan metode



dan algoritma yang digunakan (metode *K-Means clustering*, *Fuzzy Inferensi Sistem Sugeno*, dll), serta yang terkait dengan kasus ketahanan hidup penderita kanker payudara. Tujuan dilakukannya studi literatur ini yaitu untuk meningkatkan pemahaman serta mendukung proses penelitian. Dalam hal ini literatur yang digunakan sebagai acuan bisa diambil dari berbagai sumber, seperti buku, jurnal, *e-book*, internet, serta sumber-sumber lain yang bisa menjadi acuan dan dapat memberi informasi pengetahuan mengenai topic penelitian.

### 3.2 Data Penelitian

Data yang digunakan dalam penelitian ini yaitu data ketahanan hidup penderita kanker payudara (*Haberman's Survival*) yang telah menjalani operasi dalam kurun waktu antara 1958-1970. Data ini diambil dari hasil penelitian di University of Chicago's Billings Hospital, dan data ini disumbangkan oleh Tjen-Sien Lim. Data ini digunakan untuk memprediksi apakah penderita mampu bertahan hidup dalam kurun waktu lima tahun atau lebih atau penderita tidak mampu bertahan hidup kurang dari lima tahun.

Data Haberman's Survival ini terdiri dari 306 data yang mempunyai empat atribut yaitu umur (umur penderita ketika melakukan operasi), tahun (tahun penderita ketika melakukan operasi, node (jumlah node yang positif terdeteksi kanker), dan status ketahanan hidup (yang merupakan dari kelas penentu apakah penderita dapat bertahan hidup atau tidak).

### 3.3 Analisa dan Perancangan Sistem

#### 3.3.1 Deskripsi Umum Sistem

Secara garis besar, sistem terdiri dari dua proses sistem utama. Yaitu proses pembangkitan aturan *fuzzy* (terdiri dari proses *clustering*, perhitungan standar deviasi, sampai dihasilkan nilai ideal dari *cluster* tersebut), dan proses pengujian keakuratan klasifikasi data menggunakan *Fuzzy Inferensi Sistem Sugeno* dengan rule dari hasil pembangkitan aturan metode *K-Means clustering*. Dari uraian gambaran secara umum sistem, nantinya diharapkan hasil pembangkitan aturan nanti bisa digunakan untuk memprediksi dan menguji data suatu kasus yang mana tidak lagi tergantung dengan pakar dalam penentuan rule.

Adapun uraian proses-proses dalam membangun sistem yaitu :

1) Proses Pelatihan

Yakni proses dimana mengelompokkan data menggunakan metode *K-Means clustering*. Dimana proses pengelompokkan ini membutuhkan beberapa data set yang digunakan sebagai data training dalam proses pembangkitan aturan. Pada proses *clustering* bisa dikelompokkan dari dua sampai  $n-1$  *cluster* yang mana jumlah *cluster* yang nantinya digunakan sama dengan jumlah aturan yang akan dibangkitkan. Pada proses pengelompokkan data ini nantinya akan dihitung juga standar deviasi, varian, serta akan dihitung juga nilai kelayakan dari *cluster* sehingga akan diketahui jumlah *cluster* yang ideal yang akan dibentuk dengan membandingkan nilai ideal dengan jumlah *cluster* yang berbeda.

Dari hasil perhitungan standar deviasi dan perhitungan nilai ideal suatu cluster, dilanjutkan dengan menghitung masing-masing nilai derajat keanggotaan dari data training yang hasilnya nanti digunakan dalam proses LSE untuk mendapatkan nilai koefisien output dari masing-masing aturan sebelum melanjutkan pada proses pengujian data menggunakan *Fuzzy Inferensi Sistem Sugeno Orde-Satu*.

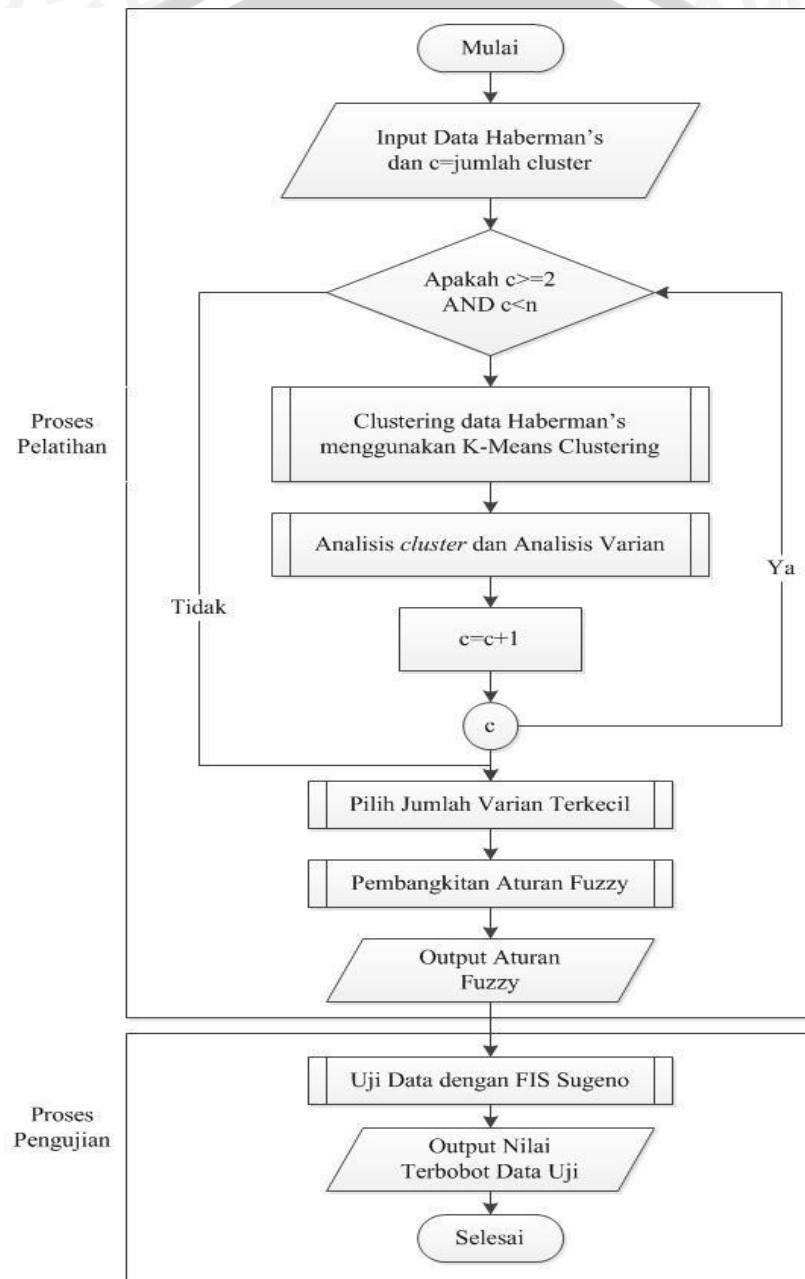
2) Proses Pengujian

Dari proses pengelompokkan, pembangkitan aturan, sampai dihasilkan nilai koefisien output dari masing-masing aturan. Dilanjutkan proses menghitung keakuratan rule dari hasil pembangkitan aturan dengan menguji beberapa data set untuk diklasifikasi menggunakan *Fuzzy Inferensi Sistem Sugeno* dengan menghitung nilai derajat keanggotaan pada setiap atribut menggunakan fungsi Gauss di setiap data uji. Dilanjutkan menghitung nilai *fire strength* ( $\alpha$ -predikat) dan  $z$  di masing-masing aturan yang terbentuk. Pada proses tahap akhir dihitunglah *wighted average* dengan menggunakan  $\alpha$ -predikat dan  $z$ , yang mana nilai bobot ini akan digunakan sebagai acuan apakah data tergolong pada data dengan kelas bertahan atau tidak. Kemudian dari hasil data pengujian, data hasil uji dicocokkan dengan data hasil penelitian pada data Haberman's Survival.

Apakah data hasil uji sistem dan hasil penelitian mempunyai hasil yang sama atau tidak.

### 3.3.2 Perancangan Proses

Pada perancangan proses, terdiri dari dua proses utama yaitu proses pelatihan dan pengujian. Masing-masing proses akan diuraikan dengan flowchart pada Gambar 3.2.



Gambar 3.2 Alur Perancangan Sistem

Pada proses pelatihan terdiri dari empat proses utama, yaitu proses *clustering* menggunakan *K-Means* data Haberman's Survival, proses analisis *cluster* dan analisis varian, proses pemilihan *cluster* dengan jumlah varian terkecil, proses pembangkitan aturan *fuzzy* yang nantinya akan menghasilkan *output* berupa aturan *fuzzy* yang nanti digunakan untuk menguji keakuratan klasifikasi. Sedangkan pada proses pengujian data sendiri terdapat satu proses pengujian data yaitu proses ujia data dengan *Fuzzy Inference System* Sugeno.

Adapun tahap-tahapan secara keseluruhan dari proses sistem nantinya adalah sebagai berikut :

1. Input data ketahanan hidup penderita kanker payudara (data Haberman's Survival) yang digunakan sebagai data latih untuk menghasilkan aturan. Adapun atribut pendukung (parameter input data) yang digunakan antara lain umur penderita, tahun operasi, jumlah node positif, dan status penderita.
2. Dilanjutkan proses pengelompokkan (*clustering*) data Haberman's menggunakan metode *K-Means clustering*. Tentukan jumlah *cluster* yang nantinya ingin dijadikan sebagai acuan jumlah kelompok data. Banyaknya jumlah *cluster* akan menentukan jumlah aturan yang dibangkitkan. Karena jumlah *cluster* sama dengan jumlah rule. Keakuratan *clustering* sangat menentukan keakuratan hasil dari pengujian sistem. Maka dari itu nantinya perlu dihitung jumlah nilai ideal *cluster* yang layak digunakan. Dari proses *clustering* sendiri nantinya akan dihasilkan pusat *cluster* baru (*centroid*) yang nantinya digunakan untuk menghitung mean dan standar deviasi masing-masing *cluster*.
3. Selanjutnya dilakukan proses *clustering* dengan jumlah *cluster* yang bervariasi. Analisis dan hitung jumlah *cluster* yang paling ideal dari masing-masing jumlah *cluster*.
4. Dari proses perhitungan analisis varian masing-masing *cluster* dihasilkan nilai varian. Dan nilai varian *cluster* terkecil yang akan dipilih sebagai *cluster* yang paling ideal.
5. Tahapan berlanjut ke proses pembangkitan aturan *fuzzy* dari jumlah *cluster* dan *centroid* yang didapatkan. Seperti yang sudah dibahas jumlah aturan yang terbentuk sama dengan jumlah *cluster* yang ditentukan.

6. Menghitung derajat keanggotaan masing-masing data latih dengan menggunakan fungsi Gauss. Dilanjutkan proses pembentukan matriks U dan normalisasi matriks U. Sampai dilanjutkan ke proses perhitungan koefisien output masing-masing rule menggunakan LSE.
7. Setelah koefisien output masing-masing rule telah ditentukan. Proses berlanjut ke dalam pengujian data. Dimana data set yang lain (data Haberman's Survival yang tidak dijadikan sebagai data latih) digunakan sebagai data uji sistem.
8. Data uji diinputkan, dan sistem menguji data uji menggunakan *Fuzzy Inference System Sugeno* Orde-Satu dan mencari nilai terbobot tiap data uji pada masing-masing aturan.
9. Dari proses pengujian data, didapatkan hasil nilai output berupa nilai crisp yang mana nilai tersebut menentukan apakah data tersebut tergolong ke dalam status bertahan hidup atau sebaliknya. Adapun alur dari perhitungan dan perancangan proses bisa dilihat pada Gambar 3.2.

### 3.3.2.1 Proses Pelatihan

Tahapan proses pelatihan data terdiri dari beberapa tahapan, antara lain yaitu proses *clustering*, menghitung mean dan standar deviasi, analisis varian, pemilihan jumlah *cluster* dengan varian terkecil, pembangkitan aturan, menghitung derajat keanggotaan, proses perhitungan koefisien output, proses normalisasi, dan proses perhitungan LSE. Untuk langkah-langkah proses pelatihnnya diuraikan sebagai berikut :

#### a) Proses Clustering

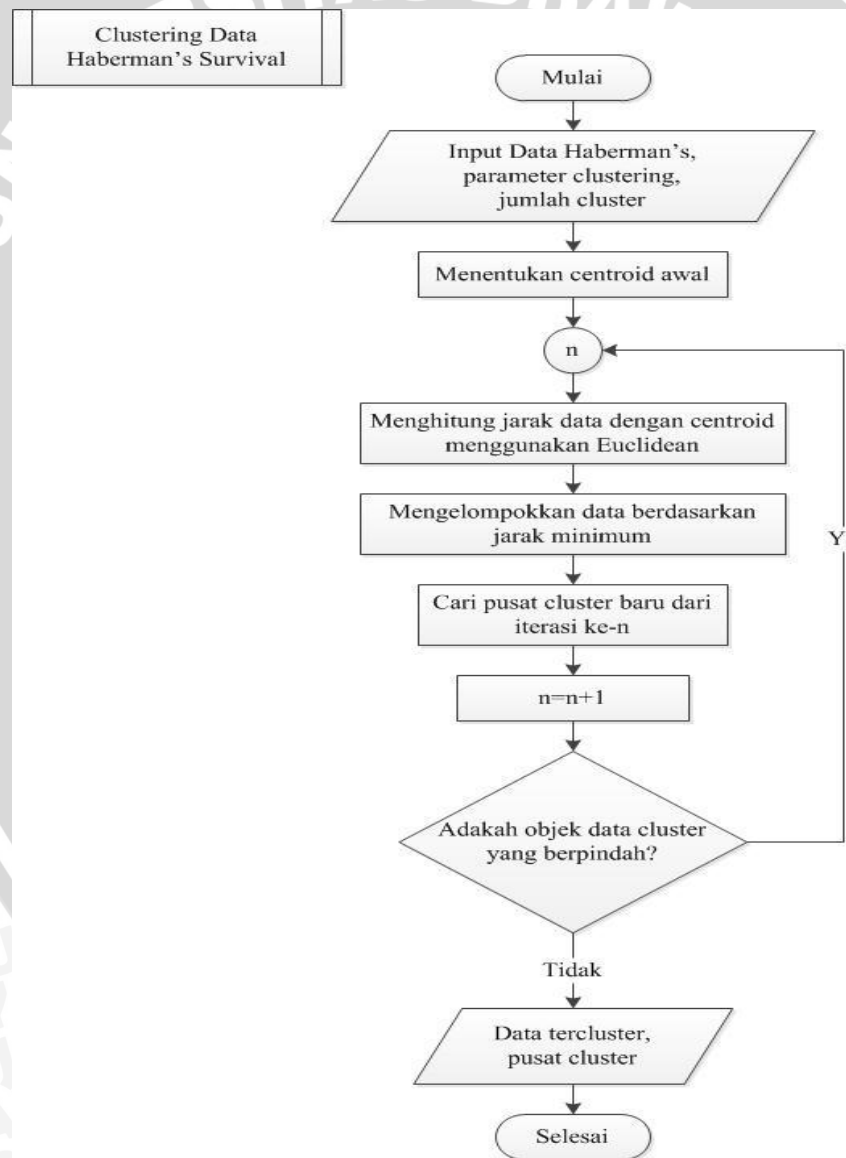
Alur tahapan proses pengelompokkan data Haberman's Survival menggunakan *K-Means clustering* adalah sebagai berikut :

1. Input data haberman's, jumlah *cluster*.
2. Tentukan centroid awal secara random sejumlah banyaknya cluster.
3. Menghitung jarak *cluster* masing-masing data latih terhadap centroid dengan menggunakan rumus *euclidian distance* seperti persamaan 2.1.
4. Tentukan anggota tiap *cluster* masing-masing data, dengan melihat jarak masing-masing data dari hasil perhitungan menggunakan

*Euclidean*. Hasil perhitungan jarak terkecil di semua *cluster* menentukan data tersebut masuk ke dalam *cluster* mana.

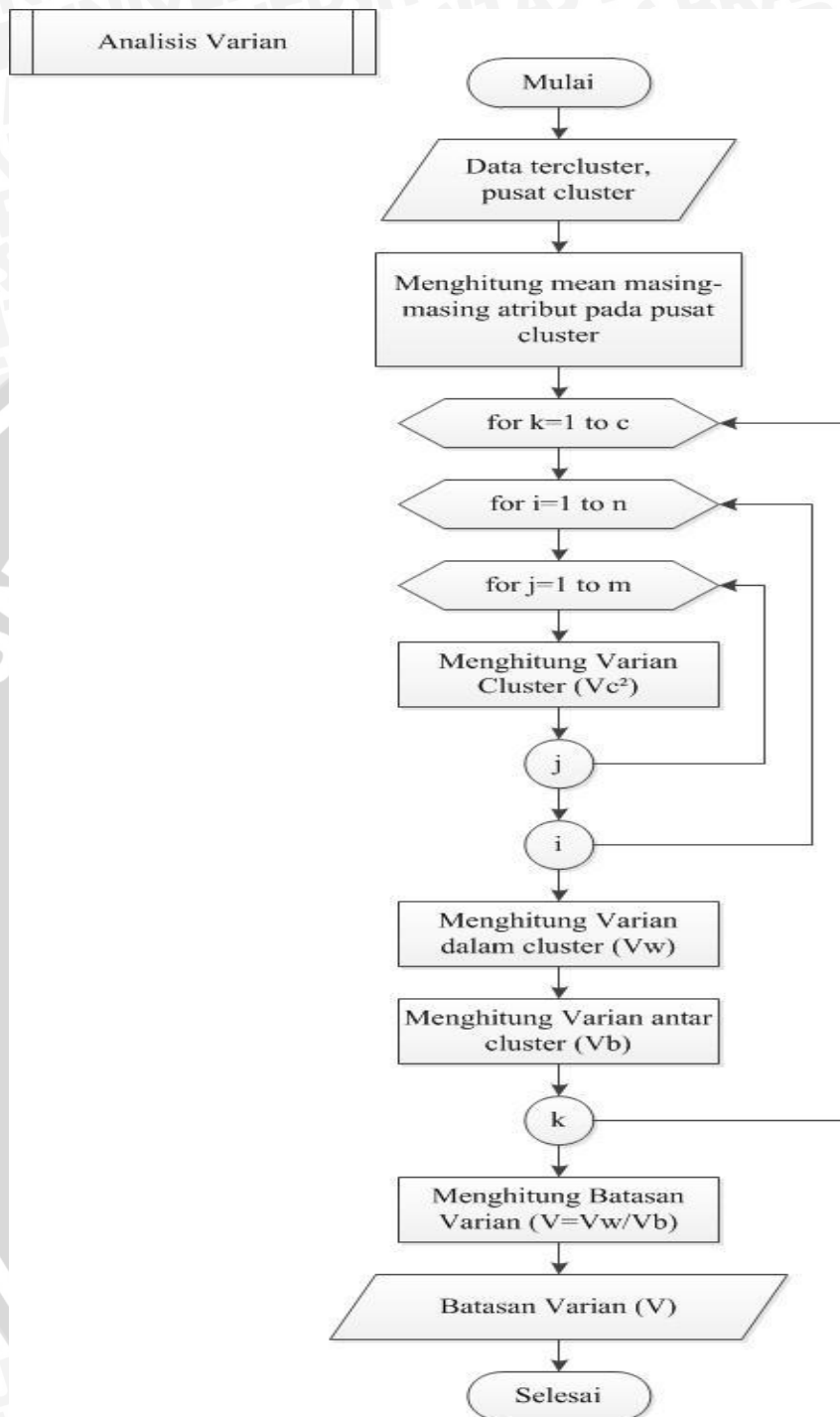
5. Setelah ditentukan tiap data masuk ke dalam suatu *cluster*, kemudian dicari centroid terbaru tiap *cluster* dengan menghitung mean masing-masing *cluster*.
6. Lakukan proses perhitungan pada tahapan ke 3-5 sampai anggota tiap *cluster* tidak berubah. Dengan tidak adanya anggota *cluster* yang berubah maka proses iterasi berakhir.

Proses pengelompokan digambarkan dengan flowchart pada Gambar 3.3.



Gambar 3.3 Alur proses pengelompokan data dengan *K-Means clustering*

b) Analisis Varian



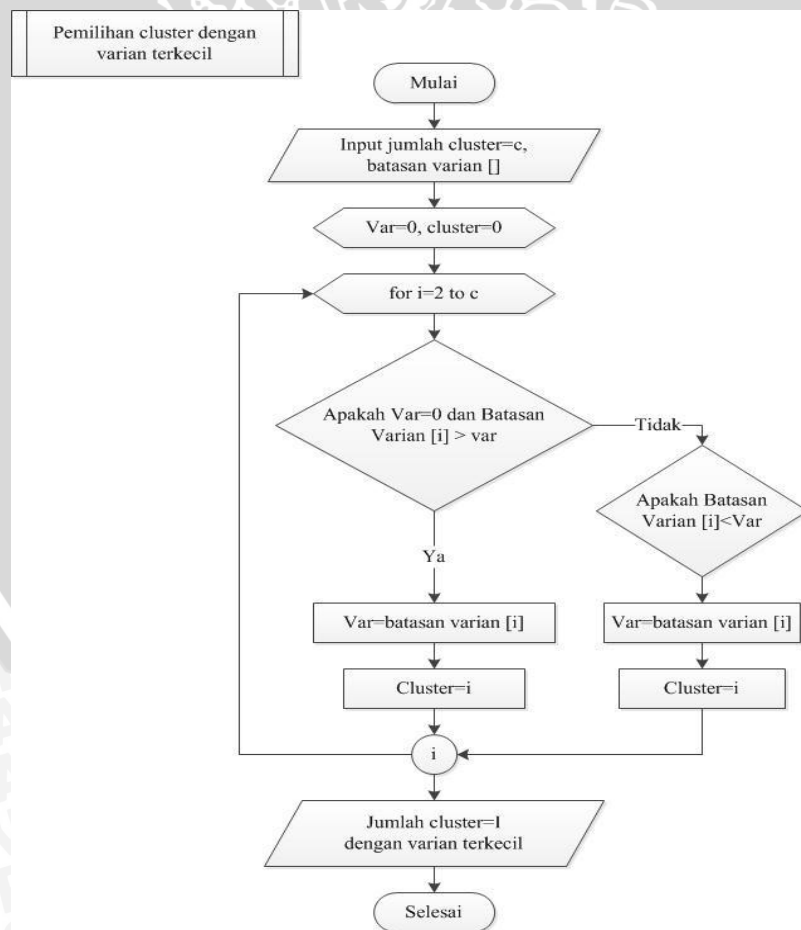
Gambar 3.4 Alur proses perhitungan batasan varian

Setelah proses *clustering* berakhir, proses dilanjutkan dengan tahapan perhitungan analisa varian. Tahapan dalam menganalisa varian sebagai berikut :

1. Menghitung rata-rata nilai pusat cluster berdasarkan atribut.
2. Menghitung nilai varian untuk setiap *cluster* dengan menggunakan persamaan 2.4, yang hasilnya nanti akan digunakan untuk membantu perhitungan *variance within cluster* ( $V_w$ ).
3. Menghitung  $V_w$  dengan menggunakan persamaan 2.5.
4. Menghitung *variance between cluster* ( $V_b$ ) dengan menggunakan persamaan 2.6.
5. Dari hasil proses perhitungan  $V_w$  dan  $V_b$ , dihitung batasan varian menggunakan persamaan 2.7.

Dari hasil perhitungan varian inilah kita bisa mengetahui *cluster* ideal dari beberapa perhitungan varian dengan jumlah *cluster* yang bervariasi dengan memilih hasil perhitungan varian terkecil. Alur tahapan proses dari analisis varian akan dijelaskan dengan Gambar 3.4.

**c) Memilih Jumlah Cluster dengan Varian Terkecil**



Gambar 3.5 Alur proses pemilihan *cluster* dengan varian terkecil



Tahapan proses pemilihan jumlah *cluster* dengan varian terkecil yaitu sebagai berikut :

1. Inputkan batasan varian dari jumlah *cluster*  $c$  yang dibangkitkan.
2. Inisialisasi awal nilai var dan cluster masing-masing dengan nilai awal 0.
3. Lakukan proses perulangan dengan mengecek dari *cluster* 2 hingga  $c$  *cluster*. Apakah batasan varian *cluster* ke- $i$  lebih besar dari nilai var dan apakah var sama dengan 0. Jika iya maka akan dilakukan proses dimana nilai var = batasan varian ke- $i$  dan cluster =  $i$ , jika tidak maka nilai var = batasan varian ke- $i$  dan cluster =  $i$  melanjutkan iterasi menuju ke konektor  $i$ .
4. Setelah selesai akhir iterasi maka akan didapatkan jumlah cluster dengan varian terkecil.

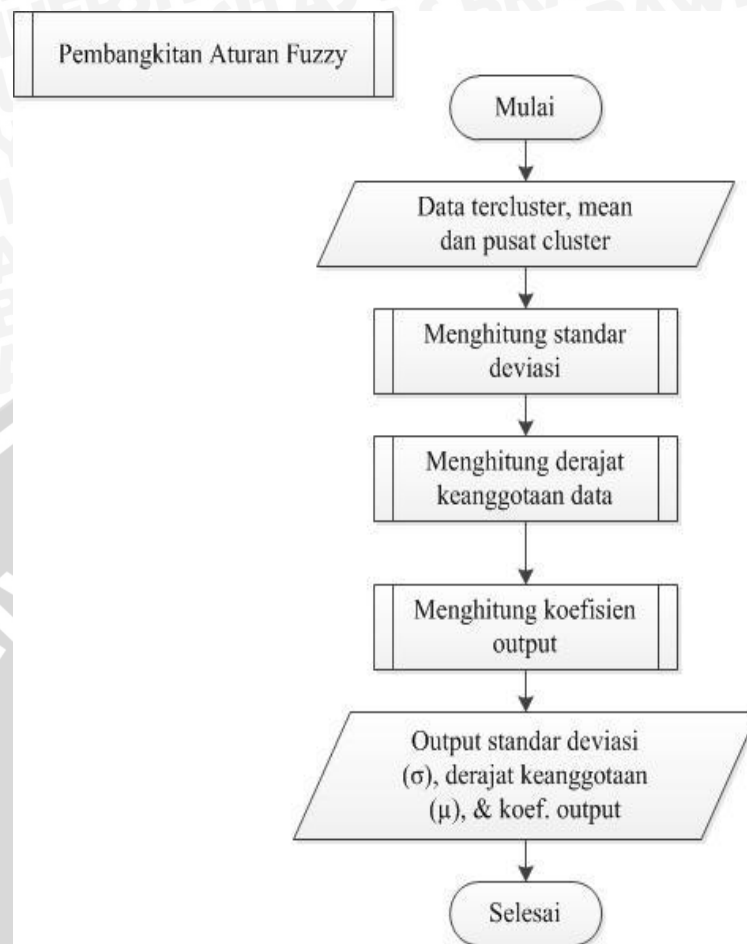
Proses tahapan pemilihan jumlah cluster ideal dengan varian terkecil akan digambarkan pada alur flowchart Gambar 3.5

#### d) Pembangkitan Aturan

Tahap pembangkitan aturan *fuzzy* sebagai berikut :

1. Hasil dari pengelompokan dan perhitungan *clustering* yaitu berupa centroid yang selanjutnya digunakan untuk menghitung standar deviasi dengan menggunakan persamaan 2.3.
2. Hasil dari perhitungan standar deviasi ini digunakan untuk menghitung derajat keanggotaan masing-masing data dan nantinya akan digunakan untuk menghitung koefisien output untuk setiap rule yang dibangkitkan. Dan nantinya hasil dari perhitungan koefisien output tersebut digunakan untuk menghitung nilai uji dari data yang menghasilkan nilai crisp dari nilai kategori data penderita kanker payudara.

Alur proses tahapan pembangkitan aturan *fuzzy* bisa dilihat pada Gambar 3.6.

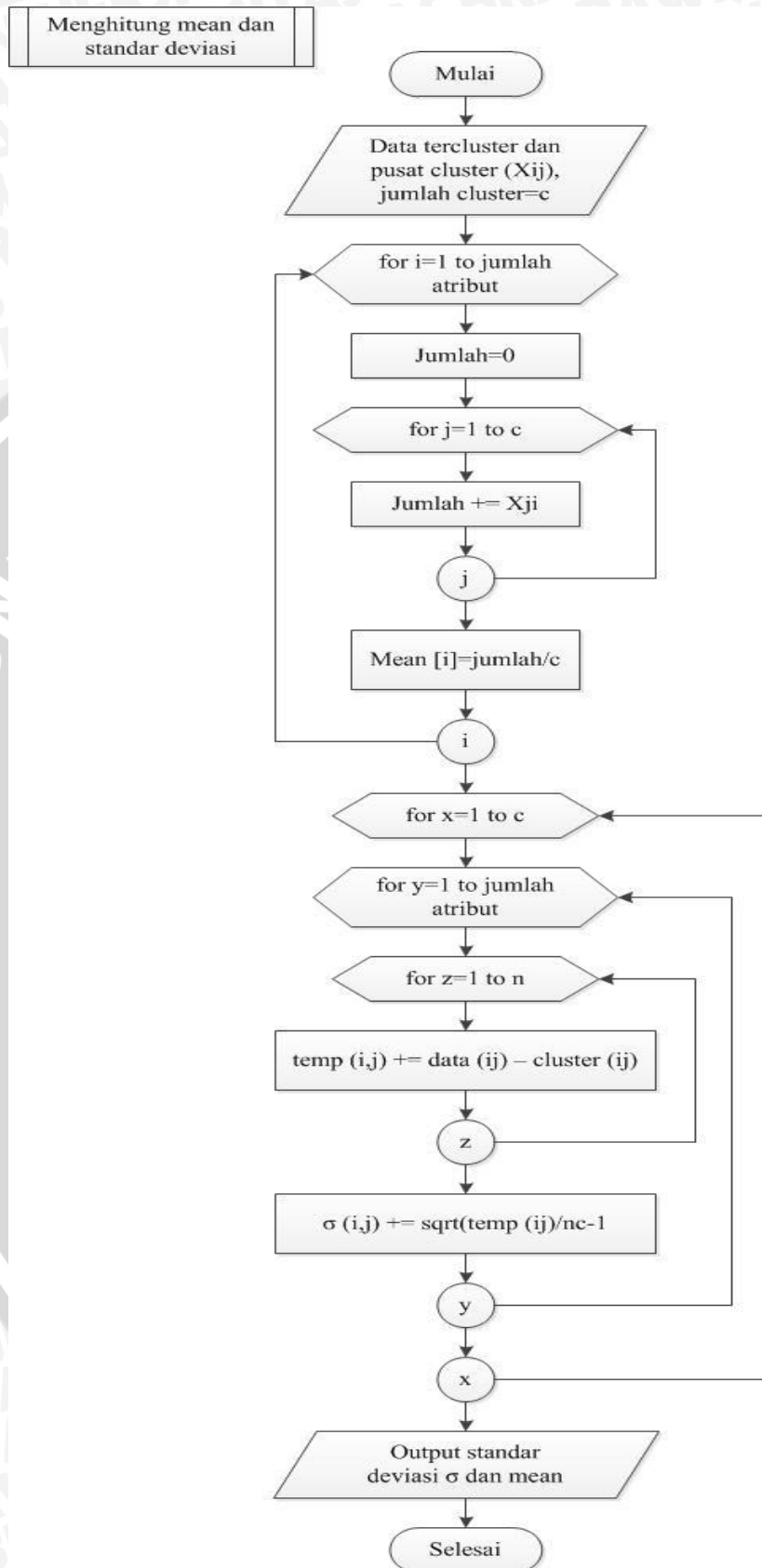


Gambar 3.6 Alur proses pembangkitan aturan *fuzzy*

#### e) Menghitung Mean dan Standar Deviasi

Untuk menghitung mean dan standar deviasi, perlu dilakukan tahapan sebagai berikut :

1. Input yang digunakan yaitu data Haberman's Survival yang sudah di *cluster* berupa pusat *cluster*.
2. Hitung mean data berdasarkan atribut tiap *cluster*-nya dengan menggunakan persamaan 2.2.
3. Hitung standar deviasi menggunakan persamaan 2.3.
4. Hasil dari perhitungan mean dan standar deviasi berguna untuk menghitung derajat keanggotaan masing-masing data dan nantinya akan digunakan untuk menghitung koefisien output untuk setiap rule yang dibangkitkan.



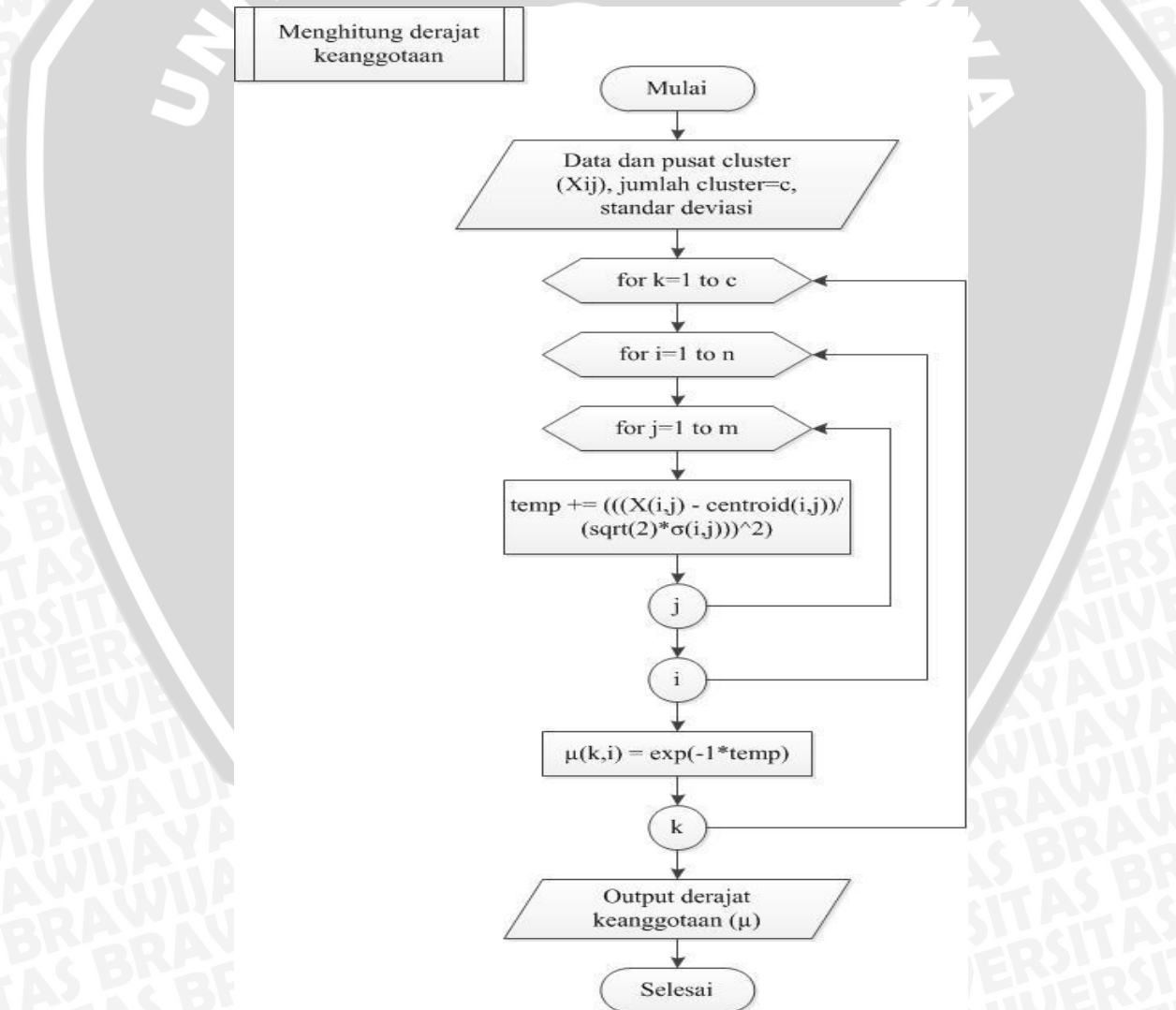
Gambar 3.7 Alur proses menghitung mean dan standar deviasi

**f) Menghitung Derajat Keanggotaan**

Tahapan menghitung derajat keanggotaan akan memerlukan inputan dari proses-proses sebelumnya, secara lengkap tahapan sebagai berikut :

1. Inputan yang digunakan yaitu centroid akhir dari proses *clustering*, standar deviasi, dan data latih Haberman's Survival.
2. Menghitung derajat keanggotaan masing-masing data menggunakan persamaan 2.17 dilakukan sebanyak iterasi yang ada.
3. Hasil output dari perhitungan yaitu derajat keanggotaan tiap *cluster* sejumlah n data.

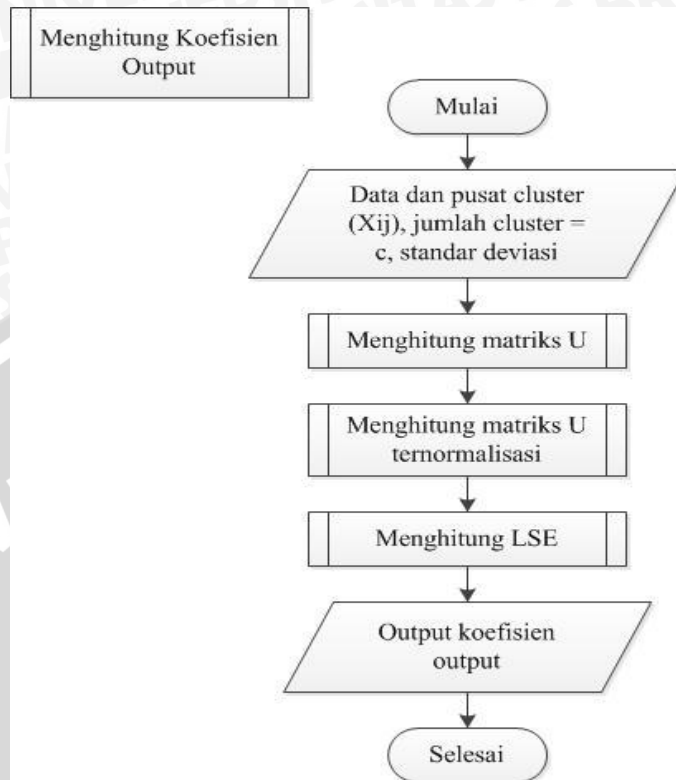
Untuk alur flowchart akan ditunjukkan pada Gambar 3.8.



Gambar 3.8 Alur proses menghitung derajat keanggotaan ( $\mu$ )



## g) Perhitungan Koefisien Output



Gambar 3.9 Alur proses menghitung koefisien output

Tahapan proses perhitungan koefisien output membutuhkan input berupa data latih  $X_{ij}$ , mean, dan derajat keanggotaan  $\mu_{ik}$ . Hasil outputnya berupa koefisien output tiap rule. Proses perhitungan koefisien output sendiri ada tiga tahapan, yaitu :

## 1. Normalisasi

Tahapan ini menormalisasi data dengan bentuk matriks dari hasil perhitungan derajat keanggotaan tiap *cluster* dan jumlah  $\mu_{ik}$ .

## 2. Pembentukan matriks U

Proses ini membentuk matriks  $n \times m$  yang berisi data hasil normalisasi. Dimana  $n$  merupakan jumlah data dan  $m$  sebesar jumlah atribut dikalikan jumlah aturan yang terbentuk.

3. Perhitungan LSE (*Least Square Error*)

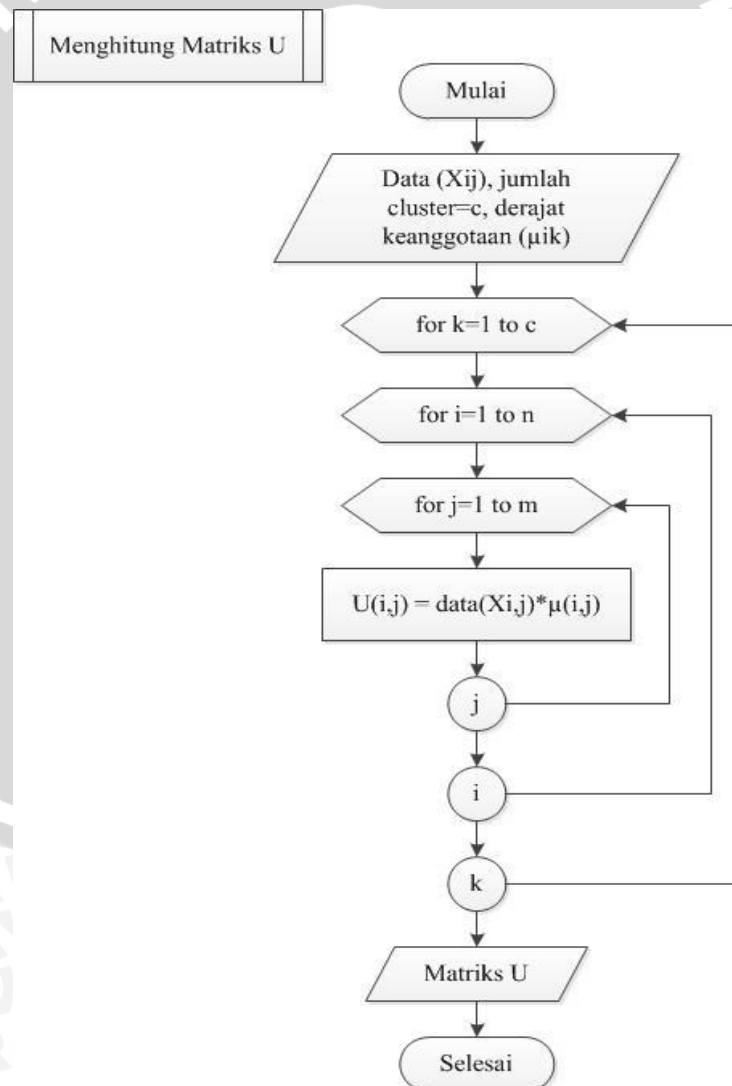
Tahapan ini berguna untuk menghitung koefisien output untuk tiap rule yang dibangkitkan, untuk menentukan nilai output hasil pengujian data.

### h) Proses Membentuk Matriks U

Proses pembentukan matriks U ini membutuhkan inputan berupa data ( $X_{ij}$ ), derajat keanggotaan. Dan hasil dari matriks U ini sendiri nantinya digunakan untuk menghitung matriks U ternormalisasi. Langkah-langkahnya sebagai berikut :

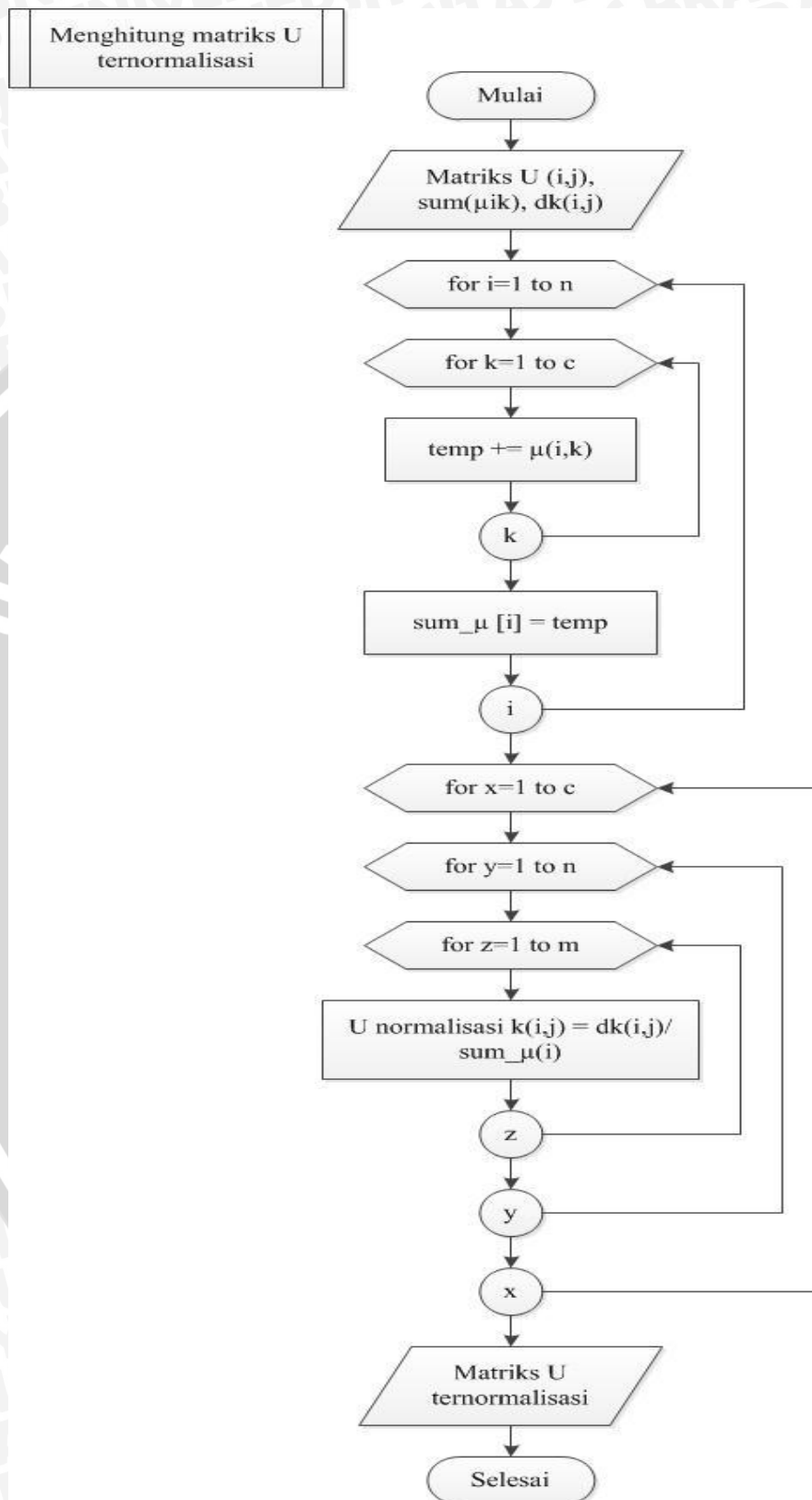
1. Mengalikan derajat keanggotaan dengan setiap data ( $X_{ij}$ ) pada masing-masing cluster.
2. Lakukan pembentukan matriks U dari langkah ke-1 sebanyak jumlah data ( $n$ ) x jumlah atribut ( $m$ ) dikali aturan yang dibentuk.

Proses pembentukan matriks U sendiri akan ditunjukkan pada Gambar 3.10.



Gambar 3.10 Alur proses pembentukan matriks U

i) Proses Normalisasi



Gambar 3.11 Alur proses pembentukan matriks U ternormalisasi

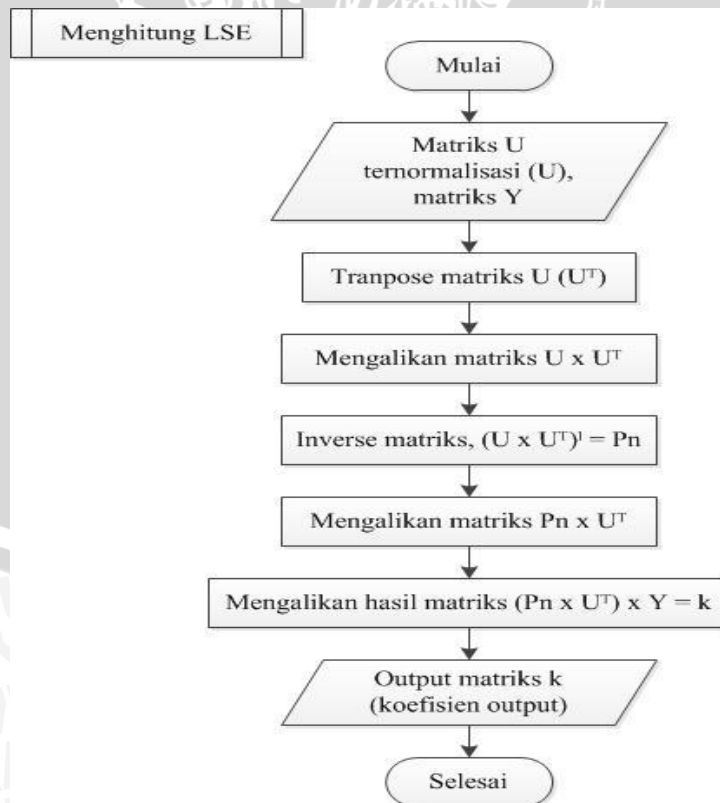
Tahapan normalisasi memerlukan inputan berupa data latih  $X_{ij}$ , pusat mean, dan derajat keanggotaan  $\mu_{ik}$ . Output hasil prosesnya berupa matriks  $d^k_{ij}$  yang berisi normalisasi data perkalian  $X_{ij}$  dengan  $\mu_{ik}$  dan data  $\mu_{ik}$ . Tahapan-tahapannya yaitu :

1. Menjumlahkan derajat keanggotaan ( $\mu_{ik}$ ) semua *cluster*.
2. Membagi  $d^k_{ij}$  dengan hasil penjumlahan derajat keanggotaan.

Proses pembentukan matriks U ternormalisasi sendiri akan ditunjukkan pada Gambar 3.11.

**j) Proses Perhitungan LSE**

Proses perhitungan LSE memerlukan input berupa matriks U dan matriks Y. dimana matriks U didapatkan dari hasil normalisasi, sedangkan matriks Y berisi atribut terakhir  $X_{ij}$  yang merupakan status ketahanan hidup penderita kanker. Perhitungan ini menghasilkan output berupa matriks k yang berisi koefisien output setiap aturan yang dibangkitkan. Perhitungan LSE sendiri menggunakan persamaan 2.12. proses alur perhitungan LSE sendiri ditunjukkan pada Gambar 3.12.



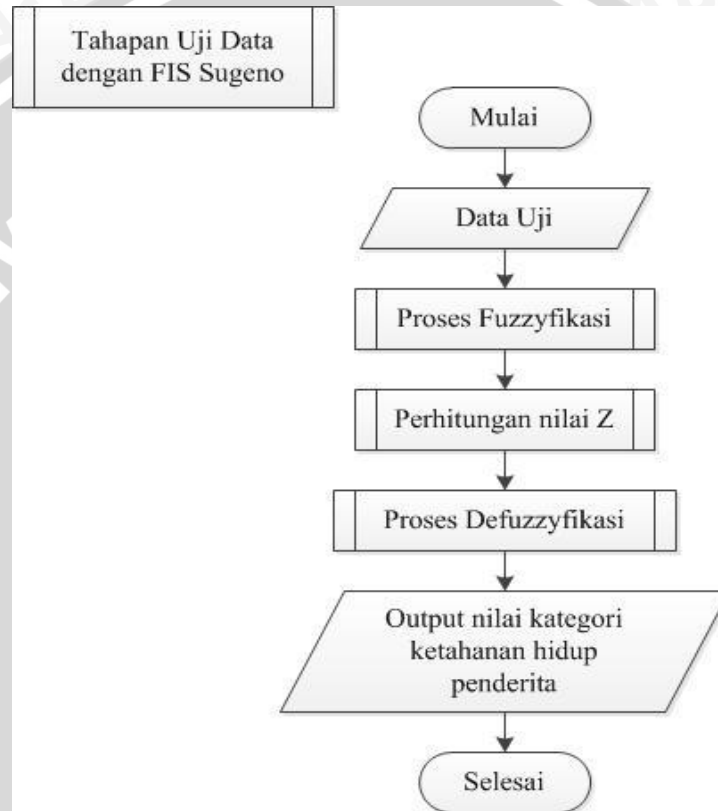
Gambar 3.12 Alur proses menghitung LSE



### 3.3.2.2 Proses Pengujian

Proses pengujian sendiri masih harus melewati beberapa tahapan yaitu : proses inferensi sistem menggunakan *Fuzzy Inferensi Sistem Sugeno*, proses fuzzifikasi, proses perhitungan  $z$ , dan proses defuzzifikasi. Adapun tahapan-tahapan tersebut akan diuraikan sebagai berikut :

#### a) Sistem Inferensi *Fuzzy Model Sugeno*



Gambar 3.13 Alur uji data menggunakan *Fuzzy Inference System Sugeno*

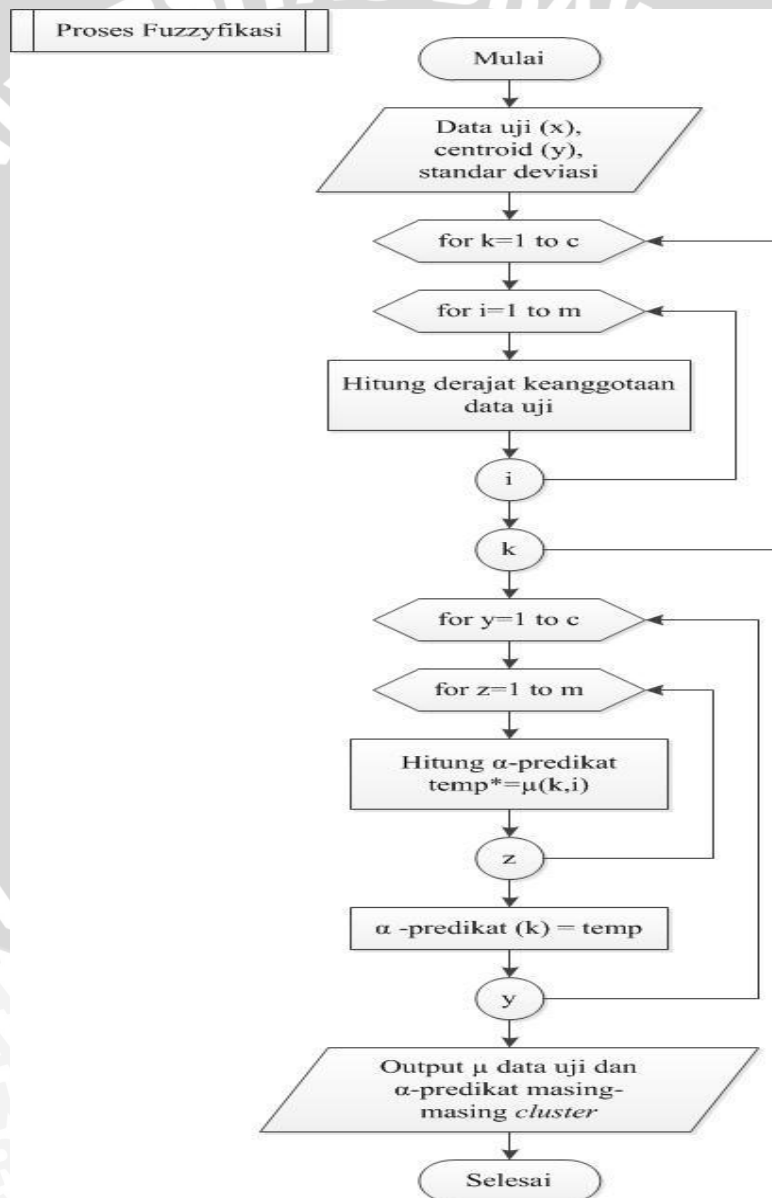
Tahapan yang bisa dikatakan terakhir yaitu proses perhitungan nilai status ketahanan hidup penderita dengan menggunakan *Fuzzy Inference System Sugeno*. Tahapan ini dikatakan sebagai tahapan pengujian keakuratan aturan dalam melakukan klasifikasi data. Adapun tahapannya sebagai berikut :

1. Menginputkan data uji Haberman's Survival (selain data yang termasuk data training).
2. Menghitung derajat keanggotaan tiap-tiap atribut input seperti pada persamaan 2.17.

3. Menghitung  $\alpha$ -predikat untuk setiap aturan dengan fungsi implikasi product sebagai operator anteseden seperti pada persamaan 2.20.
4. Tentukan nilai  $z$  untuk setiap aturan dengan menggunakan persamaan 2.21.
5. Lakukan proses defuzzifikasi dengan menggunakan metode *weighted Average* seperti pada persamaan 2.22.

Proses tahapan dalam Sistem Inferensi *Fuzzy* Sugeno sendiri ditunjukkan pada Gambar 3.13.

**b) Proses Fuzzifikasi**



Gambar 3.14 Alur proses fuzzifikasi

Fuzzifikasi adalah proses mengubah bentuk data nilai *crisp* ( $X_{uji}$ ) menjadi nilai *fuzzy* ( $\mu_{uji}$ ) berupa derajat keanggotaan. Perubahan bentuk data ini dengan menggunakan fungsi Gauss yang bersesuaian dibantu dengan centroid yang terbentuk dari proses *clustering* dan standar deviasi dari proses *clustering*. Proses ini menghasilkan output berupa *fire strength*  $\alpha_k$ . Adapun tahapannya sebagai berikut dan ditunjukkan pada flowchart Gambar 3.14:

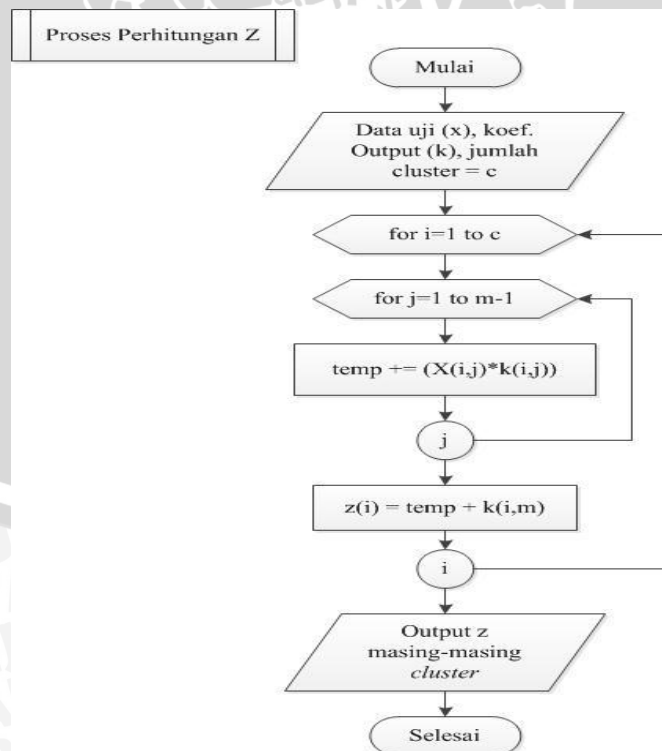
1. Menghitung Derajat Keanggotaan

Derajat keanggotaan dihitung dengan menggunakan fungsi Gauss yang telah dibangkitkan melalui proses pelatihan. Dimana masing-masing data uji dihitung derajat keanggotaannya pada masing-masing atribut di setiap *cluster* atau aturan yang terbentuk.

2. Menghitung Fire Strength

*Fire Strength* ( $\alpha$ -predikat) diperoleh dari perkalian derajat keanggotaan pada setiap data uji di setiap atribut input untuk setiap aturan atau *cluster* yang dibentuk ( $\alpha$ -predikat =  $\mu_1 * \mu_2 * \mu_3 * \dots * \mu_n$ ).

c) Proses Perhitungan Nilai Z

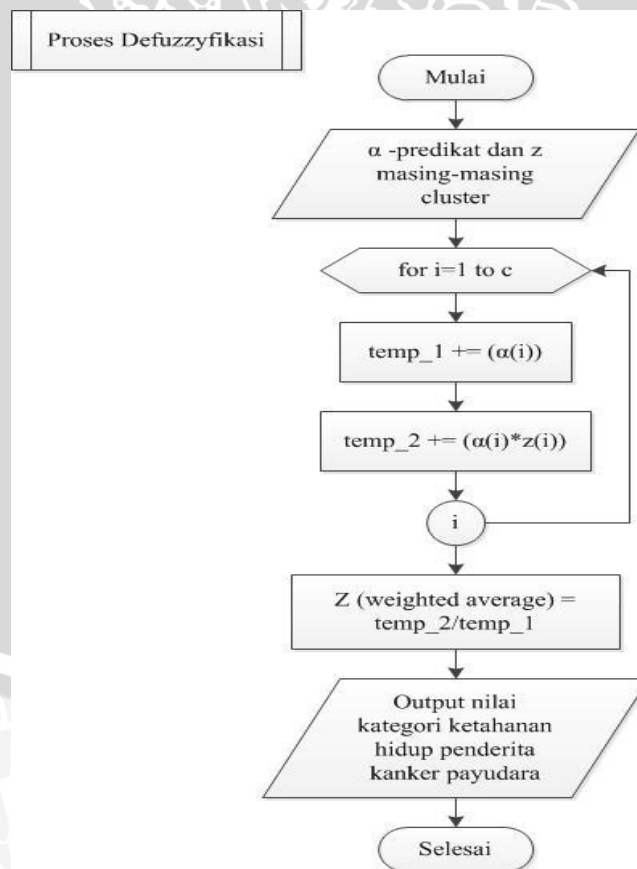


Gambar 3.15 Alur proses perhitungan z masing-masing cluster

Proses perhitungan nilai  $Z$  melibatkan input berupa koefisien output ( $k_{cj}$ ) dan data uji  $X_{uji}$ , serta menghasilkan output berupa nilai  $z$  untuk setiap aturan. Proses perhitungan nilai  $z$  ini digunakan sebagai bobot nilai dari masing-masing aturan untuk setiap data uji. Dimana proses ini menghitung total penjumlahan dari perkalian koefisien output dengan input data uji ke- $j$  untuk setiap *cluster*. Alur tahapan proses perhitungan nilai  $z$  ditunjukkan pada Gambar 3.15.

#### d) Proses Defuzzifikasi

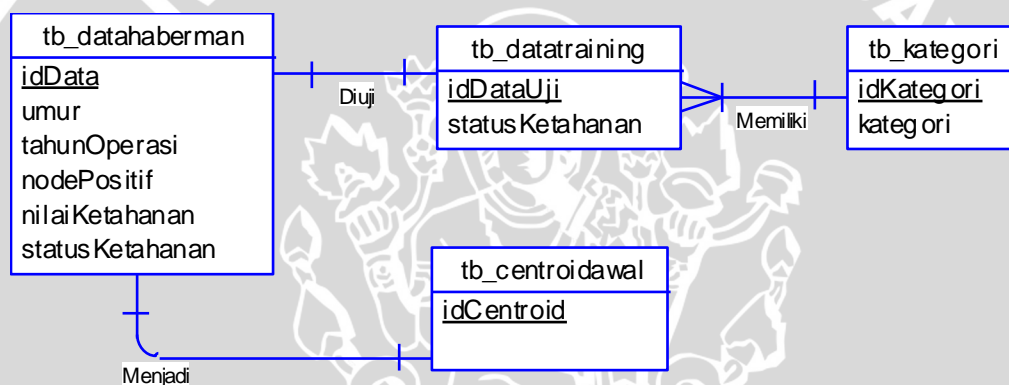
Proses defuzzifikasi melibatkan input berupa nilai  $Z$  masing-masing *cluster*, *Fire Strength* ( $\alpha$ -predikat), dan jumlah total dari  $\alpha$ -predikat yang terhitung dari semua *cluster*. Dari perhitungan ini menghasilkan output berupa nilai status ketahanan hidup penderita kanker payudara tergolong bertahan atau sebaliknya. Adapun alur tahapan perhitungan proses defuzzifikasi digambarkan dengan Gambar 3.16.



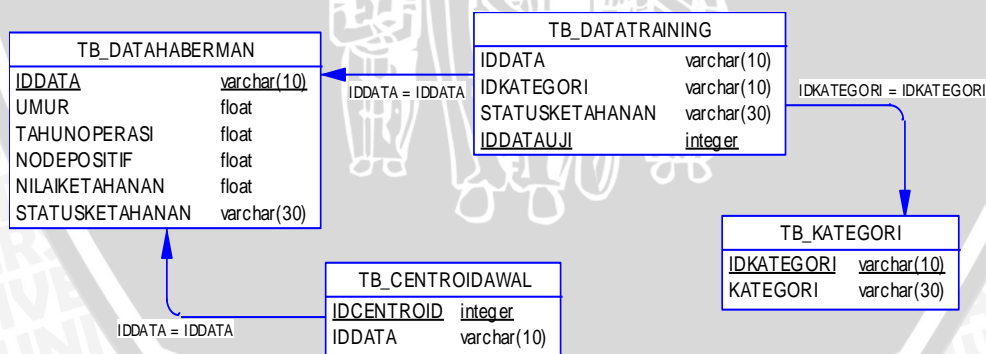
Gambar 3.16 Alur proses menghitung nilai kategori ketahanan penderita kanker payudara

### 3.3.3 Perancangan Tabel

Dalam penelitian ini digunakan empat tabel mulai dari proses *clustering*, pembangkitan aturan, sampai pengujian data. Tabel-tabel tersebut yaitu tabel data Haberman's, tabel data training, tabel kategori, dan tabel centroid awal. Dalam penelitian ini tidak semua hasil *clustering* disimpan, hanya beberapa hasil dari proses perhitungan seperti hasil dari proses uji data yang nantinya menentukan kategori data penderita yang akan disimpan pada tabel data training. Untuk jelasnya relasi antar tabel dijelaskan pada Gambar 3.17. Sedangkan untuk physical data modelnya dijelaskan pada Gambar 3.18. Yang mana physical data model ini menggambarkan hubungan dan gambaran tipe-tipe data apa yang digunakan dalam tabel.



Gambar 3.17 Conceptual Data Model



Gambar 3.18 Physical Data Model

Dari gambar di atas bisa dilihat ketiga tabel yaitu `tb_datahaberman`, `tb_datatraining` dan `tb_kategori` mempunyai relasi. Untuk relasi tabel `tb_datatraining` dengan `tb_kategori` mempunyai relasi *one to many* (1 to m), dimana satu kategori bisa mempunyai satu atau lebih dari satu data training (banyak data). Sedangkan relasi untuk tabel `tb_datahaberman` dengan



tb\_datatraining mempunyai relasi *one to one* (1 to 1). Dimana satu data haberman akan diuji ke dalam satu data training. Untuk penjelasan struktur masing-masing tabel akan dijelaskan pada uraian sebagai berikut :

1. Tabel tb\_datahaberman

Tabel ini digunakan untuk menyimpan seluruh data Haberman's Survival sebanyak 306 data. Dimana dalam tabel ini terdapat enam field yang digunakan untuk menyimpan Id Data (idData), Umur Pasien (umur), Tahun Operasi Pasien (tahunOperasi), Node Positif Terdeteksi Kanker (nodePositif), Kelas Ketahanan Pasien (nilaiKetahanan), dan Status Ketahanan Pasien (statusKetahanan). Untuk lebih jelasnya rancangan tabel tb\_datahaberman bisa dilihat pada Tabel 3.1.

Tabel 3.1 Perancangan Tabel tb\_datahaberman

Nama Kolom	Tipe Data	Keterangan
idData	Varchar	Merupakan id tiap data Haberman's yang memiliki panjang field sepuluh karakter.
Umur	Float	Merupakan umur pasien ketika melakukan operasi.
tahunOperasi	Float	Merupakan tahun dimana pasien melakukan operasi.
nodePositif	Float	Merupakan titik bagian penderita yang didiagnosa positif mengidap kanker.
nilaiKetahanan	Float	Merupakan kelas ketahanan penderita berupa angka 1.0 dan 2.0.
statusKetahanan	Varchar	Merupakan status ketahanan dari kelas nilai ketahanan yaitu 1.0 (bertahan hidup lebih dari 5 tahun) dan 2.0 (hidup kurang dari 5 tahun).

2. Tabel tb\_kategori

Tabel tb\_kategori ini berfungsi untuk menyimpan kategori status penderita yaitu kategori data bertahan hidup lebih dari lima tahun atau

sebaliknya. Penjelasan dari perancangan tabel `tb_kategori` dijelaskan pada Tabel 3.2.

Tabel 3.2 Perancangan Tabel `tb_kategori`

Nama Kolom	Tipe Data	Keterangan
<code>idKategori</code>	Varchar	Merupakan id tiap kategori yang digunakan untuk menilai status data Haberman's Survival dan merupakan primary key.
<code>nilaiCluster</code>	Float	Merupakan kelas <i>cluster</i> sebagai hasil dari pengelompokan data.

### 3. Tabel `tb_datatraining`

Seperti halnya tabel `tb_datahaberman`, tabel ini berfungsi untuk menyimpan data Haberman's Survival yang digunakan sebagai data training. Data ini mereference data dari tabel `tb_datahaberman` dengan menggunakan `idData` sebagai penghubung tabel. Tabel ini juga menyimpan hasil penentuan kategori dari data tergolong ke dalam kategori data yang mampu bertahan hidup lebih dari lima tahun atau sebaliknya. Struktur perancangan `tb_datatraining` dijelaskan pada Tabel 3.3.

Tabel 3.3 Perancangan Tabel `tb_datatraining`

Nama Kolom	Tipe Data	Keterangan
<code>idData</code>	Varchar	Merupakan id tiap data Haberman's yang memiliki panjang field sepuluh karakter dan merupakan foreign key.
<code>idKategori</code>	Varchar	Merupakan id tiap kategori yang digunakan untuk menilai status data Haberman's Survival dan merupakan foreign key.
<code>statusKetahanan</code>	Varchar	Merupakan status ketahanan dari kelas nilai ketahanan yaitu 1.0 (bertahan hidup lebih dari 5 tahun) dan 2.0 (hidup kurang dari 5 tahun).

#### 4. Tabel tb\_centroidawal

Tabel ini berfungsi untuk menampung data Haberman's Survival yang dijadikan sebagai bilangan random untuk centroid awal. Dimana nantinya data centroid awal ini digunakan untuk proses *clustering*. Struktur perancangan tb\_centroidawal dijelaskan pada Tabel 3.4.

Tabel 3.4 Perancangan Tabel tb\_centroidawal

Nama Kolom	Type Data	Keterangan
idCentroid	Varchar	Merupakan id data centroid yang diambil dari data Haberman's yang mempunyai tipe integer dan merupakan primery key.
idData	Varchar	Merupakan id tiap data Haberman's yang memiliki panjang field sepuluh karakter dan merupakan foreign key.

#### 3.3.4 Perancangan Antarmuka

Secara umum gambaran dari sistem untuk membangkitkan aturan *fuzzy* pada klaisifikasi ketahanan hidup penderita kanker payudara terdiri dari dua bagian utama, yaitu :

##### 1. Proses Pelatihan

Proses ini merupakan proses dimana sistem menjalankan program dan melakukan proses *clustering* termasuk perhitungan varian untuk cluster paling ideal. Proses ini bertujuan untuk mengelompokkan data berdasarkan atribut yang telah ditentukan serta untuk mendapatkan nilai koefisien output setiap rule yang dibangkitkan. Untuk desain antarmuka dari proses pelatihan bisa dilihat pada Gambar 3.19.

##### 2. Proses Pengujian

Pada proses pengujian ini bertujuan untuk melakukan pengujian terhadap data uji dengan menggunakan aturan yang dihasilkan dari proses pelatihan. Dengan proses pengujian menggunakan rule yang dibangkitkan akan dilihat sejauh mana keakuratan rule yang dibangkitkan dan sistem yang digunakan dalam mengklasifikasi data uji apakah kategori data uji



menggunakan sistem sesuai dengan data hasil penelitian pada data Haberman's Survival. Proses pengujian ini harus melalui tahapan proses seperti perhitungan derajat keanggotaan menggunakan fungsi Gauss, proses fuzzifikasi, perhitungan  $z$  masing-masing *cluster*, dan proses defuzzifikasi yang nantinya akan dihasilkan nilai crisp berupa nilai kategori apakah data uji termasuk kategori data pasien kanker payudara yang bertahan hidup lebih lima tahun atau tidak. Untuk desain antarmuka dari proses pengujian bisa dilihat pada Gambar 3.19.

The screenshot shows a web-based application interface for breast cancer patient survival analysis. The main window is titled "Form Utama Aplikasi Ketahanan Hidup Penderita Kanker Payudara". It features a tabbed interface with "Data Haberman Survival" and "Data Uji" tabs. The central area is titled "PEMBANGKITAN ATURAN FUZZY MENGGUNAKAN METODE K-MEANS CLUSTERING". Under "Penentuan Cluster", users can select the number of clusters from a dropdown menu, enter the number of training data points, and specify a variance value. There are "Proses" and "Reset" buttons for this section. Below this, there are sections for "Hasil Pembangkitan Aturan" and "Hasil Data Tercluster", both currently empty. A table header is visible with columns: "Umur", "Tahun Operasi", "Node Positif", "Nilai Hasil Uji (Z)", "Status Uji", and "Status Penelitian". The "Hasil Uji Semua Data Dan Akurasi Perhitungan" section is also empty. At the bottom, the "Proses Uji Data" section includes a "Load" button, input fields for "Umur Pasien", "Tahun Operasi" (pre-filled with 19), "Node Positif", and "Hasil Perhitungan Klasifikasi", and a "Hasil Keputusan Status Data Uji" field with "Proses" and "Reset" buttons.

Gambar 3.19 Tampilan Antarmuka Sistem

Dalam tampilan antarmuka sistem tersebut terdapat dua bagian utama, yaitu bagian proses pelatihan dan bagian proses uji data. Seperti yang telah dijelaskan bahwa pada proses pelatihan ini juga dilakukan perhitungan *cluster* ideal berdasarkan varian terkecil sampai dilakukan proses pembangkitan aturan yang nantinya digunakan sebagai rule acuan untuk menguji data uji. Untuk lebih jelasnya bagian-bagian dari proses pelatihan dan proses pengujian akan dijelaskan dengan Gambar 3.20, Gambar 3.21, dan Gambar 3.22.

Form Utama Aplikasi Ketahanan Hidup Penderita Kanker Payudara

**C** Data Haberman Survival **D** Data Uji

PEMBANGKITAN ATURAN FUZZY MENGGUNAKAN METODE K-MEANS CLUSTERING

Penentuan Cluster (A)

Jumlah Cluster : Pilih Cluster

Jumlah Data Latih :

Hasil Varian :

Proses Reset

Hasil Pembangkitan Aturan

Hasil Data Tercluster

Umur	Tahun Operasi	Node Positif	Nilai Hasil Uji (Z)	Status Uji	Status Penelitian
Hasil Uji Semua Data Dan Akurasi Perhitungan					

Proses Uji Data (B)

Load Umur Pasien :  Tahun Operasi : 19  Node Positif :  Hasil Perhitungan Klasifikasi :

Hasil Keputusan Status Data Uji :  Proses Reset

Gambar 3.20 Tampilan Bagian-Bagian Proses Sistem

Keterangan :

- A = Proses Pelatihan
- B = Proses Pengujian Data
- C = Menu Form Data Habermans Survival
- D = Menu Form Data Uji

Pada bagian selanjutnya akan dijelaskan bagian-bagian berdasarkan masing-masing proses pelatihan dan proses pengujian data.

Gambar 3.21 Tampilan Bagian-Bagian Proses Pelatihan

Keterangan :

1. Bagian proses perhitungan varian dan *cluster* ideal dengan varian terkecil.
2. Text area yang digunakan untuk menampilkan aturan *fuzzy* yang terbentuk dari proses pembangkitan aturan *fuzzy*.
3. Bagian text area yang berfungsi untuk menampilkan hasil data yang telah dikelompokkan.
4. Bagian antarmuka yang berfungsi untuk menampilkan hasil pengujian seluruh data dan menampilkan hasil perhitungan akurasi dari hasil pengujian.
5. Combo box yang digunakan sebagai pilihan jumlah *cluster* yang akan dihitung nilai variannya, yang nantinya digunakan sebagai jumlah *cluster* ideal yang akan digunakan.
6. Text field yang digunakan sebagai media inputan jumlah data yang digunakan sebagai data training.
7. Text field yang berfungsi sebagai output dari hasil perhitungan varian.
8. Tombol proses berfungsi sebagai tombol utama untuk melakukan beberapa proses seperti, proses perhitungan *cluster* dengan jumlah varian terkecil setelah semua nilai dan parameter terpenuhi, melakukan proses *clustering*, sampai proses pembangkitan aturan.

9. Tombol reset yang digunakan sebagai pengaturan nilai-nilai dari objek-objek antarmuka di set null.

The screenshot shows a web form titled "Proses Uji Data". It contains several input fields and buttons. Callout 1 points to a "Load" button. Callout 2 points to the "Umur Pasien" input field. Callout 3 points to the "Tahun Operasi" input field. Callout 4 points to the "Node Positif" input field. Callout 5 points to the "Hasil Perhitungan Klasifikasi" output field. Callout 6 points to the "Hasil Keputusan Status Data Uji" output field. Callout 7 points to a "Proses" button. Callout 8 points to a "Reset" button.

Gambar 3.22 Tampilan Bagian-Bagian Proses Pengujian Data

Keterangan :

1. Tombol Load digunakan untuk menampilkan form data yang ingin diuji satu persatu.
2. Text field ini berfungsi sebagai inputan berupa umur pasien ketika melakukan operasi kanker.
3. Text field tahun operasi ini digunakan untuk inputan berupa tahun dimana pasien melakukan operasi.
4. Text field node positif merupakan inputan berupa jumlah node positif yang terdeteksi kanker payudara.
5. Text field ke empat berfungsi sebagai output dari pengujian data berupa nilai crisp dari perhitungan menggunakan rule hasil pembangkitan.
6. Text field kriteria status pasien merupakan output berupa kategori sebagai hasil dari perhitungan pengujian data berupa kategori apakah data yang diuji tergolong dalam penderita yang bertahan hidup lebih dari sama dengan lima tahun atau penderita yang hidup kurang dari lima tahun.
7. Tombol proses berfungsi sebagai tombol aksi untuk melakukan proses pengujian data.
8. Tombol reset berfungsi untuk mengembalikan field inputan dan output pada proses pengujian data bernilai null.

Untuk mengetahui data-data Haberman's Survival bisa dilihat dengan menggunakan Form Data Haberman melalui menu bar Data Haberman Survival seperti pada Gambar 3.23.

Gambar 3.23 Tampilan Form Data Haberman

Sedangkan untuk melihat data uji dan data training bisa dilihat pada Form Data Uji melalui menu bar Data Uji seperti pada Gambar 3.24.

Keterangan :

\*Tabel data Haberman’s Survival

Gambar 3.24 Tampilan Form Data Uji

Keterangan :

1. Text field id data uji untuk menambahkan data uji pada proses pembangunan aturan dan proses uji data.
2. Text field id data untuk menampilkan id data Haberman.

3. Text field ini berfungsi sebagai inputan berupa umur pasien ketika melakukan operasi kanker.
4. Text field tahun operasi ini digunakan untuk inputan berupa tahun dimana pasien melakukan operasi.
5. Text field node positif merupakan inputan berupa jumlah node positif yang terdeteksi kanker payudara.
6. Text field status pasien merupakan output berupa kategori sebagai kategori apakah data yang tersebut tergolong dalam penderita yang bertahan hidup lebih dari sama dengan lima tahun atau penderita yang hidup kurang dari lima tahun.
7. Tombol v digunakan untuk menampilkan form data Haberman's Survival.
8. Tombol Tambah digunakan untuk menambahkan data uji yang digunakan untuk proses pembangkitan aturan dan proses uji data.
9. Tombol Hapus digunakan untuk menghapus data uji.

### 3.3.5 Perhitungan Manual

#### 3.3.5.1 Proses Perhitungan K-Means Clustering

Pada proses perhitungan K-Means *clustering* ini diperlukan data training yang diambil dari data set Haberman's Survival. Dimana data Haberman's tersebut terdiri dari empat atribut pendukung diantaranya umur, tahun operasi, node positif, dan kategori penderita (bertahan hidup lebih dari lima tahun bernilai 0 atau bertahan hidup kurang dari lima tahun bernilai 1). Contoh data training yang digunakan dalam proses K-Means *clustering* seperti pada Tabel 3.5.

Tabel 3.5 Contoh Data Penderita Kanker Payudara Pada Data Haberman's Survival

No.	Umur Pasien	Tahun Pasien Operasi	Node Positif	Kategori
1	34	1960	0	0
2	54	1966	0	0
3	63	1961	9	0
4	59	1964	4	0
5	55	1966	0	0
6	57	1961	5	1
7	35	1964	13	0
8	53	1965	12	1
9	41	1967	0	1
10	49	1964	10	1

Berdasarkan tabel di atas, disajikan sejumlah data penderita kanker payudara sejumlah sepuluh data ( $n = 10$ ). Disamping itu terdapat empat jumlah atribut ( $m = 4$ ) sebagai parameter pendukung seperti (umur, tahun, node, dan kategori pasien). Misalkan dari sepuluh data tersebut akan di kelompokkan ke dalam 3 *cluster* (yang selanjutnya kita sebut  $c$ ) dengan menggunakan metode K-Means *clustering*. Selanjutnya dilakukan proses pembangkitan pusat *cluster* (centroid) awal (sejumlah *cluster* yang diinginkan) yang digunakan dalam proses *clustering* di setiap atributnya. Misalkan centroid awal hasil random ditunjukkan data sebagai berikut :

centroid awal :	Umur	Tahun Operasi	Node	Kategori
c1:	47	60	4	0
c2:	69	67	8	1

Setelah dilakukan proses pembangkitan centroid awal pada masing-masing *cluster*, dilakukan perhitungan jarak semua data training dengan pusat *cluster* di setiap *cluster*-nya. Dalam perhitungan jarak antar data dengan pusat *cluster* menggunakan rumus Euclidean (persamaan 2.1) seperti yang sudah dijelaskan pada sub bab 2.5.4.2.

Pada iterasi pertama, dengan menggunakan rumus Euclidean (persamaan 2.1), dihasilkan 2 pusat *cluster*. Untuk perhitungan lengkapnya bisa dilihat pada Tabel 3.6.

Tabel 3.6 Data Hasil Perhitungan Jarak Data Training Dengan Centroid Awal

Iterasi Ke-1			
No.	Distance Ke-C1	Distance Ke-C2	Cluster
1	13.60147051	36.59234893	1
2	10.04987562	17.05872211	1
3	16.79285562	8.602325267	2
4	12.64911064	11.22497216	2
5	10.77032961	16.18641406	1
6	10.14889157	13.74772708	1
7	15.5241747	34.51086785	1
8	11.22497216	16.61324773	1
9	10.09950494	29.12043956	1
10	7.549834435	20.32240143	1

Dari hasil perhitungan di Tabel 3.6 bisa diketahui untuk kesemua data digolongkan ke dalam *cluster* 1 atau *cluster* 2. Penentuan data menjadi anggota *cluster* 1 atau *cluster* 2 diketahui dari membandingkan hasil perhitungan jarak di setiap *cluster* dan diambil dengan jarak terkecil. Seperti contoh pada data ke-1, hasil perhitungan Euclidean pada *cluster* c1 lebih kecil dari hasil perhitungan Euclidean pada *cluster* c2. Sehingga dapat ditentukan data ke-1 termasuk golongan *cluster* yang pertama. Proses tersebut dilakukan sampai kesemua data masuk ke dalam kategori *cluster*.

Proses selanjutnya, setelah diketahui anggota tiap *cluster*. Maka dihitung centroid baru dari iterasi ke-1 di setiap *cluster* dengan mencari mean data anggota *cluster* di setiap *cluster*. Misal kita hitung centroid baru untuk atribut umur, pada Tabel 3.6 bisa kita lihat anggota pada *cluster* c1 sebanyak 8 data, sedangkan pada *cluster* c2 mempunyai anggota sebanyak 2 data. Jadi bisa kita hitung centroid baru sebagai berikut :



$$c1 = \frac{34 + 54 + 55 + 57 + 35 + 53 + 41 + 49}{8} = 47.25$$

$$c2 = \frac{63 + 59}{2} = 61$$

Begitu seterusnya untuk menghitung centroid baru pada atribut yang lain di setiap *cluster*-nya. Sehingga hasil perhitungan centroid baru didapatkan sebagai berikut :

centroid baru iterasi ke-1 :	Umur	Tahun Operasi	Node	Kategori
<b>c1:</b>	47.25	64.125	5	0.5
<b>c2:</b>	61	62.5	6.5	0

Dari proses perhitungan centroid baru di atas dilakukan lagi proses perhitungan jarak Euclidean pada iterasi ke-2. Seperti halnya menghitung jarak pada iterasi ke-1 digunakan rumus Euclidean pada persamaan 2.1. hasil dari proses perhitungan jarak pada iterasi ke-2 bisa dilihat pada Tabel 3.7.

Tabel 3.7 Data Hasil Perhitungan Jarak Pada Iterasi Ke-2

Iterasi Ke-2			
No.	Distance Ke-C1	Distance Ke-C2	Cluster
1	15.21281024	28.5323036	1
2	9.185749867	10.97992359	1
3	17.05494007	4.59372673	2
4	12.32790475	4.870828693	2
5	9.979902535	10.32891156	1
6	10.73438478	5.936979626	2
7	15.16088057	27.47746109	1
8	9.710978177	10.86799189	1
9	9.092724011	22.19505439	1
10	5.989821319	13.40978752	1

Hasil proses perhitungan jarak pada iterasi ke-2 ini bisa dilihat terjadi perubahan anggota *cluster* yaitu perubahan anggota tersebut terjadi pada data ke-6. Dimana data ke-6 pada proses iterasi ke-1 menjadi anggota *cluster* c1, tetapi pada perhitunga jarak di iterasi ke-2 berubah menjadi anggota *cluster* ke-

2. Karena masih adanya perubahan anggota pada *cluster*, maka dilakukan lagi perhitungan centroid baru dari anggota *cluster* baru seperti pada iterasi ke-2. Proses iterasi dan perhitungan centroid baru terus dilakukan sampai anggota pada masing-masing *cluster* tidak ada yang berubah.

Dari tabel iterasi ke-2 diperoleh jumlah anggota *cluster* c1 menjadi 7 data dan *cluster* c2 menjadi 3 data. Proses menghitung centroid baru pada iterasi ke-2, ditunjukkan sebagai berikut :

$$c1 = \frac{34 + 54 + 55 + 35 + 53 + 41 + 49}{7} = 45.8571428571428$$

$$c2 = \frac{63 + 59 + 57}{3} = 59.6666666666666$$

Dengan proses dan cara yang serupa dihitung centroid baru untuk atribut tahun operasi, node positif, dan kategori di masing-masing *cluster* dan akan dihasilkan centroid baru sebagai berikut :

centroid baru iterasi ke-2 : **Umur Tahun Operasi Node Kategori**

<b>c1:</b>	45.857	64.571	5	0.428571
<b>c2:</b>	59.667	62	6	0.333333

Dari hasil proses perhitungan centroid baru pada iterasi ke-2, digunakan centroid baru tersebut untuk menghitung jarak Euclidean pada iterasi ke-3. Hasil dari perhitungan jarak tersebut bisa dilihat pada Tabel 3.8.

Tabel 3.8 Data Hasil Perhitungan Jarak Pada Iterasi Ke-3

Iterasi Ke-3			
No.	Distance Ke-C1	Distance Ke-C2	Cluster
1	14.20866917	26.97074698	1
2	10.1349081	9.757158696	2
3	18.43559781	5.080939771	2
4	13.65916552	3.664203085	2
5	10.99073378	9.177025687	2
6	12.16467356	3.93605579	2
7	14.0551996	26.24943841	1
8	10.49959954	10.0377727	2
9	7.994313638	20.78662895	1
10	6.44363512	12.15039134	1

Dan karena pada proses *clustering* data pada iterasi ke-3 ini masih ada perubahan anggota pada masing-masing *cluster*. Yaitu perubahan pada data ke-2, 5, dan 8 menjadi anggota *cluster* ke-2 atau c2. Sehingga jumlah anggota pada *cluster* c1 menjadi 4 data, sedangkan pada *cluster* c2 menjadi 6 data. Dengan begitu dilakukan ulang perhitungan centroid baru untuk iterasi ke-4 sebagai berikut :

$$c1 = \frac{34 + 35 + 41 + 49}{4} = 39.75$$

$$c2 = \frac{54 + 63 + 59 + 55 + 57 + 53}{6} = 56.83333333333333$$

Dilakukan pula perhitungan centroid untuk setiap atribut di masing-masing *cluster* sehingga menghasilkan centroid baru pada iterasi ke-3 sebagai berikut :

centroid baru iterasi ke-3 : **Umur Tahun Operasi Node Kategori**

**c1:** 39.75 63.75 5.75 0.5

**c2:** 56.8333 63.8333 5 0.333333

Seperti pada iterasi-iterasi sebelumnya dihitung ulang jarak Euclidean pada iterasi ke-4 dengan menggunakan centroid baru hasil dari iterasi ke-3. Hasil perhitungan iterasi ke-4 bisa dilihat pada Tabel 3.9.

Tabel 3.9 Data Hasil Perhitungan Jarak Pada Iterasi Ke-4

Iterasi Ke-4			
No.	Distance Ke-C1	Distance Ke-C2	Cluster
1	9.728626274	24.25154456	1
2	15.86103427	6.898586234	2
3	24.02858918	8.200463794	2
4	19.68747484	3.08180733	2
5	16.79071868	6.507203113	2
6	17.83541907	3.526479234	2
7	9.460058118	23.81349978	1
8	15.0644316	8.691246901	2
9	7.309706809	17.52008264	1
10	10.50624267	9.949056472	2

Dari proses perhitungan jarak pada iterasi ke-4, hasil pada Tabel 3.5 memperlihatkan bahwa masih ada perubahan anggota *cluster* pada data ke-10. Sehingga perubahan jumlah anggota *cluster* c1 menjadi 3 data dan *cluster* c2 menjadi 7 data. Karena masih adanya perubahan anggota maka dilakukan lagi perhitungan centroid baru pada iterasi ke-4 untuk setiap *cluster* sehingga menghasilkan centroid baru seperti berikut :

centroid baru iterasi ke-4 : **Umur Tahun Operasi Node Kategori**

**c1:** 36.6667 63.6667 4.333 0.33333

**c2:** 55.7143 63.8857 5.7143 0.42857

Selanjutnya dilakukan kembali perhitungan jarak Euclidean pada iterasi ke-5. Untuk hasil perhitungan jarak pada iterasi ke-5 ditampilkan pada Tabel 3.10.

Tabel 3.10 Data Hasil Perhitungan Jarak Pada Iterasi Ke-5

Iterasi Ke-5			
No.	Distance Ke-C1	Distance Ke-C2	Cluster
1	7.012351297	23.31257245	1
2	18.45640301	6.875022413	2
3	27.32165869	8.967996074	2
4	22.77715052	4.125600907	2
5	19.42254265	6.666781729	2
6	20.95900435	3.765573058	2
7	9.357946849	22.49846321	1
8	18.51599874	7.563164085	2
9	7.511379229	16.63993478	1
10	13.97440591	8.587489437	2

Dari hasil proses *clustering* pada iterasi ke-5 bisa dilihat pada Tabel 3.10 jumlah dan anggota *cluster* c1 dan c2 sama dengan hasil pengelompokan data pada Tabel 3.9. Dengan begitu diketahui sudah tidak ada perubahan anggota *cluster*, sehingga tidak diperlukan lagi perhitungan centroid baru pada iterasi ke-5 ini. Dengan demikian proses *clustering* menggunakan metode K-Mean *clustering* sudah selesai.

Dari proses berakhirnya *clustering* menggunakan K-Means, dilanjutkan proses perhitungan standar deviasi pada setiap atribut di masing-masing *cluster*. Dalam perhitungan standar deviasi ini dihitung dengan menggunakan persamaan 2.3. Dalam kasus ketahanan hidup penderita kanker payudara ini dilakukan perhitungan standar deviasi untuk atribut umur, tahun operasi, node positif, dan kategori penderita. Missal dihitung standar deviasi atribut umur (sebut saja  $\sigma_{11}$ ) pada *cluster* c1 (*cluster* ke-1) sebagai berikut :

$$\sigma_{11} = \sqrt{\frac{(34 - 36.67)^2 + (35 - 36.67)^2 + (41 - 36.67)^2}{3 - 1}} = 4.390647$$

Proses perhitungan standar deviasi dilakukan di setiap *cluster* pada masing-masing atribut. Sehingga hasil proses perhitungan standar deviasi :

Standar Deviasi	Umur	Tahun Operasi	Node	Kategori
c1 :	4.390647	4.371626	10.16257	0.6666667
c2 :	9.15754	5.180747	11.0509	1.2009067

### 3.3.5.2 Analisis Varian

Setelah data dikelompokkan berdasarkan jarak terdekat dengan *centroid* dan setelah dilakukan perhitungan standar deviasi untuk setiap *cluster*. Dilanjutkan lagi menghitung rata-rata data *centroid* hasil *clustering* untuk setiap atributnya ( $\bar{y}$ ) yang dihitung dari rata-rata data setiap *clusternya* ( $\bar{y}_i$ ). Hasil perhitungan mean data *centroid* disajikan dalam Tabel 3.11.

Tabel 3.11 Nilai Rata-Rata Setiap Atribut Pada *Centroid* Akhir Data Training

Cluster		Rata-Rata ( $\bar{y}_i$ )			
		Umur	Tahun	Node	Kategori
1	Y1	36.66667	63.66667	4.333333	0.333333
2	Y2	55.71429	63.85714	5.714286	0.428571
<b>Rata-Rata (<math>\bar{y}</math>)</b>		<b>46.19048</b>	<b>63.7619</b>	<b>5.02381</b>	<b>0.380952</b>

Setelah rata-rata ( $\bar{y}_i$ ) dan rata-rata ( $\bar{y}$ ) diketahui, langkah selanjutnya yaitu menghitung varian setiap *cluster* ( $V_c$ ) dengan menggunakan persamaan 2.4. Karena dalam pengelompokan data menggunakan 2 *cluster*, maka akan dihitung 2 nilai varian setiap *cluster* yaitu  $V_1$  dan  $V_2$ . Contoh perhitungan varian *cluster* c1 adalah sebagai berikut :

$$V_1^2 = \frac{1}{3-1} \sum_{i=1}^3 (y_i - \bar{y}_c)^2$$

$$V_1^2 = \frac{1}{3-1} * ((34 - 36.6667)^2 + (35 - 36.6667)^2 + (41 - 36.6667)^2 + (60 - 63.6667)^2 + (64 - 63.6667)^2 + (67 - 63.6667)^2 + (0 - 4.3333)^2 + (13 - 4.3333)^2 + (0 - 4.3333)^2 + (0 - 0.3333)^2 + (0 - 0.3333)^2 + (0 - 0.3333)^2 + (1 - 0.3333)^2) = 83.333333333333$$

Dengan menggunakan rumus yang sama dihitung juga varian setiap *cluster* untuk *cluster* c2. Dari perhitungannya dihasilkan nilai varian setiap *cluster* yaitu  $V_1=83.333333333333$  dan  $V_2=47.9047619047619$ .

Setelah perhitungan varian setiap *cluster* selesai selanjutnya dihitung *variance within cluster* ( $V_w$ ) dan *variance between cluster* ( $V_b$ ). Untuk menghitung  $V_w$  digunakan rumus persamaan 2.5, contoh perhitungan  $V_w$  sebagai berikut :

$$V_w = \frac{1}{10-2} \sum_{i=1}^3 (n_i - 1) \cdot V_i^2$$

$$V_w = \frac{1}{10-2} * ((3-1) * (83.33333333333333) + (7-1) * (47.9047619047619))$$

$$= 56.7619047619048$$

Dilanjutkan menghitung  $V_b$  dengan menggunakan persamaan 2.6, contoh perhitungan  $V_b$  sebagai berikut :

$$V_b = \frac{1}{3-1} \sum_{i=1}^3 (\bar{y}_i - \bar{y})^2$$

$$V_b = \frac{1}{3-1} * ((3) * (36.6667 - 46.190)^2 + (7) * (55.7143 - 46.190)^2 +$$

$$(3) * (63.6667 - 63.762)^2 + (7) * (63.85714 - 63.762)^2 +$$

$$(3) * (4.3333 - 5.024)^2 + (7) * (5.7143 - 5.024)^2 +$$

$$(3) * (0.3333 - 0.381)^2 + (7) * (0.42857 - 0.381)^2)$$

$$= 911.910430839002$$

Setelah nilai  $V_w$  dan  $V_b$  diketahui, maka batasan varian ( $V$ ) dapat dihitung dengan persamaan 2.7, hasil perhitungan  $V$  sebagai berikut :

$$V = \frac{V_w}{V_b} = \frac{56.7619047619048}{911.910430839002} = 0.0622450438577405$$

Suatu *cluster* dikatakan layak jika  $V_w < V_b$ . semakin kecil  $V_w$  dan semakin besar  $V_b$  maka *cluster* semakin baik. Dari perhitungan  $V_w$  dan  $V_b$  di atas bisa diketahui nilai  $V_w < V_b$ . ini berarti untuk penggunaan *cluster* dengan jumlah *cluster* ( $c=2$ ) sudah layak untuk digunakan.

### 3.3.5.3 Ekstraksi Aturan Fuzzy

Setelah dilakukan analisis varian dan pengelompokan data, maka diketahui jumlah *cluster* ideal yang terbentuk berdasarkan nilai varian terkecil. Proses pemilihan *cluster* ideal ini dilakukan beberapa kali dengan merubah parameter jumlah dari *cluster*. Dengan menentukan jumlah *cluster* yang paling ideal tersebut, maka ditentukan pula jumlah aturan yang ingin dbangkitkan.

Pada contoh proses perhitungan *cluster*, dilakukan perhitungan varian dengan menggunakan *cluster* dengan jumlah *cluster* sama dengan 2. Jika dimisalkan analisis varian dengan *cluster* 2 mempunyai nilai varian terkecil. Maka nantinya hasil rule dari pembangkitan aturan dengan jumlah *cluster* 2 sama dengan 2 aturan. Aturan *fuzzy* dapat diekstraksi dari pusat *cluster* yang telah dicari pada proses *clustering*.

centroid : **Umur Tahun Operasi Node Kategori**

**c1:** 36.6667 63.6667 4.333 0.33333

**c2:** 55.7143 63.8857 5.7143 0.42857

Dari nilai-nilai centroid hasil proses *clustering* dapat diekstrak aturan sebagai berikut :

1. IF umur = in\_C<sub>11</sub> (umur) AND tahun operasi = in\_C<sub>12</sub> (tahun operasi) AND node = in\_C<sub>13</sub> (node) THEN kategori = p<sub>11</sub>x<sub>1</sub>+ p<sub>12</sub>x<sub>2</sub>+ p<sub>13</sub>x<sub>3</sub>+ p<sub>14</sub>
2. IF umur = in\_C<sub>21</sub> (umur) AND tahun operasi = in\_C<sub>22</sub> (tahun operasi) AND node = in\_C<sub>23</sub> (node) THEN kategori = p<sub>21</sub>x<sub>1</sub>+ p<sub>22</sub>x<sub>2</sub>+ p<sub>23</sub>x<sub>3</sub>+ p<sub>24</sub>

Dari hasil ekstraksi aturan di atas, bisa ditentukan missal pada aturan ke-1. Jika umur termasuk in\_C<sub>11</sub>(umur) yang artinya bahwa data umur dekat dengan pusat *cluster* 36.6667; dan tahun operasi termasuk in\_C<sub>11</sub>(tahun operasi) yang berarti data tahun operasi dekat dengan pusat *cluster* 63.6667, dan node termasuk in\_C<sub>13</sub>(node) maka bisa diartikan bahwa data node dekat dengan pusat *cluster* 4.333; maka nilai kategori data penderita adalah dengan menghitung fungsi kategori = p<sub>11</sub>x<sub>1</sub>+ p<sub>12</sub>x<sub>2</sub>+ p<sub>13</sub>x<sub>3</sub>+ p<sub>14</sub>, dimana p adalah koefisien output masing-masing atribut di setiap *cluster* dan x adalah data input yang diuji.

#### 3.3.5.4 Perhitungan Fungsi Keanggotaan

Setelah melakukan ekstraksi aturan, langkah selanjutnya yaitu menghitung fungsi keanggotaan (*membership function*) dari setiap aturan. Proses perhitungan derajat keanggotaan ini menggunakan persamaan 2.17 yaitu fungsi Gauss. Agar dapat membentuk fungsi keanggotaan Gauss diperlukan nilai centroid dan standar deviasi yang sudah dihitung pada proses *clustering*. Untuk menghitung derajat keanggotaan pada *cluster* ke-1 dengan data ke-i, i=1...n (n=jumlah data *cluster*) menggunakan :



$$\mu_{i1} = e^{(-1) * \left( \frac{(x_{i1} - 36.6667)^2}{2 * (3.78594)^2} + \frac{(x_{i2} - 63.6667)^2}{2 * (3.51188)^2} + \frac{(x_{i3} - 4.3333)^2}{2 * (7.50553)^2} + \frac{(x_{i4} - 0.3333)^2}{2 * (0.57733)^2} \right)}$$

Sedangkan untuk menghitung derajat keanggotaan pada data ke- $i$ ,  $i=1 \dots n$  menggunakan :

$$\mu_{2i} = e^{(-1) * \left( \frac{(x_{i1} - 55.7143)^2}{2 * (4.49867)^2} + \frac{(x_{i2} - 63.8571)^2}{2 * (2.11570)^2} + \frac{(x_{i3} - 5.71428)^2}{2 * (4.78589)^2} + \frac{(x_{i4} - 0.42857)^2}{2 * (0.53452)^2} \right)}$$

Dari proses perhitungan fungsi keanggotaan, diperoleh hasil pada Tabel 3.12.

Tabel 3.12 Hasil Proses Perhitungan Fungsi Keanggotaan

Menghitung Derajat Keanggotaan			
$\mu_{ik}$	1	2	sum $\mu_{ik}$
1	0.324185	5.89E-07	3.24E-01
2	1.61E-05	0.197948	1.98E-01
3	1.63E-11	0.062015	6.20E-02
4	2.34E-08	0.519659	5.20E-01
5	4.65E-06	0.210188	2.10E-01
6	2.09E-07	0.2154	2.15E-01
7	0.392689	5.65E-06	3.93E-01
8	2.58E-05	0.17173	1.72E-01
9	0.143873	4.37E-04	1.44E-01
10	0.001907	0.12388	1.26E-01

### 3.3.5.5 Menghitung Koefisien Output

Perhitungan koefisien output dilakukan dengan menggunakan metode kuadrat terkecil (*least square*). Langkah awal dalam menghitung koefisien output yaitu membentuk matriks U dengan ukuran 10x8. Dimana 10 = jumlah baris merupakan banyaknya data *cluster*, sedangkan jumlah kolom = 8 merupakan jumlah parameter (atribut input uji+1) dikalikan jumlah aturan terbentuk. Elemen-elemen matriks U terbentuk dari  $d^k_{ij}$  dan  $d^k_{i(m+1)}$  dengan cara :

$$d^k_{ij} = x_{ij} * \mu_{ik}$$

$$d^k_{i(m+1)} = \mu_{ik}$$

Missal akan dihitung matriks U untuk matriks (0,0) dengan  $k=1$ ,  $i=1$ ,  $j=1$  :

$$d^k_{ij} = 34 * 0.324185176552709 = 11.0223$$

$$d^k_{i(m+1)} = \mu_{ik} = 0.324185176552709$$

Proses perhitungan tersebut dilakukan sampai data matriks U lengkap.

Matriks U :

11.0223	19.45111	0	0.32419	0.00002	.....	0
0.00087	0.00106	0	0.00002	10.68918	.....	0.19795
0	0	0	0	3.90694	.....	0.06201
0	0	0	0	30.65987	.....	0.51966
0.00026	0.00031	0	0	11.56035	.....	0.21019
0.00001	0.00001	0	0	12.27778	.....	0.2154
13.74412	25.1321	5.10496	0.39269	0.0002	.....	0.00001
0.00137	0.00167	0.00031	0.00003	9.1017	.....	0.17173
5.8988	9.63951	0	0.14387	0.0179	.....	0.00044
0.09343	0.12203	0.01907	0.00191	6.07013	.....	0.12388

Selanjutnya setelah dibentuk matriks U, maka dibentuk matriks U ternormalisasi dengan menggunakan persamaan 2.9 :

$$d^k_{ij} = \frac{d^k_{ij}}{\sum_{k=1}^c \mu_{ki}}$$

Contoh perhitungan matriks U ternormalisasi :

$$d^1_{11} = \frac{d^1_{11}}{0.324185765364633} = \frac{11.0223}{0.324185765364633} = 33.99994$$

Seperti halnya pembentukan matriks U, proses perhitungan matriks U ternormalisasi juga dilakukan sampai matriks U ternormalisasi lengkap.

Matriks U ternormalisasi (matriks U) :

33.99994	59.99989	0	1	.....	0
0.0044	0.00538	0	0.00008	.....	0.99992
0	0	0	0	.....	1
0	0	0	0	.....	1
0.00122	0.00146	0	0.00002	.....	0.99998
0.00006	0.00006	0	0	.....	1



34.9995	63.99908	12.99981	0.99999	.....	0.00001
0.00795	0.00975	0.0018	0.00015	.....	0.99985
40.87598	66.79733	0	0.99698	.....	0.00302
0.74275	0.97012	0.15158	0.01516	.....	0.98484

Selanjutnya dilakukan proses perhitungan dengan menggunakan metode kuadrat terkecil. Terlebih dahulu dibentuk matriks U ternormalisasi transpose :

Matriks U ternormalisasi transpose ( $U^T$ ):

33.99994	0.0044	0	0	.....	0.74275
59.99989	0.00538	0	0	.....	0.97012
0	0	0	0	.....	0.15158
1	0.00008	0	0	.....	0.01516
0.00006	53.9956	63	59	.....	48.25725
0.00011	65.99462	61	64	.....	63.02988
0	0	9	4	.....	9.84842
0	0.99992	1	1	.....	0.98484

Setelah itu proses dilanjutkan dengan mengalikan matriks  $U^T$  dengan matriks U. Dari hasil perkalian antara ( $U^T \times U$ ) kemudian diinverse sehingga menghasilkan matriks berukuran 8 x 8.

Hasil Inverse Perkalian ( $U^T \times U$ ) :

77.99184	-77.9855	17.99234	2027.403	.....	-149.855
-77.98554	78.02019	-17.9925	-2029.84	.....	149.832
17.99234	-17.9925	4.15969	467.7754	.....	-34.5749
2027.40257	-2029.84	467.7754	52867.98	.....	-3894.84
1.21058	-1.21039	0.27931	31.4639	.....	-4.88342
1.26742	-1.26723	0.29242	32.94118	.....	-9.36251
-0.14529	0.14532	-0.03354	-3.77927	.....	-1.2486
-149.85547	149.832	-34.5749	-3894.84	.....	881.6881

Dilanjutkan dari hasil inverse matriks dikalikan dengan matriks  $U^T$  sehingga akan menghasilkan matriks berukuran 8 x 10. Dan hasil dari perkalian tersebut dikalikan dengan matriks data output (Y) yang mempunyai ukuran (10 x 1) dan akan menghasilkan koefisien output. Hasil perhitung koefisien output (k) :



Koef.Output

-4.11085

4.25321

-0.99246

-115.42353

-0.174

-0.23351

0.03892

24.93611

Hasil dari perhitungan koefisien output bisa ditampilkan :

Umur	Tahun Operasi	Node	Kategori
-4.11085	4.25321	-0.99246	-115.42353
-0.174	-0.23351	0.03892	24.93611

Data koefisien ini nantinya digunakan sebagai perhitungan data uji pada aturan yang dibangkitkan yang didefinisikan sebagai koefisien p. dari koefisien p inilah nanti akan dihitung parameter-parameter output di setiap aturannya :

1. IF umur = in\_C<sub>11</sub> (umur) AND tahun operasi = in\_C<sub>12</sub> (tahun operasi) AND node = in\_C<sub>13</sub> (node) THEN kategori = p<sub>11</sub>x<sub>1</sub>+ p<sub>12</sub>x<sub>2</sub>+ p<sub>13</sub>x<sub>3</sub>+ p<sub>14</sub>
2. IF umur = in\_C<sub>21</sub> (umur) AND tahun operasi = in\_C<sub>22</sub> (tahun operasi) AND node = in\_C<sub>23</sub> (node) THEN kategori = p<sub>21</sub>x<sub>1</sub>+ p<sub>22</sub>x<sub>2</sub>+ p<sub>23</sub>x<sub>3</sub>+ p<sub>24</sub>

Sehingga hasil perhitungan dari Z untuk setiap aturan :

$$Z_1 = p_{11}x_1 + p_{12}x_2 + p_{13}x_3 + p_{14}$$

$$= (-4.11085 * x_1) + (4.25321 * x_2) + (-0.99246 * x_3) + (-115.42353)$$

$$Z_2 = p_{21}x_1 + p_{22}x_2 + p_{23}x_3 + p_{24}$$

$$= (-0.174 * x_1) + (-0.23351 * x_2) + (0.03892 * x_3) + (24.93611)$$

### 3.3.5.6 Proses Uji Data

Dalam proses pengujian data, diperlukan beberapa proses untuk dapat menguji suatu data uji. Proses-proses yang harus dilakukan dalam pengujian data :

- a) Proses Fuzzifikasi

Pada proses ini diawali dengan menghitung derajat keanggotaan setiap data uji dengan menggunakan fungsi Gauss pada persamaan 2.17. Proses



ini dilakukan di setiap atribut data uji. Missal ingin menguji dan menghitung derajat keanggotaan data ke-1 dengan parameter data input :

- Umur penderita : 34 tahun
- Tahun Operasi : 60 (tahun 1960)
- Node Positif : 0

Dari data di atas dihitung untuk setiap parameter atributnya menggunakan fungsi Gauss :

$$gaussian(x; c; \sigma) = e^{-\frac{(x-c)^2}{2*(\sigma)^2}}$$

$$\mu_{Umur}(R1)[34] = e^{-\frac{(34-36.6667)^2}{2*(3.78594)^2}} = 0.7803115503635186$$

$$\mu_{Umur}(R2)[34] = e^{-\frac{(34-55.71423)^2}{2*(4.49867)^2}} = 8.727115716694185E - 6$$

Proses perhitungan derajat keanggotaan dilakukan pada semua atribut data.

Untuk hasil derajat keanggotaan secara lengkap pada Tabel 3.13.

Tabel 3.13 Hasil Perhitungan Derajat Keanggotaan Data Uji

Hasil Derajat Keanggotaan Atribut Input ke-			
Aturan	1	2	3
1	0.7803115504	0.57981566493	0.846481724891
2	8.7271157E-6	0.18978676940	0.490269169691

b) Proses Perhitungan *Fire Strenght* ( $\alpha$ -predikat)

Proses pengujian selanjutnya setelah melakukan perhitungan derajat keanggotaan, dilakukan proses perhitungan ( $\alpha$ -predikat). Dimana digunakan operator *product* pada antesedennya :

$$[\text{Rule 1}] \alpha_1 = \mu_{Umur}(R1)[34] * \mu_{Tahun\ Operasi}(R1)[60] * \mu_{Node}(R1)[0]$$

$$= 0.7803 * 0.57982 * 0.846482 = 0.38297953401723095$$

$$[\text{Rule 2}] \alpha_2 = \mu_{Umur}(R2)[34] * \mu_{Tahun\ Operasi}(R2)[60] * \mu_{Node}(R2)[0]$$

$$= 8.72712E-6 * 0.189787 * 0.49027 = 8.120284614142145E-7$$

c) Proses Perhitungan Nilai Z Setiap Aturan

Perhitungan nilai Z untuk setiap aturan yang dibangkitkan sebagai berikut :

$$\begin{aligned} Z_1 &= p_{11}x_1 + p_{12}x_2 + p_{13}x_3 + p_{14} \\ &= (-4.11085*34) + (4.25321*60) + (-0.99246*0) + (-115.42353) \\ &= -9.699256565909309E-6 \end{aligned}$$

$$\begin{aligned} Z_2 &= p_{21}x_1 + p_{22}x_2 + p_{23}x_3 + p_{24} \\ &= (-0.174*34) + (-0.23351*60) + (0.03892*0) + (24.93611) \\ &= 5.009702230827433 \end{aligned}$$

d) Proses Defuzzifikasi

Proses defuzzifikasi dilakukan dengan menghitung nilai bobot menggunakan *weighted average* pada data uji.

$$Z = \frac{(\alpha_1 * Z_1) + (\alpha_2 * Z_2)}{(\alpha_1 + \alpha_2)} = 9.227732920646529E - 7$$

Dengan hasil Z dari proses defuzzifikasi di atas, bisa dilihat bahwa Z bernilai mendekati 0. Dengan demikian hasil perhitungan nilai kategori dengan menggunakan aturan hasil pembangkitan K-Means *clustering* sesuai dengan klasifikasi data penelitian Haberman's Survival dengan kategori hasil pengujian bertahan hidup lebih dari 5 tahun.

### 3.3.6 Sistematika Pengujian

Proses pengujian dilakukan dalam dua hal yaitu proses pengujian *cluster* dengan menganalisa varian terkecil untuk jumlah *cluster* ideal dan menguji tingkat akurasi dari hasil data uji dengan menggunakan aturan yang dibangkitkan. Proses analisa *cluster* dilakukan dengan menghitung varian yang digunakan untuk mengetahui jumlah *cluster* yang terbaik untuk digunakan dalam proses pembangkitan aturan. Sedangkan proses pengujian terhadap kategori ketahanan hidup penderita digunakan untuk mengetahui sejauh mana akurasi dari aturan yang dibangkitkan untuk menguji data dengan hasil yang sesuai dengan data set. Berikut dijelaskan dalam proses sistematika pengujian :

#### 3.3.6.1 Uji Jumlah Cluster Ideal

Proses analisa jumlah *cluster* ideal ini dilakukan dengan mengelompokkan data berdasarkan jumlah *cluster* uji, serta menghitung varian dari *cluster* uji tersebut. Untuk itu dilakukan pengujian jumlah *cluster* data uji

dengan beberapa variasi jumlah *cluster*. Jumlah *cluster* yang bisa di uji adalah  $2 \leq c < n$ . Dalam penelitian ini dilakukan pengujian jumlah *cluster* ideal sebanyak tiga kali dengan jumlah *cluster* uji  $c=2,3$ , dan 4 dengan jumlah data uji  $n=10$  data.

Proses pengujian jumlah *cluster* ideal ini, diuji dengan membandingkan hasil perhitungan varian di masing-masing *cluster*. Dengan jumlah varian terkecil merupakan jumlah *cluster* ideal untuk proses pengelompokan data. Untuk mempermudah proses perbandingan, hasil data uji varian dicatat pada Tabel 3.14.

Tabel 3.14 Tabel Data Hasil Perhitungan Batasan Varian

	Jumlah <i>Cluster</i> Uji			
	2	3	...	c
<b>Batasan Varian</b>				
<b>Varian Terkecil</b>				

### 3.3.6.2 Uji Tingkat Akurasi

Proses pengujian akurasi ini dilakukan pada proses perhitungan kategori data penderita kanker payudara apakah tergolong yang bertahan hidup lebih dari 5 tahun atau sebaliknya. Tujuan dari pengujian ini digunakan untuk mengetahui akurasi hasil dari rule yang dibangkit melalui K-Means *clustering* dengan mencocokkan hasil data uji perhitungan dengan data set Haberman's Survival. Dengan proses tersebut bisa diketahui hasil perhitungan data uji. Berapa data yang sesuai dengan data Haberman's Survival dan berapa data yang tidak sesuai dengan data Haberman's Survival. Proses perhitungan akurasi menggunakan rumus persamaan 2.24. Tabel uji coba akurasi sistem ketahanan hidup penderita kanker payudara ditunjukkan Tabel 3.15.

Tabel 3.15 Tabel Data Hasil Proses Perhitungan Akurasi Pengujian Data

Cluster	Percobaan Ke-i,	Hasil Perhitungan Kategori		Akurasi	Rata-Rata
		Hasil Benar	Hasil Salah		
	1				
	2				
	...				
	5				
		$\sum data\ benar$	$\sum data\ salah$		





## BAB IV IMPLEMENTASI

### 4.1 Lingkungan Implementasi

Dalam menerapkan metode penelitian ke dalam sistem serta dalam pengembangan sistem, perlu adanya beberapa faktor yang perlu diperhatikan dalam menerapkan metode dan pengembangan sistem. Tujuannya yaitu untuk memenuhi kebutuhan dari sistem yang nantinya akan dikembangkan dan metode yang diimplementasikan. Beberapa aspek yang perlu diperhatikan yaitu dari segi perangkat keras (*hardware*) maupun perangkat lunak (*software*).

#### 4.1.1 Lingkungan Implementasi Perangkat Keras

Dalam mengembangkan sistem dan penerapan metode dalam penelitian mengenai implementasi metode K-Means *clustering* untuk pembangkitan aturan *fuzzy* pada klasifikasi ketahanan hidup penderita kanker payudara digunakan beberapa komponen perangkat keras sebagai berikut :

1. Prosesor Intel Core i5-2430M CPU @ 2.40 GHz (4 CPUs).
2. Memory 4096MB RAM.
3. Hard disk 500GB.
4. Keyboard.
5. Mouse.
6. VGA AMD Radeon HD 6630M 2GB.

#### 4.1.2 Lingkungan Implementasi Perangkat Lunak

Pengembangan sistem dan penelitian mengenai implementasi metode K-Means *clustering* untuk pembangkitan aturan *fuzzy* pada klasifikasi ketahanan hidup penderita kanker payudara juga membutuhkan beberapa perangkat lunak. Beberapa perangkat lunak yang digunakan sebagai berikut :

1. Sistem operasi yang digunakan Windows 7 Ultimate
2. Aplikasi pembangun GUI dan code menggunakan NetBeans IDE 6.7.1
3. Bahasa pemrograman yang dipakai yaitu bahasa pemrograman java.
4. Komponen java yang digunakan yaitu JDK 1.6.0.
5. Basis data yang digunakan yaitu MySQL Server Versi 5.1.33 dengan menggunakan tool XAMPP Versi 1.7.1.

6. Microsoft Word 2010, Microsoft Excel 2010, Microsoft Visio 2010.

## 4.2 Implementasi Program

Pada implementasi program, program dibuat dengan menggunakan method-method. Di mana method-method tersebut merupakan tahapan proses yang dilakukan dalam program. Method-method ini mewakili proses dari proses pelatihan sampai proses pengujian data. Proses implementasi program ini dimulai dari load data dari database, proses *clustering*, proses pembangkitan aturan *fuzzy*, sampai proses pengujian data, dan perhitungan akurasi.

### 4.2.1 Tahapan Proses Pelatihan

Pada tahapan ini dilakukan beberapa proses dengan menggunakan data training yang diambil dari data set Haberman's Survival. Proses-proses yang dilakukan dalam tahapan pelatihan ini antara lain proses input data, proses hitung data, proses *clustering*, proses hitung derajat keanggotaan, proses hitung alva predikat, hitung *Z percluster*, dan proses defuzzifikasi.

#### 4.2.1.1 Proses Input Data

Pada tahapan ini dilakukan proses pengambilan data training dari database yang diambil dari *tb\_datahaberman*. Data input ini digunakan untuk data training dalam proses *clustering*. Tahapan proses ini digambarkan dengan Kode Program 4.1.

```
public static void inputDataDb()
{
    int countData=0, countCentroid=0;
    try{
        String sql = "select tb_datahaberman.umur as umur,
tb_datahaberman.tahunOperasi as tahunOperasi,
tb_datahaberman.nodePositif as nodePositif,
tb_datahaberman.nilaiKetahanan as nilaiKetahanan,
tb_datahaberman.statusKetahanan as statusKetahanan
from tb_datahaberman, tb_datatraining where
tb_datahaberman.idData=tb_datatraining.idData";
        ResultSet rs = statement.executeQuery(sql);
        rs.beforeFirst();
        while(rs.next()){
            countData++;
        }
        jumData=countData;
        rs.beforeFirst();
        int id=0;
        while (rs.next()) {
            atrib[id][0]= rs.getDouble("umur");
            atrib[id][1]= rs.getDouble("tahunOperasi");
            atrib[id][2]=rs.getDouble("nodePositif");
```

```

        atrib[id][3]=rs.getDouble("nilaiKetahanan");
        tatusAsli[id]=rs.getString("statusKetahanan");
        id++;
    }
    String Ssql = "select tb_datahaberman.
umur as umur,tb_datahaberman.tahunOperasi as
tahunOperasi,tb_datahaberman.nodePositif
as nodePositif,
tb_datahaberman.nilaiKetahanan as nilaiKetahanan
from tb_datahaberman, tb_datacentroidawal where
b_datahaberman.idData=tb_datacentroidawal.idData";
ResultSet rsCentroid=statement.executeQuery(Ssql);
while(rsCentroid.next()){
    countCentroid++;
}
rsCentroid.beforeFirst();
int idCentroid=0;
while(rsCentroid.next()){
    if(idCentroid<jumCluster)
    {
        bilRand[idCentroid][0]=
            rsCentroid.getDouble("umur");
        bilRand[idCentroid][1]=
            rsCentroid.getDouble("tahunOperasi");
bilRand[idCentroid][2]=
rsCentroid.getDouble("nodePositif");

bilRand[idCentroid][3]=rsCentroid.getDouble("nilaiKetahanan");
    }
    idCentroid++;
}
for(int x=0; x<jumData; x++)
{
    tempAnggotaCluster[x]=0.0;
}
for(int x=0; x<jumCluster; x++)
{
    jumCluster_Array[x]=0.0;
    tempCluster_Array[x]=0.0;
}
}catch(Exception e){
    System.out.println("Errorrrrrrrrrrrrrrrrrrrrr");
}
}

```

Kode Program 4.1 Method Input Data

#### 4.2.1.2 Proses K-Means Clustering

Proses *clustering* merupakan salah satu tahapan proses pelatihan, di mana proses ini melakukan pengelompokkan data dengan menghitung jarak data training dengan centroid menggunakan persamaan Euclidean seperti persamaan 2.1. Proses pengelompokkan data dilihat dari hasil perhitungan jarak. Semakin kecil jarak data dengan centroid di masing-masing *cluster*,

maka semakin dekat kemiripan data dengan centroid yang menunjukkan data tersebut nantinya tergolong ke dalam *cluster* yang mana. Selanjutnya akan dicari centroid baru yang nantinya digunakan sebagai centroid untuk menghitung jarak antar data jika proses pengelompokan masih berlanjut. Proses *clustering* ini dilakukan berulang kali sampai tidak ada lagi anggota *cluster* yang berubah. Kode program proses *clustering* secara umum digambarkan pada Kode Program 4.2.

```
public static void prosesClustering()
{
    boolean satu=true, dua=true, tiga=true;
    for(int i=0; i<jumData; i++)
    {
        for(int x=0; x<jumCluster; x++)
        {
            if (jumCluster_Array[x]==0.0 &&
                tempCluster_Array[x]==0.0)
            {
                if(satu=false)
                {
                    satu=false;
                }else {
                    satu=true;
                }
            }else{
                satu=false;
            }
        }
        if(satu=true)
        {
            prosesEuclidean();
            prosesPengelompokanData();
            prosesCariCentroid();
        }else
        {
            for(int a=0; a<jumCluster; a++)
            {
                if(jumCluster_Array[a]==
                    tempCluster_Array[a])
                {
                    if(dua=false)
                    {
                        dua=false;
                    }else
                    {
                        dua=true;
                    }
                }else
                {
                    dua=false;
                }
            }
            for(int b=0; b<jumData; b++)
```

```
{
    if(anggotaCluster[b]==
tempAnggotaCluster[b])
    {
        if(tiga=false)
        {
            tiga=false;
        }else
        {
            tiga=true;
        }
    }else
    {
        tiga=false;
    }
}
if((dua=true) && (tiga=true))
{
    break;
}else
{
    for(int c=0;c<jumCluster;c++)
    {
        tempCluster_Array[c]=
jumCluster_Array[c];
    }
    for(int j=0; j<jumData; j++)
    {
        tempAnggotaCluster[j]=
anggotaCluster[j];
    }
    prosesEuclidean();
    prosesPengelompokkanData();
    prosesCariCentroid();
}
}
}
prosesHitungMeanDanStandarDeviasi();
prosesAnalisisVarian();
prosesHitungVarianTerkecil();
prosesHitungDerajatKeanggotaan();
prosesHitungMatriksU();
prosesHitungUTernormalisasi();
prosesHitungMatriksUTernormalisasi();
prosesFuzzifikasi();
prosesHitungAlvaPredikat();
prosesHitungZ();
prosesDefuzzifikasi();
}
```

Kode Program 4.2 Proses K-Means *Clustering*

Dan proses selanjutnya, data akan dikelompokkan pada *cluster* yang memiliki hasil perhitungan jarak terkecil pada masing-masing *cluster*. Proses

perhitungan Euclidean dan proses pengelompokkan data akan ditunjukkan pada

### Kode Program 4.3.

```
public static void prosesEuclidean(){
    double sumTemp=0.0;
    for(int k=0; k<jumCluster; k++)
    {
        for(int i=0; i<jumData; i++)
        {
            for (int j=0; j<jumAtribut; j++)
            {
                temp [i][k]+=((Math.pow((atrib[i][j]-
                bilRand[k][j]), 2.0)));
                sumTemp = temp[i][k];
            }
            temp [i][k] = Math.sqrt(sumTemp);
        }
    }
}

public static void prosesPengelompokkanData(){

    for(int i=0; i<jumData; i++)
    {
        for (int k=0; k<1; k++)
        {
            if(temp [i][k]< temp[i][k+1])
            {
                anggotaCluster[i]=1;
            } else if(temp [i][k]> temp[i][k+1])
            {
                anggotaCluster[i]=2;
            }
        }
    }
}
```

### Kode Program 4.3 Proses Perhitungan Jarak (Euclidean)

Dan untuk pembentukan centroid baru untuk masing-masing iterasi akan digambarkan dengan Kode Program 4.4.

```
public static void prosesCariCentroid()
{
    double subTemp_1 =0.0;
    double subTemp_2 =0.0;
    double x =0.0;
    double n=0.0;
    for(int k=0; k<jumCluster; k++)
    {
        for(int j=0; j<jumAtribut; j++)
        {
            bilRand [k][j]=0.0;
            subTemp_1=0.0;
            subTemp_2=0.0;
            x=0.0;
            n=0.0;
            for (int i=0; i<jumData; i++)
```

```
        {
            if((anggotaCluster[i]==1.0) && (k==0))
            {
                bilRand [k][j]=subTemp_1+atrib
[i][j];
                subTemp_1 = bilRand [k][j];
                x++;
                jumCluster_1 = x;
            }else if((anggotaCluster[i]==2.0) &&
(k==1))
            {
                bilRand [k][j]+= atrib [i][j];
                subTemp_2 = bilRand [k][j];
                n++;
                jumCluster_2 = n;
            }
        }
    }
    if(k==0)
    {
        if(x==0)
        {
            bilRand [k][j]=0.0;
            centroidAkhir [k][j]=bilRand[k][j];
        }else
        {
            bilRand [k][j]=subTemp_1/x;
            centroidAkhir [k][j]=bilRand[k][j];
        }
    }else if(k==1)
    {
        if(n==0)
        {
            bilRand [k][j]=0.0;
            centroidAkhir [k][j]=bilRand[k][j];
        }else
        {
            bilRand [k][j]=subTemp_2/n;
            centroidAkhir [k][j]=bilRand[k][j];
        }
    }
}
}
```

Kode Program 4.4 Proses Perhitungan Centroid Baru

#### 4.2.1.3 Proses Perhitungan Mean dan Standar Deviasi

Proses perhitungan mean dan standar deviasi dilakukan setelah proses K-Means *clustering* menghasilkan centroid akhir dari hasil perhitungan pengelompokan data. Hasil perhitungan mean dan standar deviasi ini nantinya digunakan untuk proses analisis varian dan proses perhitungan derajat keanggotaan baik data training maupun data uji pada proses pengujian data.

Proses perhitungan mean dan standar deviasi digambarkan dengan Kode Program 4.5 dengan menggunakan persamaan 2.2 dan 2.3.

```

public static void prosesHitungMeanDanStandarDeviasi(){
    double tempSD=0.0, a_1=0.0;
    for (int k=0; k<jumCluster; k++)
    {
        for(int j=0; j<jumAtribut; j++)
        {
            a_1=0.0;
            tempStandarDeviasi=0.0;
            for (int i=0; i<jumData; i++)
            {
                if((anggotaCluster[i]==1.0) && (k==0))
                {
                    standarDeviasi[i][j]=(Math.pow((atrib[i][j]-
                    centroidAkhir[k][j]),2));
                    tempStandarDeviasi=standarDeviasi[i][j]+tempStandarDeviasi;
                    a_1++;
                }else if((anggotaCluster[i]==2.0)
                && (k==1))
                {
                    standarDeviasi[i][j]+=(Math.pow((atrib[i][j]-
                    centroidAkhir[k][j]),2));
                    tempStandarDeviasi=standarDeviasi[i][j]+tempStandarDeviasi;
                    a_1++;
                }
            }
            if (a_1==0.0)
            {
                standarDeviasi[k][j]=0.0;
            }else if (a_1-1==0.0)
            {
                standarDeviasi[k][j]=0.0;
            }else
            {
                standarDeviasi[k][j]=Math.sqrt(((tempStandarDeviasi)/(a_1-1)));
            }
        }
    }
    double tempRata =0.0;
    for (int x=0; x<jumAtribut; x++)
    {
        tempRata=0.0;
        for (int y=0; y<jumCluster; y++)
        {
            tempRata=centroidAkhir[y][x]+tempRata;
        }
        rataRata[x]=tempRata/jumCluster;
    }
}

```

Kode Program 4.5 Proses Perhitungan Mean dan Standar Deviasi



#### 4.2.1.4 Proses Analisis Varian

Proses ini dilakukan untuk menghitung varian di masing-masing centroid. Hasil perhitungan varian nantinya akan digunakan untuk menghitung varian total di masing-masing *cluster*. Proses perhitungannya digambarkan dengan Kode Program 4.6.

```
public static void prosesAnalisisVarian()
{
    double tempSD=0.0, v_1=0.0, v_2=0.0;
    for (int k=0; k<jumCluster; k++)
    {
        for(int j=0; j<jumAtribut; j++)
        {
            v_1=0.0;
            v_2=0.0;
            tempVPC=0.0;
            for (int i=0; i<jumData; i++)
            {
                if((anggotaCluster[i]==1.0) && (k==0))
                {
                    varianPerCluster[k][j]=
                    (Math.pow((atrib[i][j]-
                    centroidAkhir[k][j]),2));
                    tempVPC=varianPerCluster[k][j]+tempVPC;
                    varianPerCluster[k][j]=tempVPC;
                    v_1++;
                    jumDC_1=v_1;
                }else if((anggotaCluster[i]==2.0)
                && (k==1))
                {
                    varianPerCluster[k][j]=
                    Math.pow((atrib[i][j]-
                    centroidAkhir[k][j]),2);
                    tempVPC=varianPerCluster[k][j]+tempVPC;
                    varianPerCluster[k][j]=tempVPC;
                    v_2++;
                    jumDC_2=v_2;
                }
            }
        }
    }
    double    tmpVPC_1=0.0,    tmpVPC_2=0.0,    vpc_1=0.0,
    vpc_2=0.0;
    for (int x=0; x<jumCluster; x++)
    {
        for (int y=0; y<jumAtribut; y++)
        {
            if (x==0)
            {
                tmpVPC_1=    varianPerCluster
                [x][y]+tmpVPC_1;
            } else if (x==1)
            {
                tmpVPC_2=    varianPerCluster
```

```

[x][y]+tmpVPC_2;
    }
}
if (jumDC_1==0.0)
{
    vpc_1=0.0;
}else if (jumDC_1-1==0.0)
{
    vpc_1=0.0;
}else
{
    vpc_1=(tmpVPC_1)*((1/(jumDC_1-1)));
}

if (jumDC_2==0.0)
{
    vpc_2=0.0;
}else if (jumDC_2-1==0.0)
{
    vpc_2=0.0;
}else
{
    vpc_2=(tmpVPC_2)*((1/(jumDC_2-1)));
}
}

```

Kode Program 4.6 Analisis Varian

#### 4.2.1.5 Proses Perhitungan Varian Cluster

Dari proses analisis varian, dilanjutkan proses perhitungan varian terkecil. Perhitungan varian terkecil ini digunakan untuk menentukan nilai kelayakan dari suatu *cluster* dalam melakukan proses *clustering* dan proses pengujian. Dari hasil perhitungan varian nantinya akan disesuaikan apakah proses pengujian data dengan beberapa skenario bisa menghasilkan tingkat akurasi yang terbaik untuk setiap *cluster*. Untuk gambaran proses perhitungan varian ditunjukkan dengan Kode Program 4.7. Pada proses perhitungan varian digunakan persamaan 2.5, persamaan 2.6, dan persamaan 2.7.

```

public static void prosesHitungVarianTerkecil()
{
    double tempVarian=0.0, tempVar_1=0.0, tempVar_2=0.0;
    for (int x=0; x<jumCluster; x++)
    {
        tempVarian=0.0;
        for (int y=0; y<jumAtribut; y++)
        {
            if (x==0)
            {
                tempVarian=(jumDC_1*(Math.pow
                ((centroidAkhir[x][y]-rataRata[y]),2)))+
                tempVarian;
            }
        }
    }
}

```

```

        tempVar_1=tempVarian;
    }else if (x==1)
    {
        tempVarian=(jumDC_2*(Math.pow
        ((centroidAkhir[x][y]-rataRata[y]), 2)))+
        tempVarian;
        tempVar_2=tempVarian;
    }
    }
}
varianBetween =((1.0/(jumCluster-1.0))*
(tempVar_1+tempVar_2));
varianTotal=varianWithin/varianBetween;
}

```

Kode Program 4.7 Proses Perhitungan Varian

#### 4.2.1.6 Proses Perhitungan Derajat Keanggotaan

Proses selanjutnya dilakukan proses perhitungan derajat keanggotaan dari data training. Proses ini menggunakan fungsi Gauss dengan persamaan 2.17. Proses ini melibatkan standar deviasi untuk proses perhitungannya. Hasil proses perhitungan ini digunakan untuk proses pembentukan matriks U. Tahap prosesnya ditunjukkan pada Kode Program 4.8.

```

public static void prosesHitungDerajatKeanggotaan()
{
    double tempDerj=0.0, tempDer_1=0.0, tempDer_2=0.0;
    for (int x=0; x<jumCluster; x++)
    {
        for(int y=0; y<jumData; y++)
        {
            tempDerj=0.0;
            for (int z=0; z<jumAtribut; z++)
            {
                if (standarDeviasi [x][z]==0.0)
                {
                    tempDerj+=0.0;
                }else{
                    tempDerj=(Math.pow(atrib[y][z]-
                    centroidAkhir[x][z],2)/(2*Math.pow(standarDeviasi[x][z],2)))+
                    tempDerj;
                }
            }
            derajat[y][x]=Math.exp((-1)*(tempDerj));
        }
    }
}

```

Kode Program 4.8 Proses Perhitungan Derajat Keanggotaan

#### 4.2.1.7 Proses Pembentukan Matriks U

Dari proses perhitungan nilai derajat keanggotaan untuk setiap *cluster* digunakan untuk membentuk matriks U dengan ukuran matriks  $m \times n$ . Di mana  $m$  adalah jumlah data dan  $n$  adalah jumlah atribut utama ditambah satu dikali jumlah aturan yang akan dibentuk. Contoh akan dikelompokkan 10 data menggunakan dua *cluster* berarti akan dibentuk dua aturan. Maka ukuran matriks U nantinya  $(10 * (2 \text{ aturan} * (3 \text{ atribut} + 1)))$  sehingga ukuran matriks U  $10 \times 8$ . Pembentukan matriks U menggunakan persamaan 2.8 di mana. Untuk gambaran proses pembentukan matriks U ditunjukkan pada Kode Program 4.9.

```
public static void prosesHitungMatriksU()
{
    for (int k=0; k<jumCluster; k++)
    {
        for (int i=0; i<jumData; i++)
        {
            for (int j=0; j<jumAtribut-1; j++)
            {
                d [k][i][j] = atrib [i][j]*derajat[i][k];
            }
            d[k][i][jumAtribut-1]=derajat[i][k];
        }
    }
    int a=0,b=0;
    for(int i=0;i<jumData;i++){
        for(int k=0;k<jumCluster;k++){
            for(int j=0;j<jumAtribut;j++){
                dki[j][a][b]=d[k][i][j];
                if(b==dki[j][a].length-1){
                    b=0;
                    a++;
                }
                else b++;
            }
        }
    }
    SimpleMatrix u = new SimpleMatrix(dki[j]);
}
```

Kode Program 4.9 Proses Perhitungan Matriks U

Setelah proses pembentukan matriks U, dilanjutkan proses perhitungan LSE. Di mana proses ini menghasilkan *output* berupa matriks  $k$  yaitu matriks koefisien output setiap atribut di masing-masing *cluster* yang akan dibangkitkan. Proses perhitungan LSE ini melibatkan operasi-operasi matriks seperti perkalian, transpose, dan invers matriks dari matriks U. Proses perhitungan LSE sendiri menggunakan persamaan 2.12 seperti yang sudah

dijelaskan pada Bab II. Untuk proses perhitungan LSE ditunjukkan dengan Kode Program 4.10.

```

public static void prosesHitungUTernormalisasi()
{
    double tempMiu=0.0;
    for(int i=0; i<jumData; i++)
    {
        tempMiu=0.0;
        for(int j=0; j<jumAtribut; j++)
        {
            tempMiu=derajat[i][j]+tempMiu;
            sumMiu[i]=tempMiu;
        }
    }
    for(int x=0; x<jumData; x++)
    {
        for(int y=0; y<(jumAtribut*jumCluster); y++)
        {
            uTernormalisasi[x][y]=dkij[x][y]/sumMiu[x];
        }
    }
    SimpleMatrix uT = new SimpleMatrix(uTernormalisasi);
}

public static void prosesHitungMatriksUTernormalisasi()
{
    SimpleMatrix u = new SimpleMatrix(uTernormalisasi);
    SimpleMatrix uTranpose = u.transpose();
    SimpleMatrix uttXut = uTranpose.mult(u);
    for(int i=0; i<(jumAtribut*jumCluster); i++)
    {
        for(int j=0; j<(jumAtribut*jumCluster); j++)
        {
            pN[i][j]=uttXut.get(i, j);
        }
    }
    SimpleMatrix p = new SimpleMatrix(pN);
    SimpleMatrix pnInverse = p.pseudoInverse();

    SimpleMatrix x = pnInverse.mult(uTranpose);
    double Y [][]=new double[jumData][1];
    for (int b= 0; b<jumData; b++)
    {
        Y [b][0]= atrib[b][jumAtribut-1];
    }
    SimpleMatrix y = new SimpleMatrix(Y);
    SimpleMatrix kO = x.mult(y);
    double koef [][]=
    new double[kO.numRows()][kO.numCols()];
    for (int i=0; i<koef.length; i++)
    {
        for (int j=0; j<koef[0].length; j++)
        {
            koef [i][j]=kO.get(i,j);
        }
    }
}

```

```

    }
    int a=0;
    for (int i=0; i<koefOutput.length; i++)
    {
        for (int j=0; j<koefOutput[0].length; j++)
        {
            koefOutput [i][j]=koef[a][0];
            a++;
        }
    }
}

```

Kode Program 4.10 Proses *Least Square Estimator*

Dari proses ini akan dihasilkan nilai koefisien output yang nantinya digunakan sebagai fungsi yang terbentuk sebagai konsekuen di setiap rule yang dibangkitkan. Dan nantinya koefisien output ini digunakan untuk menghitung bobot nilai untuk setiap *cluster*. Sehingga secara lengkap proses pembangkitan aturan telah dilengkapi dengan centroid sebagai anteseden untuk setiap *clusternya*, dan koefisien output sebagai konsekuen untuk setiap *clusternya*.

## 4.2.2 Tahap Proses Pengujian

### 4.2.2.1 Proses Fuzzifikasi

Tahapan ini merupakan proses awal uji data, di mana akan dihitung derajat keanggotaan untuk data uji. Proses ini mengubah data input bilangan *crisp* menjadi bilangan *fuzzy*. Proses perhitungan derajat keanggotaan dilakukan untuk setiap atribut dan di masing-masing aturan yang terbentuk. Untuk proses perhitungan derajat keanggotaan ditunjukkan dengan Kode Program 4.11.

```

public static void prosesFuzzifikasi()
{
    double tempDerj=0.0, tempDer_1=0.0, tempDer_2=0.0;
    for (int x=0; x<jumCluster; x++)
    {
        for (int z=0; z<(jumAtribut-1); z++)
        {
            if (standarDeviasi [x][z]==0.0)
            {
                tempDerj=0.0;
            }else{tempDerj=(Math.pow(dataUji[0][z]-
            centroidAkhir[x][z],2)/(2.0*Math.pow(standarDeviasi[x][z],2)));
            }
            miuDataUji[x][z]=Math.exp((-1)*(tempDerj));
        }
    }
}

```

Kode Program 4.11 Proses Fuzzifikasi

Dari perhitungan derajat keanggotaan, proses dilanjutkan dengan implikasi produk di mana merupakan operasi dari anteseden di setiap rule yang nantinya akan dihasilkan  $\alpha$ -predikat untuk setiap rule yang dibangkitkan. Proses perhitungan  $\alpha$ -predikat ditunjukkan dengan Kode Program 4.12.

```
public static void prosesHitungAlvaPredikat()
{
    sumAlva=0.0;
    for (int i=0; i<jumCluster;i++)
    {
        alvaPredikat[i]=0.0;
    }
    double tempAlva=1.0;
    for (int i=0; i<jumCluster; i++)
    {
        tempAlva=1.0;
        for (int j=0; j<jumAtribut-1; j++)
        {
            tempAlva = miuDataUji[i][j]*tempAlva;
        }
        alvaPredikat [i]=tempAlva;
    }
    for(int x=0; x<jumCluster; x++)
    {
        sumAlva = alvaPredikat [x]+sumAlva;
    }
}
```

Kode Program 4.12 Proses Perhitungan  $\alpha$ -Predikat

#### 4.2.2.2 Proses Perhitungan Z Untuk Setiap Aturan

Proses perhitungan z di sini berfungsi untuk menghitung nilai konsekuen dari masing-masing *cluster*. Hasil dari perhitungan tersebut nantinya digunakan untuk menghitung nilai bobot pada proses defuzzifikasi. Proses perhitungan konsekuen untuk setiap aturan ditunjukkan pada Kode Program 4.13.

```
public static void prosesHitungZ()
{
    for (int i=0; i<jumCluster;i++)
    { zKecil[i]=0.0;}double tempZKecil=0.0;
    for (int i=0; i<jumCluster; i++)
    { tempZKecil=0.0;
        for (int j=0; j<(jumAtribut-1);j++)
        { tempZKecil =
            (dataUji[0][j]*kofOutput[i][j])+tempZKecil;
        }
        zKecil[i]=tempZKecil+kofOutput[i][3];
    }
}
```

Kode Program 4.13 Proses Perhitungan Z Untuk Setiap Aturan

### 4.2.2.3 Proses Defuzzifikasi

Proses selanjutnya dalam menguji data yaitu proses defuzzifikasi. Di mana proses ini mengubah kembali bilangan *fuzzy* yang terbentuk dari proses fuzzifikasi sampai proses perhitungan konsekuen untuk setiap aturan, diubah menjadi bilangan *crisp*. Proses ini menghitung nilai rata-rata terbobot dengan menggunakan metode *weighted average* menggunakan persamaan 2.22 di mana hasil dari perhitungannya digunakan untuk menentukan nilai data uji tergolong data penderita dengan ketahanan hidup lebih dari 5 tahun atau data tergolong dari data penderita dengan ketahanan hidup kurang dari 5 tahun. Proses defuzzifikasi ditunjukkan dengan Kode Program 4.14.

```
public static void prosesDefuzzifikasi()
{
    double tempZTotal=0.0;zTotal=0.0;
    for (int x=0; x<jumCluster; x++)
    {

tempZTotal=(alvaPredikat[x]*zKecil[x])+tempZTotal;
    }
    zTotal=tempZTotal/sumAlva;
    System.out.println("");
    System.out.println("Z Total = "+zTotal);
}
}
```

Kode Program 4.14 Proses Defuzzifikasi

Pada proses pengujian, data uji yang dibutuhkan hanya 3 atribut utama yaitu umur, tahun, dan node positif. Data atribut tersebut digunakan dimulai dari proses perhitungan derajat keanggotaan sampai proses defuzzifikasi, sehingga didapatkan hasil nilai rata-rata bobot dari data untuk menentukan kategori data.

## 4.3 Implementasi Antarmuka

Seperti yang sudah dijelaskan pada sub bab 3.3.4, bahwa sistem secara umum ada dua proses bagian utama yaitu proses pelatihan dan proses pengujian data. Di mana kedua proses tersebut berada dalam satu form yang sama. Sistem yang dibuat ini terdiri dari tiga form yaitu form utama, form data Haberman's Survival, dan form data uji.



### 4.3.1 Form Utama

Form utama yang dibuat ini terdiri dari dua bagian proses utama yaitu proses pelatihan dan proses pengujian. Dalam proses pelatihan, tidak ada inputan yang dilakukan. Hanya proses yang harus dipilih dan dioperasikan. Sedangkan untuk proses pengujian perdata, bisa dilakukan dengan menggunakan inputan secara otomatis yang diambil dari data Haberman's Survival atau bisa juga dilakukan inputan manual pada field text yang disediakan. Form utama dan bagian-bagian dari form ditunjukkan pada Gambar 4.1.

The screenshot shows a software interface for breast cancer survival analysis. The main window is titled "Form Utama Aplikasi Ketahanan Hidup Penderita Kanker Payudara". It contains several functional areas: a "Penentuan Cluster" section with a dropdown menu for "Jumlah Cluster", input fields for "Jumlah Data Latih" and "Hasil Varian", and "Proses" and "Reset" buttons; a "Hasil Pembangkitan Aturan" section with a large empty text box; a "Hasil Data Tercluster" section with another large empty text box; a header row with labels: "Umur || Tahun Operasi || Node Positif || Nilai Hasil Uji (Z) || Status Uji || Status Penelitian"; a "Hasil Uji Semua Data Dan Akurasi Perhitungan" section with a large empty text box; and a "Proses Uji Data" section with a "Load" button, input fields for "Umur Pasien", "Tahun Operasi: 19", "Node Positif", and "Hasil Perhitungan Klasifikasi", and a "Hasil Keputusan Status Data Uji" field with "Proses" and "Reset" buttons.

Gambar 4.1 Form Utama

Proses untuk menjalankan dan melakukan proses pelatihan dan pengujian data pengguna harus memilih jumlah *cluster* yang ingin digunakan untuk mengelompokkan (dengan asumsi semua data training sudah diinputkan). Selanjutnya pengguna akan melihat pada text field jumlah data, bahwa data yang nantinya akan digunakan untuk proses pengelompokkan menggunakan K-Means *clustering* sebanyak jumlah yang tertera pada text field jumlah data. Untuk menjalankan proses pelatihan dan pembangkitan aturan pengguna dilanjutkan menekan tombol Proses pada bagian Penentuan *cluster*. Jika proses berjalan

maka tampilan form utama setelah ditekan tombol proses ditunjukkan pada Gambar 4.2.

Gambar 4.2 Tampilan Form Utama Setelah Proses Pelatihan Dilakukan

Misal dicontohkan pengguna memilih jumlah *cluster* tiga dan data yang digunakan untuk pelatihan sejumlah 12 data. Dari hasil proses pelatihan, pengguna bisa melihat pada text area hasil pembangkitan aturan. Terdapat tiga aturan yang terbentuk, di mana jumlah hasil pembangkitan aturan sesuai jumlah *cluster* yang terbentuk. Dan pada text area hasil uji semua data dan akurasi perhitungan, pengguna bisa melihat bahwa pada text area dilakukan pengujian data training sendiri. Di mana data yang diuji diluar dari data uji skenario. Dari uji data training dengan menggunakan aturan yang dibangkitkan, dihitung akurasi dari data training yang diuji. Hasil perhitungan akurasi bisa dilihat pada bagian bawah text area hasil semua uji data dan akurasi perhitungan.

Dari proses pelatihan, hasil dari pembangkitan aturan *fuzzy* bisa digunakan untuk proses pengujian perdata dengan jumlah *cluster* tertentu. Misal dengan menggunakan jumlah *cluster* tiga dan menggunakan aturan yang dibangkitkan pengguna ingin menguji suatu data. Pengguna bisa menekan

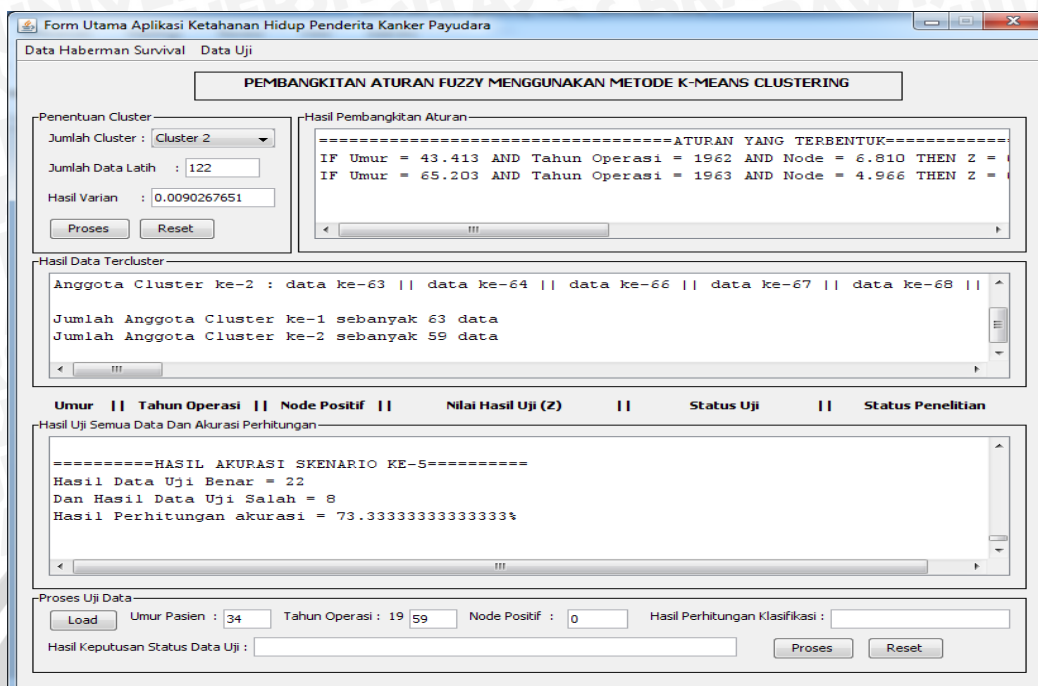
tombol Load dan akan muncul form data Haberman's Survival yang ditunjukkan pada Gambar 4.3.

The screenshot shows a software application window titled "Form Utama Aplikasi Ketahanan Hidup Penderita Kanker Payudara". The main window has a tab "Data Haberman Survival" and a sub-tab "Data Uji". The main window contains a section titled "PEMBANGKITAN ATURAN FUZZY MENGGUNAKAN METODE K-MEANS CLUSTERING" with a "Load" button. A modal window titled "Form Data Uji Haberman Survival" is open, displaying a table of patient data for testing. The table has the following columns: ID DATA, UMUR, TAHUN OPERASI, NODE, NILAI KETAHANAN, and STATUS KETAHANAN. The data is as follows:

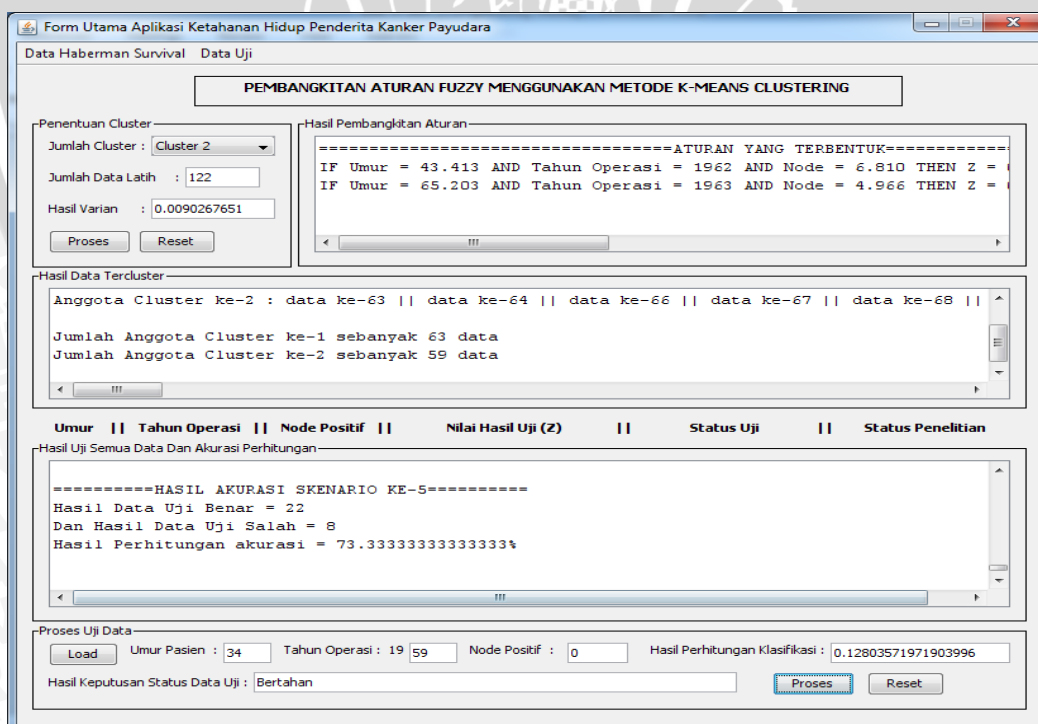
ID DATA	UMUR	TAHUN OPERASI	NODE	NILAI KETAHANAN	STATUS KETAHANAN
H001	30	64	1	0	Bertahan
H002	30	62	3	0	Bertahan
H003	30	65	0	0	Bertahan
H004	31	59	2	0	Bertahan
H005	31	65	4	0	Bertahan
H006	33	58	10	0	Bertahan
H007	33	60	0	0	Bertahan
H008	34	59	0	1	Tidak Bertahan
H009	34	66	9	1	Tidak Bertahan
H010	34	58	30	0	Bertahan
H011	34	60	1	0	Bertahan
H012	34	61	10	0	Bertahan
H013	34	67	7	0	Bertahan
H014	34	60	0	0	Bertahan

Gambar 4.3 Form Data Uji Haberman's Survival Setelah Tombol Load Ditekan

Untuk menginputkan data yang ingin diuji, pengguna bisa memilih data pada tabel Data Haberman pada form data uji dengan meng-klik tabel tepat pada data yang ingin diuji. Kemudian klik tombol *close* (x), dan data telah terinput pada proses data uji. Hasil inputan data bisa dilihat pada Gambar 4.4. Proses input data uji bisa juga dilakukan langsung input pada text field umur pasien, tahun operasi, dan node positif.



Gambar 4.4 Hasil Proses Inputan Menggunakan Form Data Haberman's Setelah inputan lengkap, pengguna bisa memproses data uji dengan menekan tombol proses pada proses uji data. Hasil dari proses pengujian data tersebut ditunjukkan dengan Gambar 4.5.



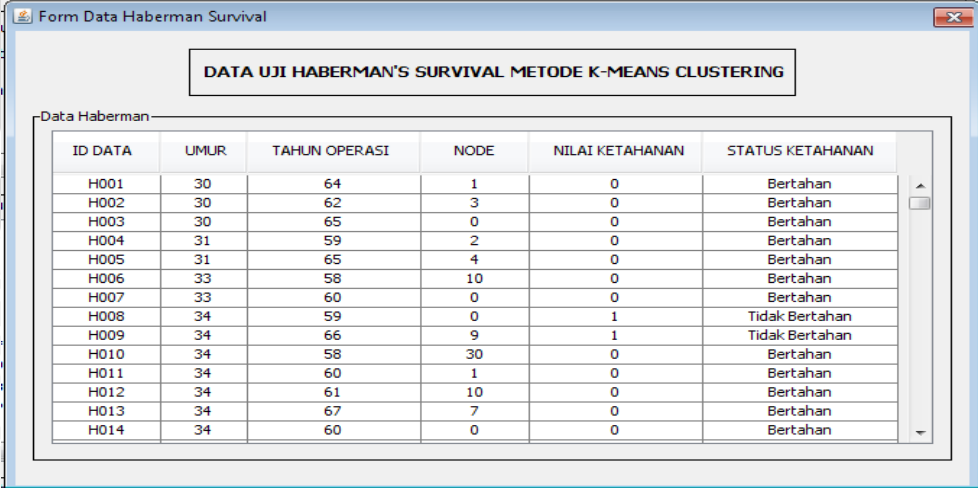
Gambar 4.5 Hasil Proses Pengujian Data Haberman's Survival

Bisa dilihat bahwa data uji dengan data umur 34, tahun operasi 1959, dan node positif 0. Hasil pengujian data nilai rata-rata terbobot yaitu sebesar 0.12803571971903996, di mana nilai tersebut lebih mendekati dengan kategori kelas01. Maka data tersebut tergolong data penderita kanker payudara dengan kriteria status penderita yaitu bertahan.

Untuk menghitung akurasi dari pengujian data, dari contoh Gambar 4.5, perlu dibandingkan hasil pengujian klasifikasi dengan data penelitian asli pada data Haberman's Survival. Bisa kita lihat pada Gambar 4.3 bahwa data dengan idData H008, umur 34, tahun operasi 1959, dan node positif 0 mempunyai nilai ketahanan 1 dan kategori status tidak bertahan. Artinya pengujian untuk data pertama tidak tepat, dengan membandingkan seperti itu dan menguji sebanyak n data, sehingga nantinya bisa dihitung seberapa banyak data uji yang benar dari data semua data uji dan bisa dihitung akurasi data seluruh pengujian.

#### 4.3.2 Form Data Haberman's Survival

Form ini digunakan untuk melihat seluruh data Haberman's Survival. Form ini bisa juga sebagai media input data pengujian klasifikasi data Haberman's Survival. Tampilan dari Form Data Haberman's Survival ditunjukkan Gambar 4.6.



ID DATA	UMUR	TAHUN OPERASI	NODE	NILAI KETAHANAN	STATUS KETAHANAN
H001	30	64	1	0	Bertahan
H002	30	62	3	0	Bertahan
H003	30	65	0	0	Bertahan
H004	31	59	2	0	Bertahan
H005	31	65	4	0	Bertahan
H006	33	58	10	0	Bertahan
H007	33	60	0	0	Bertahan
H008	34	59	0	1	Tidak Bertahan
H009	34	66	9	1	Tidak Bertahan
H010	34	58	30	0	Bertahan
H011	34	60	1	0	Bertahan
H012	34	61	10	0	Bertahan
H013	34	67	7	0	Bertahan
H014	34	60	0	0	Bertahan

Gambar 4.6 Form Data Haberman's Survival

#### 4.3.3 Form Data Uji

Form data uji berfungsi untuk melihat, menambah, dan menghapus data training. Form ini bisa diakses pada menu bar Data Uji. Pada form ini

ditampilkan data training dengan menampilkan id data, umur, tahun, node, nilai ketahanan, dan status kategori data. Untuk tampilan dari form data uji ditunjukkan pada Gambar 4.7.

ID DATA UJI	ID DATA	UMUR	TAHUN OPERASI	NODE	NILAI KETAHANAN	STATUS KETAHANAN
1	H014	34	60	0	0	Bertahan
2	H174	54	66	0	0	Bertahan
3	H252	63	61	9	0	Bertahan
4	H220	59	64	4	0	Bertahan
5	H187	55	66	0	0	Bertahan
6	H198	57	61	5	1	Tidak Bertahan
7	H015	35	64	13	0	Bertahan
8	H162	53	65	12	1	Tidak Bertahan
9	H046	41	67	0	1	Tidak Bertahan
10	H011	34	60	1	0	Bertahan
11	H020	37	63	0	0	Bertahan
12	H037	39	67	0	0	Bertahan

Gambar 4.7 Form Data Uji

#### 4.4 Implementasi Uji Coba

##### 4.4.1 Skenario Uji Coba

Dalam proses pengujian, nantinya terdiri dari 5 skenario dan diuji pada *cluster* 2, 3, 4, 5, 6, 7, 8, dan 9. Serta akan dilakukan pencatatan nilai varian terkecil untuk setiap proses pengelompokan data di masing-masing *cluster*. Di mana setiap skenario pengujian menggunakan 30 data yang diambil dari data Haberman's Survival. pengujian data untuk setiap skenario menggunakan data yang berbeda. Misal untuk jumlah *cluster* 2 pengujian data skenario ke satu sebanyak 30 data, nantinya akan diuji data dengan jumlah yang sama tetapi dengan data berbeda pada skenario ke dua, begitu selanjutnya. Untuk nilai varian *cluster* nantinya dicatat pada Tabel 3.14, di mana nantinya akan dianalisis apakah nilai varian terkecil yang dihasilkan di setiap *cluster* sesuai dengan nilai akurasi tertinggi yang dihasilkan dari proses perhitungan akurasi pada proses pengujian.

Sedangkan proses pengujian data dilakukan untuk menghitung nilai bobot data yang dihitung berdasarkan proses pengujian. Dan nantinya hasil perhitungan bobot rata-rata data tersebut bisa menghasilkan kelas di mana menunjukkan data uji tergolong ke kategori 0 (data penderita yang bertahan

hidup lebih dari 5 tahun) atau kategori 1 (data penderita yang meninggal kurang dari 5 tahun). Hasil dari pengujian data dibandingkan dengan kategori asli dari penelitian apakah hasil uji sama dengan hasil penelitian pada data Haberman's Survival. Proses ini dilakukan sebanyak  $n$  data uji (pada penelitian digunakan 30 data uji) untuk setiap *cluster*nya.

Dari proses pengujian seluruh data masing-masing skenario di setiap *cluster* dicatat berapa jumlah data uji benar yang mempunyai hasil uji kelas sesuai dengan data penelitian Haberman's Survival pada Tabel 3.15 yang telah dijelaskan pada sub bab 3.3.6.2. Pencatatan ini digunakan untuk menghitung akurasi uji data dari jumlah data uji setiap skenario. Dilanjutkan dari masing-masing pengujian untuk setiap skenario dihitung rata-rata global akurasi untuk setiap seluruh skenario uji di setiap *cluster*. Dari rata-rata global ini yang dijadikan seberapa besar akurasi yang dihasilkan dari setiap aturan yang dibangkitkan di setiap *cluster*.

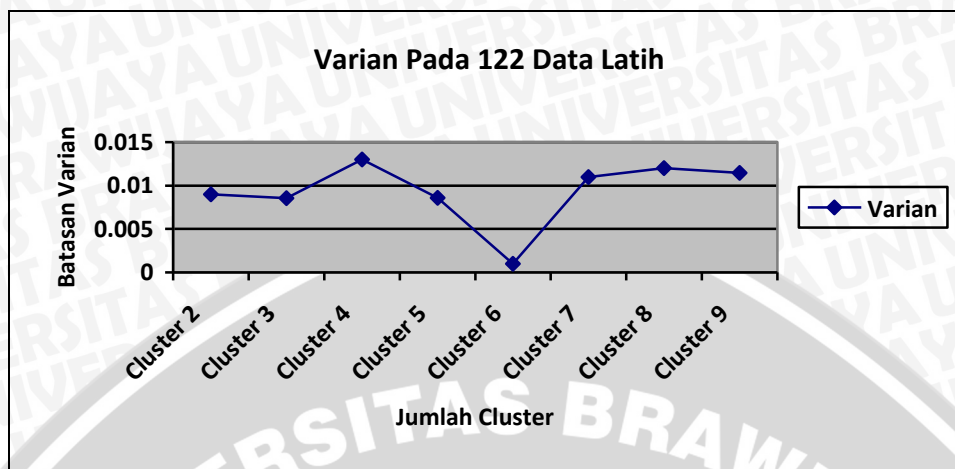
Data perhitungan varian untuk masing-masing *cluster* ditunjukkan pada Tabel 4.1.

Tabel 4.1 Hasil Perhitungan Varian Masing-Masing *Cluster*

Hasil Perhitungan Varian <i>Cluster</i> Uji								
K	2	3	4	5	6	7	8	9
V	0.009	0.00856	0.013	0.0086	0.001	0.011	0.012	0.01145

Hasil perhitungan varian pada Tabel 4.1 dihasilkan dari proses pengelompokkan data sejumlah 122 data latih. Di mana data latih tersebut dikelompokkan ke dalam beberapa *cluster* yaitu 2, 3, 4, 5, 6, 7, 8, dan 9.

Berdasarkan Tabel 4.1, maka grafik hubungan antara nilai varian dengan jumlah *cluster* dapat dibuat seperti pada Gambar 4.8.



Gambar 4.8 Grafik Nilai Varian Pada Proses Pengelompokkan Dari Tabel 4.1 dan grafik Gambar 4.8 bisa dilihat jumlah varian terkecil pada *cluster* 6 yaitu dengan  $V= 0.001$  mempunyai kandidat perhitungan akurasi terbaik. Dan tidak memungkinkan bahwa pada *cluster* lain yang memiliki jumlah varian lebih besar memiliki keakuratan pengujian data lebih bagus.

#### 4.4.2 Hasil Uji Coba

##### 4.4.2.1 Uji Coba 1

Untuk uji coba ke-1 ini, akan diuji 30 data untuk setiap skenario pengujian. Dari pengujian 30 data awal ini nantinya juga dilihat seberapa tepat perhitungan varian untuk menentukan jumlah *cluster* yang paling ideal. Jumlah *cluster* ideal ini dipilih dari perhitungan varian terkecil untuk semua *cluster* uji. Di mana pada pembahasan uji coba penelitian ini dibahas dengan menggunakan jumlah *cluster* uji yaitu 3, 5, 6, dan 7 yang mempunyai nilai varian terkecil dalam proses pengelompokkan. Hasil pengujian data uji 30 data pertama ditunjukkan pada Tabel 4.2. Data uji untuk proses uji coba 1 sampai dengan 5 dicantumkan di lampiran.

Tabel 4.2 Hasil Pengujian 30 Data Pertama

Id Data	Cluster 3		Cluster 5		Cluster 6		Cluster 7	
	Uji	Asli	Uji	Asli	Uji	Asli	Uji	Asli
H302	1	0	0	0	0	0	0	0
H153	1	0	0	0	0	0	0	0
H21	0	0	0	0	0	0	0	0
H133	0	0	0	0	0	0	0	0





H63	0	1	0	1	0	1	0	1
H39	0	0	0	0	0	0	0	0
H236	1	0	0	0	0	0	0	0
H169	0	1	0	1	0	1	0	1
H135	0	0	0	0	0	0	0	0
H178	1	0	0	0	1	0	0	0
H98	0	1	0	1	0	1	0	1
H36	0	0	0	0	0	0	0	0
H16	0	0	0	0	0	0	0	0
H119	0	0	0	0	0	0	0	0
H13	0	0	0	0	0	0	0	0
H167	1	0	0	0	0	0	0	0
H75	0	1	1	1	1	1	1	1
H74	0	1	0	1	1	1	1	1
H123	0	0	0	0	0	0	0	0
H215	1	0	0	0	0	0	0	0
H285	0	0	0	0	0	0	0	0
H82	0	1	1	1	1	1	1	1
H94	0	0	0	0	0	0	0	0
H124	0	0	0	0	0	0	0	0
H11	0	0	0	0	0	0	0	0
H105	0	0	0	0	0	0	0	0
H239	0	1	1	1	1	1	1	1
H95	0	0	0	0	0	0	0	0
H206	1	0	0	0	0	0	0	0
H41	0	0	0	0	0	0	0	0

Keterangan : Uji = kategori data hasil uji

Asli = kategori data Haberman's Survival

0 = kategori penderita bertahan hidup lebih dari 5 tahun

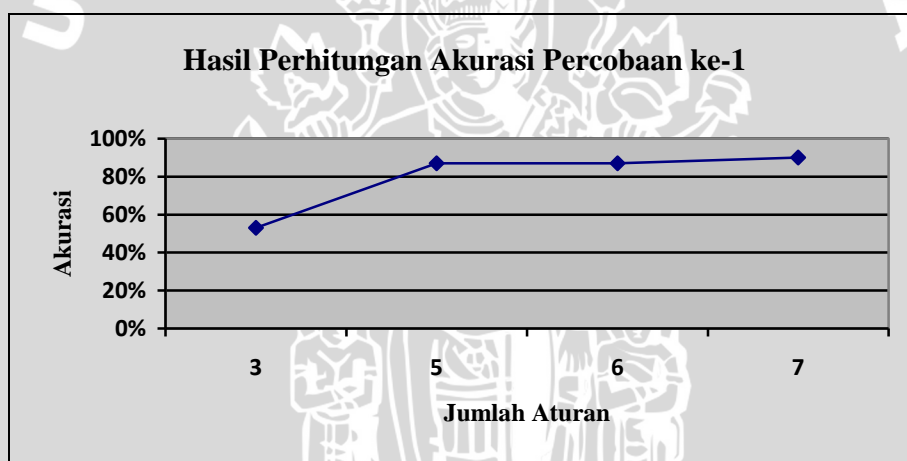
1 = kategori penderita meninggal kurang dari 5 tahun

Dari hasil data pengujian pada Tabel 4.2, dilakukan perhitungan data pengujian data yang menghasilkan kategori yang sesuai dengan data asli (data Haberman's Survival). Hasil dari perhitungan data kategori yang sesuai dengan data asli ditunjukkan pada Tabel 4.3.

Tabel 4.3 Hasil Perhitungan Kategori Percobaan ke-1

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
1	3	16	14	53%
	5	26	4	87%
	6	26	4	87%
	7	27	3	90%

Akurasi yang di dapat pada Tabel 4.3 merupakan hasil perhitungan akurasi untuk skenario ke-1 pada masing-masing *cluster*. Di mana skenario dilakukan sebanyak 5 percobaan. Grafik hasil perhitungan akurasi untuk percobaan ke-1 ditunjukkan pada Gambar 4.9.



Gambar 4.9 Grafik Nilai Perhitungan Akurasi Percobaan ke-1

Grafik pada Gambar 4.9 menunjukkan perhitungan akurasi tertinggi pada skenario percobaan ke-1 terjadi pada pembangkitan aturan 7 sebesar 90%. Perhitungan akurasi dengan 7 aturan ini lebih besar dari perhitungan akurasi dengan *cluster* ideal dengan varian terkecil pada jumlah aturan 6. Di mana hanya menghasilkan jumlah akurasi sebesar 87%.

#### 4.4.2.2 Uji Coba 2

Pada percobaan ke-2 dilakukan pengujian dengan 30 data uji seperti halnya pengujian pada skenario percobaan ke-1. Percobaan ke-2 ini 30 data uji

yang digunakan berbeda dengan data uji yang diuji pada percobaan ke-1. Data dan hasil pengujian pada percobaan ke-2 ditunjukkan pada Tabel 4.4.

Tabel 4.4 Hasil Pengujian 30 Data Kedua

Id Data	Cluster 3		Cluster 5		Cluster 6		Cluster 7	
	Uji	Asli	Uji	Asli	Uji	Asli	Uji	Asli
H142	0	0	0	0	0	0	0	0
H127	0	0	0	0	0	0	0	0
H176	1	0	0	0	0	0	0	0
H56	0	0	0	0	0	0	0	0
H150	0	0	0	0	0	0	0	0
H270	1	1	1	1	1	1	1	1
H306	0	1	1	1	1	1	1	1
H283	0	0	0	0	0	0	0	0
H186	0	0	0	0	0	0	0	0
H214	0	0	0	0	0	0	0	0
H233	0	0	0	0	0	0	0	0
H37	0	0	0	0	0	0	0	0
H24	0	0	0	0	0	0	0	0
H265	0	0	0	0	0	0	0	0
H118	0	0	0	0	0	0	0	0
H73	0	0	0	0	0	0	0	0
H272	0	0	0	0	0	0	1	0
H158	0	1	0	1	0	1	0	1
H131	0	0	0	0	0	0	0	0
H146	0	1	0	1	0	1	0	1
H102	0	0	0	0	0	0	0	0
H64	0	1	0	1	0	1	0	1
H49	0	0	0	0	0	0	0	0
H114	0	0	0	0	0	0	0	0
H3	0	0	0	0	0	0	0	0
H32	0	0	1	0	1	0	0	0
H298	1	0	0	0	0	0	0	0
H286	0	1	0	1	0	1	0	1
H226	1	0	0	0	0	0	0	0
H218	1	0	0	0	0	0	0	0

Keterangan : Uji = kategori data hasil uji

Asli = kategori data Haberman's Survival

0 = kategori penderita bertahan hidup lebih dari 5 tahun

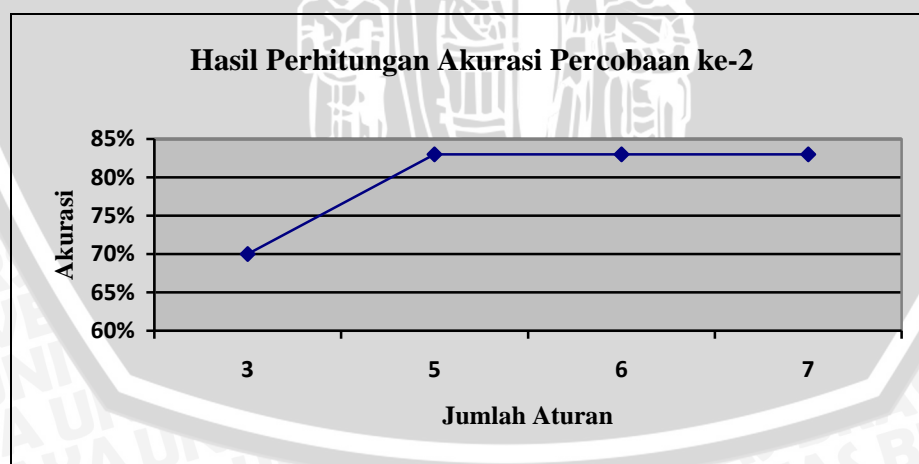
1 = kategori penderita meninggal kurang dari 5 tahun

Proses selanjutnya dilakukan proses perhitungan data kategori pada hasil data uji yang sesuai dengan kategori data asli. Untuk hasil perhitungan data kategori yang sesuai ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil Perhitungan Kategori Percobaan ke-2

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
2	3	21	9	70%
	5	25	5	83%
	6	25	5	83%
	7	25	5	83%

Dari perhitungan kategori pada Tabel 4.5 bisa dilihat untuk perhitungan akurasi pada *cluster* ke 5, 6, dan 7 mengalami penurunan akurasi di skenario ke-2. Sedangkan pada *cluster* 3 terjadi kenaikan akurasi untuk pengujian data pada skenario ke-2. Grafik perhitungan data kategori untuk percobaan ke-2 ditunjukkan pada Gambar 4.10.



Gambar 4.10 Grafik Nilai Perhitungan Akurasi Percobaan ke-2

Dilihat pada grafik Gambar 4.9 terjadi perubahan nilai akurasi tertinggi pada pengujian dengan 7 aturan sebesar 83% mengalami penurunan 7% dari percobaan pada skenario ke-1. Berbanding terbalik dengan pengujian dengan 6



H42	0	0	0	0	0	0	0	0
H182	0	1	0	1	1	1	1	1
H76	1	1	0	1	0	1	1	1
H27	0	0	0	0	0	0	0	0
H246	0	1	0	1	0	1	1	1
H263	0	0	0	0	0	0	0	0
H299	0	0	0	0	0	0	0	0
H10	0	0	0	0	0	0	0	0
H48	0	0	0	0	1	0	1	0

Keterangan : Uji = kategori data hasil uji

Asli = kategori data Haberman's Survival

0 = kategori penderita bertahan hidup lebih dari 5 tahun

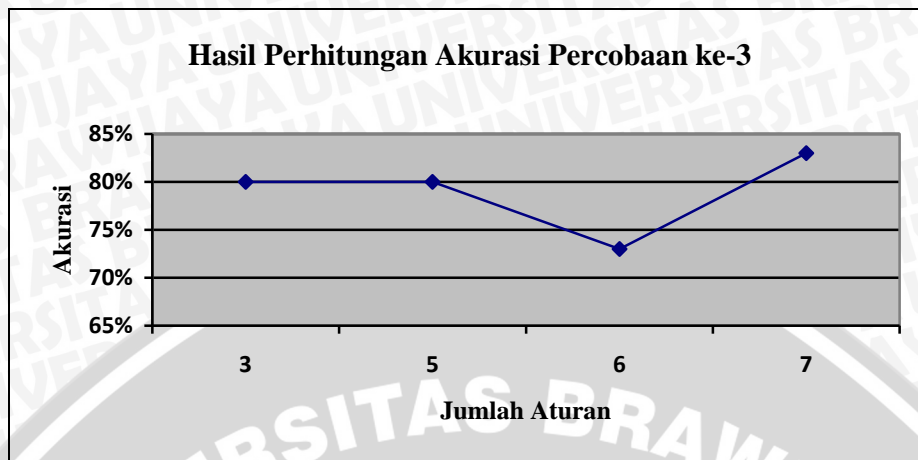
1 = kategori penderita meninggal kurang dari 5 tahun

Seperti pengujian pada skenario sebelumnya, dilakukan juga perhitungan kategori hasil uji. Hasil perhitungan kategori data uji ditunjukkan pada Tabel 4.7.

Tabel 4.7 Hasil Perhitungan Kategori Percobaan ke-3

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
3	3	24	6	80%
	5	24	6	80%
	6	22	8	73%
	7	25	5	83%

Dari Tabel 4.7 bisa dibuat grafik hasil perhitungan akurasi yang ditunjukkan pada Gambar 4.11.



Gambar 4.11 Grafik Nilai Perhitungan Akurasi Percobaan ke-3

Dari Tabel 4.7 dan grafik pada Gambar 4.11 diketahui terjadi kenaikan akurasi pengujian pada pengujian menggunakan 3 aturan sebesar 80%. Dan pengujian pada aturan 5, dan 6 mengalami penurunan. Tetapi untuk pengujian dengan menggunakan 3 aturan walaupun mengalami kenaikan hasil pengujian akurasi, namun masih kurang maksimal dilihat dari perhitungan variannya.

#### 4.4.2.4 Uji Coba 4

Pada skenario pengujian ke-4, dilakukan pengujian 30 data yang berbeda. Untuk data uji pada skenario ke-4 ditunjukkan pada Tabel 4.8.

Tabel 4.8 Hasil Pengujian 30 Data Keempat

Id Data	Cluster 3		Cluster 5		Cluster 6		Cluster 7	
	Uji	Asli	Uji	Asli	Uji	Asli	Uji	Asli
H70	0	0	0	0	0	0	0	0
H62	0	0	0	0	0	0	0	0
H109	0	1	1	1	1	1	1	1
H31	0	0	0	0	0	0	0	0
H99	0	1	0	1	0	1	0	1
H45	0	1	0	1	0	1	0	1
H106	0	0	0	0	0	0	1	0
H141	0	0	0	0	0	0	0	0
H284	0	0	0	0	0	0	0	0
H53	0	0	0	0	0	0	0	0
H2	1	0	0	0	0	0	0	0

H161	1	1	0	1	0	1	0	1
H223	0	0	0	0	0	0	0	0
H183	0	0	0	0	0	0	0	0
H277	0	0	0	0	0	0	0	0
H113	0	0	0	0	0	0	0	0
H205	1	0	0	0	0	0	0	0
H294	1	1	0	1	0	1	0	1
H260	1	1	0	1	0	1	1	1
H208	0	0	0	0	0	0	0	0
H145	0	1	0	1	0	1	0	1
H159	0	1	0	1	0	1	0	1
H120	0	0	0	0	0	0	0	0
H250	1	0	0	0	0	0	0	0
H30	0	0	0	0	0	0	0	0
H68	0	0	0	0	0	0	0	0
H152	1	0	0	0	0	0	0	0
H202	0	0	0	0	0	0	0	0
H85	0	0	0	0	0	0	0	0
H220	1	0	0	0	0	0	0	0

Keterangan : Uji = kategori data hasil uji

Asli = kategori data Haberman's Survival

0 = kategori penderita bertahan hidup lebih dari 5 tahun

1 = kategori penderita meninggal kurang dari 5 tahun

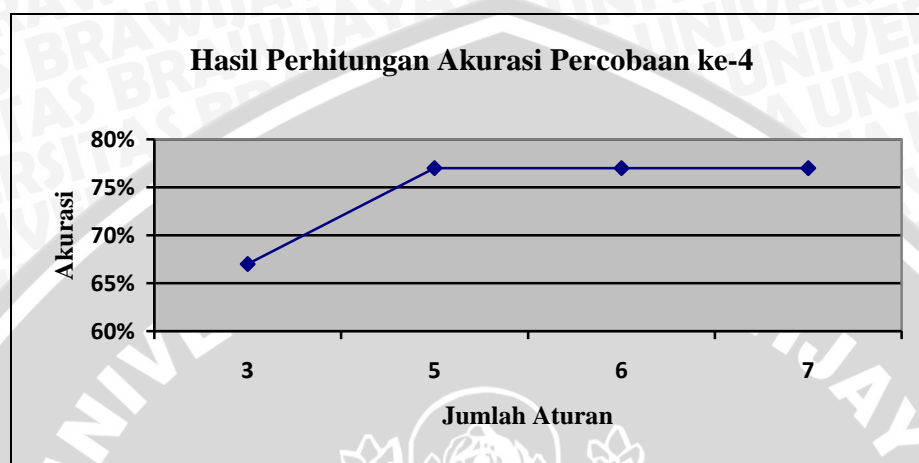
Untuk hasil proses perhitungan kategori data uji ditunjukkan pada Tabel 4.9.

Tabel 4.9 Hasil Perhitungan Kategori Percobaan ke-4

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
4	3	20	10	67%
	5	23	7	77%
	6	23	7	77%
	7	23	7	77%



Berdasarkan Tabel 4.9 bisa dibuat grafik mengenai hasil perhitungan akurasi pada skenario percobaan ke-4. Hasil perhitungan akurasi pada Tabel 4.9 mengalami penurunan. Untuk grafik hasil pengujian pada skenario ke-4 ditunjukkan pada Gambar 4.12.



Gambar 4.12 Grafik Nilai Perhitungan Akurasi Percobaan ke-4

Bisa dilihat dari hasil perhitungan akurasi pada grafik Gambar 4.12, menunjukkan bahwa perhitungan akurasi pada keempat *cluster* mengalami penurunan. Di mana pengujian dengan menggunakan aturan 5, 6, dan 7 mengalami penurunan yang sama dengan akurasi menjadi 77%. Sedangkan untuk pengujian menggunakan 6 aturan masih kurang maksimal, di mana seharusnya pada pengujian dengan menggunakan aturan 6 ini menghasilkan perhitungan akurasi yang paling maksimal.

#### 4.4.2.5 Uji Coba 5

Pada pengujian yang terakhir, dilakukan pengujian 30 data yang kelima. Seperti halnya pada pengujian di skenario 1, 2, 3, dan 4, data yang diuji untuk setiap skenario berbeda dengan skenario yang lain. Data uji yang digunakan pada pengujian skenario pada skenario kelima ditunjukkan pada Tabel 4.10.

Tabel 4.10 Hasil Pengujian 30 Data Kelima

Id Data	Cluster 3		Cluster 5		Cluster 6		Cluster 7	
	Uji	Asli	Uji	Asli	Uji	Asli	Uji	Asli
H136	0	0	0	0	0	0	0	0
H273	0	0	0	0	0	0	0	0
H211	1	0	0	0	0	0	0	0

H1	1	0	0	0	0	0	0	0
H274	1	1	1	1	1	1	1	1
H168	1	1	1	1	1	1	1	1
H78	0	0	0	0	0	0	0	0
H280	0	0	0	0	0	0	0	0
H61	0	0	0	0	0	0	0	0
H28	0	0	0	0	0	0	0	0
H165	0	0	0	0	0	0	0	0
H108	0	1	1	1	1	1	1	1
H139	0	0	0	0	1	0	0	0
H137	0	1	1	1	1	1	1	1
H242	1	0	0	0	0	0	0	0
H40	0	0	0	0	0	0	0	0
H25	0	1	0	1	0	1	0	1
H241	1	1	0	1	0	1	0	1
H259	0	1	0	1	0	1	0	1
H238	0	0	0	0	0	0	0	0
H248	1	0	0	0	0	0	0	0
H179	0	0	0	0	0	0	0	0
H44	1	1	0	1	0	1	0	1
H132	0	0	0	0	0	0	0	0
H15	1	0	0	0	0	0	0	0
H65	0	1	0	1	0	1	0	1
H219	1	0	0	0	0	0	0	0
H112	0	0	0	0	0	0	0	0
H117	0	0	0	0	0	0	0	0
H292	1	0	0	0	0	0	0	0

Keterangan : Uji = kategori data hasil uji

Asli = kategori data Haberman's Survival

0 = kategori penderita bertahan hidup lebih dari 5 tahun

1 = kategori penderita meninggal kurang dari 5 tahun

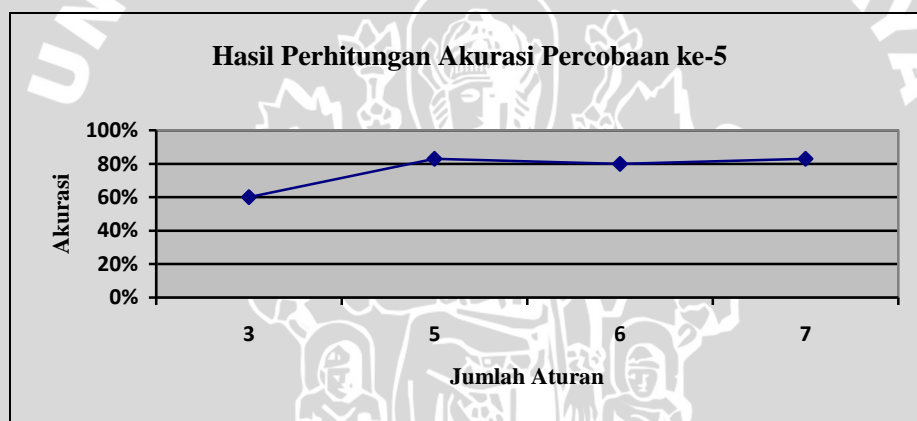
Selanjutnya seperti proses pengujian pada skenario yang lain juga dilakukan perhitungan kategori yang sesuai dengan data asli kemudian dibuat grafik

untuk hasil perhitungan kategori. Untuk data perhitungan kategori ditunjukkan pada Tabel 4.11.

Tabel 4.11 Hasil Perhitungan Kategori Percobaan ke-5

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
5	3	18	12	60%
	5	25	5	83%
	6	24	6	80%
	7	25	5	83%

Sedangkan grafik hasil perhitungan ditunjukkan pada Gambar 4.13.



Gambar 4.13 Grafik Nilai Perhitungan Akurasi Percobaan ke-5

Dari Tabel 4.11 dan grafik Gambar 4.13 bisa dilihat masih terjadi kekurangan dalam perhitungan akurasi pada percobaan dengan menggunakan 6 aturan. Terbukti dari hasil pengujian akurasi untuk masing-masing aturan yang dibangkitkan, pada pengujian dengan 7 aturan lebih bagus dari pada pengujian menggunakan 6 aturan. Di mana pada proses pengelompokkan menunjukkan bahwa seharusnya *cluster* 6 lebih memiliki hasil akurasi yang lebih maksimal jika dilihat dari potensial perhitungan variannya. Karena pada jumlah *cluster* 6 memiliki varian terkecil untuk seluruh proses pengelompokkan data.

## BAB V

### ANALISA DAN PEMBAHASAN

Pada bab ini dilakukan analisis hasil pengujian sistem, pengujian data, dan perhitungan akurasi rule yang dibangkitkan dengan menggunakan metode K-Means *clustering* dan *Fuzzy Inference System Sugeno*. Proses analisa dimulai dari proses perhitungan varian guna memilih jumlah *cluster* yang paling ideal, dan proses pengujian data yang dilakukan dalam beberapa skenario pengujian.

Seperti yang telah dilakukan proses pengujian varian dan pengujian data untuk 5 skenario di setiap *cluster* pada bab IV, akan ditunjukkan tabel hasil pengujian secara lengkap untuk seluruh varian pada Tabel 5.1 dan perhitungan kategori data uji yang ditunjukkan pada Tabel 5.2.

Tabel 5.1 Hasil Perhitungan Varian Masing-Masing *Cluster*

Hasil Perhitungan Varian <i>Cluster</i> Uji								
K	2	3	4	5	6	7	8	9
V	0.009	0.00856	0.013	0.0086	0.001	0.011	0.012	0.01145

Dari proses pengelompokkan, dihasilkan perhitungan varian seperti pada Tabel 5.1. Hasil proses perhitungan varian ini digunakan untuk mengetahui jumlah *cluster* mana yang mempunyai nilai varian terkecil. Di mana jika sebuah *cluster* mempunyai nilai varian terkecil kemungkinan besar jumlah *cluster* tersebut tepat untuk mengelompokkan data. Namun tidak memungkinkan juga adanya jumlah *cluster* lain yang memiliki jumlah nilai varian lebih besar namun untuk proses pengelompokkannya lebih bagus dari *cluster* yang memiliki nilai varian terkecil. Pada Tabel 5.1 bisa dilihat, dari delapan *cluster* yang digunakan untuk proses pengelompokkan data, *cluster* yang memiliki nilai varian paling kecil yaitu pada *cluster* 6 sebesar  $V=0.001$ . Dan *cluster* terdekat yang memiliki nilai varian terkecil dari *cluster* 6 yaitu pada *cluster* 3 dengan besar varian  $V=0.00856$ . Untuk membuktikannya maka pada bab IV dilakukan pengujian akurasi data, fungsinya untuk mengetahui seberapa akurat dari pengujian dengan menggunakan rule hasil pembangkitan aturan *fuzzy* menggunakan K-Means.

Tabel 5.2 Hasil Perhitungan Kategori Percobaan Untuk Seluruh Skenario

Percobaan ke-i	Cluster	Hasil Perhitungan Kategori		Akurasi
		$\sum$ data benar	$\sum$ data salah	
1	2	24	6	80%
	3	16	14	53%
	4	25	5	83%
	5	26	4	87%
	6	26	4	87%
	7	27	3	90%
	8	27	3	90%
	9	24	6	80%
	2	2	26	4
3		21	9	70%
4		24	6	80%
5		25	5	83%
6		25	5	83%
7		25	5	83%
8		24	6	80%
9		24	6	80%
3		2	25	5
	3	24	6	80%
	4	23	7	77%
	5	24	6	80%
	6	22	8	73%
	7	25	5	83%
	8	25	5	83%
	9	23	7	77%

4	2	23	7	77%
	3	20	10	67%
	4	23	7	77%
	5	23	7	77%
	6	23	7	77%
	7	23	7	77%
	8	23	7	77%
	9	23	7	77%
	5	2	22	8
3		18	12	60%
4		23	7	77%
5		25	5	83%
6		24	6	80%
7		25	5	83%
8		23	7	77%
9		23	7	77%

Tabel 5.3 Tabel Data Hasil Proses Perhitungan Akurasi Pengujian Data

Cluster	Percobaan Ke-i,	Hasil Perhitungan Kategori		Akurasi	Rata-Rata
		Hasil Benar	Hasil Salah		
2	1	24	6	80%	80%
	2	26	4	87%	
	3	25	5	83%	
	4	23	7	77%	
	5	22	8	73%	
3	1	16	14	53%	66%
	2	21	9	70%	
	3	24	6	80%	

	4	20	10	67%	
	5	18	12	60%	
4	1	25	5	83%	79%
	2	24	6	80%	
	3	23	7	77%	
	4	23	7	77%	
	5	23	7	77%	
	6	23	7	77%	
5	1	26	4	87%	82%
	2	25	5	83%	
	3	24	6	80%	
	4	23	7	77%	
	5	25	5	83%	
6	1	26	4	87%	80%
	2	25	5	83%	
	3	22	8	73%	
	4	23	7	77%	
	5	24	6	80%	
7	1	27	3	90%	83%
	2	25	5	83%	
	3	25	5	83%	
	4	23	7	77%	
	5	25	5	83%	
8	1	27	3	90%	81%
	2	24	6	80%	
	3	25	5	83%	
	4	23	7	77%	
	5	23	7	77%	

9	1	24	6	80%	78%
	2	24	6	80%	
	3	23	7	77%	
	4	23	7	77%	
	5	23	7	77%	

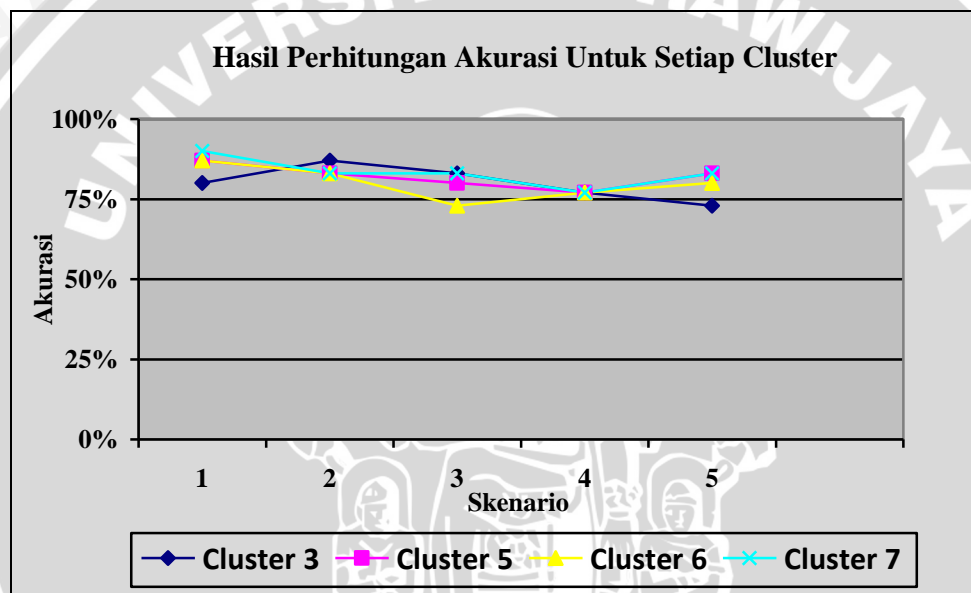
Dari pengujian data yang dilakukan pada bab IV, digunakan 30 data uji untuk setiap skenario di masing-masing aturan yang digunakan (3, 5, 6, dan 7 aturan). Pada Tabel 5.2 diketahui bahwa akurasi yang ditampilkan merupakan akurasi dari perhitungan pengujian data untuk setiap skenario di masing-masing *cluster*. Dari hasil perhitungan akurasi untuk setiap skenario tidak digunakan sebagai akurasi pengujian data di setiap *cluster*. Tetapi untuk mengetahui akurasi global pengujian data di setiap *cluster* perlu dicari rata-rata akurasi dari masing-masing *cluster*. Untuk data akurasi lengkap ditunjukkan pada Tabel 5.3.

Bisa kita lihat dari hasil perhitungan varian yang digunakan untuk menentukan *cluster* ideal dengan menggunakan sistem pada Tabel 5.1. Menunjukkan *cluster* 6 merupakan jumlah *cluster* yang memiliki varian terkecil dan berpotensi menjadi jumlah *cluster* yang paling ideal. Tetapi berbanding pada hasil proses pengujian akurasi pada Tabel 5.2. Pada percobaan ke-1 menurut perhitungan akurasi, *cluster* yang menghasilkan jumlah akurasi tertinggi terjadi pada pengujian dengan menggunakan 7 *cluster* sebesar 90% di mana mempunyai hasil akurasi lebih tinggi dari *cluster* 6 yang hanya menghasilkan nilai akurasi 87%. Pada percobaan ke-2 hasil perhitungan akurasi menunjukkan hasil paling maksimal untuk perhitungan akurasi dilakukan dengan menggunakan jumlah *cluster* 5, 6, dan 7 sebanyak 83%. Pada percobaan ke-3, perhitungan akurasi dengan menggunakan jumlah *cluster* 6 masih belum menghasilkan perhitungan akurasi maksimal. Pada percobaan ke-3 ini hasil perhitungan akurasi paling maksimal terjadi pada jumlah *cluster* 7 sebesar 83%. Pada percobaan ke-4 ada 3 *cluster* yang dapat menghasilkan perhitungan akurasi maksimal yang terjadi pada penggunaan jumlah *cluster* 5, 6, dan 7 dengan hasil akurasi sebesar 77%. Dan pada percobaan yang ke-5, bisa dilihat perhitungan akurasi terjadi pada *cluster* 5 dan 7 sebesar 83%, sedangkan *cluster* 6 menghasilkan perhitungan akurasi sebesar



80%. Dari hasil pengujian jumlah *cluster* ideal dari 5 kali pengujian bisa diketahui bahwa jumlah *cluster* 7 memiliki hasil jumlah perhitungan akurasi lebih tinggi dari pada percobaan dengan menggunakan jumlah *cluster* 6. Perbedaan hasil penentuan jumlah *cluster* ideal dari hasil pengujian sistem dengan perhitungan varian, ini membuktikan bahwa penentuan jumlah *cluster* yang paling ideal belum bisa optimal dengan menggunakan nilai varian.

Setelah proses baik perhitungan varian dan proses pengujian akurasi dilakukan, bisa dibandingkan untuk setiap akurasi global dengan menggunakan bentuk grafik yang ditunjukkan pada Gambar 5.1.



Gambar 5.1 Grafik Hasil Perhitungan Akurasi Untuk Setiap *Cluster*

Tidak hanya melihat dari perhitungan akurasi dari masing-masing skenario, diperlukan juga melihat hasil dari perhitungan akurasi global untuk keseluruhan skenario di masing-masing *cluster*. Pada Tabel 5.3 bisa dilihat untuk rata-rata akurasi global di setiap *cluster*, menunjukkan bahwa pengujian akurasi paling maksimal dilakukan dengan menggunakan *cluster* 7 dengan rata-rata akurasi globalnya berkisar 83% di setiap percobaan. 3% lebih tinggi dari jumlah *cluster* 6 yang mana merupakan jumlah *cluster* yang memiliki nilai varian terkecil dari semua *cluster* yang lain. Dan dari grafik Gambar 5.1 juga menunjukkan pada semua proses percobaan, pengujian akurasi dengan jumlah *cluster* 7 menghasilkan perhitungan paling maksimal. Dari hasil perhitungan rata-rata akurasi global dan

gambaran grafik Gambar 5.1, semakin jelas bahwa penentuan jumlah *cluster* yang paling ideal belum bisa optimal dengan menggunakan nilai varian.

Berdasarkan hasil uji coba yang dilakukan beberapa kali untuk setiap *cluster*, dapat dilihat untuk pengujian data akurasi menggunakan jumlah data uji yang sama, namun dengan data yang berbeda menghasilkan beragam hasil akurasi yang berbeda-beda dan cenderung naik turun. Ketidakstabilan hasil proses perhitungan akurasi ini, dipengaruhi dari faktor utama proses pengelompokkan. Proses utama yang sangat berpengaruh pada proses *clustering* terjadi pada penentuan awal (proses randomisasi) centroid awal. Di mana proses ini akan mempengaruhi nilai-nilai proses pembentukan aturan *fuzzy* terutama pada proses LSE yang menghasilkan fungsi di mana akan berpengaruh pada nilai-nilai konsekuennya (koefisien outputnya).

Dapat dilihat dari hasil proses pengujian akurasi, hasil yang kurang tepat ini bisa disebabkan karena dalam proses pengujian data diambil secara acak, kurang banyaknya data latih dan keragaman data latih sehingga sistem cenderung hanya mengenali pola data dari data latih yang ada. Proses hasil pengujian akurasi juga dipengaruhi karena tidak adanya *Preprocessing Data Mining*, serta diperlukan adanya optimalisasi penentuan data random harus bisa representatif. Proses ini mempunyai tujuan mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif untuk kebutuhan pemakai yang nantinya dapat berpengaruh pada hasil yang lebih akurat, pengurangan waktu komputasi, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi yang ada.

Dari beberapa kali percobaan yang dilakukan, pembangkitan aturan *fuzzy* menggunakan metode K-Means *clustering* menghasilkan akurasi di atas rata-rata. Di mana hasil proses pengujian aturan beberapa *cluster* menghasilkan akurasi minimal 66%. Jumlah hasil akurasi tersebut sudah melebihi 50% dari hasil akurasi untuk dikatakan melebihi pengujian batas minimal akurasi. Walaupun ada kekurangan dalam perhitungan varian untuk menentukan jumlah *cluster* ideal yang tidak sesuai dengan proses perhitungan akurasi yang dilakukan. Tetapi metode ini dapat menghasilkan aturan yang tepat dalam membangkitkan aturan sebagai tujuan dari penelitian.

## BAB VI PENUTUP

### 6.1 Kesimpulan

Berdasarkan hasil dari penelitian tentang implementasi metode K-Means *clustering* untuk pembangkitan aturan *fuzzy* pada klasifikasi ketahanan hidup penderita kanker payudara, dapat diambil kesimpulan :

1. Pembangkitan aturan *fuzzy* menggunakan metode K-Means *clustering* merupakan metode yang tepat. Terbukti dari hasil perhitungan akurasi pengujian data uji setiap skenario dan akurasi pengujian data uji skenario global sama-sama menghasilkan akurasi minimal di atas 50%.
2. Dari hasil perhitungan nilai varian dan hasil perhitungan akurasi. Bisa diambil kesimpulan, proses penentuan jumlah *cluster* paling ideal masih belum optimal dengan menggunakan nilai varian terkecil. Maka dari itu diperlukan adanya optimasi penentuan *cluster* ideal untuk memaksimalkan hasil pembangkitan aturan *fuzzy*.

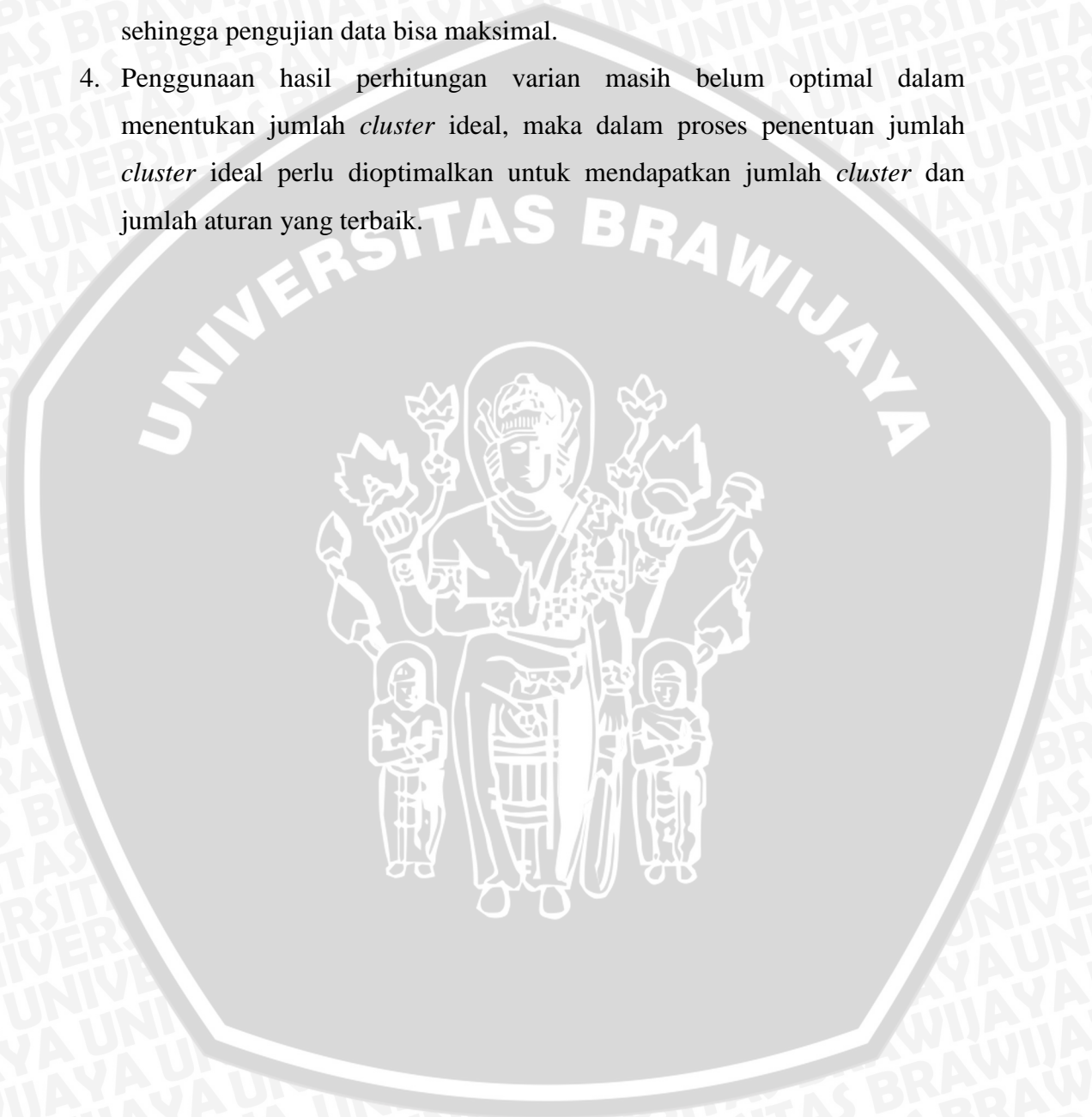
### 6.2 Saran

Saran yang bisa diberikan dari hasil penelitian yang digunakan sebagai bahan pengembangan penelitian yang terkait dengan penelitian ini, yaitu :

1. Perlu adanya pencarian data yang lebih detil terhadap atribut yang berpengaruh dalam menentukan data termasuk kategori bertahan atau tidak bertahan. Seperti tahun operasi dan node positif masih terlalu umum dan informasi atribut tersebut masih terlalu kurang akurat untuk menjadi atribut penentu kategori data. Lebih dicari referensi detilnya apakah tahun operasi ini mempunyai perlakuan yang khusus pada penderita. Sedangkan untuk node positif disini kurang diketahui apakah node positif ini hasil dari proses seperti kemoterapi, pemeriksaan, atau menunjukkan suatu hasil dari proses yang harus diberikan perlakuan khusus.
2. Keragaman atribut yang digunakan agar lebih diperbanyak lagi untuk mendapatkan hasil pengelompokkan data yang lebih tepat dan mendapatkan akurasi hasil pembangkitan aturan yang lebih baik lagi. Seperti detil node

positif lebih diperinci, apakah node tersebut merupakan hasil deteksi suatu proses kemoterapi atau hasil dari proses suatu medis.

3. Proses pelatihan data sebaiknya menggunakan data yang lebih banyak dengan tujuan mendapatkan pola data yang lebih baik dan lebih banyak sehingga pengujian data bisa maksimal.
4. Penggunaan hasil perhitungan varian masih belum optimal dalam menentukan jumlah *cluster* ideal, maka dalam proses penentuan jumlah *cluster* ideal perlu dioptimalkan untuk mendapatkan jumlah *cluster* dan jumlah aturan yang terbaik.



## DAFTAR PUSTAKA

- [ABH-95] Anderson, R.E., Black, W.C., Hair, J.F. dan Thatam, R.I. 1995, “*Multivariate Data Analysis with Readings 4<sup>th</sup> ed*”, Prentice-Hall Inc, New Jersey.
- [AND-07] Andayani, Sri. 2007, “Pembentukan Cluster dalam *Knowledge Discovery in Database* dengan Algoritma K-Means”, FMIPA UNY, Yogyakarta.
- [AGU-07] Agusta, Yudi. 2007, “K-Means – Penerapan, Permasalahan dan Metode Terkait”, *Jurnal Sistem dan Informatika*, Vol. 3, hal. 47-60.
- [AKH-10] Arapoglou, Roi., Kolomvatsos, Kostas., Hadjiefthymiades, Stathes. 2010, “*Buyer Agent Decision Process Based on Automatic Fuzzy Rules Generation Methods*”.  
[http://p-comp.di.uoa.gr/pubs/WCCI\\_f427.pdf](http://p-comp.di.uoa.gr/pubs/WCCI_f427.pdf) [1 Desember 2012].
- [BAD-05] Badriyah, Tessy. 2005, “*Clustering Analysis*”. <http://lecturer.eepis-its.edu/~iwanarif/kuliah/dm/5Clustering.pdf> [27 November 2012].
- [BUD-12] Budiman, Irwan. 2012, “Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma”, Universitas Diponegoro, Semarang.
- [BUD-02] Budiyanto, M. A. K. 2002, “Gizi dan Kesehatan”, Bayu Media dan UMM Press, Malang.
- [COR-01] Corwin, J., Elizabeth. 2001, “*Buku Saku Patofisiologi*”, Kedokteran EGC, Jakarta.
- [COX-05] Cox, Earl. 2005, “*Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*”, Morgan Kauffman Publishers, San Francisco.
- [DHS-12] Dhanasaputra, Nalendra., Santosa, Budi. 2012, “Pengembangan Algoritma *Cat Swarm Optimization* (CSO) Untuk Klasifikasi”, ITS, Surabaya.
- [DUY-06] Duyff, Roberta, L. 2006, “*Complete Food and Nutrition Guide*”, American Dietetic Association, USA.

- [FAR-08] Fariska, M. Andy. 2008, "Peramalan Multi Atribut Dengan Menggunakan *Fuzzy Clustering*".  
<http://repo.eepis-its.edu/425/1/972.pdf>. [30 November 2012].
- [FPS-96] Fayyad, U., Piatetsky, Shapiro, G., dan Smyth, P. 1996, "*From Data Mining to Knowledge Discovery in Databases*", AAAI and The MIT Press, hal. 37-53.
- [HAK-06] Han, Jiawei., and Kamber, Micheline. 2006, "Data Mining Concepts and Techniques", Morgan Kauffman Publishers, Second Edition, San Francisco.
- [HAS-09] Hasnida, Namora, Lumongga, Lubis. 2009, "Dukungan Sosial pada Pasien Kanker, Perlukah?", USU Press, Medan.
- [JAI-09] Jain, A., K. 2009, "*Data Clustering : 50 years beyond K-means*", *Pattern Recognition Letters*, 31(8), 651-666.  
<http://dx.doi.org/10.1016/j.patrec.2009.09.011> [25 November 2012].
- [JSM-97] Jang, J.S.R., Sun, C.T., Mizutani, E. 1997, "*Neuro-Fuzzy and Soft Computing*", Prentice-Hall International, New Jersey.
- [KAR-08] Karhendana, A. 2008, "Pemanfaatan *Document Clustering* pada Agregator Berita", ITB, Bandung.
- [KUP-10] Kusumadewi, Sri., Purnomo, Hari. 2010, "Aplikasi Logika Fuzzy untuk Pendukung Keputusan", Graha Ilmu, Yogyakarta.
- [KUS-07] Kusnawi. 2007, "Pengantar Solusi Data Mining", Seminar Nasional Teknologi, Yogyakarta.
- [KON-08] Konginan, Agustina. 2008, "Depresi Pada Penderita Kanker", RSU Dr. Soetomo, Surabaya.
- [KOS-82] Koestedjo, R., dan Soemartono, H.R. 1982, "Diagnosa Dini Penyakit Kanker Dan Cara Menanggulangnya", Kantor Wilayah Departemen Kesehatan Propinsi Jawa Barat, Bandung.
- [LAR-06] Larose D, T., 2006, "*Data Mining Methods and Models*", Jhon Wiley & Sons, Inc, Hoboken New Jersey.
- [LIM-99] Lim, Tjen, Sien. 1999, "*Haberman's Survival Data*", University of Chicago's Billings Hospital, USA.

- [MAH-12] Mahleda, Maulia, I. P., Hartini, Nurul. 2012, “*Post-traumatic Growth* pada Pasien Kanker Payudara Pasca Mastektomi Usia Dewasa Madya”, *Jurnal Psikologi Klinis dan Kesehatan Mental*, Vol. 1, No. 02, hal. 67-70.
- [MAN-09] Mangan, Y. 2009, “Solusi Sehat Mencegah dan Mengatasi Kanker”, PT. Agromedia Pustaka, Jakarta.
- [MAR-04] Mardiana, Lina. 2004, “Kanker pada Wanita”, Penebar Swadaya, Jakarta.
- [MEN-00] Medelsohn, J. 2000, “*Prinsip Neoplasma. Di dalam: Horrison Prinsip – prinsip Ilmu Penyakit Dalam Volume 4*”, Kedokteran EGC, Jakarta.
- [MOR-07] Moreno, Juan, E., et al. 2007, “*Data Mining for extraction of fuzzy IF-THEN rules using Mamdani and Takagi-Sugeno-Kang FIS*”.  
[http://www.engineeringletters.com/issues\\_v15/issue\\_1/EL\\_15\\_1\\_13.pdf](http://www.engineeringletters.com/issues_v15/issue_1/EL_15_1_13.pdf) [5 Desember 2012].
- [MRA-10] Martiana, E., Rosyid, N., Agusetia, U. 2010, “Mesin Pencari Dokumen dengan Pengklasteran Secara Otomatis”, *TELKOMNIKA* 8 (1), hal. 41-48.
- [NAN-09] Nani, Desiyani. 2009, “Analisis Faktor-Faktor Yang Berhubungan Dengan Kejadian Kanker Payudara di Rumah Sakit Pertamina Cilacap”, *Jurnal Keperawatan Soedirman (The Soedirman Journal of Nursing)*, Vol. 4, No.2, hal. 61-66.
- [NAP-08] Nasca., Philip, C. 2008, “*Fundamentals Of Cancer Epidemiology*”, Jones and Bartlett Publishers, Canada.
- [NOH-09] Noor, M. Helmy., Hariadi, Moch. 2009, “*Image Cluster Berdasarkan Warna Untuk Identifikasi Kematangan Buah Tomat Dengan Metode Valley Tracing*”, Institut Teknologi Sepuluh November, Surabaya.
- [PRE-94] Prett., William, B., et al. 1994, “*The Cancer Problem*”, *The Anticancer Drugs Second Edition*, Oxford University Press, New York, hal 3-16.
- [PRI-07] Priyono, Agus., dkk. 2007, “*Generation Of Fuzzy Rules With Subtractive Clustering*”.  
<http://eprints.utm.my/1451/1/JTDIS43D%5B10%5Dnew.pdf> [30 November 2012].

- [RAM-06] Ramdhany, Dhany, Nugraha., dkk. 2006, “Diagnosis Gangguan Sistem Urinari Pada Anjing dan Kucing Menggunakan VFI 5”, Jurnal Ilmu Komputer dan Informasi, Vol. 2, No. 2, ISSN 1979-0732, hal. 86-94.
- [RUP-02] Ramli, H. Muchlis., Umbas, Rainy., dan Panigoro, Sonar, S. 2002, “Deteksi Dini Kanker”, Balai Penerbit FKUI, Jakarta.
- [SAM-06] Sambamoorthi, N. 2010, “*Hierarchical Cluster Analysis: Some Basics and Algorithms*”, CRMportals Inc, Englishtown.
- [SAN-07] Santoso, B. 2007, “Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis”, Graha Ilmu, Yogyakarta.
- [TIO-09] Tiolena, Ristarolas. 2009, “Faktor-faktor yang Mempengaruhi Keterlambatan Pengobatan Pada Wanita Penderita Kanker Payudara RSUP H. Adam Malik Medan Tahun 2008”, USU Repository, Universitas Sumatera Utara, Medan.
- [TUR-05] Turban, et al. 2005, “*Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas)* Edisi Ketujuh, Andi, Yogyakarta.
- [VEL-10] Velmurugan, T. 2010, “*Performance Evaluation of K-means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points*” European Journal of Scientific Research ISSN 1450-216X, Vol. 46, NO. 3, hal. 320-330.
- [VOR-80] Vorherr, Helmuth. 1980, “*Breast Cancer, Epidemiology, Endocrinology, Biochemistry, and Pathobiology*”, Urban & Scharzenberger, Baltimore Munich.
- [WAH-06] Wahyuni, Arlinda, Sari. 2006, “Hubungan Jenis Histologi dengan Ketahanan Hidup 5 Tahun Penderita Kanker Payudara”, Majalah Kedokteran Nusantara, Vol. 39, No. 1, hal. 7-11.
- [WID-07] Widhiastiwi, Yuni. 2007, “Model Fuzzy”, Bina Widya, Vol.18, No. 02, hal. 88-94.



## Lampiran ke-1

## Data Haberman's Survival

Umur	Tahun Operasi	Edge Positif	Kategori Ketahanan	Kategori Penelitian
30	64	1	1	0
30	62	3	1	0
30	65	0	1	0
31	59	2	1	0
31	65	4	1	0
33	58	10	1	0
33	60	0	1	0
34	59	0	2	1
34	66	9	2	1
34	58	30	1	0
34	60	1	1	0
34	61	10	1	0
34	67	7	1	0
34	60	0	1	0
35	64	13	1	0
35	63	0	1	0
36	60	1	1	0
36	69	0	1	0
37	60	0	1	0
37	63	0	1	0
37	58	0	1	0
37	59	6	1	0
37	60	15	1	0
37	63	0	1	0
38	69	21	2	1
38	59	2	1	0
38	60	0	1	0
38	60	0	1	0
38	62	3	1	0
38	64	1	1	0
38	66	0	1	0
38	66	11	1	0
38	60	1	1	0
38	67	5	1	0
39	66	0	2	1
39	63	0	1	0
39	67	0	1	0

39	58	0	1	0
39	59	2	1	0
39	63	4	1	0
40	58	2	1	0
40	58	0	1	0
40	65	0	1	0
41	60	23	2	1
41	64	0	2	1
41	67	0	2	1
41	58	0	1	0
41	59	8	1	0
41	59	0	1	0
41	64	0	1	0
41	69	8	1	0
41	65	0	1	0
41	65	0	1	0
42	69	1	2	1
42	59	0	2	1
42	58	0	1	0
42	60	1	1	0
42	59	2	1	0
42	61	4	1	0
42	62	20	1	0
42	65	0	1	0
42	63	1	1	0
43	58	52	2	1
43	59	2	2	1
43	64	0	2	1
43	64	0	2	1
43	63	14	1	0
43	64	2	1	0
43	64	3	1	0
43	60	0	1	0
43	63	2	1	0
43	65	0	1	0
43	66	4	1	0
44	64	6	2	1
44	58	9	2	1
44	63	19	2	1
44	61	0	1	0
44	63	1	1	0
44	61	0	1	0

44	67	16	1	0
45	65	6	2	1
45	66	0	2	1
45	67	1	2	1
45	60	0	1	0
45	67	0	1	0
45	59	14	1	0
45	64	0	1	0
45	68	0	1	0
45	67	1	1	0
46	58	2	2	1
46	69	3	2	1
46	62	5	2	1
46	65	20	2	1
46	62	0	1	0
46	58	3	1	0
46	63	0	1	0
47	63	23	2	1
47	62	0	2	1
47	65	0	2	1
47	61	0	1	0
47	63	6	1	0
47	66	0	1	0
47	67	0	1	0
47	58	3	1	0
47	60	4	1	0
47	68	4	1	0
47	66	12	1	0
48	58	11	2	1
48	58	11	2	1
48	67	7	2	1
48	61	8	1	0
48	62	2	1	0
48	64	0	1	0
48	66	0	1	0
49	63	0	2	1
49	64	10	2	1
49	61	1	1	0
49	62	0	1	0
49	66	0	1	0
49	60	1	1	0
49	62	1	1	0

49	63	3	1	0
49	61	0	1	0
49	67	1	1	0
50	63	13	2	1
50	64	0	2	1
50	59	0	1	0
50	61	6	1	0
50	61	0	1	0
50	63	1	1	0
50	58	1	1	0
50	59	2	1	0
50	61	0	1	0
50	64	0	1	0
50	65	4	1	0
50	66	1	1	0
51	59	13	2	1
51	59	3	2	1
51	64	7	1	0
51	59	1	1	0
51	65	0	1	0
51	66	1	1	0
52	69	3	2	1
52	59	2	2	1
52	62	3	2	1
52	66	4	2	1
52	61	0	1	0
52	63	4	1	0
52	69	0	1	0
52	60	4	1	0
52	60	5	1	0
52	62	0	1	0
52	62	1	1	0
52	64	0	1	0
52	65	0	1	0
52	68	0	1	0
53	58	4	2	1
53	65	1	2	1
53	59	3	2	1
53	60	9	2	1
53	63	24	2	1
53	65	12	2	1
53	58	1	1	0

53	60	1	1	0
53	60	2	1	0
53	61	1	1	0
53	63	0	1	0
54	60	11	2	1
54	65	23	2	1
54	65	5	2	1
54	68	7	2	1
54	59	7	1	0
54	60	3	1	0
54	66	0	1	0
54	67	46	1	0
54	62	0	1	0
54	69	7	1	0
54	63	19	1	0
54	58	1	1	0
54	62	0	1	0
55	63	6	2	1
55	68	15	2	1
55	58	1	1	0
55	58	0	1	0
55	58	1	1	0
55	66	18	1	0
55	66	0	1	0
55	69	3	1	0
55	69	22	1	0
55	67	1	1	0
56	65	9	2	1
56	66	3	2	1
56	60	0	1	0
56	66	2	1	0
56	66	1	1	0
56	67	0	1	0
56	60	0	1	0
57	61	5	2	1
57	62	14	2	1
57	64	1	2	1
57	64	9	1	0
57	69	0	1	0
57	61	0	1	0
57	62	0	1	0
57	63	0	1	0

57	64	0	1	0
57	64	0	1	0
57	67	0	1	0
58	59	0	1	0
58	60	3	1	0
58	61	1	1	0
58	67	0	1	0
58	58	0	1	0
58	58	3	1	0
58	61	2	1	0
59	62	35	2	1
59	60	0	1	0
59	63	0	1	0
59	64	1	1	0
59	64	4	1	0
59	64	0	1	0
59	64	7	1	0
59	67	3	1	0
60	59	17	2	1
60	65	0	2	1
60	61	1	1	0
60	67	2	1	0
60	61	25	1	0
60	64	0	1	0
61	62	5	2	1
61	65	0	2	1
61	68	1	2	1
61	59	0	1	0
61	59	0	1	0
61	64	0	1	0
61	65	8	1	0
61	68	0	1	0
61	59	0	1	0
62	59	13	2	1
62	58	0	2	1
62	65	19	2	1
62	62	6	1	0
62	66	0	1	0
62	66	0	1	0
62	58	0	1	0
63	60	1	2	1
63	61	0	1	0

63	62	0	1	0
63	63	0	1	0
63	63	0	1	0
63	66	0	1	0
63	61	9	1	0
63	61	28	1	0
64	58	0	1	0
64	65	22	1	0
64	66	0	1	0
64	61	0	1	0
64	68	0	1	0
65	58	0	2	1
65	61	2	2	1
65	62	22	2	1
65	66	15	2	1
65	58	0	1	0
65	64	0	1	0
65	67	0	1	0
65	59	2	1	0
65	64	0	1	0
65	67	1	1	0
66	58	0	2	1
66	61	13	2	1
66	58	0	1	0
66	58	1	1	0
66	68	0	1	0
67	64	8	2	1
67	63	1	2	1
67	66	0	1	0
67	66	0	1	0
67	61	0	1	0
67	65	0	1	0
68	67	0	1	0
68	68	0	1	0
69	67	8	2	1
69	60	0	1	0
69	65	0	1	0
69	66	0	1	0
70	58	0	2	1
70	58	4	2	1
70	66	14	1	0
70	67	0	1	0

70	68	0	1	0
70	59	8	1	0
70	63	0	1	0
71	68	2	1	0
72	63	0	2	1
72	58	0	1	0
72	64	0	1	0
72	67	3	1	0
73	62	0	1	0
73	68	0	1	0
74	65	3	2	1
74	63	0	1	0
75	62	1	1	0
76	67	0	1	0
77	65	3	1	0
78	65	1	2	1
83	58	2	2	1

**Keterangan :**

**Umur** : Umur pasien penderita kanker payudara

**Tahun Operasi** : Tahun pasien ketika menjalani operasi kanker payudara

**Node Positif** : Jumlah node yang positif terdeteksi kanker payudara

**Kategori Ketahanan** : Kelas kategori ketahanan hidup penderita kanker payudara (1 : bertahan hidup lebih dari 5 tahun, 2 : meninggal kurun waktu kurang dari 5 tahun).

**Kategori Penelitian** : Kelas kategori ketahan hidup yang digunakan untuk penelitian skripsi. (0 : bertahan hidup lebih dari 5 tahun, 1 : meninggal kurun waktu kurang dari 5 tahun).



## IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK PEMBANGKITAN ATURAN FUZZY PADA KLASIFIKASI KETAHANAN HIDUP PENDERITA KANKER PAYUDARA

**Khoirul Sholeh, Budi Darma Setiawan, Imam Cholissodin**

Program Studi Ilmu Komputer, Jurusan Ilmu Komputer  
Fakultas Program Teknologi Informasi dan Ilmu Komputer  
Jl. Veteran No 8, Malang 65145, Indonesia  
E-mail : [khoirul.sholeh@gmail.com](mailto:khoirul.sholeh@gmail.com)

### ABSTRAK

Untuk membentuk aturan fuzzy diperlukan seorang pakar. Pakar adalah seorang ahli yang berpengalaman dalam suatu bidang yang mampu menjelaskan suatu aturan yang terkait dengan suatu bidang. Dalam penelitian ini dibentuk suatu rule secara otomatis yang tidak bergantung dengan seorang pakar. Aturan fuzzy yang dibangkitkan bisa diperoleh dari beberapa teknik seperti proses clustering. Metode yang digunakan dalam membangkitkan aturan fuzzy ini yaitu metode K-Means Clustering. Kelebihan metode ini mempunyai algoritma yang sederhana dan efisien, di mana algoritma akan menghasilkan pengelompokan data dengan kondisi yang optimum. K-Means clustering digunakan untuk mengelompokkan data dan membangkitkan aturan pada kasus penderita kanker payudara. Hasil dari pembangkitan aturan fuzzy digunakan untuk proses inferensi fuzzy menggunakan metode Fuzzy Inference System Sugeno. Metode Sugeno menghasilkan output (konsekuen) berupa konstanta atau persamaan linier. Dalam penelitian ini, digunakan 122 data training dan dilakukan proses pengujian sebanyak 30 data dalam lima skenario percobaan untuk setiap jumlah cluster. Hasil pengujian digunakan untuk mengetahui akurasi aturan yang terbentuk. Dari hasil penelitian menunjukkan bahwa hasil akurasi tertinggi dengan menggunakan pembangkitan aturan sejumlah lima aturan yaitu sebesar 83% dari rata-rata pengujian seluruh skenario uji masing-masing cluster.

**Kata kunci :** penderita kanker payudara, aturan fuzzy, fuzzy inference system sugeno, K-Means, clustering

### 1. PENDAHULUAN

#### 1.1 Latar Belakang

Kanker merupakan jenis penyakit yang paling membunuh di dunia. Salah satu jenis penyakit kanker yang berbahaya yaitu kanker payudara. Kanker ini merupakan momok bagi wanita-wanita diseluruh dunia [16].

Dengan adanya banyak kasus yang terjadi dan semakin meningkatnya kasus mengenai kanker payudara. Diperkirakan dalam dua puluh tahun terakhir ini, telah terjadi kemajuan yang sangat menyolok dalam penemuan dini, pengobatan dan penelitian yang berhasil pada beberapa tipe keganasan yang di masa lampau selalu dianggap fatal [14]. Salah satunya yaitu penelitian ketahanan hidup penderita kanker payudara yang dilakukan oleh Tjen-Sien Lim yang disebut dengan data Haberman's Survival.

Merupakan data hasil dari penelitian mengenai ketahanan pasien yang telah menjalani operasi kanker payudara di University of Chicago's Billings Hospital.

Ketahanan hidup seorang penderita kanker payudara pasti berbeda antara penderita yang satu dengan penderita yang lain. Masalah perbedaan ini tentunya ada beberapa faktor yang mempengaruhi daya tahan penderita kanker payudara tersebut. Faktor-faktor tersebut bisa dijadikan sebagai karakteristik penderita kanker payudara. Cara paling mudah dalam mengetahui karakteristik penderita kanker payudara ini yaitu dengan mengelompokkan data penderita berdasarkan klasifikasi atau kelompok tertentu. Salah satu metode *clustering* ini yaitu metode *K-Means*. Metode ini digunakan untuk mengolah data untuk membentuk aturan-aturan yang diguna-

kan untuk menguji data daya tahan hidup penderita kanker payudara apakah tergolong dari kategori (bertahan hidup lebih dari lima tahun atau tidak) dengan menggunakan metode *Fuzzy Inference Sugeno*.

### 1.2 Rumusan Masalah

Dari uraian latar belakang yang diuraikan, dapat dibuat suatu rumusan masalah yaitu :

- 1) Bagaimana cara mengimplementasikan metode *K-Means Clustering* untuk membangkitkan aturan *fuzzy* pada proses *clustering* data ketahanan hidup penderi-ta kanker payudara
- 2) Bagaimana perhitungan akurasi hasil pembangkitan aturan *fuzzy* menggunakan metode *K-Means Clustering* untuk pengujian data ketahanan hidup penderita kanker payudara menggunakan *Fuzzy Inferensi Sistem Sugeno* pada data *Haberman's Survival*.

### 1.3 Batasan Masalah

Masalah yang dibahas pada penelitian ini dibatasi pada :

- 1) Data yang dijadikan data latih dan data uji pada skripsi ini merupakan dataset *Haberman's Survival* yang merupakan hasil penelitian di *University Chicago's Billings Hospital* oleh *Tjen-Sien Lim* mengenai *Ketahanan Hidup Penderita Kanker Payudara*, diambil dari :  
<http://archive.ics.uci.edu/ml/machine-learning-databases/haberman/haberman.data>
- 2) Jika nantinya ingin menguji dengan data kasus yang berbeda, maka faktor-faktor lain (internal maupun eksternal) yang mempengaruhi kondisi penderita di-abaikan.

### 1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian skripsi ini adalah :

- 1) Menerapkan algoritma *K-Means Clustering* untuk pembangkitan aturan *fuzzy* pada proses *clustering* data ketahanan hidup penderita kanker payudara berdasarkan perhitungan

kemiripan data pada data *Haberman's Survival*.

- 2) Untuk mengetahui hasil akurasi uji data dan klasifikasi yang dilakukan dengan menggunakan metode *Fuzzy Inferensi Sistem Sugeno* dengan pembangkitan aturan *fuzzy* menggunakan metode *K-Means Clustering* terhadap kasus ketahanan hidup penderita kanker payudara menggunakan data *Haberman's Survival*.

### 1.5 Manfaat Penelitian

Manfaat yang dapat diambil dalam penelitian ini yaitu menghasilkan aplikasi yang digunakan untuk mengklasifikasi ketahanan hidup penderita kanker payudara, serta dapat mengimplementasikan metode *K-Means clustering* sebagai alat untuk membangkitkan aturan secara otomatis sebagai ganti rule yang didefinisikan oleh seorang pakar.

## 2. Tinjauan Pustaka

### 2.1 Kanker

Kanker adalah suatu penyakit, dimana terdapat pertumbuhan sel dan jaringan yang tidak normal, cepat, dan tidak terkendali. Bila pertumbuhan itu tidak segera dihentikan, maka ia akan berkembang terus sehingga akan mengganggu dan menghancurkan jaringan dan organ yang terkena, serta dapat menyebar ke bagian tubuh yang lainnya, yang akhirnya dapat menyebabkan kematian penderitanya [2].

### 2.2 Kanker Payudara

Kanker payudara adalah tumor ganas yang menyerang jaringan payudara. Jaringan payudara tersebut terdiri dari kelenjar susu (kelenjar pembuat air susu), saluran kelenjar (saluran air susu), dan jaringan penunjang payudara [8].

Kanker payudara merupakan penyakit yang paling ditakuti oleh kaum wanita. Meskipun demikian, berdasarkan penemuan terakhir, kaum pria pun bisa terkena kanker payudara. Dari hasil penelitian para ahli menunjukkan bahwa satu dari delapan wanita terkena kanker payudara. Kanker payudara umumnya menyerang wanita yang telah berumur lebih dari 40 tahun. Namun demikian,

wanita muda pun bisa terserang kanker ini [8].

Beberapa faktor risiko pada kanker payudara yang sudah diterima secara luas oleh kalangan “*oncologist*” di dunia adalah sebagai berikut [14]:

1. Umur lebih dari 30 tahun mempunyai kemungkinan yang lebih besar untuk mendapat kanker payudara dan risiko ini akan bertambah sampai umur 50 tahun dan setelah menopause.
2. Tidak kawin/Nulipara risikonya 2-4 kali lebih tinggi daripada wanita yang kawin dan punya anak.
3. Anak pertama lahir setelah 35 tahun risikonya 2 kali lebih besar.
4. “*Menarche*” kurang dari 12 tahun risikonya 1,7-3,4 kali lebih tinggi daripada wanita dengan “*menarche*” yang datang pada usia normal atau lebih dari 12 tahun.
5. Menopause datang terlambat lebih dari 55 tahun, risikonya 2,5-5 kali lebih tinggi.
6. Pernah mengalami infeksi, trauma atau operasi tumor jinak payudara, risikonya 3-9 kali lebih besar.
7. Adanya kanker pada payudara kontralateral, risikonya 3-9 kali lebih besar.
8. Pernah mengalami operasi ginekologis tumor ovarium, risikonya 3-4 kali lebih tinggi.
9. Yang mengalami radiasi di dinding dada risikonya 2-3 kali lebih tinggi.
10. Riwayat keluarga ada yang menderita kanker payudara pada ibu, saudara perempuan ibu, saudara perempuan, adik/kakak, risikonya 2-3 kali lebih tinggi.
11. Kontrasepsi oral pada penderita tumor payudara jinak seperti kelainan fibrokistik yang ganas akan meningkatkan risiko untuk mendapatkan kanker payudara 11 kali lebih tinggi.

### 2.3 Ketahanan Hidup Penderita Kanker Payudara

Ketahanan hidup (*survival rate*) seorang penderita kanker payudara bisa dipengaruhi dari adanya program diagnosis dini pada penderita kanker payudara. Tujuan dari suatu program

seperti ini bukan saja hanya memperbaiki angka ketahanan hidup (*survival rate*), tetapi juga perbaikan angka penyembuhan (*free of disease*) [14]. Tolak ukur keberhasilan pengobatan penyakit kanker lazim digambarkan dengan angka ketahanan hidup (*year survival rate*) [11].

### 2.4 Data Haberman’s Survival

Data Haberman’s Survival merupakan data hasil dari penelitian mengenai ketahanan pasien yang telah menjalani operasi kanker payudara di University of Chicago’s Billings Hospital yang dilakukan pada tahun 1958-1970. Data ini disumbangkan oleh Tjen-Sien Lim. Data *Haberman’s Survival* digunakan untuk memprediksi apakah seorang pasien yang telah menjalani operasi kanker payudara akan bertahan hidup atau tidak [3]. Masing-masing data terdiri dari empat atribut yaitu : umur pasien ketika menjalani operasi, tahun pasien menjalani operasi, jumlah node positif yang terdeteksi, dan status ketahanan hidup. Dan terdiri dari 306 kasus [7].

### 2.5 Data Mining

#### 2.5.1 Definisi Data Mining

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan dalam database besar [17]. Secara sederhana, data mining dapat diartikan sebagai proses atau “menggali” knowledge yang ada pada sekumpulan (sejumlah) data [5].

#### 2.5.2 K-Means Clustering

Algoritma K-means adalah algoritma yang terbaik dalam algoritma *partitional clustering* dan yang paling sering digunakan diantara algoritma *clustering* lainnya karena kesederhanaan dan efisiensinya . Algoritma *K-Means* dinilai cukup efisien, yang ditunjukkan dengan kompleksitasnya  $O(nKT)$ , dengan catatan  $n$  adalah banyaknya obyek data,  $k$  adalah jumlah cluster yang dibentuk, dan  $t$  banyaknya iterasi. Biasanya, nilai  $k$  dan  $t$  jauh lebih kecil daripada nilai  $n$ . selain itu, dalam

iterasinya, algoritma ini akan berhenti dalam kondisi optimum [1].

### 2.5.3 Algoritma K-Means Clustering

Algoritma clustering K-means dapat diringkas sebagai berikut [15]:

1. Pilih jumlah cluster
  2. Inisialisasi k pusat cluster (diberi nilai-nilai random)
  3. Tempatkan setiap data/obyek ke cluster terdekat. Kedekatan dua obyek ditentukan berdasarkan jarak kedua obyek tersebut. Jarak paling dekat antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Dalam proses perhitungan jarak, digunakan rumus Euclidean seperti persamaan 2.1.
- $$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$
4. Hitung kembali pusat cluster dengan anggota cluster yang sekarang. Pusat cluster adalah rata-rata semua data/obyek dalam cluster
  5. Tugaskan lagi setiap obyek memakai pusat cluster yang baru. Jika pusat Cluster sudah tidak berubah lagi, maka proses pengclustering selesai.
  6. Kembali ke langkah 3 sampai pusat cluster tidak berubah lagi.

### 2.5.4 Analisa Cluster

Sebelum proses analisa cluster, terlebih dahulu dilakukan perhitungan mean dan standar deviasi untuk data yang sudah dikelompokkan dengan menggunakan persamaan 2.2 dan persamaan 2.3.

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2.2)$$

$$\sigma_{ki} = \sqrt{\frac{\sum_{j=1}^n (x_j - \mu_{ki})^2}{n - 1}} \quad (2.3)$$

Dimana : k = jumlah cluster 1...k  
 i = atribut 1...i  
 n = jumlah data  
 x<sub>j</sub> = data ke-j  
 μ = rata-rata cluster

Analisa cluster bisa diperoleh dari kepadatan cluster yang dibentuk (*cluster density*). Kepadatan suatu cluster dapat

ditentukan dengan *variance within cluster* (V<sub>w</sub>) dan *variance between cluster* (V<sub>b</sub>). Varian tiap tahap pembentukan cluster dihitung dengan persamaan [9]:

$$V_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^n (y_i - \bar{y}_c)^2 \quad (2.4)$$

dengan : V<sub>c</sub><sup>2</sup> = varian pada cluster c  
 c = 1..k, dimana k = jumlah cluster  
 n<sub>c</sub> = jumlah data pada cluster c  
 y<sub>i</sub> = data ke-i pada suatu cluster  
 $\bar{y}_c$  = rata-rata dari data pada suatu cluster

Selanjutnya dari nilai varian diatas, kita bisa menghitung nilai *variance within cluster* (V<sub>w</sub>) dengan rumus :

$$V_w = \frac{1}{N - c} \sum_{i=1}^c (n_i - 1)v_i^2 \quad (2.5)$$

dimana : N = jumlah semua data  
 c = 1..k, dimana k = jumlah cluster  
 n<sub>i</sub> = jumlah data pada cluster  
 v<sub>i</sub> = varian pada cluster

Dan nilai *variance between cluster* (V<sub>b</sub>) dengan rumus:

$$V_b = \frac{1}{c - 1} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 \quad (2.6)$$

dimana  $\bar{y}$  = rata-rata dari  $\bar{y}_i$

Salah satu metode yang digunakan untuk menentukan cluster yang ideal adalah nilai varian, yaitu dengan menghitung kepadatan cluster berupa *variance within cluster* (V<sub>w</sub>) dan *variance between cluster* (V<sub>b</sub>) dengan menggunakan persamaan 2.7.

$$V = \frac{V_w}{V_b} \quad (2.7)$$

### 2.6 Logika Fuzzy

Logika fuzzy pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dasar logika fuzzy adalah teori himpunan fuzzy. Pada teori himpunan fuzzy, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Nilai keanggotaan atau derajat keanggotaan atau *membership function* menjadi ciri utama dari penalaran dengan logika fuzzy tersebut [6].



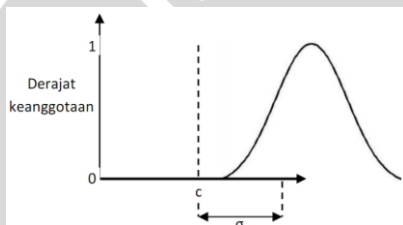
**2.6.1 Fungsi Keanggotaan**

Fungsi keanggotaan (*membership function*) adalah suatu kurva yang menunjukkan pemetaan titik-titik input data kedalam nilai keanggotaannya (sering juga disebut dengan derajat keanggotaan) yang memiliki interval antara 0 sampai 1. Salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi [6].

**2.6.2 Fungsi Keanggotaan Gauss**

Fungsi keanggotaan Gaussian ditentukan dengan 2 parameter {*c*,  $\sigma$ } dengan mengikuti persamaan :

$$gaussian(x; c; \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (2.8)$$



Gambar 2.1 Kurva Gaussian

Fungsi keanggotaan gaussian ditentukan oleh *c* dan  $\sigma$ . *c* merepresentasikan titik tengah (center) dan  $\sigma$  merepresentasikan lebar dari fungsi keanggotaan. Fungsi keanggotaan Gaussian dapat diwujudkan seperti pada gambar 2.1.

**2.7 Fuzzy Inference System Sugeno**

Sistem inferensi fuzzy merupakan proses pengolahan data dalam bentuk crisp input yang melalui beberapa tahapan dalam sistem fuzzy untuk menghasilkan data dalam bentuk crisp output [18].

Model sugeno sering digunakan karena memiliki proses perhitungan yang sederhana sehingga membutuhkan waktu relatif cepat dan efisien dan metode ini tidak terlalu membebani sistem dalam melakukan komputasi [12]. Penalaran dengan Metode Sugeno hampir sama dengan penalaran Metode Mamdani, hanya saja *output* (konsekuen) sistem tidak berupa himpunan fuzzy, melainkan berupa konstanta atau persamaan linear [6].

Secara umum bentuk model fuzzy Sugeno Orde-Satu adalah :

IF ( $x_1$  is  $A_1$ ) o ... o ( $x_N$  is  $A_N$ ) THEN  $z = p_1 * x_1 + \dots + p_N * x_N + q$

dengan  $A_i$  adalah himpunan fuzzy ke-*i* sebagai anteseden, dan  $p_i$  adalah suatu konstanta (tegas) ke-*i* dan *q* juga merupakan konstanta dalam konsekuen.

**2.8 Ekstraksi Aturan Fuzzy**

Aturan Fuzzy kadang-kadang dapat diperoleh dari ahli manusia. Akuisisi pengetahuan, bagaimanapun juga adalah tugas yang rumit, dan beberapa bagian sistem tidak diketahui, sistem pakar belum tersedia [10].

Data mining dapat didefinisikan sebagai proses otomatis mencari ukuran pola-pola data yang besar. Data mining menggunakan proses pencarian melalui ukuran data yang besar menggunakan teknik clustering (*K-means*, *Fuzzy K-means*, *Subtractive*) untuk memperoleh data yang relevan dan signifikan dalam pengenalan pola, dan logika fuzzy dari fuzzy inferensi sistem (Mamdani dan TSK) merupakan teknik berbasis untuk mengesktraksi pembuatan aturan-aturan (IF-THEN) [10].

Proses pembangkitan aturan fuzzy dilakukan dengan proses pencarian nilai output melalui langkah-langkah berikut [6] :

- Derajat keanggotaan setiap data *i* dalam cluster *k* dikalikan dengan setiap atribut *j* dari data *i*, yang dapat dimisalkan dengan  $d^{k_{ij}}$  dan dihitung berdasarkan persamaan (2.8).

$$d^{k_{ij}} = x_{ij} * \mu_{ki}^k, d^{k_{i(m+1)}} = \mu_{ki}^k \quad (2.8)$$

- Proses normalisasi dilakukan dengan cara membagi  $d^{k_{ij}}$  dan  $d^{k_{i(m+1)}}$  dengan jumlah derajat keanggotaan setiap titik data *i* pada cluster *k* menggunakan persamaan (2.9) untuk  $d^{k_{ij}}$  da persamaan (2.10) untuk  $d^{k_{i(m+1)}}$ .

$$d^{k_{ij}} = \frac{d^{k_{ij}}}{\sum_{k=1}^c \mu_{ki}} \quad (2.9)$$

$$d^{k_{i(m+1)}} = \frac{d^{k_{i(m+1)}}}{\sum_{k=1}^c \mu_{ki}} \quad (2.10)$$

- Langkah selanjutnya adalah memben-tuk matriks *U* yang berukuran *n* x



$$U = \begin{pmatrix} u_{11} & u_{12} & u_{1m} & u_{1(m+1)} & \dots & u_{1(c*(m+1))} \\ u_{21} & u_{22} & u_{2m} & u_{2(m+1)} & \dots & u_{2(c*(m+1))} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & u_{nm} & u_{n(m+1)} & \dots & u_{n(c*(m+1))} \end{pmatrix}$$

untuk mencari nilai koefisien *output* tiap-tiap aturan pada setiap variabel ( $k_{ij}$ ,  $i=1,2,\dots,n$ ; dan  $j=1,2,\dots,m+1$ ). Karena matriks  $U$  bukan matriks bujursangkar, maka digunakan metode kuadrat terkecil (*least square*).

- Selanjutnya, dilakukan perhitungan kuadrat terkecil untuk membentuk matriks  $K$  dengan persamaan (2.11) [4].

$$[k] = ([U]^T \cdot [U])^{-1} \cdot [U]^T \cdot [Y] \quad (2.11)$$

Keterangan :  $k$  = koefisien *output*

$Y$  = nilai target *output*

$U$  = matriks  $U$

- Untuk mempermudah komputasi, matriks  $K$  yang berukuran  $c \times (m+1)$  :

$$K = \begin{pmatrix} k_{11} & k_{12} & k_{13} & \dots & k_{1m} & k_{10} \\ k_{21} & k_{22} & k_{23} & \dots & k_{2m} & k_{20} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{c1} & k_{c2} & k_{c3} & \dots & k_{cm} & k_{c0} \end{pmatrix}$$

disusun menjadi satu vektor  $k$  :

$$k = [k_{11} \dots k_{1m} k_{10} k_{21} \dots k_{2m} k_{20} \dots k_{c1} \dots k_{cm} k_{c0}]^T$$

Dari proses di atas terbentuklah koefisien *output* yang digunakan sebagai nilai *output* (konsekuen) dari inferensi Sugeno.

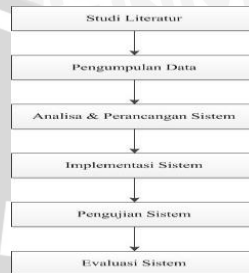
### 2.9 Perhitungan Akurasi

Hasil yang diamati pada penelitian ini adalah seberapa tingkat akurasi metode *K-means clustering* dalam membangkitkan aturan *fuzzy* dalam pengujian data ketahanan hidup penderita kanker payudara. Dalam penelitian ini akurasi dihitung dari jumlah uji ketahanan hidup yang tepat dibagi dengan jumlah data. Akurasi merupakan seberapa dekat suatu angka hasil pengukuran terhadap angka sebenarnya (*true value* atau *reference value*). Tingkat akurasi diperoleh dengan perhitungan sesuai dengan persamaan sebagai 2.12 [13].

$$Akurasi = \frac{\sum \text{data uji benar}}{\sum \text{total data uji}} \times 100\% \quad (2.12)$$

### 3. Metodologi

Berikut ini bagan alur yang digunakan dalam penelitian penggunaan metode *K-Means clustering* untuk pembangkitan aturan pada kasus ketahanan hidup penderita kanker payudara :

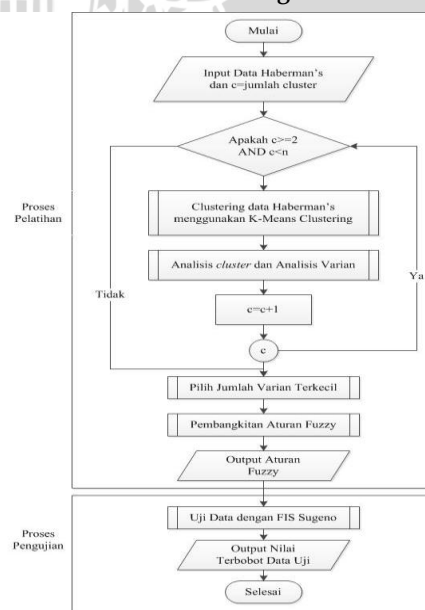


Gambar 3.1 Alur Penelitian

### 3.1 Data Penelitian

Data yang digunakan dalam penelitian ini yaitu data ketahanan hidup penderita kanker payudara (*Haberman's Survival*) yang telah menjalani operasi dalam kurun waktu antara 1958-1970. Data ini diambil dari hasil penelitian di University of Chicago's Billings Hospital, dan data ini disumbangkan oleh Tjen-Sien Lim. Data ini digunakan untuk memprediksi apakah penderita mampu bertahan hidup dalam kurun waktu lima tahun atau lebih atau penderita tidak mampu bertahan hidup kurang dari lima tahun.

### 3.2 Analisa dan Perancangan Sistem



Gambar 3.2 Alur Perancangan Sistem

Secara garis besar, sistem terdiri dari dua proses sistem utama. Yaitu proses pembangkitan aturan *fuzzy* (terdiri dari proses *clustering*, perhitungan standar deviasi, sampai dihasilkan nilai ideal dari *cluster* tersebut), dan proses pengujian keakuratan klasifikasi data menggunakan *Fuzzy Inferensi Sistem Sugeno* dengan rule dari hasil pembangkitan aturan metode *K-Means clustering*. Gambar 3.2 menunjukkan alur perancangan sistem secara umum.

#### 4. Implementasi

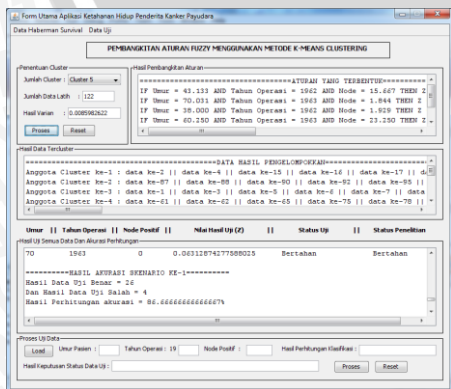
##### 4.1 Implementasi Program

Pada implementasi program, program dibuat dengan menggunakan *method-method*. Di mana *method-method* tersebut merupakan tahapan proses yang dilakukan dalam program. *Method-method* ini mewakili proses dari proses pelatihan sampai proses pengujian data.

##### 4.2 Implementasi Antarmuka

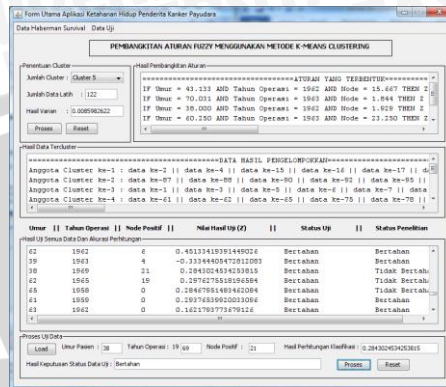
Sistem secara umum ada dua proses bagian utama yaitu proses pelatihan dan proses pengujian data. Di mana kedua proses tersebut berada dalam satu form yang sama. Sistem yang dibuat ini terdiri dari tiga form yaitu form utama, form data Haberman's Survival, dan form data uji.

Form utama yang dibuat ini terdiri dari dua bagian proses utama yaitu proses pelatihan dan proses pengujian. Dalam proses pelatihan, tidak ada inputan yang dilakukan. Hanya proses yang harus dipilih dan dioperasikan sehingga sistem akan menghasilkan aturan yang dibangkitkan seperti Gambar 3.3.



Gambar 3.3 Tampilan Form Utama Saat Proses Pembangkitan Aturan

Sedangkan untuk proses pengujian data, hasil pengujian ditunjukkan pada Gambar 3.4.



Gambar 4.4 Tampilan Hasil Pengujian Data

##### 4.3 Implementasi Uji Coba

Dalam proses pengujian dilakukan dalam 5 skenario untuk masing-masing *cluster* uji (2, 3, 4, 5, 6, 7, 8, dan 9). Dimasing-masing skenario nanti diuji sebanyak 30 data yang berbeda untuk tiap skenarionya.

Sedangkan untuk proses pembangkitan aturan *fuzzy* digunakan sebanyak 122 data latih yang diambil dari data Haberman's Survival. Dan dilakukan perhitungan nilai varian untuk setiap jumlah *cluster* yang diuji. Nilai akurasi dilakukan dengan mencari rata-rata akurasi global untuk seluruh nilai akurasi skenario pada masing-masing jumlah *cluster* uji.

Dari hasil proses pelatihan nantinya dicatat hasil perhitungan varian dan dibandingkan dengan nilai akurasi global untuk seluruh skenario di masing-masing jumlah *cluster* apakah nilai varian terkecil dari jumlah *cluster* memiliki nilai akurasi tertinggi.

#### 5. Analisa dan Pembahasan

Dari skenario percobaan yang telah dilakukan, proses pengujian varian dan pengujian data untuk 5 skenario di setiap *cluster*, ditunjukkan pada tabel hasil pengujian secara lengkap untuk seluruh varian pada Tabel 5.1 dan perhitungan kategori data uji yang ditunjukkan pada Tabel 5.2.

Bisa kita lihat dari hasil perhitungan varian yang digunakan untuk menentukan

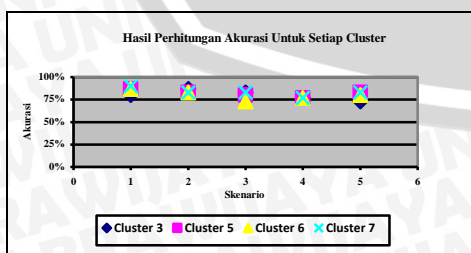
cluster ideal dengan menggunakan sistem pada Tabel 5.1. Berbeda dengan data pada Tabel 5.2 bisa dilihat untuk rata-rata akurasi global di setiap cluster, menunjukkan bahwa pengujian akurasi paling maksimal dilakukan dengan menggunakan cluster 7 dengan rata-rata akurasi globalnya berkisar 83% di setiap percobaan. 3% lebih tinggi dari jumlah cluster 6 yang merupakan jumlah cluster yang memiliki nilai varian terkecil dari semua cluster yang lain.

Tabel 5.1 Hasil Perhitungan Varian Masing-Masing Cluster

Hasil Perhitungan Varian Cluster Uji								
K	2	3	4	5	6	7	8	9
V	0.009	0.00856	0.013	0.0086	0.001	0.011	0.012	0.01145

Tabel 5.2 Data Hasil Perhitungan Akurasi

Cluster	Percobaan Ke-1	Hasil Perhitungan Kategori		Akurasi	Rata-Rata
		Hasil Benar	Hasil Salah		
2	1	24	6	80%	80%
	2	26	4	87%	
	3	25	5	83%	
	4	23	7	77%	
	5	22	8	73%	
3	1	16	14	53%	66%
	2	21	9	70%	
	3	24	6	80%	
	4	20	10	67%	
	5	18	12	60%	
4	1	25	5	83%	79%
	2	24	6	80%	
	3	23	7	77%	
	4	23	7	77%	
	5	23	7	77%	
5	1	26	4	87%	82%
	2	25	5	83%	
	3	24	6	80%	
	4	23	7	77%	
	5	25	5	83%	
6	1	26	4	87%	80%
	2	25	5	83%	
	3	22	8	73%	
	4	23	7	77%	
	5	24	6	80%	
7	1	27	3	90%	83%
	2	25	5	83%	
	3	25	5	83%	
	4	23	7	77%	
	5	25	5	83%	
8	1	27	3	90%	81%
	2	24	6	80%	
	3	25	5	83%	
	4	23	7	77%	
	5	23	7	77%	
9	1	24	6	80%	78%
	2	24	6	80%	
	3	23	7	77%	
	4	23	7	77%	
	5	23	7	77%	



Gambar 5.1 Grafik Hasil Perhitungan Akurasi Untuk Setiap Cluster

Dari grafik Gambar 5.1 juga menunjukkan pada semua proses percobaan, pengujian akurasi dengan jumlah aturan 7 menghasilkan perhitungan paling maksimal. Hasil akurasi maksimal dengan menggunakan jumlah cluster 7 terjadi pada skenario ke-1. Dari hasil perhitungan rata-rata akurasi global dan gambaran grafik Gambar 5.1, semakin jelas bahwa penentuan jumlah cluster yang paling ideal belum bisa optimal dengan menggunakan nilai varian.

## 6. Kesimpulan

Berdasarkan hasil dari penelitian tentang implementasi metode K-Means clustering untuk pembangkitan aturan fuzzy pada klasifikasi ketahanan hidup penderitanya kanker payudara, dapat diambil kesimpulan :

1. Pembangkitan aturan fuzzy menggunakan metode K-Means clustering merupakan metode yang tepat. Terbukti dari hasil perhitungan akurasi pengujian data uji setiap skenario dan akurasi pengujian data uji skenario global sama-sama menghasilkan akurasi minimal di atas 50%.
2. Dari hasil perhitungan nilai varian dan hasil perhitungan akurasi. Bisa diambil kesimpulan, proses penentuan jumlah cluster paling ideal masih belum optimal dengan menggunakan nilai varian terkecil. Dengan melihat hasil penelitian diperlukan adanya optimalisasi dalam menentukan cluster paling ideal.

## 7. Daftar Pustaka

- [1] Andayani, Sri. 2007, "Pembentukan Cluster dalam Knowledge Discovery in Database dengan Algoritma K-Means", FMIPA UNY, Yogyakarta.
- [2] Budiyo, M. A. K. 2002, "Gizi dan Kesehatan", Bayu Media dan UMM Press, Malang.
- [3] Dhanasaputra, Nalendra., Santosa, Budi. 2012, "Pengembangan Algoritma Cat Swarm Optimization (CSO) Untuk Klasifikasi", ITS, Surabaya.
- [4] Fariska, M. Andy. 2008, "Peramalan Multi Atribut Dengan Menggunakan Fuzzy Clustering". <http://repo.eepis-its.edu/425/1/972.pdf>. [30 November 2012].



- [5] Han, Jiawei., and Kamber, Micheline. 2006, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, Second Edition, San Francisco.
- [6] Kusumadewi, Sri., Purnomo, Hari. 2010, "Aplikasi Logika Fuzzy untuk Pendukung Keputusan", Graha Ilmu, Yogyakarta.
- [7] Lim, Tjen, Sien. 1999, "Haberman's Survival Data", University of Chicago's Billings Hospital, USA.
- [8] Mardiana, Lina. 2004, "Kanker pada Wanita", Penebar Swadaya, Jakarta.
- [9] Martiana, E., Rosyid, N., Aguseta, U. 2010, "Mesin Pencari Dokumen dengan Pengklasteran Secara Otomatis", TELKOMNIKA 8 (1), hal. 41-48.
- [10] Moreno, Juan, E., et al. 2007, "Data Mining for extraction of fuzzy IF-THEN rules using Mamdani and Takagi-Sugeno-Kang FIS". [http://www.engineeringletters.com/issues\\_v15/issue\\_1/EL\\_15\\_1\\_13.pdf](http://www.engineeringletters.com/issues_v15/issue_1/EL_15_1_13.pdf) [5 Desember 2012].
- [11] Prett, William, B., et al. 1994, "The Cancer Problem", The Anticancer Drugs Second Edition, Oxford University Press, New York, hal 3-16.
- [12] Priyono, Agus., dkk. 2007, "Generation Of Fuzzy Rules With Subtractive Clustering". <http://eprints.utm.my/1451/1/JTDIS43D%5B10%5Dnew.pdf> [30 November 2012].
- [13] Ramdhany, Dhany, Nugraha., dkk. 2006, "Diagnosis Gangguan Sistem Urinari Pada Anjing dan Kucing Menggunakan VFI 5", Jurnal Ilmu Komputer dan Informasi, Vol. 2, No. 2, ISSN 1979-0732, hal. 86-94.
- [14] Ramli, H. Muchlis., Umbas, Rainy., dan Panigoro, Sonar, S. 2002, "Deteksi Dini Kanker", Balai Penerbit FKUI, Jakarta.
- [15] Santoso, B. 2007, "Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis", Graha Ilmu, Yogyakarta.
- [16] Tiolena, Ristarolas. 2009, "Faktor-faktor yang Mempengaruhi Keterlambatan Pengobatan Pada Wanita Penderita Kanker Payudara RSUP H. Adam Malik Medan Tahun 2008", USU Repository, Universitas Sumatera Utara, Medan.
- [17] Turban, et al. 2005, "Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas) Edisi Ketujuh, Andi, Yogyakarta.
- [18] Widhiastiwi, Yuni. 2007, "Model Fuzzy", Bina Widya, Vol.18, No. 02, hal. 88-94.