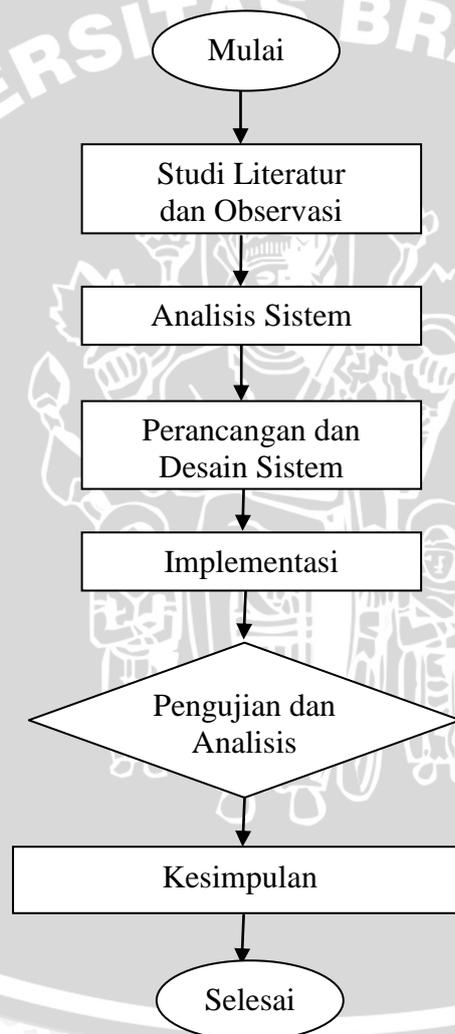


BAB III

METODOLOGI DAN PERANCANGAN SISTEM

3.1 Metodologi Penelitian

Dalam pembuatan tugas akhir skripsi ini menggunakan beberapa metode untuk merancang sebuah sistem yang dapat mendiagnosa awal penyakit asma yang dapat dilihat pada Gambar 3.1.



Gambar 3.1 Flowchart Metodologi Perancangan

Berikut merupakan beberapa langkah – langkah dari metodologi perancangan sistem yang digunakan pada penelitian ini:

1. Pengumpulan data – data yang diperlukan :

a. Studi literatur

Dalam studi literatur dilakukan dengan mempelajari dan mengumpulkan data – data dari berbagai macam literatur seperti buku serta *website* yang memberikan informasi yang relevan untuk membantu penyusunan skripsi ini.

Hal – hal yang menjadi bahan studi literatur antara lain :

- Penyakit asma
- Konsep klasifikasi
- Konsep dan teori dari algoritma ID3
- Konsep *Case Based Reasoning* (CBR)

b. Observasi

Metode observasi dilakukan dengan cara melakukan pengamatan langsung terhadap data yang diteliti. Penelitian ini menggunakan data sekunder, yaitu data yang sudah ada dan diambil dari Rumah Sakit Umum Situbondo. Data yang digunakan berjumlah 100 data dan terdapat 6 kriteria yang dijadikan sebagai atribut yaitu Batuk, Dahak, Darah, Sesak Napas, Nyeri Dada, dan Demam. Selain itu terdapat dua jenis diagnosa yang terdapat pada data rekam medis tersebut yaitu Asma Bronkial dan Bronkitis Akut.

2. Analisis sistem

Analisis sistem adalah proses analisis terhadap permasalahan dan mendefinisikan model penyelesaian, termasuk melakukan analisis terhadap spesifikasi perangkat lunak yang akan dibangun. Perancangan sistem dilakukan dengan mengolah data yang telah diperoleh dari hasil observasi dengan menjadikannya lebih terstruktur dan dilakukan pembentukan aturan atau *rule* ID3 dan dikembangkan pada sistem CBR untuk menambah pengetahuan baru bagi sistem yang nantinya akan digunakan pada aplikasi yang akan dibuat.

3. Perancangan dan desain sistem

Pada tahap ini dilakukan perancangan sistem dengan melakukan pemodelan sistem yaitu dengan merancang *flowchart* yang nantinya akan membantu proses selanjutnya.

4. Implementasi

Pada tahap ini dilakukan pembuatan aplikasi sesuai dengan desain sistem yang telah dirancang pada tahap sebelumnya ke dalam kode – kode program sesuai dengan algoritma ID3 untuk mendeteksi dini penyakit asma pada sistem CBR yang dikembangkan dengan bahasa java.

5. Pengujian dan analisis

Pengujian dilakukan pada sistem secara keseluruhan untuk memastikan bahwa sistem yang dibuat sudah benar sesuai dengan semua karakteristik yang telah ditetapkan sehingga tidak terjadi kesalahan.

6. Kesimpulan

Penarikan kesimpulan dilakukan berdasarkan analisis hasil yang didapatkan setelah uji coba dan evaluasi yang dilakukan untuk menyelesaikan permasalahan yang telah dibahas dan di selesaikan melalui laporan ini.

3.2 Analisa Kebutuhan Sistem

Tahap ini digunakan untuk mengetahui permasalahan sehingga dapat dilakukan pemodelan terhadap penyelesaian. Setelah perangkat lunak ditetapkan selanjutnya dilakukan perancangan sistem yang meliputi pengolahan data. Perancangan arsitektur sistem deteksi dini penyakit asma menggunakan algoritma ID3 pada sistem CBR. Secara keseluruhan yang dibutuhkan sistem dalam pembuatan sistem ini meliputi:

1. Kebutuhan *Hardware* :

- Komputer PC atau Laptop

2. Kebutuhan *Software* :

- *Microsoft Windows 7 Ultimate* sebagai sistem operasi
- Netbeans IDE 7.0.1 untuk mengembangkan aplikasi
- JDK 7.0.1 development kit untuk menulis program java
- Ms. Excel

3. Data yang dibutuhkan :

- Data pasien penderita asma yang berjumlah 100 data

3.3 Perancangan Sistem

3.3.1 Deskripsi Umum Sistem

Sesuai dengan kebutuhan sistem yang diterapkan, akan digunakan dua parameter *input* yaitu data pasien dan data hasil diagnosa pasien. User akan memasukkan nilai dari variabel Batuk, Dahak, Darah, Sesak Napas, Nyeri Dada, dan Demam yang akan menjadi inputan dalam proses perhitungan *Information Gain* dan *Entropy*. Hasil hitung *Information Gain* dan *Entropy* akan dilanjutkan dalam proses pengambilan keputusan dengan menggunakan algoritma ID3 hasilnya akan menjadi *output* terhadap *user* berupa hasil diagnosa sistem yaitu asma bronkial atau bronkitis akut.

Hasil diagnosa yang muncul akan menjadi pengetahuan yang akan disimpan di *case base* untuk selanjutnya digunakan jika ada kasus baru yang mirip dengan kasus sebelumnya. Metode yang digunakan untuk permasalahan tersebut adalah metode *Case Based Reasoning* (CBR). Dengan penambahan metode *Case Based Reasoning* pada ID3 maka sistem akan memiliki kelebihan yaitu mampu menyimpan kasus lama dengan solusi yang dapat digunakan pada permasalahan baru jika ada kemiripan, sehingga dapat memberikan solusi dengan cepat dan efisien serta memiliki akurasi yang baik.

3.3.2 Batasan Sistem

Sistem ini akan dikembangkan dengan batasan – batasan :

1. Tida ada data *missing value*
2. Tidak ada hasil yang salah pada data latih dan data uji

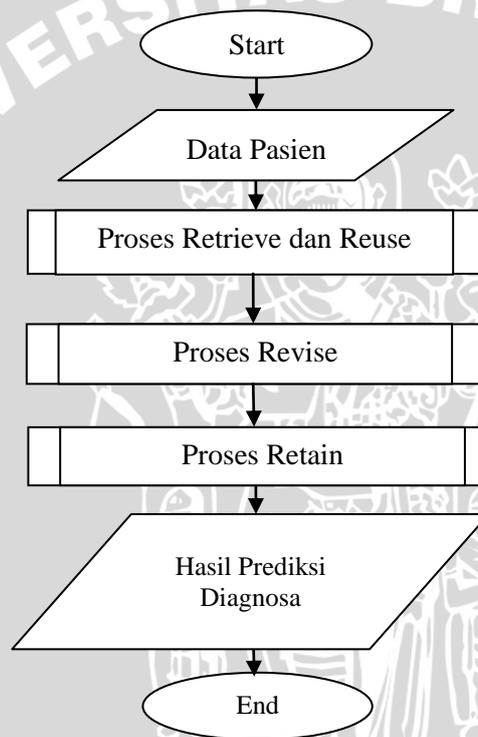
3.3.3 Case Based Reasoning (CBR)

Gambar 3.2 menjelaskan beberapa tahapan pada metode *Case Based Reasoning* (CBR). Beberapa tahapan tersebut akan dijelaskan pada *flowchart* Gambar 3.2 :

1. Proses memasukkan data pasien
2. Proses *Retrieve* yaitu melihat permasalahan yang ada dengan mengidentifikasi kasus lama apakah ada kemiripan dengan kasus baru atau tidak. Tetapi proses *retrieve* bisa terjadi jika pengetahuan dari ID3 sudah terbentuk sebelumnya.

Sedangkan pada proses reuse sistem akan menggunakan informasi dari kasus lama untuk menyelesaikan kasus baru.

3. Proses *Revise* yaitu solusi yang telah ditemukan dari kasus baru yang muncul akan dievaluasi dan dikonfirmasi untuk menentukan apakah sistem telah benar dalam mengidentifikasi dan memberikan solusi.
4. Proses *Retain* yaitu mengambil hasil konfirmasi dari sebuah solusi untuk dijadikan sebagai pengetahuan baru. Dalam proses ini terdapat proses ID3 dengan membentuk pohon aturan untuk selanjutnya diperbaharui jika ada kasus baru.



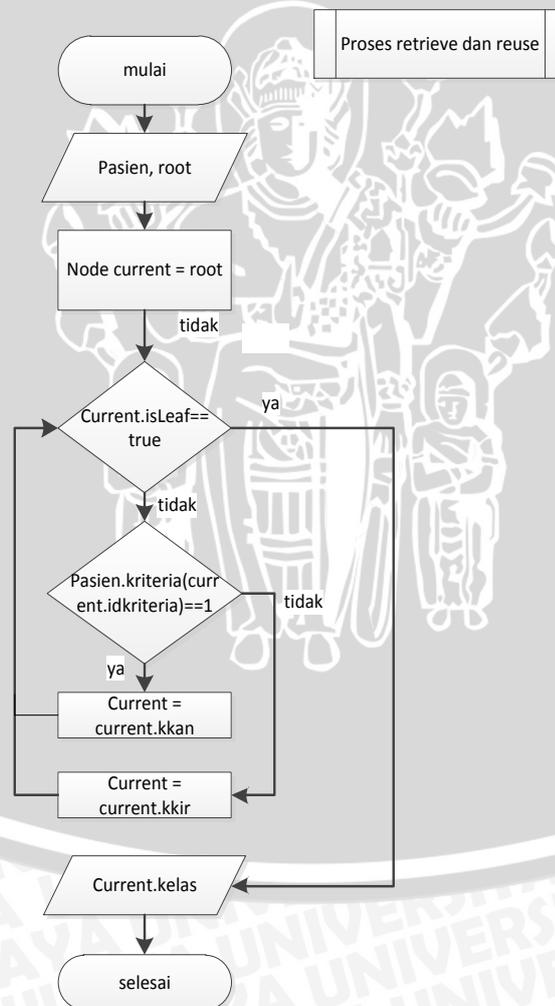
Gambar 3.2 Flowchart Metode *Case Based Reasoning*

3.3.4 Iterative Dichotomiser Tree (ID3)

Algoritma *Iterative Dichotomiser 3* (ID3) merupakan algoritma *decision tree* yang membangun *decision tree* secara *top-down* dengan mengecek atribut mana yang harus diletakkan pada *root* dengan cara menghitung nilai *entropy* dan *information gain* untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data.

3.3.4.1 Proses Retrieve dan Reuse

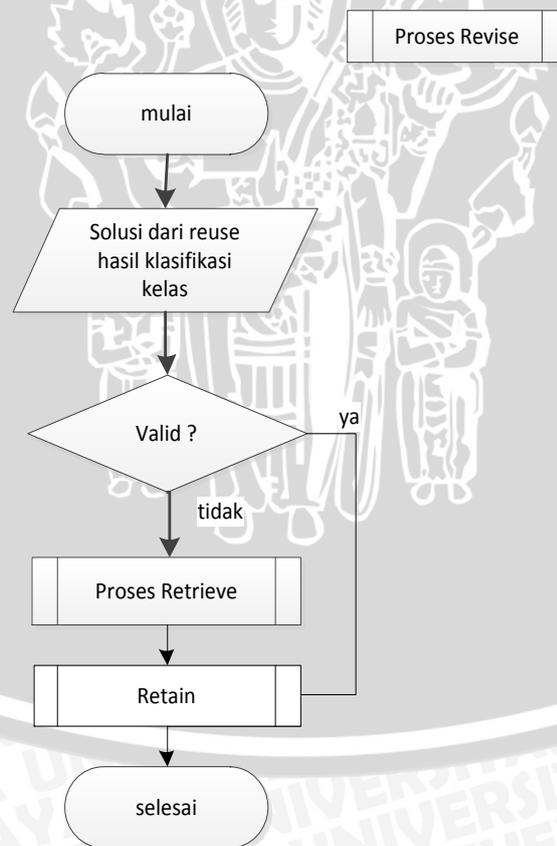
Proses *retrieve* dapat terjadi jika pengetahuan ID3 sudah terbentuk sebelumnya. Sedangkan proses *reuse* adalah menggunakan pohon keputusan yang telah dibentuk untuk mengklasifikasikan data pasien baru yang akan digunakan. Proses klasifikasi dilakukan dengan menelusuri aturan yang telah dibentuk dan mencari kecocokan berdasarkan aturan yang ada untuk mendapatkan hasil diagnosa berdasarkan gejala pada pasien baru yang diuji cobakan. Pencocokan dilakukan dengan membandingkan nilai pada gejala tertentu di tingkatan secara terurut pada *tree*. Gambar 3.3 adalah *Flowchart* proses *retrieve* dan *reuse* atau proses mendapatkan hasil dari gejala yang dimasukkan untuk menentukan kelasnya.



Gambar 3.3 *Flowchart* Penentuan Kelas

3.3.4.2 Proses *Revise*

Solusi yang telah ditemukan dari kasus baru yang muncul akan dievaluasi oleh pakar atau admin dan dikonfirmasi untuk menentukan apakah sistem telah benar dalam mengidentifikasi dan memberikan solusi yang benar menurut pakar atau admin. Proses *Revise* adalah proses evaluasi dan konfirmasi untuk menentukan apakah pohon aturan sebelumnya telah dapat diterapkan untuk menyelesaikan kasus yang baru. Admin akan melakukan proses evaluasi setiap 3 hari sekali. Jadi proses evaluasi dan konfirmasi akan dilakukan sekaligus pada beberapa kasus. Jika ‘tidak’ adalah konfirmasi dari pakar atau admin maka sistem akan melakukan proses *retain* untuk membentuk pohon aturan baru dari penggabungan data pasien baru dengan data pasien yang telah ada dan disimpan untuk digunakan pada kasus berikutnya. Gambar 3.4 merupakan *flowchart* dari proses *revise*.



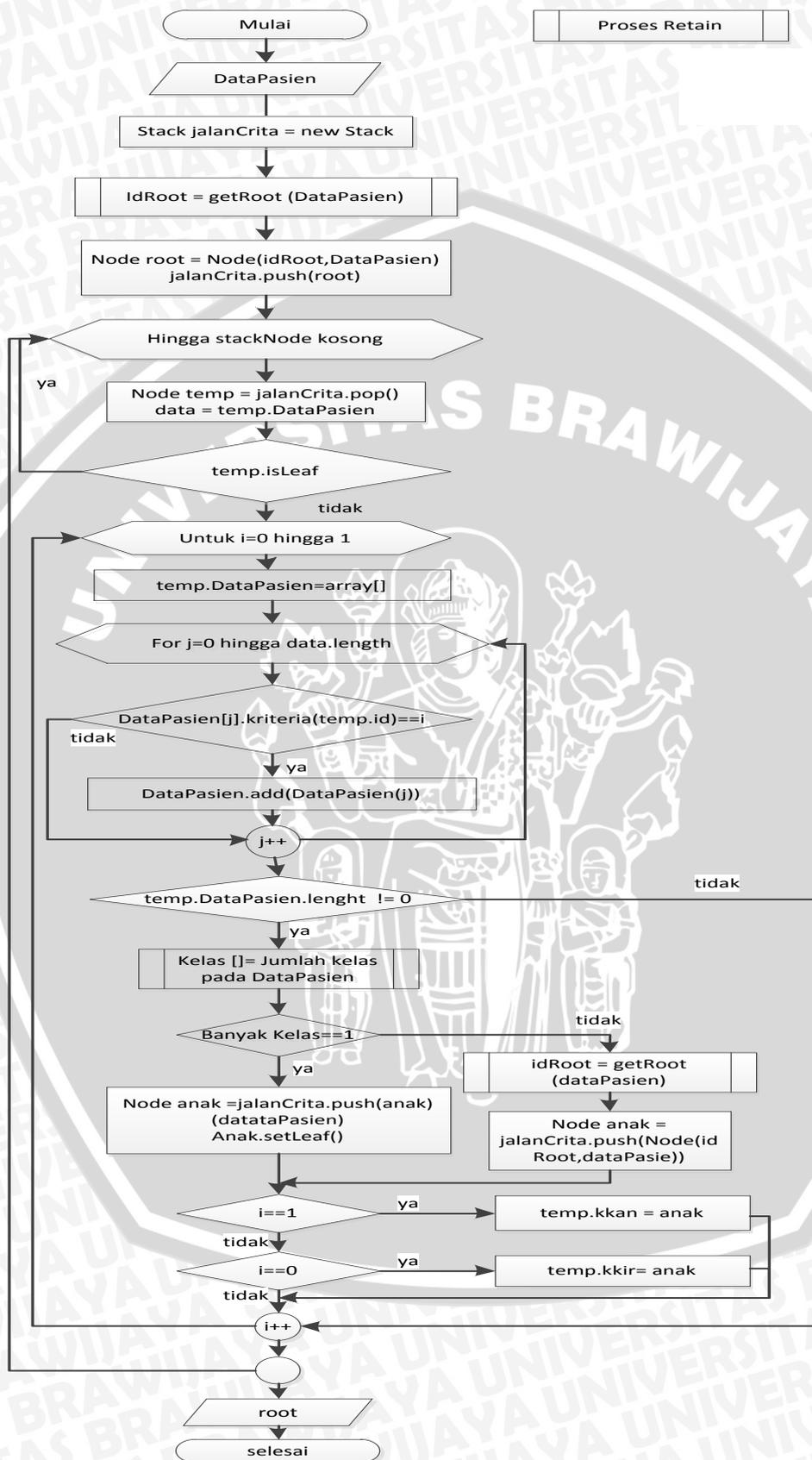
Gambar 3.4 Flowchart *Revise*

3.3.4.3 Proses *Retain*

Pada proses *retain* terjadi pengambilan hasil yang telah valid dari proses *revise* sebagai pengetahuan baru dan disimpan sehingga pohon aturan akan berubah untuk selanjutnya digunakan menyelesaikan kasus baru yang memiliki kemiripan di masa mendatang. Pohon aturan baru dibentuk dari penggabungan data pasien baru dengan data pasien yang telah ada dan disimpan untuk digunakan pada kasus berikutnya. Terdapat beberapa tahap yaitu pembentukan *tree*, penentuan *root*, perhitungan *entropy* gejala, perhitungan *entropy* total, dan perhitungan *information gain*.

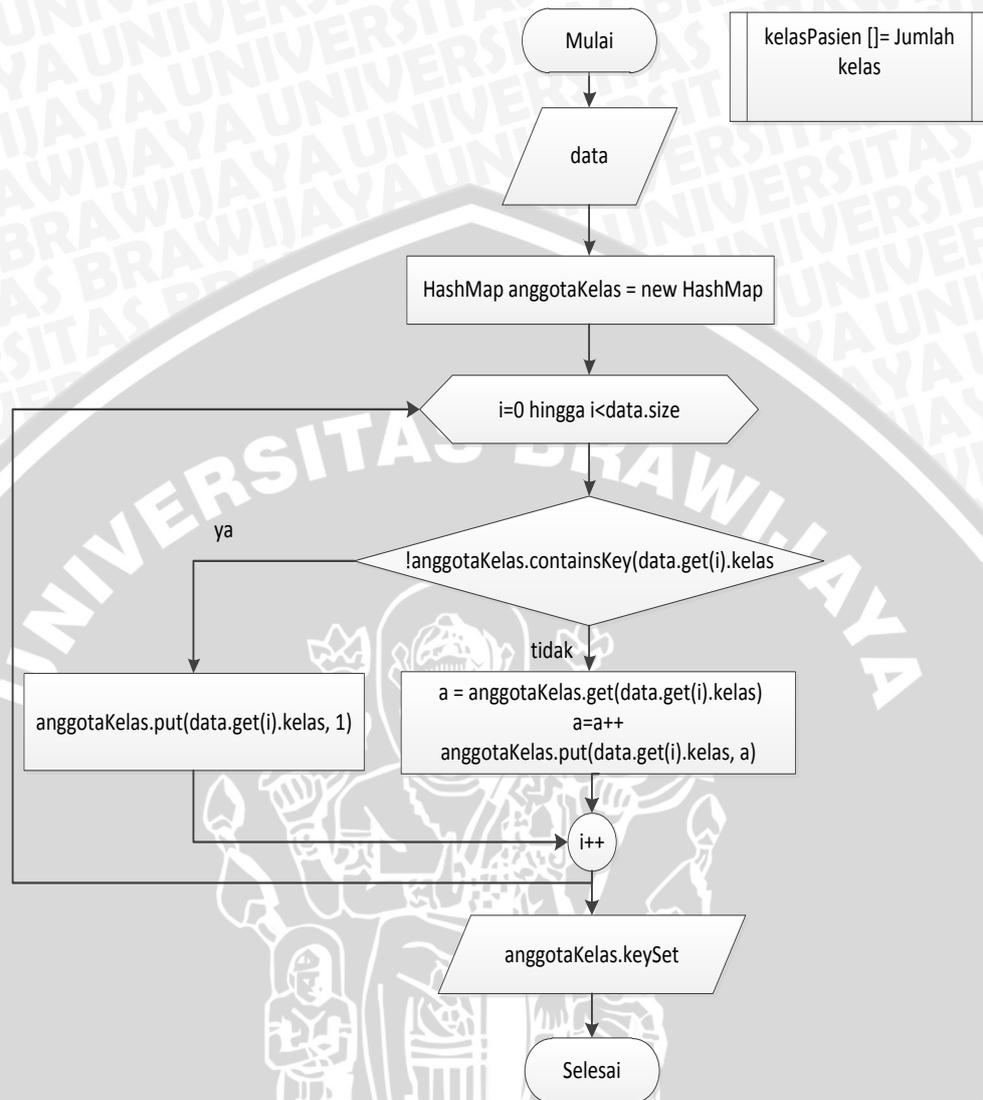
Proses awal yaitu proses pembentukan *tree* yang menjadi proses utama dari metode ID3. Setiap *node* yang dipilih sebagai *root* adalah nilai *information gain* terbesar yang didapatkan dari *entropy* gejala yang diderita pasien. Proses dimulai dengan mendapatkan *information gain* dari setiap gejala pasien, kemudian dipilih yang terbesar untuk ditempatkan sebagai *root* dari *tree* yang akan dibentuk. Struktur data yang dipilih adalah *stack* atau tumpukan untuk menyelesaikan pembentukan *tree* dimulai dari yang paling dalam. Hasil penentuan *root* yang dilakukan sebelumnya dimasukkan ke dalam *stack* dan dilakukan perulangan hingga *stack* kosong, dimana setiap perulangan *stack* akan diambil satu persatu.

Penentuan *node* berikutnya dimulai dengan memilah data pasien sesuai dengan nilai pada atribut gejala yang dipilih sebagai *root*. Nilai 'ya' untuk anak sebelah kanan dan nilai 'tidak' untuk anak sebelah kiri. Hasil filtrasi masing - masing data pasien kemudian dihitung kembali *information gain*nya dan ditentukan gejala dengan nilai *entropy* tertinggi untuk dipilih sebagai *node* berikutnya dengan memasukkannya ke dalam *stack*. Suatu *node* dipilih sebagai daun dan tidak lagi dimasukan ke dalam *stack* apabila data hasil filtrasi memiliki kelas yang sama. Gambar 3.5 adalah *flowchart* pembentukan *tree*. Hasil dari pembentukan *tree* adalah *node* pada posisi *root*. Dan pada Gambar 3.6 menunjukkan *flowchart* dari kelas Pasien.



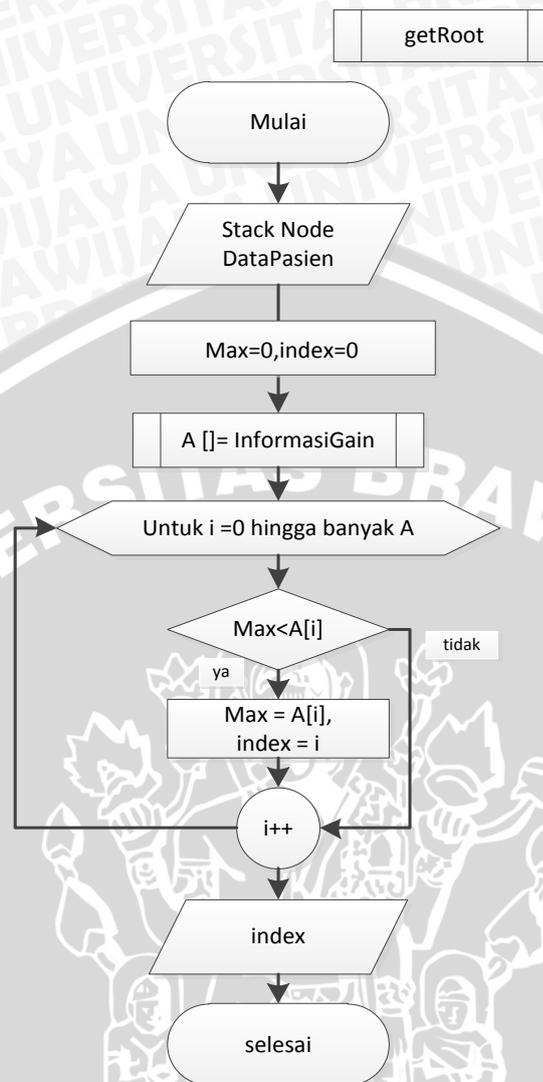
Gambar 3.5 Flowchart Pembentukan Tree





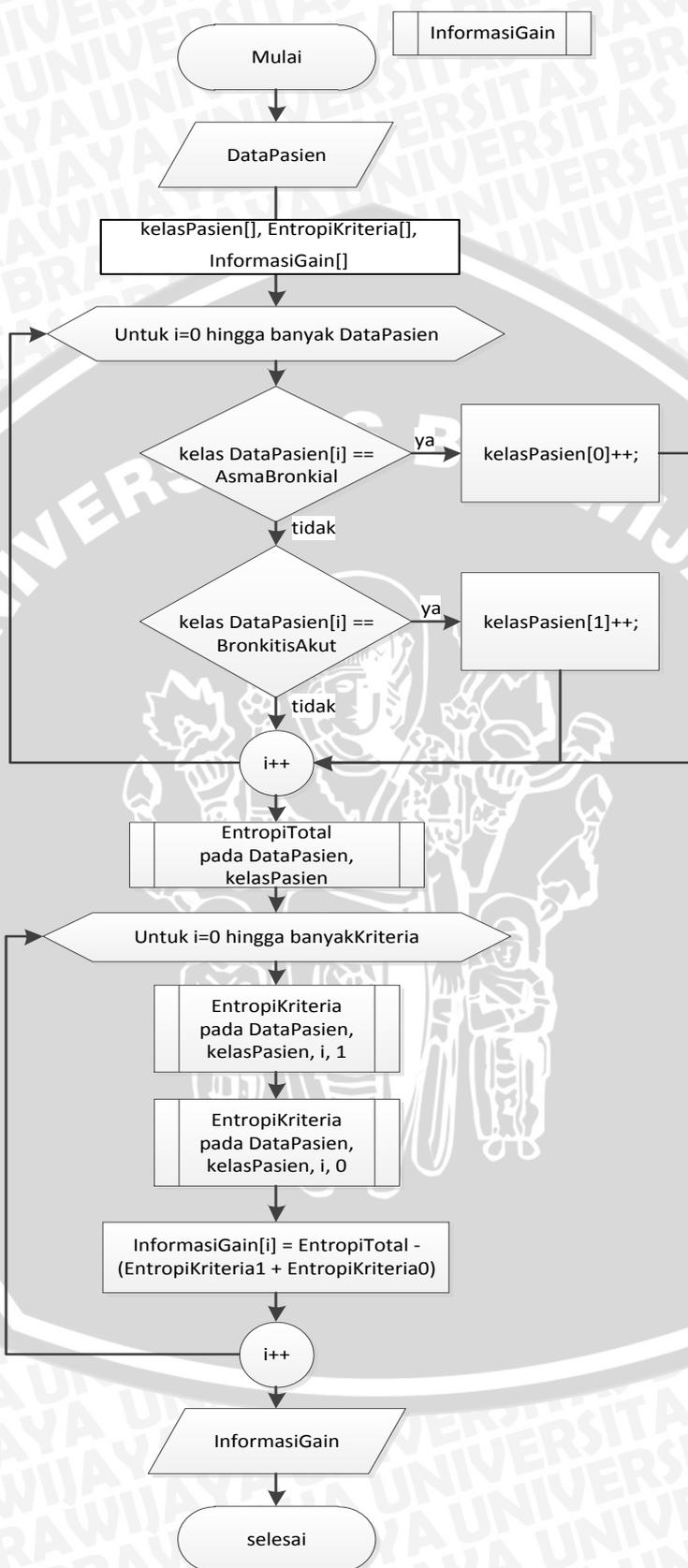
Gambar 3.6 Flowchart Jumlah Kelas

Tahap selanjutnya yaitu penentuan *root*, yaitu tahap menentukan gejala dengan nilai yang besar berdasarkan *information gain*. *Information Gain* dihitung pada masing- masing gejala berdasarkan data pasien pada kondisi saat itu. Gejala yang dipilih untuk *node perent* adalah gejala yang memiliki *Information gain* terbesar. Gambar 3.7 adalah *flowchart* untuk penentuan *root*



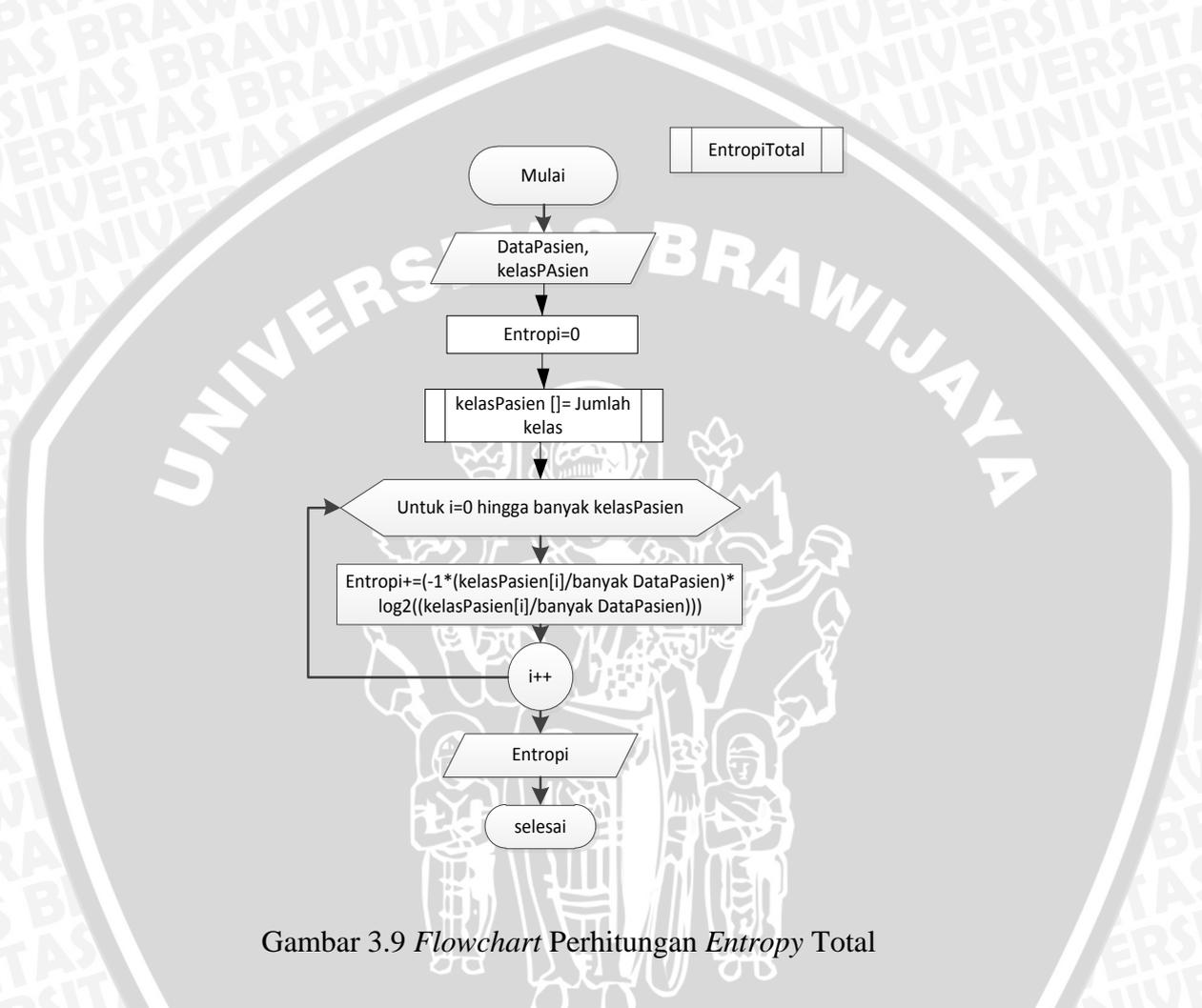
Gambar 3.7 Flowchart Penentuan Root

Selanjutnya *information gain* dihitung dari pengurangan *entropy* total seluruh gejala dengan *entropy* masing - masing nilai gejala pada variabel gejala yang diukur. Pengukuran variabel gejala dilakukan untuk semua gejala hingga didapatkan *information gain* dari setiap gejala. Gambar 3.8 adalah flowchart perhitungan *information gain* untuk keseluruhan gejala.



Gambar 3.8 Flowchart Perhitungan Information Gain

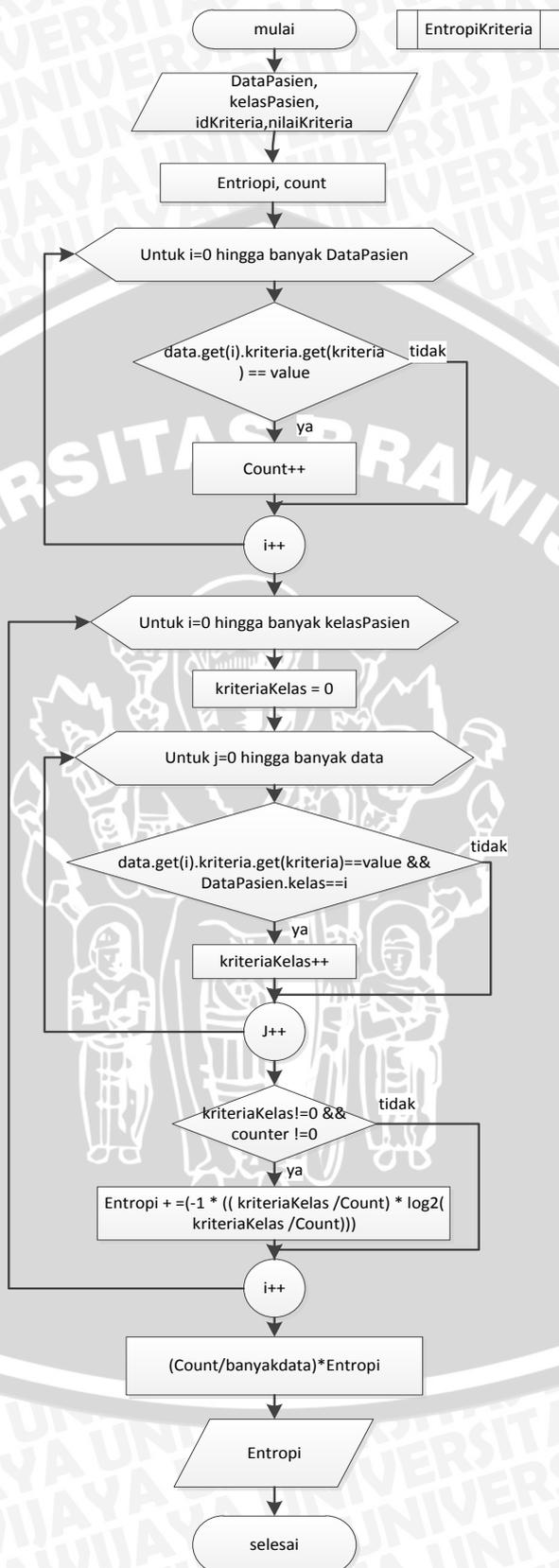
Information Gain didapatkan dari pengurangan *entropy* total dengan nilai *entropy* masing - masing gejala. *Entropy* total dihitung dari jumlah nilai *entropy* dari masing masing kelas. Gambar 3.9 adalah *flowchart* untuk perhitngan *entropy* total.



Gambar 3.9 *Flowchart* Perhitungan *Entropy* Total

Pada perhitungan *entropy* gejala, dihitung nilai tertentu pada variabel gejala. Nilai *entropy* gejala dengan nilai tertentu dihitung dari jumlah nilai *entropy* pada gejala tertentu dengan nilai tertentu pada kelas ‘asma bronkial’ dan ‘bronkitis akut’. Nilai *entropy* yang dikeluarkan nantinya akan dikalikan dengan nilai perbandingan banyaknya pasien pada gejala tertentu dan nilai gejala tertentu dengan jumlah seluruh pasien. Gambar 3.10 adalah *flowchart* perhitungan *entropy* gejala.





Gambar 3.10 Flowchart Perhitungan Entropy Gejala



3.4 Perhitungan Manual

Perhitungan manual ini menggunakan data pasien berjumlah 100 data yang dilampirkan pada Lampiran 1, dan parameter yang digunakan sebanyak 6 yaitu batuk, dahak, darah, sesak, nyeri dan demam yang masing – masing memiliki dua nilai yaitu ya dan tidak. Pada perhitungan ini akan diambil 10 data acak dari jumlah keseluruhan 100 data untuk membentuk aturan – aturan (*rule*). Cara perhitungan data tersebut adalah dengan menentukan nilai *entropy* dan nilai *information gainnya*.

Nilai *information gain* tertinggi akan digunakan sebagai root pada pohon keputusan yang terbentuk, kemudian perhitungan akan dilanjutkan pada rekursi selanjutnya. Berikut merupakan perhitungan manual untuk membentuk aturan dengan menggunakan ID3 pada sistem CBR :

1. Proses perhitungan dengan metode ID3 untuk membentuk aturan dengan 6 parameter dan 2 alternatif, yaitu :

- Parameter / atribut yang digunakan :

A1 = Batuk dikelompokkan dalam 2 kategori (ya, tidak).

A2 = Dahak dikelompokkan dalam 2 kategori (ya, tidak).

A3 = Darah dikelompokkan dalam 2 kategori (ya, tidak).

A4 = Sesak dikelompokkan dalam 2 kategori (ya, tidak).

A5 = Nyeri dada dikelompokkan dalam 2 kategori (ya, tidak).

A6 = Demam dikelompokkan dalam 2 kategori (ya, tidak).

- Alternatif yang diseleksi :

a. AB = Asma Bronkial

b. BA = Bronkitis Akut

Tabel 3.1 Data Training

Pasien	Gejala						Diagnosa
	A1	A2	A3	A4	A5	A6	
IDP15	Ya	Tidak	Ya	Ya	Tidak	Tidak	BA
IDP16	Ya	Ya	Ya	Ya	Ya	Tidak	BA
IDP17	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA

IDP18	Ya	Ya	Tidak	Ya	Ya	Ya	AB
IDP19	Tidak	Tidak	Tidak	Ya	Tidak	Tidak	BA
IDP20	Tidak	Tidak	Tidak	Ya	Tidak	Ya	BA
IDP21	Ya	Tidak	Ya	Ya	Ya	Tidak	BA
PID22	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	BA
IDP23	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA
IDP24	Ya	Ya	Ya	Ya	Tidak	Tidak	AB

Untuk penentuan node pertama dalam pembentukan pohon keputusan maka harus dihitung nilai entropy terlebih dahulu. Jumlah dari penderita AB = 2 dan penderita BA = 8, maka P1 = 2 dan P2 = 8. Dengan menggunakan persamaan 2-1 maka dilakukan penghitungan entropy total :

$$\begin{aligned} \text{Entropy}(S) &= P_a \log_2 P_a - P_b \log_2 P_b, \\ &= \left(-\frac{2}{10} \times \log_2 \left(\frac{2}{10} \right) \right) + \left(-\frac{8}{10} \log_2 \left(\frac{8}{10} \right) \right) = 0,7219 \end{aligned}$$

Dari table di atas atribut AB merupakan sampel positif (+) dan atribut BA merupakan sampel negative (-), sehingga :

a. Values (A1) = S_{Ya}, S_{Tidak}

$$S_{Ya} = [2+, 5-], |S_{Ya}| = 7$$

$$S_{Tidak} = [0+, 3-], |S_{Tidak}| = 3$$

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{2}{7} \times \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{5}{7} \times \log_2 \left(\frac{5}{7} \right) \right) = 0,8631$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{0}{3} \times \log_2 \left(\frac{0}{3} \right) \right) + \left(-\frac{3}{3} \times \log_2 \left(\frac{3}{3} \right) \right) = 0$$

Information Gain (S_{A1}) :

$$= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,7219 - \left(\frac{7}{10} \right) 0,8631 - \left(\frac{3}{10} \right) 0$$

$$= 0,7219 - 0,6042 - 0$$

$$= 0,1177$$

b. Values (A2) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{2}{5} \times \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2 \left(\frac{3}{5}\right)\right) = 0,9709$$

$$\text{Entropy } (S_{\text{Tidak}}) = \left(-\frac{0}{5} \times \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{5}{5} \times \log_2 \left(\frac{5}{5}\right)\right) = 0$$

Information Gain (S_{A2}) :

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

$$= 0,7219 - (5/10)*0,9709 - (5/10)*0$$

$$= 0,7219 - 0,4855 - 0$$

$$= 0,2356$$

c. Values ($A3$) = S_{Ya} , S_{Tidak}

$$S_{Ya} = [1+, 3-], |S_{Ya}| = 4$$

$$S_{\text{Tidak}} = [1+, 5-], |S_{\text{Tidak}}| = 6$$

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{1}{4} \times \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2 \left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{Entropy } (S_{\text{Tidak}}) = \left(-\frac{1}{6} \times \log_2 \left(\frac{1}{6}\right)\right) + \left(-\frac{5}{6} \times \log_2 \left(\frac{5}{6}\right)\right) = 0,65$$

Information Gain (S_{A3}) :

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

$$= 0,7219 - (4/10)*0,8113 - (6/10)*0,65$$

$$= 0,7219 - 0,3245 - 0,39$$

$$= 0,0074$$

d. Values ($A4$) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{2}{9} \times \log_2 \left(\frac{2}{9}\right)\right) + \left(-\frac{7}{9} \times \log_2 \left(\frac{7}{9}\right)\right) = 0,7642$$

$$\text{Entropy } (S_{\text{Tidak}}) = \left(-\frac{0}{1} \times \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2 \left(\frac{1}{1}\right)\right) = 0$$

Information Gain (S_{A4}) :

$$\begin{aligned}
 &= Entropy(S) - \sum Ya, Tidak \frac{|Sv|}{|S|} Entropy(Sv) \\
 &= 0,7219 - (9/10)*0,7642 - (1/10)*0 \\
 &= 0,7219 - 0,6878 - 0 \\
 &= 0,0341
 \end{aligned}$$

e. Values ($A5$) = S_{Ya}, S_{Tidak}

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

$$S_{Tidak} = [1+, 6-], |S_{Tidak}| = 7$$

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

$$Entropy(S_{Ya}) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,9183$$

$$Entropy(S_{Tidak}) = \left(-\frac{1}{7} \times \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} \times \log_2\left(\frac{6}{7}\right)\right) = 0,5917$$

Information Gain (S_{A5}) :

$$\begin{aligned}
 &= Entropy(S) - \sum Ya, Tidak \frac{|Sv|}{|S|} Entropy(Sv) \\
 &= 0,7219 - (3/10)*0,9183 - (7/10)*0,5917 \\
 &= 0,7219 - 0,2755 - 0,4142 \\
 &= 0,0323
 \end{aligned}$$

f. Values ($A6$) = S_{Ya}, S_{Tidak}

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

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

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

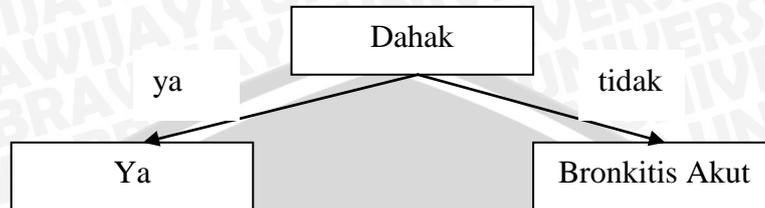
$$Entropy(S_{Ya}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(S_{Tidak}) = \left(-\frac{1}{8} \times \log_2\left(\frac{1}{8}\right)\right) + \left(-\frac{7}{8} \times \log_2\left(\frac{7}{8}\right)\right) = 0,5436$$

Information Gain (S_{A6}) :

$$\begin{aligned}
 &= Entropy(S) - \sum Ya, Tidak \frac{|Sv|}{|S|} Entropy(Sv) \\
 &= 0,7219 - (2/10)*1 - (8/10)*0,5436 \\
 &= 0,8813 - 0,2 - 0,4349 = 0,0871
 \end{aligned}$$

Dari perhitungan di atas maka dapat diketahui bahwa nilai gain tertinggi adalah pada atribut A2 yaitu Dahak sebesar 0,2365, sehingga atribut tersebut akan dijadikan root seperti pada Gambar 3.11.



Gambar 3.11 Root Atribut Dahak

Kemudian data dipilih dan dilanjutkan dengan dahak pada nilai “ya” seperti pada Tabel 3.2 berikut :

Tabel 3.2 Hasil dari filter atribut dahak dengan nilai “ya”

Pasien	Gejala						Diagnosa
	Batuk	Dahak	Darah	Sesak Napas	Nyeri Dada	Demam	
IDP16	Ya	Ya	Ya	Ya	Ya	Tidak	BA
IDP17	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA
IDP18	Ya	Ya	Tidak	Ya	Ya	Ya	AB
IDP23	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA
IDP24	Ya	Ya	Ya	Ya	Tidak	Tidak	AB

Dengan menggunakan persamaan 2-1 maka dilakukan penghitungan entropy total:

$$\begin{aligned}
 Entropy(S) &= P_a \log_2 P_a - P_b \log_2 P_b, \\
 &= \left(-\frac{2}{5} \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \log_2 \left(\frac{3}{5}\right)\right) = 0,9709
 \end{aligned}$$

Dari table di atas atribut AB merupakan sampel positif (+) dan atribut BA merupakan sampel negative (-), sehingga

- Values (A1) = S_{Ya}, S_{Tidak}
 $S_{Ya} = [2+, 3-], |S_{Ya}| = 5$
 $S_{Tidak} = [0+, 0-], |S_{Tidak}| = 0$

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) = 0,9709$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A1}) :

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

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

$$= 0$$

b. Values (A2) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) = 0,9709$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A2}) :

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

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

$$= 0$$

c. Values (A3) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,9183$$

Information Gain (S_{A3}) :

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

$$= 0,9709 - (2/5)*1 - (3/5)*0,9183$$

$$= 0,9709 - 0,4 - 0,551$$

$$= 0,02$$

d. Values (A4) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) = 0,9709$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A4}) :

$$\begin{aligned} &= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|Sv|}{|S|} \text{Entropy}(Sv) \\ &= 0,9709 - (5/5)*0,9709 - (0/5)*0 \\ &= 0 \end{aligned}$$

e. Values (A5) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy}(S_{Tidak}) = \left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,9183$$

Information Gain (S_{A5}) :

$$\begin{aligned} &= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|Sv|}{|S|} \text{Entropy}(Sv) \\ &= 0,9709 - (2/5)*1 - (3/5)*0,9183 \\ &= 0,02 \end{aligned}$$

f. Values (A6) = S_{Ya}, S_{Tidak}

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

$$S_{Tidak} = [1+, 3-], |S_{Tidak}| = 4$$

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

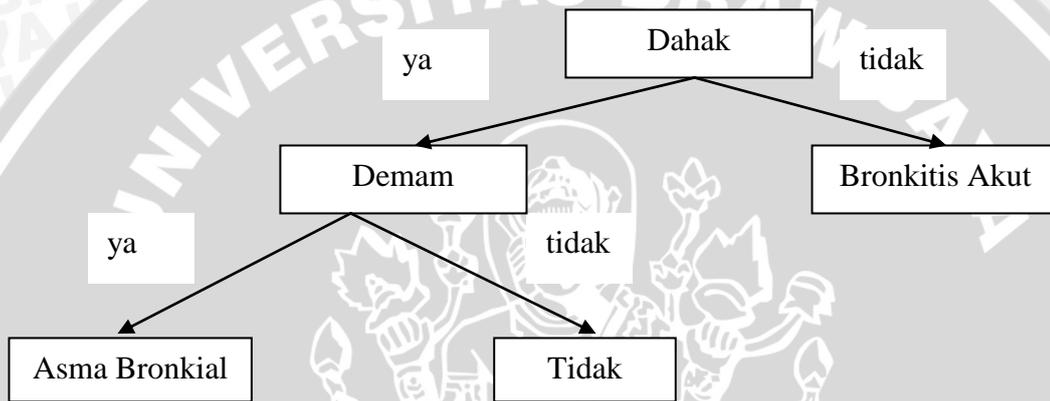
$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0,8113$$

Information Gain (S_{A6}) :

$$\begin{aligned}
 &= Entropy (S) - \sum Ya, Tidak \frac{|Sv|}{|S|} Entropy (Sv) \\
 &= 0,9709 - (1/5)*0 - (4/5)*0,8113 \\
 &= 0,9709 - 0 - 0,649 = 0,3219
 \end{aligned}$$

Dari perhitungan di atas diperoleh nilai gain tertinggi pada atribut Demam sebesar 0,3219 . Maka pada Gambar 3.12 akan ditunjukkan child dari 2 parameter Demam dengan information gain terbesar.



Gambar 3.12 Root Atribut Demam

Kemudian data dipilih kembali dan dilanjutkan pada atribut Demam pada nilai “tidak” . Berikut Tabel 3.3 merupakan hasil filter atribut demam dengan nilai “tidak”.

Tabel 3.3 Hasil filter atribut demam dengan nilai “tidak”.

Pasien	Gejala						Diagnosa
	Batuk	Dahak	Darah	Sesak Napas	Nyeri Dada	Demam	
IDP16	Ya	Ya	Ya	Ya	Ya	Tidak	BA
IDP17	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA
IDP23	Ya	Ya	Tidak	Ya	Tidak	Tidak	BA
IDP24	Ya	Ya	Ya	Ya	Tidak	Tidak	AB



Dengan menggunakan persamaan 2-1 maka dilakukan penghitungan entropy total :

$$\begin{aligned} \text{Entropy}(S) &= P_a \log_2 P_a - P_b \log_2 P_b, \\ &= \left(-\frac{1}{4} \times \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \log_2 \left(\frac{3}{4}\right)\right) = 0,8113 \end{aligned}$$

Dari table di atas atribut AB merupakan sampel positif (+) dan atribut BA merupakan sampel negative (-), sehingga :

a. Values (A1) = S_{Ya}, S_{Tidak}

$$S_{Ya} = [1+, 3-], |S_{Ya}| = 4$$

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{4} \times \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2 \left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A1}) :

$$\begin{aligned} &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\ &= 0,8113 - (4/4) \times 0,8113 - (0/4) \times 0 \\ &= 0 \end{aligned}$$

b. Values (A2) = S_{Ya}, S_{Tidak}

$$S_{Ya} = [1+, 3-], |S_{Ya}| = 4$$

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{4} \times \log_2 \left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2 \left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A2}) :

$$\begin{aligned} &= \text{Entropy}(S) - \sum Y_{a,Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \\ &= 0,8113 - (4/4) \times 0,8113 - (0/4) \times 0 \\ &= 0 \end{aligned}$$

c. Values (A3) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \log_2\left(\frac{2}{2}\right)\right) = 0$$

Information Gain (S_{A3}) :

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

$$= 0,8113 - (2/4)*1 - (2/4)*0$$

$$= 0,8113 - 0,5 - 0$$

$$= 0,3113$$

d. Values ($A4$) = S_{Ya} , S_{Tidak}

$$S_{Ya} = [1+, 3-], |S_{Ya}| = 4$$

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0,8113$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A4}) :

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

$$= 0,8113 - (4/4)*0,8113 - (0/4)*0$$

$$= 0$$

e. Values ($A5$) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,9183$$

Information Gain (S_{A5}) :

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

$$= 0,8113 - (1/4)*0 - (3/4)*0,9183$$

$$= 0,8113 - 0 - 0,6887$$

$$= 0,1226$$

f. Values (A6) = S_{Ya}, S_{Tidak}

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

$$S_{Tidak} = [1+, 3-], |S_{Tidak}| = 4$$

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

$$\text{Entropy}(S_{Ya}) = 0$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0,8113$$

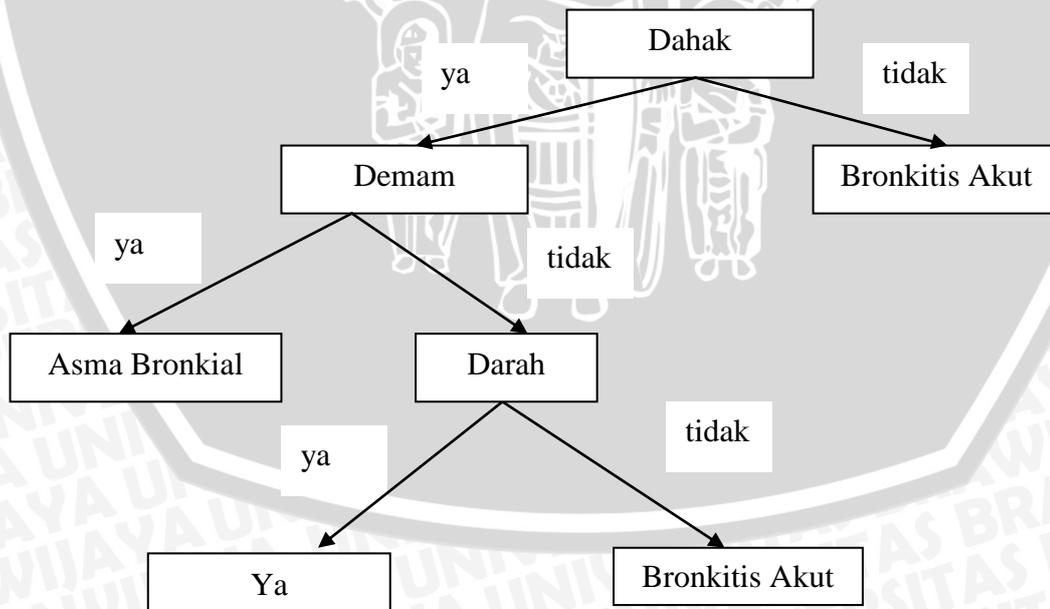
Information Gain (S_{A6}) :

$$= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|Sv|}{|S|} \text{Entropy}(Sv)$$

$$= 0,8113 - (0/4)*0 - (4/4)*0,8113$$

$$= 0$$

Dari perhitungan di atas diperoleh nilai gain tertinggi pada atribut darah sebesar 0,3113. Maka pada Gambar 3.13 akan ditunjukkan child dari 2 parameter Demam dengan information gain terbesar.



Gambar 3.13 Root Atribut Darah



Kemudian data dipilih kembali dan dilanjutkan pada atribut Darah pada nilai “ya”. Berikut Tabel 3.4 merupakan hasil filter atribut Darah dengan nilai “ya”.

Tabel 3.4 Hasil filter atribut Darah dengan nilai “ya”.

Pasien	Gejala						Diagnosa
	Batuk	Dahak	Darah	Sesak Napas	Nyeri Dada	Demam	
IDP16	Ya	Ya	Ya	Ya	Ya	Tidak	BA
IDP24	Ya	Ya	Ya	Ya	Tidak	Tidak	AB

Dengan menggunakan persamaan 2-1 maka dilakukan penghitungan entropy total :

$$\begin{aligned} \text{Entropy}(S) &= P_a \log_2 P_a - P_b \log_2 P_b, \\ &= \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) = 1 \end{aligned}$$

Dari table di atas atribut AB merupakan sampel positif (+) dan atribut BA merupakan sampel negative (-), sehingga :

- a. Values (A1) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy}(S_{Tidak}) = 0$$

Information Gain (S_{A1}) :

$$= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 1 - (2/2)*1 - (0/2)*0$$

$$= 0$$

- b. Values (A2) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy } (S_{Tidak}) = 0$$

Information Gain (S_{A2}) :

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

$$= 1 - (2/2)*1 - (0/2)*0$$

$$= 0$$

c. Values (A3) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy } (S_{Tidak}) = 0$$

Information Gain (S_{A3}) :

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

$$= 1 - (2/2)*1 - (0/2)*0$$

$$= 0$$

d. Values (A4) = S_{Ya} , S_{Tidak}

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

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

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

$$\text{Entropy } (S_{Ya}) = \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2 \left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy } (S_{Tidak}) = 0$$

Information Gain (S_{A4}) :

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

$$= 1 - (2/2)*1 - (0/2)*0$$

$$= 0$$

e. Values (A5) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 0$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 0$$

Information Gain (S_{A5}) :

$$= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|Sv|}{|S|} \text{Entropy}(Sv)$$

$$= 1 - 0 - 0$$

$$= 1$$

f. Values (A6) = S_{Ya}, S_{Tidak}

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

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

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

$$\text{Entropy}(S_{Ya}) = 0$$

$$\text{Entropy}(S_{Tidak}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

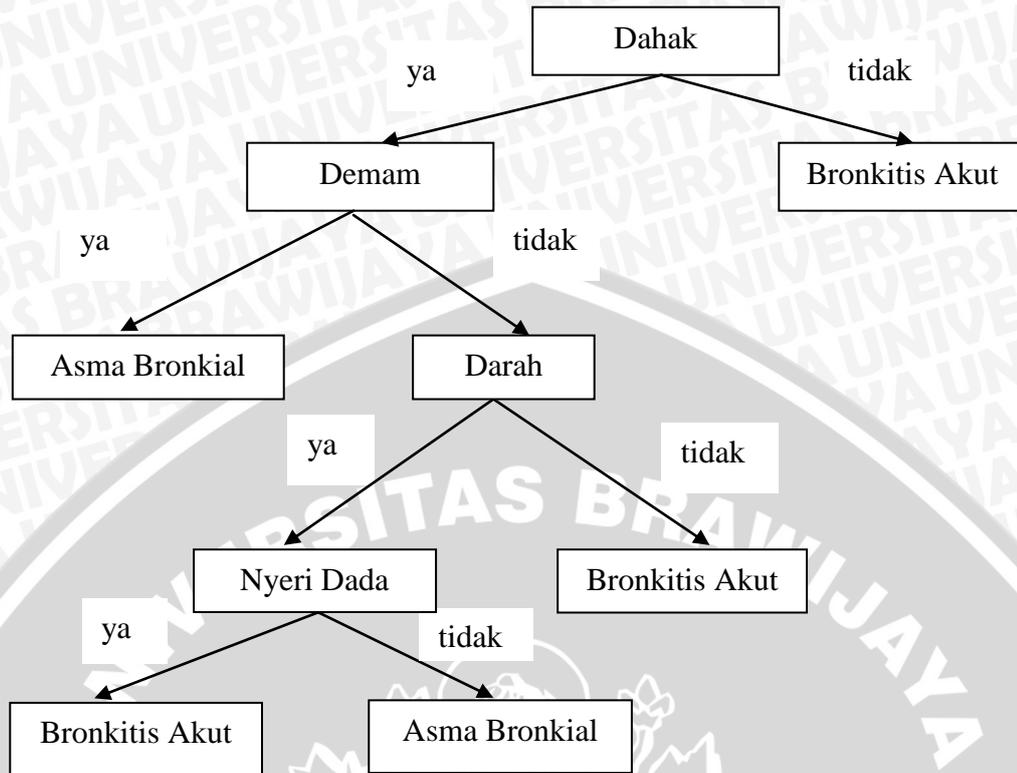
Information Gain (S_{A6}) :

$$= \text{Entropy}(S) - \sum_{Ya, Tidak} \frac{|Sv|}{|S|} \text{Entropy}(Sv)$$

$$= 1 - 0 - (2/2)*1$$

$$= 0$$

Dari perhitungan di atas diperoleh nilai gain tertinggi pada atribut Nyeri Dada sebesar 1. Maka pada Gambar 3.14 akan ditunjukkan child dari 2 parameter Nyeri Dada dengan information gain terbesar.



Gambar 3.14 Root Atribut Nyeri Dada

3.5 Skenario Pengujian

Pada tahap pengujian akan dilakukan pengujian terhadap sistem yang telah dibuat untuk mengetahui kesesuaian dan akurasi sistem terhadap pengguna dan kasus – kasus yang asli. Sistem ini akan diuji dengan menggunakan 100 data yang telah diketahui hasilnya dan akan dilihat berapa tingkat akurasinya.

Dalam penelitian ini untuk menentukan tingkat akurasinya akan dibuat beberapa skenario pengujian dengan membedakan perbandingan banyak data training dan penggunaan 10 data testing yang sama untuk masing – masing pengujian dan diuji sebanyak 5 kali dari 100 data seperti pada Tabel 3.5 berikut ini.

Tabel 3.5 Skenario Pengujian

Perbandingan (Data Training : Data Testing)	Pengujian	Akurasi	Akurasi Rata- Rata
90:10	1.		
	1.		
80:10	2.		
	3.		
	4.		
	5.		
	1.		
70:10	2.		
	3.		
	4.		
	5.		
	1.		
60:10	2.		
	3.		
	4.		
	5.		
	1.		