

**SISTEM PENDIAGNOSA PENYAKIT HIPERKOLESTEROL
MENGUNAKAN ALGORITMA C4.5**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika

Oleh :

DESTRI ARIANTI
10251020348



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU**

2010

SISTEM PENDIAGNOSA PENYAKIT HIPERKOLESTEROL MENGUNAKAN ALGORITMA C4.5

**DESTRI ARIANTI
10251020348**

Tanggal Sidang : 05 Februari 2010
Periode Wisuda : Juli 2010

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Kesehatan merupakan hal yang begitu penting bagi manusia. Tidak banyak orang mengetahui bahwa penyakit hiperkolesterol merupakan faktor risiko penyebab kematian di usia muda. Pemeriksaan kolesterol biasanya dilakukan dari hasil rekam medis pasien dan dari hasil laboratorium. Pembacaan hasil tersebut biasanya dilakukan oleh dokter. Hal ini menyebabkan pasien harus menunggu jadwal yang ditentukan oleh dokter tersebut. Untuk itu, perlu sebuah sistem yang membantu dalam meningkatkan dan menambah basis pengetahuan, dalam hal ini sistem pakar berperan sebagai rekan kerja (*partner*).

Sistem yang dirancang adalah sistem pendiagnosa penyakit hiperkolesterol menggunakan metode pohon keputusan (*Decision tree*) dengan menggunakan Algoritma C4.5, sistem ini memberikan hasil berdasarkan beberapa kriteria yaitu IMT (Indeks Massa Tubuh), LDL (*Low Density Lipoprotein*), HDL (*High Density Lipoprotein*), TC (*Total Cholesterol*), TGA (*Trigeliserida*), dan tingkatan perokok. Sistem ini dirancang dan dibangun dengan menggunakan bahasa pemrograman VB 6.0 dan *database Microsoft access*.

Setelah dilakukan analisa dan pengujian, sistem pendiagnosa penyakit hiperkolesterol ini dapat memberikan hasil berupa penggolongan tingkatan kolesterol yaitu: Hiperkilomikronemia, Hiperkolesterol, Hiperlipoproteinemia tipe III, Hiperlipoproteinemia tipe IV dan Hiperlipoproteinemia Tipe V. Dari data Training yang berjumlah 900 data dan data testing berjumlah 20 data diperoleh *error* sebesar 5%.

Kata kunci : Algoritma C4.5, Penyakit Hiperkolesterol, Pohon Keputusan

HYPERCHOLESTEROLEMIA DIAGNOSIS SYSTEM USING C4.5 ALGORITHM

**DESTRI ARIANTI
10251020348**

*Final Exam Date : February 5th, 2010
Graduation Ceremony Period : July, 2010*

*Informatics Engineering Department
Faculty of Sciences and Technology
State Islamic University of Sultan Syarif Kasim Riau*

ABSTRACT

Health is important for humans. Not many people know that the hypercholesterolemia disease is a risk factor because of death at younger ages. Cholesterol inspection is usually done from the results of patient medical records and laboratory results. The reading of these results is usually performed by a doctor. This causes the patient to wait for the schedule prescribed by the doctor. For that, need a system that helps in improving and increasing the knowledge base, in this case the expert system acts as a partner.

Designed system is hypercholesterolemia diagnosis system using a decision tree using C4.5 algorithm, this system provides results based on several criteria: Body Mass Index (BMI), Low Density Lipoprotein (LDL), High Density Lipoprotein (HDL), Total Cholesterol (TC), Trigeliserida (TGA), and levels of smokers. This system was designed and built using Visual Basic 6.0 programming language and Microsoft Access databases.

After analysis and testing, hypercholesterolemia diagnosis system can give results in the form of classification levels of cholesterol are: hyperchylomicronemia, hypercholesterolemia, hyperlipoproteinemia type III , hyperlipoproteinemia type IV and hyperlipoproteinemia type V. From Of the total 900 Training data and testing data are 20 data obtained for 5% error.

Keywords: C4.5 algorithm, Decision tree, Hypercholesterolemia

DAFTAR ISI

	Halaman
<i>COVER</i>	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL.....	xvi
DAFTAR LAMPIRAN	xvii
DAFTAR SIMBOL.....	xviii
BAB I PENDAHULUAN	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Batasan Masalah.....	I-3
1.4 Tujuan Penelitian	I-3
1.5 Sistematika Penulisan	I-3
BAB II LANDASAN TEORI.....	II-1
2.1 Penyakit Hiperkolesterol.....	II-1
2.1.1 Faktor penyebab tingginya kadar Kolesterol	II-3
2.1.2 Jenis-jenis Penyakit Kolesterol	II-4
2.2 Sistem Pakar.....	II-8
2.2.1 Definisi Sistem Pakar.....	II-8
2.2.2 Karakteristik dan Nilai Guna Sistem Pakar	II-9

2.2.3 Konsep Dasar Sistem Pakar	II-10
2.2.4 Bentuk Sistem Pakar	II-11
2.2.5 Struktur Sistem Pakar.....	II-12
2.2.6 Komponen Sistem Pakar	II-13
2.2.7 Basis Pengetahuan (<i>Knowlegde Base</i>)	II-14
2.2.8 Aturan Basis Pengetahuan.....	II-15
2.2.9 Motor Inferensi.....	II-16
2.3 <i>Learning</i>	II-16
2.3.1 Metode-metode Learning	II-16
2.4 Pohon Keputusan (<i>Decision Tree</i>)	II-17
2.4.1 Algoritma-algoritma dalam Metode Pohon Keputusan	II-18
2.4.2 Algoritma C4.5.....	II-19
2.4.3 Tiga Prinsip Kerja Algoritma C4.5	II-19
2.4.3.1 Pembuatan Pohon Keputusan	II-20
2.4.3.1.1 <i>Information Gain</i>	II-20
2.4.3.1.2 Penanganan Atribut Kotinyu	II-25
2.4.3.2 Pemangkasan Pohon Keputusan.....	II-26
2.4.3.2.1 <i>Pruning Tree</i>	II-26
2.4.3.3 Pembuatan aturan-aturan dari Pohon Keputusan ...	II-27
BAB III METODOLOGI PENELITIAN.....	III-1
3.1 Studi Pustaka.....	III-2
3.2 Identifikasi Masalah	III-2
3.3 Perumusan Masalah	III-2
3.4 Pemilihan Metode	III-3
3.5 Pengumpulan data	III-3
3.6 Analisis Sistem.....	III-4
3.7 Perancangan Sistem	III-6
3.8 Implementasi	III-6
3.9 Pengujian.....	III-7
3.10 Kesimpulan dan Saran.....	III-7
BAB IV ANALISA DAN PERANCANGAN	IV-1

4.1	Analisa Sistem.....	IV-1
4.1.1	Analisa Lama	IV-1
4.1.2	Analisa Sistem Baru	IV-3
4.1.3	Analisa Kebutuhan data	IV-5
4.1.4	Analisa Penyelesaian Masalah	IV-7
4.1.4.1	Langkah-langkah Membangun Pohon Keputusan IV-8	
4.1.4.1.1	Data Dari Tabel Keputusan Menjadi Model Pohon Keputusan	IV-8
4.1.4.1.2	Data Pohon Keputusan dilakukan Perhitungan <i>Pruning Tree</i>	IV-16
4.1.4.1.3	Pembuatan Aturan Pohon Keputusan dari Pohon Keputusan	IV-19
4.1.5	Analisa Fungsional Sistem	IV-20
4.1.5.1	Diagram Konteks (<i>Context Diagram</i>).....	IV-21
4.1.5.2	<i>Data Flow Diagram</i> (DFD)	IV-22
4.1.6	Analisa data sistem.....	IV-24
4.2	Perancangan Sistem.....	IV-26
4.2.1	Perancangan <i>Flowchat</i>	IV-27
4.2.2	Perancangan Database.....	IV-27
4.2.3	Perancangan Struktur Menu Sistem	IV-31
4.2.4	Perancangan Antar Muka Sistem.....	IV-31
BAB V IMPLEMENTASI DAN PENGUJIAN		V-1
5.1	Implementasi Sistem.....	V-1
5.1.1	Analisa Pemilihan Perangkat Lunak	V-1
5.1.2	Batas Implementasi	V-1
5.1.3	Lingkungan implementasi	V-2
5.2	Implementasi Sistem Pendiagnosa Penyakit Hiperkolesterol	V-3
5.3	Pengujian Sistem	V-12
5.3.1	Lingkungan Pengujian	V-13
5.3.2	Pengujian.....	V-13

5.3.2.1 Pengujian Dengan Menggunakan <i>Blakbox</i>	V-14
5.3.2.2 Pengujian Dengan Menggunakan <i>User Acceptance Test</i>	V-15
5.3.2.3 Pengujian Dengan Menggunakan Performansi Algoritma C4.5	V-16
5.3.2.4 Kesimpulan Pengujian	V-17
BAB VI PENUTUP	VI-1
6.1. Kesimpulan	VI-1
6.2. Saran.....	VI-2
DAFTAR PUSTAKA	
LAMPIRAN	
DAFTAR RIWAYAT HIDUP	

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan merupakan hal yang begitu penting bagi manusia. Ironisnya banyak sekali penyakit-penyakit yang ada pada akhirnya terlambat didiagnosa sehingga mencapai tahap kronis yang membuatnya sulit untuk ditangani. Tidak banyak yang mengetahui bahwa hiperkolesterol merupakan faktor risiko penyebab kematian di usia muda. Berdasarkan laporan Badan Kesehatan Dunia pada tahun 2002, tercatat sebanyak 4,4 juta kematian akibat hiperkolesterol atau sebesar 7,9% dari jumlah total kematian di usia muda (www.medicastore.com, 2008).

Secara normal kolesterol dapat diproduksi oleh tubuh sesuai dengan kebutuhan namun dapat meningkat karena makan yang dimakan mengandung kolesterol. Kolesterol yang berlebihan dapat membuat penumpukan di pembuluh darah dan dapat menimbulkan penyempitan. Karena hal tersebut dapat menyebabkan seseorang terserang stroke dan jantung.

Pemeriksaan kolesterol biasanya dilakukan dari hasil rekam medis pasien dan dari hasil laboratorium. Pembacaan hasil tersebut biasanya dilakukan oleh dokter. Hal ini menyebabkan pasien harus menunggu jadwal yang ditentukan oleh dokter tersebut. Untuk itu, perlu sebuah sistem yang membantu dalam meningkatkan dan menambah basis pengetahuan, dalam hal ini sistem pakar berperan sebagai rekan kerja (*partner*).

Berdasarkan tinjauan yang dilakukan, diperoleh variabel atribut yang berhubungan dengan penyebab hiperkolesterol, yaitu : Jenis Kelamin, Usia,

Tekanan Darah, Indeks Massa Tubuh (IMT), *Low Density Lipoprotein* (LDL), *High Density Lipoprotein* (HDL), *Total cholesterol* (TC), *Trigliserida* (TGA) dan masih banyak lagi yang bisa dijadikan pertimbangan agar prediksi menjadi akurat. Untuk memprediksi suatu kasus tentu diperlukan seorang pakar, yang bisa membantu dalam memprediksi diagnosa penyakit hiperkolesterol. Tingginya kasus penderita penyakit hiperkolesterol membuat perlu adanya penanganan khusus.

Dari permasalahan diatas, maka dibutuhkan sebuah sistem komputerisasi yang dapat membantu dalam pendiagnosaan penyakit hiperkolesterol. Sistem tersebut memerlukan suatu metode yang mampu memberikan hasil kesimpulan penyakit dan perkiraan kolesterol. Salah satu metode yang dapat digunakan adalah pohon keputusan (*Decision Tree*). Pohon Keputusan (*Decision Tree*) merupakan salah satu metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-data yang terdapat kesalahan serta mampu mempelajari fungsi-fungsi *disjunctive*. (Suyanto, 2007). Metode ini dibangun menggunakan algoritma C4.5.

Algoritma C4.5 adalah salah satu dari jenis pohon keputusan yang sangat populer. Algoritma C4.5 berusaha membangun pohon keputusan secara *top-down* yang disebut juga *Top Down Induction Decision Tree* (TDIDT). Pohon Keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variable target.

1.2 Rumusan Masalah

Untuk dapat memprediksikan hasil yang diperoleh pada penyakit hiperkolesterol maka rumusan masalah nya adalah “bagaimana merancang dan membangun sebuah sistem untuk mendiagnosa Penyakit Hiperkolesterol menggunakan Algoritma C4.5”.

1.3 Batasan Masalah

Ruang lingkup yang membatasi permasalahan yang dibahas pada tugas akhir ini antara lain :

1. Kriteria-kriteria yang menjadi atribut dan nilai untuk penyakit hiperkolesterol adalah : Indeks Massa Tubuh (IMT), *Low Density Lipoprotein* (LDL), *High Density Lipoprotein* (HDL), *Total Cholesterol* (TC), *Trigeliserida* (TGA), tingkatan perokok.
2. Menggunakan penalaran berbasis aturan (*Rule-Based Reasoning*)

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai oleh penulis dari tugas akhir ini adalah untuk Merancang dan membangun suatu sistem untuk diagnosa penyakit hiperkolesterol menggunakan Algoritma C4.5.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini dibagi menjadi 6 (enam) bab. Berikut penjelasan tentang masing- masing bab :

BAB I Pendahuluan

Bab ini, membahas latar belakang, rumusan masalah, batasan masalah, tujuan, dan sistematika penulisan Tugas Akhir yang dibuat.

BAB II Landasan Teori

Bab ini menguraikan tentang penyakit hiperkolesterol dan faktor-faktor penyebab hiperkolesterol, Teori sistem pakar, Teori *Decission Tree* dan Teori Algoritma C4.5

BAB III Metodologi

Bab ini berisi tentang studi pendahuluan, Studi pustaka, Identifikasi Masalah, Perumusan Masalah, Pemilihan Metode, Pengumpulan Data, Analisa Sistem, Perancangan Sistem, Implementasi, Pengujian dan Kesimpulan.

BAB IV Analisa dan Perancangan

Bab ini berisi analisis penggunaan metode Algoritma C4.5 untuk mendiagnosa penyakit Hiperkolesterol, analisa sistem lama, analisa sistem baru, analisa kebutuhan data, analisa penyelesaian masalah, analisa fungsional sistem, dan analisa data sistem. Perancangan sistem terdiri dari perancangan *flowchart*, perancangan *database*, perancangan struktur menu dan perancangan antarmuka.

BAB V Implementasi

Bab ini akan membahas implementasi Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Algoritma C4.5, lingkungan

implementasi, batasan implementasi, pengujian sistem, hasil pengujian dan kesimpulan pengujian.

BAB VI Penutup

Bab ini akan menjelaskan beberapa kesimpulan yang didapatkan dari pembahasan penggunaan metode Algoritma C4.5 untuk mendiagnosa Penyakit Hiperkolesterol pada bab sebelumnya, serta beberapa saran sebagai hasil akhir penelitian yang dilakukan.

BAB II

LANDASAN TEORI

2.1 Penyakit Hiperkolesterol

Kolesterol adalah salah satu komponen dari lemak dan lemak merupakan salah satu zat gizi yang diperlukan oleh tubuh di samping zat gizi yang lain. Lemak menjadi salah satu sumber energi yang memiliki kalori paling tinggi dan pada lemak khususnya kolesterol berfungsi untuk membentuk dinding-dinding sel dalam tubuh, dan kolesterol juga merupakan bahan utama dari pembentukan hormon steroid (<http://www.pjnhk.go.id>, 2008)

Kolesterol yang berlebihan dapat membuat penumpukan di pembuluh darah dan dapat menimbulkan penyempitan, karena itulah dapat menyebabkan seseorang terserang stroke dan jantung. Untuk mengukur kolesterol seseorang dilakukan suatu tes darah. Setiap orang dewasa di atas umur 20 tahun perlu mengetahui tingkatan kolesterol mereka dengan mengecek sedikitnya sekali dalam 5 tahun. *Total Cholesterol* (TC) harus dianalisa untuk memperhatikan tingkatan *Low Density Lipopretin* (LDL), *High Density Lipopretin* (HDL) dan *Trigliserida* (TGA) (Ohio State University Medical Center, 2007)

Kolesterol diangkut oleh lipoprotein menuju sel-sel yang membutuhkan termasuk jantung dan otak, sedangkan kelebihan kolesterol diangkut oleh HDL menuju hati yang kemudian diurai dan masuk ke empedu sebagai asam empedu. LDL memiliki kandungan lemak yang lebih banyak dibandingkan dengan HDL. LDL dianggap lemak yang jahat karena dapat menempel di dinding

pembuluh darah sedangkan HDL dianggap lemak yang baik karena ia membawa lemak yang berlebih menuju hati.

Untuk mengukur kadar kolesterol dari hasil yang diperoleh atas tes darah, dapat diketahui dengan melihat tabel 2.1. Hiperkolesterol adalah satu dari sekian banyak faktor utama pemupukan lemak dan pengerasan arteri (aterosklerosis) (Arora, 2007). Kadar kolesterol yang tinggi (hiperkolesterol) dapat membuat darah menjadi kental sehingga mengurangi oksigen dan gejalanya dapat dirasakan seperti sakit kepala dan pegal-pegal, namun tidak sedikit yang tanpa gejala.

Kadar kolesterol harus dikontrol, kadar total kolesterol jangan lewat ambang standar yaitu 200 mg/dL, kadar TGA kurang dari 150 mg/dL, kadar HDL harus tinggi yaitu lebih dari 50 mg/dL, dan kadar LDL jangan lebih dari 100mg/dL.

Tabel 2.1 Klasifikasi LDL dan HDL Kolesterol,

Total Kolesterol dan Trigliserida

LDL (Kolesterol Jahat)	
<100 mg/dL	Normal
100 – 159 mg/dL	Sedang
Lebih dari 160mg/dL	Tinggi
HDL (Kolesterol Baik)	
0 – 34 mg/dL	Rendah
35 – 55 mg/dL	Tinggi
TC (Total Kolesterol)	
0 – 200 mg/dL	Normal
Lebih dari 200 mg/dL	Tinggi
TGA (Trigliserida)	
100 – 150 mg/dL	Normal
151 – 190 mg/dL	Sedang
191 – 200 mg/dL	Tinggi

Sumber : Dr. Irwan SP.JP

2.1.1 Faktor penyebab tingginya kadar Kolesterol

Faktor penyebab tingginya kadar Kolesterol, sebagai berikut (http://medicastore.com, 2009):

1. Berat Badan

Cara menilai berat badan dengan menggunakan IMT (Indeks Masa Tubuh), dengan rumus :

$$IMT = \frac{Berat\ Badan(kg)}{Tinggi\ Badan(m)^2} \dots\dots\dots(1)$$

Penilaian berat badan berdasarkan IMT menggunakan batas ambang.

Jika berat badan 67 kg dan tinggi badan 1,73 m, maka :

$$IMT = \frac{67}{(1.73)^2} = \frac{67}{2.9929} = 22.3863$$

Kategori dengan IMT = 22.3863 adalah Normal

Tabel 2.2 Kategori Batas Ambang IMT

	Kategori	Batas Ambang
Kurus	Kekurangan berat badan tingkat berat	< 17,0
	Kekurangan berat badan tingkat ringan	17,0 – 18,5
Normal		> 18,5 – 25,0
Gemuk	Kelebihan berat badan tingkat ringan	>25,0 – 27,0
	Kelebihan berat badan tingkat berat	> 27,0

2. Merokok

Merokok adalah kebiasaan buruk yang salah satunya dapat memicu penebalan atau penyempitan pembuluh darah. Rokok mengandung nikotin yang

memacu pengeluaran zat-zat seperti Adrenalin, yang merangsang peningkatan denyut jantung dan tekanan darah

Rokok juga mengandung karbon mono-oksida (CO) yang memiliki kemampuan jauh lebih kuat daripada sel darah merah (*hemoglobin*) untuk mengikat oksigen. CO menurunkan kapasitas sel darah merah membawa oksigen ke jaringan-jaringan termasuk jantung. Hal ini perlu diperhatikan terutama bagi penderita hiperkolesterol, karena pembuluh darahnya sudah terdapat plak dengan aliran darah yang sudah sangat berkurang.

3. Kurang Olah raga

Dalam suatu studi, aerobik yang rutin dapat meningkatkan kadar kolesterol HDL 3-9%. Lakukanlah aerobik sedikitnya 30-60 menit sesering mungkin setiap minggu. Sebaiknya, olahraga setiap hari atau jogging setiap hari, naik sepeda atau berenang. Jika tidak bisa melakukan olahraga dalam satu waktu khusus, maka anda dapat melakukannya dalam beberapa waktu dalam satu hari.

4. Makanan Tidak Sehat

Menjaga makanan dengan diet rendah lemak. Mengontrol kolesterol bukan artinya tidak bisa makan enak, tetapi mengurangi atau menghindari makanan berlemak.

2.1.2 Jenis-Jenis Penyakit Kolesterol

Terdapat 5 jenis hiperlipoproteinemia yang masing-masing memiliki gambaran lemak darah serta resiko yang berbeda (menurut DR. Irawan, SP.JP):

1. Hiperlipoproteinemia tipe I

Disebut juga *hiperkilomikronemia familial*, merupakan penyakit keturunan yang jarang terjadi dan ditemukan pada saat lahir.

Dimana tubuh penderita tidak mampu membuang kilomikron dari dalam darah.

Anak-anak dan dewasa muda dengan kelainan ini mengalami serangan berulang dari nyeri perut. Hati dan limpa membesar, pada kulitnya terdapat pertumbuhan lemak berwarna kuning-pink (*xantoma eruptif*). Pemeriksaan darah menunjukkan kadar trigliserida yang sangat tinggi. Penyakit ini tidak menyebabkan terjadi *aterosklerosis* tetapi bisa menyebabkan *pankreatitis*, yang bisa berakibat fatal. Penderita diharuskan menghindari semua jenis lemak (baik lemak jenuh, lemak tak jenuh maupun lemak tak jenuh ganda).

2. **Hiperlipoproteinemia tipe II**

Disebut juga *hiperkolesterolemia familial*, merupakan suatu penyakit keturunan yang mempercepat terjadinya *aterosklerosis* dan kematian dini, biasanya karena serangan jantung. Kadar kolesterol LDLnya tinggi. Endapan lemak membentuk pertumbuhan *xantoma* di dalam *tendon* dan kulit.

1 diantara 6 pria penderita penyakit ini mengalami serangan jantung pada usia 40 tahun dan 2 diantara 3 pria penderita penyakit ini mengalami serangan jantung pada usia 60 tahun. Penderita wanita juga memiliki resiko, tetapi terjadinya lebih lambat. 1 dari 2 wanita penderita penyakit ini akan mengalami serangan jantung pada usia 55 tahun. Orang yang memiliki 2 *gen* dari penyakit ini (jarang terjadi) bisa memiliki kadar kolesterol total sampai 500-1200 mg/dL dan seringkali meninggal

karena *penyakit arteri koroner* pada masa kanak-kanak. Tujuan pengobatan adalah untuk menghindari faktor resiko, seperti merokok, dan *obesitas*, serta mengurangi kadar kolesterol darah dengan mengkonsumsi obat-obatan. Penderita diharuskan menjalani diet rendah lemak atau tanpa lemak, terutama lemak jenuh dan kolesterol serta melakukan olah raga secara teratur. Menambahkan bekatul gandum pada makanan akan membantu mengikat lemak di usus. Seringkali diperlukan obat penurun lemak.

3. **Hiperlipoproteinemia tipe III.**

Merupakan penyakit keturunan yang jarang terjadi, yang menyebabkan tingginya kadar kolesterol VLDL dan trigliserida. Pada penderita pria, tampak pertumbuhan lemak di kulit pada masa dewasa awal. Pada penderita wanita, pertumbuhan lemak ini baru muncul 10-15 tahun kemudian. Baik pada pria maupun wanita, jika penderitanya mengalami *obesitas*, maka pertumbuhan lemak akan muncul lebih awal. Pada usia pertengahan, aterosklerosis seringkali menyumbat *arteri* dan mengurangi aliran darah ke tungkai. Pemeriksaan darah menunjukkan tingginya kadar kolesterol total dan trigliserida. Kolesterol terutama terdiri dari VLDL. Penderita seringkali mengalami *diabetes* ringan dan peningkatan kadar *asam urat* dalam darah. Pengobatannya meliputi pencapaian dan pemeliharaan berat badan ideal serta mengurangi asupan kolesterol dan lemak jenuh. Biasanya diperlukan obat penurun kadar

lemak. Kadar lemak hampir selalu dapat diturunkan sampai normal, sehingga memperlambat terjadinya aterosklerosis.

4. **Hiperlipoproteinemia tipe IV.**

Merupakan penyakit umum yang sering menyerang beberapa anggota keluarga dan menyebabkan tingginya kadar trigliserida. Penyakit ini bisa meningkatkan resiko terjadinya aterosklerosis. Penderita seringkali mengalami kelebihan berat badan dan diabetes ringan. Penderita dianjurkan untuk mengurangi berat badan, mengendalikan diabetes dan menghindari alkohol. Bisa diberikan obat penurun kadar lemak darah.

5. **Hiperlipoproteinemia tipe V.**

Merupakan penyakit keturunan yang jarang terjadi, dimana tubuh tidak mampu memetabolisme dan membuang kelebihan trigliserida sebagaimana mestinya. Selain diturunkan, penyakit ini juga bisa terjadi akibat:

- penyalahgunaan alcohol
- kebanyakan merokok
- diabetes yang tidak terkontrol dengan baik
- gagal ginjal
- makan setelah menjalani puasa selama beberapa waktu.

Jika diturunkan, biasanya penyakit ini muncul pada masa dewasa awal.

Ditemukan sejumlah besar pertumbuhan lemak (xantoma) di kulit, pembesaran hati dan limpa serta nyeri perut.

Biasanya terjadi diabetes ringan dan peningkatan asam urat. Banyak penderita yang mengalami kelebihan berat badan. Komplikasi utamanya adalah pankreatitis, yang seringkali terjadi setelah penderita makan lemak dan bisa berakibat fatal. Pengobatannya berupa penurunan berat badan, menghindari lemak dalam makanan dan menghindari rokok. Bisa diberikan obat penurun kadar lemak.

2.2 Sistem Pakar (*Expert System*)

Sistem pakar merupakan salah satu bentuk sistem basis pengetahuan yang paling populer.

2.2.1 Definisi Sistem Pakar

Secara umum, sistem pakar (*expert system*) adalah sistem yang berusaha mengadopsi pengetahuan manusia ke komputer, agar komputer dapat menyelesaikan masalah seperti yang biasa dilakukan oleh para ahli (Kusumadewi, 2003).

Sistem pakar ini diharapkan dapat menyelesaikan suatu permasalahan tertentu dengan meniru kerja ahli. Sehingga orang awam pun dapat menyelesaikan masalah yang cukup rumit yang sekalipun tanpa bantuan ahli dalam bidang tersebut. Bagi para ahli, sistem pakar ini dapat digunakan sebagai asisten yang berpengalaman.

Sistem pakar merupakan cabang dari bidang kecerdasan buatan (*Artificial Intelligence*) yang cukup tua karena sistem ini telah mulai dikembangkan pada pertengahan tahun 1960. Sistem pakar yang muncul pertama kali adalah *General-*

purpose problem solver (GPS) yang dikembangkan oleh Newl dan Simon. Sampai saat ini sudah banyak sistem pakar yang dibuat, seperti MYCIN, DENDRAL, XCON & XSEL, SOPHIE, Prospector, FOLIO, DELTA, dan sebagainya (Kusumadewi, 2003).

2.2.2 Karakteristik dan Nilai Guna Sistem Pakar

Sistem pakar memiliki beberapa karakteristik keahlian yang dipindahkan kedalam program komputer. Karakteristik-karakteristik dari sistem pakar mencakup hal-hal adalah sebagai berikut (Kusumadewi,2003):

1. Basis pengetahuan mudah diperbaharui.
2. Kemampuan mempelajari fakta atau kejadian baru dari pengalamannya sendiri.
3. Mudah dipakai siapa saja.
4. Mampu menjelaskan proses pemikiran dan langkah-langkah yang dilakukan untuk mencapai kesimpulan.
5. Menggunakan pengetahuan untuk menyelesaikan permasalahan.
6. Mampu menangani masalah yang kompleks.

Banyak manfaat yang dapat diambil dengan adanya sistem pakar. Hal itu berupa keuntungan dan manfaat dari sistem pakar yang meliputi (Kusumadewi, 2003):

1. Memungkinkan orang awam bisa mengerjakan pekerjaan para ahli.
2. Bisa melakukan proses secara berulang secara otomatis.
3. Menyimpan pengetahuan dan keahlian para pakar.
4. Meningkatkan output dan produktivitas.

5. Meningkatkan kualitas.
6. Mampu mengambil dan melestarikan keahlian para pakar.
7. Mampu beroperasi dalam lingkungan yang berbahaya.
8. Memiliki kemampuan untuk mengakses pengetahuan.
9. Memiliki reabilitas.
10. Meningkatkan kapabilitas sistem komputer.
11. Memiliki kemampuan untuk bekerja dengan informasi yang tidak lengkap yang mengandung ketidakpastian.
12. Sebagai media pelengkap dalam pelatihan.
13. Meningkatkan kapabilitas dalam penyelesaian masalah.
14. Menghemat waktu dalam pengambilan keputusan.

Selain memiliki keuntungan dan manfaat, sistem pakar juga memiliki keterbatasan, antara lain (Kusumadewi, 2003):

1. Biaya yang diperlukan untuk membuat dan memeliharanya sangat mahal.
2. Sulit dikembangkan. Hal ini berkaitan dengan ketersediaan pakar dibidangnya.
3. Sistem pakar tidak 100% bernilai benar.

2.2.3 Konsep Dasar Sistem Pakar

Konsep dasar sistem pakar mengandung keahlian, ahli, pengalihan keahlian, inferensi, aturan dan kemampuan menjelaskan (Turban, 1995 dikutip dari Kusumadewi, 2003).

Keahlian adalah suatu kelebihan penguasaan pengetahuan dibidang tertentu yang diperoleh dari pelatihan, membaca atau pengalaman (Kusumadewi,2003).

Seorang ahli adalah seseorang yang mampu menjelaskan suatu tanggapan, mempelajari hal-hal baru seputar topik permasalahan (domain), menyusun kembali pengetahuan jika dipandang perlu, memecah aturan-aturan jika dibutuhkan dan menentukan relevansi tidaknya keahlian mereka (Kusumadewi,2003).

Pengalihan keahlian dari para ahli ke komputer untuk kemudian dialihkan lagi ke orang lain yang bukan ahli merupakan tujuan utama dari sistem pakar. Proses ini membutuhkan 4 aktivitas, yaitu: tambahan pengetahuan, representasi pengetahuan, inferensi pengetahuan dan pengalihan pengetahuan ke *user* (Kusumadewi,2003).

2.2.4 Bentuk Sistem Pakar

Ada empat bentuk sistem pakar, yaitu (Kusumadewi,2003):

1. Berdiri sendiri. Merupakan software yang berdiri sendiri tidak bergabung dengan software yang lainnya.
2. Tergabung. Merupakan bagian program yang terkandung didalam suatu algoritma (konvensional) atau merupakan program didalamnya memanggil algoritma subrutin lain (konvensional).
3. Menghubungkan ke *software* lain. Merupakan sistem pakar yang menghubungkan kesuatu paket program tertentu, misalnya dengan DBMS.

4. Sistem mengabdikan. Merupakan bagian dari komputer khusus yang dihubungkan dengan suatu fungsi tertentu.

2.2.5 Struktur Sistem Pakar

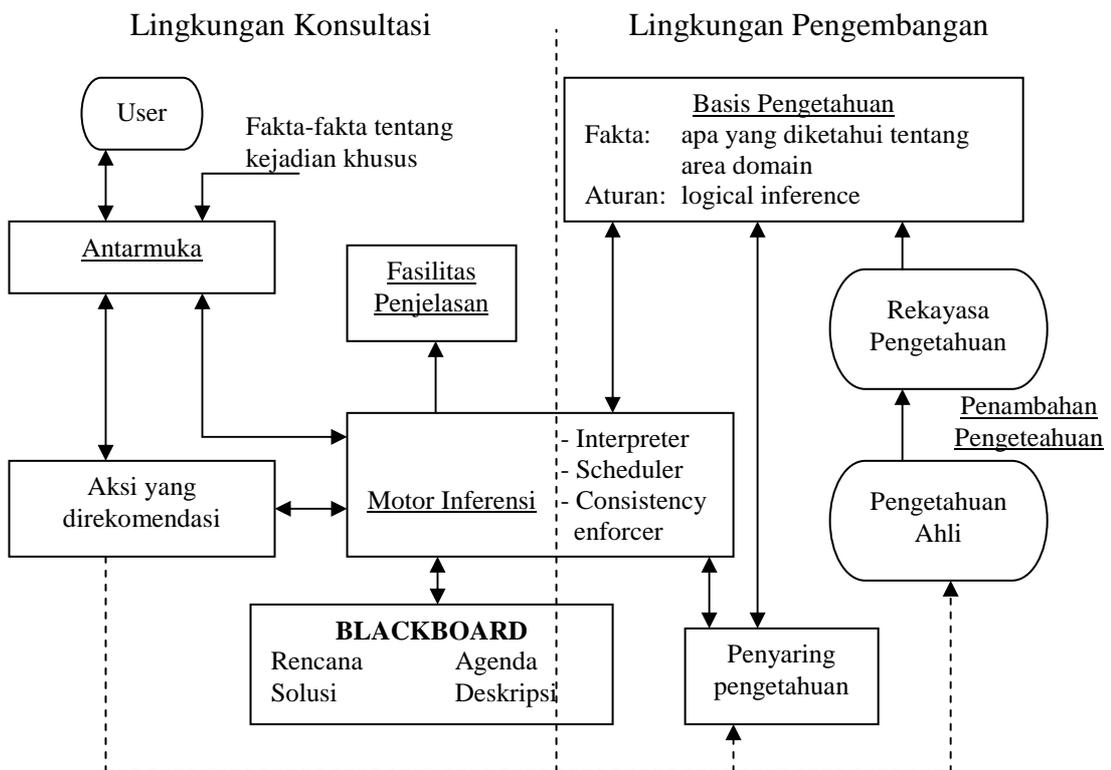
Sistem Pakar terdiri dari dua bagian pokok, yaitu (Kusumadewi,2003):

1. Lingkungan Pengembangan (Development Environment)

Lingkungan pengembangan digunakan sebagai pembangun sistem pakar.

2. Lingkungan Konsultasi (Consultation Environment)

Lingkungan konsultasi digunakan oleh seorang yang bukan ahli untuk berkonsultasi.



Gambar 2.1 Struktur Sistem Pakar (Sumber: Kusumadewi, 2003)

2.2.6 Komponen Sistem Pakar

Komponen-komponen yang ada pada sistem pakar terlihat pada gambar

2.1 adalah (Kusumadewi,2003):

1. Subsistem Penambahan Pengetahuan.

Bagian ini digunakan untuk memasukkan pengetahuan, mengkonstruksi atau memperluas pengetahuan dalam basis pengetahuan. Pengetahuan itu berasal dari ahli, buku, basisdata, penelitian, dan gambar.

2. Basis Pengetahuan.

Berisi pengetahuan-pengetahuan yang dibutuhkan untuk memahami, menformulasikan, dan menyelesaikan masalah.

3. Motor Inferensi (*Inference Engine*).

Program yang berisi metodologi yang digunakan untuk melakukan penalaran terhadap informasi-informasi dalam basis pengetahuan dan blackboard, serta digunakan untuk menformulasikan konklusi. Ada 3 elemen utama dalam motor inferensi, yaitu:

- a. *Interpreter* adalah mengeksekusi item-item yang terpilih dengan menggunakan aturan-aturan dalam basis pengetahuan yang sesuai.
- b. *Scheduler* akan mengontrol agenda.
- c. *Consistency Enforcer* akan berusaha memelihara konsistensi dalam merepresentasikan solusi yang bersifat darurat.

4. Blackboard.

Merupakan area dalam memori yang digunakan untuk merekam kejadian yang sedang berlangsung termasuk keputusan sementara. Ada 3 tipe keputusan yang dapat direkam yaitu:

- a. Rencana: bagaimana menghadapi masalah.
- b. Agenda: aksi-aksi yang potensial yang sedang menunggu untuk dieksekusi.
- c. Solusi: calon aksi yang akan dibangkitkan.

5. Antarmuka.

Digunakan untuk media komunikasi antara *user* dan program.

6. Subsistem Penjelasan.

Digunakan untuk melacak respon dan memberikan penjelasan tentang kelakuan sistem pakar secara interaktif melalui pertanyaan.

7. Sistem Penyaring Pengetahuan.

Sistem ini digunakan untuk mengevaluasi kinerja sistem pakar itu sendiri untuk melihat apakah pengetahuan-pengetahuan yang ada masih cocok untuk digunakan dimasa mendatang.

2.2.7 Basis Pengetahuan (*Knowlegde Base*)

Knowledge base memuat fakta-fakta yang menjelaskan area masalah dan juga teknik menerangkan masalah yang menjelaskan bagaimana fakta-fakta tersebut cocok satu dengan yang lainnya dalam urutan yang logis. Istilah problem domain digunakan untuk menjelaskan area masalah (Mcleod, 1998).

Basis pengetahuan berisi pengetahuan-pengetahuan dalam penyelesaian masalah didalam domain tertentu. Bentuk pendekatan basis pengetahuan yang umum digunakan adalah (Kusumadewi,2003):

1. Penalaran berbasis aturan (*Rule-Based Reasoning*)

Pengetahuan direpresentasikan dengan menggunakan aturan berbentuk: *IF-THEN*. Bentuk ini digunakan apabila memiliki sejumlah pengetahuan pakar pada suatu permasalahan tertentu, dan si pakar dapat menyelesaikan masalah tersebut secara berurutan. Disamping itu, bentuk ini juga digunakan apabila dibutuhkan penjelasan tentang jejak (langkah-langkah) pencapaian solusi.

2. Penalaran berbasis kasus (*Case-Based Reasoning*)

Basis pengetahuan akan berisi solusi-solusi yang telah dicapai sebelumnya, kemudian akan diturunkan suatu solusi keadaan yang terjadi sekarang (fakta yang ada). Bentuk ini digunakan apabila *user* menginginkan untuk tahu lebih banyak lagi pada kasus-kasus yang hampir serupa (mirip). Selain itu, bentuk ini juga digunakan apabila kita telah memiliki sejumlah situasi atau kasus tertentu dalam basis pengetahuan.

2.2.8 Aturan Basis Pengetahuan

Teknik menerangkan masalah (*knowledge representation technique*) yang populer adalah penggunaan aturan. Aturan menentukan apa yang harus dilakukan dalam situasi tertentu dan terdiri dari dua bagian, yaitu: suatu kondisi yang mungkin benar, mungkin tidak dan tindakan yang harus diambil jika kondisi benar. Semua aturan yang ada dalam sistem pakar disebut perangkat aturan (*rule set*). Perangkat aturan dapat bervariasi dari selusin aturan untuk sistem pakar yang

sederhana 10 sampai 500, 1000 sampai 10.000 aturan untuk sistem pakar yang rumit (McLeod, 1998).

2.2.9 Motor Inferensi

Inference engine adalah bagian dari sistem pakar yang melakukan penalaran dengan menggunakan isi *knowledge base* berdasarkan urutan tertentu. Selama konsultasi, *inference engine* menguji aturan-aturan dari *knowledge base* satu demi satu, dan saat kondisi aturan itu benar tindakan tertentu diambil. Dalam terminologi sistem pakar aturan itu ditembakkan (*fired*) saat tindakan diambil (McLeod, 2001).

2.3 Learning

Defenisi *learning* (Pembelajaran) komputer adalah setiap program komputer yang sanggup meningkatkan performansinya melalui pengalaman (*experience*) (Suyanto, 2007).

2.3.1 Metode-metode Learning

Metode-metode yang ada pada learning sebagai berikut (Suyanto, 2007):

1. *Decision Tree Learning*

Pembelajaran Pohon Keputusan (*Decision Tree Learning*) adalah suatu metode untuk memperkirakan fungsi target yang bernilai diskrit, di mana fungsi yang dipelajari diwakili oleh suatu alur keputusan. pohon yang dipelajari juga dapat kembali diwakilkan seperti aturan *if-then* untuk meningkatkan kemampuan membaca manusia. metode pembelajaran ini adalah termasuk paling populer dari algoritma induktif inferensi dan telah dengan sukses berlaku untuk suatu

jangkauan luas dari tugas dari pelajaran untuk mendiagnosa kasus medis, belajar untuk menilai risiko kredit dari pelamar/peminta pinjaman (Mitchell, 1997)

2. Jaringan Syaraf Tiruan (JST)

JST adalah prosesor tersebar paralel (paralel distributed prosesor) yang sangat besar yang memiliki kecenderungan untuk menyimpan pengetahuan yang bersifat pengalaman dan membuatnya siap untuk menggunakan. JST menyerupai otak manusia dalam dua hal, yaitu : Pengetahuan diperoleh jaringan melalui proses belajar; Kekuatan hubungan antar sel syaraf (*neuron*) yang dikenal sebagai bobot-bobot sinaptik digunakan untuk menyimpan pengetahuan (Alexander dan Morton, 1994 dikutip dari Suyanto, 2007).

3. Algoritma Genetika

Algoritma genetika (AG) adalah algoritma pencarian heuristik yang didasarkan atas mekanisme evolusi biologi. Keberagaman pada evolusi biologi adalah variasi dari kromosom antar individu organisme. Variasi kromosom ini akan mempengaruhi laju reproduksi dan tingkat kemampuan organisme untuk tetap hidup. Individu yang lebih kuat (*fit*) akan memiliki tingkat survival dan tingkat reproduksi yang lebih tinggi jika dibandingkan dengan individu yang kurang *fit*. Pada kurun waktu tertentu (sering dikenal dengan istilah generasi), populasi secara keseluruhan akan lebih banyak memuat organisme yang *fit*.

2.4 Pohon Keputusan (*Decision Tree*)

Pohon keputusan adalah suatu pohon di mana tiap-tiap cabang pohon masing-masing menghadirkan suatu pilihan antar sejumlah alternatif, dan masing-masing simpul daun menghadirkan suatu keputusan (Wei, 2008).

Proses dalam pohon keputusan yaitu : mengubah bentuk data (tabel) menjadi model *tree*, mengubah model *tree* menjadi *rule* dan menyederhanakan *rule* (*pruning*).

2.4.1 Algoritma-algoritma dalam Metode Pohon Keputusan

Beberapa algoritma yang ada pada metode pohon keputusan, sebagai berikut :

1. Algoritma ID3 (*Iterative Dychotomizer version 3*)

ID3 adalah suatu algoritma pelajaran pohon keputusan yang sederhana yang dikembangkan oleh Ross Quinlan (1983). Algoritma ID3 berusaha membangun pohon keputusan secara *top-down* yang disebut juga *Top Down Induction Decision Tree* (TDIDT) (Suyanto, 2007). Algoritma TDIDT memerlukan 2 langkah pengerjaan yaitu membangun pohon keputusan (*Construction Decision Tree*) dan menyederhanakan pohon keputusan yang dibuat (*Pruning Decision Tree*).

2. Algoritma Assistant

Algoritma Assistant ini juga termasuk kepada keluarga ID3 yang merupakan sistem induksi atas-kebawah, dari pohon keputusan. (Quinlan, 1986). Sistem secara berulang membangun suatu pohon keputusan yang biner. simpul dari pohon sesuai dengan atribut, dan daun-daun (simpul terminal) ke kelas diagnostik /peramalan. Pada setiap langkah yang berulang tentang konstruksi pohon keputusan, atribut “paling informatif” (suatu atribut yang memperkecil jumlah test yang diharapkan yang diperlukan untuk penggolongan dari kasus yang baru) terpilih dan suatu subtree dibangun. Assistant menyamaratakan lebih lanjut

di atribut dihargai-bilangan bulat dari ACLS dengan mengizinkan atribut dengan nilai kontinyu. Assistant tidak membentuk suatu pohon keputusan secara berulang-ulang seperti cara ID3, tetapi meliputi algoritma untuk pilih pelatihan bernilai baik dari objek tersedia.

3. Algoritma C4.5

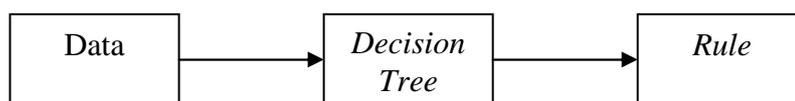
C4.5 adalah ekstensi dari algoritma decision-tree ID3. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah *decision tree* berdasarkan *training* data yang telah disiapkan. Algoritma ini mempunyai inputan berupa *training samples* dan *samples* (Heryanti, 2005). *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya.

2.4.2 Algoritma C4.5

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Kusrini, 2008):

1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.4.3 Tiga Prinsip Kerja Algoritma C4.5



Gambar 2.2 Alur Diagram Pohon Keputusan (Sumber : Achmad, 2003)

Tiga prinsip kerja dari algoritma C4.5, adalah sebagai berikut :

1. Pembuatan Pohon Keputusan
2. Pemangkasan Pohon Keputusan dan Evaluasi
3. Pembuatan aturan-aturan dari Pohon Keputusan

2.4.3.1 Pembuatan Pohon Keputusan

C4.5 membangun pohon keputusan dengan strategi *divide* dan *conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan Gain Ratio, kemudian pada node-node yang terbentuk di level berikutnya, algoritma *divide* dan *conquer* akan diterapkan kembali.

2.4.3.1.1 Information Gain

Information gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih test atribut tiap node pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai test atribut dari suatu node. Ada 2 kasus berbeda pada saat penghitungan *Information Gain*, pertama untuk kasus penghitungan atribut tanpa *missing value* dan kedua, penghitungan atribut dengan *missing value*.

- Perhitungan *Information Gain* tanpa *Missing Value*

Misalkan S berisi *s* data samples. Anggap atribut untuk class memiliki *m* nilai yang berbeda, C_i (untuk $i = 1, \dots, I$). anggap s_i menjadi jumlah samples S pada class C_i . Maka besar *Entropy*-nya dapat dihitung dengan :

$$E(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i * \log_2(p_i) \dots\dots\dots (2)$$

m adalah jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi).

Dimana $p_i = \frac{s_i}{s}$ adalah probabilitas dari sample yang mempunyai class C_i .

Misalkan atribut A mempunyai v nilai yang berbeda, $\{a_1, a_2, \dots, a_v\}$. Atribut A dapat digunakan untuk mempartisi S menjadi v subset, $\{S_1, S_2, \dots, S_v\}$, dimana S_j berisi samples pada S yang mempunyai nilai a_j dari A. Jika A terpilih menjadi test atribut (yaitu, best atribut untuk splitting), maka subset-subset akan berhubungan dengan pertumbuhan node-node cabang yang berisi S. Anggap s_{ij} sebagai jumlah samples class C_i pada subset S_j . *Entropy*, atau nilai information dari subset A adalah :

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_1, s_2, \dots, s_m) \dots \dots \dots (3)$$

$\frac{s_{1j} + \dots + s_{mj}}{s}$ adalah bobot dari subset j th dan jumlah samples pada subset (yang mempunyai nilai a_j dari A) dibagi dengan jumlah total samples pada S. Untuk subset S_j ,

$$E(s_{1j}, s_{2j}, \dots, s_{mj}) = - \sum_{i=1}^m p_{ij} * \log_2(p_{ij}) \dots \dots \dots (4)$$

Dimana $p_{ij} = \frac{s_{ij}}{|s_j|}$ adalah probabilitas sample S_j yang mempunyai class C_i .

Maka nilai information gain atribut A pada subset S adalah

$$\text{Gain}(A) = E(s_1, s_2, \dots, s_m) - E(A) \dots \dots \dots (5)$$

Sebagai contoh : kita ingin mencari apakah pemain baseball akan masuk class yes play atau no play berdasarkan data berikut:

Tabel 2.3 Contoh Data Bermain Baseball

Outlook	Temperature	Humidity	Windy	Play ball
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Sumber : <http://www.cise.ufl.edu>, di akses 2008

Data-data diatas diklasifikasikan berdasarkan atribut outlook, temperature, humidity dan windy. Pembagian hasilnya ada 2, yaitu Yes dan No untuk Play ball. Dari tabel diketahui bahwa ada 14 data, 5 menyatakan No Play ball, 9 menyatakan Yes Play ball. Maka,

$$E(s1, s2) = E(5,9) = -\frac{5}{14} \log_2 \frac{5}{14} - \frac{9}{14} \log_2 \frac{9}{14} = 0,940$$

Entropy untuk atribut outlook:

$$\begin{aligned} E(\text{outlook}) &= \frac{5}{14} \times E(2,3) + \frac{4}{14} \times E(4,0) + \frac{5}{14} \times E(3,2) \\ &= 0,694 \end{aligned}$$

Nilai Gain untuk Atribut outlook:

$$\begin{aligned} \text{Gain}(\text{outlook}) &= E(s1, s2) - E(\text{outlook}) \\ &= 0,940 - 0,694 \\ &= 0,246 \end{aligned}$$

dengan menggunakan cara yang sama, Gain dari semua atribut dapat dicari.

$$\text{Gain (Outlook) } = 0,246$$

$$\text{Gain (Humidity)} = 0,151$$

$$\text{Gain (Windy)} = 0,048$$

$$\text{Gain (Temperature)} = 0,029$$

Setelah nilai information gain pada semua atribut dihitung, maka atribut yang mempunyai nilai information gain terbesar yang dipilih menjadi test atribut.

- Perhitungan *Information Gain* dengan *Missing Value*

Untuk atribut dengan *missing value* penghitungan information gain-nya diselesaikan dengan Gain Ratio. Sebelum menghitung gain ratio terlebih dahulu dihitung $E (s_1, s_2, \dots, s_m)$ dan $E(A)$.

$$E(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i * \log_2 (p_i) \dots\dots\dots (6)$$

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I (s_1, s_2, \dots, s_m) \dots\dots\dots (7)$$

Dimana penghitungan $E (s_1, s_2, \dots, s_m)$ dan $E(A)$ hanya dilakukan pada atribut yang ada nilainya.

Kemudian untuk mencari gain dari atribut A dihitung dengan rumus sebagai berikut :

$$\text{Gain (A) = Prob S yang diketahui * E(A) } \dots\dots\dots (8)$$

Dimana,

A = atribut dengan missing value yang sedang dicari nilai gain-nya,

S = jumlah samples pada subset A yang diketahui nilainya.

Sedangkan nilai split pada atribut A dinyatakan dengan :

$$\text{Split(A) = } -u * \log_2 u - \sum_{i=1}^m p_j * \log_2 (p_j) \dots\dots\dots (9)$$

Dimana,

u adalah prob samples pada atribut A yang merupakan *missing values*.

$$p_j = \frac{s_j}{|s|}$$

adalah probabilitas sample S_j yang diketahui nilainya

Nilai Gain Ratio pada atribut A :

$$\text{Gain Ratio}(A) = \text{Gain}(A) / \text{Split}(A) \dots \dots \dots (10)$$

Sebagai contoh : kita ingin mencari apakah pemain baseball akan masuk class yes play atau no play berdasarkan data berikut:

Tabel 2.4 Contoh Data Bermain Baseball

Outlook	Temp	Humidity	Windy	Play ball
sunny	75	70	yes	Yes
sunny	80	90	yes	No
sunny	85	85	no	No
sunny	72	95	no	No
sunny	69	70	no	Yes
?	72	90	yes	Yes
cloudy	83	78	no	Yes
cloudy	64	65	yes	Yes
cloudy	81	75	no	Yes
rain	71	80	yes	No
rain	65	70	yes	No
rain	75	80	no	Yes
rain	68	80	no	Yes
rain	70	96	no	Yes

Sumber : <http://www.cise.ufl.edu>, di akses 2008

Pertama, menghitung frekuensi pada OUTLOOK sebagai berikut:

Tabel 2.5 Frekuensi Outlook

	Play	Don't play	Total
Sunny	2	3	5
Cloudy	3	0	3
Rain	3	2	5
Total	8	5	13

Sumber : <http://www.cise.ufl.edu>, di akses 2008

Untuk menghitung informasi gainnya adalah sebagai berikut:

$$E(s1, s2) = E(8,5) = -\frac{8}{13} \log_2 \frac{8}{13} - \frac{5}{13} \log_2 \frac{5}{13} = 0,961$$

$$E(\text{outlook}) = \frac{5}{13} \times \left(-\frac{2}{5} \times \log_2 \frac{2}{5} - \frac{3}{5} \times \log_2 \frac{3}{5} \right) + \frac{3}{13} \times \left(-\frac{3}{3} \times \log_2 \frac{3}{3} - \frac{0}{3} \times \log_2 \frac{0}{3} \right) + \frac{5}{13} \times \left(-\frac{3}{5} \times \log_2 \frac{3}{5} - \frac{2}{5} \times \log_2 \frac{2}{5} \right) = 0,747$$

$$\text{Gain}(\text{outlook}) = \frac{13}{14} \times (0,961 - 0,747) = 0,199$$

$$\text{Split} = -\frac{5}{14} \log_2 \frac{5}{14} - \frac{3}{14} \log_2 \frac{3}{14} - \frac{5}{14} \log_2 \frac{5}{14} - \frac{1}{14} \log_2 \frac{1}{14} = 1,809$$

$$\text{Gain Ratio} = \frac{0,199}{1,809} = 0,110$$

Setelah semua Gain diketahui maka dapat ditentukan atribut mana yang layak menjadi root. Atribut yang layak adalah atribut yang mempunyai Gain terbesar

2.4.3.1.2 Penanganan Atribut Kotinyu

Algoritma C4.5 juga menangani masalah atribut kontinyu. Salah satu cara adalah dengan *Entropy-Based Discretization* yang melibatkan penghitungan class entropy.

Misalkan T membagi S example menjadi subset S1 dan S2. Umpakan ada k class C1, C2, ..., Ck. Misal P(Ci, Sj) menjadi perbandingan dari example pada Sj yang mempunyai class i.

Maka class entropy dari subset Sj didefinisikan dengan :

$$\text{Ent}(S) = - \sum_{i=1}^k P(Ci, S) \log(P(Ci, S)) \dots \dots \dots (11)$$

Dan class information entropy E(A, T;S)

$$E(A, T;S) = \frac{|S1|}{|S|} \text{Ent}(S1) + \frac{|S2|}{|S|} \text{Ent}(S2) \dots \dots \dots (12)$$

Dimana $Ent(S_j)$ = class entropy dari subset S_j

S_j = subset dari S

C_i = class i

$P(C_i, S_j)$ = perbandingan instance dari S_j yang berada pada class C_i

$E(A, T_A; S)$ = class information entropy partisi dengan cut point T_A di A

A = atribut

$|S_k|$ = jumlah instance di S_k

Cut point yang terbaik adalah yang memberikan class information entropy yang paling kecil diantara semua kandidat cut point.

2.4.3.2 Pemangkasan Pohon Keputusan

Karena pohon yang dibangun dapat berukuran besar dan tidak mudah “dibaca”, C4.5 dapat menyederhanakan pohon dengan melakukan pemangkasan berdasarkan nilai tingkat kepercayaan (*confidence level*). Selain untuk pengurangan ukuran pohon, pemangkasan juga bertujuan untuk mengurangi tingkat kesalahan prediksi pada kasus baru.

2.4.3.2.1 Pruning Tree

Pruning tree adalah melakukan suatu kegiatan untuk mengganti suatu *subtree* dengan suatu *leaf*. Penggantian dilakukan jika *error rate* pada *subtree* lebih besar jika dibandingkan dengan *single leaf* (Kumar, 2006).

Pada C4.5 perkiraan *error* untuk satu *node* dihitung dengan :

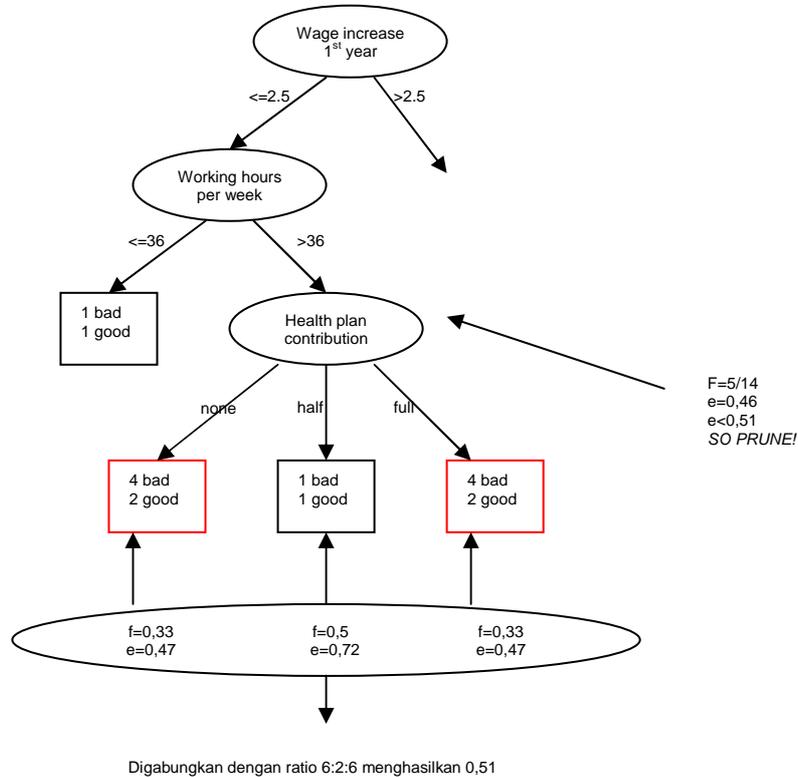
$$e(N, f, z) = \frac{f + \frac{z^2}{2N} + z \sqrt{\frac{f(1-f)}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}} \dots\dots\dots (13)$$

Jika $c = 25\%$ (default untuk C4.5) maka $z = 0,69$ (dari distribusi normal)

f = error pada data *training*

N = jumlah *instance* pada satu *leaf*

Contoh :



Gambar 2.3 *Tree Pruning* (Sumber : COM9417, 2008)

2.4.3.3 Pembuatan Aturan-Aturan Dari Pohon Keputusan

Dari gambar 2.2, baru dapat dilakukan pembuatan aturan-aturan dari pohon keputusan. Aturan-aturan dalam bentuk *if-then* diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar sampai ke daun. Setiap simpul dan syarat pencabangnya akan diberikan di *if*, sedangkan nilai pada daun akan menjadi ditulis di *then*. Setelah semua aturan dibuat, maka aturan akan disederhanakan

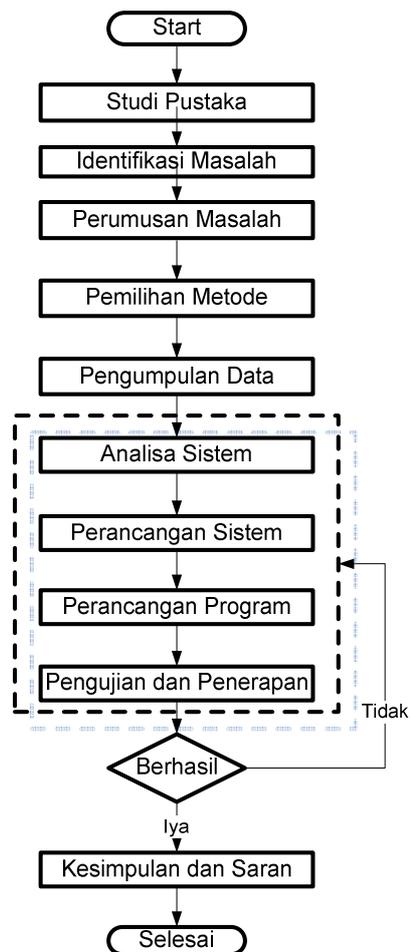
Jika aturan-aturan dari pohon tidak dibuat, maka klasifikasi kasus baru dapat dilakukan dengan menggunakan pohon keputusan.

Keakuratan dari aturan-aturan ini dinamakan kesalahan klasifikasi dan didefinisikan sebagai persentase dari kasus-kasus yang diklasifikasikan dengan salah. Penjelasan rinci tentang algoritma C4.5 dapat ditemukan pada (Quinlan, 1993) dan (Ruggieri, 2001).

BAB III

METODOLOGI PENELITIAN

Dalam metodologi penelitian dijabarkan tahap-tahap yang dilakukan dalam penelitian. Metodologi penelitian yang baik terdiri dari beberapa tahapan yang saling terkait secara sistematis. Tahapan ini diperlukan untuk memudahkan dalam melakukan penelitian. Tahapan yang dilakukan dalam penelitian ini tergambar dalam skema metodologi penelitian (Gambar 3.1). Tahapan penelitian yang dilakukan adalah sebagai berikut :



Gambar 3.1 Skema Metode Penelitian

3.1 Studi Pustaka

Studi Pustaka dilakukan untuk mendapatkan teori serta konsep yang mendukung dalam penelitian dan berkaitan dengan masalah yang akan diangkat dalam penelitian. Topik yang dipelajari dalam studi pustaka antara lain: Penyakit hiperkolesterol, definisi kolesterol, Faktor-faktor penyebab kolesterol, penalaran berbasis aturan, definisi Sistem pakar, Definisi *Decision Tree* (Pohon Keputusan) dan Algoritma C4.5. Teori serta konsep tersebut diperoleh dengan membaca buku serta referensi yang terkait sehingga memudahkan dalam menyelesaikan permasalahan yang ada.

3.2 Identifikasi Masalah

Dari Studi pendahuluan yang dilakukan, diketahui bahwa penyakit hiperkolesterol ini disebabkan oleh beberapa faktor berat badan, LDL (*Low Density Lipoprotein*), HDL (*High Density Lipoprotein*), TC (*Total Cholesterol*), TGA (*Trigeliserida*), tingkatan perokok. Dari faktor-faktor tersebut diperolehlah sebuah kesimpulan hasil diagnosa.

3.3 Perumusan Masalah

Berdasarkan identifikasi masalah yang telah dijelaskan sebelumnya, maka perlu dilakukan merancang dan membangun sistem pendignosa penyakit hiperkolestrol menggunakan Algoritma C4.5.

3.4 Pemilihan Metode

Metode yang digunakan untuk membangun sistem pendiagnosa penyakit hiperkolesterol ini adalah sistem pakar dan *Decision Tree Learning* menggunakan Algoritma C4.5. Alasan pemilihan metode *Decision Tree* ini adalah :

- a. Metode *Decision Tree* merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit
- b. Metode *Decision Tree* adalah suatu gambaran pemodelan dari suatu persoalan yang terdiri dari serangkaian keputusan yang mengarah ke solusi.
- c. Konsep Metode *Decision Tree* adalah mengubah data yang ada pada tabel keputusan menjadi sebuah pohon keputusan dan aturan-aturan keputusan (*Rule*).

Pada sistem yang akan dirancang ini secara garis besar metode *decision tree* ini akan melalui dua tahap yaitu tahap pembangunan pohon keputusan dan tahap *pruning tree*.

3.5 Pengumpulan data

Pengumpulan data dilakukan dengan cara :

- a. Pengumpulan data

Merupakan pemilihan sejumlah item tertentu dari seluruh item yang ada dengan tujuan mempelajari sebagian item tersebut untuk mewakili seluruh itemnya. Data yang dikumpulkan adalah data-data rekam medis (dapat dilihat pada Lampiran B). Rekam medis adalah berkas yang berisikan catatan dan dokumen tentang identitas pasien,

pemeriksaan, pengobatan, tindakan, dan pelayanan lain yang telah diberikan kepada pasien.

b. Wawancara

Wawancara adalah suatu percakapan langsung dengan tujuan-tujuan tertentu dengan menggunakan format tanya jawab yang terencana. Wawancara memungkinkan untuk mendengar tujuan-tujuan, perasaan, pendapat dan prosedur-prosedur informal dalam wawancara dengan petugas laboratorium, dokter di Rumah Sakit yaitu Dr. Irwan SP.JP (Dokter Spesialis Jantung) dan Dr. Yossi Eka Putri.

Wawancara juga digunakan untuk mengobservasi tempat kerja, serta untuk mengumpulkan data-data yang berhubungan dengan kelengkapan informasi.

3.6 Analisis Sistem

Analisa dilakukan atas sistem yang diterapkan untuk mendiagnosa penyakit hiperkolesterol menggunakan algoritma C4.5. Hal ini dilakukan untuk mengetahui apakah sistem tersebut sudah dapat memenuhi kebutuhan dalam hal diagnosa penyakit hiperkolesterol. Pada tahapan analisa sistem yang akan dilakukan adalah:

1. Analisa Sistem Lama

Analisa permasalahan berisi bagaimana cara memahami masalah-masalah yang ada pada sistem yang sedang berjalan.

2. Analisa Sistem Baru

Analisa dilakukan terhadap sistem baru berdasarkan data yang diperoleh dari rekam medis kemudian data-data tersebut digunakan dalam membangun sistem dengan menggunakan metode pohon keputusan.

3. Analisa Kebutuhan data

Analisa kebutuhan data berisi data-data yang dibutuhkan dalam pembuatan sistem pendiagnosa penyakit hiperkolesterol.

4. Analisa penyelesaian masalah

Analisa penyelesaian masalah berisi bagaimana menyelesaikan masalah yang ada dengan menggunakan algoritma C4.5.

5. Analisa fungsional sistem

Analisa fungsional sistem adalah suatu analisa yang menggambarkan bagaimana masukan diproses oleh aplikasi menjadi keluaran yang diharapkan oleh pengguna sistem. Analisa fungsional memuat beberapa diagram alir data yang memperlihatkan aliran data dari luar sistem yang diproses (kadang melibatkan penyimpanan sistem) kemudian menghasilkan keluaran yang berguna. Analisa fungsional sistem akan membahas tentang DFD (*Data Flow Diagram*).

6. Analisa data sistem

Analisa data sistem merupakan notasi grafik untuk objek data dan hubungannya dalam sistem. Analisa data sistem akan membahas tentang ERD (*Entity Relationship Diagram*)

3.7 Perancangan Sistem

Pada tahap perancangan sistem, dilakukan perancangan ulang dan perbaikan yang dianggap perlu setelah dilakukan analisa sistem yang ada. Tahap ini dibagi menjadi 4, yaitu:

1. Perancangan *Flowchart*

Perancangan *Flowchart* berisi tentang *flowchart* yang akan digunakan dalam sistem pendiagnosa penyakit hiperkolesterol.

2. Perancangan database

Perancangan database berisi tentang database yang akan digunakan dalam sistem pendiagnosa penyakit hiperkolesterol.

3. Perancangan struktur menu

Perancangan struktur menu digunakan untuk memudahkan pemakai sistem, diperlukan susunan daftar menu agar pengguna yang belum terbiasa dengan sistem juga dapat menggunakannya.

4. Perancangan antarmuka

Perancangan antarmuka dilakukan agar pengguna lebih mudah mengoperasikan aplikasi sistem yang akan dibuat.

3.8 Implementasi

Setelah perancangan dikerjakan, perancangan yang dibuat dituangkan kedalam bentuk program komputer dengan menggunakan bahasa pemrograman *Visual Basic* dengan *database* dari program *Microsoft Access*.

3.9 Pengujian

Setelah implementasi sistem selesai, maka selanjutnya dilakukan pengujian sistem, setelah selesai dan sistem bisa bekerja sesuai dengan tujuan.

3.10 Kesimpulan dan Saran

Tahapan akhir dari penelitian adalah penarikan kesimpulan berdasarkan hasil yang telah diperoleh dari tahap sebelumnya, serta memberikan saran-saran untuk pengguna sistem serta untuk penyempurnaan dan pengembangan penelitian ini.

BAB IV

ANALISA DAN PERANCANGAN

Pada perancangan sistem, analisa memegang peranan penting dalam membuat rincian sistem baru. Analisa perangkat lunak merupakan langkah pemahaman persoalan sebelum mengambil tindakan atau keputusan penyelesaian hasil utama, sedangkan tahap perancangan sistem adalah membuat rincian sistem hasil dari analisa menjadi bentuk perancangan agar di mengerti oleh pengguna.

Setelah mempelajari teori-teori tentang metode *Decision Tree* (Pohon Keputusan) khususnya Algoritma C4.5 dan sistem pakar pada bab sebelumnya, bab ini akan lebih difokuskan pada penjelasan mengenai analisa dan perancangan perangkat lunak yang nantinya akan diimplementasikan yang diberi nama “Sistem Pendiagnosa Penyakit Hiperkolesterol”.

Pembahasan mengenai analisa dan perancangan sistem akan dimulai dengan terlebih dahulu menganalisa sistem lama kemudian membandingkan dan melakukan analisa sistem yang akan dirancang atau dibuat.

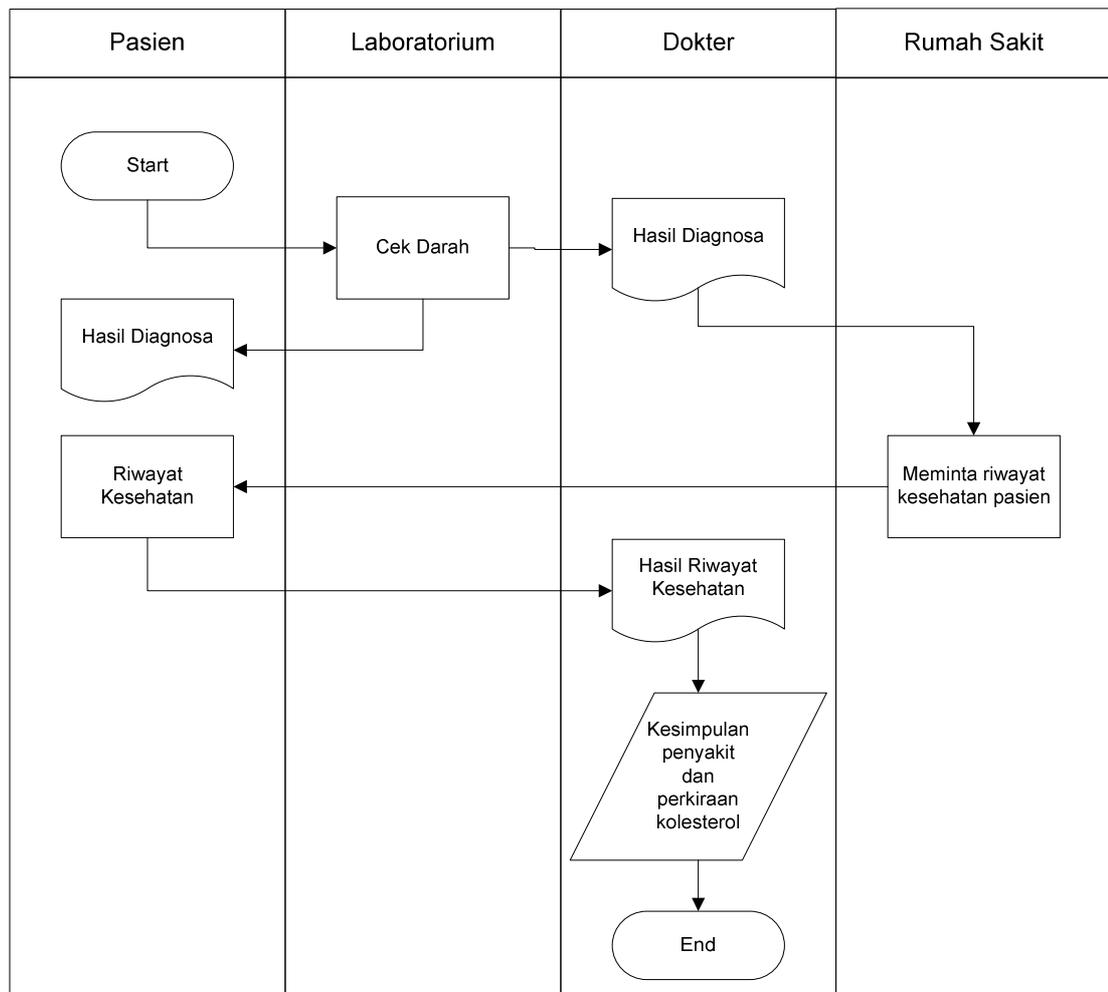
4.1 Analisa Sistem

Pada tahapan ini akan dianalisa mengenai sistem yang akan dikembangkan, kebutuhan pengguna serta menganalisa kebutuhan sistem itu sendiri.

4.1.1 Analisa Sistem Lama

Analisa sistem lama dilakukan untuk memperoleh informasi penting dan menjadi sumber daya bagi sistem yang akan dikembangkan agar mampu

mengatasi kelemahan-kelemahan terhadap sistem lama. Berdasarkan wawancara dan survey yang dilakukan, analisa yang dilakukan dapat dilihat dalam *flowchart* berikut ini:



Gambar 4.1 *Flowchart* Analisa Sistem Lama

Keterangan:

- Pasien melakukan cek darah ke laboratorium (Lab)
- Hasil Lab tersebut berisi tentang hasil diagnosa dari pasien
- Kemudian dokter melakukan pembacaan terhadap hasil Lab. Selain itu pasien juga ditanyakan tentang riwayat kesehatan terdahulu

- Dari kedua hasil dokter menyimpulkan jenis dari penyakit pasien dan memberikan alternatif untuk perbaikan kolesterol pasien

Pasien berjumpa langsung dengan dokter, akan lebih menguntungkan, karena dokter lebih mudah dalam mengidentifikasi gejala penyakitnya dan bertanya langsung kepada pasien, tetapi berkonsultasi dengan dokter membutuhkan waktu dan biaya tidak sedikit, sehingga ini menjadi permasalahan utama yang terjadi.

4.1.2 Analisa Sistem Baru

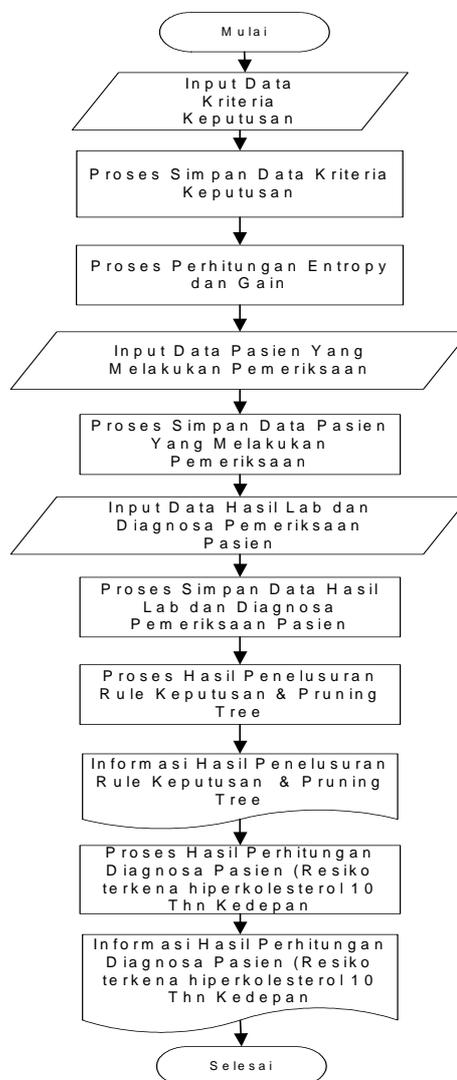
Analisa pada sistem baru meliputi analisa kebutuhan data, analisa permasalahan, analisa penyelesaian masalah, analisa fungsional sistem dan analisa data sistem. Sistem baru yang akan dibuat ini memanfaatkan sistem pakar dalam menemukan kesimpulan permasalahan serta solusi pencegahannya. Karena sistem pakar dapat menyelesaikan masalah tertentu seperti yang dilakukan oleh para ahli.

Sistem baru yang akan dibangun berdasarkan pengembangan dari sistem yang sudah ada yaitu :

1. Petugas Lab melakukan input terhadap data kriteria keputusan dan melakukan proses penyimpanan data kriteria.
2. Perawat melakukan penelusuran terhadap data-data kriteria pasien pada pohon keputusan yang telah dibangun menggunakan metode pohon keputusan (*Decision Tree*).
3. Petugas Lab melakukan Input data-data hasil Lab pasien dan diagnosa pasien, serta melakukan penyimpanan data-data tersebut.
4. Perawat melakukan proses rule keputusan dan proses pruning.

5. Kemudian Perawat melakukan proses perhitungan diagnosa pasien tentang resiko 10 tahun kemudian.
6. Perawat memperoleh laporan dari informasi hasil penelusuran rule keputusan dan proses pruning, serta informasi diagnosa pasien tentang resiko 10 tahun kemudian.

Sistem melakukan proses untuk pengalambihan keputusan menggunakan analisa resiko.



gambar 4.2 *Flowchart* Analisa Sistem Baru

4.1.3 Analisa Kebutuhan Data

Beberapa data yang dibutuhkan untuk memulai pembuatan sistem adalah sebagai berikut:

1. Data Pengguna, yaitu data-data siapa saja yang akan menggunakan sistem yang akan dibangun.
2. Data Pasien, data ini berupa masukan oleh pasien yang akan didiagnosa oleh sistem.
3. Data Kriteria

- a. Berat badan

Kriteria berat badan diperoleh dari hasil perhitungan Indeks Masa Tubuh (IMT) pasien. Adapun standar penilaiannya adalah (Irwan, 2010):

- Kurus tingkat berat < 17,0
- Kurus tingkat rendah 17,0 – 18,5
- Normal > 18,5 – 25,0
- Gemuk tingkat rendah >25,0 – 27,0
- Gemuk tingkat tinggi > 27,0

- b. LDL (*Low Density Lipoprotein*)

Tingkatan Kadar LDL dalam darah adalah(Irwan, 2010):

- Normal <100 mg/dL
- Sedang 100 – 159 mg/dL
- Tinggi > 160mg/dL

- c. HDL (*High Density Lipoprotein*)

Tingkatan Kadar HDL dalam darah adalah (Irwan, 2010):

- Rendah 0 – 34 mg/dL
- Tinggi 35 – 55 mg/dL

d. TC (*Total Cholesterol*)

Tingkatan kadar TC dalam darah adalah (Irwan, 2010):

- Normal 0 – 200 mg/dL
- Tinggi > 200 mg/dL

e. TGA (*Trigliserida*)

Tingkatan kadar TGA dalam darah adalah (Irwan, 2010):

- Normal 100 – 150 mg/dL
- Sedang 151 – 190 mg/dL
- Tinggi 191 – 200 mg/dL

f. Tingkatan Perokok

Tingkatan Perokok berdasarkan skor adalah (Irwan, 2010):

- Tidak merokok 1
- Tidak merokok tetapi hidup dan bekerja dengan perokok 2
- Berhenti merokok selama 3 tahun terakhir 3
- Merokok 4
- Merokok dan hidup atau bekerja dengan perokok 5

4.1.4 Analisa Penyelesaian Masalah

Penyelesaian masalah yang digunakan dalam analisa sistem pendiagnosa penyakit hiperkolesterol ini adalah model pohon keputusan dan untuk membangun pohon keputusan tersebut dengan menggunakan algoritma C45.

Pohon keputusan adalah salah satu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit (Suyanto,2007). Adapun yang menjadi konsep pohon keputusan adalah mengubah data yang ada pada tabel keputusan menjadi pohon keputusan, serta dari pohon keputusan akan diubah kedalam bentuk aturan-aturan keputusan (*Rule*) (Achmad,2003).

Tahapan penyelesaian masalah sistem pendiagnosa penyakit hiperkolesterol adalah urutan langkah-langkah yang harus dikerjakan dalam melakukan penyelesaian diagnose penyakit hiperkolesterol. Secara garis besar proses-proses yang dilakukan dalam perangkat lunak ini dimulai dengan pembuatan pohon keputusan. Selanjutnya adalah pemangkasan pohon keputusan. Setelah itu, pembuatan aturan-aturan dari pohon keputusan.

Dalam pembuatan sistem ini, contoh kasus yang diambil adalah pada salah satu rumah sakit yang dihadapkan pada permasalahan pemeriksaan pasien pada rumah sakit tersebut.

Tabel 4.1 Data Menentukan Pasien Mengalami Hiperkolestrol Atau Tidak

No	Berat Badan	LDL	HDL	TC	TGA	Tingkatan Perokok	Hasil Analisa
1	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 1	Tidak
2	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 2	Tidak
3	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 3	Tidak
4	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 4	Tidak

5	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 5	Tidak
6	Kurus Berat	Normal	Rendah	Normal	Sedang	Tingkat 1	Tidak
7	Kurus Berat	Normal	Rendah	Normal	Sedang	Tingkat 2	Tidak
8	Kurus Berat	Normal	Rendah	Normal	Sedang	Tingkat 3	Tidak
9	Kurus Berat	Normal	Rendah	Normal	Sedang	Tingkat 4	Tidak
10	Kurus Berat	Normal	Rendah	Normal	Sedang	Tingkat 5	Tidak
11	Kurus Berat	Normal	Rendah	Normal	Tinggi	Tingkat 1	Tidak
12	Kurus Berat	Normal	Rendah	Normal	Tinggi	Tingkat 2	Tidak
13	Kurus Berat	Normal	Rendah	Normal	Tinggi	Tingkat 3	Tidak
14	Kurus Berat	Normal	Rendah	Normal	Tinggi	Tingkat 4	Hiperkilomikronemia
15	Kurus Berat	Normal	Rendah	Normal	Tinggi	Tingkat 5	Hiperkilomikronemia
16	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 1	Tidak
17	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 2	Tidak
18	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 3	Tidak
19	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 4	Tidak
20	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 5	Hiperkolesterol

Lanjutan tabel berikutnya dapat dilihat di lampiran B

4.1.4.1 Langkah-Langkah Membangun Pohon Keputusan

Adapun langkah-langkah dalam membangun pohon keputusan adalah sebagai berikut, (Achmad,2003):

4.1.4.1.1 Data Dari Tabel Keputusan Menjadi Model Pohon Keputusan

Untuk mengubah data dari tabel keputusan menjadi pohon keputusan dengan menggunakan algoritma C4.5. Adapun cara kerja algoritma C4.5 yaitu sebagai berikut:

Rekusi Level 0 Iterasi 1

Memanggil fungsi c4.5 dengan kumpulan sampel berupa semua sampel data = [488,412], atribut target = 'hiperkolesterol, hiperkilimikronemia, hiperlipoproteinemia tipe III, IV dan V' dan kumpulan atribut = { IMT, LDL, HDL, TC, TGA, Tingkatan Perokok}. Untuk menentukan atribut yang merupakan

the best classifier dan diletakkan sebagai *Root* perlu menghitung *information gain* untuk semua atribut tersebut.

Information Gain semua Atribut:

Jumlah data semuanya = 900

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 488

Tidak = 412

$$\begin{aligned}
 \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
 &= -(488/900) \log_2 (488/900) - (412/900) \log_2 (412/900) \\
 &= -(488/900) \log (488/900) / \log 2 - (412/900) \log (412/900) / \log 2 \\
 &= 0,99485
 \end{aligned}$$

a. Atribut IMT

Jumlah data IMT untuk S_{kurus} tingkat tinggi = 180

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 74

Tidak = 106

$$\begin{aligned}
 \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
 &= -(74/180) \log_2 (74/180) - (106/180) \log_2 (106/180) \\
 &= -(74/180) \log (74/180) / \log 2 - (106/180) \log (106/180) / \log 2 \\
 &= 0,97708
 \end{aligned}$$

Jumlah data IMT untuk S_{kurus} tingkat rendah = 180

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 74

Tidak = 106

$$\text{Entropy (S)} = -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})}$$

$$\begin{aligned}
&= -(74/180) \log_2 (74/180) - (106/180) \log_2 (106/180) \\
&= -(74/180) \log (74/180) / \log 2 - (106/180) \log (106/180) / \log 2 \\
&= 0,97708
\end{aligned}$$

Jumlah data IMT untuk $S_{\text{normal}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 74

Tidak = 106

$$\begin{aligned}
\text{Entropy (S)} &= -\mathbf{P}_{(\text{Yes})} \log_2 \mathbf{P}_{(\text{Yes})} - \mathbf{P}_{(\text{No})} \log_2 \mathbf{P}_{(\text{No})} \\
&= -(74/180) \log_2 (74/180) - (106/180) \log_2 (106/180) \\
&= -(74/180) \log (74/180) / \log 2 - (106/180) \log (106/180) / \log 2 \\
&= 0,97708
\end{aligned}$$

Jumlah data IMT untuk $S_{\text{gemuk tingkat rendah}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 126

Tidak = 54

$$\begin{aligned}
\text{Entropy (S)} &= -\mathbf{P}_{(\text{Yes})} \log_2 \mathbf{P}_{(\text{Yes})} - \mathbf{P}_{(\text{No})} \log_2 \mathbf{P}_{(\text{No})} \\
&= -(126/180) \log_2 (126/180) - (54/180) \log_2 (54/180) \\
&= -(126/180) \log (126/180) / \log 2 - (54/180) \log (54/180) / \log 2 \\
&= 0,881291
\end{aligned}$$

Jumlah data IMT untuk $S_{\text{gemuk tingkat tinggi}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 140

Tidak = 40

$$\begin{aligned}
\text{Entropy (S)} &= -\mathbf{P}_{(\text{Yes})} \log_2 \mathbf{P}_{(\text{Yes})} - \mathbf{P}_{(\text{No})} \log_2 \mathbf{P}_{(\text{No})} \\
&= -(140/180) \log_2 (140/180) - (40/180) \log_2 (40/180) \\
&= -(140/180) \log (140/180) / \log 2 - (40/180) \log (40/180) / \log 2
\end{aligned}$$

$$= 0,76420$$

$$\text{Gain (S,IMT)} = \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v))$$

$$= 0,99485 - (180/900) \times 0,97708 - (180/900) \times 0,97708 - (180/900) \times 0,97708 - (180/900) \times 0,60373 - (180/900) \times 0,76420$$

$$= \mathbf{0,079503}$$

b. Atribut LDL (*Low Density Lipoprotein*)

Jumlah data LDL untuk $S_{\text{normal}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 124

Tidak = 176

$$\text{Entropy (S)} = -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})}$$

$$= -(124/300) \log_2 (124/300) - (176/300) \log_2 (176/300)$$

$$= -(124/300) \log (124/300) / \log 2 - (176/300) \log (176/300) / \log 2$$

$$= 0,97822$$

Jumlah data LDL untuk $S_{\text{sedang}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 128

Tidak = 172

$$\text{Entropy (S)} = -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})}$$

$$= -(128/300) \log_2 (128/300) - (172/300) \log_2 (172/300)$$

$$= -(128/300) \log (128/300) / \log 2 - (172/300) \log (172/300) / \log 2$$

$$= 0,98443$$

Jumlah data LDL untuk $S_{\text{tinggi}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 236

Tidak = 64

$$\begin{aligned}
\text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
&= -(236/300) \log_2 (236/300) - (64/300) \log_2 (64/300) \\
&= -(236/300) \log (236/300) / \log 2 - (64/300) \log (64/300) / \log 2 \\
&= 0,74781
\end{aligned}$$

$$\begin{aligned}
\text{Gain (S,LDL)} &= \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v)) \\
&= 0,99485 - (300/900) \times 0,97822 - (300/900) \times 0,98443 - (300/900) \\
&\quad \times 0,74781 \\
&= \mathbf{0,091362}
\end{aligned}$$

c. Atribut HDL (*High Density Lipoprotein*)

Jumlah data HDL untuk $S_{\text{rendah}} = 450$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 241

Tidak = 209

$$\begin{aligned}
\text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
&= -(241/450) \log_2 (241/450) - (209/450) \log_2 (209/450) \\
&= -(241/450) \log(241/450) / \log 2 - (209/450) \log (209/450) / \log 2 \\
&= 0,99635
\end{aligned}$$

Jumlah data HDL untuk $S_{\text{tinggi}} = 450$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 247

Tidak = 203

$$\begin{aligned}
\text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
&= -(247/450) \log_2 (247/450) - (203/450) \log_2 (203/450) \\
&= -(247/450) \log(247/450) / \log 2 - (203/450) \log (203/450) / \log 2 \\
&= 0,99309
\end{aligned}$$

$$\begin{aligned}
\text{Gain (S,HDL)} &= \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v)) \\
&= 0,99485 - (450/900) \times 0,99635 - (450/900) \times 0,99309 \\
&= \mathbf{0,00013}
\end{aligned}$$

d. Atribut TC (*Total Cholesterol*)

Jumlah data TC untuk $S_{\text{normal}} = 450$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 150

Tidak = 300

$$\begin{aligned}
\text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
&= -(150/450) \log_2 (150/450) - (300/450) \log_2 (300/450) \\
&= -(150/450) \log(150/450) / \log 2 - (300/450) \log (300/450) / \log 2 \\
&= 0,918295
\end{aligned}$$

Jumlah data TC untuk $S_{\text{tinggi}} = 450$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V = 338

Tidak = 112

$$\begin{aligned}
\text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\
&= -(338/450) \log_2 (338/450) - (112/450) \log_2 (112/450) \\
&= -(338/450) \log(338/450) / \log 2 - (112/450) \log (112/450) / \log 2 \\
&= 0,80951
\end{aligned}$$

$$\begin{aligned}
\text{Gain (S,TC)} &= \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v)) \\
&= 0,99485 - (450/900) \times 0,918245 - (450/900) \times 0,80851 \\
&= \mathbf{0,130948}
\end{aligned}$$

e. Atribut TGA (*Trigliserida*)

Jumlah data TGA untuk $S_{\text{normal}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 138

Tidak = 162

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(138/300) \log_2 (138/300) - (162/300) \log_2 (162/300) \\ &= -(138/300) \log(138/300) / \log 2 - (162/300) \log (162/300) / \log 2 \\ &= 0,995378 \end{aligned}$$

Jumlah data TGA untuk $S_{\text{sedang}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 140

Tidak = 160

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(140/300) \log_2 (140/300) - (160/300) \log_2 (160/300) \\ &= -(140/300) \log(140/300) / \log 2 - (160/300) \log (160/300) / \log 2 \\ &= 0,99679 \end{aligned}$$

Jumlah data TGA untuk $S_{\text{tinggi}} = 300$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 210

Tidak = 90

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(210/300) \log_2 (210/300) - (90/300) \log_2 (90/300) \\ &= -(210/300) \log (210/300) / \log 2 - (90/300) \log (90/300) / \log 2 \\ &= 0,88129 \end{aligned}$$

$$\begin{aligned} \text{Gain (S,TGA)} &= \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v)) \\ &= 0,99485 - (300/900) \times 0,995378 - (300/900) \times 0,99679 - \\ &\quad (300/900) \times 0,88129 \end{aligned}$$

$$= 0,03703$$

f. Atribut Tingkatan Perokok

Jumlah data tingkatan perokok untuk $S_{\text{tingkat 1}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 69

Tidak = 111

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(69/180) \log_2 (69/180) - (111/180) \log_2 (111/180) \\ &= -(69/180) \log (69/180) / \log 2 - (111/180) \log (111/180) / \log 2 \\ &= 0,96036 \end{aligned}$$

Jumlah data tingkatan perokok untuk $S_{\text{tingkat 2}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 68

Tidak = 112

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(68/180) \log_2 (68/180) - (112/180) \log_2 (112/180) \\ &= -(68/180) \log (68/180) / \log 2 - (112/180) \log (112/180) / \log 2 \\ &= 0,95646 \end{aligned}$$

Jumlah data tingkatan perokok untuk $S_{\text{tingkat 3}} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 72

Tidak = 108

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(72/180) \log_2 (72/180) - (108/180) \log_2 (108/180) \\ &= -(72/180) \log (72/180) / \log 2 - (108/180) \log (108/180) / \log 2 \\ &= 0,97095 \end{aligned}$$

Jumlah data tingkatan perokok untuk $S_{\text{tingkat } 4} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 139

Tidak = 41

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(139/180) \log_2 (139/180) - (41/180) \log_2 (41/180) \\ &= -(139/180) \log (139/180) / \log 2 - (41/180) \log (41/180) / \log 2 \\ &= 0,77412 \end{aligned}$$

Jumlah data tingkatan perokok untuk $S_{\text{tingkat } 5} = 180$

Hiperkolesterol, Hiperkilomikronemia, Hiperlipoproteinemia III, IV dan V= 140

Tidak = 40

$$\begin{aligned} \text{Entropy (S)} &= -P_{(\text{Yes})} \log_2 P_{(\text{Yes})} - P_{(\text{No})} \log_2 P_{(\text{No})} \\ &= -(140/180) \log_2 (140/180) - (40/180) \log_2 (40/180) \\ &= -(140/180) \log (140/180) / \log 2 - (40/180) \log (40/180) / \log 2 \\ &= 0,76420 \end{aligned}$$

$$\begin{aligned} \text{Gain (S,TGA)} &= \text{Entropy (S)} - \sum ((|S_v| / |S|) \times \text{Entropy (S}_v)) \\ &= 0,99485 - (180/900) \times 0,96036 - (180/900) \times 0,95646 - \\ &\quad (180/900) \times 0,97095 - (180/900) \times 0,77412 - (180/900) \times 0,76420 \\ &= \mathbf{0,109632} \end{aligned}$$

Perhitungan berikutnya dapat dilihat dilampiran D

4.1.4.1.2 Data Pohon Keputusan dilakukan Perhitungan *Pruning Tree*

Pruning tree adalah melakukan suatu kegiatan untuk mengganti suatu *subtree* dengan suatu *leaf*. Penggantian dilakukan jika *error rate* pada *subtree* lebih besar jika dibandingkan dengan *single leaf* (kumar, 2006).

$$e(N, f, z) = \frac{f + \frac{z^2}{2 \cdot N} + z \sqrt{\frac{f(1-f)}{N} + \frac{z^2}{4 \cdot N^2}}}{1 + \frac{z^2}{N}}$$

Pruning pada cabang Tingkatan Perokok untuk TC = “Normal”, TGA = “Normal”, LDL=”Tinggi”, Tingkatan Perokok=”Tingkat 1,2 dan 3”

$$e\left(10, \frac{8}{10}, 0.69\right) = \frac{\frac{8}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{8}{10}\left(1 - \frac{8}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,87$$

$$e\left(10, \frac{8}{10}, 0.69\right) = \frac{\frac{8}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{8}{10}\left(1 - \frac{8}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,87$$

$$e\left(10, \frac{7}{10}, 0.69\right) = \frac{\frac{7}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{7}{10}\left(1 - \frac{7}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,78$$

Digabungkan dengan rasio 10:10:10 menghasilkan e=0,84

Jadi untuk $e < 0,84$ harus di pruning sehingga hasilnya menjadi daun yaitu “Hiperkolesterol”.

Pruning pada cabang Tingkatan Perokok untuk TC = “Normal”, TGA = “Sedang”, LDL=”Tinggi”, Tingkatan Perokok=”Tingkat 1, 2 dan 3”

$$e\left(10, \frac{8}{10}, 0.69\right) = \frac{\frac{8}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{8}{10}\left(1 - \frac{8}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,87$$

$$e\left(10, \frac{8}{10}, 0.69\right) = \frac{\frac{8}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{8}{10}\left(1 - \frac{8}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,87$$

$$e\left(10, \frac{7}{10}, 0.69\right) = \frac{\frac{7}{10} + \frac{0.69^2}{2 \cdot 10} + 0.69 \sqrt{\frac{\frac{7}{10}\left(1 - \frac{7}{10}\right)}{10} + \frac{0.69^2}{4 \cdot 10^2}}}{1 + \frac{0.69^2}{10}} = 0,78$$

Digabungkan dengan rasio 10:10:10 menghasilkan $e=0,84$

Jadi untuk $e < 0,84$ harus di pruning sehingga hasilnya menjadi daun yaitu “Hiperkolesterol”.

Pruning pada cabang Tingkatan Perokok untuk TC = “Tinggi”, Tingkatan Perokok = “Tingkat 1”, LDL=“Normal,Sedang”

$$e\left(30, \frac{19}{30}, 0.69\right) = \frac{\frac{19}{30} + \frac{0.69^2}{2 \cdot 30} + 0.69 \sqrt{\frac{\frac{19}{30}\left(1 - \frac{19}{30}\right)}{30} + \frac{0.69^2}{4 \cdot 30^2}}}{1 + \frac{0.69^2}{30}} = 0,68$$

$$e\left(30, \frac{18}{30}, 0.69\right) = \frac{\frac{18}{30} + \frac{0.69^2}{2 \cdot 30} + 0.69 \sqrt{\frac{\frac{18}{30}\left(1 - \frac{18}{30}\right)}{30} + \frac{0.69^2}{4 \cdot 30^2}}}{1 + \frac{0.69^2}{30}} = 0,65$$

Digabungkan dengan rasio 30:30 menghasilkan $e=0,67$

Jadi untuk $e < 0,67$ harus di pruning sehingga hasilnya menjadi daun yaitu “Hiperkolesterol”.

Pruning pada cabang Tingkatan Perokok untuk TC = “Tinggi”, Tingkatan Perokok = “Tingkat 2”, LDL=“Normal,Sedang”

$$e\left(30, \frac{20}{30}, 0.69\right) = \frac{\frac{20}{30} + \frac{0.69^2}{2 \cdot 30} + 0.69 \sqrt{\frac{\frac{20}{30}\left(1 - \frac{20}{30}\right)}{30} + \frac{0.69^2}{4 \cdot 30^2}}}{1 + \frac{0.69^2}{30}} = 0,72$$

$$e\left(30, \frac{18}{30}, 0.69\right) = \frac{\frac{18}{30} + \frac{0.69^2}{2 \cdot 30} + 0.69 \sqrt{\frac{\frac{18}{30}\left(1 - \frac{18}{30}\right)}{30} + \frac{0.69^2}{4 \cdot 30^2}}}{1 + \frac{0.69^2}{30}} = 0,65$$

Digabungkan dengan rasio 30:30 menghasilkan $e=0,68$

Jadi untuk $e < 0,68$ harus di pruning sehingga hasilnya menjadi daun yaitu “Hiperkolesterol”.

4.1.4.1.3 Pembuatan Aturan Pohon Keputusan dari Pohon Keputusan

Pada tahap ini akan dilakukan perubahan dari pohon keputusan menjadi kedalam bentuk *Rule – rule*. Berikut hasil *rule-rule* yang diubah dari pohon keputusan diatas.

R1 if (total cholesterol = “normal”)[^](TGA= “normal”) ^ (LDL= “normal”) then
keterangan= “tidak”

R2 if (total cholesterol = “normal”)[^](TGA= “normal”) ^ (LDL= “sedang”) then
keterangan= “tidak”

R3 if (total cholesterol = “normal”)[^](TGA= “normal”) ^ (LDL= “tinggi”)[^](tingkatan perokok=”tingkat 1”)[^](IMT=” Kurus tk. berat”) then
keterangan=”tidak”

R4 if (total cholesterol = “normal”)[^](TGA= “normal”) ^ (LDL= “tinggi”)[^](tingkatan perokok=”tingkat 1”)[^](IMT=” Kurus tk. rendah”) then
keterangan=”tidak”

- R5 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 1")^(IMT="normal") then keterangan="tidak"
- R6 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 1")^(IMT=" Gemuk tk. rendah") then keterangan="hiperkolesterol"
- R7 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 1")^(IMT=" Gemuk tk. berat") then keterangan="hiperkolesterol"
- R8 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 2")^(IMT=" Kurus tk. berat") then keterangan="tidak"
- R9 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 2")^(IMT=" Kurus tk. rendah") then keterangan="tidak"
- R10 if (total cholesterol = "normal")^(TGA= "normal") ^ (LDL= "tinggi")^(tingkatan perokok="tingkat 2")^(IMT=" Normal") then keterangan="tidak"

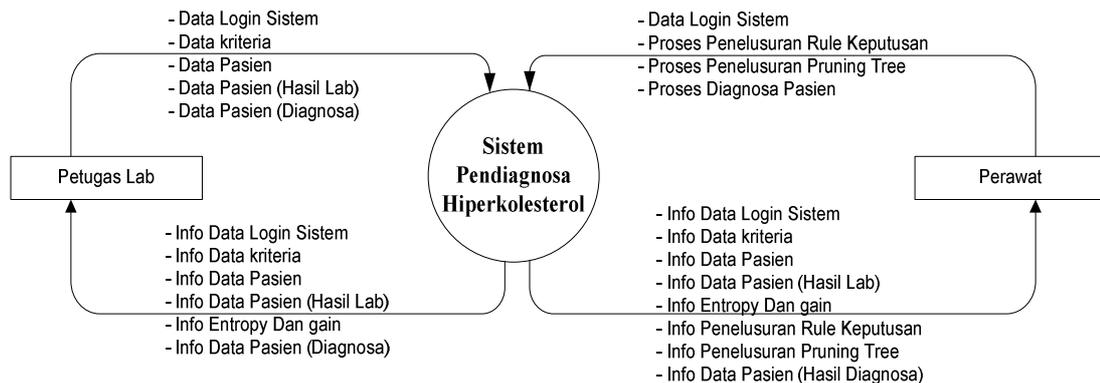
Rule-rule selanjutnya dapat dilihat pada Lampiran E.

4.1.5 Analisa Fungsional Sistem

Analisa Fungsional data terdiri dari Diagram Konteks (*Context Diagram*), DFD (*Data Flow Diagram*), ERD (*Entity Relationship Diagram*) dan kamus data. Masing-masing subsistem tersebut dapat dijelaskan sebagai berikut:

4.1.5.1 Diagram Konteks (*Context Diagram*)

Context Diagram merupakan gambaran umum dari sistem yang akan dibangun. Sistem ini memiliki dua buah entitas yaitu Perawat dan Petugas Lab



Gambar 4.5 *Context Diagram*

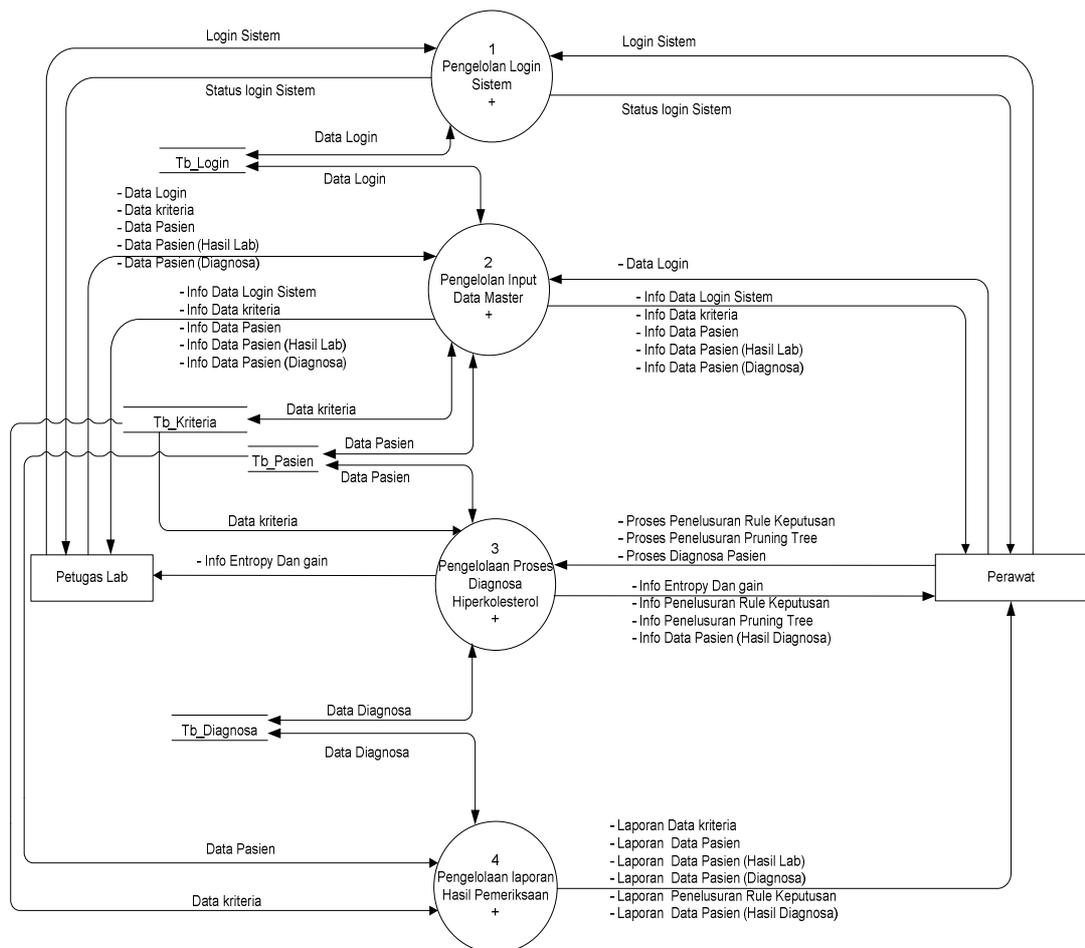
Entitas luar yang berinteraksi dengan sistem adalah:

1. Petugas Lab, yang memiliki peran antara lain:
 - a. Melakukan login sistem dan memasukkan data login Petugas Lab kedalam sistem
 - b. Memasukkan data kriteria
 - c. Memasukkan data pasien
 - d. Memasukan data hasil lab pasien
 - e. Memasukan data diagnosa pasien.
 - f. Melakukan proses perhitungan data Entropy dan gain, penelusuran rule keputusan, perhitungan dan perangkaian nilai Deference
2. Perawat, yang memiliki peran antara lain:
 - a. Melakukan login sistem dan memasukkan data login perawat kedalam sistem

- b. Melakukan proses perhitungan data penelusuran rule keputusan
- c. Melakukan proses perhitungan data penelusuran pruning tree
- d. Melakukan proses perhitungan diagnosa pasien

4.1.5.2 Data Flow Diagram (DFD)

Data flow diagram (DFD) sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa mempertimbangkan lingkungan fisik dimana data tersebut mengalir, atau lingkungan fisik dimana data tersebut tersimpan.



Gambar 4.6 Data Flow Diagram (DFD) Level 1 Sistem Pendiagnosa Penyakit

Hiperkolesterol

Dipecah menjadi 4 (empat) buah proses dan beberapa buah aliran data.

Untuk keterangan masing-masing dapat dilihat kamus data pada tabel berikut ini.

Tabel 4.2 Keterangan proses pada DFD level 1

No	Nama proses	Masukan	Keluaran	Deskripsi
1	Pengelolaan Login Sistem	- Login Sistem	- Status login Sistem	Proses untuk login kedalam sistem bagi pengguna
2	Pengelolaan Input Data Master	- Data Login - Data kriteria - Data Pasien - Data Pasien (Hasil Lab) - Data Pasien (Diagnosa)	- Info Data Login - Info Data kriteria - Info Data Pasien - Info Data Pasien (Hasil Lab) - Info Data Pasien (Diagnosa)	Proses untuk melakukan entri data utama sistem
3	Pengelolaan Proses Diagnosa Hiper kolesterol	- Proses Penelusuran Rule Keputusan - Proses Penelusuran Pruning Tree - Proses Diagnosa Pasien	- Info Entropy Dan gain - Info Proses Penelusuran Rule Keputusan - Info Proses Penelusuran Pruning Tree - Info Proses Diagnosa Pasien	Proses untuk melakukan proses Diagnosa hiperkolesterol
4	Pengelolaan Laporan Hasil Pemeriksaan	-	- Laporan Data kriteria - Laporan Data Pasien - Laporan Data Pasien (Hasil Lab) - Laporan Data Pasien (Diagnosa) - Laporan Proses Penelusuran Pruning Tree - Laporan Pasien (Hasil Diagnosa)	Proses pelaporan Laporan Hasil Pemeriksaan

Tabel 4.3 Keterangan Aliran data pada DFD level 1

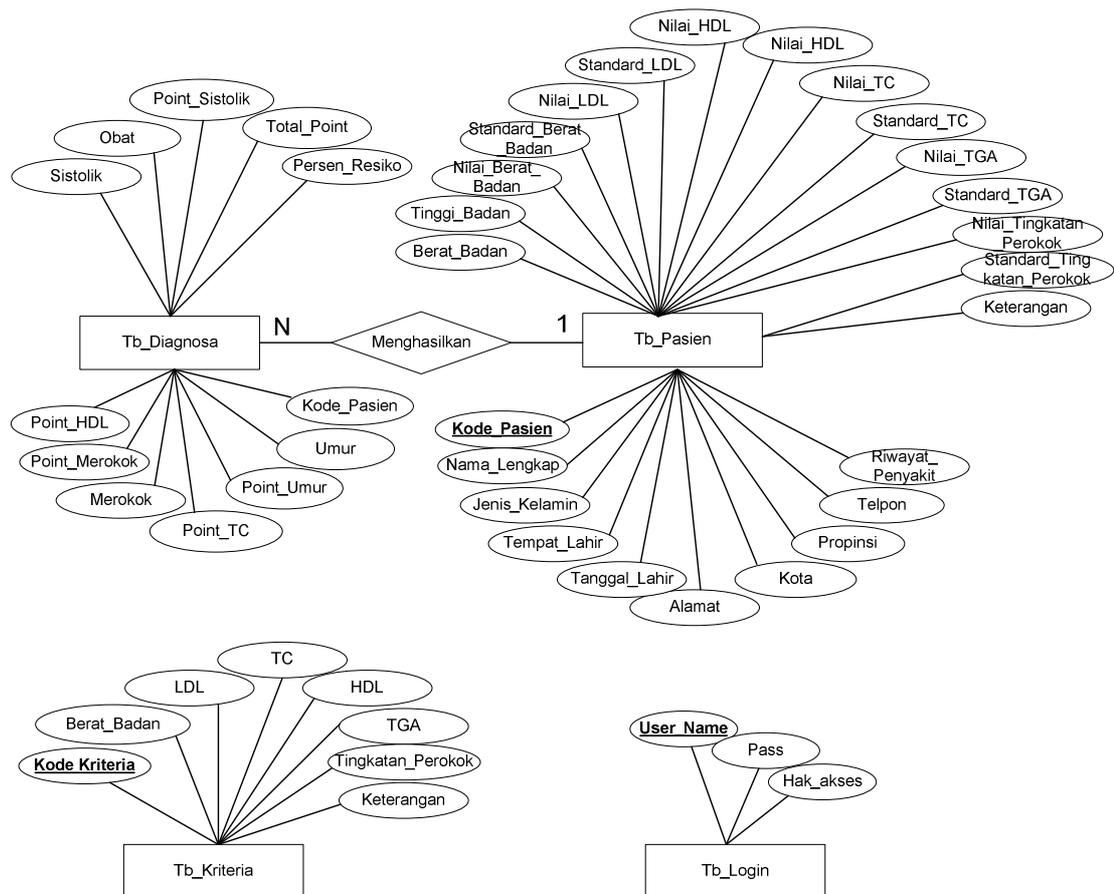
No	Nama	Deskripsi
1	Data Login	Log in nama dan password
2	Data Pasien	Proses Data pasien, Data Nilai kriteria pasien, data nilai hasil perhitungan penelusuran
3	Data Kriteria	Proses Data Kriteria, data keterangan keputusan
4	Data Diagnosa	Proses memasukkan data diagnosa pasien.

Untuk DFD yang selanjutnya dapat dilihat pada lampiran A.

4.1.6 Analisa data sistem

Analisa data sistem merupakan notasi grafik untuk objek data dan hubungannya dalam sistem. Adapun cakupan dari analisa data sistem dapat dilihat pada *Entity Relationship Diagram (ERD)*

Pada model data aplikasi ini, komposisi masing-masing objek data dan atribut yang menggambarkan objek tersebut serta hubungan antara masing-masing objek data dan objek lainnya dapat dilihat di *Entity Relationship Diagram (ERD)*. Adapun ERD dari aplikasi ini adalah pada gambar sebagai berikut:



Gambar 4.7 *Entity Relationship Diagram (ERD)*

Tabel 4.4 Keterangan entitas pada ERD

No	Nama	Deskripsi	Atribut	Primary key
1.	Tb_Pasien	Berisi data alternatif calon karyawan yang akan melakukan seleksi program rekrut	<ul style="list-style-type: none"> - Kode_Pasien - Nama_Lengkap - Jenis_Kelamin - Tempat_Lahir - Tanggal_Lahir - Alamat - Kota - Propinsi - Telpon - Riwayat_Penyakit - Berat_Badan - Tinggi_Badan - Nilai_Berat_Badan - Standard_Berat_Badan - Nilai_LDL - Standard_LDL - Nilai_HDL - Standard_HDL - Nilai_TC - Standard_TC - Nilai_TGA - Standard_TGA - Nilai_Tingkatan_Perokok - Standard_Tingkatan_Perokok - Keterangan 	Kode_Pasien
2.	Tb_Kriteria	Berisi data keputusan dari perpasangan criteria	<ul style="list-style-type: none"> - Kode_Kriteria * - Berat_Badan - LDL - HDL - TC - TGA - Tingkatan_Perokok - Keterangan 	Kode_Kriteria
3.	Tb_Diagnosa	Berisi data jabatan yang akan diisi oleh calon karyawan	<ul style="list-style-type: none"> - Kode_Pasien* - Umur - Point_Umur - Point_TC - Merokok - Point_Merokok - Point_HDL - Sistolik - Obat - Point_Sistolik - Total_Point 	Kode_Pasien

			- Persen_Resiko	
4.	Tb_Login	Berisi data User Login sistem	- User_Name* - Pass - Hak_Akses	Username

Tabel 4.5 Keterangan hubungan pada ERD

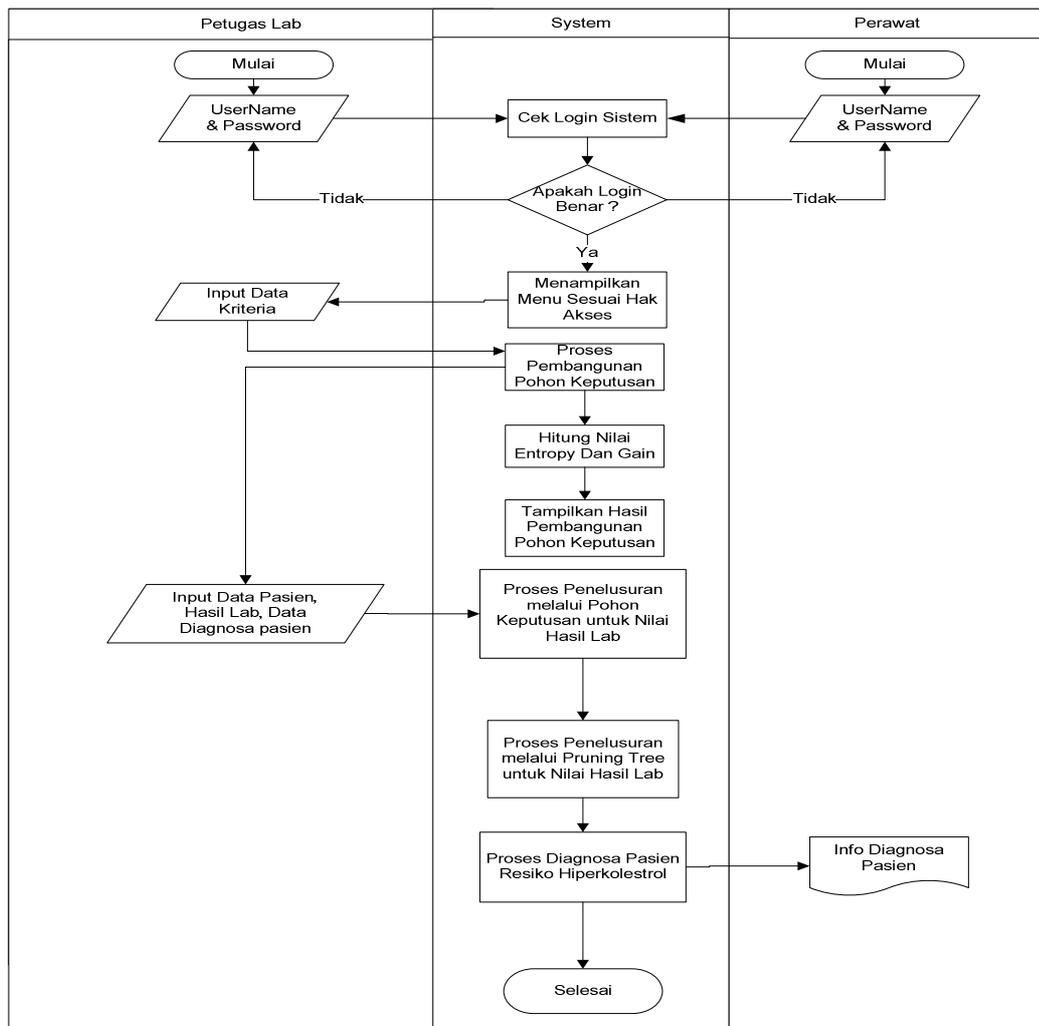
No	Nama	Deskripsi
1.	Menghasilkan	Hubungan entitas Tb_Pasien dengan Entitas Tb_Diagnosa

4.2 Perancangan Sistem

Setelah melakukan analisa terhadap sistem yang akan dikembangkan, maka langkah berikutnya adalah merancang sistem berdasarkan hasil analisa yang telah dilakukan sebelumnya.

4.2.1 Perancangan *Flowchat*

Perancangan *flowchart* digunakan untuk memudahkan dalam implementasi sistem.



Gambar 4.8 *Flowchart* Sistem

4.2.2 Perancangan Database

Deskripsi tabel yang dianalisa pada basis data berdasarkan ERD yang telah dibuat diatas adalah sebagai berikut:

1. Tabel Pasien

Nama : Tb_Pasien

Deskripsi isi : Berisi data pasien yang akan melakukan pemeriksaan hipperkolesterol

Primary key : Kode_Calon_Karyawan

Foreign key : -

Tabel 4.6 Tabel Pasien

Nama Field	Type dan Length	Deskripsi	Boleh Null	Default
Kode_Pasien	Number	Identifier Kode Pasien	No	AutoInc
Nama_Lengkap	Text,200	Nama Lengkap	No	-
Jenis_Kelamin	Text,20	Jenis Kelamin	No	-
Tempat_Lahir	Text,50	Tempat Lahir	No	-
Tanggal_Lahir	Date/Time	Tanggal Lahir	No	Date()
Alamat	Text,100	Alamat	Yes	-
Kota	Text,100	Kota	Yes	-
Propinsi	Text,100	Propinsi	Yes	-
Telpon	Text,25	Telpon	Yes	-
Riwayat_Penyakit	Text,255	Riwayat Penyakit	Yes	-
Berat_Badan	Number	Berat Badan	Yes	-
Tinggi_Badan	Number	Tinggi Badan	Yes	-
Nilai_Berat_Badan	Number	Nilai Berat Badan	Yes	0
Standard_Berat_Badan	Text,50	Standard Berat Badan	Yes	-
Nilai_LDL	Number	Nilai LDL	Yes	0
Standard_LDL	Text,50	Standard LDL	Yes	-
Nilai_HDL	Number	Nilai HDL	Yes	0
Standard_HDL	Text,50	Standard HDL	Yes	-
Nilai_TC	Number	Nilai TC	Yes	0
Standard_TC	Text,50	Standard TC	Yes	-
Nilai_TGA	Number	Nilai TGA	Yes	0
Standard_TGA	Text,50	Standard TGA	Yes	-
Nilai_Tingkatan_Perokok	Text,50	Nilai Tingkatan Perokok	Yes	-
Standard_Tingkatan_Perokok	Text,50	Standard Tingkatan Perokok	Yes	-
Keterangan	Text,15	Keterangan hasil	Yes	-

2. Tabel Kriteria

Nama : Tb_ Kriteria

Deskripsi isi : Berisi data keputusan dari perpasangan kriteria

Primary key : Kode_ Kriteria

Foreign key : -

Tabel 4.7 Tabel Kriteria

Nama Field	Type dan Length	Deskripsi	Boleh Null	Default
Kode_Kriteria *	Auto Number	Identifier Kode Kriteria	No	Auto Inc
Berat_Badan	text, 25	Nilai Nilai Berat Badan	No	-
LDL	text, 25	Nilai LDL	No	-
HDL	text, 25	Nilai HDL	No	-
TC	text, 25	Nilai TC	No	-
TGA	text, 25	Nilai TGA	No	-
Tingkatan_Perokok	text, 25	Nilai Tingkatan Perokok	No	-
Keterangan	text, 25	Hasil keputusan	No	-

3. Tabel Diagnosa

Nama : Tb_ Diagnosa

Deskripsi isi : Berisi data diagnosa pasien.

Primary key : -

Foreign key : Kode_Pasien

Tabel 4.8 Tabel Diagnosa

Nama Field	Type dan Length	Deskripsi	Boleh Null	Default
Kode_Pasien	Number	Kode Pasien	No	-
Umur	Number	Umur	Yes	-
Point_Umur	Number	Point Umur	Yes	-
Point_TC	Number	Point TC	Yes	-
Merokok		Merokok	Yes	-
Point_Merokok	Number	Point Merokok	Yes	-
Point_HDL	Number	Point HDL	Yes	-
Sistolik	Number	Sistolik	Yes	-

Obat	text, 10	Obat	Yes	-
Point_Sistolik	Number	Point Sistolik	Yes	-
Total_Point	Number	Total Point	Yes	-
Persen_Resiko	Number	Persen Resiko	Yes	-

4. Tabel Login

Nama : Tb_Login

Deskripsi isi : Berisi data login user

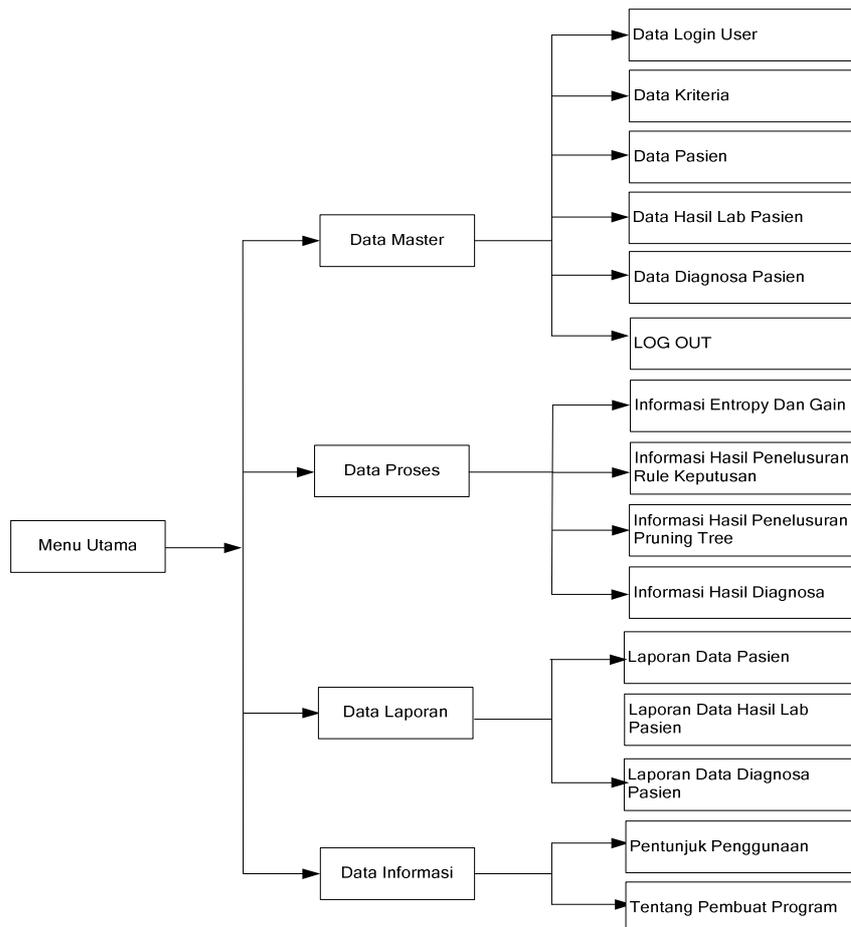
Primary key : -

Tabel 4.9 Tabel Login

Nama Field	Type dan Length	Deskripsi	Boleh Null	Default
User_name	text,40	Nama User login	No	-
Password	Text,40	Password user	No	-
Hak Akses	Text,40	Hak Akses user	No	-

4.2.3 Perancangan Struktur Menu Sistem

Berikut adalah perancangan Struktur menu dari sistem yang dirancang agar memudahkan didalam melakukan integrasi antar modul atau *form*.



Gambar 4.9 Struktur Menu Sistem

4.2.4 Perancangan Antar Muka Sistem

Perancangan antar muka sistem bertujuan untuk menggambarkan sistem yang akan dibuat. Menu utama dari aplikasi ini berisi menu Data Master, Data Proses, Data Laporan, dan Data Informasi. Di halaman utama ini juga berisi informasi tentang tujuan dari pembuatan sistem dan bagaimana cara pengguna menggunakan sistem ini.

Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan (Decision Tree)			
<GAMBAR HEADER>			
[Data Master]	[Data Proses]	[Data Laporan]	[Data Informasi]
Data Login User	Informasi Entropy Dan Gain	Laporan Data Kriteria	Petunjuk Penggunaan
Data Kriteria	Informasi Hasil Penelusuran Rule Keputusan	Laporan Data Pasien	Tentang Pembuat Program
Data Pasien	Informasi Hasil Penelusuran Pruning Tree	Laporan Data Hasil Lab Pasien	
Data Hasil Lab	Informasi Hasil Diagnosa	Laporan Data Hasil Diagnosa Pasien	
Data Hasil Diagnosa			
Log Out			

Gambar 4.10 Perancangan Menu Utama Sistem Pendiagnosa Hiperkolesterol

Perancangan antar muka detail beserta penjelasannya akan dijelaskan pada lampiran C.

BAB V

IMPLEMENTASI DAN PENGUJIAN

5.1 Implementasi Sistem

Implementasi merupakan tahapan pembuatan sistem yang dilakukan berdasarkan hasil analisa dan perancangan yang telah dilakukan sebelumnya. Pada tahapan implementasi sistem ini diharapkan sistem yang telah dirancang siap untuk dioperasikan pada keadaan yang sebenarnya, sehingga akan diketahui apakah sistem yang dibuat benar-benar dapat menghasilkan tujuan yang diinginkan.

5.1.1 Analisa Pemilihan Perangkat Lunak

Perangkat lunak yang digunakan dalam implementasi dan pengujian yaitu *visual basic 6.0* dan *Microsoft Access 2003*. Alasan pemilihan perangkat lunak ini adalah :

1. Secara umum *Visual Basic 6.0* adalah menyediakan komponen-komponen yang memungkinkan untuk membuat program aplikasi yang sesuai dengan tampilan dan cara kerja *windows*.
2. *Microsoft Access 2003* perangkat lunak pengolahan database yang cocok untuk mengelola informasi dalam jumlah yang banyak dan saat ini banyak digunakan. Dengan menggunakan *Microsoft Access 2003*, programmer dapat merancang, membuat dan mengelola database dengan mudah.

5.1.2 Batas Implementasi

Batasan implementasi dari sistem pendiagnosa penyakit hiperkolesterol menggunakan Algoritma C4.5 adalah :

1. Aplikasi sistem pendukung keputusan hanya bisa dioperasikan pada sistem operasi *windows*.
2. Sistem dibuat menggunakan pemrograman *Visual Basic 6.0* dan *database Microsoft Access 2003*
3. sistem mempunyai 2 hak akses yaitu Petugas LAB dan Perawat.
4. Menampilkan data penyakit pasien berdasarkan perhitungan algoritma C4.5 dan hasil diagnosa penyakit.

5.1.3 Lingkungan implementasi

Pada prinsipnya setiap desain sistem yang telah dirancang memerlukan sarana pendukung yaitu berupa peralatan-peralatan yang sangat berperan dalam menunjang penerapan sistem yang didesain terhadap pengolahan data. Komponen-komponen yang dibutuhkan antara lain *hardware*, yaitu kebutuhan perangkat keras komputer dalam pengolahan data kemudian *software*, yaitu kebutuhan akan perangkat lunak berupa sistem untuk mengoperasikan sistem yang telah didesain.

1. Perangkat Keras Komputer dengan spesifikasi:
 - a. Processor : Intel Pentium 4 CPU 3.06 GHz
 - b. Memory : 256 MB
 - c. Harddisk : 40 GB
2. Perangkat Lunak dengan spesifikasi:
 - a. Sistem Operasi : Windows XP Profesional
 - b. Bahasa Pemrograman : Basic
 - c. *Tools* : Visual Basic 6.0

d. DBMS : MS. SQL Server

5.2 Implementasi Sistem Pendiagnosa Penyakit Hiperkolesterol

1. Modul Informasi Data Kriteria Keputusan

No.	Berat Badan	LDL	HDL	TC	TGA	Tingkatan
1	Kurus Tingkat Berat	Normal	Rendah	Normal	Normal	Tingkat 1
2	Kurus Tingkat Berat	Normal	Rendah	Normal	Normal	Tingkat 2
3	Kurus Tingkat Berat	Normal	Rendah	Normal	Normal	Tingkat 3
4	Kurus Tingkat Berat	Normal	Rendah	Normal	Normal	Tingkat 4
5	Kurus Tingkat Berat	Normal	Rendah	Normal	Normal	Tingkat 5
6	Kurus Tingkat Berat	Normal	Rendah	Normal	Sedang	Tingkat 1
7	Kurus Tingkat Berat	Normal	Rendah	Normal	Sedang	Tingkat 2
8	Kurus Tingkat Berat	Normal	Rendah	Normal	Sedang	Tingkat 3
9	Kurus Tingkat Berat	Normal	Rendah	Normal	Sedang	Tingkat 4
10	Kurus Tingkat Berat	Normal	Rendah	Normal	Sedang	Tingkat 5
11	Kurus Tingkat Berat	Normal	Rendah	Normal	Tinggi	Tingkat 1

Gambar 5.1 Modul Informasi Data Kriteria Keputusan

Modul ini berfungsi untuk memasukkan data Kriteria Keputusan, informasi table keputusan berfungsi untuk memasukkan data kriteria yang terdiri dari nilai-nilai dari masing-masing kriteria dan keputusan, form ini berfungsi untuk membangun pohon keputusan. Dalam menu ini juga diberikan fasilitas untuk melakukan perubahan dan penghapusan data kriteria yang telah diinputkan ke dalam sistem

2. Modul Informasi Data Pasien

The screenshot shows a software window titled "Form Input Data Pasien" with a menu bar containing "DATA MASTER", "DATA PROSES", "DATA LAPORAN", and "INFORMASI". The form includes fields for "Nama Lengkap", "Jenis Kelamin", "Tpt. Lahir", "Tgl. Lahir", "Alamat", "Kota", "Propinsi", "Telpon", and "Rwy. Penyakit". Below the form are buttons for "TAMBAH", "SIMPAN", "HAPUS", and "KELUAR". A table below the buttons lists patient data:

No.	NAMA LENGKAP	JENIS KEL.	TPT. LAHIR	TGL. LAHIR	ALAMAT
1	linawati	Perempuan	zxc	23/11/1978	zxz
2	sofyan	Laki-Laki	x	23/11/1982	xz
3	ahmed	Laki-Laki	bcv	23/11/1960	cxk
4	alia	Perempuan	dsd	23/11/1988	dsd

At the bottom of the window, a status bar displays: "Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan (Decision Tree) | ECCL-TIE-FST-UIN SUSKA RIAU | 5:57 | 30/01/2010".

Gambar 5.2 Modul Informasi Data Pasien

Modul ini berfungsi untuk memasukkan data pasien yang melakukan pemeriksaan lab kedalam sistem, nama pasien harus dimasukkan kedalam sistem supaya dapat dilakukan proses penelusuran rule keputusan terhadap pasien tersebut. data field harus dimasukkan semua kedalam sistem sesuai dengan nama fieldnya. dalam modul Informasi data pasien menu ini juga diberi fasilitas untuk melakukan perubahan dan penghapusan data pasien yang telah dimasukkan kedalam sistem.

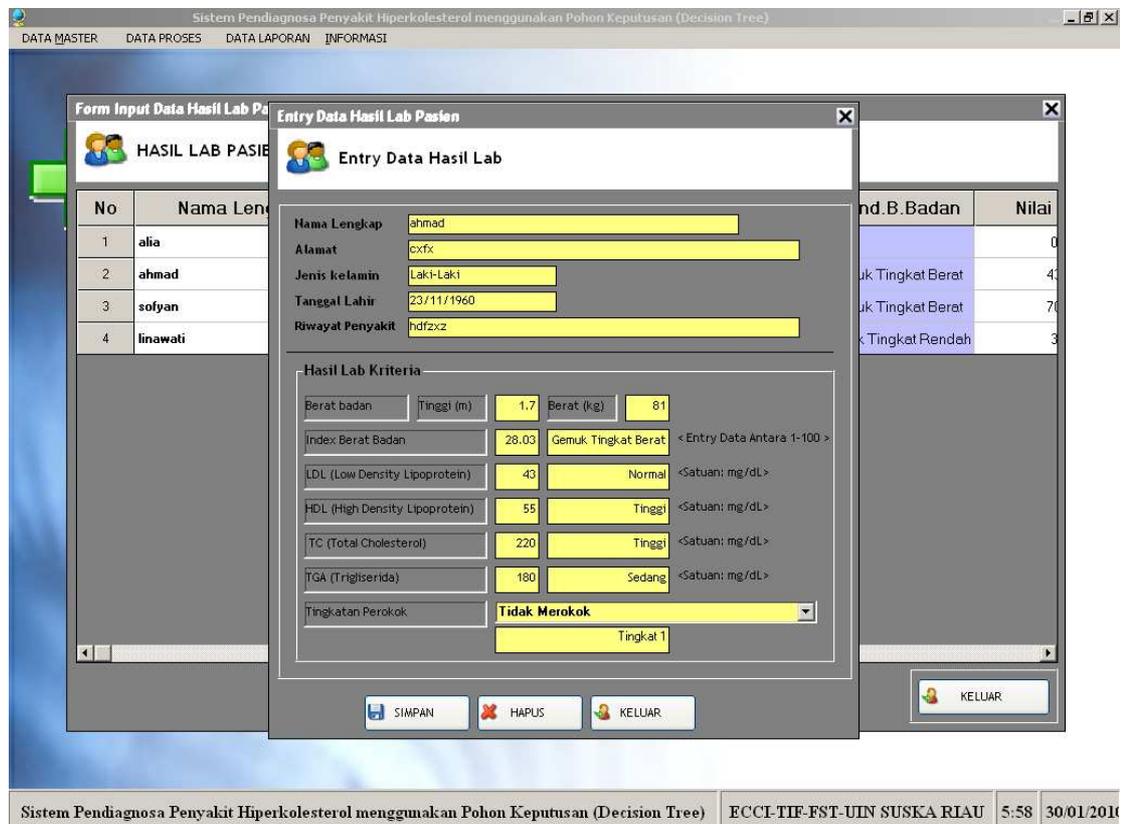
3. Modul Informasi Data Hasil Pemeriksaan Lab Pasien

No	Nama Lengkap	Tinggi	Berat	Nilai B.Badan	Stand.B.Badan	Nilai
1	alia	0	0	0.00		0
2	ahmad	1.7	81	28.03	Gemuk Tingkat Berat	4
3	sofyan	1.7	90	31.14	Gemuk Tingkat Berat	7
4	linawati	1.7	78	26.99	Gemuk Tingkat Rendah	3

Gambar 5.3 Modul Informasi Data Hasil Pemeriksaan Lab Pasien

Modul ini adalah modul informasi data hasil pemeriksaan pasien, informasi yang ditampilkan sesuai dengan variabel yang dibutuhkan dalam pembuatan sistem ini, variabel tersebut meliputi Berat Badan, LDL (*Low Density Lipoprotein*), HDL (*High Density Lipoprotein*), TC (*Total cholesterol*), TGA (*Trigliserida*).

4. Modul Informasi Entri Data Hasil Lab



Gambar 5.4 Modul Informasi Entri Data Hasil Lab

Modul ini berfungsi untuk memasukkan data hasil pemeriksaan lab pada pasien. Langkah yang dilakukan adalah dengan cara melakukan klik pada baris tabel nama pasien yang akan diinputkan data hasil pemeriksaan labnya. Selanjutnya akan muncul informasi entri data hasil lab seperti pada gambar 5.10, lakukan pengisian data sesuai dengan nama fieldnya kemudian lakukan penyimpanan maka data hasil pemeriksaan lab pada pasien pada tabel akan berubah sesuai dengan data variabel yang diinputkan. Data yang diinputkan dapat dirubah kembali jika ada kesalahan.

5. Modul Informasi Data Hasil Diagnosa Pasien

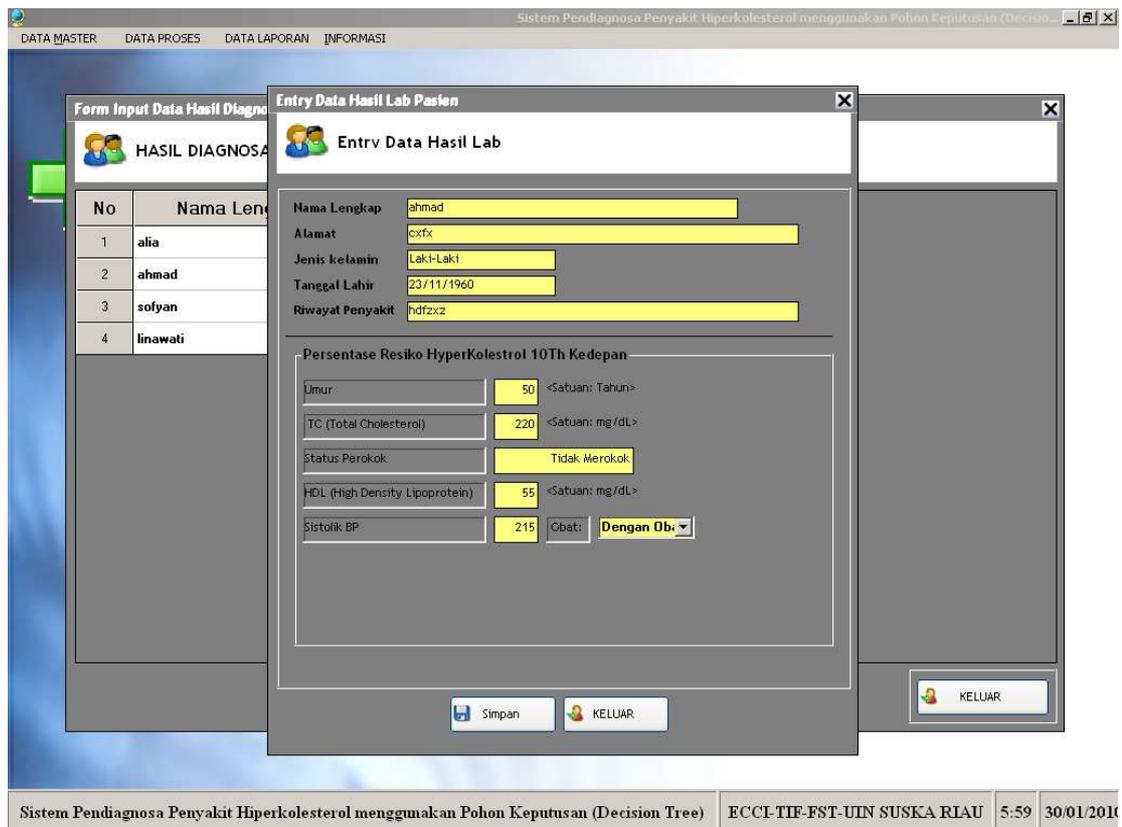
The screenshot shows a software window titled "Form Input Data Hasil Diagnosa" with a menu bar containing "DATA MASTER", "DATA PROSES", "DATA LAPORAN", and "INFORMASI". The window contains a table with the following data:

No	Nama Lengkap	Umur	Sistolik	Obat
1	alia	0	0	
2	ahmad	50	215	Dengan Obat
3	sofyan	28	200	Tanpa Obat
4	linawati	32	225	Tanpa Obat

A "KELUAR" button is located at the bottom right of the form.

Gambar 5.5 Modul Informasi Data Hasil Diagnosa Pasien

Modul ini berfungsi untuk menampilkan informasi data yang akan digunakan dalam perhitungan diagnosa pasien. Data yang diinputkan adalah data sistolik dengan kondisi menggunakan obat atau tidak. Untuk melakukan input data sistolik dengan cara melakukan klik pada baris tabel sesuai dengan nama pasien. Maka akan muncul form dibawah ini.



Gambar 5.6 Modul Input Data Hasil Diagnosa Pasien

Modul ini berfungsi untuk memasukkan data sistolik pada pasien yang akan digunakan untuk proses perhitungan diagnosa pasien. lakukan pengisian data nama field sistolik BP dan kondisi penggunaan obat kemudian lakukan penyimpanan maka data diagnosa pada pasien pada tabel akan berubah sesuai dengan data sistolik BP yang diinputkan. Untuk data umur, TC, status perokok dan HDL akan muncul secara otomatis karena diambil dari modul hasil pemeriksaan lab. Data yang diinputkan dpat dirubah kembali jika ada kesalahan.

6. Modul Informasi Data Entropy Dan Gain

Gambar 5.7 Modul Informasi Data Entropy Dan Gain

Modul ini berfungsi untuk melihat informasi entropy dan gain, , fungsi dari proses entropy dan gain adalah untuk mengubah data dari tabel keputusan menjadi pohon keputusan dengan menggunakan algoritma C45. dalam modul ini terdiri dari enam nama tab, yang berarti terdiri dari enam kali iterasi untuk menjadikan informasi pohon keputusan.

7. Modul Informasi Proses hasil Penelusuran Rule Keputusan

No	Nama Lengkap	Stand.B.Badan	Stand.LDL	Stand.HDL	Stand.TC	Status
1	alia					
2	ahmad	Gemuk Tingkat Berat	Normal	Tinggi	Tinggi	Se
3	sofyan	Gemuk Tingkat Berat	Normal	Rendah	Normal	N
4	linawati	Gemuk Tingkat Rendah	Normal	Rendah	Normal	T

Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan (Decision Tree) | ECCL-TIF-FST-UIN SUSKA RIAU | 6:00 | 30/01/2010

Gambar 5.8 Modul Informasi Proses hasil Penelusuran Rule Keputusan

Modul ini berfungsi untuk melakukan proses penelusuran rule keputusan dinama hasilnya adalah informasi berupa apakah pasien tersebut hiperkolesterol atau tidak. Dalam tabel tersebut informasi yang ditampilkan adalah nama pasien, nilai kriteria standar dari hasil pemeriksaan lab pada pasien tersebut dan hasil informasi berupa apakah pasien tersebut hiperkolesterol atau tidak.

8. Modul Informasi Proses Penelusuran Pruning Tree

The screenshot shows a software window titled "Form Pencarian Nilai Hasil Penelusuran Rule Keputusan Pruning Tree" with a subtitle "INFORMASI HASIL PENELUSURAN PRUNING TREE". The window is part of a system called "Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan". The interface includes the following elements:

- Navigation Menu:** DATA MASTER, DATA PROSES, DATA LAPORAN, INFORMASI.
- Form Fields:**
 - Nama Lengkap: solyan
 - Jenis Kelamin: Laki-Laki
 - Tpt. Lahir: x
 - Tgl. Lahir: 23/11/1982
 - Alamat: xz
 - Berat badan: Gemuk Tingkat Berat
 - HDL (High Density Lipoprotein): Rendah
 - TGA (Trigliserida): Normal
 - LDL (Low Density Lipoprotein): Normal
 - TC (Total Cholesterol): Normal
 - Tingkatan Perokok: Tingkat 5
 - Keterangan Hasil: Hiperkolesterol
- Rule Penelusuran Pruning Tree:**
 - Penelusuran 1: Penelusuran 1 TC:Normal [133,317]
 - Penelusuran 2: Penelusuran 2 Tingkat Perokok:Tingkat 5 [62,28]
 - Penelusuran 3: Penelusuran 3 LDL:Normal [9,21]
 - Penelusuran 4: Penelusuran 4 Berat Badan:Gemuk Tingkat Berat [3,3]
 - Penelusuran 5: Penelusuran 5 HDL:Rendah [3,0]
 - Penelusuran 6: (empty field)
- Hasil:** Hiperkolesterol
- Buttons:** KELUAR

At the bottom of the window, there is a status bar with the text: "Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan (Decision Tree) ECCL-TIE-FST-UIN SUSKA RIAU 6:00 30/01/2010".

Gambar 5.9 Modul Informasi Proses Penelusuran Pruning Tree

Modul ini berfungsi untuk melakukan proses Penelusuran Pruning Tree dimana hasilnya adalah informasi berupa apakah pasien tersebut hiperkolesterol atau tidak dan juga informasi kedalaman penelusuran mulai dari penelusuran 1 sampai dengan penelusuran 6. informasi yang dihasilkan harus sama antara penelusuran Pruning Tree dan penelusuran rule keputusan

9. Modul Informasi Proses Hasil Diagnosa Pasien

The screenshot displays a software window titled "Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan". The interface includes a menu bar with "DATA MASTER", "DATA PROSES", "DATA LAPORAN", and "INFORMASI". A table on the left lists patient names: alia, ahmad, sofyan, and linawati. The main window, "Entry Data Hasil Lab Pasien", contains the following fields:

- Nama Lengkap: ahmad
- Alamat: cxf
- Jenis kelamin: Laki-Laki
- Tanggal Lahir: 23/11/1960
- Riwayat Penyakit: hdfzrz

Below these fields is a section titled "Persentase Resiko Hyperkolestrol 10Th Kedepan" with the following inputs:

- Umur: 50 (Satuan: Tahun) - Point Umur: []
- TC (Total Cholesterol): 220 (Satuan: mg/dL) - Point TC: []
- Status Perokok: Tidak Merokok - Point Rokok: []
- HDL (High Density Lipoprotein): 55 (Satuan: mg/dL) - Point HDL: []
- Sistolik BP: 215 - Obat: Dengan Obat - Point Sistolik: []
- Total Point: []
- Persentase Resiko 10Th: []

Buttons for "Proses" and "KELUAR" are visible at the bottom of the form. A footer bar contains the text: "Sistem Pendiagnosa Penyakit Hiperkolesterol menggunakan Pohon Keputusan (Decision Tree) ECCE-TIE-FST-UIN SUSKA RIAU 6:15 30/01/2010".

Gambar 5.10 Modul Informasi Proses Hasil Diagnosa Pasien

Modul ini berfungsi untuk melakukan proses perhitungan hasil diagnosa pasien. Informasi yang dihasilkan adalah prediksi 10 tahun kedepan terhadap pasien tersebut apakah akan mengalami resiko hiperkolesterol atau tidak. Hasil diagnosanya berupa besarnya persentase resiko hiperkolesterol dalam 10 tahun kedepan. Proses implementasi Sistem Pendiagnosa Penyakit Hiperkolestrol secara rinci serta dokumentasinya ada pada lampiran F.

5.3 Pengujian Sistem

Pengujian dilakukan untuk melihat hasil implementasi, apakah berjalan sesuai tujuan atau masih terdapat kesalahan-kesalahan. Pengujian sistem

pendiagnosa penyakit hiperkolesterol dilakukan pada lingkungan pengujian sesuai dengan lingkungan implementasi. Pengujian dilakukan dengan menguji fungsi-fungsi per modul.

5.3.1 Lingkungan Pengujian

Perangkat keras dan perangkat lunak yang digunakan, data dan responden yang menguji sistem, serta bentuk observasi yang dilakukan. sistem pendiagnosa hiperkolesterol diuji dengan menggunakan komputer sebagai berikut:

1. Perangkat Keras
 - a. Processor : AMD Sempron 2600 MHz
 - b. Memory : 256 MB
 - c. Harddisk : 40 GB
2. Perangkat Lunak
 - a. Sistem Operasi : Windows XP Profesional
 - b. Bahasa Pemrograman : Visual Basic
 - c. *Tools* : Visual Basic 6.0
 - d. DBMS : MS. SQL Server

5.3.2 Pengujian

Pengujian yang dilakukan terdiri dari *Black Box*, *User Acceptance Test* (UAT), dan Performansi Algoritma C4.5. Metode pengujian *Black Box* memfokuskan pada keperluan fungsional dari software. Karena itu ujicoba *Black Box* memungkinkan pengembang *software* untuk membuat himpunan kondisi input yang akan melatih seluruh syarat-syarat fungsional suatu program. UAT

adalah proses dimana pengguna melakukan verifikasi program/aplikasi yang dibuat. Pada akhir dari fase UAT, para pengguna akan diminta untuk melakukan *sign-off* pada UAT. Untuk Performansi Algoritma C4.5 dilakukan pengujian tingkat persenan *error* data. Kelas uji pada identifikasi pengajuan secara rinci sebagai berikut:

5.3.2.1 Pengujian Dengan Menggunakan *Blakbox*

Pengujian dengan menggunakan *blackbox* yaitu pengujian yang dilakukan untuk antarmuka perangkat lunak, pengujian ini dilakukan untuk memperlihatkan bahwa fungsi-fungsi bekerja dengan baik dalam artian masukkan diterima dengan benar dan keluaran yang dihasilkan benar-benar tepat, pengintegrasian eksternal data berjalan dengan baik.

Prekondisi :

1. Dapat dibuka dari layar menu utama bagian admin
2. Ditabel keputusan telah diisi data Login

Tabel 5.1 Butir Uji Pengujian Modul Pengelolaan Login

Deskripsi	Prekondisi	Prosedur Pengujian	Masukkan	Keluaran yang diharapkan	Kriteria Evaluasi Hasil	Hasil yang didapat	kesimpulan
Pengujian Login	Tampil layar menu utama aplikasi	1.Masukan username, password dan hak akses. 2.Klik tombol login untuk masuk ke menu utama 3.Tampil menu utama	Data username, password dan hak akses	Data Berhasil diproses tampilan menu utama dan tidak ada intruksi error	Data berhasil diproses, tampilan menu utama dan tidak ada intruksi error	Data berhasil diproses, tampilan menu utama dan tidak ada intruksi error	Diterima

Prekondisi :

1. Dapat dibuka dari layar menu utama bagian admin
2. Ditabel keputusan telah diisi data kriteria

Tabel 5.2 Butir Pengujian Modul Input Tabel Keputusan.

Deskripsi	Prekondisi	Prosedur Pengujian	Masukkan	Keluaran yang diharapkan	Kriteria Evaluasi Hasil	Hasil yang didapat	kesimpulan
Pengujian input data table keputusan	Tampil layar menu utama bagian admin	1.Klik tombol 'Tambah' 2.Isi data sesuai dengan fieldnya 3.Klik tombol simpan	Data seluruh kriteria yang diperlukan	Data Berhasil diproses tampilan menu utama dan tidak ada intruksi error	Data berhasil diproses, tampilan menu utama dan tidak ada intruksi error	Data berhasil diproses, tampilan menu utama dan tidak ada intruksi error	Diterima

Dokumentasi riciannya ada pada lampiran G

5.3.2.2 Pengujian Dengan Menggunakan *User Acceptance Test*

User acceptance test adalah pengujian akhir yang dilakukan oleh calon pengguna atau sistem yang telah siap kita ajukan. Hasil dari pengujian tersebut dilampirkan berupa kuesioner yang diisi oleh calon pengguna dalam hal ini yaitu Petugas Lab dan Perawat Berikut data para responden:

Tabel 5.3 Data Responden

No	Nama	Jabatan
1	Dr. Yossi Eka Putri	Dokter
2	Nurhayati, S. Psi	Perawat
3	Srimarina, A.Md	Petugas Lab
4	Astuti, S.Kep	Perawat

Berikut data hasil pengujian pada sistem ini didapat berdasarkan atas penilaian responden terhadap sitem yang akan digunakan data sebagai berikut :

Tabel 5.4 Hasil Data Responden

Pertanyaan	Jawaban	
	Ya	Tidak
1. Setelah anda melihat dan menggunakan sistem ini, menurut anda apakah sistem ini bisa digunakan sebagai alat Bantu untuk memudahkan dalam melihat hasil keputusan yang akurat?	4 Orang	0
2. Menurut anda apakah sistem ini sudah memenuhi syarat untuk digunakan secara nyata menjadi fasilitas anda?	3 Orang	1 Orang
3. Menurut anda apakah sistem ini mudah digunakan (<i>User Friendly</i>)?	4 Orang	0 Orang
4. Secara umum apakah informasi yang dihasilkan sistem ini telah sesuai dengan kebutuhan didalam memberikan hasil keputusan diagnosa penyakit hiperkolesterol?	3 Orang	1 Orang
5. sejauh yang anda tahu, apakah sebelumnya sudah ada aplikasi yang sama dengan sistem ini ?	2 Orang	2 Orang

5.3.2.3 Pengujian Dengan Menggunakan Performansi Algoritma C4.5

Data *Training Test* diambil dari data pasien pada tabel keputusan.

Tabel 5.5 Data *Training Test*

No	Berat Badan	LDL	HDL	TC	TGA	Tingkatan Perokok	Hasil Analisa
1	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 1	Tidak
2	Normal	Normal	Tinggi	Tinggi	Normal	Tingkat 2	Tidak
3	Gemuk Tinggi	Normal	Rendah	Normal	Normal	Tingkat 1	Tidak
4	Gemuk Rendah	Normal	Rendah	Normal	Tinggi	Tingkat 5	Hiperlipoproteinemia Tipe V
5	Normal	Tinggi	Rendah	Normal	Tinggi	Tingkat 3	Tidak
6	Gemuk Rendah	Normal	Tinggi	Normal	Tinggi	Tingkat 3	Hiperlipoproteinemia Tipe IV
7	Gemuk Tinggi	Sedang	Tinggi	Tinggi	Normal	Tingkat 5	Hiperkolesterol
8	Gemuk Rendah	Normal	Rendah	Tinggi	Sedang	Tingkat 5	Hiperkolesterol
9	Normal	Normal	Rendah	Tinggi	Tinggi	Tingkat 1	Tidak
10	Gemuk Tinggi	Tinggi	Tinggi	Tinggi	Sedang	Tingkat 4	Hiperkolesterol
11	Gemuk Rendah	Tinggi	Rendah	Normal	Tinggi	Tingkat 1	Hiperlipoproteinemia Tipe IV
12	Normal	Sedang	Tinggi	Normal	Sedang	Tingkat 2	Tidak
13	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 4	Hiperkolesterol
14	Gemuk Rendah	Tinggi	Tinggi	Normal	Sedang	Tingkat 5	Hiperkolesterol
15	Kurus Berat	Normal	Rendah	Tinggi	Tinggi	Tingkat 4	Hiperkilomikronemia
16	Gemuk Rendah	Tinggi	Rendah	Normal	Tinggi	Tingkat 5	Hiperlipoproteinemia Tipe V
17	Normal	Normal	Rendah	Normal	Normal	Tingkat 3	Tidak
18	Kurus Ringan	Sedang	Rendah	Tinggi	Normal	Tingkat 1	Tidak
19	Gemuk Tinggi	Sedang	Rendah	Tinggi	Sedang	Tingkat 2	Hiperkolesterol
20	Gemuk Rendah	Tinggi	Rendah	Tinggi	Sedang	Tingkat 5	Hiperkolesterol

Data yang diambil dalam pengujian ini adalah data baru dari 5 pasien.

Tabel 5.6 Data Hasil Pasien.

No	Berat Badan	LDL	HDL	TC	TGA	Tingkatan Perokok	Hasil Analisa
1	Kurus Berat	Normal	Rendah	Normal	Normal	Tingkat 1	Tidak
2	Normal	Normal	Tinggi	Tinggi	Normal	Tingkat 2	Tidak
3	Gemuk Tinggi	Normal	Rendah	Normal	Normal	Tingkat 1	Tidak
4	Gemuk Rendah	Normal	Rendah	Normal	Tinggi	Tingkat 5	Hiperlipoproteinemia Tipe V
5	Normal	Tinggi	Rendah	Normal	Tinggi	Tingkat 3	Tidak
6	Gemuk Rendah	Normal	Tinggi	Normal	Tinggi	Tingkat 3	Hiperlipoproteinemia Tipe IV
7	Gemuk Tinggi	sedang	Rendah	Tinggi	Normal	Tingkat 2	Hiperkolesterol
8	Gemuk Rendah	Normal	Rendah	Tinggi	Sedang	Tingkat 5	Hiperkolesterol
9	Normal	Normal	Rendah	Tinggi	Tinggi	Tingkat 1	Tidak
10	Gemuk Tinggi	Tinggi	Tinggi	Tinggi	Sedang	Tingkat 4	Hiperkolesterol
11	Gemuk Rendah	Tinggi	Rendah	Normal	Tinggi	Tingkat 1	Hiperlipoproteinemia Tipe IV
12	Normal	Sedang	Tinggi	Normal	Sedang	Tingkat 2	Tidak
13	Kurus Berat	Normal	Rendah	Tinggi	Normal	Tingkat 4	Tidak
14	Gemuk rendah	Tinggi	Tinggi	Normal	Sedang	Tingkat 5	Hiperkolesterol
15	Kurus Berat	Normal	Rendah	Tinggi	Tinggi	Tingkat 4	Hiperkilomikronemia
16	Gemuk Rendah	Tinggi	Rendah	Normal	Tinggi	Tingkat 5	Hiperlipoproteinemia Tipe V
17	Normal	Normal	Rendah	Normal	Normal	Tingkat 3	Tidak
18	Kurus Ringan	Sedang	Rendah	Tinggi	Normal	Tingkat 1	Tidak
19	Gemuk Tinggi	Sedang	Rendah	Tinggi	Sedang	Tingkat 2	Hiperkolesterol
20	Gemuk Rendah	Tinggi	Rendah	Tinggi	Sedang	Tingkat 5	Hiperkolesterol

Data-data tersebut melalui proses perhitungan *entropy* dan *pruning*, setelah proses tersebut diperoleh hasil perhitungannya. Dan dapat di hitung tingkat error data. Pada data diatas tingkat error data nya adalah $(1/20 \times 100\% = 5\%)$.

5.3.2.4 Kesimpulan Pengujian

Setelah melakukan pengujian sistem terhadap kasus sistem pendiagnosa penyakit hiperkolesterol ini berdasarkan hasil pengujian *black box*, hasil pengujian menggunakan *user acceptance test*, dan hasil pengujian dengan menggunakan performansi algoritma C4.5 maka dapat diambil kesimpulan:

1. Setelah melakukan beberapa pengujian, *output* yang dihasilkan implementasi pada sistem pendiagnosa penyakit hiperkolesterol ini sesuai dengan analisa dan perancangan.
2. Dari tabel hasil data responden diatas dapat diambil kesimpulan bahwa sistem pendiagnosa penyakit hiperkolesterol ini dapat diterima oleh dokter dan perawat (pengguna sistem) karena sistem ini lebih efisien, mudah untuk digunakan (*user friendly*), ekonomis dan akurat dibandingkan sistem yang digunakan sekarang ini yang masih bersifat manual.
3. Hasil pengujian performansi algoritma C4.5, dari data *Training Test* yang berjumlah 20 data diperoleh *error* sebesar 5%.

BAB VI

PENUTUP

6.1 Kesimpulan

Dengan adanya sistem pendiagnosa penyakit hiperkolesterol menggunakan pohon keputusan (*decision tree*), dapat diambil suatu kesimpulan sebagai berikut:

1. Sistem Pendiagnosa Penyakit Hiperkolesterol berhasil dibangun sesuai dengan kriteria penentuan hiperkilomikronemia, hiperkolesterol, hiperlipoproteinemia tipe III, hiperlipoproteinemia tipe IV, dan hiperlipoproteinemia tipe V, sesuai dengan kriteria-kriteria sebagai berikut:
 - a. Indeks Massa Tubuh (IMT)
 - b. *Low Density Lipoprotein* (LDL)
 - c. *High Density Lipoprotein* (HDL)
 - d. *Total Cholesterol* (TC)
 - e. *Trigeliserida* (TGA)
 - f. tingkatan perokok.
2. Berdasarkan hasil pengujian dengan menggunakan *user acceptance test* bahwa Sistem Pendiagnosa penyakit hiperkolesterol ini dapat diterima oleh dokter dan perawat (pengguna sistem) karena sistem ini lebih efisien, mudah untuk digunakan (*user friendly*) dan akurat dibandingkan sistem yang digunakan sekarang ini yang masih bersifat manual.

6.2 Saran

Agar sistem ini dapat bermanfaat baik untuk sekarang maupun akan datang, maka penulis memberikan saran, sebagai berikut:

1. Untuk kebutuhan yang akan datang, sistem ini perlu penambahan variabel/kriteria penyakit hiperkolesterol, agar tingkat akurasi datanya lebih baik.
2. Sistem ini perlu dibandingkan dengan metode lain, untuk mendapatkan tingkat error yang lebih kecil.

DAFTAR PUSTAKA

- Almatsier, Sunita, "*Penuntun Diet*", Edisi baru, halaman 21-22, Gramedia, Jakarta, 2005.
- Arora, Anjali, "*5 langkah memahami kolesterol*", halaman 1-123, BIP, Jakarta, 2007
- Department of Medical-Surgical Nursing, "*Learning about Cholesterol*", halaman 1-4, Ohio State University Medical Center, USA, 2007
- Itje, Enny Sela, "*Knowledge Discovery pada Risked Costumer's Bank menggunakan Decision Tree*", KOMMIT, halaman 5-7, 2006.
- Kumar, "*Introduction to Data Mining*", Addison-wesley, halaman 145-205, 2006.
- Kusumadewi, Sri, "*Artificial Intelligence (Teknik dan Aplikasinya)*", halaman 115-116, Graha Ilmu, Bandung, 2003.
- McLeod, Raymond, "*Sistem Informasi Manajemen Jilid II*", halaman 110-120, PT Prenhallindo, Jakarta, 1996.
- Mitchell, Tom M, "*Machine Learning*", Halaman 52-80, McGraw Hill, Singapore, 1997.
- Oetoro, Samuel MS, "*Cara Cerdas Menyikapi Kolesterol*" [online] Available http://www.medicastore.com/kolesterol/Cara_cerdas_menyikapi_kolesterol.htm, diakses 23 September 2008.
- Peng, Wei. "*An Implementation of ID3 --- Decision Tree Learning Algorithm*", halaman 1-20, University of New South Wales, Sydney, 2008
- Suyanto, "*Artificial Intelligence : Searching, Reasoning, Planning dan Learning*", halaman 137-163, Informatika, Bandung, 2007.
- Quinlan, J. Ross, "*Induction of Decision Tree : Machine Learning vol. I*", Halaman 81-106, Kluwer Academic Publishers, Boston, 1986
- Quinlan, J. Ross, "*Improved Use of Continuous Attributes in C4.5*", Halaman 77-90, Journal of Artificial Intelligence Research 4, 1996
- Zelič, Igor, "*Impact of Machine Learning to the Diagnosis and Prognosis of First Cerebral Paroxysm*", Halaman 1-3, Slovenia, 1999.

_____, "*The ID3 Algorithm*" [online] available
<http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm>, diakses 3
Februari 2008

_____, "*Hiperkolesterol*" [online] available
[http://www.pjnhk.go.id/index2.php?option=com_content&do_pdf=1&id=16
93](http://www.pjnhk.go.id/index2.php?option=com_content&do_pdf=1&id=1693), diakses tanggal 23 September 2008