

BAB III

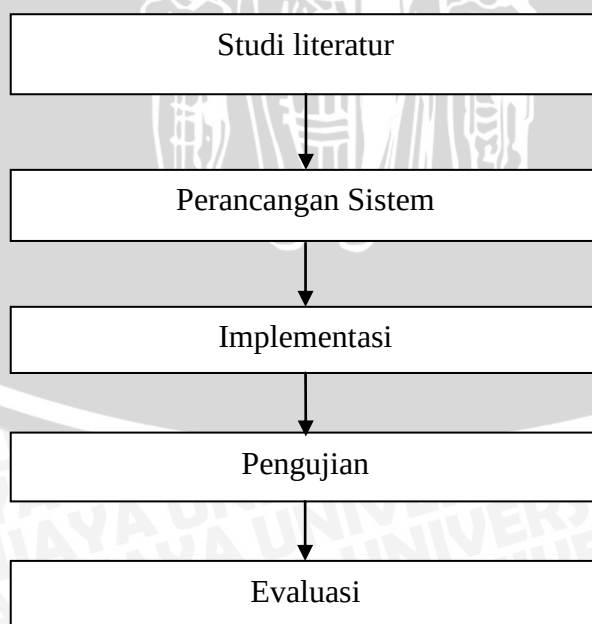
METODOLOGI DAN PERANCANGAN SISTEM

Pada bab metodologi dan perancangan dibahas tentang metode yang digunakan dalam penelitian, Rancangan yang akan dibuat untuk membangun perangkat lunak klasifikasi penyakit thypus menggunakan *Algoritma Induction Decision Tree* (ID-3)

Penelitian yang dilakukan dengan tahapan sebagai berikut:

1. Membaca dan mempelajari literatur yang terkait dengan metode *decision tree* serta algoritma ID-3 dan juga penyakit thypus.
2. Merancang system.
3. Membuat perangkat lunak sesuai dengan analisa dan perancangan yang dilakukan.
4. Menguji perangkat lunak.
5. Mengevaluasi hasil analisa yang dilakukan.

Langkah langkah yang dilakukan dapat dilihat pada Gambar 3.1



Gambar 3.1 Langkah - Langkah Penelitian

3.1 Studi Literatur

Pada tahap ini dilakukan studi literatur sebagai tahap teori untuk mendukung dalam penyelesaian masalah pada penelitian. Teori-teori tersebut mengenai penyakit thypus, decision tree, data mining, dan algoritma ID-3 yang digunakan sebagai dasar teori yang didapat dari beberapa jurnal, pakar dan browsing internet.

3.2 Perancangan Sistem

3.2.1 Keterangan Data

Data yang digunakan adalah data yang diperoleh dari Rumah sakit umum besuki. Pada data ini terdapat 8 atribut dan 2 kelas, antara lain;

1. Demam
2. Pusing
3. Lesu
4. Infeksi Tenggorakan
5. Perut tidak enak
6. Lidah tengah kotor
7. Ujung lidah merah
8. Ngigau
9. Class (positif dan negatif)

Data yang tersedia sebanyak 100 data, tetapi terdapat 6 data dengan missing value. Pada sistem ini missing value tetap digunakan, karena missing value akan diisi dengan KNN dimana 6 data missing value akan diklasifikasikan kelasnya dengan memilih data sejumlah k yang letaknya terdekat dari data ini kemudian *class* terbanyak dari data terdekat sejumlah k tersebut akan dipilih sebagai *class* yang diprediksikan untuk data ini. Berikut adalah data yang ditampilkan ditabel 3.1

3.2.2 Deskripsi Sistem

Sistem yang digunakan merupakan perangkat lunak untuk pengklasifikasian positif dan negatif penyakit thypus. Sistem mengimplementasikan metode *decision tree* dengan algoritma ID-3 dalam

pengklasifikasiannya. Pengklasifikasian penyakit thypus ini dibuat dua tahap. Tahap pertama proses pelatihan terhadap sekumpulan data latih dan tahap selanjutnya adalah proses percobaan menggunakan data uji yang kemudian diklasifikasikan menggunakan aturan - aturan yang sudah terbentuk dari proses pembelajaran.

3.2.3 Batasan Sistem

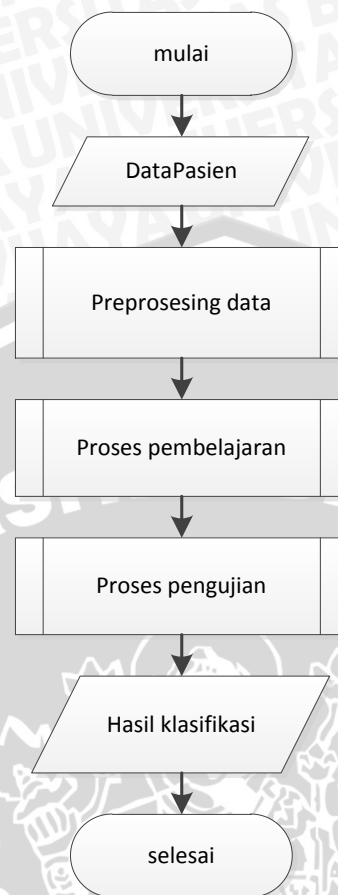
Sistem yang dibuat memiliki batasan sebagai berikut :

1. Data latih dan data uji tidak memiliki missing value.
2. Penanganan missing value.
3. Data yang digunakan adalah data dengan jumlah kelas =2.

3.2.4 Perancangan Proses

Pada sistem yang dibuat terdapat dua proses utama, yaitu proses pengklasifikasian data dan proses penentuan positif dan negatif penyakit thypus. Berdasarkan Gambar 3.2 tersebut, rancangan sistem meliputi beberapa proses utama, antara lain adalah ;

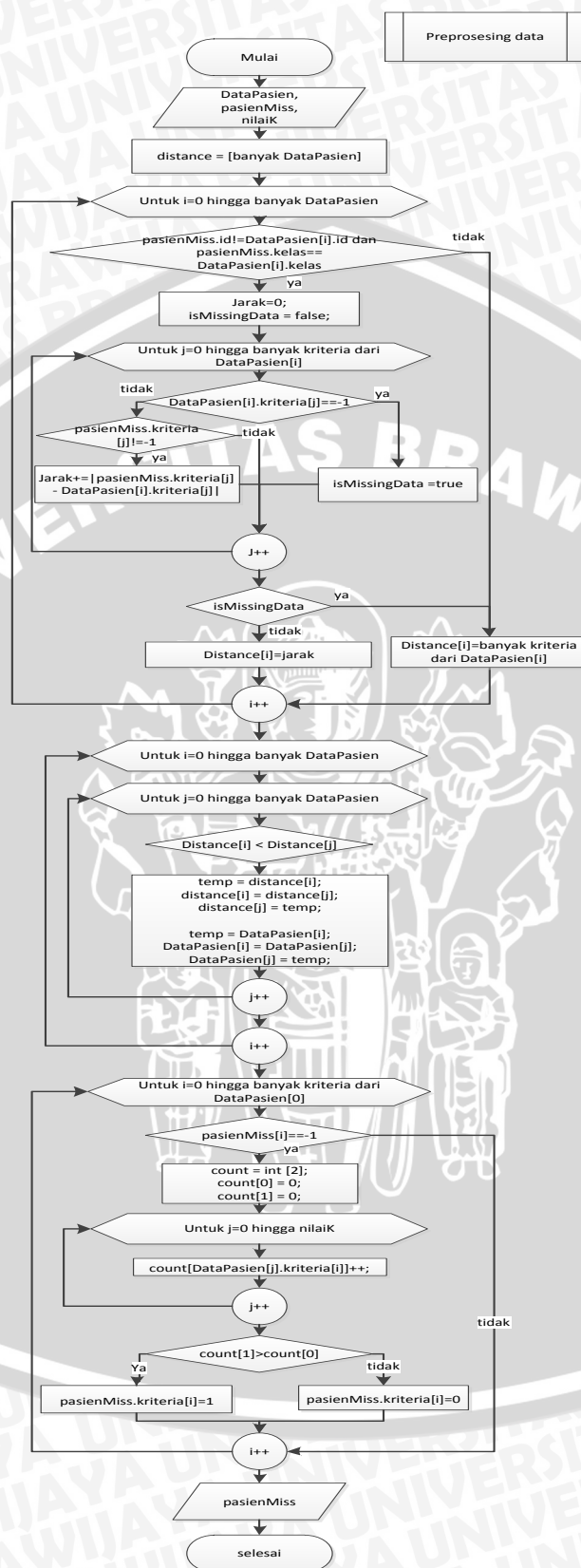
1. Proses memasukkan data pasien berupa data latih.
2. Proses pembelajaran untuk membangun *decision tree* dan membentuk aturan-aturan untuk pengklasifikasian data uji,
3. Proses percobaan untuk menguji data uji yang dimasukkan.
4. Hasil klasifikasi data uji.



Gambar 3.2 Rancangan sistem

3.2.4.1 Preprosesing Data

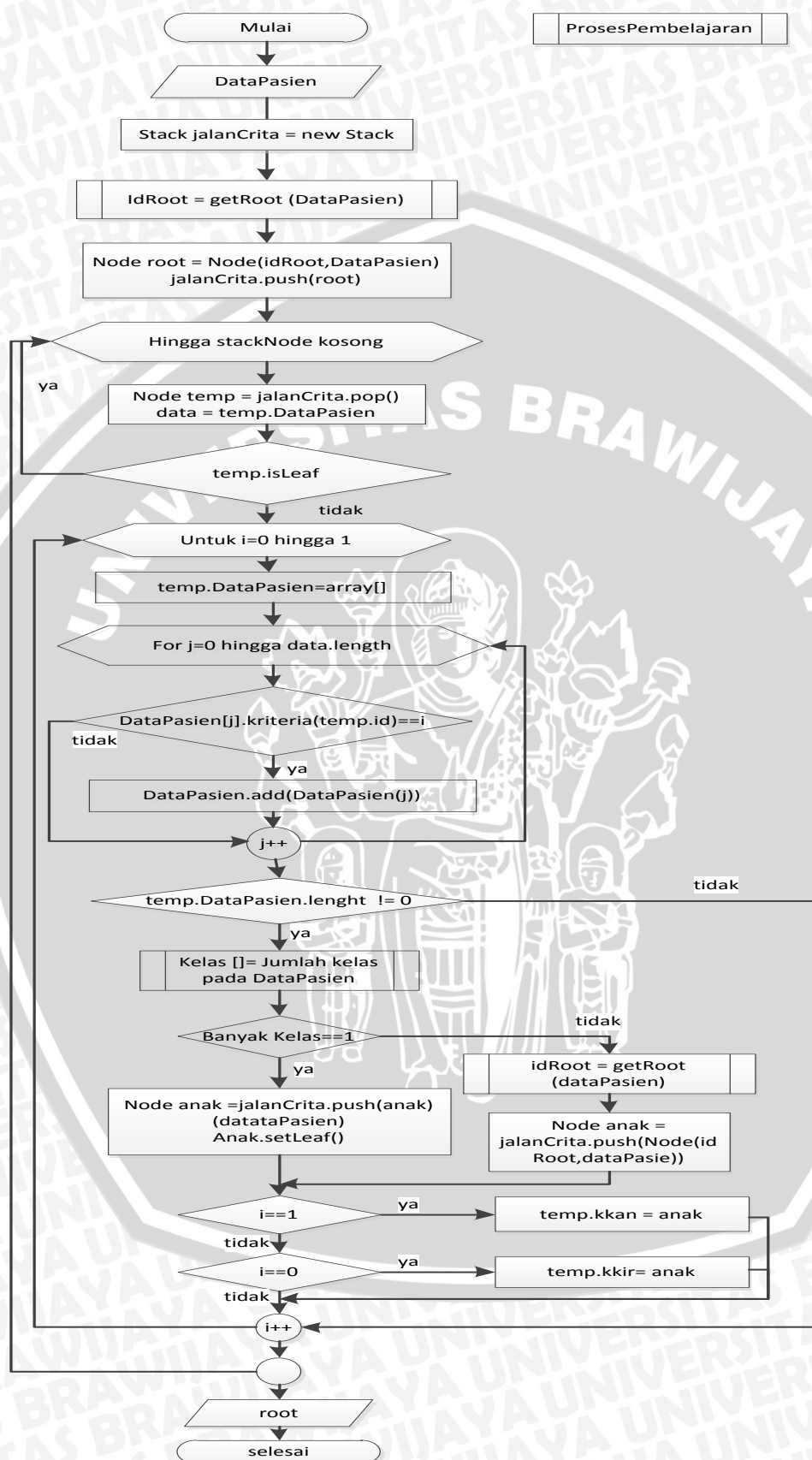
Preprosesing data bertujuan untuk mengisi nilai kriteria pada data pasien yang tidak utuh (*missing value*). *Preprosesing data* menerapkan metode KNN untuk menentukan nilai yang digunakan untuk mengisi kriteria yang kosong tersebut. Konsep dasar KNN yang diterapkan adalah mencari jarak antar data yaitu jarak pasien yang memiliki nilai kriteria yang kosong (*missing value*) dengan pasien lain yang masih dalam satu kelas. Dicari jarak terkecil dari hasil perhitungan jarak, pasien yang memiliki jarak terkecil (mirip) terhadap pasien yang memiliki *missing value* akan diambil nilai kriteria pada posisi kriteria pasien yang *missing value*. Hasilnya nilai *miss* pada pasien tertentu akan terpenuhi dengan mengisi nilai sama dengan pasien lain yang mirip dan masih dalam satu kelas. Pada gambar 3.3 ditunjukkan diagram alir *preprosesing data*.



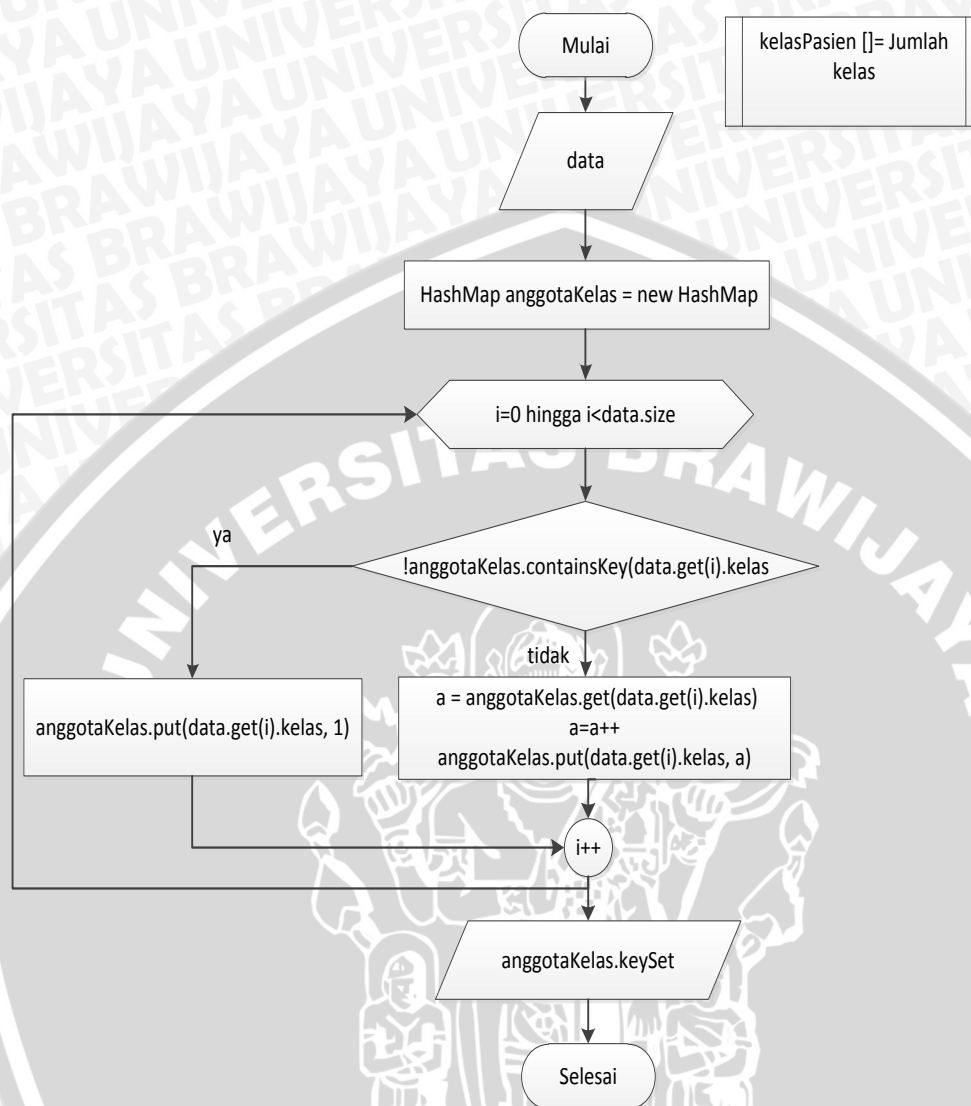
Gambar 3.3 Diagram Alir Preprocessing Data

3.2.4.2 Proses Pembelajaran

Proses pembelajaran berupa pembentukan *tree* adalah proses utama pada metode ID3. Setiap node yang dipilih sebagai *parent* adalah nilai entropi terbesar dari gejala yang diderita pasien. Proses dimulai dengan mendapatkan *information gain* dari setiap gejala pasien, kemudian dipilih yang terbesar untuk ditempatkan sebagai *root* dari *tree* yang akan dibentuk. Struktur data yang dipilih adalah *stack*/tumpukan untuk menyelesaikan pembentukan *tree* dimulai dari yang paling dalam. Hasil penentuan *root* yang dilakukan sebelumnya dimasukan kedalam *stack* dan dilakukan perulangan hingga *stack* kosong, dimana didalam setiap perulangannya *stack* akan diambil satu persatu. Penentuan node berikutnya dimulai dengan memilah data pasien sesuai dengan nilai pada atribut gejala yang dipilih sebagai *parent*. Nilai 'ya' untuk anak sebelah kanan dan nilai 'tidak' untuk anak sebelah kiri. Hasil filtrasi masing - masing data pasien kemudian dihitung kembali *informasi gain* dan ditentukan gejala dengan nilai entropi tertinggi untuk dipilih sebagai node berikutnya dengan memasukan kedalam *stack*. Suatu node dipilih sebagai daun dan tidak lagi dimasukkan ke dalam *stack* apabila data hasil filtrasi memiliki kelas yang sama. Gambar 3.4 adalah diagram alir pembentukan *tree*. Hasil dari pembentukan *tree* adalah node pada posisi *root*. Dan pada gambar 3.5 menunjukkan diagram alir dari jumlah kelas pasien.



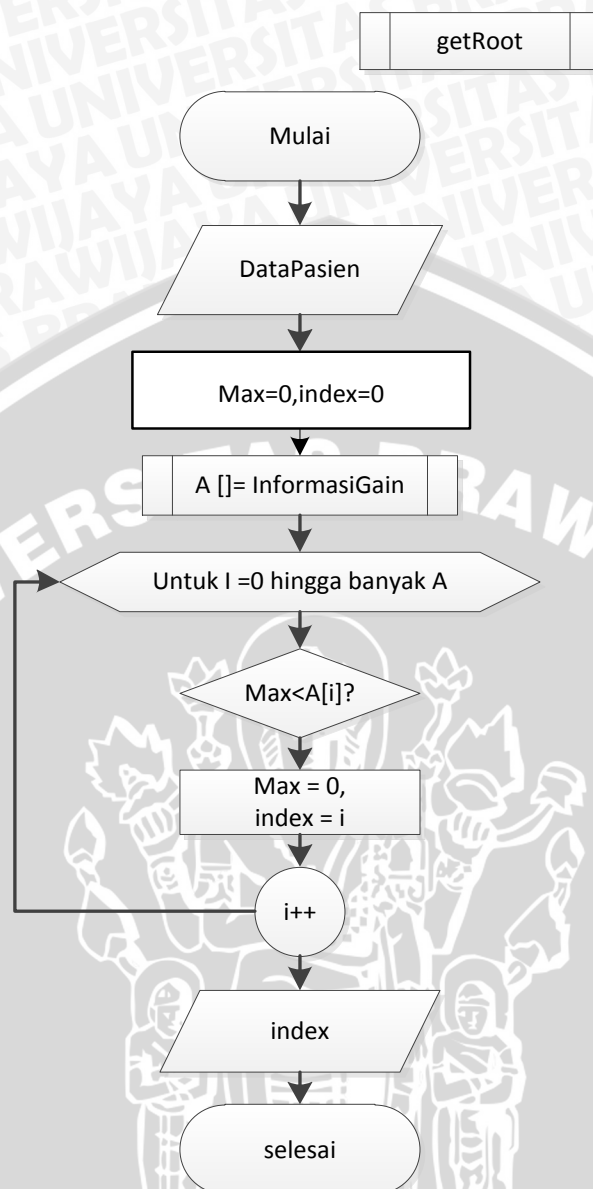
Gambar 3.4 Diagram Alir Pembentukan Tree



Gambar 3.5 Diagram Alir Jumlah Kelas

3.2.4.3 Penentuan Parent

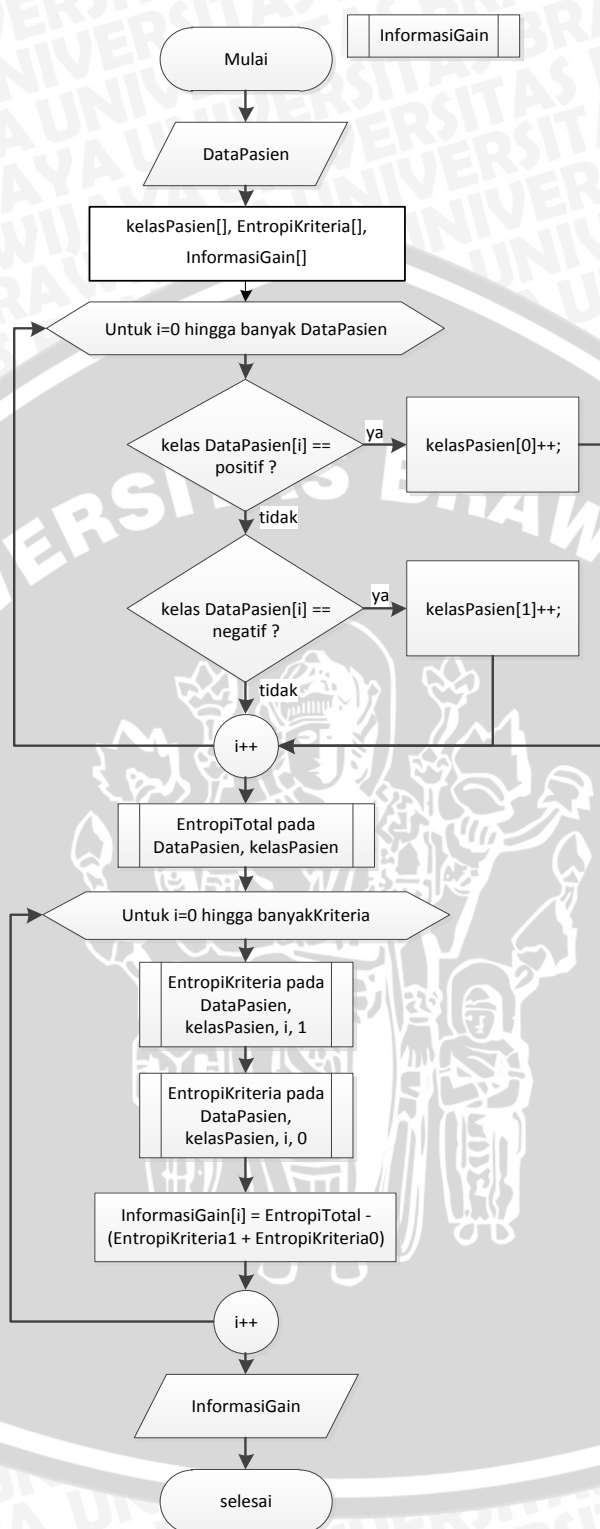
Penentuan *parent* adalah proses untuk menentukan gejala yang dipilih berdasarkan *information gain*. *Informasi Gain* dihitung pada masing - masing gejala berdasarkan data pasien pada saat itu. Gejala yang dipilih untuk node *parent* adalah gejala yang memiliki *Informasi Gain* terbesar. gambar 3.6 adalah digram alir untuk penentuan *parent*.



Gambar 3.6 Diagram Alir Penentuan Parent

3.2.4.4 Perhitungan Information Gain

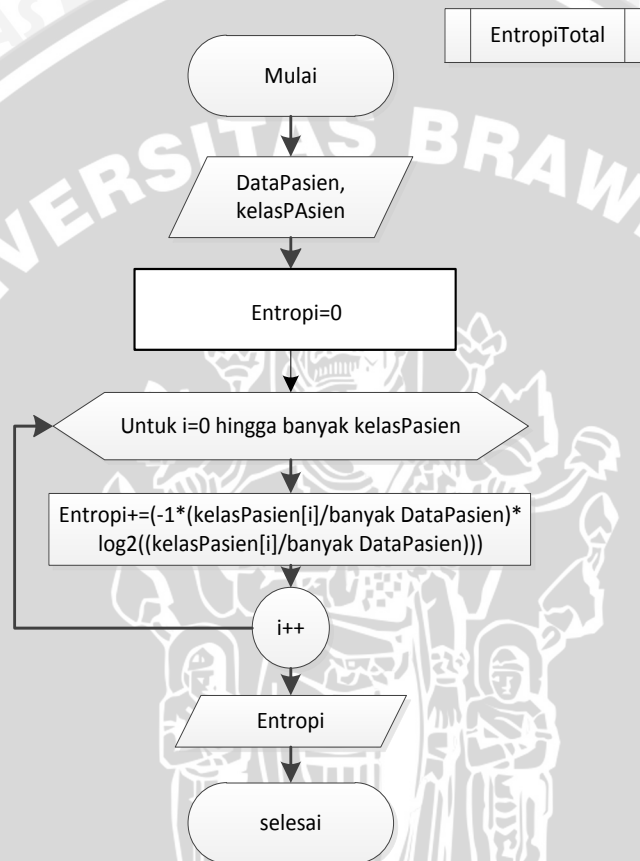
Information Gain dihitung dari pengurangan entropi total seluruh gejala dengan entropi masing - masing nilai gejala pada variabel gejala yang diukur. Pengukuran variabel gejala dilakukan untuk semua gejala hingga didapatkan *Informasi Gain* dari setiap gejala. Gambar 3.7 adalah diagram alir perhitungan *informasi gain* untuk keseluruhan gejala.



Gambar 3.7 Diagram alir Perhitungan Informasi Gain

3.2.4.5 Perhitungan Entropi Total

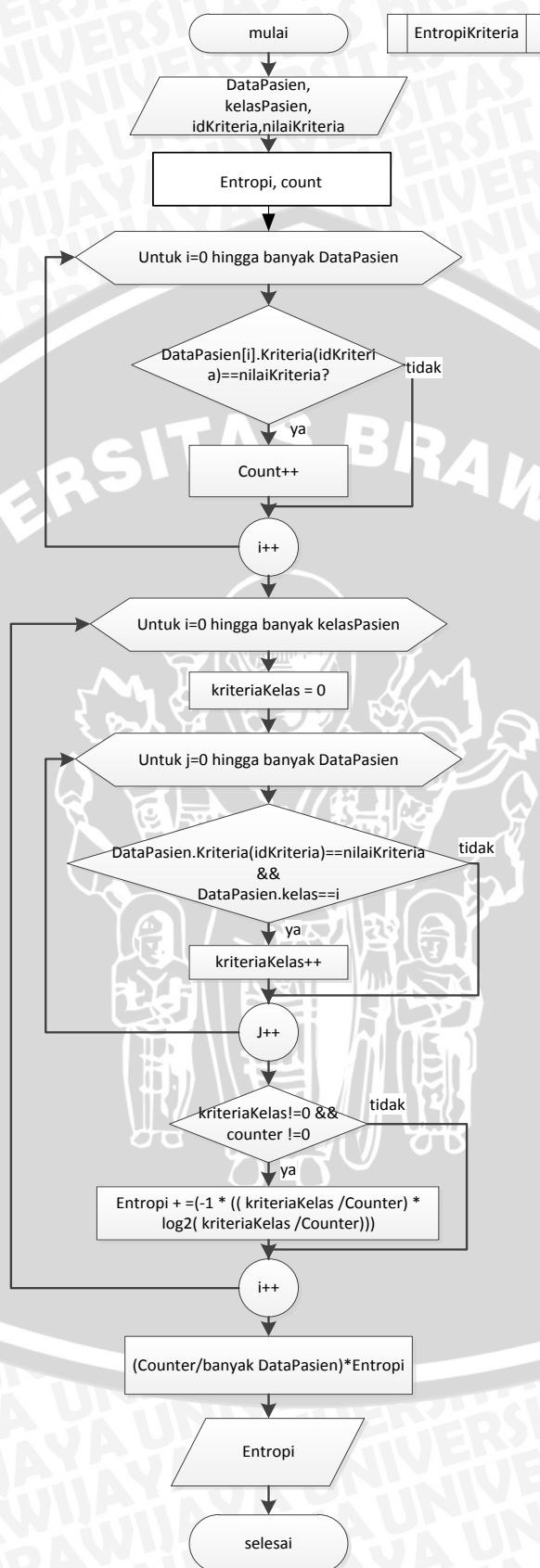
Proses penentuan gejala untuk digunakan sebagai *parent* ditentukan dengan membandingkan nilai *Information Gain*. *Information Gain* didapatkan dari pengurangan entropi total dengan nilai entropi masing - masing gejala diukur. Entropi total dihitung dari jumlah nilai entropi dari masing - masing kelas. Gambar 3.8 adalah diagram alir untuk perhitungan entropi total.



Gambar 3.8 Diagram Alir Perhitungan Entropi Total

3.2.4.6 Perhitungan Entropi Gejala

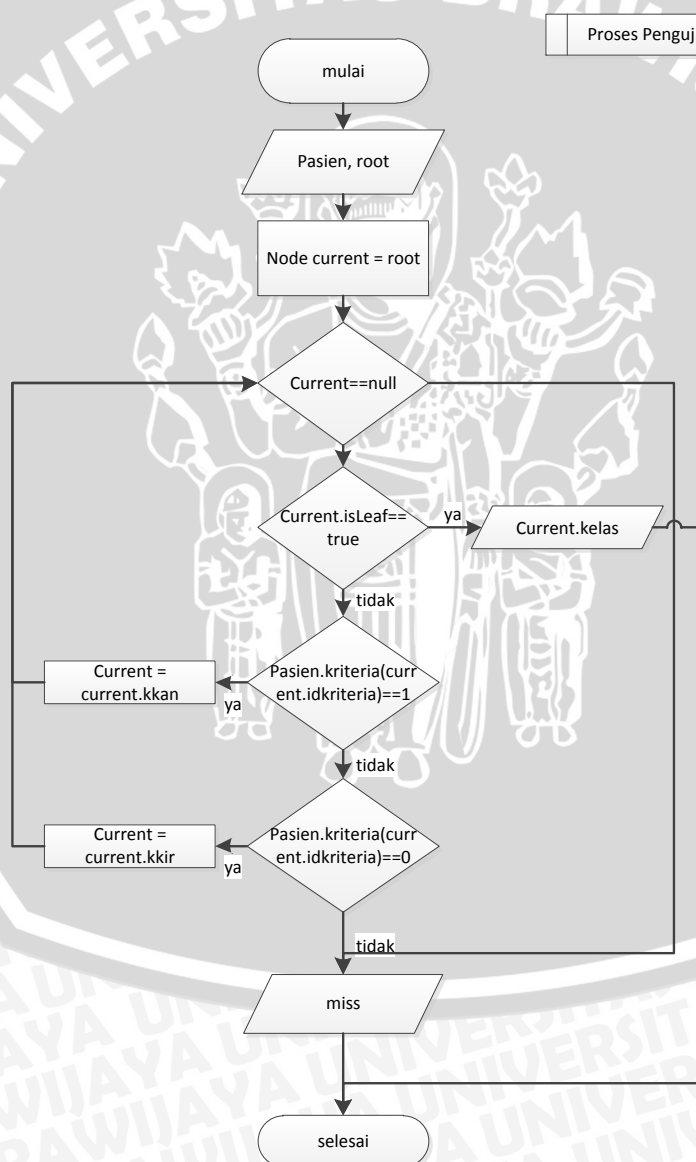
Entropi gejala adalah nilai entropi untuk nilai tertentu pada variabel gejala. Nilai entropi gejala dengan nilai tertentu dihitung dari jumlah nilai entropi pada gejala tertentu dengan nilai tertentu pada kelas 'positif' dan 'negatif'. Nilai entropi yang dikeluarkan nantinya akan dikalikan dengan nilai perbandingan banyaknya pasien pada gejala tertentu dan nilai gejala tertentu dengan jumlah seluruh pasien. Gambar 3.9 adalah diagram alir perhitungan entropi gejala.



Gambar 3.9 Diagram Alir Perhitungan Entropi Gejala

3.2.4.7 Proses Pengujian Tree

Hasil dari proses pembentukan pembentukan *tree* adalah node *root*. Alur pengujian memanfaatkan node *root* untuk ditelusuri hingga mencapai daun (*leaf*) pada *tree*. Proses pengujian dimulai dengan menelusuri node *root*, dilanjutkan ke anak kanan atau kiri menyesuaikan dengan data ujinya. Anak kanan ditelusuri jika data uji memiliki gejala tertentu sesuai node pada saat itu bernilai ‘ya’ atau 1 dan anak kiri jika bernilai ‘tidak’ atau 0. Berikutnya proses penelusuran dilanjutkan pada node anaknya. Jika gejala data uji pada saati itu bernilai miss maka hasil diagnosanya adalah miss. Gambar 3.10 adalah diagram alir proses pengujian *tree*.



Gambar 3.10 Diagram Alir Pengujian Tree

3.3 Implementasi

3.3.1 Memasukkan Data

Berdasarkan perhitungan manual ini hanya 10 data sampel yang digunakan dari 100 data yang sudah ada. Tabel 3.1 adalah contoh 10 data latih.

Keterangan :

D = Demam

P = Pusing

L = Lesu

IT = Infeksi Tenggorokan

PTE = Perut Tidak Enak

LTK = Lidah Tengah Kotor

UJM = Ujung Lidah Merah

N = Ngigau

Tabel 3.1 Contoh Data Latih

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P1	V	V	X	V	V	X	V	V	Positif
P2	V	V	X	V	V	X	V	X	Negatif
P3	V	V	V	V	X	X	V	V	Positif
P4	V	X	V	V	X	X	V	V	Positif
P5	V	V	X	V	V	X	V	V	Positif
P6	V	V	X	X	V	-	V	X	Negatif
P7	V	X	X	V	X	V	V	X	Positif
P8	X	V	X	V	X	X	V	X	Positif
P9	V	X	V	V	V	V	V	X	Positif
P10	V	X	V	X	X	X	V	V	Positif

Data rekam medis diatas merupakan data penderita penyakit Typus di Rumah Sakit Umum Besuki terdapat 10 record data pasien. diagnosa awal

penyakit Typus dengan memperhatikan 8 parameter/ atribut diagnosa. Dimana $V = \text{ya}$ dan $X = \text{tidak}$. Delapan parameter/ atribut sebagai berikut :

1. Demam dikelompokkan dalam 2 kategori (ya, tidak).
2. Pusing dikelompokkan dalam 2 kategori (ya, tidak).
3. Lesu dikelompokkan dalam 2 kategori (ya, tidak).
4. Infeksi Tenggorokan dikelompokkan dalam 2 kategori (ya, tidak).
5. Perut tidak enak dikelompokkan dalam 2 kategori (ya, tidak).
6. Lidah Tengahan Kotor dikelompokkan dalam 2 kategori (ya, tidak).
7. Ujung Lidah Merah dikelompokkan dalam 2 kategori (ya, tidak).
8. Ngigau dikelompokkan dalam 2 kategori (ya, tidak).

Pasien P6 pada data contoh pasien diatas memiliki nilai *miss* pada atribut Lidah Tengahan Kotor. Preprosesing data digunakan untuk mengisi nilai tersebut berdasarkan kedekatan data pasien tertentu dengan pasien lain yang masih dalam satu kelas. Pasien P6 memiliki kelas negatif kemudian dibandingkan dengan pasien lain (dalam kasus ini hanya pasien P2). Berikut adalah contoh perhitungan jarak dengan *Manhattan*.

$$D_{p6,p2} = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3| + |x_4 - y_4| + |x_5 - y_5| + |x_7 - y_7| + |x_8 - y_8|$$

$$D_{p6,p2} = |1 - 1| + |1 - 1| + |0 - 0| + |0 - 1| + |1 - 1| + |1 - 1| + |0 - 0|$$

$$D_{p6,p2} = 1$$

Karena tidak terdapat pasien lain dari kelas negatif maka pasien P2 dianggap mirip dengan pasien P6 dengan jarak 1, Sehingga nilai atribut Lidah Tengahan Kotor pasien P6 diberi nilai yang sama dengan atribut Lidah Tengahan Kotor pada pasien P2 yaitu 0 (tidak). Table 3.2 dibawah ini adalah hasil preprosesing data.

Tabel 3.2 Contoh Data Latih

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P1	V	V	X	V	V	X	V	V	Positif
P2	V	V	X	V	V	X	V	X	Negatif
P3	V	V	V	V	X	X	V	V	Positif
P4	V	X	V	V	X	X	V	V	Positif
P5	V	V	X	V	V	X	V	V	Positif
P6	V	V	X	X	V	X	V	X	Negatif
P7	V	X	X	V	X	V	V	X	Positif
P8	X	V	X	V	X	X	V	X	Positif
P9	V	X	V	V	V	V	V	X	Positif
P10	V	X	V	X	X	X	V	V	Positif

Tahap awal prosesing atau pembentukan pohon adalah menghitung jumlah pasien positif dan negatif, serta jumlah pasien pada setiap nilai pada setiap atribut untuk memudahkan perhitungan Entropi. Tabel 3.3 dibawah adalah contoh perhitungan jumlah pasien.

Tabel 3.3 Hasil Perhitungan Jumlah Masing Masing Atribut

Atribut	Nilai Atribut	Positif	Negatif	Jumlah
Demam	Ya	7	2	9
	Tidak	1	0	1
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
pusing	Ya	4	2	6
	Tidak	4	0	4
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
lesu	Ya	4	0	4
	Tidak	4	2	6

Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
IT	Ya	8	2	10
	Tidak	0	0	0
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
PTE	Ya	3	2	5
	Tidak	5	0	5
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
LTK	Ya	1	0	1
	Tidak	7	2	9
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
ULM	Ya	8	2	10
	Tidak	0	0	0
Total		8	2	10
Atribut	Nilai Atribut	Positif	Negatif	Jumlah
Ngigau	Ya	5	0	5
	Tidak	3	2	5
Total		8	2	10

Setelah semua didapat nilai maka setiap nilai tersebut akan di hitung nilai entropy dan information gain nya, sebagai langkah awal yang harus di hitung adalah nilai entropy total sebagai berikut:

$$Entropy(S) = P_a \log_2 P_a - P_b \log_2 P_b,$$

Pada data rekam medis, jumlah diagnosa adalah sampel 1 ('Positif') adalah 8 dan 2 ('Negatif') adalah 2. Jadi $P_1 = 8$ dan $P_2 = 2$. Dengan demikian entropy untuk Kumpulan Sampel data S adalah :

$$Entropy(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

Dari tabel diatas misal atribut diagnosa = 'Positif' merupakan sampel (+), dan atribut diagnosa = 'Negatif Typus' merupakan sampel (-), dari sampel data pada tabel didapat :

a. Values (Demam) = S_{Ya}, S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [7+, 2-], |S_{Ya}| = 9$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy}(S_{Ya}) = - (7/9) \log_2 (7/9) - (2/9) \log_2 (2/9) = 0.2819$$

$$\text{Entropy}(S_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0.4822$$

Information Gain (S_{Demam}) :

$$\begin{aligned} &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\ &= 0,7219 - (9/10)*0,2819 - (1/10)*0,4822 = 0,0341 \end{aligned}$$

b. Values (Pusing) = S_{Ya}, S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [4+, 2-], |S_{Ya}| = 6$$

$$S_{Tidak} = [4+, 0-], |S_{Tidak}| = 4$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy}(S_{Ya}) = - (4/6) \log_2 (4/6) - (2/6) \log_2 (2/6) = 0.389$$

$$\text{Entropy}(S_{Tidak}) = - (4/4) \log_2 (4/4) - (0/4) \log_2 (0/4) = 0.5283$$

Information Gain (S_{Pusing}):

$$\begin{aligned} &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\ &= 0,7219 - (6/10) 0.3899 - (4/10)(0.5283) \\ &= 0,1709 \end{aligned}$$

c. Values (Lesu) = S_{Ya}, S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [4+, 0-], |S_{Ya}| = 4$$

$$S_{Tidak} = [4+, 2-], |S_{Tidak}| = 6$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy (S}_{Ya}) = - (4/4) \log_2 (4/4) - (0/4) \log_2 (0/4) = 0,3899$$

$$\text{Entropy(S}_{Tidak}) = - (4/6) \log_2 (4/6) - (2/6) \log_2 (2/6) = 0,5283$$

Information Gain (S_{Lesu}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,7219 - (4/10)0,3899 - (6/10)0,5283$$

$$= 0,1709$$

d. Values (Infeksi Tenggorokan) = S_{Ya} , S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [7+, 1-], |S_{Ya}| = 8$$

$$S_{Tidak} = [1+, 1-], |S_{Tidak}| = 2$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy (S}_{Ya}) = - (7/8) \log_2 (7/8) - (1/8) \log_2 (1/8) = 0,2575$$

$$\text{Entropy(S}_{Tidak}) = - (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2) = 0,4643$$

Menunjukkan entropy minimum karena jumlah sampel pada salah satu kelas adalah = 0 (keberagaman data minimum).

Information Gain ($S_{Infeksi Tenggorokan}$) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,7219 - 8/10)0,2575 - (2/10)0,4643$$

$$= 0$$

e. Values (Perut Tidak Enak) = S_{Ya} , S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [3+, 2-], |S_{Ya}| = 5$$

$$S_{Tidak} = [5+, 0-], |S_{Tidak}| = 5$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy (S}_{Ya}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,4421$$

$$\text{Entropy(S}_{Tidak}) = - (5/5) \log_2 (5/5) - (0/5) \log_2 (0/5) = 0,5287$$

Information Gain ($S_{\text{Perut Tidak Enak}}$) :

$$\begin{aligned}
 &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\
 &= 0,7219 - (5/10)0,4421 - (5/10)0,5287 \\
 &= \underline{0,2364}
 \end{aligned}$$

f. Values (Lidah Tengah Kotor) = S_{Y_a} , S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Y_a} = [1+, 0-], |S_{Y_a}| = 1$$

$$S_{Tidak} = [7+, 2-], |S_{Tidak}| = 9$$

Nilai-nilai Entropy untuk S_{Y_a} , S_{Tidak} :

$$\text{Entropy}(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy}(S_{Y_a}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0,2819$$

$$\text{Entropy}(S_{Tidak}) = - (7/9) \log_2 (7/9) - (2/9) \log_2 (2/9) = 0,4822$$

Information Gain ($S_{\text{Lidah Tengah Kotor}}$) :

$$\begin{aligned}
 &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\
 &= 0,7219 - (1/10)0,2819 - (9/10)0,4822 = 0,0341
 \end{aligned}$$

g. Values (Ujung Lidah Merah) = S_{Y_a} , S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Y_a} = [8+, 2-], |S_{Y_a}| = 10$$

$$S_{Tidak} = [0+, 0-], |S_{Tidak}| = 0$$

Nilai-nilai Entropy untuk S_{Y_a} , S_{Tidak} :

$$\text{Entropy}(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy}(S_{Y_a}) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,2575$$

$$\text{Entropy}(S_{Tidak}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0,4643$$

Information Gain ($S_{\text{Ujung Lidah merah}}$) :

$$\begin{aligned}
 &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\
 &= 0,7219 - (10/10)0,2572 - (0/10)0,4643 \\
 &= 0
 \end{aligned}$$

h. Values (Ngigau) = S_{Y_a} , S_{Tidak}

$$S = [8+, 2-], |S| = 10$$

$$S_{Ya} = [5+, 0-], |S_{Ya}| = 5$$

$$S_{Tidak} = [3+, 2-], |S_{Tidak}| = 5$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy}(S) = - (8/10) \log_2 (8/10) - (2/10) \log_2 (2/10) = 0,7219$$

$$\text{Entropy}(S_{Ya}) = - (5/5) \log_2 (5/5) - (0/5) \log_2 (0/5) = 0.4421$$

$$\text{Entropy}(S_{Tidak}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,5287$$

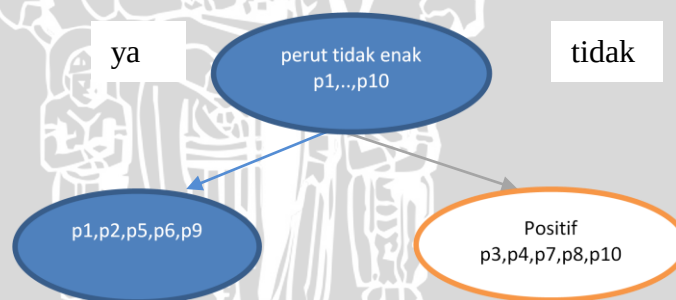
Information Gain (S_{Ngigau}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,7219 - (5/10)0,7219 - (5/10)0,421$$

$$= 0,2364$$

Dari perhitungan diatas diperoleh gain tertinggi yaitu pada atribut perut tidak enak sebesar 0,2364, sehingga atribut ‘perut tidak enak’ menjadi root pada tree. Setelah root diperoleh, maka selanjutnya dilakukan pengecekan apakah semua parameter pada atribut “perut tidak enak” sudah terklasifikasi dengan sempurna. Atribut “perut tidak enak” beserta turunan parameternya dapat dilihat pada gambar 3.11.



Gambar 3.11 Root Atribut “perut tidak enak” dan Parameter Turunannya.

Kemudian data dipilih dan dilanjutkan dengan pada atribut ‘perut tidak enak’ pada nilai ‘ya’ hasilnya dapat dilihat pada Tabel 3.4.

Tabel 3.4 Hasil Filter Atribut ‘perut tidak enak’ dengan Nilai ‘ya’

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P1	V	V	X	V	V	X	V	V	Positif
P2	V	V	X	V	V	X	V	X	Negatif
P5	V	V	X	V	V	X	V	V	Positif
P6	V	V	X	X	V	X	V	X	Negatif
P9	V	X	V	V	V	V	V	X	Positif

$$\text{Entropy (S)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

Dari tabel diatas misal atribut diagnosa = ‘Positif’ merupakan sampel (+), dan atribut diagnosa = ‘Negatif Typus’ merupakan sampel (-), dari sampel data pada tabel didapat :

- Values (Demam) = S_{Ya}, S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [3+, 2-], |S_{Ya}| = 5$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy (S)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy (S}_{Ya}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,4421$$

$$\text{Entropy (S}_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0,5287$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,9709 - (3/5)*0,4421 - (2/5)*0,5287 = 0$$

- Values (Pusing) = S_{Ya}, S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [2+, 2-], |S_{Ya}| = 4$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy (S}_{Ya}) = - (2/5) \log_2 (2/5) - (2/5) \log_2 (2/5) = 0.5$$

$$\text{Entropy (S}_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0.5$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|Sv|}{|S|} \text{Entropy (Sv)}$$

$$= 0,9709 - (4/5)*0,5 - (1/5)*0,5 = 0.8$$

- Values (Lesu) = S_{Ya} , S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [1+, 0-], |S_{Ya}| = 1$$

$$S_{Tidak} = [2+, 2-], |S_{Tidak}| = 4$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy (S}_{Ya}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0.5$$

$$\text{Entropy (S}_{Tidak}) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 0.5$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|Sv|}{|S|} \text{Entropy (Sv)}$$

$$= 0,9709 - (1/5)*0,5 - (4/5)*0,5 = 0.8$$

- Values (Infeksi tenggorokan) = S_{Ya} , S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [3+, 2-], |S_{Ya}| = 5$$

$$S_{Tidak} = [0+, 0-], |S_{Tidak}| = 0$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy (S}_{Ya}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0.4421$$

$$\text{Entropy (S}_{Tidak}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0.5287$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|Sv|}{|S|} \text{Entropy (Sv)}$$

$$= 0,9709 - (5/5)*0,4421 - (0/5)*0,5287 = 0.$$

- Values (Lidah Tengah Kotor) = S_{Ya}, S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [0+, 0-], |S_{Ya}| = 0$$

$$S_{Tidak} = [3+, 2-], |S_{Tidak}| = 5$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy}(S_{Ya}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0.4421$$

$$\text{Entropy}(S_{Tidak}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0.5287$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (0/5)*0,4421 - (5/5)*0,5287 = 0$$

- Values (Ujung Lidah merah) = S_{Ya}, S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [3+, 2-], |S_{Ya}| = 5$$

$$S_{Tidak} = [0+, 0-], |S_{Tidak}| = 0$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy}(S_{Ya}) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0.4421$$

$$\text{Entropy}(S_{Tidak}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0.5287$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (5/5)*0,4421 - (0/5)*0,5287 = 0$$

- Values (Ngigau) = S_{Ya}, S_{Tidak}

$$S = [3, 2-], |S| = 5$$

$$S_{Ya} = [2+, 0-], |S_{Ya}| = 2$$

$$S_{Tidak} = [1+, 2-], |S_{Tidak}| = 3$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (3/5) \log_2 (3/5) - (2/5) \log_2 (2/5) = 0,9709$$

$$\text{Entropy}(S_{Ya}) = - (2/2) \log_2 (2/2) - (0/2) \log_2 (0/2) = 0.5283$$

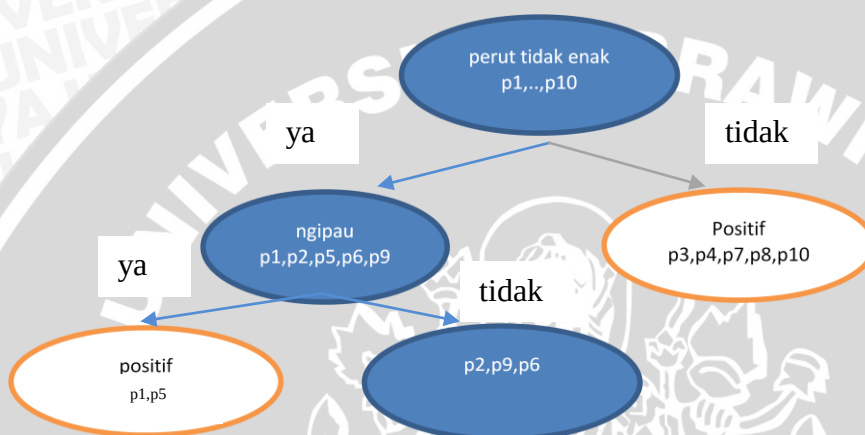
$$\text{Entropy}(S_{Tidak}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0.3899$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (2/5) * 0,5283 - (3/5) * 0,3899 = \mathbf{0.4199}$$

Dari perhitungan diatas, Gambar 3.12 menunjukkan child dari parameter 2 yaitu atribut “ngigau” dengan information gain terbesar.



Gambar 3.12 Root Atribut Ngigau

Kemudian data dipilih kembali dan dilanjutkan dengan data pada atribut ‘ngigau’ pada nilai ‘tidak’ hasilnya dapat dilihat pada Tabel 3.5.

Tabel 3.5 Hasil Filter Atribut ‘ngigau’ dengan Nilai ‘tidak’

Pasien	Gejala								Typus
	D	P	L	IT	PTE	LTK	UJM	N	
P2	V	V	X	V	V	X	V	X	Negatif
P6	V	V	X	X	V	X	V	X	Negatif
P9	V	X	V	V	V	V	V	X	Positif

$$\text{Entropy}(S) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

Dari tabel diatas misal atribut diagnosa = ‘Positif’ merupakan sampel (+), dan atribut diagnosa = ‘Negatif Typus’ merupakan sampel (-), dari sampel data pada tabel didapat :

- Values (Demam) = S_{Ya}, S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [1+, 2-], |S_{Ya}| = 3$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy}(S_{Ya}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,5823$$

$$\text{Entropy}(S_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0,3899$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (3/3)*0,5823 - (1/3)*0,3899 = 0$$

- Values (Pusing) = S_{Ya}, S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [2+, 0-], |S_{Ya}| = 2$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya}, S_{Tidak} :

$$\text{Entropy}(S) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy}(S_{Ya}) = - (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) = 0,$$

$$\text{Entropy}(S_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0,$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (2/3)*0 - (1/3)*0 = 0,9182$$

- Values (Lesu) = S_{Ya}, S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [1+, 0-], |S_{Ya}| = 1$$

$$S_{Tidak} = [0+, 2-], |S_{Tidak}| = 2$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy (S}_{Ya}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0$$

$$\text{Entropy (S}_{Tidak}) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,9709 - (1/3)*0 - (2/3)*0 = 0,9182$$

- Values (Infeksi Tenggorokan) = S_{Ya} , S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [1+, 2-], |S_{Ya}| = 3$$

$$S_{Tidak} = [0, 0-], |S_{Tidak}| = 0$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy (S}_{Ya}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,5283$$

$$\text{Entropy (S}_{Tidak}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0,3899$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,9709 - (3/3)*0,5283 - (1/0)*0,3899 = 0$$

- Values (Lidah tengah kotor) = S_{Ya} , S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [0+, 0-], |S_{Ya}| = 0$$

$$S_{Tidak} = [1+, 2-], |S_{Tidak}| = 3$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy (S)} = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy (S}_{Ya}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0,5823$$

$$\text{Entropy (S}_{Tidak}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,3899$$

Information Gain (S_{Demam}) :

$$= \text{Entropy (S)} - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy (S}_v)$$

$$= 0,9709 - (0/0)*0,5823 - (1/3)*0,3899 = 0$$

- Values (Ujung Lidah merah) = S_{Ya} , S_{Tidak}

$$S = [1, 2-], |S| = 3$$

$$S_{Ya} = [1+, 2-], |S_{Ya}| = 3$$

$$S_{Tidak} = [1+, 0-], |S_{Tidak}| = 1$$

Nilai-nilai Entropy untuk S_{Ya} , S_{Tidak} :

$$\text{Entropy}(S) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,9182$$

$$\text{Entropy}(S_{Ya}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0.5823$$

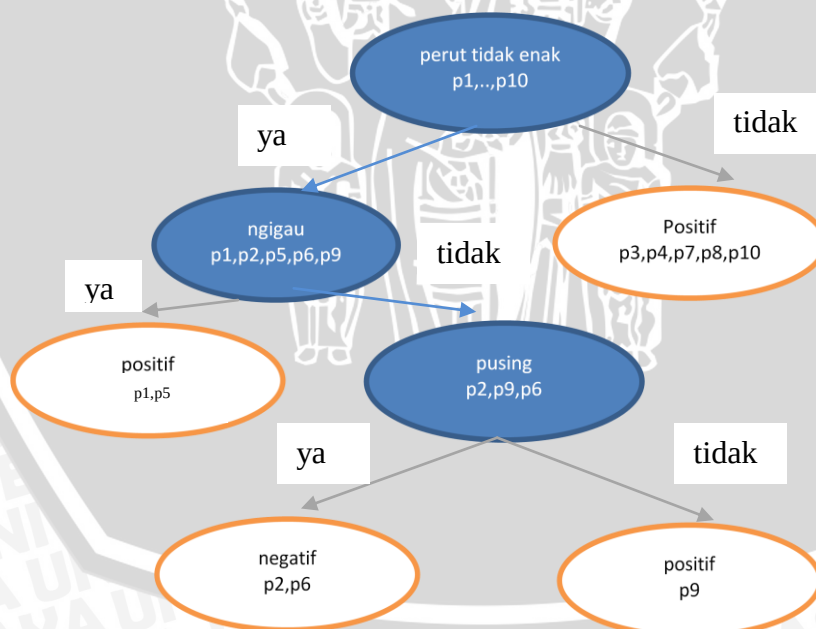
$$\text{Entropy}(S_{Tidak}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0.3899$$

Information Gain (S_{Demam}) :

$$= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9709 - (3/3)*0,5823 - (1/3)*0,3899 = 0$$

Dari perhitungan diatas, Gambar 3.13 menunjukkan child dari parameter 3 yaitu atribut “pusing” dengan *information gain* terbesar.



Gambar 3.13 Root Atribut Pusing

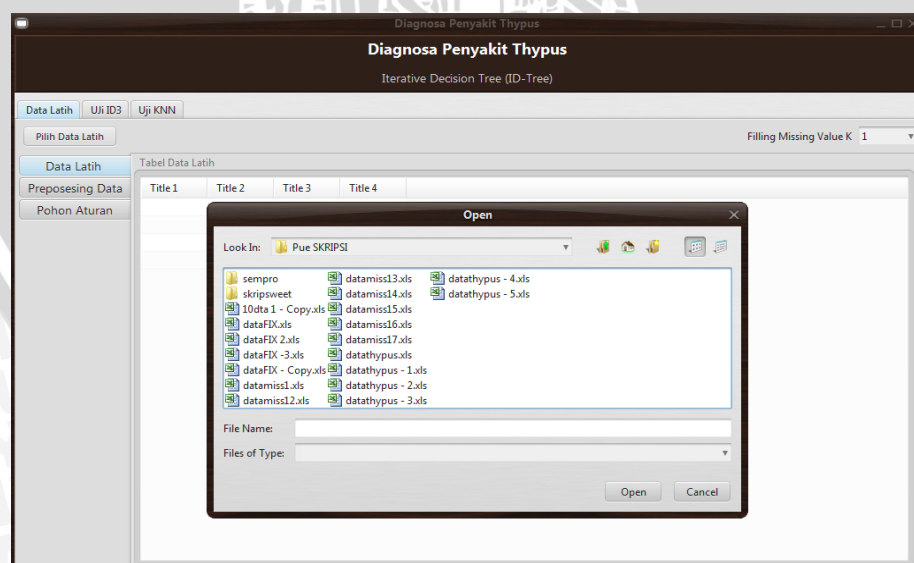
3.4 Desain Interface Sistem

Pada Gambar 3.14 merupakan halaman pertama dari dari aplikasi diagnosa penyakit thypus yang terdiri dari button data latih untuk proses pelatihan dan pembtukan *tree*, data uji untuk menentukan prediksi kelas pasien dan mengukur akurasi metode.



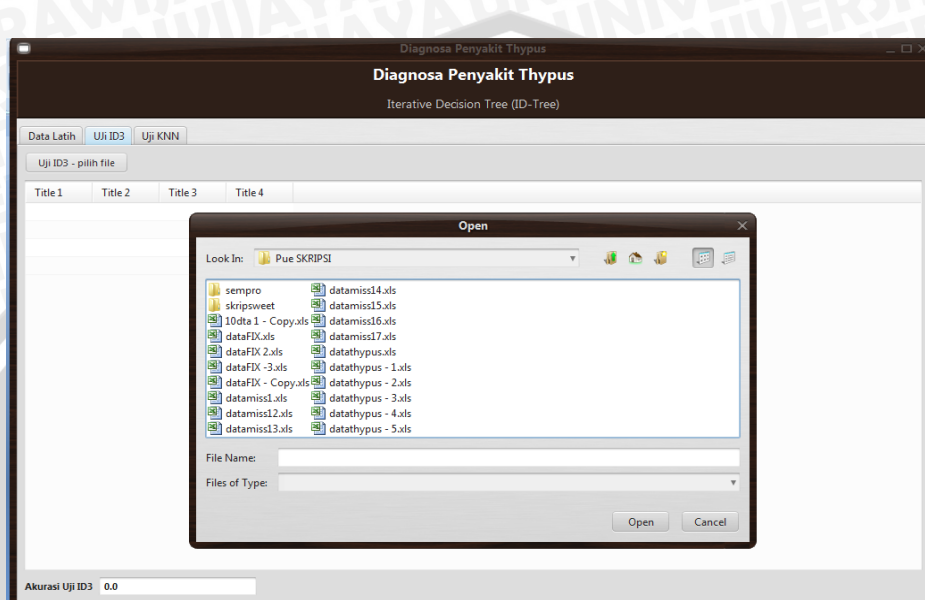
Gambar 3.14 Desain Interface

Untuk memilih data latih yang akan digunakan maka dapat dilihat pada Gambar 3.15. Format file yang dapat digunakan adalah “.xls” pada tombol Pilih Data Latih.



Gambar 3.15 Tampilan pemilihan data latih

Data latih yang akan digunakan dipilih pada tombol data uji dengan format file yang dapat digunakan adalah “.xls” pada tombol Pilih Data Uji seperti pada Gambar 3.16 .



Gambar 3.16 Tampilan pemilihan data uji