

# **IMPLEMENTASI ALGORITMA C-45 UNTUK MEMPREDIKSI PENYAKIT JANTUNG MELALUI GEJALA**

## **PENGANTAR**

Memprediksi penyakit jantung dapat dilakukan dengan berbagai macam cara secara komputer berdasarkan gejalanya, salah satunya menggunakan algoritma C-45 data mining untuk memprediksi apakah seseorang memiliki penyakit jantung atau tidak. Untuk memprediksi apakah seseorang memiliki penyakit jantung atau tidak yang diperlukan untuk memprediksinya adalah berat badan (body mass index), apakah merokok, apakah sering meminum alkohol, apakah pernah mengalami stroke, jenis kelamin, umur, apakah memiliki diabetes, bagaimana kondisi kesehatan secara keseluruhan, waktu tidur, apakah memiliki asma.

Salah satu algoritma yang dapat digunakan untuk memprediksi penyakit jantung adalah algoritma C-45. Atribut yang terdapat didalam data penyakit jantung ini menjadi studi kasus untuk laporan akhir mata kuliah ini. Data diambil dari <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease> yang berisi data tentang indikator untuk penyakit jantung (Personal Key Indicators of Heart Disease). Data yang terekam disimpan dalam bentuk format CSV sebanyak kurang lebih 319765 data dan juga terdapat 17 kolom. Namun pada studi kasus ini saya menggunakan 11 kolom.

Hasil akhir yang ingin dicapai dari studi kasus ini adalah dapat memprediksi apakah seseorang memiliki penyakit jantung atau tidak dan juga melihat hasil akhir implementasi digunakan untuk memprediksi seberapa akurat untuk memprediksi penyakit jantung. pengujian data yang dilakukan dengan cara ujicoba sebanyak 6 kali ujicoba dengan jumlah data yang berbeda beda (100, 200, 500, 800, 1000, 1500 data) untuk melihat keakurasiannya.

## **RUMUSAN MASALAH**

1. Apakah algoritma C-45 dapat melakukan prediksi apakah seseorang memiliki penyakit jantung atau tidak?
2. Berapa tingkat akurasi dari implementasi algoritma C45 untuk prediksi penyakit jantung?

## **BATASAN MASALAH**

1. Implementasi algoritma C-45 memiliki yang digunakan sebanyak 200 data
2. Ada 9 atribut(kolom) yang digunakan dalam implementasi algoritma ini
3. uji coba dilakukan sebanyak 6 kali percobaan untuk melihat akurasi hasil akhir
4. implementasi menggunakan SQL database

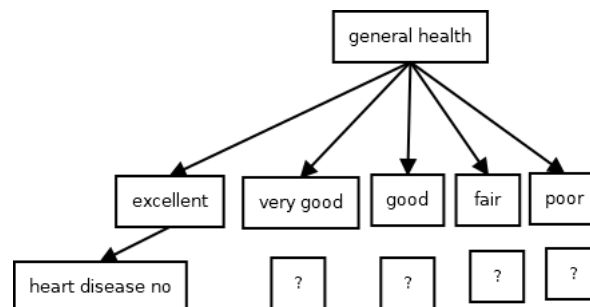
## DESAIN

Desain dari implementasi algoritma ini menggunakan 50 data sample untuk dibahas secara detail dalam algoritma C45.

Pertama tama me-iterasi 1

		jumlah data	no	yes	inf	gain	kolom bantu
totaldata		50	25	25	1		
BMI	<25	15	7	8	0,996792	0,001375	0,29903745
	25-50	35	18	17	0,999411		0,699587745
smoking	yes	29	13	17	0,970581	0,023267	0,562936904
	no	21	12	9	0,985228		0,413795817
alcohol	yes	1	0	1	0	0,020294	0
	no	49	25	24	0,9997		0,979705552
stroke	yes	1	0	1	0	0,020294	0
	no	49	25	24	0,9997		0,979705552
sex	male	24	9	15	0,954434	0,042029	0,458128321
	female	26	16	10	0,961237		0,499843034
age	dibawah 50	11	10	1	0,439497	0,153629	0,096689337
	50 sampai 79	34	13	21	0,959687		0,652587088
	diatas 80	5	2	3	0,970951		0,097095055
diabetes	yes	12	2	10	0,650022	0,108475	0,156005381
	no	38	23	15	0,967788		0,735519232
general health	excellent	7	7	0	0	0,284832	0
	very good	14	10	4	0,863121		0,241673755
	good	17	6	11	0,936667		0,31846691
	fair	5	1	4	0,721928		0,072192805
	poor	7	1	6	0,591673		0,082834185
sleep time	kurang dari 6	4	3	1	0,811278	0,030499	0,06490225
	6 sampai 8	43	21	23	0,987792		0,849501167
	lebih dari 8	3	1	2	0,918296		0,05509775
asma	yes	9	4	5	0,991076	0,001958	0,178393691
	no	41	21	20	0,999571		0,819648088

disini angka gain yang paling tinggi yaitu general health yang berada di angka 0,284 sehingga menghasilkan grafik



selanjutnya melakukan iterasi ke 2 untuk general health very good

		jumlah data	no	yes	inf	gain	kolom bantu
totaldata		17	6	11	0,93667		
BMI	<25	6	2	4	0,9183	0,00067	0,32410441
	25-50	11	4	7	0,94566		0,61189784
smoking	yes	13	5	8	0,96124	0,01072	0,73506329
	no	4	1	3	0,81128		0,19088897
alcohol	yes	1	0	1	0	0,03838	0
	no	16	6	10	0,95443		0,89829083
stroke	no	17	6	11	0,93667	0	0,93666738
sex	male	8	2	6	0,81128	0,0302	0,38177794
	female	9	4	5	0,99108		0,52468733
age	dibawah 50	1	0	1	0	0,09365	0
	50 sampai 79	13	4	9	0,89049		0,6809642
	diatas 80	3	2	1	0,9183		0,16205221
diabetes	yes	3	0	3	0	0,1253	0
	no	14	6	8	0,98523		0,81136435
sleeptime	6 sampai 8	12	5	10	0,74546	0,29281	0,52620683
	lebih dari 8	2	1	1	1		0,11764706
asma	yes	1	0	1	0	0,09334	0
	no	16	5	11	0,89604		0,8433301

disini angka gain yang paling tinggi adalah 0,371 yang berada pada umur Lalu selanjutnya untuk iterasi ke 2 general health good

		jumlah data	no	yes	inf	gain	kolom bantu
totaldata		17	6	11	0,93667		
BMI	<25	6	2	4	0,9183	0,00067	0,32410441
	25-50	11	4	7	0,94566		0,61189784
smoking	yes	13	5	8	0,96124	0,01072	0,73506329
	no	4	1	3	0,81128		0,19088897
alcohol	yes	1	0	1	0	0,03838	0
	no	16	6	10	0,95443		0,89829083
stroke	no	17	6	11	0,93667	0	0,93666738
sex	male	8	2	6	0,81128	0,0302	0,38177794
	female	9	4	5	0,99108		0,52468733
age	dibawah 50	1	0	1	0	0,09365	0
	50 sampai 79	13	4	9	0,89049		0,6809642
	diatas 80	3	2	1	0,9183		0,16205221
diabetes	yes	3	0	3	0	0,1253	0
	no	14	6	8	0,98523		0,81136435
sleeptime	6 sampai 8	15	5	10	0,9183	0,00876	0,81026103
	lebih dari 8	2	1	1	1		0,11764706
asma	yes	1	1	0	0	0,09334	0
	no	16	5	11	0,89604		0,8433301

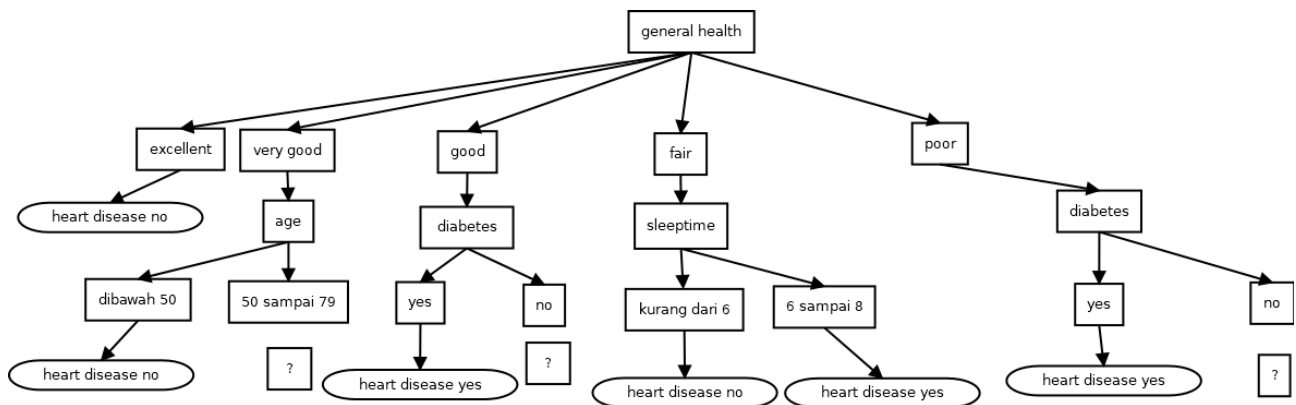
disini angka gain yang paling tinggi adalah 0,125 yang berada pada diabetes. Lalu selanjutnya untuk iterasi ke 2 general health fair

		jumlah da	no	yes	inf	gain	kolom bantu
totaldata		5	1	4	0,721928		
BMI	<25	1	0	1	0	0,072906	0
	25-50	4	1	3	0,811278		0,6490225
smoking	yes	4	1	3	0,811278	0,072906	0,6490225
	no	1	0	1	0		0
alcohol	no	5	1	4	0,721928	0	0,721928095
stroke	no	5	1	4	0,721928	0	0,721928095
sex	male	2	0	2	0	0,170951	0
	female	3	1	2	0,918296		0,5509775
age	50 sampai 79	5	1	4	0,721928	0	0,721928095
diabetes	yes	2	1	1	1	0,321928	0,4
	no	3	0	3	0		0
sleeptime	kurang dari 6	1	1	0	0		0
	6 sampai 8	4	0	4	0	0,721928	0
asma	yes	1	0	1	0	0,072906	0
	no	4	1	3	0,811278		0,6490225

disini angka gain yang paling tinggi adalah 0,721 yang berada pada waktu tidur. Lalu selanjutnya untuk iterasi ke 2 general health poor

		jumlah data	no	yes	inf	gain	kolom bantu
totaldata		7	1	6	0,591673		
BMI	25-50	7	1	6	0,591673	0	0,591672779
smoking	yes	4	1	3	0,811278	0,128085	0,4635875
	no	3	0	3	0		0
alcohol	no	7	1	6	0,591673	0	0,591672779
stroke	yes	1	0	1	0	0,034511	0
	no	6	1	5	0,650022		0,557162076
sex	male	3	0	3	0	0,128085	0
	female	4	1	3	0,811278		0,4635875
age	50 sampai 79	5	1	4	0,721928	0,07601	0,515662925
	diatas 80	2	0	2	0		0
diabetes	yes	5	0	5	0	0,305958	0
	no	2	1	1	1		0,285714286
sleeptime	kurang dari 6	1	0	1	0	0,034511	0
	6 sampai 8	6	1	5	0,650022		0,557162076
asma	yes	4	1	3	0,811278	0,128085	0,4635875
	no	3	0	3	0		0

disini angka gain yang paling tinggi adalah 0,306 yang berada pada diabetes. Setelah iterasi ke 2 selesai semua menghasilkan grafik sebagai berikut



pada selanjutnya iterasi ke 3. yang pertama dimulai dari general health very good age 50 sampai 79

		data	no	yes	inf	gain	kolom bantu
totaldata		7	3	4	0,985228		
BMI	<25	3	2	1	0,918296	0,128085	0,393555357
	25-50	4	1	3	0,811278		0,4635875
smoking	yes	4	2	2	1	0,020244	0,571428571
	no	3	1	2	0,918296		0,393555357
alcohol	no	7	3	4	0,985228	0	0,985228136
stroke	no	7	3	4	0,985228	0	0,985228136
sex	male	6	2	4	0,918296	0,198117	0,787110715
	female	1	1	0	0		0
diabetes	yes	1	0	1	0	0,128085	0
	no	6	3	3	1		0,857142857
sleeptime	6 sampai 8	6	3	3	1	0,128085	0,857142857
	lebih dari 8	1	0	1	0		0
asma	yes	1	0	1	0	0,128085	0
	no	6	3	3	1		0,857142857

disini angka gain yang paling tinggi adalah 0,198 yang berada pada gender. Lalu selanjutnya iterasi ke 3 untuk general health good diabetes no

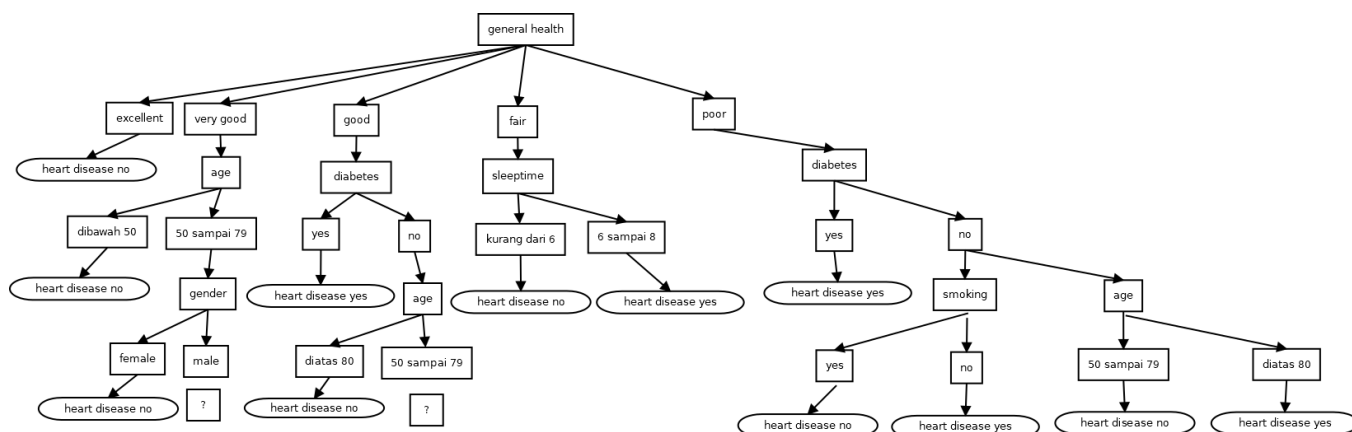
		jumlah da	no	yes	inf	gain	kolom bantu
totaldata		14	6	8	0,985228		
BMI	<25	6	2	4	0,918296	0,020244	0,393555357
	25-50	8	4	4	1		0,571428571
smoking	yes	12	5	7	0,979869	0,002483	0,839887506
	no	2	1	1	1		0,142857143
alcohol	no	14	6	8	0,985228	0	0,985228136
stroke	no	14	6	8	0,985228	0	0,985228136
sex	male	6	2	4	0,918296	0,020244	0,393555357
	female	8	4	4	1		0,571428571
age	50 sampai	12	4	8	0,918296	0,198117	0,787110715
	diatas 80	2	2	0	0		0
sleeptime	6 sampai 8	12	5	7	0,979869	0,002483	0,839887506
	lebih dari	2	1	1	1		0,142857143
asma	yes	1	1	0	0	0,092651	0
	no	13	5	8	0,961237		0,892576847

disini angka gain yang paling tinggi adalah 0,198 yang berada pada age. Lalu selanjutnya iterasi ke 3 untuk general health poor diabetes no

		jumlah data	no	yes	inf	gain	kolom bantu
totaldata		2	1	1	1		
BMI	25-50	2	1	1	1	0	1
smoking	yes	1	1	0	0	1	0
	no	0	0	1	0		0
alcohol	no	2	1	1	1	0	1
stroke	no	2	1	1	1	0	1
sex	female	2	1	1	1	0	1
age	50 sampai 79	1	1	0	0	1	0
	diatas 80	1	0	1	0		0
sleeptime	6 sampai 8	2	1	1	1	0	1
asma	yes	2	1	1	1	0	1

disini angka gain yang paling tinggi yaitu smoking dan sleeptime sebesar 1. iterasi 3 telah selesai sehingga menjadikan grafik iterasi ke 3 menjadi





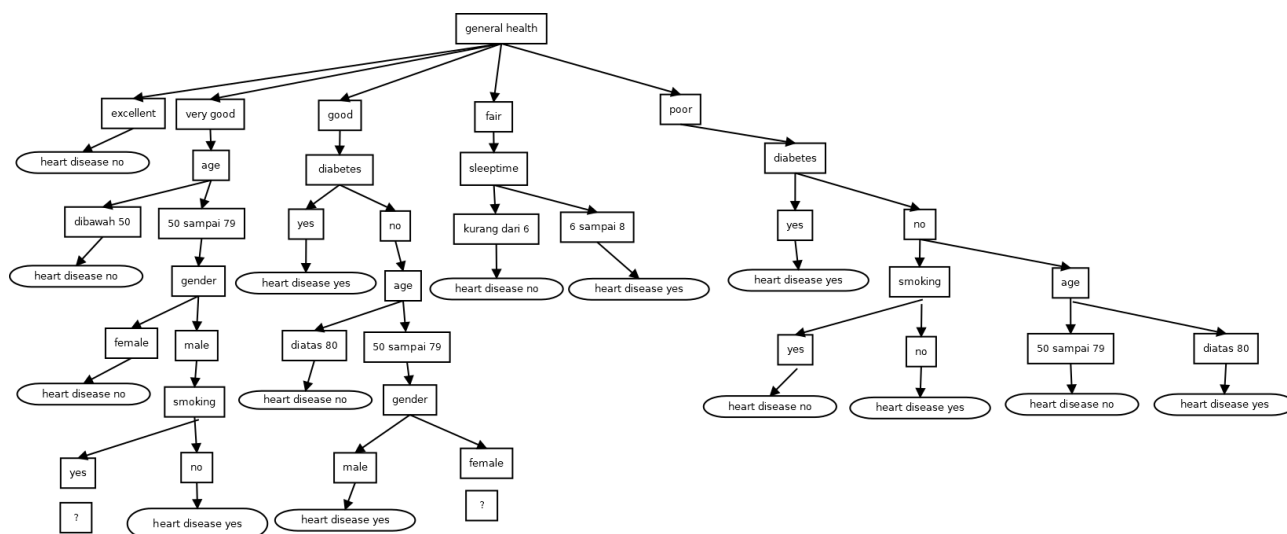
lalu selanjutnya untuk iterasi ke 4 yang pertama pada general health very good, age 50 sampai 79, gender male

		data	no	yes	inf	gain	kolom bantu
totaldata		6	2	4	0,918296		
BMI	<25	4	1	3	0,811278		0,540852083
	25-50	2	1	1	1	0,04411	0,333333333
smoking	yes	4	2	2	2		0,666666667
	no	2	0	2	0	0,251629	0
alcohol	no	6	2	4	0,918296	0	0,918295834
stroke	no	6	2	4	0,918296	0	0,918295834
diabetes	yes	1	0	1	0		0
	no	5	2	3	0,970951	0,10917	0,809125495
sleeptime	6 sampai 8	5	2	3	0,970951		0,809125495
	lebih dari 8	1	0	1	0	0,10917	0
asma	yes	1	0	1	0		0
	no	5	2	3	0,970951	0,10917	0,809125495

disini angka gain yang paling tinggi adalah 0,252 yang berada pada smoking. Lalu selanjutnya iterasi ke 4 untuk general health good, diabetes no, age 50 sampai 79

		jumlah da	no	yes	inf	gain	kolom bantu
totaldata		12	4	8	0,918296		
BMI	<25	6	2	4	0,918296		0,459147917
	25-50	6	2	4	0,918296	0	0,459147917
smoking	yes	10	3	7	0,881291		0,734409083
	no	2	1	1	1	0,01722	0,166666667
alcohol	no	12	4	8	0,918296	0	0,918295834
stroke	no	12	4	8	0,918296	0	0,918295834
sex	male	4	0	4	0		0
	female	8	4	4	1	0,251629	0,666666667
sleeptime	6 sampai 8	10	3	7	0,881291		0,734409083
	lebih dari 8	2	1	1	1	0,01722	0,166666667
asma	no	12	4	8	0,918296	0	0,918295834

disini angka gain yang paling tinggi adalah 0,252 yang berada pada gender/jenis kelamin. Iterasi 4 telah selesai sehingga menjadikan grafik iterasi ke 4 menjadi



lalu selanjutnya untuk iterasi ke 5 yang pertama berada di general health very good, age 50 sampai 79, gender male, smoking yes

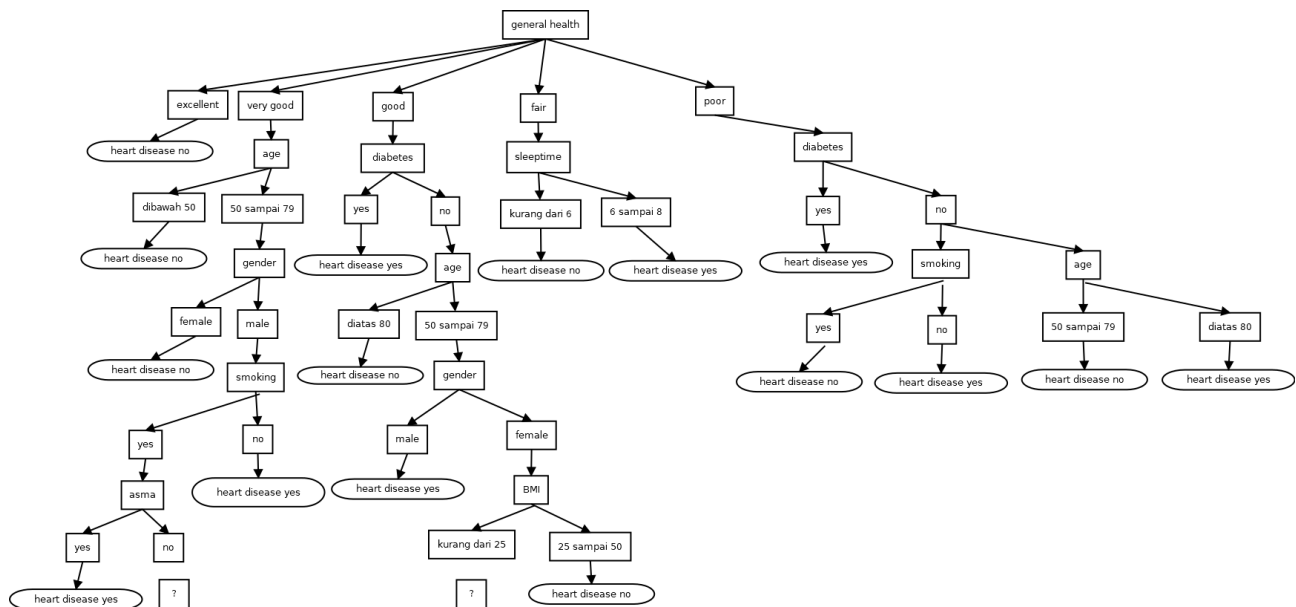
		data	no	yes	inf	gain	kolom bantu
totaldata			4	2	2	1	
BMI	<25		2	1	1	1	0,5
	25-50		2	1	1	0	0,5
alcohol	no		4	2	1	1	0
stroke	no		4	2	1	1	0
diabetes	no		4	2	1	1	0
sleeptime	6 sampai 8		4	2	1	1	0
asma	yes		1	0	1	0	0
	no		3	2	1	0,918296	0,311278
							0,688721876

disini angka gain yang paling tinggi adalah 0,311 yang berada pada asma. Lalu selanjutnya iterasi ke 5 untuk general health good, diabetes no, age 50 sampai 79, gender female

		jumlah da	no	yes	inf	gain	kolom bantu
totaldata			8	4	4	1	
BMI	<25		6	2	4	0,918296	0,688721876
	25-50		2	2	0	0	0,311278
smoking	yes		7	3	4	0,985228	0,862074619
	no		1	1	0	0	0,137925
alcohol	no		8	4	4	1	0
stroke	no		8	4	4	1	0
sleeptime	6 sampai 8		7	3	4	0,985228	0,862074619
	lebih dari 8		1	1	0	0	0,137925
asma	no		8	4	4	1	0



disini angka gain yang paling tinggi adalah 0,311 yang berada pada BMI. Iterasi 5 telah selesai sehingga menjadikan grafik iterasi ke 5 menjadi



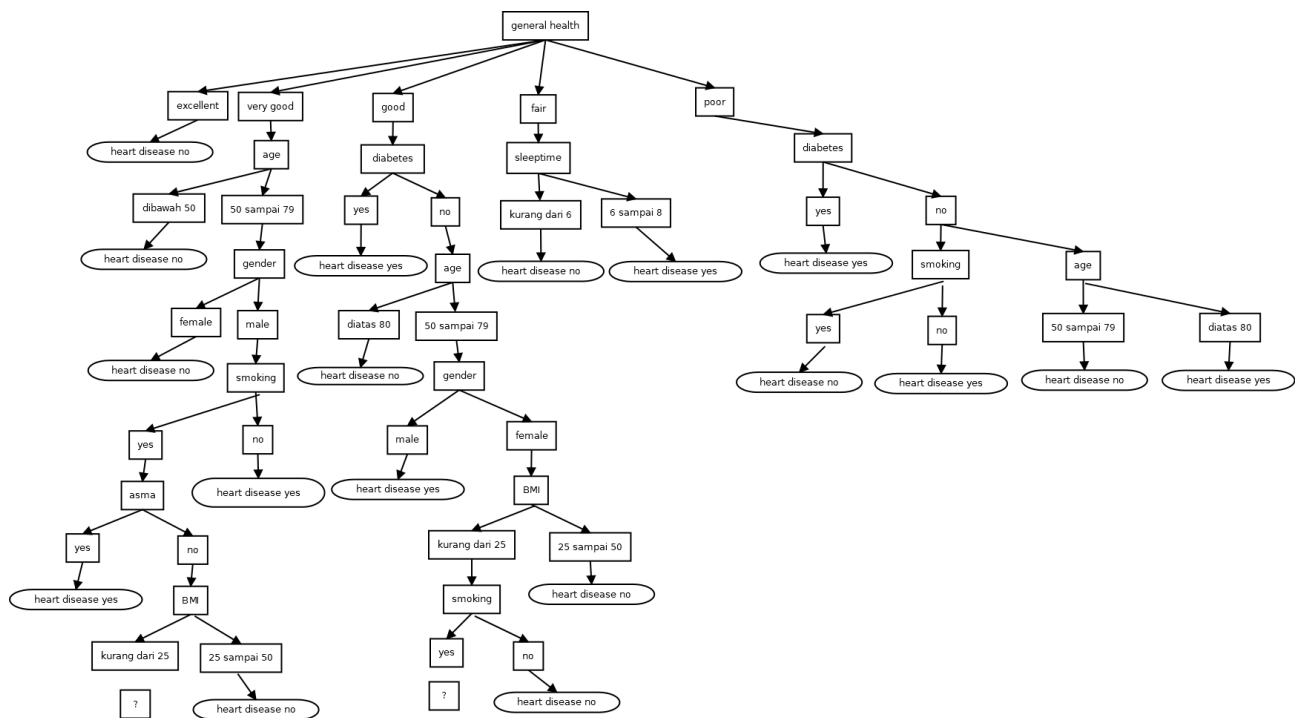
lalu selanjutnya untuk iterasi ke 6 yang pertama berada di general health very good, age 50 sampai 79, gender male, smoking yes, asma no

		data	no	yes	inf	gain	kolom bantu
totaldata		3	2	1	0,918296		
BMI	<25	2	1	1	1		0,666666667
	25-50	1	1	0	0	0,251629	0
alcohol	no	3	2	1	0,918296	0	0,918295834
stroke	no	3	2	1	0,918296	0	0,918295834
diabetes	no	3	2	1	0,918296	0	0,918295834
sleeptime	6 sampai 8	3	2	1	0,918296	0	0,918295834

disini angka gain yang paling tinggi adalah 0,252 yang berada pada BMI. Lalu selanjutnya iterasi ke 6 untuk general health good, diabetes no, age 50 sampai 79, gender female, BMI kurang dari 25

		jumlah da	no	yes	inf	gain	kolom bantu
totaldata		6	2	4	0,918296		
smoking	yes	5	1	4	0,721928		0,601606746
	no	1	1	0	0	0,316689	0
alcohol	no	6	2	4	0,918296	0	0,918295834
stroke	no	6	2	4	0,918296	0	0,918295834
sleeptime	6 sampai 8	6	2	4	0,918296	0	0,918295834
asma	no	6	2	4	0,918296	0	0,918295834

disini angka gain yang paling tinggi adalah 0,317 yang berada pada smoking. Iterasi 6 telah selesai sehingga menjadikan grafik iterasi ke 6 menjadi



iterasi berhenti di iterasi ke 6 karena data yang digun akan untuk ujicoba hingga iterasi ke 6 hingga tidak dapat dipecah kembali berdasarkan kolom. Jika diiterasi ke 7 dari general health very good, age 50 sampai 79, gender male, smoking yes, asma no , BMI kurang dari 25 dilakukan maka tabel iterasi yang terbentuk menjadi

		data	no	yes	inf	gain	kolom bantu
totaldata		2	1	1	1		
alcohol	no	2	1	1	1	0	1
stroke	no	2	1	1	1	0	1
diabetes	no	2	1	1	1	0	1
sleeptime	6 sampai 8	2	1	1	1	0	1

dan untuk iterasi ke 7 dari general health good, diabetes no, age 50 sampai 79, gender female, BMI kurang dari 25, smoking yes dilakukan maka tabel iterasi yang terbentuk menjadi

		jumlah da	no	yes	inf	gain	kolom bantu
totaldata		5	1	4	0,721928		
alcohol	no	5	1	4	0,721928	0	0,721928095
stroke	no	5	1	4	0,721928	0	0,721928095
sleeptime	6 sampai 8	5	1	4	0,721928	0	0,721928095
asma	no	5	1	4	0,721928	0	0,721928095

## IMPLEMENTASI

Pada implementasi saya menggunakan SQL database dengan jumlah data sebesar 100. Pertama tama membuat database terlebih dahulu

```
drop database if exists dbProject100;  
create database dbProject100;  
use dbProject100;
```

lalu selanjutnya membuat tabel untuk memasukan data

```
create table data100(  
    heartdisease varchar(5),  
    BMI double,  
    smoking varchar(5),  
    alcohol varchar(5),  
    stroke varchar(5),  
    physicalheatlh int,  
    mentalhealth int,  
    diffwalking varchar(5),  
    gender varchar(6),  
    age varchar(20),  
    race varchar(15),  
    diabetes varchar(30),  
    physicalactivity varchar(5),  
    generalhealth varchar(10),  
    sleeptime int,  
    asthma varchar(5),  
    penyakitginjal varchar(5),  
    penyakitkanker varchar(5)  
);
```

lalu selanjutnya memasukan data kedalam table

```
load data local infile 'D:/datamining/100data.csv'  
into table data100  
fields terminated by ';' '  
enclosed by '' '  
ignore 1 lines;
```

setelah memasukan 100 data ke table data 100 maka langkah selanjutnya adalah membuang kolom yang tidak akan digunakan yaitu physical health, mental health, physical activity, kanker, race(ras), diffwalking, dan penyakit ginjal

```
alter table data100 drop column physicalheatlh;  
alter table data100 drop column mentalhealth;  
alter table data100 drop column physicalactivity;  
alter table data100 drop column penyakitkanker;  
alter table data100 drop column race;  
alter table data100 drop column diffwalking;  
alter table data100 drop column penyakitginjal;
```

lalu selanjutnya memasukan data yang belum berupa range untuk menjadi range disini yang diubah adalah bmi(berat badan) menjadi 3 yaitu kurang dari 25, 25 sampai 50, dan juga lebih dari 50. selain bmi yang diubah adalah umur menjadi 3 yaitu dibawah 50, 50 hingga 79 dan diatas 80. Lalu untuk diabetes diubah menjadi ya dan tidak. lalu range untuk waktu tidur menjadi 3 yaitu kurang dari 6, 6 hingga 8, dan diatas 8.

```

/*memasukan berat ke range*/
alter table data100 add column rangebmi varchar(20);
update data100 set rangebmi="25_sampai_50" where 25<BMI<50;
update data100 set rangebmi="kurang_dari_25" where BMI<25;
update data100 set rangebmi="lebih_dari_50" where BMI>50;
select * from data100 limit 5;

/*membuat range umur*/
alter table data100 add column rangeumur varchar(20);
update data100 set rangeumur="dibawah_50" where age="30-34" or
age="35-39" or age="40-44" or age="45-49" or age="25-29" or age="18-24";
update data100 set rangeumur="50_sampai_79" where age="50-54" or
age="55-59" or age="60-64" or age="65-69" or age="70-74" or age="75-79";
update data100 set rangeumur="diatas_80" where age="80 or older";
select * from data100 limit 5;
/*meenjadikan diabetes menjadi yes dan no*/
alter table data100 add column simplediabetes varchar(20);
update data100 set simplediabetes="yes" where diabetes="yes" or
diabetes="Yes (during pregnancy)" or diabetes="No, borderline diabetes";
update data100 set simplediabetes="no" where diabetes="no";

```

```

/*membuat range umur*/
alter table data100 add column rangeumur varchar(20);
update data100 set rangeumur="dibawah_50" where age="30-34" or
age="35-39" or age="40-44" or age="45-49" or age="25-29" or age="18-24";
update data100 set rangeumur="50_sampai_79" where age="50-54" or
age="55-59" or age="60-64" or age="65-69" or age="70-74" or age="75-79";
update data100 set rangeumur="diatas_80" where age="80 or older";
select * from data100 limit 5;

/*meenjadikan diabetes menjadi yes dan no*/
alter table data100 add column simplediabetes varchar(20);
update data100 set simplediabetes="yes" where diabetes="yes" or
diabetes="Yes (during pregnancy)" or diabetes="No, borderline diabetes";
update data100 set simplediabetes="no" where diabetes="no";

/*membuat range waktu tidur*/
alter table data100 add column waktutidur varchar(20);
update data100 set waktutidur="6_sampai_8" where 6<=sleeptime<=8;
update data100 set waktutidur="kurang_dari_6" where sleeptime<6;
update data100 set waktutidur="lebih_dari_8" where sleeptime>8;

```

lalu selanjutnya setelah diubah data yang telah diubah sudah tidak digunakan sehingga dibuang

```

alter table data100 drop column BMI;
alter table data100 drop column age;
alter table data100 drop column diabetes;
alter table data100 drop column sleeptime;

```

lalu selanjutnya membuat tabel untuk menampung hasil iterasi

```
create table tbliterasi(
    atribut varchar(20),
    informasi varchar(20),
    jumlahdata int,
    tidak int,
    ya int,
    nilaiI double,
    gain double
);
```

lalu selanjutnya membuat procedure untuk menghitung hasil iterasi yang berinputkan nama table yang akan dihitung

pertama tama setelah membuat procedure adalah mendeklarasikan variable yang akan dipakai untuk menjalankan procedure yang digunakan adalah command yang digunakan untuk menyimpan command yang akan digunakan untuk menyimpan code sql yang akan diexecute. Variable lain yang digunakan yaitu pengganti yang digunakan untuk menyimpan command yang akan digunakan untuk mengganti nama atribut yang akan digunakan saat memasukan data. Lalu variable i untuk menyimpan nomor untuk looping. Lalu atributs yang digunakan untuk menyimpan atribut apa yang akan diinput

```
delimiter &&
create procedure iter(dbname varchar(200))
begin
    DECLARE command NVARCHAR(3000);
    declare pengganti nvarchar(3000);
    declare i int;
    declare atributs varchar(100);
    set i:=0;
    truncate tbliterasi;
```

setelah dideklarasikan maka kita perlu untuk mengosongkan terlebih dahulu tabel iterasi jika terdapat data didalamnya dengan cara truncate

```
truncate tbliterasi;
```

selanjutnya setelah kosong tabel iterasi maka kita set command untuk menghitung jumlah data sesuai dengan tabel yang diinput dari procedure. Setelah memasukan command lalu selanjutnya menyiapkan command tersebut dengan cara prepare statement from command. Setelah prepare maka selanjutnya execute untuk menjalankan command yang telah di prepare dengan perintah EXECUTE statement. Setelah dijalankan maka perlu untuk melepas prepare sehingga dapat digunakan oleh command yang lain dengan perintah DEALLOCATE prepare statement

```
set command:=concat('select @jumlah_data:=count(*)from ',dbname);
PREPARE stmt2 FROM command;
EXECUTE stmt2;
DEALLOCATE PREPARE stmt2;
```

setelah menghitung jumlah data maka lakukan hal yang sama untuk menghitung jumlah yes heart disease dengan set command, prepare statement, execute, deallocate prepare

```
set command:=concat('select @yes:=count(*)from ',dbname,' where heartdisease="yes";');
PREPARE stmt2 FROM command;
EXECUTE stmt2;
DEALLOCATE PREPARE stmt2;
```

setelah menghitung jumlah yes maka lakukan hal yang sama untuk menghitung jumlah no heart disease dengan set command, prepare statement, execute, deallocate prepare

```
set command:=concat('select @no:=count(*)from ',dbname,' where heartdisease="no";');
PREPARE stmt2 FROM command;
EXECUTE stmt2;
DEALLOCATE PREPARE stmt2;
```

setelah menghitung jumlah no maka menghitung nilai informasi untuk total data dan masukan data ke table iterasi

```
select @nilaiI:=(-(@no/@jumlah_data)*log2(@no/@jumlah_data))
+(-(@yes/@jumlah_data)*log2(@yes/@jumlah_data))as inf;
insert into tbliterasi(atribut,jumlahdata, tidak, ya,nilaiI) values
('total data',@jumlah_data,@no,@yes,@nilaiI);
```

setelah memasukan nilai data ke table maka selanjutnya memulai looping untuk menghitung nilai per kolom table karena nilai table terdapat 9 maka while i <> 10 kolom dicari dimasukan kedalam variable atribut. jika i =0 maka nilai yang diinput ke atributs adalah range bmi, jika i =1 maka nilai yang diinput ke atributs adalah smoking, jika i=2 maka nilai yang diinput ke atributs adalah alcohol, jika i=3 maka nilai yang diinput ke atributs adalah stroke, jika i=4 maka nilai yang diinput ke atributs adalah gender. Jika i=5 maka nilai yang diinput ke atributs adalah range umur, jika i=6 maka nilai yang diinput ke atributs adalah diabetes. Jika i=7 maka nilai yang diinput ke atributs adalah general health. Jika i=8 maka nilai yang diinput ke atributs adalah waktu tidur. Selain itu maka nilai yang diinput ke atributs adalah asma.

```
atribute: while i<>10 do
  if(i=0) then
    set atributs='rangebmi';
  else
    if(i=1)then set atributs='smoking';
    else
      if(i=2) then set atributs='alcohol';
      else
        if(i=3) then set atributs='stroke';
        else
          if(i=4) then set atributs='gender';
          else
            if(i=5)then set atributs='rangeumur';
            else
              if(i=6) then set atributs='simplifiediabetes';
              else
                if(i=7) then set atributs='generalhealth';
                else
                  if(i=8) then set atributs='waktutidur';
                  else
                    set atributs='asthma';
                  end if;
                end if;
              end if;
            end if;
          end if;
        end if;
      end if;
    end if;
  end if;
end if;
```



Setelah menentukan atributs maka selanjutnya set command untuk memasukan data ke table iterasi sesuai atribut yang telah ditentukan dan juga table yang diinput pada procedure yang dimasukan adalah informasi atribut,jumlah data,dan jumlah yes dan no per atribut. Setelah setting command maka selanjutnya setting pengganti untuk mengganti atribut dengan atribut yang dipakai. Lalu selanjutnya prepare pengganti dan command . Setelah di prepare maka execute statement. Setelah diexecute maka deallocate kedua prepare statement untuk pengganti dan command

```

end if;
set command:= concat('insert into tbliterasi( informasi,jumlahdata, tidak,ya) select a.',atributs,'
, count(*)as jumlahdata, (select count(*) from ',dbname,' as b where b.heartdisease="no" and
b.',atributs,'=a.',atributs,')as jawabno,(select count(*)from ',dbname,' as c where c.heartdisease
="yes" and c.',atributs,'=a.',atributs,')as jawabyes from ',dbname,' as a group by a.',atributs,');');
set pengganti:=concat('update tbliterasi set atribut = "',atributs,'" where atribut is null;');
PREPARE stmt FROM pengganti;
PREPARE stmt2 FROM command;
EXECUTE stmt2;
EXECUTE stmt;
DEALLOCATE PREPARE stmt;
DEALLOCATE PREPARE stmt2;
set i:=i+1;
end while atribute;

```

setelah looping selesai setting nilai i yang kosong pada semua tabel

```

update tbliterasi set nilaiI=
(-(tidak/jumlahdata)*log2(tidak/jumlahdata))+
(-(ya/jumlahdata)*log2(ya/jumlahdata));

```

selanjutnya membuat tabel sementara untuk menampung gain dan menghitung nilai gain. Setelah dihitung nilai gain maka memasukan nilai gain ke table iterasi dan membulatkan nilai Informasi dan juga nilai gain lalu selanjutnya menampilkan hasil output table iterasi dan stored procedure selesai

```

update tbliterasi set nilaiI=0 where nilaiI is null;
insert into tbltampung (atribut,gain)
select atribut,@nilaiI-sum((jumlahdata/@jumlah_data)*nilaiI) as nilaigain from tbliterasi group by atribut;
update tbliterasi set gain=(select gain from tbltampung where tbltampung.atribut=tbliterasi.atribut);
update tbliterasi set nilaiI=round(nilaiI,3);
update tbliterasi set gain=round(gain,3);
select dbname;
select * from tbliterasi;
end; &&
delimiter ;

```

setelah procedure selesai maka memanggil procedure untuk mengeluarkan hasil iterasi pertama

```

call iter('data100');

```

hasil yang dikeluarkan

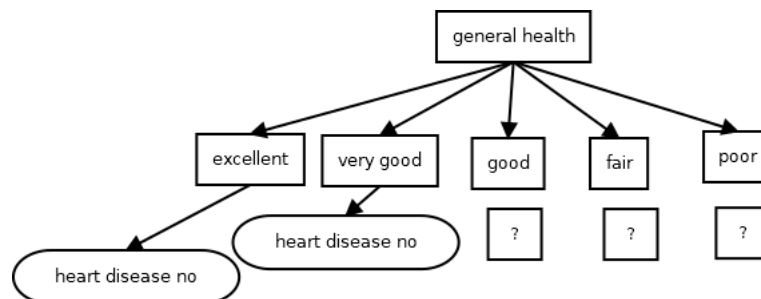
```

+-----+
| dbname |
+-----+
| data100 |
+-----+
1 row in set (0.044 sec)

```

atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	100	92	8	0.402	0
rangebmi	25_sampai_50	66	59	7	0.488	0.015
rangebmi	kurang_dari_25	34	33	1	0.191	0.015
smoking	No	51	49	2	0.239	0.018
smoking	Yes	49	43	6	0.536	0.018
alcohol	No	99	91	8	0.405	0.001
alcohol	Yes	1	1	0	0	0.001
stroke	No	94	87	7	0.382	0.004
stroke	Yes	6	5	1	0.65	0.004
gender	Female	70	64	6	0.422	0.001
gender	Male	30	28	2	0.353	0.001
rangeumur	50_sampai_79	74	67	7	0.452	0.009
rangeumur	diatas_80	22	21	1	0.267	0.009
rangeumur	dibawah_50	4	4	0	0	0.009
simplediabetes	no	72	69	3	0.25	0.033
simplediabetes	yes	28	23	5	0.677	0.033
generalhealth	Excellent	9	9	0	0	0.125
generalhealth	Fair	18	16	2	0.503	0.125
generalhealth	Good	31	29	2	0.345	0.125
generalhealth	Poor	8	4	4	1	0.125
generalhealth	Very good	34	34	0	0	0.125
waktutidur	6_sampai_8	73	70	3	0.247	0.054
waktutidur	kurang_dari_6	14	13	1	0.371	0.054
waktutidur	lebih_dari_8	13	9	4	0.89	0.054
asthma	No	81	74	7	0.424	0.002
asthma	Yes	19	18	1	0.297	0.002

nilai gain paling tinggi diperoleh general health sebesar 0.125 sehingga membuat tree sebagai berikut



lalu karena general health excellent dan verygood hanya memiliki no maka iterasi selanjutnya yaitu iterasi 2 dimulai dari general health good. Pertama tama kita membuat table mirip dengan table data100 lalu input ke table sesuai dengan data100 namun hanya general health good

```

create table data100_2_generalheatlh_good like data100;
insert into data100_2_generalheatlh_good select* from data100 where generalhealth="good";
call iter('data100_2_generalheatlh_good');

```

```

+-----+
| dbname |
+-----+
| data100_2_generalheatlh_good |
+-----+
1 row in set (0.042 sec)

```

atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	31	29	2	0.345	0
rangebmi	25_sampai_50	24	22	2	0.414	0.025
rangebmi	kurang_dari_25	7	7	0	0	0.025
smoking	No	21	19	2	0.454	0.038
smoking	Yes	10	10	0	0	0.038
alcohol	No	31	29	2	0.345	0
stroke	No	31	29	2	0.345	0
gender	Female	22	20	2	0.439	0.033
gender	Male	9	9	0	0	0.033
rangeumur	50_sampai_79	23	22	1	0.258	0.013
rangeumur	dias_80	8	7	1	0.544	0.013
simplifiediabetes	no	20	19	1	0.286	0.004
simplifiediabetes	yes	11	10	1	0.439	0.004
generalhealth	Good	31	29	2	0.345	0
waktutidur	6_sampai_8	23	21	2	0.426	0.029
waktutidur	kurang_dari_6	4	4	0	0	0.029
waktutidur	lebih_dari_8	4	4	0	0	0.029
asthma	No	25	23	2	0.402	0.021
asthma	Yes	6	6	0	0	0.021

disini nilai gain paling tinggi diperoleh oleh smoking yaitu 0.038 selanjutnya kal yang serupa untuk general health fair

```

create table data100_2_generalheatlh_fair like data100;
insert into data100_2_generalheatlh_fair select* from data100 where generalhealth="fair";
call iter('data100_2_generalheatlh_fair');

```

```

+-----+
| dbname |
+-----+
| data100_2_generalheatlh_fair |
+-----+
1 row in set (0.043 sec)

```

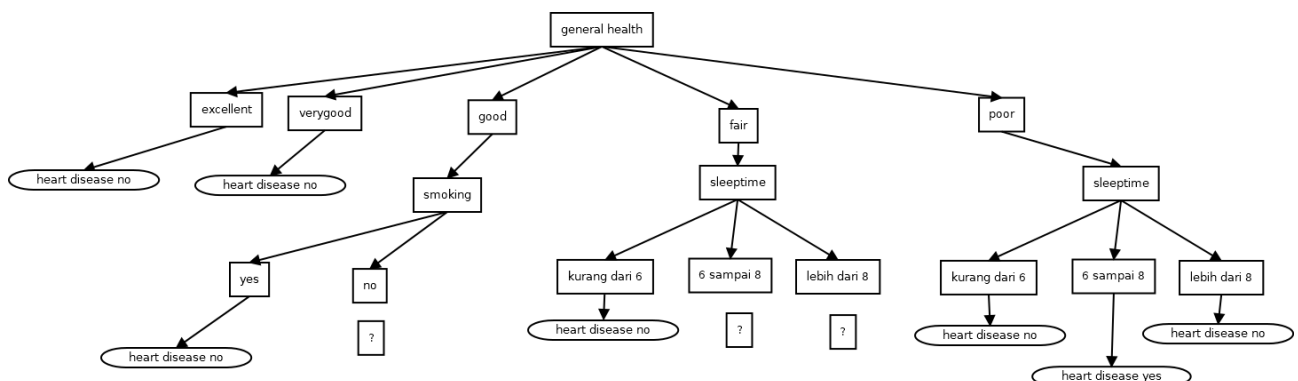
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	18	16	2	0.503	0
rangebmi	25_sampai_50	14	12	2	0.592	0.043
rangebmi	kurang_dari_25	4	4	0	0	0.043
smoking	No	7	7	0	0	0.085
smoking	Yes	11	9	2	0.684	0.085
alcohol	No	18	16	2	0.503	0
stroke	No	17	15	2	0.523	0.01
stroke	Yes	1	1	0	0	0.01
gender	Female	12	10	2	0.65	0.07
gender	Male	6	6	0	0	0.07
rangeumur	50_sampai_79	12	10	2	0.65	0.07
rangeumur	dias_80	6	6	0	0	0.07
simplifiediabetes	no	11	10	1	0.439	0.005
simplifiediabetes	yes	7	6	1	0.592	0.005
generalhealth	Fair	18	16	2	0.503	0
waktutidur	6_sampai_8	11	10	1	0.439	0.124
waktutidur	kurang_dari_6	5	5	0	0	0.124
waktutidur	lebih_dari_8	2	1	1	1	0.124
asthma	No	12	10	2	0.65	0.07
asthma	Yes	6	6	0	0	0.07

nilai gain paling tinggi diperoleh oleh waktu tidur sebesar 0.124. selanjutnya general health poor

```
create table data100_2_generalheatlh_poor like data100;
insert into data100_2_generalheatlh_poor select* from data100 where generalhealth="poor";
call iter('data100_2_generalheatlh_poor');
```

atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	8	4	4	1	0
rangebmi	25_sampai_50	6	3	3	1	0
rangebmi	kurang_dari_25	2	1	1	1	0
smoking	Yes	8	4	4	1	0
alcohol	No	8	4	4	1	0
stroke	No	6	3	3	1	0
stroke	Yes	2	1	1	1	0
gender	Female	4	2	2	1	0
gender	Male	4	2	2	1	0
rangeumur	50_sampai_79	8	4	4	1	0
simplifiediabetes	no	4	3	1	0.811	0.189
simplifiediabetes	yes	4	1	3	0.811	0.189
generalhealth	Poor	8	4	4	1	0
waktutidur	6_sampai_8	4	4	0	0	1
waktutidur	kurang_dari_6	1	0	1	0	1
waktutidur	lebih_dari_8	3	0	3	0	1
asthma	No	5	2	3	0.971	0.049
asthma	Yes	3	2	1	0.918	0.049

nilai gain paling tinggi diperoleh oleh waktu tidur sebesar 1. iterasi 2 telah selesai maka tree yang terbentuk adalah



selanjutnya iterasi ke 3 dimulai pertama general health good smoking no. Data diambil dari iterasi sebelumnya yang memiliki smoking no

```
create table data100_3_generalheatlh_good_smoking_no like data100;
insert into data100_3_generalheatlh_good_smoking_no
select* from data100_2_generalheatlh_good where smoking="no";
call iter('data100_3_generalheatlh_good_smoking_no');
```

dbname							
data100_3_generalheatlh_good_smoking_no							
1 row in set (0.036 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	21	19	2	0.454	0	
rangebmi	25_sampai_50	15	13	2	0.567	0.049	
rangebmi	kurang_dari_25	6	6	0	0	0.049	
smoking	No	21	19	2	0.454	0	
alcohol	No	21	19	2	0.454	0	
stroke	No	21	19	2	0.454	0	
gender	Female	15	13	2	0.567	0.049	
gender	Male	6	6	0	0	0.049	
rangeumur	50_sampai_79	16	15	1	0.337	0.025	
rangeumur	diatas_80	5	4	1	0.722	0.025	
simplifiediabetes	no	14	13	1	0.371	0.009	
simplifiediabetes	yes	7	6	1	0.592	0.009	
generalhealth	Good	21	19	2	0.454	0	
waktutidur	6_sampai_8	17	15	2	0.523	0.031	
waktutidur	kurang_dari_6	1	1	0	0	0.031	
waktutidur	lebih_dari_8	3	3	0	0	0.031	
asthma	No	19	17	2	0.485	0.014	
asthma	Yes	2	2	0	0	0.014	

nilai gain paling tinggi diperoleh oleh bmi dan gender sebesar 0.049. selanjutnya iterasi 3 untuk general health fair waktu tidur 6 sampai 8

```
create table data100_3_generalheatlh_fair_sleep_6s8 like data100;
insert into data100_3_generalheatlh_fair_sleep_6s8
select* from data100_2_generalheatlh_fair where waktutidur="6_sampai_8";
call iter('data100_3_generalheatlh_fair_sleep_6s8');
```

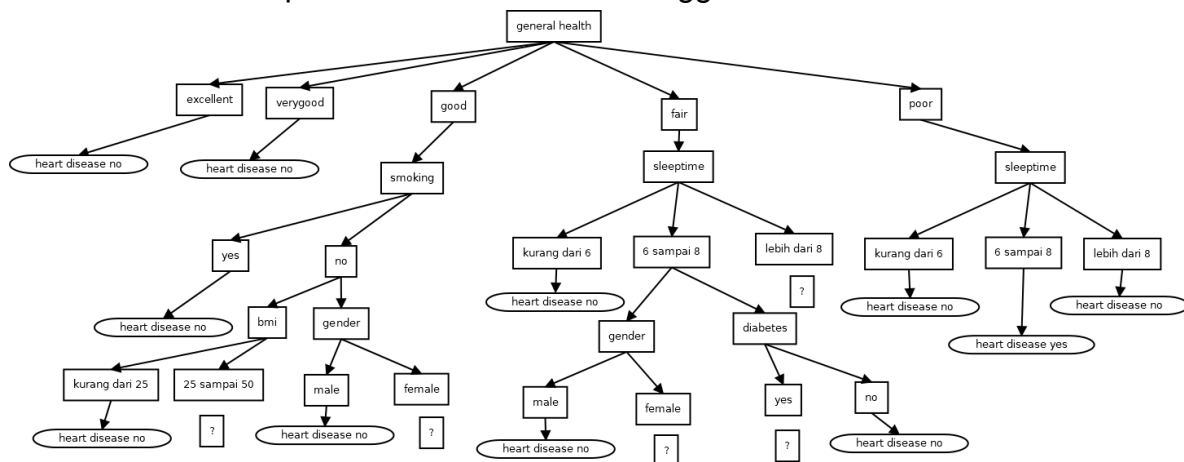
dbname							
data100_3_generalheatlh_fair_sleep_6s8							
1 row in set (0.038 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	11	10	1	0.439	0	
rangebmi	25_sampai_50	10	9	1	0.469	0.013	
rangebmi	kurang_dari_25	1	1	0	0	0.013	
smoking	No	4	4	0	0	0.063	
smoking	Yes	7	6	1	0.592	0.063	
alcohol	No	11	10	1	0.439	0	
stroke	No	10	9	1	0.469	0.013	
stroke	Yes	1	1	0	0	0.013	
gender	Female	5	4	1	0.722	0.111	
gender	Male	6	6	0	0	0.111	
rangeumur	50_sampai_79	6	5	1	0.65	0.085	
rangeumur	diatas_80	5	5	0	0	0.085	
simplifiediabetes	no	6	6	0	0	0.111	
simplifiediabetes	yes	5	4	1	0.722	0.111	
generalhealth	Fair	11	10	1	0.439	0	
waktutidur	6_sampai_8	11	10	1	0.439	0	
asthma	No	7	6	1	0.592	0.063	
asthma	Yes	4	4	0	0	0.063	

nilai gain paling tinggi diperoleh oleh gender dan diabetes sebesar 0,111 selanjutnya iterasi 3 untuk general health fair sleep lebih dari 8

```
create table data100_3_generalheatlh_fair_sleep_l8 like data100;
insert into data100_3_generalheatlh_fair_sleep_l8
select* from data100_2_generalheatlh_fair where waktutidur="lebih_dari_8";
call iter('data100_3_generalheatlh_fair_sleep_l8');
```

dbname						
data100_3_generalheatlh_fair_sleep_l8						
1 row in set (0.034 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	2	1	1	1	0
rangebmi	25_sampai_50	2	1	1	1	0
smoking	Yes	2	1	1	1	0
alcohol	No	2	1	1	1	0
stroke	No	2	1	1	1	0
gender	Female	2	1	1	1	0
rangeumur	50_sampai_79	2	1	1	1	0
simplifiediabetes	no	2	1	1	1	0
generalhealth	Fair	2	1	1	1	0
waktutidur	lebih_dari_8	2	1	1	1	0
asthma	No	2	1	1	1	0

karena nilai gain sama semua dan tidak ada lagi yang berbeda maka iterasi 3 general health fair sleep lebih dari 8 berhenti sehingga membuat tree



selanjutnya iterasi 4 untuk general health good smoking no bmi 25sampai 50

```
create table data100_4_generalheatlh_good_smoking_no_bmi_25s50 like data100;
insert into data100_4_generalheatlh_good_smoking_no_bmi_25s50
select* from data100_3_generalheatlh_good_smoking_no where rangebmi="25_sampai_50";
call iter('data100_4_generalheatlh_good_smoking_no_bmi_25s50');
```

dbname						
data100_4_generalheatlh_good_smoking_no_bmi_25s50						
1 row in set (0.054 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	15	13	2	0.567	0
rangebmi	25_sampai_50	15	13	2	0.567	0
smoking	No	15	13	2	0.567	0
alcohol	No	15	13	2	0.567	0
stroke	No	15	13	2	0.567	0
gender	Female	9	7	2	0.764	0.108
gender	Male	6	6	0	0	0.108
rangeumur	50_sampai_79	13	12	1	0.391	0.094
rangeumur	diatas_80	2	1	1	1	0.094
simplifiediabetes	no	9	8	1	0.503	0.005
simplifiediabetes	yes	6	5	1	0.65	0.005
generalhealth	Good	15	13	2	0.567	0
waktutidur	6_sampai_8	11	9	2	0.684	0.065
waktutidur	kurang_dari_6	1	1	0	0	0.065
waktutidur	lebih_dari_8	3	3	0	0	0.065
asthma	No	13	11	2	0.619	0.03
asthma	Yes	2	2	0	0	0.03

nilai gain paling tinggi diperoleh oleh gender sebesar 0.108. selanjutnya iterasi 4 general health good smoking no gender female



```
create table data100_4_generalheatlh_good_smoking_no_gender_female like data100;
insert into data100_4_generalheatlh_good_smoking_no_gender_female
select* from data100_3_generalheatlh_good_smoking_no where gender="female";
call iter('data100_4_generalheatlh_good_smoking_no_gender_female');
```

dbname						
data100_4_generalheatlh_good_smoking_no_gender_female						
1 row in set (0.038 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	15	13	2	0.567	0
rangebmi	25_sampai_50	9	7	2	0.764	0.108
rangebmi	kurang_dari_25	6	6	0	0	0.108
smoking	No	15	13	2	0.567	0
alcohol	No	15	13	2	0.567	0
stroke	No	15	13	2	0.567	0
gender	Female	15	13	2	0.567	0
rangeumur	50_sampai_79	10	9	1	0.469	0.013
rangeumur	diatas_80	5	4	1	0.722	0.013
simplifiediabetes	no	12	11	1	0.414	0.052
simplifiediabetes	yes	3	2	1	0.918	0.052
generalhealth	Good	15	13	2	0.567	0
waktutidur	6_sampai_8	13	11	2	0.619	0.03
waktutidur	kurang_dari_6	1	1	0	0	0.03
waktutidur	lebih_dari_8	1	1	0	0	0.03
asthma	No	14	12	2	0.592	0.014
asthma	Yes	1	1	0	0	0.014

nilai gain paling tinggi diperoleh oleh bmi yaitu sebesar 0.108 selanjutnya iterasi 4 untuk general health fair waktu tidur 6sampai 8 gender female

```
create table data100_4_generalheatlh_fair_sleep_6s8_gender_female like data100;
insert into data100_4_generalheatlh_fair_sleep_6s8_gender_female
select* from data100_3_generalheatlh_fair_sleep_6s8 where gender="female";
call iter('data100_4_generalheatlh_fair_sleep_6s8_gender_female');
```

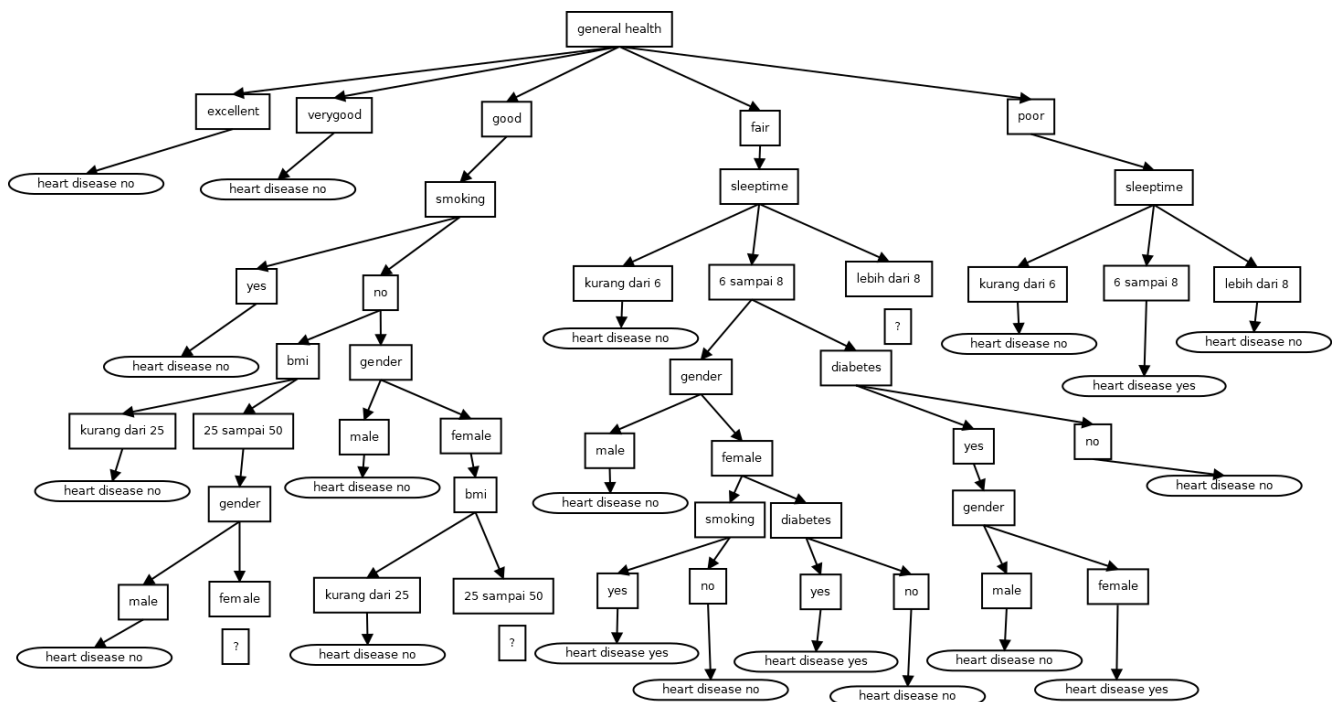
dbname						
data100_4_generalheatlh_fair_sleep_6s8_gender_female						
1 row in set (0.036 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	5	4	1	0.722	0
rangebmi	25_sampai_50	4	3	1	0.811	0.073
rangebmi	kurang_dari_25	1	1	0	0	0.073
smoking	No	4	4	0	0	0.722
smoking	Yes	1	0	1	0	0.722
alcohol	No	5	4	1	0.722	0
stroke	No	5	4	1	0.722	0
gender	Female	5	4	1	0.722	0
rangeumur	50_sampai_79	3	2	1	0.918	0.171
rangeumur	diatas_80	2	2	0	0	0.171
simplifiediabetes	no	4	4	0	0	0.722
simplifiediabetes	yes	1	0	1	0	0.722
generalhealth	Fair	5	4	1	0.722	0
waktutidur	6_sampai_8	5	4	1	0.722	0
asthma	No	4	3	1	0.811	0.073
asthma	Yes	1	1	0	0	0.073

nilai gain paling tinggi diperoleh oleh bmi dan juga diabetes sebesar 0.722 lalu selanjutnya iterasi 4 untuk general health fair sleep sampai 8 diabetes yes

```
create table data100_4_generalheatlh_fair_sleep_6s8_diabetes_yes like data100;
insert into data100_4_generalheatlh_fair_sleep_6s8_diabetes_yes
select* from data100_3_generalheatlh_fair_sleep_6s8 where simplifiediabetes="yes";
call iter('data100_4_generalheatlh_fair_sleep_6s8_diabetes_yes');
```

dbname							
data100_4_generalheatlh_fair_sleep_6s8_diabetes_yes							
1 row in set (0.035 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	5	4	1	0.722	0	
rangebmi	25_sampai_50	5	4	1	0.722	0	
smoking	Yes	5	4	1	0.722	0	
alcohol	No	5	4	1	0.722	0	
stroke	No	5	4	1	0.722	0	
gender	Female	1	0	1	0	0.722	
gender	Male	4	4	0	0	0.722	
rangeumur	50_sampai_79	3	2	1	0.918	0.171	
rangeumur	diasas_80	2	2	0	0	0.171	
simplifiediabetes	yes	5	4	1	0.722	0	
generalhealth	Fair	5	4	1	0.722	0	
waktutidur	6_sampai_8	5	4	1	0.722	0	
asthma	No	4	3	1	0.811	0.073	
asthma	Yes	1	1	0	0	0.073	

nilai gain paling tinggi diperoleh oleh gender sebesar 0.722. iterasi 4 telah selesai semua sehingga menghasilkan grafik tree



selanjutnya iterasi 5 untuk general health good smoking no bmi 25 sampai 50 gender female

```

create table data100_5_generalheatlh_good_smoking_no_bmi_25s50_gender_female like data100;
insert into data100_5_generalheatlh_good_smoking_no_bmi_25s50_gender_female
select* from data100_4_generalheatlh_good_smoking_no_bmi_25s50 where gender="female";
call iter('data100_5_generalheatlh_good_smoking_no_bmi_25s50_gender_female');

```

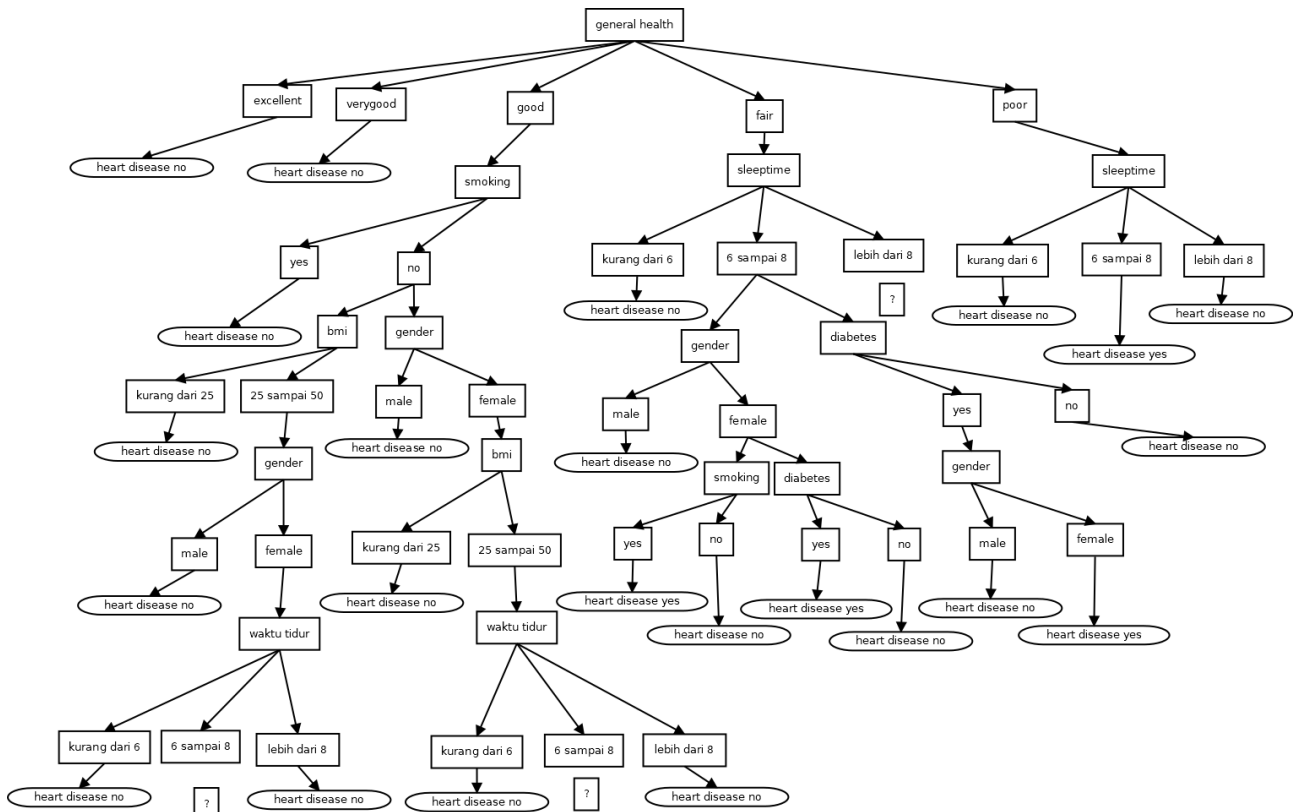
dbname						
data100_5_generalheatlh_good_smoking_no_bmi_25s50_gender_female						
1 row in set (0.037 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	9	7	2	0.764	0
rangebmi	25_sampai_50	9	7	2	0.764	0
smoking	No	9	7	2	0.764	0
alcohol	No	9	7	2	0.764	0
stroke	No	9	7	2	0.764	0
gender	Female	9	7	2	0.764	0
rangeumur	50_sampai_79	7	6	1	0.592	0.082
rangeumur	diatas_80	2	1	1	1	0.082
simplédiabetes	no	7	6	1	0.592	0.082
simplédiabetes	yes	2	1	1	1	0.082
generalhealth	Good	9	7	2	0.764	0
waktutidur	6_sampai_8	7	5	2	0.863	0.093
waktutidur	kurang_dari_6	1	1	0	0	0.093
waktutidur	lebih_dari_8	1	1	0	0	0.093
asthma	No	8	6	2	0.811	0.043
asthma	Yes	1	1	0	0	0.043

nilai gain paling tinggi diperoleh oleh waktu tidur sebesar 0.093. selanjutnya iterasi ke 5 untuk general health good smoking no gender female bmi 25sampai 50

```
create table data100_5_generalheatlh_good_smoking_no_gender_female_bmi_25s50 like data100;
insert into data100_5_generalheatlh_good_smoking_no_gender_female_bmi_25s50
select* from data100_4_generalheatlh_good_smoking_no_gender_female where rangebmi="25_sampai_50";
call iter('data100_5_generalheatlh_good_smoking_no_gender_female_bmi_25s50');
```

dbname						
data100_5_generalheatlh_good_smoking_no_gender_female_bmi_25s50						
1 row in set (0.034 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	9	7	2	0.764	0
rangebmi	25_sampai_50	9	7	2	0.764	0
smoking	No	9	7	2	0.764	0
alcohol	No	9	7	2	0.764	0
stroke	No	9	7	2	0.764	0
gender	Female	9	7	2	0.764	0
rangeumur	50_sampai_79	7	6	1	0.592	0.082
rangeumur	diatas_80	2	1	1	1	0.082
simplédiabetes	no	7	6	1	0.592	0.082
simplédiabetes	yes	2	1	1	1	0.082
generalhealth	Good	9	7	2	0.764	0
waktutidur	6_sampai_8	7	5	2	0.863	0.093
waktutidur	kurang_dari_6	1	1	0	0	0.093
waktutidur	lebih_dari_8	1	1	0	0	0.093
asthma	No	8	6	2	0.811	0.043
asthma	Yes	1	1	0	0	0.043

nilai gain paling tinggi diperoleh oleh waktu tidur sebesar 0.093. iterasi 5 selesai sehingga menghasilkan tree



selanjutnya iterasi ke 6 untuk general health good smoking no bmi 25 sampai 50 gender female waktu tidur 6 sampai 8

```
create table data100_6_gh_good_smoking_no_bmi_25s50_gender_female_sleep_6s8 like data100;
insert into data100_6_gh_good_smoking_no_bmi_25s50_gender_female_sleep_6s8
select* from data100_5_generalheatlh_good_smoking_no_bmi_25s50_gender_female where waktutidur="6_sampai_8";
call iter('data100_6_gh_good_smoking_no_bmi_25s50_gender_female_sleep_6s8');
```

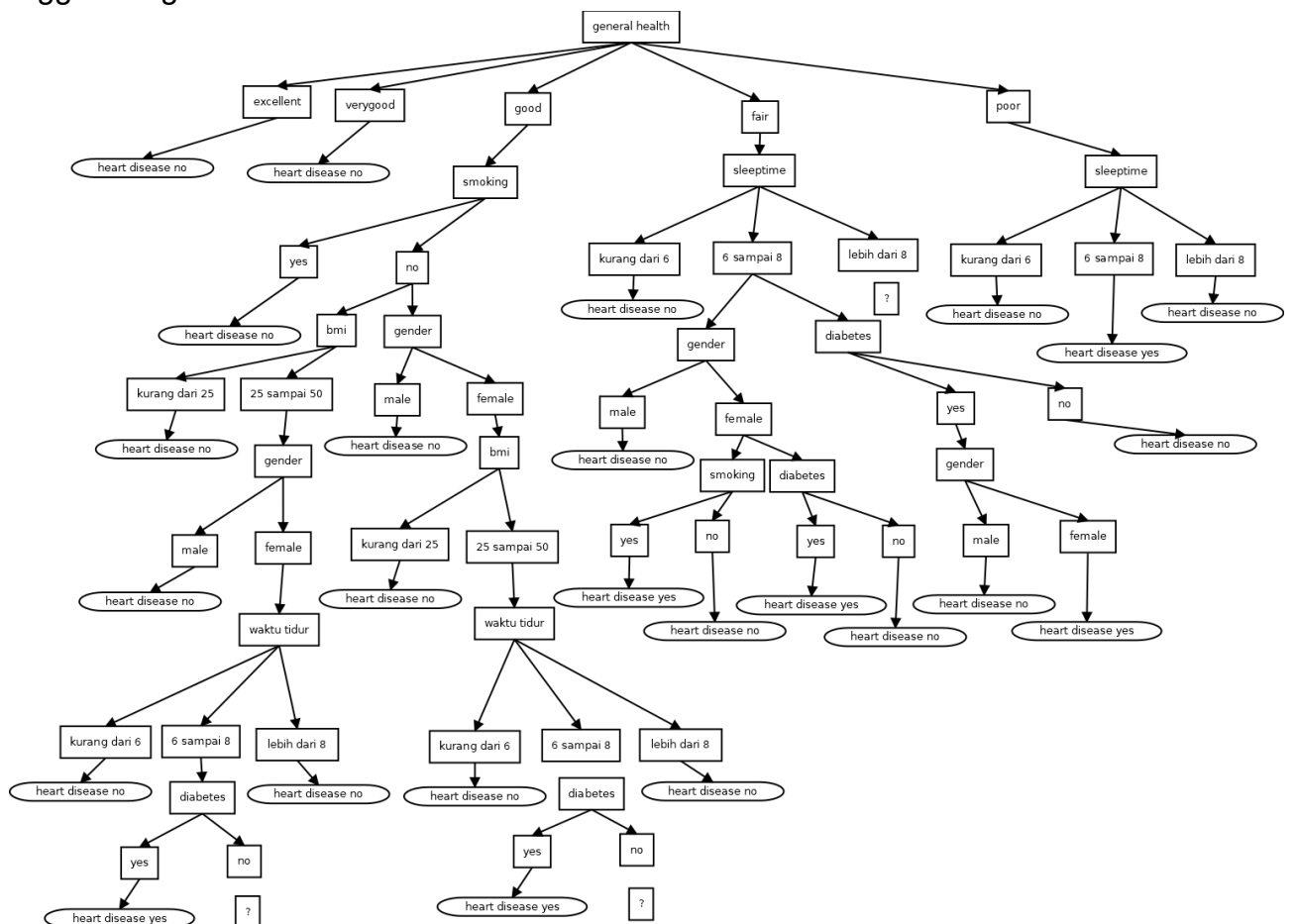
dbname							
data100_6_gh_good_smoking_no_bmi_25s50_gender_female_sleep_6s8							
1 row in set (0.034 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	7	5	2	0.863	0	
rangebmi	25_sampai_50	7	5	2	0.863	0	
smoking	No	7	5	2	0.863	0	
alcohol	No	7	5	2	0.863	0	
stroke	No	7	5	2	0.863	0	
gender	Female	7	5	2	0.863	0	
rangeumur	50_sampai_79	5	4	1	0.722	0.062	
rangeumur	diasas_80	2	1	1	1	0.062	
simplifiediabetes	no	6	5	1	0.65	0.306	
simplifiediabetes	yes	1	0	1	0	0.306	
generalhealth	Good	7	5	2	0.863	0	
waktutidur	6_sampai_8	7	5	2	0.863	0	
asthma	No	7	5	2	0.863	0	

nilai gain paling tinggi diperoleh oleh diabetes sebesar 0.306. selanjutnya iterasi 6 untuk general health good smoking no gender female bmi 25sampai50 sleep 6sampai 8

```
create table data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8 like data100;
insert into data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8
select* from data100_5_generalheatlh_good_smoking_no_gender_female_bmi_25s50 where waktutidur="6_sampai_8";
call iter('data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8');
```

dbname						
data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8						
1 row in set (0.034 sec)						
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain
total data	NULL	7	5	2	0.863	0
rangebmi	25_sampai_50	7	5	2	0.863	0
smoking	No	7	5	2	0.863	0
alcohol	No	7	5	2	0.863	0
stroke	No	7	5	2	0.863	0
gender	Female	7	5	2	0.863	0
rangeumur	50_sampai_79	5	4	1	0.722	0.062
rangeumur	dias_80	2	1	1	1	0.062
simplediabetes	no	6	5	1	0.65	0.306
simplediabetes	yes	1	0	1	0	0.306
generalhealth	Good	7	5	2	0.863	0
waktutidur	6_sampai_8	7	5	2	0.863	0
asthma	No	7	5	2	0.863	0

nilai gain paling tinggi diperoleh oleh diabetes sebesar 0.306. iterasi 6 telah selesai sehingga menghasilkan tree



selanjutnya iterasi 7 untuk general health good smoking no bmi 25sampai50 gender female waktu tidur 6sampai8 diabetes no

```
create table d_7_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no like data100;
insert into d_7_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no
select* from data100_6_gh_good_smoking_no_bmi_25s50_gender_female_sleep_6s8 where simplifiediabetes="no";
call iter('d_7_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no');
```

-----+-----							
dbname							
d_7_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no							
1 row in set (0.035 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	6	5	1	0.65	0	
rangebmi	25_sampai_50	6	5	1	0.65	0	
smoking	No	6	5	1	0.65	0	
alcohol	No	6	5	1	0.65	0	
stroke	No	6	5	1	0.65	0	
gender	Female	6	5	1	0.65	0	
rangeumur	50_sampai_79	5	4	1	0.722	0.048	
rangeumur	diatas_80	1	1	0	0	0.048	
simplifiediabetes	no	6	5	1	0.65	0	
generalhealth	Good	6	5	1	0.65	0	
waktutidur	6_sampai_8	6	5	1	0.65	0	
asthma	No	6	5	1	0.65	0	

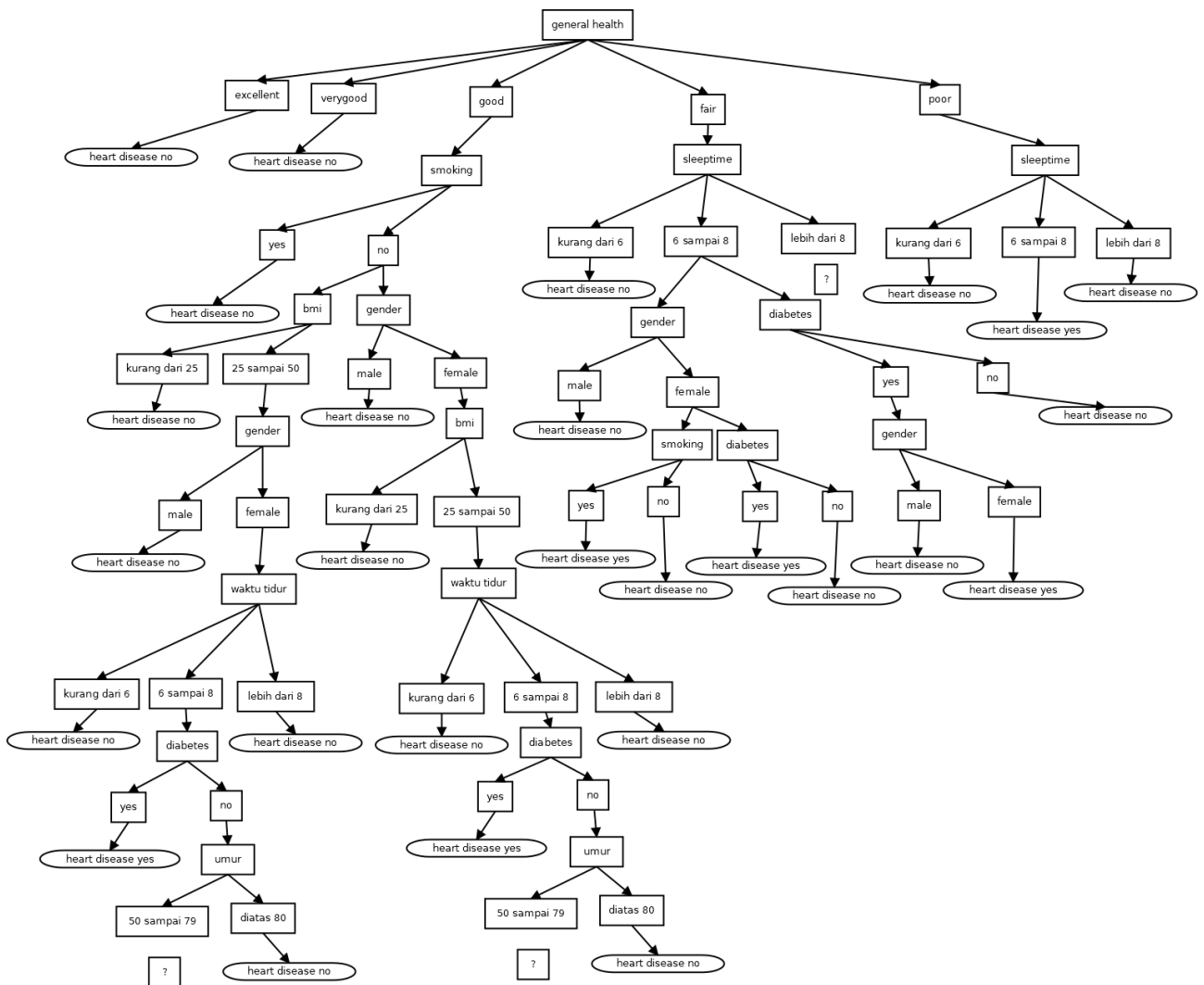
nilai gain paling tinggi diperoleh oleh umur sebesar 0.048. selanjutnya iterasi 7 untuk general health good smoking no gender female bmi 25sampai50 sleeptime 6sampai8 diabetes no

```
create table d_7_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no like data100;
insert into d_7_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no
select* from data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8 where simplifiediabetes="no";
call iter('d_7_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no');
```

-----+-----							
dbname							
d_7_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no							
1 row in set (0.034 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	6	5	1	0.65	0	
rangebmi	25_sampai_50	6	5	1	0.65	0	
smoking	No	6	5	1	0.65	0	
alcohol	No	6	5	1	0.65	0	
stroke	No	6	5	1	0.65	0	
gender	Female	6	5	1	0.65	0	
rangeumur	50_sampai_79	5	4	1	0.722	0.048	
rangeumur	diatas_80	1	1	0	0	0.048	
simplifiediabetes	no	6	5	1	0.65	0	
generalhealth	Good	6	5	1	0.65	0	
waktutidur	6_sampai_8	6	5	1	0.65	0	
asthma	No	6	5	1	0.65	0	

nilai gain paling tinggi diperoleh oleh range umur sebesar 0.048. iterasi ke 7 telah selesai sehingga menghasilkan tree





iterasi 8 general health good smoking no bmi 25sampai50 gender female waktu tidur 6sampai8 diabetes no umur 50-79

```
create table d_8_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no_u_50s79 like data100;
insert into d_8_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no_u_50s79
select* from d_7_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no where rangeumur="50_sampai_79";
call iter('d_8_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no_u_50s79');
```

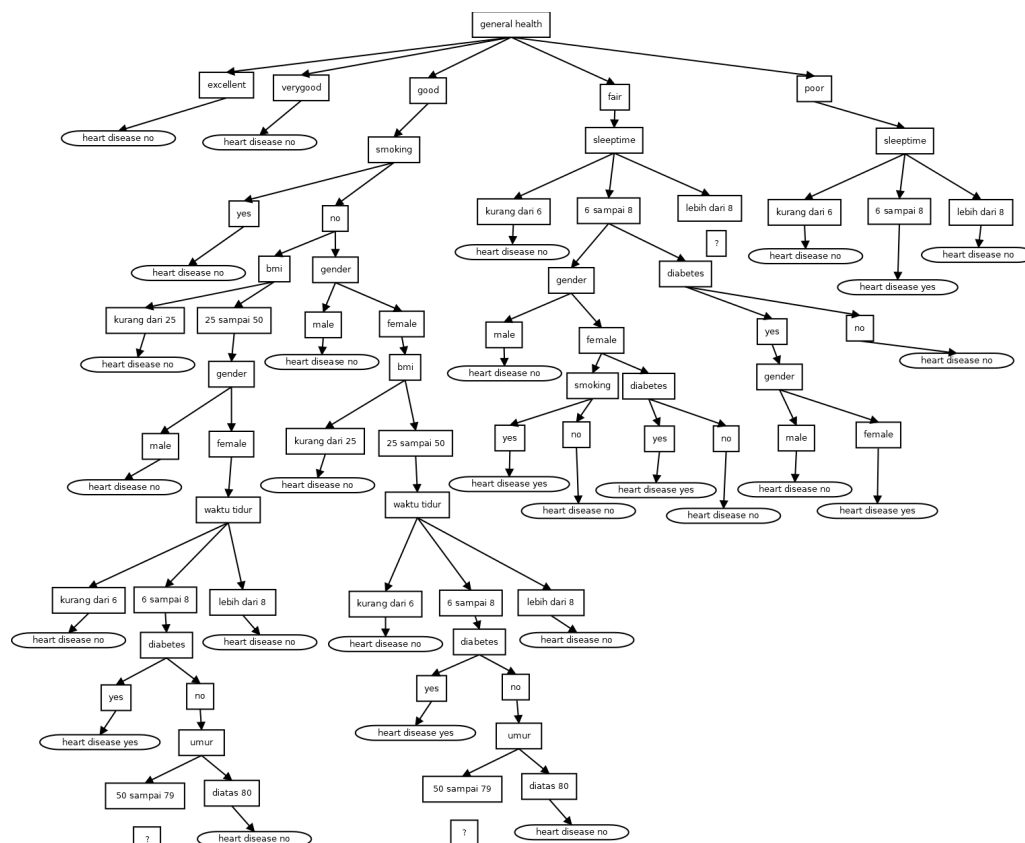
dbname							
d_8_gh_good_smo_no_bmi_25s50_gdr_f_sleep_6s8_diabet_no_u_50s79							
1 row in set (0.033 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	5	4	1	0.722	0	
rangebmi	25_sampai_50	5	4	1	0.722	0	
smoking	No	5	4	1	0.722	0	
alcohol	No	5	4	1	0.722	0	
stroke	No	5	4	1	0.722	0	
gender	Female	5	4	1	0.722	0	
rangeumur	50_sampai_79	5	4	1	0.722	0	
simplifiediabetes	no	5	4	1	0.722	0	
generalhealth	Good	5	4	1	0.722	0	
waktutidur	6_sampai_8	5	4	1	0.722	0	
asthma	No	5	4	1	0.722	0	

tidak terdapat nilai gain paling tinggi karena semua nilai gain adalah 0 maka dari itu iterasi 8 general health good smoking no bmi 25sampai50 gender female waktu tidur 6sampai8 diabetes no umur 50 sampai 79 telah selesai. Selanjutnya itrerasi 8 untuk general health good smoking no gender female bmi 25 sampai 50 waktu tidur 6sampai8 diabetes no umur 50 sampai 79

```
create table d_8_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no_u_50s79 like data100;
insert into d_8_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no_u_50s79
select* from data100_6_gh_good_smoking_no_gender_female_bmi_25s50_sleep_6s8 where rangeumur="50_sampai_79";
call iter('d_8_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no_u_50s79');
```

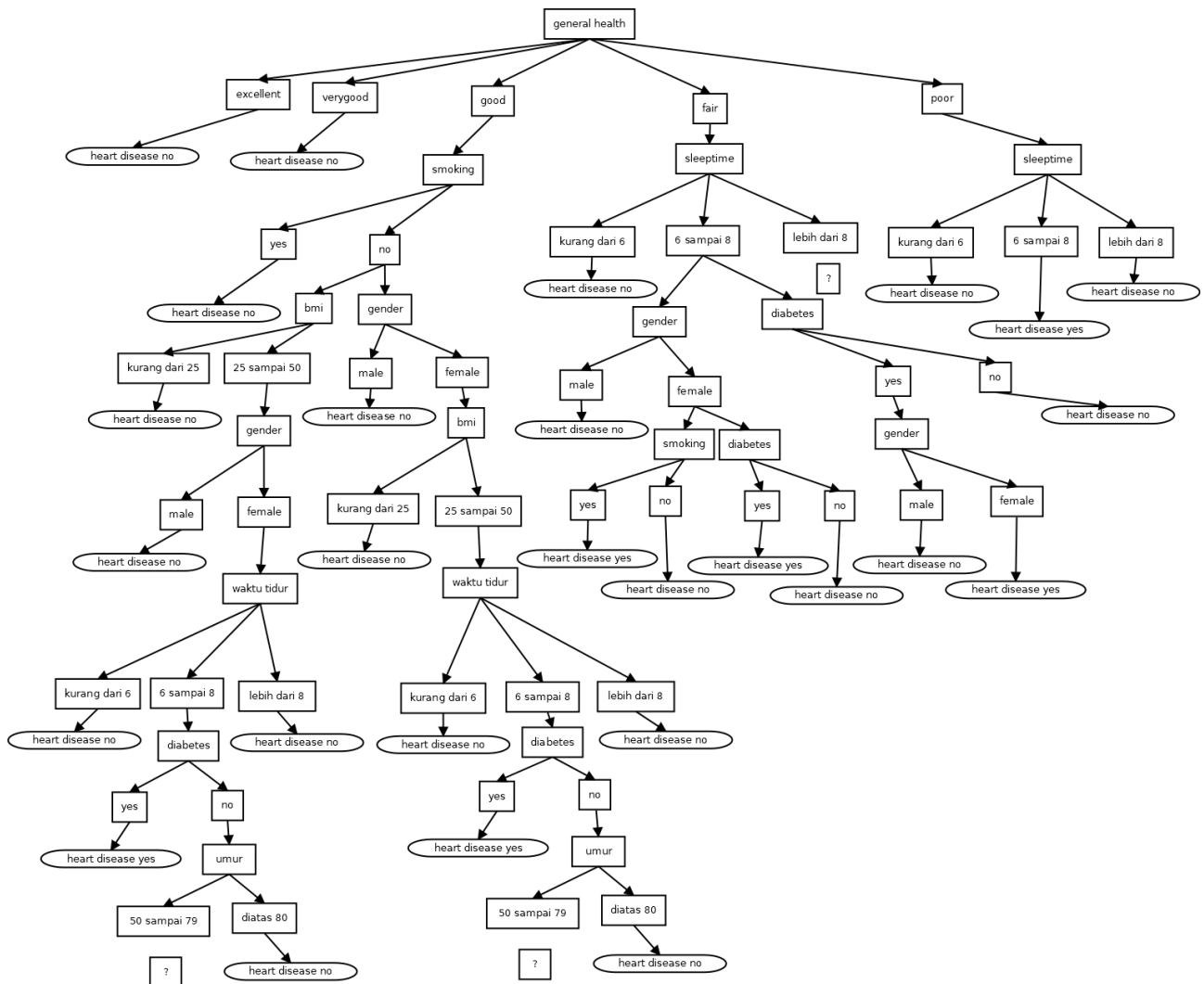
dbname							
d_8_gh_good_smo_no_gdr_f_bmi_25s50_sleep_6s8_diabet_no_u_50s79							
1 row in set (0.031 sec)							
atribut	informasi	jumlahdata	tidak	ya	nilaiI	gain	
total data	NULL	5	4	1	0.722	0	
rangebmi	25_sampai_50	5	4	1	0.722	0	
smoking	No	5	4	1	0.722	0	
alcohol	No	5	4	1	0.722	0	
stroke	No	5	4	1	0.722	0	
gender	Female	5	4	1	0.722	0	
rangeumur	50_sampai_79	5	4	1	0.722	0	
simplediabetes	no	5	4	1	0.722	0	
generalhealth	Good	5	4	1	0.722	0	
waktutidur	6_sampai_8	5	4	1	0.722	0	
asthma	No	5	4	1	0.722	0	

tidak terdapat nilai gain paling tinggi karena semua nilai gain adalah 0 maka dari itu iterasi 8 general health good smoking no bmi 25sampai50 gender female waktu tidur 6sampai8 diabetes no umur 50 sampai 79 telah selesai. Semua iterasi telah selesai sehingga menghasilkan tree akhir sama seperti iterasi 7 sebagai berikut

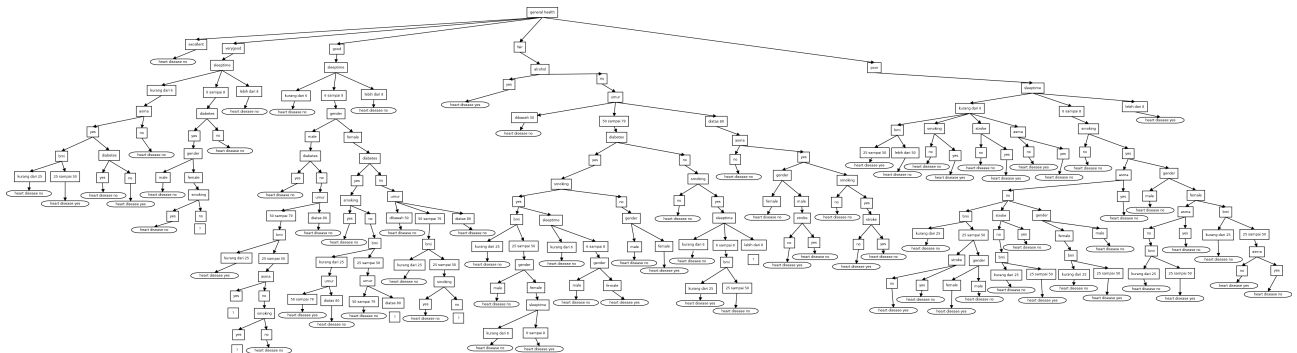


## ANALISIS & PENGUJIAN

Untuk pengujian menggunakan 50 data yang diacak dengan 2 tree yang digunakan untuk ujicoba. Ujicoba pertama menggunakan 100 data yang memiliki grafiki tree sebagai berikut



lalu untuk ujicoba kedua menggunakan 200 data yang memiliki grafik tree sebagai berikut



dari hasil pengujian 50 data secara acak dan dicocokkan kepada hasil grafik tree secara manual menghasilkan keakurasian sebagai berikut

Hasil tree	Total benar dari 50	keakurasian
100 data	42	84%
200 data	42	84%

dari hasil pengujian dapat disimpulkan bahwa model ini tergolong cukup akurat dengan keakurasian 84%. membuat data tree semakin banyak mungkin tidak akan mempengaruhi sedikit hasil keakurasian karena jumlah data antara heart disease yes dan heart disease no yang berbeda jauh menyebabkan keakurasian tidak bertambah banyak namun cabang akan semakin bercabang dan lebih panjang.

## PENUTUP

Dari hasil implementasi, maka dapat disimpulkan bahwa:

1. Algoritma C-45 digunakan untuk melakukan prediksi penyakit jantung. Dari hasil pengujian dengan data yang berbeda dapat menunjukkan bahwa akurasi selalu berada di 84%
2. Akurasi pengujian dari algoritma C-45 mencapai angka paling tinggi berada pada 84% untuk pengujian 200 data. Dari hasil uji coba, menunjukkan bahwa semakin banyak data yang diujicobakan tidak dapat mengubah terlalu banyak pada akurasi .