

## **BAB III**

### **ANALISIS DAN PERANCANGAN SISTEM**

#### **3.1 Analisis Sistem**

Analisis Sistem adalah penguraian suatu sistem yang utuh ke dalam bagian-bagian komponennya yang bertujuan untuk mengidentifikasi dan menganalisa permasalahan, kesempatan, hambatan yang terjadi dan kebutuhan yang diharapkan sehingga dapat diusulkan perbaikannya.

Sebagian besar permasalahan yang terjadi pada diri seseorang adalah malas untuk memeriksakan diri karena kesibukan sehari-hari atau karena takut mengetahui bahwa dirinya terkena resiko diabetes. Maka dari itu sebaiknya dilakukan pemeriksaan supaya bisa mengetahui apakah terkena resiko penyakit diabetes atau tidak terkena. Memprediksi tingkat resiko penyakit diabetes mellitus terdapat dua entitas yaitu pihak rumah sakit selaku admin dan pihak puskesmas selaku pengguna.

Proses prediksi dilakukan dengan menerapkan teknik data mining klasifikasi menggunakan metode *Decision Tree ID3*. Teknik tersebut membutuhkan data pembelajaran, yaitu data pasien dari hasil pemeriksaan di puskesmas yang nantinya akan diserahkan ke rumah sakit yang diinginkan dan rumah sakit akan memproses data tersebut. Proses prediksi yang dibangun akan menghasilkan nilai keluaran berupa kategori berisiko atau tidak berisiko. Untuk memilih atribut sebagai simpul akar (*root node*) atau simpul dalam (*internal node*), didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada. Sebelum perhitungan *information gain*, akan dilakukan perhitungan *entropy*. *Entropy* merupakan distribusi probabilitas dalam teori informasi dan diadopsi ke dalam algoritma *ID3* untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data (*data set*). Semakin tinggi tingkat *entropy* dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Flowchart algoritma *ID3* bisa di lihat pada gambar 2.1.

## 3.2 Hasil Analisis

Hasil analisis yang didapat adalah dengan menggunakan sistem ini pasien bisa mendapatkan informasi perkiraan kategorinya ketika sudah menjalani pemeriksaan di puskesmas yaitu berupa berisiko atau tidak berisiko. Sebelum mendapatkan informasi tersebut, pasien di puskesmas menjalani pemeriksaan untuk mendapatkan informasi apakah berisiko atau tidak berisiko terkena penyakit diabetes, dengan mengisi data tentang dirinya seperti usia, keturunan diabetes, intensitas olahraga dan index masa tubuh, maka data tersebut akan diproses menggunakan metode *Decision Tree ID3* dan akan memberikan informasi jika pasien berisiko maka akan di sarankan untuk periksa lebih lanjut di dokter spesialis penyakit dalam di rumah sakit yang sudah ditentukan karena di puskesmas tidak ada dokter spesialis penyakit dalam.

### 3.2.1 Data yang diolah

Sistem prediksi ini menggunakan teknik klasifikasi data mining yang membutuhkan data latih (data pembelajaran). Data pembelajaran akan mengambil data-data yang sudah pernah terjadi, jadi data yang diolah adalah data masa lalu.

Data yang diolah pada sistem prediksi ini adalah data dari penelitian sebelumnya penelitian sebelumnya yaitu tugas akhir Azhuardhi Azhro dengan judul “Aplikasi Perhitungan Tingkat Resiko Penyakit Diabetes Dengan Metode Sugeno” yang diambil dari puskesmas. Pengumpulan data dilakukan dengan cara membagikan kuisisioner kepada pasien-pasien yang berada di puskesmas.

Poin-poin pertanyaan kuisisioner yang dibagikan kepada responden seperti pada tabel 3.1.

**Tabel 3.1** Pertanyaan kuisioner

Kuesioner untuk pengujian data		
Tingkat Resiko penyakit Diabetes Pada Penderita diabetes dari faktor Umur, keturunan, Intensitas Olah Raga, Tingi Badan, Berat Badan		
Nama Lengkap :		
Alamat :		
Kriteria		
Usia :		Tahun
Tinggi Badan :		Cm
Berat Badan :		Kg
Keturunan Diabetes :	a. Tidak ada	b. Kakek dan nenek
	c. Paman dan Bibi	d. Kakek, nenek, paman, bibi
	e. Orang tua kandung	
Intensitas Olahraga	..... Dalam satu bulan	
Hasil Pemeriksaan	Berisiko	Tidak Berisiko

Dari data-data tersebut yang dipilih untuk dijadikan sebagai atribut adalah umur, keturunan diabetes, index Masa Tubuh (IMT) dan intensitas olahraga. Sedangkan hasil pemeriksaan menjadi label kelas atau kelas tujuan yang akan dikategorikan menjadi “Berisiko” dan “Tidak Berisiko”. Tabel tipe data atribut bisa dilihat pada tabel 3.2

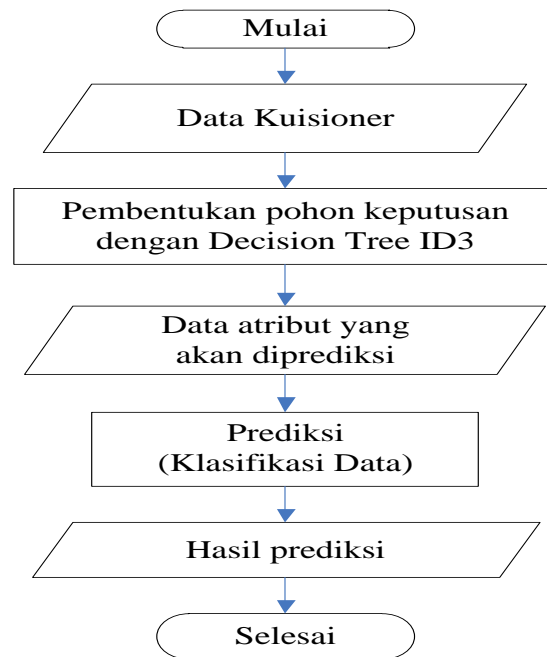
**Tabel 3.2** Tipe data atribut

Atribut	Nilai Atribut
Umur	Muda, Parobaya, Tua
Keturunan Diabtes	A = Tidak Ada, B = Kakek dan Nenek, C= Paman dan Bibi, D= Kakek, Nenek, Paman, Bibi. Dan E = Orang tua kandung
Index Masa Tubuh (IMT)	Kecil, Sedang, Besar
Intensitas Olahraga	Rajin, Sedang, Malas

### 3.2.2 Deskripsi Sistem

Sistem yang dibangun merupakan aplikasi atau *tool* prediksi tingkat resiko penyakit diabetes menggunakan teknik data mining dengan metode *Decision Tree ID3*. Sistem ini akan menghasilkan nilai keluaran berupa perkiraan hasil pemeriksaan yang akan tergolong “Beresiko” atau “Tidak beresiko”. Terdapat beberapa atribut yang dibutuhkan untuk memprediksi tingkat resiko diabetes diantaranya adalah umur, keturunan diabetes, index masa tubuh (IMT), dan Intensitas Olahraga.

Gambar 3.1 akan menjelaskan alur sistem pada aplikasi sistem prediksi tingkat resiko diabetes.



**Gambar 3.1** *Flowchart System*

Penjelasan gambar 3.1:

1. Pertama memasukkan data latih (data pasien yang di periksa) yang akan disimpan didalam *database*.
2. Pembuatan pohon keputusan dengan metode *Decision Tree ID3* berdasarkan data yang sudah disimpan didalam *database*.
3. Selanjutnya memasukkan data yang akan diprediksi (data uji).
4. Sistem melakukan klasifikasi data uji dengan menggunakan pohon keputusan yang sudah terbentuk pada proses sebelumnya.
5. Sistem mengeluarkan *output* klasifikasi atau hasil prediksi.

### 3.3 Representasi Data

Terdapat dua macam data yang akan di gunakan yaitu data latih dan data uji. Data latih berfungsi untuk pembentukan pohon keputusan sedangkan data uji adalah data untuk pengujian sistem. Data di dapat dari dari penelitian sebelumnya yaitu tugas akhir Azhuardhi Azhro dengan judul “Aplikasi Perhitungan Tingkat Resiko Penyakit Diabetes Dengan Metode Sugeno”.

Pada tabel 3.3 akan di tampilkan data keseluruhan yang akan di gunakan.

Sedangkan pada data latih pada table 3.4 dan data uji pada table 3.5

**Tabel 3.3** Data Keseluruhan

No	Usia	Keturunan diabetes	IMT	Intensitas Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
2	Muda	Kakek Nenek Paman Bibi	Sedang	Sedang	Beresiko
3	Parobaya	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
4	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
5	Muda	Paman Bibi	Kecil	Sedang	Beresiko
6	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
7	Tua	Paman Bibi	Besar	Malas	Beresiko
8	Tua	Kakek Nenek	Besar	Sedang	Beresiko
9	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
10	Muda	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
11	Muda	Paman Bibi	Besar	Rajin	Tidak Beresiko
12	Parobaya	Paman Bibi	Kecil	Rajin	Tidak Beresiko
13	Parobaya	Orang Tua	Sedang	Malas	Beresiko
14	Muda	Orang Tua	Kecil	Sedang	Beresiko
15	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
16	Tua	Kakek Nenek	Besar	Malas	Beresiko
17	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Beresiko
18	Parobaya	Kakek Nenek	Sedang	Rajin	Beresiko
19	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
20	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
21	Parobaya	Kakek Nenek Paman Bibi	Sedang	Malas	Beresiko
22	Parobaya	Paman Bibi	Kecil	Sedang	Tidak Beresiko
23	Parobaya	Orang Tua	Kecil	Malas	Beresiko
24	Tua	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
25	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
26	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko
27	Muda	Paman Bibi	Sedang	Rajin	Beresiko
28	Parobaya	Orang Tua	Besar	Sedang	Beresiko
29	Muda	Kakek Nenek Paman Bibi	Kecil	Rajin	Tidak Beresiko
30	Tua	Orang Tua	Besar	Sedang	Beresiko
31	Parobaya	Tidak Ada	Kecil	Sedang	Tidak Beresiko
32	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko
33	Parobaya	Kakek Nenek	Besar	Malas	Beresiko
34	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
35	Muda	Paman Bibi	Kecil	Sedang	Beresiko
36	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko

37	Tua	Paman Bibi	Sedang	Malas	Beresiko
38	Tua	Orang Tua	Besar	Malas	Beresiko
39	Muda	Kakek Nenek	Sedang	Rajin	Tidak Beresiko
40	Parobaya	Orang Tua	Besar	Malas	Beresiko
41	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
42	Parobaya	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko
43	Parobaya	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
44	Parobaya	Paman Bibi	Besar	Sedang	Beresiko
45	Tua	Paman Bibi	Besar	Sedang	Beresiko
46	Tua	Tidak Ada	Kecil	Sedang	Tidak Beresiko
47	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
48	Tua	Orang Tua	Sedang	Malas	Beresiko
49	Tua	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko
50	Tua	Orang Tua	Besar	Malas	Beresiko

**Tabel 3.4** Data Latih

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
2	Muda	Kakek Nenek Paman Bibi	Sedang	Sedang	Beresiko
3	Parobaya	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
4	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
5	Muda	Paman Bibi	Kecil	Sedang	Beresiko
6	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
7	Tua	Paman Bibi	Besar	Malas	Beresiko
8	Tua	Kakek Nenek	Besar	Sedang	Beresiko
9	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
10	Muda	Paman Bibi	Sedang	Rajin	Beresiko
11	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
12	Tua	Paman Bibi	Sedang	Malas	Beresiko
13	Parobaya	Orang Tua	Besar	Sedang	Beresiko
14	Muda	Kakek Nenek Paman Bibi	Kecil	Rajin	Tidak Beresiko
15	Tua	Orang Tua	Sedang	Malas	Beresiko
16	Tua	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko
17	Tua	Orang Tua	Besar	Malas	Beresiko
18	Parobaya	Orang Tua	Sedang	Malas	Beresiko
19	Muda	Orang Tua	Kecil	Sedang	Beresiko
20	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
21	Tua	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
22	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
23	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko

24	Tua	Orang Tua	Besar	Sedang	Beresiko
25	Parobaya	Tidak Ada	Kecil	Sedang	Tidak Beresiko
26	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko
27	Parobaya	Kakek Nenek	Besar	Malas	Beresiko
28	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
29	Muda	Paman Bibi	Kecil	Sedang	Beresiko
30	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
31	Muda	Paman Bibi	Besar	Rajin	Tidak Beresiko
32	Parobaya	Paman Bibi	Kecil	Rajin	Tidak Beresiko
33	Tua	Kakek Nenek	Besar	Malas	Beresiko
34	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Beresiko
35	Parobaya	Kakek Nenek	Sedang	Rajin	Beresiko
36	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
37	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
38	Parobaya	Kakek Nenek Paman Bibi	Sedang	Malas	Beresiko
39	Parobaya	Paman Bibi	Kecil	Sedang	Tidak Beresiko
40	Parobaya	Orang Tua	Kecil	Malas	Beresiko

**Tabel 3.5** Data Uji

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
2	Tua	Orang Tua	Besar	Malas	Beresiko
3	Muda	Kakek Nenek	Sedang	Rajin	Tidak Beresiko
4	Parobaya	Orang Tua	Besar	Malas	Beresiko
5	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
6	Parobaya	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko
7	Parobaya	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
8	Parobaya	Paman Bibi	Besar	Sedang	Beresiko
9	Tua	Paman Bibi	Besar	Sedang	Beresiko
10	Tua	Tidak Ada	Kecil	Sedang	Tidak Beresiko

### 3.4 Perhitungan Decision Tree ID3

Perhitungan *decision tree ID3* ini akan menggunakan data pada tabel 3.4 (*data latih*). Tabel tersebut akan diubah menjadi sebuah *tree*. Sebelum melakukan perhitungan, berikut akan dijelaskan beberapa ketentuan dalam pembentukan *tree* pada kasus ini.



- Jika semua atribut memiliki nilai yang sama, maka diambil keputusan meskipun kelas tujuan masih heterogen atau berbeda. Keputusan yang diambil adalah yang paling banyak atau mayoritas.
- Jika nilai masih sama dan kelas tujuan heterogen jumlah keputusan sama maka diambil salah satu keputusan.

1. Langkah pertama adalah memilih atribut yang akan dijadikan akar (*root node*) dengan menghitung nilai *gain* yang paling tinggi. Sebelumnya yang akan dihitung adalah nilai *entropy* semua data. Berikut adalah perhitungan *entropy* semua data.

harus di hitung dulu entropy untuk semua data terhadap komposisi kelas. Menggunakan persamaan 2.1.3

diketahui : jumlah beresiko = 23

Contoh perhitungan entropy atribut usia :

$$E(\text{semua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(23|40) \times \log_2 p(23|40)) \\ + (p(17|40) \times \log_2 p(17|40)) \end{array} \right)$$

$$= 0.9837$$

2. Selanjutnya di hitung entropy untuk setiap nilai fitur terhadap kelas, untuk entropy nilai dalam usia, didapat:

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(6|15) \times \log_2 p(6|15)) \\ + (p(9|15) \times \log_2 p(9|15)) \end{array} \right)$$

$$= 0.9710$$

$$\begin{aligned}
& E(\text{Semua}_{\text{parobaya}}) \\
&= - \left( \begin{array}{c} (p(\text{beresiko}|\text{parobaya}) \times \log_2 p(\text{beresiko}|\text{parobaya})) \\ + (p(\text{tidak beresiko}|\text{parobaya}) \times \log_2 p(\text{tidak beresiko}|\text{parobaya})) \end{array} \right) \\
& E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{c} (p(7|15) \times \log_2 p(7|15)) \\ + (p(6|15) \times \log_2 p(6|15)) \end{array} \right) \\
&= 0.9957
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{c} (p(\text{beresiko}|\text{tua}) \times \log_2 p(\text{beresiko}|\text{tua})) \\ + (p(\text{tidak beresiko}|\text{tua}) \times \log_2 p(\text{tidak beresiko}|\text{tua})) \end{array} \right) \\
& E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{c} (p(10|15) \times \log_2 p(10|15)) \\ + (p(2|15) \times \log_2 p(2|15)) \end{array} \right) \\
&= 0.6500
\end{aligned}$$

3. Entropy tersebut di hitung pada setiap fitur. Selanjutnya dihitung gain untuk setiap fitur. Dari perhitungan, di dapat gain seperti dibawah ini. Selengkapnya hasil perhitungan entropy dan gain untuk node akar di lihatkan pada tabel 3.6.

$$\begin{aligned}
G(\text{semua}, \text{usia}) &= E(\text{semua}) = \sum_{i=1}^n p(v_i | \text{semua}) \times E(\text{semua}_{\text{usia}}) \\
&= E(\text{semua}) \\
&- \left( \begin{array}{c} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{\text{muda}})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{\text{parobaya}})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{\text{tua}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{c} (p(15|40) \times (0.9170)) + (p(13|40) \times (0.9957)) \\ + (p(12|40) \times (0)) \end{array} \right) \\
&= 0.1206
\end{aligned}$$

❖ Perhitungan manual entropy atribut keturunan *diabetes* pada tabel 3.6

$$E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{array}{l} (p(10|15) \times \log_2 p(10|15)) \\ + (p(2|15) \times \log_2 p(2|15)) \end{array} \right) \\ = 0$$

$$E(\text{Semua}_{\text{kakek nenek}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{kakek nenek}}) = - \left( \begin{array}{l} (p(4|7) \times \log_2 p(4|7)) \\ + (p(3|7) \times \log_2 p(3|7)) \end{array} \right) \\ = 0.9852$$

$$E(\text{Semua}_{\text{paman bibi}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \\ \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{paman bibi}}) = - \left( \begin{array}{l} (p(5|10) \times \log_2 p(5|10)) \\ + (p(5|10) \\ \times \log_2 p(5|10)) \end{array} \right) \\ = 1$$

$$E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(7|10) \times \log_2 p(7|10)) \\ + (p(3|10) \times \log_2 p(3|10)) \end{array} \right) \\ = 0.8813$$

$$\begin{aligned}
& E(\text{Semua}_{\text{orang tua}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{orang tua}}) = - \left( \begin{array}{l} (p(7|10) \times \log_2 p(7|10)) \\ + (p(3|10) \times \log_2 p(3|10)) \end{array} \right) \\
&= 0.8813
\end{aligned}$$

❖ Perhitungan manual gain atribut keturunan *diabetes* pada tabel 3.6

$$\begin{aligned}
&= E(\text{Semua}) \\
&- \left( \begin{array}{l} (p(\text{tidak ada}|\text{semua}) \times E(\text{semua}_{\text{tidak ada}})) + (p(\text{kakek nenek}|\text{semua}) \times E(\text{semua}_{\text{kakek nenek}})) \\ + (p(\text{paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bbi}})) + \\ (p(\text{kakek nenek paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bbi}})) + \\ (p(\text{orang tua}|\text{semua}) \times E(\text{semua}_{\text{orang tua}})) \end{array} \right) \\
&= E(0.9837) - \left( \begin{array}{l} (p(3|40) \times E(0)) + (p(7|40) \times E(0.9852)) \\ + (p(10|40) \times E(1)) + \\ (p(10|40) \times E(0.8813)) + \\ (p(10|40) \times E(0.8813)) \end{array} \right) \\
&= 0.1206
\end{aligned}$$

❖ Perhitungan manual entropy atribut IMT pada tabel 3.6

$$\begin{aligned}
& E(\text{Semua}_{\text{besar}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{besar}}) = - \left( \begin{array}{l} (p(11|12) \times \log_2 p(11|12)) \\ + (p(1|12) \times \log_2 p(1|12)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{sedang}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(7|11) \times \log_2 p(7|11)) \\ + (p(4|11) \times \log_2 p(4|11)) \end{array} \right) \\
&= 0.9457
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{kecil}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{kecil}}) = - \left( \begin{array}{l} (p(5|17) \times \log_2 p(5|17)) \\ + (p(12|17) \times \log_2 p(12|17)) \end{array} \right) \\
&= 0.8740
\end{aligned}$$

❖ Perhitungan gain atribut IMT pada tabel 3.6

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{besar}|\text{semua}) \times E(\text{semua}_{\text{besar}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{kecil}|\text{semua}) \times E(\text{semua}_{\text{kecil}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(12|40) \times (0)) + (p(11|40) \times (0.9457)) \\ + (p(17|40) \times (0.8740)) \end{array} \right) \\
&= 0.3522
\end{aligned}$$

❖ Perhitungan manual entropy atribut olahraga pada tabel 3.6

$$\begin{aligned}
& E(\text{Semua}_{\text{rajin}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(2|8) \times \log_2 p(2|8)) \\ + (p(6|8) \times \log_2 p(6|8)) \end{array} \right)
\end{aligned}$$

$$= 0.8113$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(8|17) \times \log_2 p(8|17)) \\ + (p(9|17) \times \log_2 p(9|17)) \end{array} \right)$$

$$= 0.9975$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(13|15) \times \log_2 p(13|15)) \\ + (p(2|15) \times \log_2 p(2|15)) \end{array} \right)$$

$$= 0.5665$$

❖ Perhitungan manual gain atribut olahraga pada tabel 3.6.

$$= E(\text{semua}) - \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right)$$

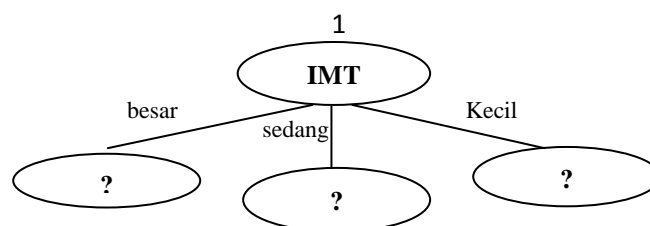
$$= E(\text{semua}) - \left( \begin{array}{l} (p(8|40) \times (0.8113)) + (p(17|40) \times (0.9975)) \\ + (p(15|40) \times (0.5665)) \end{array} \right)$$

$$= 0.1851$$

**Tabel 3.6** Hasil perhitungan entropy dan gain atribut pada node akar

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		40	23	17	0.9837	
<b>Usia</b>	Muda	15	6	9	0.9710	0.1010
	Parobaya	13	7	6	0.9957	
	Tua	12	10	2	0.6500	
<b>Keturunan</b>	Tidak ada	3	0	3	0.0000	0.1206
	Kakek Nenek	7	4	3	0.9852	
	Paman Bibi	10	5	5	1.0000	
	Kakek Nenek Paman Bibi	10	7	3	0.8813	
	Orang Tua	10	7	3	0.8813	
<b>IMT</b>	Besar	12	11	1	0.4138	0.2281
	Sedang	11	7	4	0.9457	
	Kecil	17	5	12	0.8740	
<b>Olahraga</b>	Rajin	8	2	6	0.8113	0.1851
	Sedang	17	8	9	0.9975	
	Malas	15	13	2	0.5665	

Dari hasil perhitungan *gain* pada tabel 3.6, atribut yang memiliki nilai *gain* tertinggi atribut IMT. Maka di lakukan percabangan pada atribut IMT dengan nilai besar, sedang, dan kecil. Percabangan pada atribut IMT pada gambar 3.2. Data pada kasus IMT nilai besar bisa dilihat pada tabel 3.7, kasus IMT nilai sedang pada tabel 3.8 dan kasus IMT nilai kecil pada tabel 3.9

**Gambar 3.2** Hasil pembentukan cabang pada atribut IMT**Tabel 3.7** Data pada kasus IMT nilai besar

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Parobaya	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
2	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
3	Tua	Paman Bibi	Besar	Malas	Beresiko
4	Tua	Kakek Nenek	Besar	Sedang	Beresiko
5	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
6	Parobaya	Orang Tua	Besar	Sedang	Beresiko
7	Tua	Orang Tua	Besar	Malas	Beresiko
8	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko

9	Tua	Orang Tua	Besar	Sedang	Beresiko
10	Parobaya	Kakek Nenek	Besar	Malas	Beresiko
11	Muda	Paman Bibi	Besar	Rajin	Tidak Beresiko
12	Tua	Kakek Nenek	Besar	Malas	Beresiko

**Tabel 3.8** Data pada kasus IMT nilai sedang

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
2	Muda	Kakek Nenek Paman Bibi	Sedang	Sedang	Beresiko
3	Muda	Paman Bibi	Sedang	Rajin	Beresiko
4	Tua	Paman Bibi	Sedang	Malas	Beresiko
5	Tua	Orang Tua	Sedang	Malas	Beresiko
6	Parobaya	Orang Tua	Sedang	Malas	Beresiko
7	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
8	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
9	Parobaya	Kakek Nenek	Sedang	Rajin	Beresiko
10	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
11	Parobaya	Kakek Nenek Paman Bibi	Sedang	Malas	Beresiko

**Tabel 3.9** Data pada kasus IMT nilai kecil

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
2	Muda	Paman Bibi	Kecil	Sedang	Beresiko
3	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
4	Muda	Kakek Nenek Paman Bibi	Kecil	Rajin	Tidak Beresiko
5	Tua	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko
6	Muda	Orang Tua	Kecil	Sedang	Beresiko
7	Tua	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
8	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko
9	Parobaya	Tidak Ada	Kecil	Sedang	Tidak Beresiko
10	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko
11	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
12	Muda	Paman Bibi	Kecil	Sedang	Beresiko
13	Parobaya	Paman Bibi	Kecil	Rajin	Tidak Beresiko
14	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Beresiko
15	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
16	Parobaya	Paman Bibi	Kecil	Sedang	Tidak Beresiko
17	Parobaya	Orang Tua	Kecil	Malas	Beresiko



- ❖ Perhitungan manual entropy atribut usia pada tabel 3.10

$$E(\text{semua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(14|15) \times \log_2 p(14|15)) \\ + (p(1|15) \times \log_2 p(1|15)) \end{array} \right) \\ = 0.3534$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right) \\ = 0$$

$$E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(4|4) \times \log_2 p(4|4)) \\ + (p(0|4) \times \log_2 p(0|4)) \end{array} \right) \\ = 0$$

$$E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} (p(10|10) \times \log_2 p(10|10)) \\ + (p(0|10) \times \log_2 p(0|10)) \end{array} \right) \\ = 0$$

- ❖ Perhitungan manual gain atribut usia pada tabel 3.10

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{aligned} &(p(\text{muda}|\text{semua}) \times E(\text{semua}_{\text{muda}})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{\text{parobaya}})) \\ &+ (p(\text{tua}|\text{semua}) \times E(\text{semua}_{\text{tua}})) \end{aligned} \right) \\
&= E(\text{semua}) - \left( \begin{aligned} &(p(1|15) \times (0)) + (p(4|15) \times (0)) \\ &+ (p(10|15) \times (0)) \end{aligned} \right) \\
&= 0.3534
\end{aligned}$$

❖ Perhitungan manual entropy atribut keturunan *diabetes* untuk tabel 3.10

$$\begin{aligned}
&E(\text{Semua}_{\text{tidak ada}}) \\
&= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
&E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{aligned} &(p(0|0) \times \log_2 p(0|0)) \\ &+ (p(0|0) \times \log_2 p(0|0)) \end{aligned} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
&E(\text{Semua}_{\text{kakek nenek}}) \\
&= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
&E(\text{Semua}_{\text{kakek nenek}}) = - \left( \begin{aligned} &(p(3|3) \times \log_2 p(3|3)) \\ &+ (p(0|3) \times \log_2 p(0|3)) \end{aligned} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
&E(\text{Semua}_{\text{paman bibi}}) \\
&= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \\ &\times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
&E(\text{Semua}_{\text{paman bibi}}) = - \left( \begin{aligned} &(p(3|4) \times \log_2 p(3|4)) \\ &+ (p(1|4) \\ &\times \log_2 p(1|4)) \end{aligned} \right)
\end{aligned}$$

$$= 0.8113$$

$$E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(5|5) \times \log_2 p(5|5)) \\ + (p(0|5) \times \log_2 p(0|5)) \end{array} \right) = 0$$

$$E(\text{Semua}_{\text{orang tua}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{orang tua}}) = - \left( \begin{array}{l} (p(3|3) \times \log_2 p(3|3)) \\ + (p(0|3) \times \log_2 p(0|3)) \end{array} \right) = 0$$

Perhitungan manual gain atribut keturunan *diabetes* pada tabel 3.10

$$\begin{aligned} &= E(\text{Semua}) \\ &- \left( \begin{array}{l} (p(\text{tidak ada}|\text{semua}) \times E(\text{semua}_{\text{tidak ada}})) + (p(\text{kakek nenek}|\text{semua}) \times E(\text{semua}_{\text{kakek nenek}})) \\ + (p(\text{paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ (p(\text{kakek nenek paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ (p(\text{orang tua}|\text{semua}) \times E(\text{semua}_{\text{orang tua}})) \end{array} \right) \\ &= E(0.9837) - \left( \begin{array}{l} (p(0|15) \times E(0)) + (p(3|15) \times E(0)) \\ + (p(4|15) \times E(0.8113)) + \\ (p(5|15) \times E(0)) + \\ (p(3|15) \times E(0)) \end{array} \right) \\ &= 0.1370 \end{aligned}$$

- ❖ Perhitungan manual entropy atribut olahraga pada tabel 3.10

$$\begin{aligned}
 & E(\text{Semua}_{\text{rajin}}) \\
 &= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
 & E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right) \\
 & \qquad \qquad \qquad = 0
 \end{aligned}$$

$$\begin{aligned}
 & E(\text{Semua}_{\text{sedang}}) \\
 &= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
 & E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(4|4) \times \log_2 p(4|4)) \\ + (p(0|4) \times \log_2 p(0|4)) \end{array} \right) \\
 & \qquad \qquad \qquad = 0
 \end{aligned}$$

$$\begin{aligned}
 & E(\text{Semua}_{\text{malas}}) \\
 &= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
 & E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(10|10) \times \log_2 p(10|10)) \\ + (p(0|10) \times \log_2 p(0|10)) \end{array} \right) \\
 & \qquad \qquad \qquad = 0
 \end{aligned}$$

- ❖ Perhitungan manual gain atribut olahraga pada tabel 3.10. Hasil Perhitungan dapat dilihat pada tabel 3.10

$$\begin{aligned}
 &= E(\text{semua}) \\
 &- \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right) \\
 &= E(\text{semua}) - \left( \begin{array}{l} (p(1|15) \times (0)) + (p(4|15) \times (0)) \\ + (p(10|15) \times (0)) \end{array} \right) \\
 &= 0.3534
 \end{aligned}$$

**Tabel 3.10** Hasil perhitungan *gain* pada atribut IMT nilai besar

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		15	14	1	0.3534	
<b>Usia</b>	Muda	1	0	1	0.0000	0.3534
	Parobaya	4	4	0	0.0000	
	Tua	10	10	0	0.0000	
<b>Keturunan</b>	Tidak ada	0	0	0	0.0000	0.1370
	Kakek Nenek	3	3	0	0.0000	
	Paman Bibi	4	3	1	0.8113	
	Kakek Nenek Paman Bibi	5	5	0	0.0000	
	Orang Tua	3	3	0	0.0000	
<b>Olahraga</b>	Rajin	1	0	1	0.0000	0.3534
	Sedang	4	4	0	0.0000	
	Malas	10	10	0	0.0000	

Dari hasil perhitungan diatas nilai gain sama yaitu gain atribut usia dan olahraga. Maka diambil salah satu. Terpilih gain usia. Dilakukan percabangan lagi atribut usia. Data pembagian cabang kasus usia dapat dilihat pada tabel 3.11

**Tabel 3.11** Data pembagian cabang kasus usia

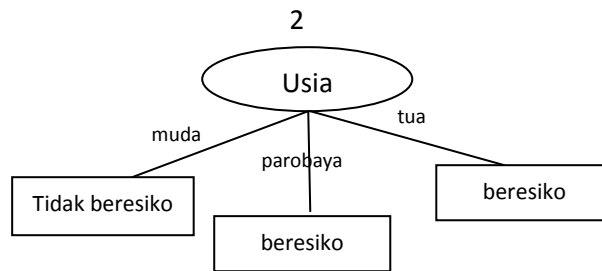
No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Besar	Rajin	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Parobaya	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
2	Parobaya	Kakek Nenek	Besar	Malas	Beresiko
3	Parobaya	Orang Tua	Besar	Malas	Beresiko
4	Parobaya	Paman Bibi	Besar	Sedang	Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
2	Tua	Paman Bibi	Besar	Malas	Beresiko
3	Tua	Kakek Nenek	Besar	Sedang	Beresiko
4	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
5	Tua	Kakek Nenek	Besar	Malas	Beresiko
6	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
7	Tua	Orang Tua	Besar	Sedang	Beresiko
8	Tua	Orang Tua	Besar	Malas	Beresiko

9	Tua	Kakek Nenek Paman Bibi	Besar	Malas	Beresiko
10	Tua	Paman Bibi	Besar	Sedang	Beresiko

Dari data pembagian cabang kasus usia sudah bisa diambil keputusan karena semua atribut dan kelas sama. Kemudian dapat di buat pohon percabangan pada gambar 3.3



**Gambar 3.3** Percabangan usia pada kasus IMT besar

Kemudian memilih atribut sebagai pemecah cabang pada usia kasus IMT sedang. Nilai *gain* tertinggi didapatkan pada atribut keturunan. Hasil perhitungan *gain* pada kasus IMT sedang bisa dilihat pada tabel 3.12

- ❖ Perhitungan manual entropy atribut usia pada tabel 3.12

$$E(\text{semua}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua}) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(5|9) \times \log_2 p(5|9)) \\ + (p(4|9) \times \log_2 p(4|9)) \end{array} \right) \\ = 0.9911$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua}) \end{array} \right)$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(2|6) \times \log_2 p(2|6)) \\ + (p(4|6) \times \log_2 p(4|6)) \end{array} \right)$$

$$= 0.9813$$

$$E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(2|2) \times \log_2 p(2|2)) \\ + (p(0|2) \times \log_2 p(0|2)) \end{array} \right) \\ = 0$$

$$E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} (p(1|1) \times \log_2 p(1|1)) \\ + (p(0|1) \times \log_2 p(0|1)) \end{array} \right) \\ = 0$$

❖ Perhitungan manual gain atribut usia pada tabel 3.12

$$= E(\text{semua}) - \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{\text{muda}})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{\text{parobaya}})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{\text{tua}})) \end{array} \right) \\ = E(\text{semua}) - \left( \begin{array}{l} (p(6|9) \times (0.9813)) + (p(2|9) \times (0)) \\ + (p(1|9) \times (0)) \end{array} \right) \\ = 0.3789$$

❖ Perhitungan manual entropy atribut keturunan *diabetes* pada tabel 3.12

$$E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{array}{l} (p(0|2) \times \log_2 p(0|2)) \\ + (p(2|2) \times \log_2 p(2|2)) \end{array} \right) \\ = 0$$

$$\begin{aligned}
& E(\text{Semua}_{\text{kakek nenek}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{kakek nenek}}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{paman bibi}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \\ \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{paman bibi}}) = - \left( \begin{array}{l} (p(2|3) \times \log_2 p(2|3)) \\ + (p(1|3) \\ \times \log_2 p(1|3)) \end{array} \right) \\
&= 0.9183
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{kakek nenek paman bibi}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(2|2) \times \log_2 p(2|2)) \\ + (p(0|2) \times \log_2 p(0|2)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{orang tua}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{orang tua}}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$



❖ Perhitungan gain atribut keturunan *diabetes* pada tabel 3.12

$$\begin{aligned}
 &= E(\text{Semua}) \\
 &= - \left( \begin{aligned} &(p(\text{tidak ada}|\text{semua}) \times E(\text{semua}_{\text{tidak ada}})) + (p(\text{kakek nenek}|\text{semua}) \times E(\text{semua}_{\text{kakek nenek}})) \\ &+ (p(\text{paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ &(p(\text{kakek nenek paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ &(p(\text{orang tua}|\text{semua}) \times E(\text{semua}_{\text{orang tua}})) \end{aligned} \right) \\
 &= E(0.9837) - \left( \begin{aligned} &(p(2|9) \times E(0)) + (p(2|9) \times E(1)) \\ &+ (p(3|9) \times E(0.9183)) + \\ &(p(2|9) \times E(0)) + \\ &(p(0|15) \times E(0)) \end{aligned} \right) \\
 &= 0.4628
 \end{aligned}$$

❖ Perhitungan manual atribut olahraga pada tabel 3.12

$$\begin{aligned}
 &E(\text{Semua}_{\text{rajin}}) \\
 &= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
 &E(\text{Semua}_{\text{rajin}}) = - \left( \begin{aligned} &(p(2|5) \times \log_2 p(2|5)) \\ &+ (p(3|5) \times \log_2 p(3|5)) \end{aligned} \right) \\
 &= 0.9710
 \end{aligned}$$

$$\begin{aligned}
 &E(\text{Semua}_{\text{sedang}}) \\
 &= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
 &E(\text{Semua}_{\text{sedang}}) = - \left( \begin{aligned} &(p(1|2) \times \log_2 p(1|2)) \\ &+ (p(1|2) \times \log_2 p(1|2)) \end{aligned} \right) \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{malas}) \\
&= - \left( \begin{aligned} & (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ & + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
& E(\text{Semua}_{malas}) = - \left( \begin{aligned} & (p(2|2) \times \log_2 p(2|2)) \\ & + (p(0|2) \times \log_2 p(0|2)) \end{aligned} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut olahraga pada tabel 3.12

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{aligned} & (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{rajin})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{sedang})) \\ & + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{malas})) \end{aligned} \right) \\
&= E(\text{semua}) - \left( \begin{aligned} & (p(5|9) \times (0.9710)) + (p(2|9) \times (1)) \\ & + (p(2|9) \times (0)) \end{aligned} \right) \\
&= 0.2294
\end{aligned}$$

**Tabel 3.12** Hasil perhitungan *gain* atribut usia di atribut IMT nilai sedang

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		9	5	4	0.9911	
<b>Usia</b>	Muda	6	2	4	0.9183	0.3789
	Parobaya	2	2	0	0.0000	
	Tua	1	1	0	0.0000	
<b>Keturunan</b>	Tidak ada	2	0	2	0.0000	0.4628
	Kakek Nenek	2	1	1	1.0000	
	Paman Bibi	3	2	1	0.9183	
	Kakek Nenek Paman Bibi	2	2	0	0.0000	
	Orang Tua	0	0	0	0.0000	
<b>Olahraga</b>	Rajin	5	2	3	0.9710	0.2294
	Sedang	2	1	1	1.0000	
	Malas	2	2	0	0.0000	

Dari hasil perhitungan diatas nilai gain tertinggi terdapat pada atribut keturunan. Maka dilakukan percabangan. Data pemecah cabang kasus keturunan di atribut IMT sedang dapat dilihat pada tabel 3.13

**Tabel 3.13** Data Pemecah Cabang Kasus Keturunan di atribut IMT Sedang

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko
2	Muda	Tidak Ada	Sedang	Rajin	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Parobaya	Kakek Nenek	Sedang	Rajin	Beresiko
2	Muda	Kakek Nenek	Sedang	Rajin	Tidak beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko
2	Muda	Paman Bibi	Sedang	Rajin	Beresiko
3	Tua	Paman Bibi	Sedang	Malas	Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Kakek Nenek Paman Bibi	Sedang	Sedang	Beresiko
2	Parobaya	Kakek Nenek Paman Bibi	Sedang	Malas	Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Tua	Orang Tua	Sedang	Malas	Beresiko
2	Parobaya	Orang Tua	Sedang	Malas	Beresiko

Di data pembagian cabang keturunan tujuan kelas masih heterogen atau berbeda yaitu di atribut keturunan kakek nenek dan paman bibi . Maka dilakukan percabangan lagi. Hasil perhitungan *gain* didapat atribut yang tertinggi yaitu atribut olahraga. Tabel Hasil Perhitungan *gain* Pemecah cabang kasus keturunan paman bibi pada gambar 3.14

❖ Perhitungan manual entropy atribut usia pada tabel 3.14

$$E(\text{semua}) = - \left( \begin{array}{l} p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ + p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua}) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(2|3) \times \log_2 p(2|3)) \\ + (p(1|3) \times \log_2 p(1|3)) \end{array} \right)$$

$$= 0.9183$$

$$E(\text{Semua}_{muda}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{muda}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right)$$

$$= 1$$

$$E(\text{Semua}_{parobaya}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{parobaya}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right)$$

$$= 0$$

$$E(\text{Semua}_{tua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{tua}) = - \left( \begin{array}{l} (p(1|1) \times \log_2 p(1|1)) \\ + (p(0|1) \times \log_2 p(0|1)) \end{array} \right)$$

$$= 0$$

❖ Perhitungan manual gain atribut usia pada tabel 3.14

$$= E(\text{semua}) - \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{muda})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{parobaya})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{tua})) \end{array} \right)$$

$$= E(\text{semua}) - \left( \begin{array}{l} (p(2|3) \times (1)) + (p(0|3) \times (0)) \\ + (p(1|3) \times (0)) \end{array} \right)$$

$$= 0.2516$$

- ❖ Perhitungan manual atribut olahraga pada tabel 3.14

$$E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(1|1) \times \log_2 p(1|1)) \\ + (p(0|1) \times \log_2 p(0|1)) \end{array} \right)$$

$$= 0$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right)$$

$$= 0$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(1|1) \times \log_2 p(1|1)) \\ + (p(0|1) \times \log_2 p(0|1)) \end{array} \right)$$

$$= 0$$

- ❖ Perhitungan manual gain atribut olahraga pada tabel 3.14

$$= E(\text{semua}) - \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right)$$

$$= E(\text{semua}) - \left( \begin{array}{l} (p(1|3) \times (0)) + (p(1|3) \times (0)) \\ + (p(1|3) \times (0)) \end{array} \right)$$

$$= 0.9183$$

**Tabel 3.14** Hasil Perhitungan *gain* Pemecah cabang kasus keturunan paman bibi

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		3	2	1	0.9183	
<b>Usia</b>	Muda	2	1	1	1.0000	0.2516
	Parobaya	0	0	0	0.0000	
	Tua	1	1	0	0.0000	
<b>Olahraga</b>	Rajin	1	1	0	0.0000	0.9183
	Sedang	1	0	1	0.0000	
	Malas	1	1	0	0.0000	

Dari hasil perhitungan diatas nilai gain tertinggi pada atribut olahraga. Maka dilakukan percabngan pada atribut olahraga. Data pemecah cabang kasus olahraga dapat dilihat pada tabel 3.15

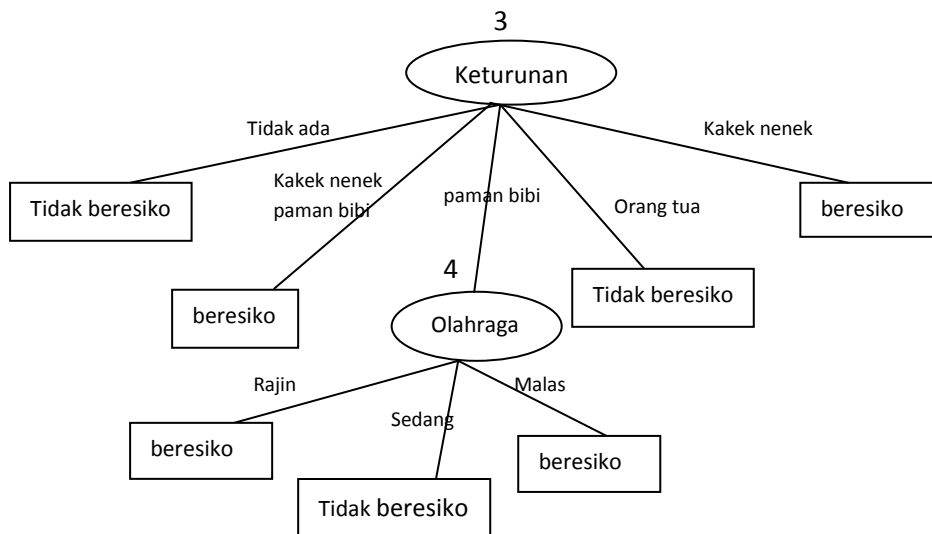
**Tabel 3.15** Data Pemecah cabang kasus olahraga

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Rajin	Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Sedang	Sedang	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Tua	Paman Bibi	Sedang	Malas	Beresiko

Dari data pembagian cabang kasus olahraga percabangan berhenti dan sudah bisa diambil keputusan karena semua atribut dan kelas sama. Atribut keturunan memiliki nilai gain tertinggi. Maka dapat di buat pohon percabangan pada gambar 3.4



**Gambar 3.4** Pohon percabangan keturunan kasus IMT sedang

Lalu memilih atribut sebagai pemecah cabang pada atribut IMT kecil. Nilai *gain* tertinggi didapatkan pada atribut usia. Data pemecah pembagian kasus IMT kecil disajikan pada tabel 3.16 dan hasil perhitungan *gain* n entropy dapat dilihat pada tabel.

- ❖ Perhitungan manual entropy atribut usia pada tabel 3.16

$$E(\text{semua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) = 1$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right) = 1$$

$$\begin{aligned}
& E(\text{Semua}_{\text{parobaya}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(1|1) \times \log_2 p(1|1)) \\ + (p(0|1) \times \log_2 p(0|1)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{tua}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut usia pada tabel 3.16

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{\text{muda}})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{\text{parobaya}})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{\text{tua}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(1|2) \times (0)) + (p(1|2) \times (0)) \\ + (p(0|2) \times (0)) \end{array} \right) \\
&= 1
\end{aligned}$$

❖ Perhitungan manual atribut olahraga pada tabel 3.14

$$\begin{aligned}
& E(\text{Semua}_{\text{rajin}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) \\
&= 0
\end{aligned}$$



$$\begin{aligned}
& E(\text{Semua}_{\text{sedang}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{malas}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut olahraga pada tabel 3.16

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(2|2) \times (1)) + (p(0|2) \times (0)) \\ + (p(0|2) \times (0)) \end{array} \right) \\
&= 0
\end{aligned}$$

**Tabel 3.16** Hasil perhitungan *gain* pada atribut IMT nilai kecil

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		2	1	1	1.0000	
<b>Usia</b>	Muda	1	0	1	0.0000	1.0000
	Parobaya	1	1	0	0.0000	
	Tua	0	0	0	0.0000	
<b>Olahraga</b>	Rajin	2	1	1	1.0000	0.0000
	Sedang	0	0	0	0.0000	
	Malas	0	0	0	0.0000	

Dari hasil perhitungan diatas nilai gain tertinggi terdapat pada atribut usia. Maka dilakukan percabangan atribut usia. Data pemecah cabang kasus usia dapat dilihat pada tabel 3.17

**Tabel 3.17** Data pemecah cabang kasus usia

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Parobaya	Kakek Nenek	Sedang	Rajin	Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Kakek Nenek	Sedang	Rajin	Tidak Beresiko

Dari data pembagian diatas percabangan berhenti sudah bisa di ambil keputusan, karena kelas menghasilkan masing-masing 1 keputusan. Data pemecah cabang kasus keturunan pada IMT kecil dapat dilihat pada tabel 3.18

**Tabel 3.18** Data Pemecah cabang kasus keturunan pada IMT kecil

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Parobaya	Tidak Ada	Kecil	Sedang	Tidak Beresiko
2	Tua	Tidak Ada	Kecil	Sedang	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
2	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Orang Tua	Kecil	Malas	Beresiko
4	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Paman Bibi	Kecil	Sedang	Beresiko
2	Parobaya	Paman Bibi	Kecil	Rajin	Tidak Beresiko
3	Parobaya	Paman Bibi	Kecil	Sedang	Tidak Beresiko
4	Muda	Paman Bibi	Kecil	Sedang	Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Beresiko
2	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Tua	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
2	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Kakek Nenek	Kecil	Rajin	Tidak Beresiko

Di data pembagian cabang keturunan tujuan kelas masih heterogen atau berbeda yaitu orang tua, paman bibi dan kakek nenek paman bibi. Maka dilakukan percabangan. Hasil perhitungan gain dan entropy dapat dilihat pada tabel 3.19

❖ Perhitungan manual entropy atribut usia pada tabel 3.19

$$E(\text{semua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(4|16) \times \log_2 p(4|16)) \\ + (p(12|16) \times \log_2 p(12|16)) \end{array} \right)$$

$$= 0.8113$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{muda}}) = - \left( \begin{array}{l} (p(3|6) \times \log_2 p(3|6)) \\ + (p(3|6) \times \log_2 p(3|6)) \end{array} \right)$$

$$= 1$$

$$\begin{aligned}
& E(\text{Semua}_{\text{parobaya}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{parobaya}}) = - \left( \begin{array}{l} (p(1|8) \times \log_2 p(1|8)) \\ + (p(7|8) \times \log_2 p(7|8)) \end{array} \right) \\
&= 0.5436
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{tua}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{tua}}) = - \left( \begin{array}{l} (p(0|2) \times \log_2 p(0|2)) \\ + (p(2|2) \times \log_2 p(2|2)) \end{array} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut usia pada tabel 3.19

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{\text{muda}})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{\text{parobaya}})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{\text{tua}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(6|16) \times (1)) + (p(8|16) \times (0.5436)) \\ + (p(2|16) \times (0)) \end{array} \right) \\
&= 0.1645
\end{aligned}$$

❖ Perhitungan manual entropy atribut keturunan *diabetes* pada tabel 3.19

$$\begin{aligned}
& E(\text{Semua}_{\text{tidak ada}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{tidak ada}}) = - \left( \begin{array}{l} (p(0|2) \times \log_2 p(0|2)) \\ + (p(2|2) \times \log_2 p(2|2)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{kakek nenek}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{kakek nenek}}) = - \left( \begin{array}{l} (p(0|3) \times \log_2 p(0|3)) \\ + (p(3|3) \times \log_2 p(3|3)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{paman bibi}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \\ \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{paman bibi}}) = - \left( \begin{array}{l} (p(2|4) \times \log_2 p(2|4)) \\ + (p(2|4) \\ \times \log_2 p(2|4)) \end{array} \right) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{kakek nenek paman bibi}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{kakek nenek paman bibi}}) = - \left( \begin{array}{l} (p(1|3) \times \log_2 p(1|3)) \\ + (p(2|3) \times \log_2 p(2|3)) \end{array} \right) \\
&= 0.9183
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{orang tua}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{orang tua}}) = - \left( \begin{array}{l} (p(1|4) \times \log_2 p(1|4)) \\ + (p(3|4) \times \log_2 p(3|4)) \end{array} \right) \\
&= 0.8113
\end{aligned}$$

❖ Perhitungan gain atribut keturunan *diabetes* pada tabel 3.19

$$\begin{aligned}
 &= E(\text{Semua}) \\
 &= - \left( \begin{aligned} &(p(\text{tidak ada}|\text{semua}) \times E(\text{semua}_{\text{tidak ada}})) + (p(\text{kakek nenek}|\text{semua}) \times E(\text{semua}_{\text{kakek nenek}})) \\ &+ (p(\text{paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ &(p(\text{kakek nenek paman bibi}|\text{semua}) \times E(\text{semua}_{\text{paman bibi}})) + \\ &(p(\text{orang tua}|\text{semua}) \times E(\text{semua}_{\text{orang tua}})) \end{aligned} \right) \\
 &= E(0.9837) - \left( \begin{aligned} &(p(2|16) \times E(0)) + (p(3|16) \times E(0)) \\ &+ (p(4|16) \times E(1)) + \\ &(p(3|16) \times E(0.9183)) + \\ &(p(4|16) \times E(0.8113)) \end{aligned} \right) \\
 &= 0.1863
 \end{aligned}$$

❖ Perhitungan manual atribut olahraga pada tabel 3.19

$$\begin{aligned}
 &E(\text{Semua}_{\text{rajin}}) \\
 &= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
 &E(\text{Semua}_{\text{rajin}}) = - \left( \begin{aligned} &(p(0|3) \times \log_2 p(0|3)) \\ &+ (p(3|3) \times \log_2 p(3|3)) \end{aligned} \right) \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 &E(\text{Semua}_{\text{sedang}}) \\
 &= - \left( \begin{aligned} &(p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ &+ (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{aligned} \right) \\
 &E(\text{Semua}_{\text{sedang}}) = - \left( \begin{aligned} &(p(3|10) \times \log_2 p(3|10)) \\ &+ (p(7|10) \times \log_2 p(7|10)) \end{aligned} \right) \\
 &= 0.8813
 \end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{\text{malas}}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(1|3) \times \log_2 p(1|3)) \\ + (p(2|3) \times \log_2 p(2|3)) \end{array} \right) \\
&= 0.9183
\end{aligned}$$

❖ Perhitungan manual gain atribut olahraga pada tabel 3.19

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(3|16) \times (0)) + (p(10|16) \times (0.8813)) \\ + (p(3|16) \times (0.9183)) \end{array} \right) \\
&= 0.0883
\end{aligned}$$

**Tabel 3.19** Hasil perhitungan *gain* kasus atribut keturunan nilai orang tua

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		16	4	12	0.8113	
<b>Usia</b>	Muda	6	3	3	1.0000	0.1645
	Parobaya	8	1	7	0.5436	
	Tua	2	0	2	0.0000	
<b>Keturunan</b>	Tidak ada	2	0	2	0.0000	0.1863
	Kakek Nenek	3	0	3	0.0000	
	Paman Bibi	4	2	2	1.0000	
	Kakek Nenek Paman Bibi	3	1	2	0.9183	
	Orang Tua	4	1	3	0.8113	
<b>Olahraga</b>	Rajin	3	0	3	0.0000	0.0883
	Sedang	10	3	7	0.8813	
	Malas	3	1	2	0.9183	

Dari hasil perhitungan di atas gain tertinggi pada atribut keturunan diabetes. Keputusan mengambil keturunan diabetes. Di atribut keturunan dilakukan percabangan. Dengan data pemecah pada kasus keturunan diabetes tabel 3.20 dan hasil perhitungan dapat dilihat pada tabel 3.21

**Tabel 3.20** Data pemecah cabang pada kasus keturunan diabetes

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Parobaya	Tidak Ada	Kecil	Sedang	Tidak Beresiko
2	Tua	Tidak Ada	Kecil	Sedang	Tidak Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
2	Muda	Orang Tua	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Orang Tua	Kecil	Malas	Beresiko
4	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Paman Bibi	Kecil	Sedang	Beresiko
2	Parobaya	Paman Bibi	Kecil	Rajin	Tidak Beresiko
3	Parobaya	Paman Bibi	Kecil	Sedang	Tidak Beresiko
4	Muda	Paman Bibi	Kecil	Sedang	Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Beresiko
2	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Kakek Nenek Paman Bibi	Kecil	Malas	Tidak Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Tua	Kakek Nenek	Kecil	Rajin	Tidak Beresiko
2	Parobaya	Kakek Nenek	Kecil	Sedang	Tidak Beresiko
3	Parobaya	Kakek Nenek	Kecil	Rajin	Tidak Beresiko

❖ Perhitungan manual entropy atribut usia pada tabel 3.21

$$\begin{aligned}
 E(\text{semua}) &= - \left( \begin{aligned} & p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua}) \\ & + p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua}) \end{aligned} \right) \\
 E(\text{semua}) &= - \left( \begin{aligned} & (p(1|4) \times \log_2 p(1|4)) \\ & + (p(3|4) \times \log_2 p(3|4)) \end{aligned} \right) \\
 &= 0.8113
 \end{aligned}$$



$$\begin{aligned}
& E(\text{Semua}_{muda}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{muda}) = - \left( \begin{array}{l} (p(0|2) \times \log_2 p(0|2)) \\ + (p(2|2) \times \log_2 p(2|2)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{parobaya}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{parobaya}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{tua}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{tua}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut usia pada tabel 3.21

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{muda})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{parobaya})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{tua})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(2|4) \times (0)) + (p(2|4) \times (1)) \\ + (p(0|4) \times (0)) \end{array} \right) \\
&= 0.3113
\end{aligned}$$

- ❖ Perhitungan manual atribut olahraga pada tabel 3.21

$$E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$\begin{aligned} E(\text{Semua}_{\text{rajin}}) &= - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\ &= 0 \end{aligned}$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$\begin{aligned} E(\text{Semua}_{\text{sedang}}) &= - \left( \begin{array}{l} (p(0|2) \times \log_2 p(0|2)) \\ + (p(2|2) \times \log_2 p(2|2)) \end{array} \right) \\ &= 0 \end{aligned}$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$\begin{aligned} E(\text{Semua}_{\text{malas}}) &= - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) \\ &= 1 \end{aligned}$$

- ❖ Perhitungan manual gain atribut olahraga pada tabel 3.21

$$\begin{aligned} &= E(\text{semua}) \\ &- \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right) \end{aligned}$$

$$= E(\text{semua}) - \left( \begin{array}{l} (p(0|4) \times (0)) + (p(2|4) \times (0)) \\ + (p(2|4) \times (1)) \end{array} \right)$$

$$= 0.3113$$

**Tabel 3.21** Hasil perhitungan *gain* kasus usia

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		4	1	3	0.8113	
<b>Usia</b>	Muda	2	0	2	0.0000	0.3113
	Parobaya	2	1	1	1.0000	
	Tua	0	0	0	0.0000	
<b>Olahraga</b>	Rajin	0	0	0	0.0000	0.3113
	Sedang	2	0	2	0.0000	
	Malas	2	1	1	1.0000	

Dari perhitungan *gain* di atas didapat *gain* yang sama. Maka dipilih salah satu. Dipilih atribut usia. Di buat data pemecah cabang pada tabel 3.22

**Tabel 3.22** Data pemecah cabang kasus usia

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Parobaya	Orang Tua	Kecil	Malas	<b>Beresiko</b>
2	Parobaya	Orang Tua	Kecil	Malas	Tidak Beresiko

No	Usia	Keturunan diabet	BMI	Olahraga	Pengidap
1	Muda	Orang Tua	Kecil	Sedang	<b>Tidak Beresiko</b>
2	Muda	Orang Tua	Kecil	Sedang	<b>Tidak Beresiko</b>

Kemudian yang terakhir di cari percabangan dari kasus kakek nenek paman bibi. Hasil perhitungan *gain* pada tabel 3.26

- ❖ Perhitungan manual entropy atribut usia pada tabel 3.23

$$E(\text{semua}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ +(p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{semua}) = - \left( \begin{array}{l} (p(1|3) \times \log_2 p(1|3)) \\ +(p(2|3) \times \log_2 p(2|3)) \end{array} \right)$$

$$= 0.9183$$

$$\begin{aligned}
& E(\text{Semua}_{muda}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{muda}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{parobaya}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{parobaya}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
& E(\text{Semua}_{tua}) \\
&= - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right) \\
& E(\text{Semua}_{tua}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right) \\
&= 0
\end{aligned}$$

❖ Perhitungan manual gain atribut usia pada tabel 3.23

$$\begin{aligned}
&= E(\text{semua}) \\
&- \left( \begin{array}{l} (p(\text{muda}|\text{semua}) \times E(\text{semua}_{muda})) + (p(\text{parobaya}|\text{semua}) \times E(\text{semua}_{parobaya})) \\ + (p(\text{tua}|\text{semua}) \times E(\text{semua}_{tua})) \end{array} \right) \\
&= E(\text{semua}) - \left( \begin{array}{l} (p(2|3) \times (1)) + (p(1|3) \times (0)) \\ + (p(0|3) \times (0)) \end{array} \right) \\
&= 0.2516
\end{aligned}$$

- ❖ Perhitungan manual entropy atribut olahraga pada tabel 3.23

$$E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{rajin}}) = - \left( \begin{array}{l} (p(0|0) \times \log_2 p(0|0)) \\ + (p(0|0) \times \log_2 p(0|0)) \end{array} \right)$$

$$= 0$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{sedang}}) = - \left( \begin{array}{l} (p(1|2) \times \log_2 p(1|2)) \\ + (p(1|2) \times \log_2 p(1|2)) \end{array} \right)$$

$$= 1$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(\text{beresiko}|\text{semua}) \times \log_2 p(\text{beresiko}|\text{semua})) \\ + (p(\text{tidak beresiko}|\text{semua}) \times \log_2 p(\text{tidak beresiko}|\text{semua})) \end{array} \right)$$

$$E(\text{Semua}_{\text{malas}}) = - \left( \begin{array}{l} (p(0|1) \times \log_2 p(0|1)) \\ + (p(1|1) \times \log_2 p(1|1)) \end{array} \right)$$

$$= 0$$

- ❖ Perhitungan manual gain atribut olahraga pada tabel 3.23

$$= E(\text{semua}) - \left( \begin{array}{l} (p(\text{rajin}|\text{semua}) \times E(\text{semua}_{\text{rajin}})) + (p(\text{sedang}|\text{semua}) \times E(\text{semua}_{\text{sedang}})) \\ + (p(\text{malas}|\text{semua}) \times E(\text{semua}_{\text{malas}})) \end{array} \right)$$

$$= E(\text{semua}) - \left( \begin{array}{l} (p(0|3) \times (1)) + (p(2|3) \times (1)) \\ + (p(1|3) \times (0)) \end{array} \right)$$

$$= 0.2516$$

**Tabel 3.23** Hasil perhitungan gain kasus usia

		Jumlah	Beresiko	Tidak Beresiko	Entropy	Gain
<b>Total</b>		3	1	2	0.9183	
<b>Usia</b>	Muda	2	1	1	1.0000	0.2516
	Parobaya	1	0	1	0.0000	
	Tua	0	0	0	0.0000	
<b>Olahraga</b>	Rajin	0	0	0	0.0000	0.2516
	Sedang	2	1	1	1.0000	
	Malas	1	0	1	0.0000	

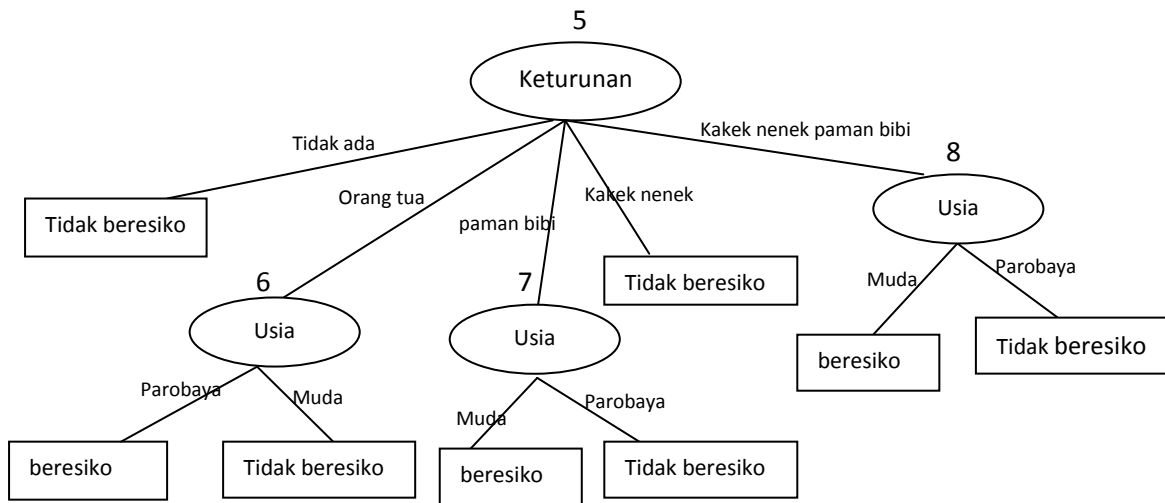
Dari perhitungan *gain* di atas nilai *gain* sama, maka diambil salah satu. Yang diambil atribut usia. Maka di buat data pemecah cabang pada tabel 3.25

**Tabel 3.24** Data pemecah cabang kasus usia

No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	<b>Beresiko</b>
2	Muda	Kakek Nenek Paman Bibi	Kecil	Sedang	Tidak Beresiko

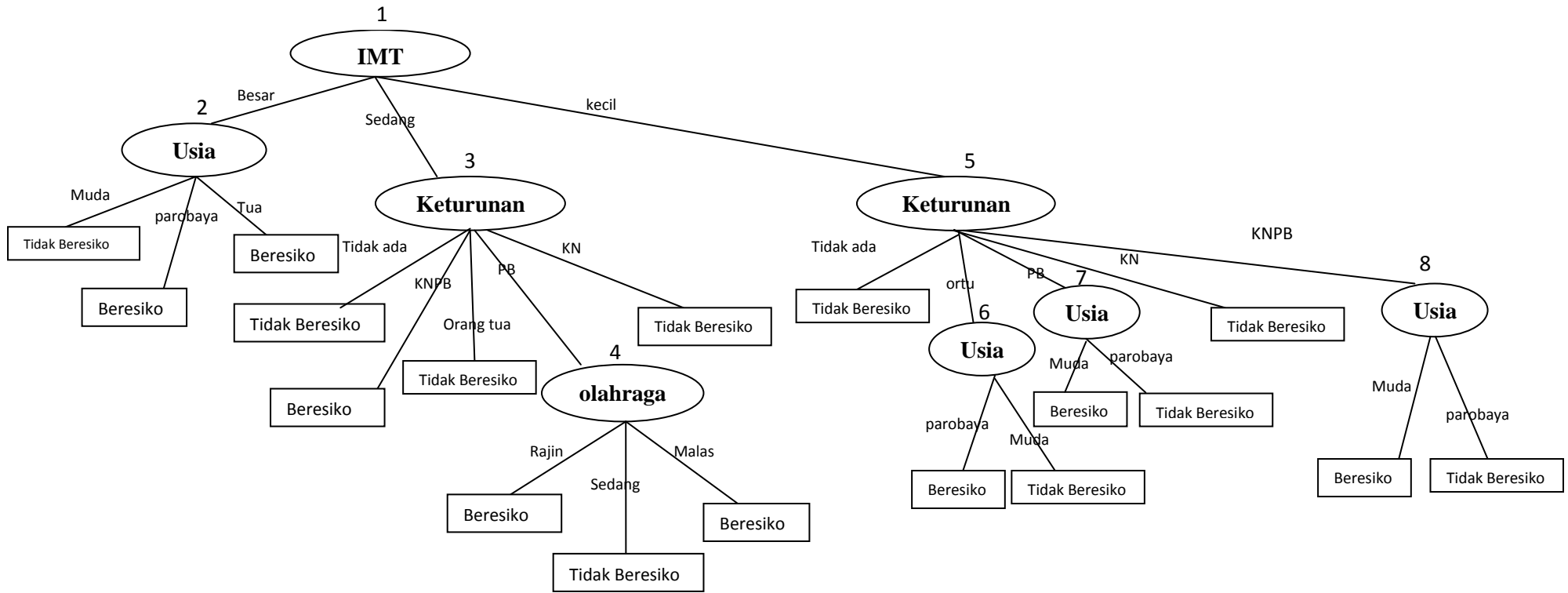
No	Usia	Keturunan diabet	IMT	Olahraga	Pengidap
1	Parobaya	Kakek Nenek Paman Bibi	Kecil	Malas	<b>Tidak Beresiko</b>

Dari hasil data pemecah cabang kasus usia diatas. Percabangan berhenti Karena sudah diperoleh hasil kelas tujuan, maka dari semua data dan hasil perhitungan diatas atribut keturunan yang memiliki nilai *gain* tertinggi, Maka atribut keturunan yang di ambil untuk dilakukan percabangan keturunan pada kasus IMT kecil yang disajikan pada gambar dapat di buat pohon percabangan pada gambar 3.5



**Gambar 3.5** Pohon percabangan keturunan kasus IMT kecil

Pembentukan pohon keputusan dinyatakan selesai, karena sudah tidak ada node yang harus diproses dan semua kasus sudah menjadi daun. Pohon keputusan yang terbentuk seperti disajikan pada gambar 3.6.



**Gambar 3.6** Pohon keputusan yang terbentuk



Keterangan :

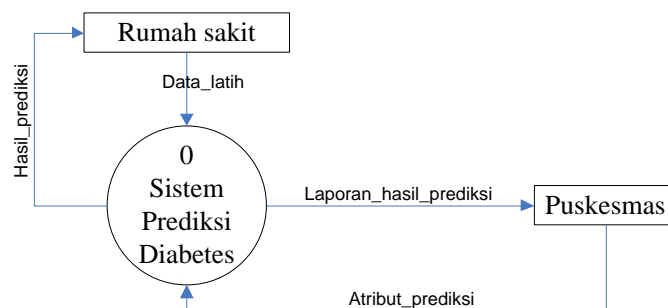
- PB : Paman bibi  
 KNPB : Kakek nenek paman bibi  
 KN : Kakek nenek  
 Ortu : Orang Tua

### 3.5 Perancangan Sistem

Bagian ini akan menjelaskan rancangan sistem seperti diagram context, diagram berjenjang dan *data flow diagram* (DFD).

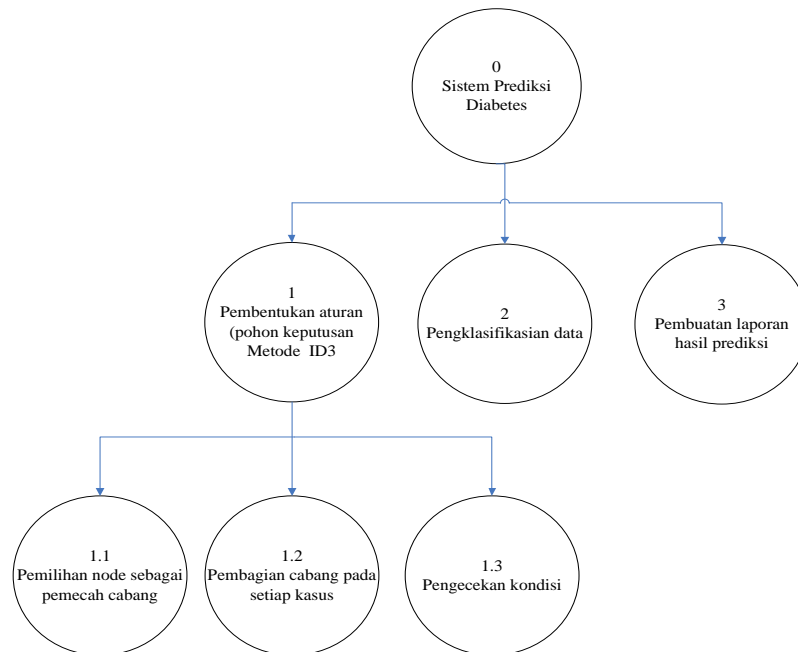
#### 3.5.1 Diagram Context

Dari gambar 3.7 terlihat bahwa yang terlibat dalam sistem (*entity*) ini adalah rumah sakit dan puskesmas. Rumah sakit memasukkan hasil kuisioner atau data yang akan diprediksi dan akan mendapatkan informasi hasil prediksi penyakit diabetes berdasarkan data yang telah dimasukkan. Sedangkan puskesmas memasukkan atribut penyakit diabetes yang terdiri dari Usia, keturunan diabetes, index masa tubuh (IMT), intensitas olahraga. Data tersebut digunakan sebagai *data training* atau data yang akan diproses untuk pembentukan pohon keputusan. Laporan hasil prediksinya dapat dilihat oleh puskesmas. *Diagram context* ditunjukkan pada gambar 3.7



**Gambar 3.7** *Diagram Context*

### 3.5.2 Diagram Berjenjang



**Gambar 3.8** Diagram Berjenjang

Diagram berjenjang disajikan pada gambar 3.8 berikut penjelasannya:

Top level : Sistem prediksi resiko diabetes

1.0 Pembentukan aturan (pohon keputusan) dengan metode ID3 berdasarkan data kuissioner, yang didalamnya terdapat tiga proses.

1.1 Pemilihan node sebagai pemecah cabang

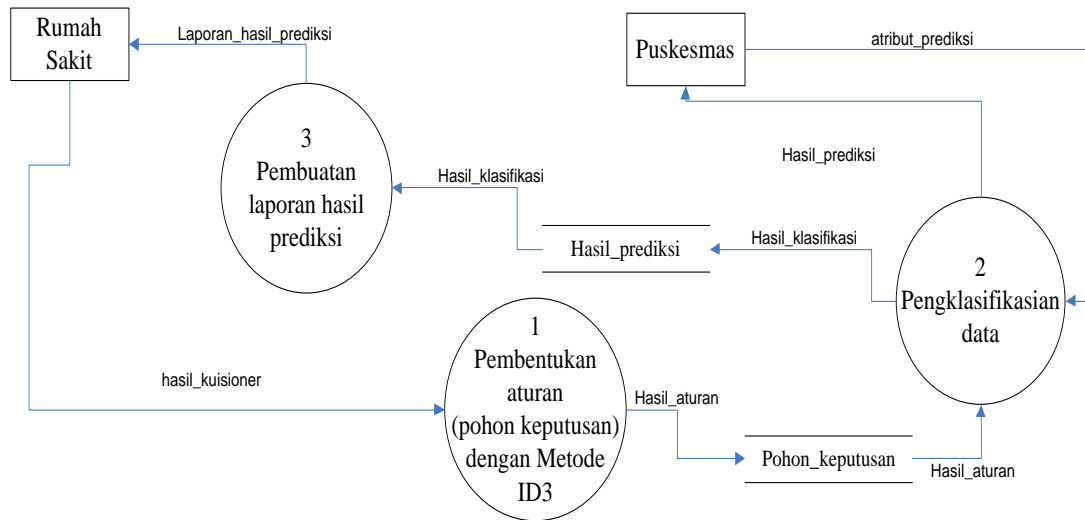
1.2 Pembagian cabang pada setiap kasus

1.3 Pengecekan kondisi, yaitu jika masih ada kasus yang memiliki kelas yang berbeda maka mengulangi.

2.0 Pengklasifikasian data uji menggunakan aturan yang sudah terbentuk

3.0 Pembuatan laporan hasil prediksi

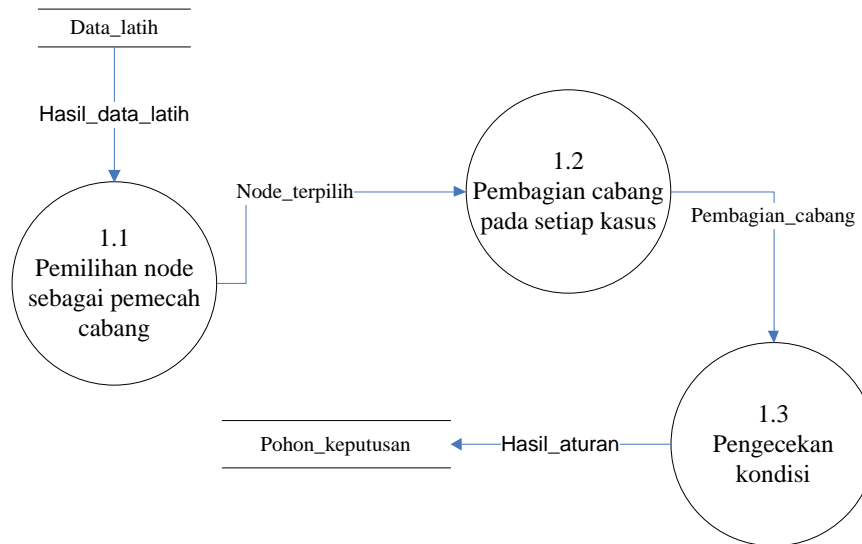
### 3.5.3 Data Flow Diagram Level 0



**Gambar 3.9** DFD Level 0 proses aliran data pada sistem

DFD level 0 pada gambar 3.9 menjelaskan aliran data pada sistem. Terdapat tiga proses didalam sistem tersebut. Proses pertama pembentukan aturan (pohon keputusan) dengan metode id3. Dari atribut kuisisioner terdapat hasil kuisisioner atau data latih yang merupakan data untuk proses pembentukan pohon keputusan. Pada proses kedua yaitu klasifikasi data uji membutuhkan masukkan dari rumah sakit berupa hasil kuisisioner data yang akan diprediksi. Pemeriksaan dilakukan di puskesmas. Hasil pemeriksaan diberikan ke rumah sakit melalui sistem yang akan disimpan ke-hasil prediksi. Proses ketiga adalah pembuatan laporan hasil prediksi penyakit diabetes yang akan diberikan kepada rumah sakit dengan mengambil data dari tabel hasil prediksi.

### 3.5.4 Data Flow Diagram Level 1



**Gambar 3.10** DFD Level 1 proses pembentukan aturan (pohon keputusan)

DFD level 1 proses pembentukan aturan pada gambar 3.10 adalah pembentukan aturan menggunakan metode *decision tree id3* ini memiliki tiga proses didalamnya yaitu, proses pemilihan node yang akan dijadikan sebagai pemecah cabang, membagi cabang pada setiap kasus, dan proses pengecekan kondisi, jika ada kasus yang memiliki kelas berbeda maka akan mengulangi pada proses pemilihan *node*. Hasil dari proses ini adalah aturan atau pohon keputusan yang akan disimpan pada *data base*.

### 3.6 Kebutuhan Pembuatan Sistem

Dalam proses pembuatan sistem ini membutuhkan beberapa komponen yang dibagi menjadi kebutuhan perangkat lunak (*software*) dan kebutuhan perangkat keras (*hardware*).

- **Kebutuhan Perangkat Lunak (*Software*)**

Perangkat lunak yang dibutuhkan dalam pembangunan aplikasi ini adalah sebagai berikut :

- Microsoft Win 7 Ultimate
- Mozilla Firefox 12
- Editor PHP (Notepad++)
- SQLyog Enterprise
- Mysql
- Apache

- **Kebutuhan Perangkat Keras (*Hardware*)**

Perangkat keras yang dibutuhkan dalam pembangunan aplikasi ini memiliki spesifikasi sebagai berikut :

- Prosesor Intel Core i3
- RAM 2 GB
- HDD 500 GB
- Monitor 14"

### 3.7 Struktur Tabel

Struktur tabel ini menjelaskan tabel atau tempat penyimpanan data yang digunakan untuk keperluan sistem yang akan dibangun. Berikut adalah struktur dari tabel-tabel yang akan digunakan.

- **Data\_latih**

Tabel ini digunakan untuk menyimpan data latih atau data yang akan diproses pada pembentukan pohon keputusan. Struktur tabel data latih dapat dilihat pada tabel 3.25

**Tabel 3.25** Struktur tabel data latih

No	Name_field	Type	Length	Key
1	Id	Int	11	Primary key
2	Usia	Varchar	11	
3	Keturunan diabetes	Varchar	25	
4	Index masa tubuh (IMT)	Varchar	15	
5	Intensitas Olahraga	Varchar	15	
6	Pengidap diabetes	Varchar	20	

- Data\_uji

Tabel ini digunakan untuk menyimpan data pengujian, yaitu untuk menguji tingkat akurasi dari pohon keputusan yang terbentuk. Strukturnya sama dengan tabel data latih dengan ditambahi *field* hasil prediksi. Struktur tabel data gaji dapat dilihat pada tabel 3.26

**Tabel 3.26** Struktur tabel data uji

No	Name_field	Type	Length	Key
1	Id	Int	11	Primary key
2	Usia	Varchar	10	
3	Keturunan diabetes	Varchar	10	
4	Index masa ubuh (IMT)	Varchar	20	
5	Intensitas olahraga	Varchar	15	
6	Pengidap_asli	Varchar	20	
7	Pengidap_prediksi	Varchar	20	

- Hasil\_prediksi

Tabel ini digunakan untuk menyimpan data hasil prediksi. Strukturnya sama dengan tabel data training namun *field* ipk diganti dengan hasil. Struktur tabel hasil prediksi dapat dilihat pada tabel 3.27

**Tabel 3.27** Struktur tabel hasil prediksi

No	Name_field	Type	Length	Key
1	Id	Int	11	Primary key
2	Usia	Varchar	11	
3	Keturunan diabetes	Varchar	25	
4	Index masa tubuh (IMT)	Varchar	15	
5	Intensitas olahraga	Varchar	15	
6	Hasil_prediksi	Varchar	20	

- User

Tabel ini digunakan untuk menyimpan data *user*, baik itu user sebagai rumah sakit atau sebagai puskesmas. Struktur tabel user dapat dilihat pada tabel 3.28

**Tabel 3.28** Struktur tabel user

No	Name_field	Type	Length	Key
1	user_id	Varchar	25	Primary key
2	Nama	Varchar	50	
3	Password	Text		
4	Level	Varchar	10	

- Pohon keputusan

Tabel ini menampung hasil dari proses pembentukan pohon keputusan, yaitu menampung aturan-aturan yang telah terbentuk. Struktur tabel aturan dapat dilihat pada tabel 3.29

**Tabel 3.29** Struktur tabel aturan

No	Name_field	Type	Length	Key
1	Id	Int	11	Primary key
2	Parent	Text		
3	Akar	Text		
4	Keputusan	Varchar	10	

### 3.8 Desain Antar Muka

Tampilan antar muka pengguna sebagai mahasiswa yaitu halaman *login*, prediksi, hasil prediksi, ubah password dan pohon keputusan. Sedangkan pengguna sebagai kaprodi adalah *login*, *mining*, pohon keputusan, uji pohon keputusan, daftar hasil prediksi, dan data *user*. Gambar rancangan halaman login dapat dilihat pada gambar 3.11

#### ➤ Halaman Login (Rumah Sakit/Puskesmas)

LOGIN	
SISTEM PREDIKSI TINGKAT RESIKO PENYAKIT DIABETES MENGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3	
<input type="text"/>	
<input type="text"/>	
USER	:
<input type="button" value="LOGIN"/>	

**Gambar 3.11** Rancangan halaman *login*

Halaman *login* diperlukan untuk mengetahui pengguna yang masuk kedalam sistem adalah pengguna sebagai rumah sakit atau sebagai puskesmas.

#### ➤ Halaman Utama (Rumah Sakit/Puskesmas)

Menu yang ditampilkan untuk pengguna sebagai puskesmas adalah menu *home*, prediksi, dan pohon keputusan. Sedangkan untuk pengguna sebagai rumah sakit yang ditampilkan adalah menu home, olah data, mining, pohon keputusan, hasil, dan data user. Gambar rancangan halaman utama dapat dilihat pada gambar 3.12

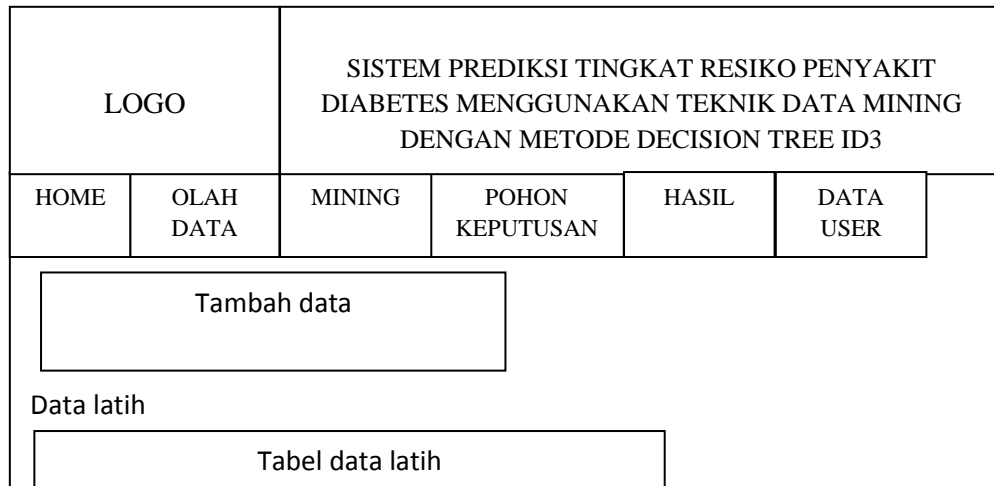
LOGO	SISTEM PREDIKSI TINGKAT RESIKO PENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3					
Home	Olah data	Mining	Pohon	Hasil	Prediksi	Data user
Anda login sebagai user/admin   Logout						
Selamat Datang di Aplikasi Sistem Prediksi Tingkat Resiko Penyakit Diabetes...						

**Gambar 3.12** Rancangan halaman utama



➤ **Halaman Olah Data (Rumah Sakit)**

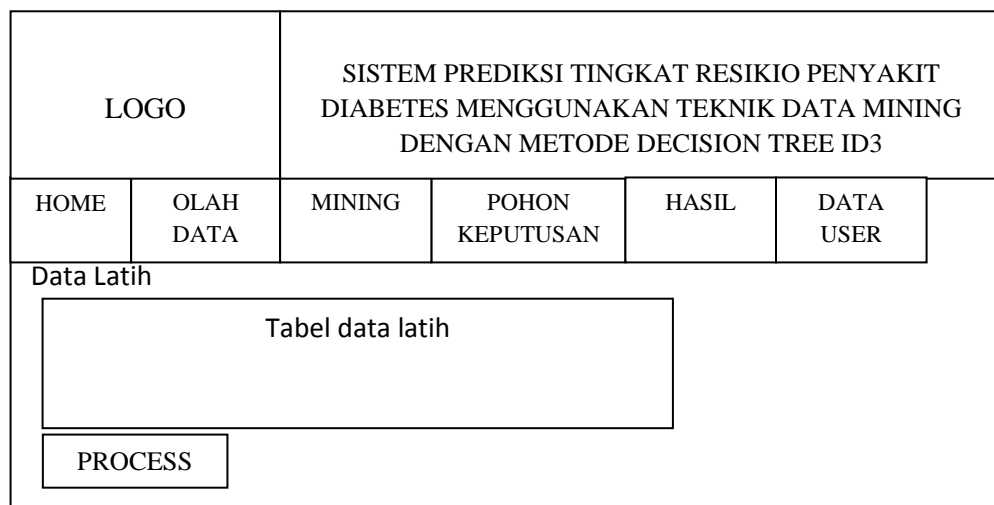
Tampilan ini adalah halaman yang akan digunakan untuk proses pembentukan pohon keputusan. Halaman ini hanya bisa diakses oleh pihak rumah sakit. Gambar rancangan halaman olah data dapat dilihat pada gambar 3.13



**Gambar 3.13** Rancangan halaman olah data

➤ **Halaman Mining (Rumah Sakit)**

Tampilan ini adalah halaman yang akan digunakan untuk proses pembentukan pohon keputusan. Halaman ini hanya bisa diakses oleh rumah. Sakit. Gambar rancangan halaman mining dapat dilihat pada gambar 3.14



**Gambar 3.14** Rancangan halaman mining

➤ **Halaman Pohon Keputusan (Rumah Sakit/Puskesmas)**

Halaman ini menampilkan pohon keputusan atau aturan yang didapat dari proses mining. Pohon keputusan ini juga dapat diuji keakurasiannya. Gambar rancangan halaman pohon keputusan dapat dilihat pada gambar 3.15

LOGO		SISTEM PREDIKSI TINGKAT RESIKIOPENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3			
HOME	OLAH DATA	MINING	POHON KEPUTUSAN	HASIL	DATA USER
<p>Rule / aturan yang terbentuk adalah sebagai berikut:</p> <p>Opsi: Hapus Pohon Keputusan</p> <p>1. JIKA (usia=muda) MAKA pengidap = beresiko</p> <p>2. JIKA (usia=parobaya) &amp; (keturunan diabetes= orang tua) MAKA pengidap = tidak beresiko</p> <p>3. JIKA (usia=tua) &amp; (keturunan diabetes = paman bibi) MAKA pengidap = Tinggi</p> <p>4. JIKA (usia=tua) &amp; (keturunan diabetes = tidak ada) &amp; (olahraga = malas) MAKA pengidap = Rendah</p>					
<div style="border: 1px solid black; padding: 5px; display: inline-block;">Uji Pohon Keputusan</div>					

**Gambar 3.15** Rancangan halaman pohon keputusan

➤ **Halaman Uji Pohon Keputusan (Rumah Sakit)**

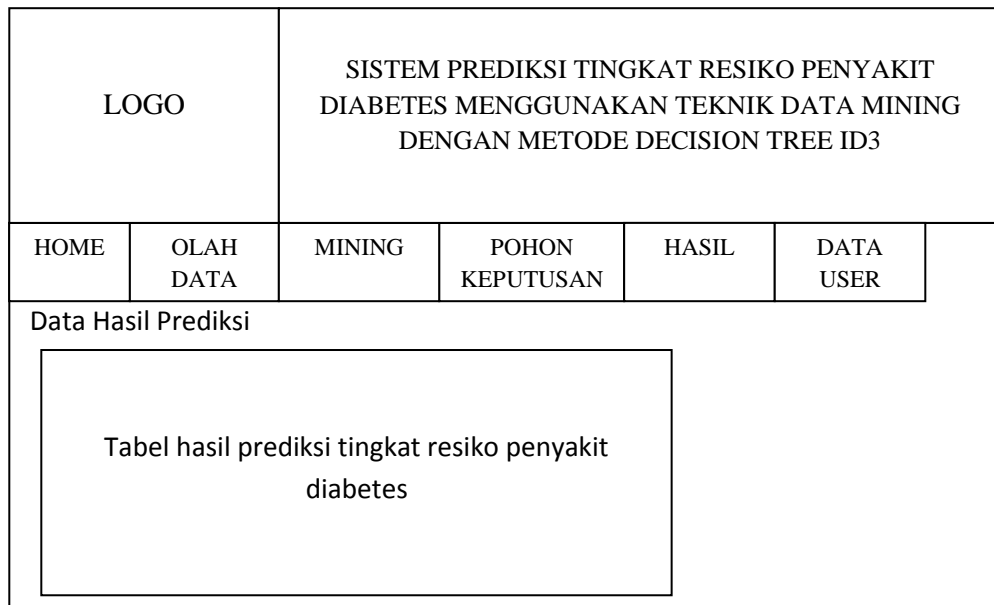
Halaman ini digunakan untuk menguji tingkat akurasi pohon keputusan gambar yang terbentuk dari proses mining. Gambar halaman uji pohon keputusan da gambar dapat dilihat pada gambar 3.16

LOGO		SISTEM PREDIKSI TINGKAT RESIKIOPENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3			
HOME	OLAH DATA	MINING	POHON KEPUTUSAN	HASIL	DATA USER
<div style="border: 1px solid black; padding: 10px; width: fit-content; margin: auto;">Tabel Data Uji</div>					
<p>Jumlah datayang diprediksi : 30</p> <p>Data yang hasil prediksinya tepat ada : 27</p> <p>Data yang hasil prediksinya tidak tepat ada : 3</p> <p>Akurasi = 90%      laju eror=10%</p>					

**Gambar 3.16** Rancangan halaman uji pohon keputusan

➤ **Halaman Hasil (Rumah Sakit)**

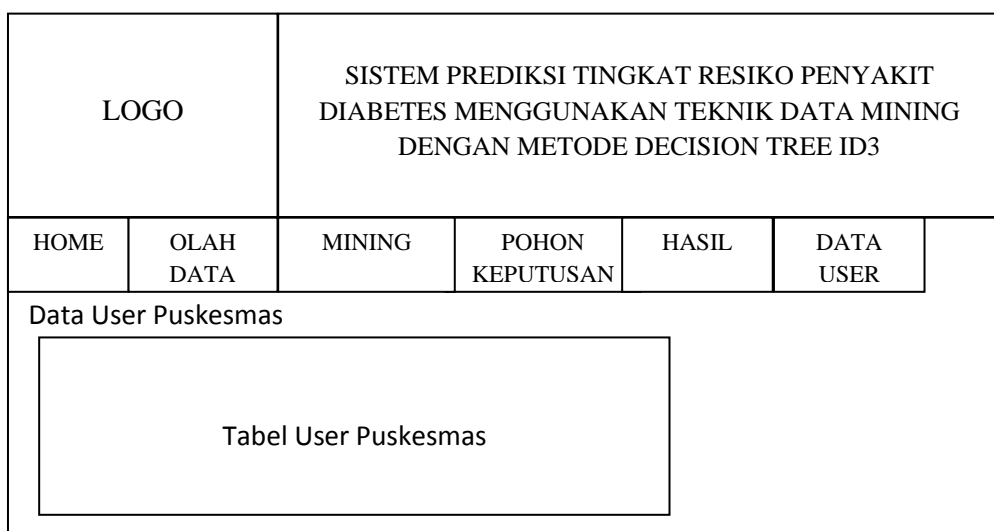
Halaman hasil ini akan menampilkan daftar hasil prediksi dari puskesmas yang telah melakukan prediksi. Gambar rancangan halaman hasil dapat dilihat pada gambar 3.17



**Gambar 3.17** Rancangan halaman hasil

➤ **Halaman Data User (Rumah Sakit)**

Halaman ini digunakan untuk mengolah data pengguna sebagai puskesmas. Rancangan halaman data user dapat dilihat pada gambar 3.18



**Gambar 3.18** Rancangan halaman data user

➤ **Halaman Prediksi (Puskesmas)**

Halaman ini digunakan puskesmas untuk memasukkan data atributnya yang akan prediksi. Gambar rancangan halaman prediksi dapat dilihat pada gambar 3.19

LOGO		SISTEM PREDIKSI TINGKAT RESIKO PENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3	
HOME	PREDIKSI	POHON KEPUTUSAN	Ubah Password
Silahkan masukkan datanya..			
Usia : <input type="radio"/> Muda <input type="radio"/> Parobaya <input type="radio"/> Tua			
Intensitas Olahraga : <input type="radio"/> Rajin <input type="radio"/> Sedang <input type="radio"/> Malas			
Keturunan diabetes : <input type="radio"/> Tidak ada <input type="radio"/> KNPR <input type="radio"/> PR <input type="radio"/> KN <input type="radio"/> Orang tua			
IMT : <input type="radio"/> Besar <input type="radio"/> Sedang <input type="radio"/> Malas			
<input type="button" value="SUBMIT"/>			

**Gambar 3.19** Rancangan halaman prediksi

➤ **Tampilan Halaman Hasil Prediksi (Puskesmas)**

Hasil prediksi ini adalah hasil prediksi pasien yang telah melakukan prediksi. Gambar rancangan halaman hasil prediksi dapat dilihat pada gambar 3.20

LOGO		SISTEM PREDIKSI TINGKAT RESIKO PENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3	
HOME	PREDIKSI	POHON KEPUTUSAN	Ubah Password
Data yang telah Anda masukkan adalah			
Usia : Muda			
Intensitas Olahraga : Rajin			
Keturunan diabetes : Kakek nenek			
IMT : Besar			
Status Kerja : Belum			
HASIL PREDIKSI RESIKO PENYAKIT DIABETES ANDA BERESIKO			

**Gambar 3.20** Rancangan halaman hasil prediksi

Halaman ubah password ini adalah halaman untuk mengganti password user sebagai puskesmas. Gambar rancangan halaman ubah password dapat dilihat pada gambar 3.21

LOGO		SISTEM SISTEM PREDIKSI TINGKAT RESIKO PENYAKIT DIABETES MENGGUNAKAN TEKNIK DATA MINING DENGAN METODE DECISION TREE ID3	
HOME	PREDIKSI	POHON KEPUTUSAN	Ubah Password
Password lama		:	<input type="text"/>
Password baru		:	<input type="text"/>
Password konfirmasi		:	<input type="text"/>
<input type="button" value="ubah"/>			

**Gambar 3.21** Rancangan halaman ubah password

### 3.8 Evaluasi Sistem

Selanjutnya adalah menentukan seberapa *classifier* tersebut akurat dalam memprediksi. Evaluasi dilakukan dengan menguji data set yang diprediksi secara benar kategori beresiko dan tidak beresiko dengan menggunakan *Confusion Matrix*. *Confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik pengklasifikasi tersebut dapat mengenali tupel dalam kelas-kelas yang berbeda. Berikut tabel *confusion matrix* dalam memprediksikan kelas atau output seperti pada tabel 3.30

**Tabel 3.30** *Confusion Matrix*

		Kelas Hasil Prediksi	
		Beresiko	Tidak beresiko
Kelas Asli	Beresiko	True Positive (TP)	False Negative (FN)
	Tidak beresiko	False Positive (FP)	True Negative (TN)

Keterangan :

TP : beresiko yang diprediksi secara benar sebagai beresiko

FP : tidak beresiko yang diprediksi secara salah sebagai beresiko

TN : tidak beresiko yang diprediksi secara benar sebagai tidak beresiko

FN : Bresiko yang diprediksi secara salah sebagai tidak beresiko

Dari tabel *confusion matrix* tersebut, dapat dihitung tingkat akurasi, laju error, sensitivitas dan spesifisitas seperti di bawah ini :

a. Akurasi Pengelompokan

Akurasi digunakan untuk mengukur prosentase pengenalan secara keseluruhan dan dihitung sebagai jumlah data uji yang dikenali dengan benar dibagi dengan jumlah seluruh data uji. Berikut rumus akurasi dan laju error berdasarkan tabel *confusion matrix*.

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \\ &= \frac{TP + TN}{TP + FN + FP + TN} \end{aligned}$$

$$\begin{aligned} \text{Laju error} &= \frac{\text{Jumlah data yang diprediksi secara salah}}{\text{jumlah prediksi yang dilakukan}} \\ &= \frac{FN + FP}{TP + FN + FP + TN} \end{aligned}$$

a. Sensitivitas dan Spesifisitas

Sensitivitas digunakan untuk mengukur proporsi positif asli yang dikenali (diprediksi) secara benar sebagai positif asli. Sedangkan Spesifisitas digunakan untuk mengukur proporsi negatif asli yang dikenali (diprediksi) secara benar sebagai negatif asli. Berikut rumus sensitivitas dan spesifisitas berdasarkan tabel *confusion matrix*. Tabel perhitungan evaluasi sitem dapat dilihat pada tabel 3.41

$$\text{Sensitivitas} = \frac{TP}{TP + FN}$$

$$\text{Spesifisitas} = \frac{TN}{FP + TN}$$

**Tabel 3.31** perhitungan evaluasi sistem

		Kelas Hasil Prediksi	
		Beresiko	Tidak beresiko
Kelas Asli	Beresiko	5	0
	Tidak beresiko	1	4

Dari hasil prediksi, diketahui:

Jumlah data yang diprediksi secara benar = 9,

Jumlah data yang diprediksi secara salah = 1,

Jumlah prediksi yang dilakukan = 10,

Maka perhitungan akurasi dan laju error-nya adalah sebagai berikut.

$$\text{Akurasi} = \frac{9}{10} = 0.9 = 90\%$$

$$\text{Laju error} = \frac{1}{10} = 0.1 = 10\%$$

Dari hasil prediksi diketahui:

TP = 5

FN = 0

TN = 4

FP = 1

Maka perhitungan sensitivitas dan spesifisitas adalah sebagai berikut.

$$\text{Sensitivitas} = \frac{5}{5 + 0} = \frac{5}{5} = 1 = 100\%$$

$$\text{Spesifisitas} = \frac{4}{4 + 1} = \frac{4}{5} = 0.8 = 80\%$$

### 3.10 Skenario Pengujian Sistem

Sebelum membuat sistem prediksi tingkat resiko penyakit diabetes menggunakan teknik data mining dengan metode decision tree ID3 ini, perlu dilakukan beberapa skenario pengujian sistem terlebih dahulu, agar sistem dapat berjalan sesuai dengan tujuan pembuatannya.

- Membuat pertanyaan kuisisioner yang sesuai dengan empat atribut yang digunakan yaitu : nama, alamat, usia, berat badan, tinggi badan, intensitas olahraga, keturunan diabetes dan index masa tubuh (IMT) dengan output beresiko dan tidak beresiko yang nantinya akan di masukan ke database sistem.
- Di sediakan 2 macam data yaitu data latih sebanyak 40 dan data uji sebanyak 10. Data latih di gunakan untuk membentuk pohon keputusan dan data uji digunakan untuk menguji akurasi. Selanjutnya membentuk pohon keputusan dari data latih yang sudah di sediakan. Kemudian data uji di klasifikasi berdasarkan pohon keputusan yang dibentuk. setelah itu dihitung akurasi. Akurasi itu menunjukkan baik atau tidaknya pohon keputusan yang sudah terbentuk.
- Akurasi rendah menunjukkan pohon keputusan masih kurang baik. Maka yang dilakukan merubah komposisi data latih. Jika akurasi tinggi maka pohon keputusan akan bisa digunakan.