

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Proses analisis yang dilakukan dalam merancang Sistem Prediksi Churn pasien, diketahui bahwa Saat ini pihak manajemen Rumah Sakit Muhammadiyah Gresik belum mengetahui bagaimana pola pasien kontrol kehamilan. Pihak Rumah Sakit perlu mengetahui pola pasien untuk menentukan strategi yang tepat agar pasien tidak beralih ke kompetitor lain. Dengan diketahuinya pola pasien maka akan membantu pihak rumah sakit untuk membuat strategi pemasaran yang efektif.

Dengan diterapkannya metode CART (*Classification and Regression Tree*) ke dalam aplikasi sistem prediksi *churn* pasien ini, diharapkan dapat membantu pihak rumah sakit mengetahui pola pasien yang melakukan *churn*.

3.2 Hasil Analisis

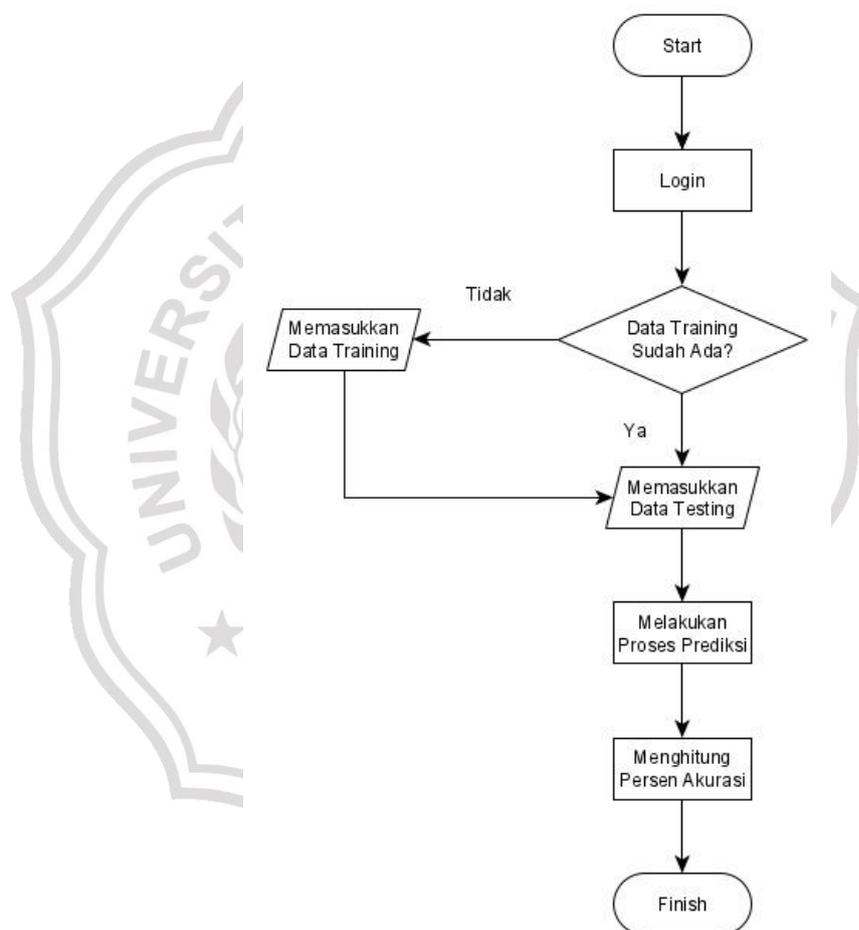
Hasil analisis *churn* pasien di Rumah Sakit Muhammadiyah Gresik dalam menyelesaikan permasalahan untuk mengetahui pola pasien maka dibutuhkan peran sebuah sistem prediksi *churn* pasien rumah sakit yang dapat membantu pihak Rumah Sakit untuk mengetahui pola pasien untuk membuat suatu strategi pemasaran agar pasien tidak melakukan *churn*.

Sistem prediksi churn pasien yang akan dibuat menggunakan metode CART (Classification and Regression Tree), dimana metode ini akan menghasilkan suatu pohon klasifikasi jika peubah responnya kategorik, dan menghasilkan pohon regresi jika peubah responnya kontinu. Tujuan utama CART (Classification and Regression Tree) adalah untuk mendapatkan suatu kelompok data yang akurat sebagai ciri dari suatu pengklasifikasian.

Sistem prediksi *churn* pasien akan menyimpan data pasien dan hasil prediksi, data pasien tersebut meliputi umur pasien, jumlah kontrol, dan

penanggung persalinan. Selanjutnya data pasien tersebut akan dihitung menggunakan metode CART(*Classification and Regression Tree*).

Sistem ini akan menghasilkan hasil prediksi apakah pasien tersebut akan melakukan *churn* atau tidak, hasil prediksi tersebut juga berisi pola pasien tersebut yang dapat digunakan sebagai pertimbangan dalam menentukan suatu strategi pemasaran. Diagram alur sistem prediksi *churn* pasien dengan menggunakan metode CART dapat dilihat pada **gambar 3.1**.

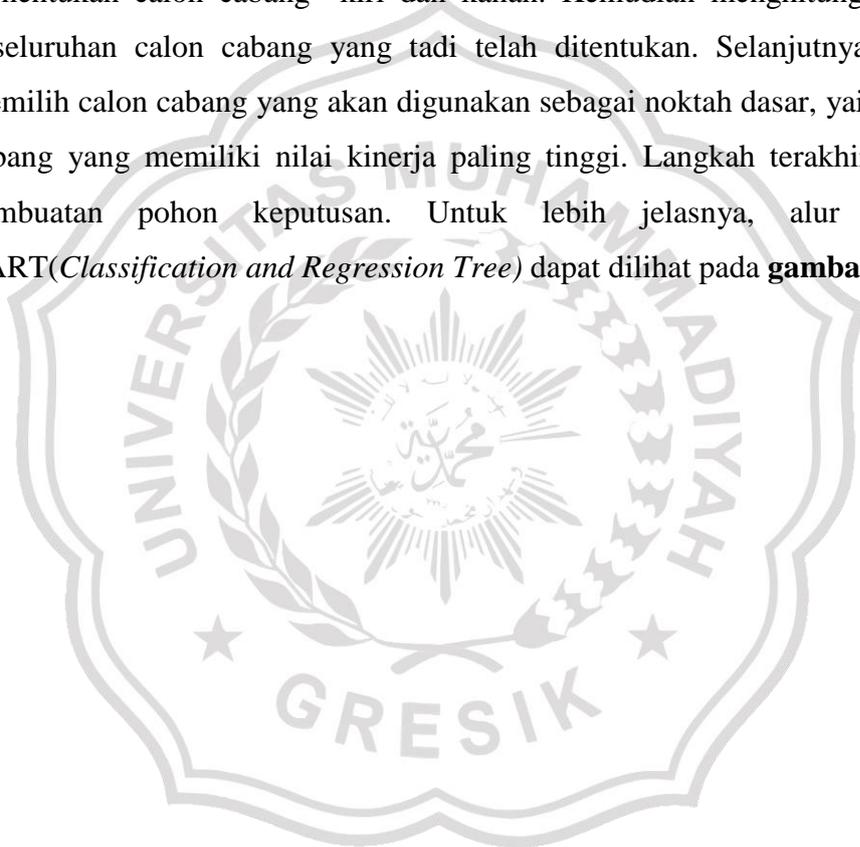


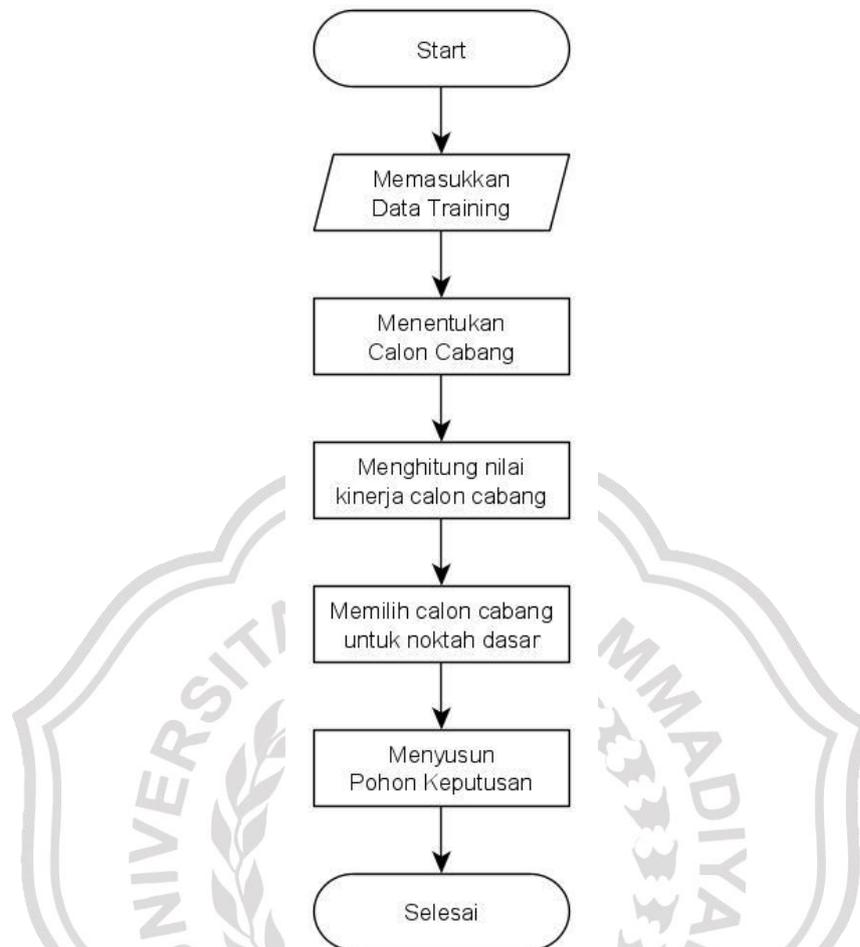
Gambar 3.1 Diagram Alur Sistem Prediksi *Churn* Pasien

Gambar 3.1 menjelaskan tahapan Sistem Prediksi *Churn* Pasien yang dimulai dengan login. Kemudian dilanjutkan dengan memasukkan data training dan data testing. Setelah itu system akan melakukan prediksi dengan menggunakan metode CART(*Classification and Regression Tree*).Dimana

akan dilakukan klasifikasi berupa pohon klasifikasi jika peubah responnya kategorik dan pohon regresi jika peubah responnya kontinu. Setelah proses klasifikasi selesai maka sistem akan melakukan output berupa pohon klasifikasi pasien tersebut, apakah pasien tersebut beresiko melakukan *churn* atau tidak.

Alur metode CART(*Classification and Regression Tree*) dimulai dengan menentukan terlebih dahulu data yang akan diproses. Selanjutnya yaitu menentukan calon cabang kiri dan kanan. Kemudian menghitung kinerja keseluruhan calon cabang yang tadi telah ditentukan. Selanjutnya adalah memilih calon cabang yang akan digunakan sebagai noktah dasar, yaitu calon cabang yang memiliki nilai kinerja paling tinggi. Langkah terakhir adalah pembuatan pohon keputusan. Untuk lebih jelasnya, alur metode CART(*Classification and Regression Tree*) dapat dilihat pada **gambar 3.2**





Gambar 3.2 Diagram Alur Metode CART(*Classification and Regression Tree*)

Gambar 3.2 Menjelaskan bahwa alur metode CART(*Classification and Regression Tree*) dimulai dengan memasukkan data training. Kemudian menentukan calon cabang kiri dan kanan. Selanjutnya menghitung nilai kinerja calon cabang yang tadi telah ditentukan. Kemudian calon cabang yang memiliki nilai kinerja paling tinggi akan dipilih sebagai noktah dasar. Langkah terakhir adalah pembuatan pohon keputusan.

Kebutuhan *stakeholder* dalam sistem ini adalah sebagai berikut:

Tabel 3.1 Kebutuhan *Stakeholder*

No.	Entitas	Informasi	Proses	Data
1.	Admin	<ul style="list-style-type: none"> - Mendapatkan laporan pola pasien - Mendapatkan laporan hasil prediksi <i>churn</i> pasien 	<ul style="list-style-type: none"> - Proses <i>input</i> data 	<ul style="list-style-type: none"> - Melakukan <i>input</i> data latih dan data uji

3.3 Representasi Model

Data yang diolah menjadi data latih pada sistem prediksi *churn* pasien ini adalah data pasien yang melakukan kontrol kehamilan di poli kandungan Rumah Sakit Muhammadiyah Gresik pada bulan Maret 2019 sampai September 2019. Data tersebut berisi 3937 data pasien. Variabel yang digunakan pada penelitian ini dapat dilihat pada **Tabel 3.1**. Data yang digunakan pada penelitian ini dapat dilihat pada **tabel 3.2**

Tabel 3.2 Variabel

Nama Variabel	Kategori	Skala
Churn	Ya	Nominal
	Tidak	
Jumlah Kontrol		Rasio
Umur		Rasio
Penanggung	Umum	Nominal
	Asuransi	

Tabel 3.3 Dataset

Nama Pasien	Penanggung	Jumlah Kontrol	Umur	Churn
Pasien 1	Umum	1	29	ya
Pasien 2	Umum	4	65	ya
Pasien 3	Umum	1	26	ya
Pasien 4	Umum	1	40	tidak
Pasien 5	Asuransi	1	39	ya
Pasien 6	Umum	2	29	ya
Pasien 7	Umum	1	31	ya
Pasien 8	Umum	1	28	ya
Pasien 9	Asuransi	2	33	tidak
Pasien 10	Umum	2	28	ya
Pasien 11	Umum	2	35	ya
Pasien 12	Umum	2	44	ya
Pasien 13	Umum	12	31	tidak
Pasien 14	Umum	1	42	ya
.....
Pasien 3937	Umum	1	26	ya

3.3.1 Indeks Gini

Tahapan pertama dalam metode CART (*Classification and Regression Tree*) adalah menghitung nilai keheterogenan simpul awal dengan menggunakan fungsi keheterogenan Indeks Gini yang telah didefinisikan pada **persamaan 2.1** Variabel yang akan dihitung menggunakan **persamaan 2.1** adalah variabel churn, karena variabel churn merupakan variabel tetap.

$$i(t) = \sum_{j=1}^n p(i|t)p(j|t) \dots(2.1)$$

1. Langkah pertama adalah mencari nilai $N(t)$, $N(t)$ adalah jumlah seluruh data. Pada penelitian ini jumlah seluruh data adalah 3937. Dengan demikian maka nilai $N(t)$ adalah 3937.

$$N(t) = 3937$$

2. Selanjutnya adalah menghitung nilai $p(i|t)$, untuk menghitung $p(i|t)$ dilakukan dengan cara menghitung nilai $N_i(t)$ dibagi nilai $N(t)$, $N_i(t)$ merupakan jumlah data yang memiliki nilai variabel tetap "Ya". Jumlah data yang memiliki variabel tetap "Ya" berjumlah 1869, oleh karena itu maka nilai $N_i(t)$ berarti 1869. Dengan demikian maka $p(i|t) =$

$$N_i(t) = 1869$$

$$p(i|t) = \frac{1869}{3937}$$

$$p(i|t) = 0,474726949$$

3. Selanjutnya adalah menghitung nilai $p(j|t)$, untuk menghitung $p(j|t)$ dilakukan dengan cara menghitung nilai $N_j(t)$ dibagi nilai $N(t)$, $N_j(t)$ merupakan jumlah data yang memiliki nilai variabel tetap "Tidak". Jumlah data yang memiliki variabel tetap "Tidak" berjumlah 2068, oleh karena itu maka nilai $N_j(t)$ berarti 2068.

Dengan demikian maka $p(j|t) =$

$$N_j(t) = 2068$$

$$p(j|t) = \frac{2068}{3937}$$

$$p(j|t) = 0,525273051$$

4. Setelah diketahui nilai $p(i|t)$ dan nilai $p(j|t)$, maka langkah berikutnya adalah menghitung nilai $i(t)$, untuk menghitung nilai $i(t)$ dilakukan dengan cara mengkalikan nilai $p(i|t)$ dan nilai $p(j|t) +$ nilai $p(j|t)$ dikali nilai $p(i|t)$.

$$i(t) = p(1|t)p(2|t) + p(2|t)p(1|t)$$

$$(i|t) = (0,474726949)(0,525273051) + (0,525273051)(0,474726949)$$

$$(i|t) = 0,498722546$$

Dengan demikian maka telah diketahui bahwa nilai $(i|t)$ adalah sebesar 0,498722546.

3.3.2 Goodness of Split

Tahapan Selanjutnya adalah menghitung *goodness of split*, perhitungan akan dilakukan pada 3 variabel prediktor, yaitu variabel Penanggung, jumlah kontrol dan umur. Perhitungan *goodness of split* dilakukan dengan menggunakan persamaan 2.2

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(i_L) - P_R i(i_R) \dots (2.2)$$

3.3.2.1 goodness of split Penanggung

Variabel Penanggung memiliki skala pengukuran nominal yang terdiri dari 2 kategori yaitu Umum dan Asuransi. Oleh karena itu maka akan dilakukan perhitungan *goodness of split* 1 kali, karena hanya terdapat 1 kemungkinan pemilah. Berikut adalah langkah-langkah menghitung nilai $\phi(s,t)$ variabel Penanggung :

1. Menentukan nilai NL ,

Nilai NL merupakan jumlah data yang nilai variabel Penanggungnya “Umum”, terdapat 2997 data untuk kategori tersebut. Oleh karena itu maka $NL = 2997$.

2. Menentukan nilai $N(1|tL)$,

Nilai $N(1|tL)$ merupakan jumlah data yang nilai variabel Penanggungnya “Umum” dan variabel Churnnya bernilai “Ya”, terdapat 1548 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 1548$.

3. Menentukan nilai $N(2|tL)$,

Merupakan jumlah data yang nilai variabel Penanggungnya “Umum” dan variabel Churnnya bernilai “Tidak”, terdapat 1449 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 1449$.

4. Menentukan nilai pL

Merupakan populasi NL dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NL dibagi jumlah seluruh data. Maka $pL = \frac{2997}{3937} = 0,761239522$.

5. Menentukan nilai $p(1|tL)$

Merupakan populasi $p(1|tL)$ dalam NL , untuk menghitung $p(1|tL)$ dilakukan dengan cara $N(1|tL)$ dibagi NL . Maka $p(1|tL) = \frac{1548}{2997} = 0,516516517$.

6. Menentukan nilai $p(2|tL)$

Merupakan populasi $p(2|tL)$ dalam NL , untuk menghitung $p(2|tL)$ dilakukan dengan cara $N(2|tL)$ dibagi NL . Maka $p(2|tL) = \frac{1449}{2997} = 0,483483483$.

7. Menghitung $i(tL)$

Untuk menghitung dilakukan dengan menggunakan persamaan $p(1|tL) * p(2|tL) + p(2|tL) * p(1|tL)$, Maka :

$$i(tL) = (0,516516517) * (0,483483483) + (0,483483483) * (0,516516517)$$

$$i(tL) = 0,499454409.$$

8. Menentukan nilai NR , nilai NR merupakan jumlah data yang nilai variabel Penanggungnya “Asuransi”, terdapat 940 data untuk kategori tersebut. Oleh karena itu maka $NR = 940$.

9. Menentukan nilai $N(1|tR)$

Nilai $N(1|tR)$ merupakan jumlah data yang nilai variabel Penanggungnya “Umum” dan variabel Churnnya bernilai “Ya”, terdapat 321 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 321$.

10. Menentukan nilai $N(2|tR)$

Merupakan jumlah data yang nilai variabel Penanggungnya “Umum” dan variabel Churnnya bernilai “Tidak”, terdapat 619 data untuk kategori tersebut. Oleh karena itu maka $N(2|tR) = 619$.

11. Menentukan nilai pR

Merupakan populasi NR dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NR dibagi jumlah seluruh data. Maka $pR = \frac{940}{3937} = 0,238760478$.

12. Menentukan nilai $p(1|tR)$

Merupakan populasi $p(1|tR)$ dalam NR , untuk menghitung $p(1|tR)$ dilakukan dengan cara $N(1|tR)$ dibagi NR . Maka $p(1|tR) = \frac{321}{940} = 0,341489362$.

13. Menentukan nilai $p(2|tR)$

Merupakan populasi $p(2|tR)$ dalam NR , untuk menghitung $p(2|tR)$ dilakukan dengan cara $N(2|tR)$ dibagi NR . Maka $p(2|tR) = \frac{619}{940} = 0,658510638$.

14. Menentukan nilai $i(tR)$

Untuk menghitung $i(tR)$ dilakukan dengan menggunakan persamaan :

$$p(1|tR) * p(2|tR) + p(2|tR) * p(1|tR)$$

Maka :

$$i(tR) = (0,341489362) * (0,658510638) +$$

$$(0,658510638) * (0,341489362)$$

$$i(tR) = 0,449748755.$$

15. Menghitung $\phi(s, t)$

Untuk menghitung dilakukan dengan menggunakan

Persamaan 2.2 :

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(i_L) - P_R i(i_R) \dots (2.2)$$

Maka :

$$\begin{aligned} \phi(s, t) &= 0,498722546 - 0,761239522(0,499454409) - \\ &0,238760478(0,449748755) \\ &= 0,011135882 \end{aligned}$$

Jadi $\phi(s, t)$ variabel Penanggung adalah sebesar 0,011135882.

Hasil perhitungan goodness of split Penanggung dapat dilihat pada **tabel 3.4**

Tabel 3.4 Perhitungan *Goodness of split* Penanggung

Simpul Kiri		Simpul Kanan	
$NL = 2997$	$pL =$ 0,761239522	$NR = 940$	$pR =$ 0,238760478
$N(1 tL) = 1548$	$p(1 tL) =$ 0,516516517	$N(1 tR) = 321$	$p(1 tR) =$ 0,341489362
$N(2 tL) = 1449$	$p(2 tL) =$ 0,483483483	$N(2 tR) = 619$	$p(2 tR) =$ 0,658510638
$i(tL) = 0,499454409$		$i(tR) = 0,449748755$	
<i>Goodness of Split</i> $\phi(s, t) = 0,011135882$			

3.3.2.2 goodness of split Jumlah Kontrol

Variabel Jumlah Kontrol memiliki skala pengukuran kontinu, karena skala pengukurannya berupa kontinu, maka threshold variabel jumlah kontrol adalah median dari 2 nilai amatan yang berurutan. Perhitungan median dari nilai amatan variabel jumlah kontrol dapat dilihat pada **tabel 3.5**

Tabel 3.5 Perhitungan nilai median variabel jumlah kontrol

No	Variabel 1	Variabel 2	Median
1	1	2	1,5
2	2	3	2,5
3	3	4	3,5
4	4	5	4,5
5	5	6	5,5
6	6	7	6,5
7	7	8	7,5
8	8	8	8,5
8	8	10	9,5
10	10	11	10,5
11	11	12	11,5
12	12	13	12,5
13	13	14	13,5
14	14	15	14,5
15	15	17	16

Perhitungan tersebut menghasilkan 15 variabel, setelah menghitung Nilai median dari seluruh nilai amatan yang muncul, maka nilai median tersebut akan digunakan sebagai threshold untuk menghitung goodness of split. Langkah berikutnya adalah menghitung proporsi pada simpul kiri dan kanan calon cabang. Berikut adalah langkah-langkah perhitungan goodness of split variabel jumlah kontrol :

1. Menentukan nilai *NL*,

Nilai *NL* merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $\leq 1,5$, terdapat 1406 data untuk kategori tersebut. Oleh karena itu maka $NL = 1406$.

2. Menentukan nilai $N(1|tL)$,

Nilai $N(1|tL)$ merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $\leq 1,5$ dan variabel Churnnya bernilai “Ya”, terdapat 899 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 899$.

3. Menentukan nilai $N(2|tL)$,

Merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $\leq 1,5$ dan variabel Churnnya bernilai “Tidak”, terdapat 507 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 507$.

4. Menentukan nilai pL

Merupakan populasi NL dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NL dibagi jumlah seluruh data. Maka $pL = \frac{1406}{3937} = 0,3571$.

5. Menentukan nilai $p(1|tL)$

Merupakan populasi $p(1|tL)$ dalam NL , untuk menghitung $p(1|tL)$ dilakukan dengan cara $N(1|tL)$ dibagi NL . Maka $p(1|tL) = \frac{899}{1406} = 0,63940$.

6. Menentukan nilai $p(2|tL)$

Merupakan populasi $p(2|tL)$ dalam NL , untuk menghitung $p(2|tL)$ dilakukan dengan cara $N(2|tL)$ dibagi NL . Maka $p(2|tL) = \frac{507}{1406} = 0,3606$.

7. Menghitung $i(tL)$

Untuk menghitung dilakukan dengan menggunakan persamaan $p(1|tL) * p(2|tL) + p(2|tL) * p(1|tL)$, Maka :

$$i(tL) = (0,63940) * (0,3606) + (0,3606) * (0,63940)$$

$$i(tL) = 0,46113.$$

8. Menentukan nilai NR

Nilai NR merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $>1,5$, terdapat 2531 data untuk kategori tersebut. Oleh karena itu maka $NR = 2531$.

9. Menentukan nilai $N(1|tR)$

Nilai $N(1|tR)$ merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $>1,5$ dan variabel Churnnya bernilai “Ya”, terdapat 970 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 970$.

10. Menentukan nilai $N(2|tR)$

Merupakan jumlah data yang memiliki nilai variabel Jumlah Kontrol sebesar $>1,5$ dan variabel Churnnya bernilai “Tidak”, terdapat 1561 data untuk kategori tersebut. Oleh karena itu maka $N(2|tR) = 1561$.

11. Menentukan nilai pR

Merupakan populasi NR dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NR dibagi jumlah seluruh data. Maka $pR = \frac{2531}{3937} = 0,642875$.

12. Menentukan nilai $p(1|tR)$

Merupakan populasi $p(1|tR)$ dalam NR , untuk menghitung $p(1|tR)$ dilakukan dengan cara $N(1|tR)$ dibagi NR . Maka $p(1|tR) = \frac{321}{940} = 0,341489$.

13. Menentukan nilai $p(2|tR)$

Merupakan populasi $p(2|tR)$ dalam NR , untuk menghitung $p(2|tR)$ dilakukan dengan cara $N(2|tR)$ dibagi NR . Maka $p(2|tR) = \frac{619}{940} = 0,658511$.

14. Menentukan nilai $i(tR)$

Untuk menghitung $i(tR)$ dilakukan dengan menggunakan persamaan :

$$p(1|tR) * p(2|tR) + p(2|tR) * p(1|tR)$$

Maka :

$$i(tR) = (0,38325) * (0,61675) + (0,61675) * (0,38325)$$

$$i(tR) = 0,472738.$$

15. Menghitung $\phi(s, t)$

Untuk menghitung $\phi(s, t)$ dilakukan dengan menggunakan persamaan 2.2:

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(i_L) - P_R i(i_R) \dots (2.2)$$

Maka :

$$\phi(s, t) = 0,498722546 - 0,3571(0,46113) -$$

$$0,642875(0,472738)$$

$$= 0,03013.$$

Jadi $\phi(s, t)$ variabel Jumlah Kontrol untuk threshold 1,5 adalah sebesar 0,03013.

Langkah- langkah tersebut akan dilakukan kembali terhadap seluruh threshold yang ada pada variabel Jumlah Kontrol. Hasil Perhitungan proporsi calon cabang kiri dapat dilihat pada tabel dan tabel 3.6.

Tabel 3.6 Perhitungan Calon cabang kiri jumlah kontrol

No	Nama Variabel	NL	N(1 tL)	N(2 tL)	pL	p(1 tL)	p(2 tL)	i(tL)
1	1,5	1406	899	507	0,3571	0,63940	0,3606	0,46113
2	2,5	2163	1269	894	0,5494	0,58669	0,4133	0,48497
3	3,5	2687	1464	1223	0,6825	0,54485	0,4552	0,49598
4	4,5	3064	1604	1460	0,7783	0,52350	0,4765	0,49890
5	5,5	3329	1699	1630	0,8456	0,51036	0,4896	0,49979

6	6,5	3515	1764	1751	0,8928	0,50185	0,4982	0,49999
7	7,5	3662	1812	1850	0,9301	0,49481	0,5052	0,49995
8	8,5	3760	1833	1927	0,9550	0,48750	0,5125	0,49969
8	9,5	3825	1849	1976	0,9716	0,48340	0,5166	0,49945
10	10,5	3863	1859	2004	0,9812	0,48123	0,5188	0,49930
11	11,5	3893	1860	2033	0,9888	0,47778	0,5222	0,49901
12	12,5	3915	1864	2051	0,9944	0,47612	0,5239	0,49886
13	13,5	3926	1865	2061	0,9972	0,47504	0,5250	0,49875
14	14,5	3933	1868	2065	0,9990	0,47496	0,5250	0,49875
15	16	3936	1869	2067	0,9997	0,47485	0,5252	0,49873

Berikutnya adalah perhitungan untuk calon cabang kanan variabel Jumlah Kontrol. Hasil perhitungan proporsi calon cabang kanan variabel Jumlah Kontrol dapat dilihat pada tabel 3.7.

Tabel 3.7 Perhitungan Calon cabang kanan jumlah kontrol

No	Nama Variabel	NR	N(1 tR)	N(2 tR)	pR	p(1 tR)	p(2 tR)	i(tR)
1	1,5	2531	970	1561	0,642875	0,38325	0,61675	0,472738
2	2,5	1774	600	1174	0,450597	0,33822	0,66178	0,447654
3	3,5	1250	405	845	0,317501	0,32400	0,67600	0,438048
4	4,5	873	265	608	0,221742	0,30355	0,69645	0,422816
5	5,5	608	170	438	0,154432	0,27961	0,72039	0,402852
6	6,5	422	105	317	0,107188	0,24882	0,75118	0,373812
7	7,5	275	57	218	0,069850	0,20727	0,79273	0,328621
8	8,5	177	36	141	0,044958	0,20339	0,79661	0,324045
8	9,5	112	20	92	0,028448	0,17857	0,82143	0,293367
10	10,5	74	10	64	0,018796	0,13514	0,86486	0,233747
11	11,5	44	9	35	0,011176	0,20455	0,79545	0,325413
12	12,5	22	5	17	0,005588	0,22727	0,77273	0,35124

13	13,5	11	4	7	0,002794	0,36364	0,63636	0,46281
14	14,5	4	1	3	0,001016	0,25000	0,75000	0,375
15	16	1	0	1	0,000254	0,00000	1,00000	0

Nilai proporsi simpul yang didapat akan digunakan untuk menghitung goodness of split. Hasil perhitungan goodness of split dapat dilihat pada tabel 3.8

Tabel 3.8 *Goodness of split* jumlah kontrol

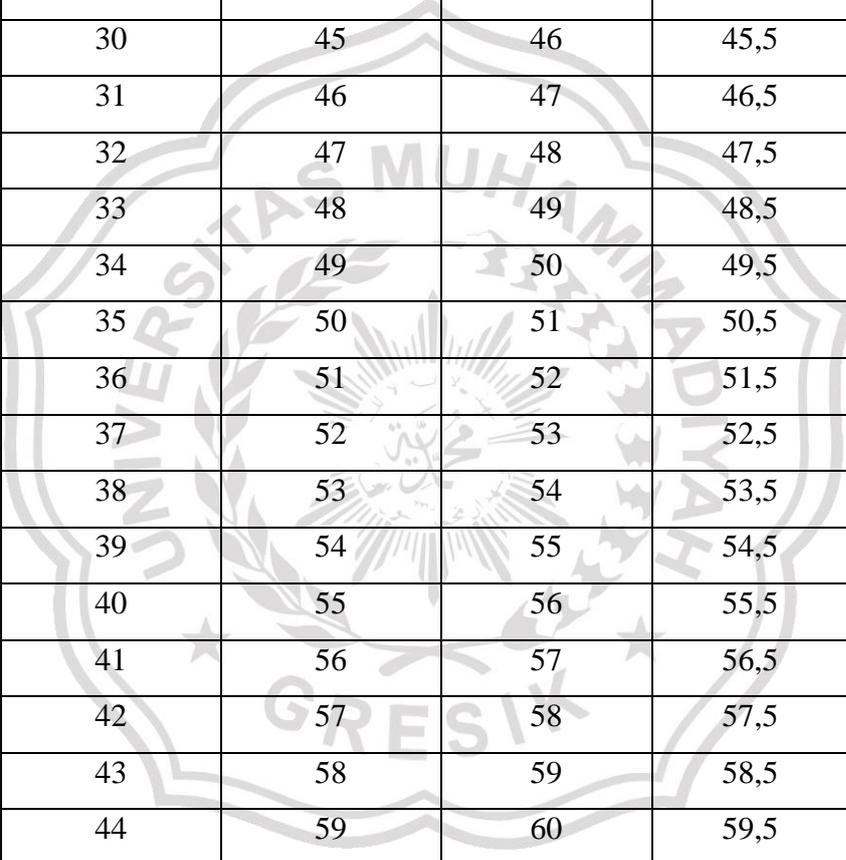
No	Nama Variabel	<i>Goodness of Split</i> $\phi(s,t)$
1	1,5	0,03013
2	2,5	0,03057
3	3,5	0,02114
4	4,5	0,01670
5	5,5	0,01391
6	6,5	0,01225
7	7,5	0,01074
8	8,5	0,00693
9	9,5	0,00514
10	10,5	0,00442
11	11,5	0,00165
12	12,5	0,00069
13	13,5	0,00007
14	14,5	0,00010
15	16	0,00011

3.3.2.3 Goodness of split Umur

Variabel Umur memiliki skala pengukuran kontinu, karena skala pengukurannya berupa kontinu, maka threshold variabel jumlah kontrol adalah median dari 2 nilai amatan yang berurutan. Perhitungan median dari nilai amatan variabel jumlah kontrol dapat dilihat pada **tabel 3.9**

Tabel 3.9 Perhitungan nilai median variabel Umur

No	Variabel 1	Variabel 2	Median
1	16	17	16,5
2	17	18	17,5
3	18	19	18,5
4	19	20	19,5
5	20	21	20,5
6	21	22	21,5
7	22	23	22,5
8	23	24	23,5
9	24	25	24,5
10	25	26	25,5
11	26	27	26,5
12	27	28	27,5
13	28	29	28,5
14	29	30	29,5
15	30	31	30,5
16	31	32	31,5
17	32	33	32,5
18	33	34	33,5
19	34	35	34,5
20	35	36	35,5
21	36	37	36,5



22	37	38	37,5
23	38	39	38,5
24	39	40	39,5
25	40	41	40,5
26	41	42	41,5
27	42	43	42,5
28	43	44	43,5
29	44	45	44,5
30	45	46	45,5
31	46	47	46,5
32	47	48	47,5
33	48	49	48,5
34	49	50	49,5
35	50	51	50,5
36	51	52	51,5
37	52	53	52,5
38	53	54	53,5
39	54	55	54,5
40	55	56	55,5
41	56	57	56,5
42	57	58	57,5
43	58	59	58,5
44	59	60	59,5
45	60	61	60,5
46	61	62	61,5
47	62	63	62,5
48	63	64	63,5
49	64	65	64,5

Perhitungan tersebut menghasilkan 49 variabel, setelah menghitung Nilai median dari seluruh nilai amatan yang muncul, maka nilai median tersebut akan digunakan sebagai threshold untuk menghitung goodness of split. Langkah berikutnya adalah menghitung proporsi pada simpul kiri dan kanan calon cabang. Berikut adalah langkah-langkah perhitungan goodness of split variabel Umur :

1. Menentukan nilai NL ,

Nilai NL merupakan jumlah data yang memiliki nilai variabel Umur sebesar $\leq 16,5$, terdapat 7 data untuk kategori tersebut. Oleh karena itu maka $NL = 7$.

2. Menentukan nilai $N(1|tL)$,

Nilai $N(1|tL)$ merupakan jumlah data yang memiliki nilai variabel Umur sebesar $\leq 16,5$ dan variabel Churnnya bernilai "Ya", terdapat 6 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 6$.

3. Menentukan nilai $N(2|tL)$,

Merupakan jumlah data yang memiliki nilai variabel Umur sebesar $\leq 16,5$ dan variabel Churnnya bernilai "Tidak", terdapat 1 data untuk kategori tersebut. Oleh karena itu maka $N(2|tR) = 1$.

4. Menentukan nilai pL

Merupakan populasi NL dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NL dibagi jumlah seluruh data. Maka $pL = \frac{7}{3937} = 0,001778$.

5. Menentukan nilai $p(1|tL)$

Merupakan populasi $p(1|tL)$ dalam NL , untuk menghitung $p(1|tL)$ dilakukan dengan cara $N(1|tL)$ dibagi NL . Maka $p(1|tL) = \frac{6}{7} = 0,857143$.

6. Menentukan nilai $p(2|tL)$

Merupakan populasi $p(2|tL)$ dalam NL , untuk menghitung $p(2|tL)$ dilakukan dengan cara $N(2|tL)$ dibagi NL . Maka $p(2|tL) = \frac{1}{7} = 0,142857$.

7. Menghitung $i(tL)$

Untuk menghitung dilakukan dengan menggunakan persamaan $p(1|tL) * p(2|tL) + p(2|tL) * p(1|tL)$, Maka :

$$i(tL) = (0,857143) * (0,142857) + (0,142857) * (0,857143)$$

$$i(tL) = 0,244898.$$

8. Menentukan nilai NR

Nilai NR merupakan jumlah data yang memiliki nilai variabel Umur sebesar $>16,5$, terdapat 3930 data untuk kategori tersebut. Oleh karena itu maka $NR = 3930$.

9. Menentukan nilai $N(1|tR)$

Nilai $N(1|tR)$ merupakan jumlah data yang memiliki nilai variabel Umur sebesar $>16,5$ dan variabel Churnnya bernilai "Ya", terdapat 1863 data untuk kategori tersebut. Oleh karena itu maka $N(1|tR) = 1863$.

10. Menentukan nilai $N(2|tR)$

Merupakan jumlah data yang memiliki nilai variabel Umur sebesar $>16,5$ dan variabel Churnnya bernilai “Tidak”, terdapat 2067 data untuk kategori tersebut. Oleh karena itu maka $N(2|tR) = 2067$.

11. Menentukan nilai pR

Merupakan populasi NR dalam keseluruhan data, untuk menghitung pL dilakukan dengan cara NR dibagi jumlah seluruh data. Maka $pR = \frac{3930}{3937} = 0,99822$.

12. Menentukan nilai $p(1|tR)$

Merupakan populasi $p(1|tR)$ dalam NR , untuk menghitung $p(1|tR)$ dilakukan dengan cara $N(1|tR)$ dibagi NR . Maka $p(1|tR) = \frac{1863}{3930} = 0,474046$.

13. Menentukan nilai $p(2|tR)$

Merupakan populasi $p(2|tR)$ dalam NR , untuk menghitung $p(2|tR)$ dilakukan dengan cara $N(2|tR)$ dibagi NR . Maka $p(2|tR) = \frac{2067}{3930} = 0,52595$.

14. Menentukan nilai $i(tR)$

Untuk menghitung $i(tR)$ dilakukan dengan menggunakan persamaan :

$$p(1|tR) * p(2|tR) + p(2|tR) * p(1|tR)$$

Maka :

$$i(tR) = (0,474046) * (0,52595) + (0,52595) * (0,474046)$$

$$i(tR) = 0,498653.$$

15. Menghitung $\phi(s, t)$

Untuk menghitung dilakukan dengan menggunakan persamaan :

$$(\phi(s, t)) = \Delta i(s, t) = i(t) - P_L i(i_L) - P_R i(i_R)$$

Maka :

$$\phi(s, t) = 0,498722546 - 0,001778 (0,244898) - 0,99822 (0,498653)$$

$$= 0,00052.$$

Jadi $\phi(s, t)$ variabel Umur untuk threshold 16,5 adalah sebesar 0,00052.

Langkah- langkah tersebut akan dilakukan kembali terhadap seluruh threshold yang ada pada variabel Jumlah Kontrol. Hasil Perhitungan proporsi calon cabang kiri dapat dilihat pada tabel dan tabel **3.10**.

Tabel 3.10 Perhitungan calon cabang kiri umur

No	Nama Variabel	NL	N(1 tL)	N(2 tL)	pL	p(1 tL)	p(2 tL)	i(tL)
1	16,5	7	6	1	0,001778	0,857143	0,142857	0,244898
2	17,5	21	17	4	0,005334	0,809524	0,190476	0,30839
3	18,5	38	29	9	0,009652	0,763158	0,236842	0,361496
4	19,5	75	49	26	0,01905	0,653333	0,346667	0,452978
5	20,5	132	82	50	0,033528	0,621212	0,378788	0,470615
6	21,5	210	126	84	0,05334	0,6	0,4	0,48
7	22,5	342	185	157	0,086868	0,540936	0,459064	0,496649
8	23,5	570	290	280	0,14478	0,508772	0,491228	0,499846
9	24,5	812	411	401	0,206248	0,506158	0,493842	0,499924

10	25,5	1102	544	558	0,279909	0,493648	0,506352	0,499919
11	26,5	1378	678	700	0,350013	0,492017	0,507983	0,499873
12	27,5	1614	785	829	0,409957	0,486369	0,513631	0,499628
13	28,5	1825	866	959	0,463551	0,474521	0,525479	0,498702
14	29,5	2043	962	1081	0,518923	0,470876	0,529124	0,498304
15	30,5	2246	1052	1194	0,570485	0,468388	0,531612	0,498001
16	31,5	2434	1136	1298	0,618237	0,466721	0,533279	0,497785
17	32,5	2608	1210	1398	0,662433	0,463957	0,536043	0,497402
18	33,5	2747	1259	1488	0,697739	0,458318	0,541682	0,496525
19	34,5	2880	1307	1573	0,731521	0,453819	0,546181	0,495735
20	35,5	3005	1355	1650	0,763272	0,450915	0,549085	0,495181
21	36,5	3136	1415	1721	0,796546	0,451212	0,548788	0,495239
22	37,5	3245	1464	1781	0,824232	0,451156	0,548844	0,495228
23	38,5	3332	1508	1824	0,84633	0,452581	0,547419	0,495503
24	39,5	3403	1550	1853	0,864364	0,45548	0,54452	0,496036
25	40,5	3462	1582	1880	0,87935	0,456961	0,543039	0,496295
26	41,5	3503	1598	1905	0,889764	0,45618	0,54382	0,49616
27	42,5	3735	1725	2010	0,948692	0,461847	0,538153	0,497089
28	43,5	3811	1774	2037	0,967996	0,465495	0,534505	0,497619
29	44,5	3832	1787	2045	0,97333	0,466336	0,533664	0,497733
30	45,5	3853	1802	2051	0,978664	0,467688	0,532312	0,497912
31	46,5	3867	1811	2056	0,98222	0,468322	0,531678	0,497993
32	47,5	3874	1817	2057	0,983998	0,469024	0,530976	0,498081
33	48,5	3883	1824	2059	0,986284	0,46974	0,53026	0,498169
34	49,5	3888	1827	2061	0,987554	0,469907	0,530093	0,498189
35	50,5	3893	1831	2062	0,988824	0,470331	0,529669	0,49824
36	51,5	3898	1835	2063	0,990094	0,470754	0,529246	0,498289
37	52,5	3906	1842	2064	0,992126	0,471582	0,528418	0,498385
38	53,5	3908	1844	2064	0,992634	0,471853	0,528147	0,498415
39	54,5	3912	1847	2065	0,99365	0,472137	0,527863	0,498447

40	55,5	3913	1848	2065	0,993904	0,472272	0,527728	0,498462
41	56,5	3915	1850	2065	0,994412	0,472542	0,527458	0,498492
42	57,5	3918	1853	2065	0,995174	0,472945	0,527055	0,498536
43	58,5	3920	1855	2065	0,995682	0,473214	0,526786	0,498565
44	59,5	3923	1858	2065	0,996444	0,473617	0,526383	0,498608
45	60,5	3924	1859	2065	0,996698	0,473751	0,526249	0,498622
46	61,5	3926	1861	2065	0,997206	0,474019	0,525981	0,49865
47	62,5	3928	1861	2067	0,997714	0,473778	0,526222	0,498625
48	63,5	3932	1864	2068	0,99873	0,474059	0,525941	0,498654
49	64,5	3936	1868	2068	0,999746	0,474593	0,525407	0,498709

Berikutnya adalah perhitungan untuk calon cabang kanan variabel Umur. Hasil perhitungan proporsi calon cabang kanan variabel Umur dapat dilihat pada tabel 3.11.

Tabel 3.11 Perhitungan calon cabang kanan umur

No	Nama Variabel	NR	N(1 tR)	N(2 tR)	pR	p(1 tR)	p(2 tR)	i(tR)
1	16,5	3930	1863	2067	0,99822	0,474046	0,52595	0,498653
2	17,5	3916	1852	2064	0,99467	0,472932	0,52707	0,498535
3	18,5	3899	1840	2059	0,99035	0,471916	0,52808	0,498423
4	19,5	3862	1820	2042	0,98095	0,471258	0,52874	0,498348
5	20,5	3805	1787	2018	0,96647	0,469645	0,53035	0,498157
6	21,5	3727	1743	1984	0,94666	0,467668	0,53233	0,497909
7	22,5	3595	1684	1911	0,91313	0,468428	0,53157	0,498006
8	23,5	3367	1579	1788	0,85522	0,468963	0,53104	0,498073
9	24,5	3125	1458	1667	0,79375	0,46656	0,53344	0,497764
10	25,5	2835	1325	1510	0,72009	0,467372	0,53263	0,497871
11	26,5	2559	1191	1368	0,64999	0,465416	0,53458	0,497608
12	27,5	2323	1084	1239	0,59004	0,466638	0,53336	0,497774

13	28,5	2112	1003	1109	0,53645	0,474905	0,52509	0,498741
14	29,5	1894	907	987	0,48108	0,478881	0,52112	0,499108
15	30,5	1691	817	874	0,42951	0,483146	0,51685	0,499432
16	31,5	1503	733	770	0,38176	0,487691	0,51231	0,499697
17	32,5	1329	659	670	0,33757	0,495862	0,50414	0,499966
18	33,5	1190	610	580	0,30226	0,512605	0,48739	0,499682
19	34,5	1057	562	495	0,26848	0,531693	0,46831	0,497991
20	35,5	932	514	418	0,23673	0,551502	0,44850	0,494695
21	36,5	801	454	347	0,20345	0,566792	0,43321	0,491078
22	37,5	692	405	287	0,17577	0,58526	0,41474	0,485461
23	38,5	605	361	244	0,15367	0,596694	0,40331	0,4813
24	39,5	534	319	215	0,13564	0,597378	0,40262	0,481035
25	40,5	475	287	188	0,12065	0,604211	0,39579	0,47828
26	41,5	434	271	163	0,11024	0,624424	0,37558	0,469037
27	42,5	202	144	58	0,05131	0,712871	0,28713	0,409372
28	43,5	126	95	31	0,03200	0,753968	0,24603	0,371
29	44,5	105	82	23	0,02667	0,780952	0,21905	0,342132
30	45,5	84	67	17	0,02134	0,797619	0,20238	0,322846
31	46,5	70	58	12	0,01778	0,828571	0,17143	0,284082
32	47,5	63	52	11	0,01600	0,825397	0,17460	0,288234
33	48,5	54	45	9	0,01372	0,833333	0,16667	0,277778
34	49,5	49	42	7	0,01245	0,857143	0,14286	0,244898
35	50,5	44	38	6	0,01118	0,863636	0,13636	0,235537
36	51,5	39	34	5	0,00991	0,871795	0,12821	0,223537
37	52,5	31	27	4	0,00787	0,870968	0,12903	0,224766
38	53,5	29	25	4	0,00737	0,862069	0,13793	0,237812
39	54,5	25	22	3	0,00635	0,88	0,12000	0,2112
40	55,5	24	21	3	0,00610	0,875	0,12500	0,21875
41	56,5	22	19	3	0,00559	0,863636	0,13636	0,235537
42	57,5	19	16	3	0,00483	0,842105	0,15789	0,265928

43	58,5	17	14	3	0,00432	0,823529	0,17647	0,290657
44	59,5	14	11	3	0,00356	0,785714	0,21429	0,336735
45	60,5	13	10	3	0,00330	0,769231	0,23077	0,35503
46	61,5	11	8	3	0,00279	0,727273	0,27273	0,396694
47	62,5	9	8	1	0,00229	0,888889	0,111111	0,197531
48	63,5	5	5	0	0,00127	1	0,00000	0
49	64,5	1	1	0	0,00025	1	0,00000	0

Nilai proporsi simpul yang didapat akan digunakan untuk menghitung goodness of split. Hasil perhitungan goodness of split dapat dilihat pada **tabel. 3.12**

Tabel 3.12 Goodness of split umur

No	Nama Variabel	Goodness of Split $\phi(s,t)$
1	16,5	0,00052
2	17,5	0,00120
3	18,5	0,00162
4	19,5	0,00124
5	20,5	0,00149
6	21,5	0,00177
7	22,5	0,00083
8	23,5	0,00039
9	24,5	0,00051
10	25,5	0,00028
11	26,5	0,00032
12	27,5	0,00019
13	28,5	0,00000
14	29,5	0,00003
15	30,5	0,00011
16	31,5	0,00021

17	32,5	0,00046
18	33,5	0,00124
19	34,5	0,00238
20	35,5	0,00366
21	36,5	0,00433
22	37,5	0,00521
23	38,5	0,00540
24	39,5	0,00472
25	40,5	0,00460
26	41,5	0,00555
27	42,5	0,00613
28	43,5	0,00516
29	44,5	0,00514
30	45,5	0,00455
31	46,5	0,00453
32	47,5	0,00400
33	48,5	0,00358
34	49,5	0,00369
35	50,5	0,00342
36	51,5	0,00315
37	52,5	0,00249
38	53,5	0,00223
39	54,5	0,00210
40	55,5	0,00197
41	56,5	0,00170
42	57,5	0,00131
43	58,5	0,00106
44	59,5	0,00069
45	60,5	0,00057
46	61,5	0,00036

47	62,5	0,00079
48	63,5	0,00070
49	64,5	0,00014

3.3.3 Pemilihan Pemilah Terbaik

Dalam menyusun pohon keputusan, pemilah terbaik dipilih berdasarkan nilai goodness of split terbesar. Pembentukan pohon keputusan dimulai dari menentukan root node terlebih dahulu. Pembentukan pohon keputusan akan dilakukan sampai diperoleh simpul yang homogen, simpul yang sudah homogen akan menjadi terminal node.

3.3.3.1 Pemilahan Pertama

Root node dipilih berdasarkan nilai goodness of split terbesar. perbandingan nilai goodness of split dapat dilihat pada tabel 3.13.

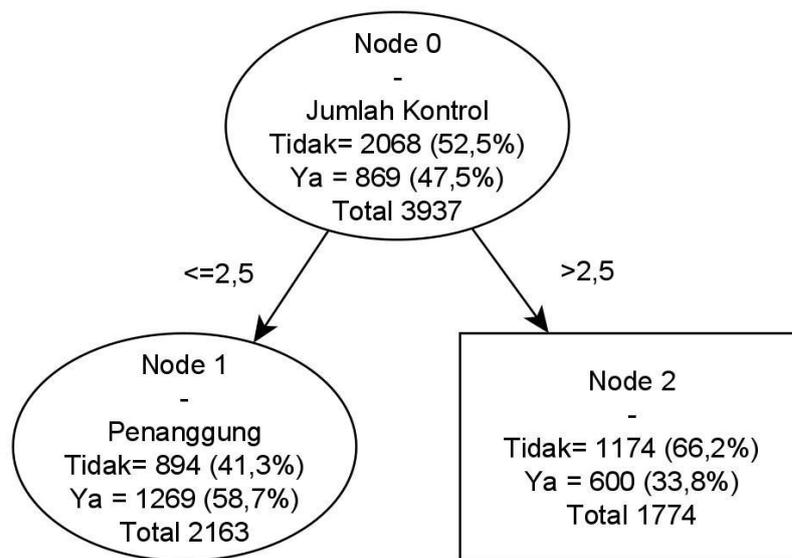
Tabel 3.13 Perbandingan Goodness of split pemilahan pertama

Nama Variabel	Nama Threshold	Goodness of Split $\phi(s,t)$
Penanggung	Umum	0,011136
	Asuransi	0,011136
Jumlah Kontrol	1,5	0,03013
	2,5	0,03057
	3,5	0,02114
	4,5	0,01670
	5,5	0,01391
	6,5	0,01225
	7,5	0,01074
	8,5	0,00693
	9,5	0,00514
	10,5	0,00442

	11,5	0,00165
	12,5	0,00069
	13,5	0,00007
	14,5	0,00010
	16	0,00011
Umur	16,5	0,00052
	17,5	0,00120
	18,5	0,00162
	19,5	0,00124
	20,5	0,00149
	21,5	0,00177
	22,5	0,00083
	23,5	0,00039
	24,5	0,00051
	25,5	0,00028
	26,5	0,00032
	27,5	0,00019
	28,5	0,00000
	29,5	0,00003
	30,5	0,00011
	31,5	0,00021
	32,5	0,00046
	33,5	0,00124
	34,5	0,00238
	35,5	0,00366
36,5	0,00433	
37,5	0,00521	
38,5	0,00540	
39,5	0,00472	
40,5	0,00460	

41,5	0,00555
42,5	0,00613
43,5	0,00516
44,5	0,00514
45,5	0,00455
46,5	0,00453
47,5	0,00400
48,5	0,00358
49,5	0,00369
50,5	0,00342
51,5	0,00315
52,5	0,00249
53,5	0,00223
54,5	0,00210
55,5	0,00197
56,5	0,00170
57,5	0,00131
58,5	0,00106
59,5	0,00069
60,5	0,00057
61,5	0,00036
62,5	0,00079
63,5	0,00070
64,5	0,00014

Berdasarkan **tabel 3.13** maka yang akan menjadi root node adalah variabel jumlah kontrol dengan nilai threshold 2,5 dengan nilai goodness of split sebesar 0,03057. Pohon klasifikasi hasil pemilahan simpul awal dapat dilihat pada **gambar 3.3**



Gambar 3.3 Pohon klasifikasi pemilahan pertama

Simpul kiri(Node 1) dari Pohon klasifikasi pemilahan simpul awal pada gambar 3 adalah noktah keputusan dan akan dilakukan pemilahan lagi Sedangkan simpul kanan (Node 2) adalah noktah terminasi karena pada node tersebut apabila dilakukan pemilahan lagi tidak terjadi penurunan keheterogenan yang signifikan.

3.3.3.2 Pemilahan Ke-dua

Pemilah Optimal dipilih berdasarkan nilai goodness of split terbesar, perbandingan nilai goodness of split dapat dilihat pada tabel 3.14.

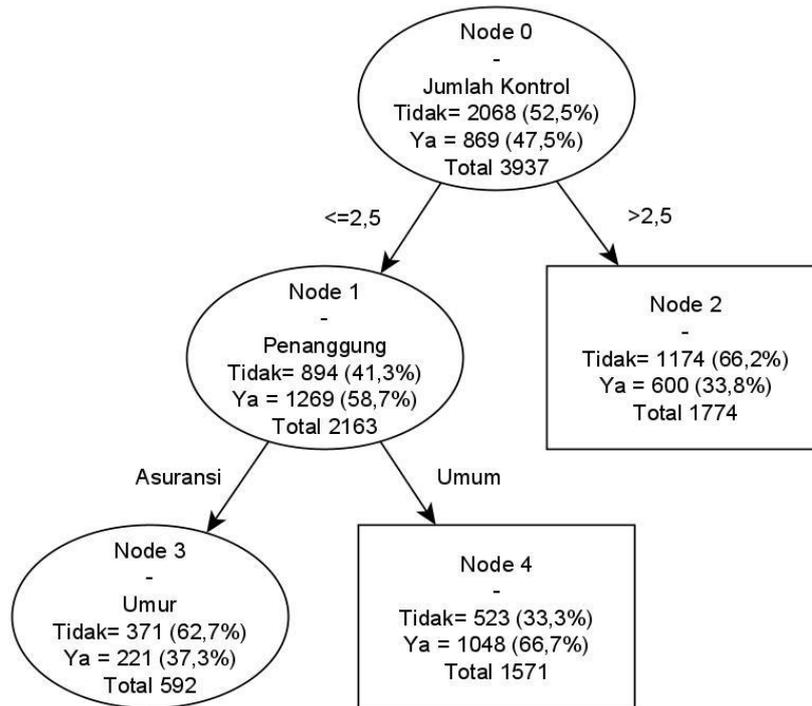
Tabel 3.14 Perbandingan goodness of split pemilahan kedua

Nama Variabel	Nama Threshold	Goodness of Split $\phi(s,t)$
Penanggung		0,082880171
Jumlah Kontrol		0,0103231446
Umur	16,5	0,000950004589
	17,5	0,001833948461

	18,5	0,002349834681
	19,5	0,001869605299
	20,5	0,001686741057
	21,5	0,002405637292
	22,5	0,000980073985
	23,5	0,000245877688
	24,5	0,000144595069
	25,5	0,000033343102
	26,5	0,000102746314
	27,5	0,000000345937
	28,5	0,000032569396
	29,5	0,000085633851
	30,5	0,000282561076
	31,5	0,000854327068
	32,5	0,001317194216
	33,5	0,003074139403
	34,5	0,005253095926
	35,5	0,006583377631
	36,5	0,006943606473
	37,5	0,007878752438
	38,5	0,009700972882
	39,5	0,008255984308
	40,5	0,008603478822
	41,5	0,010778572787
	42,5	0,006551221992
	43,5	0,004915619344
	44,5	0,004674266367
	45,5	0,004869425574
	46,5	0,005746476838
	47,5	0,004946438225
	48,5	0,004492133820

	49,5	0,004013329432
	50,5	0,003537604463
	51,5	0,003065501847
	52,5	0,002621857538
	53,5	0,002305376188
	54,5	0,001833948461
	55,5	0,001677915270
	56,5	0,001522581260
	57,5	0,001062273292
	58,5	0,000762812814
	60	0,000338077826
	61,5	0,000210492055
	62,5	0,000632627391
	63,5	0,000315837656

Berdasarkan tabel **3.14** maka yang akan menjadi Pemilah adalah variabel 0,082880171 dengan nilai goodness of split sebesar 0,082880171. Pohon klasifikasi hasil pemilahan dapat dilihat pada **gambar 3.4**



Gambar 3.4 Klasifikasi pemilahan kedua

Simpul kiri(Node 3) dari Pohon klasifikasi pemilahan simpul kedua pada gambar 3 adalah noktah keputusan dan akan dilakukan pemilahan lagi. Sedangkan simpul kanan(Node 4) adalah noktah keputusan karena pada node tersebut apabila dilakukan pemilahan lagi tidak terjadi penurunan keheterogenan yang signifikan.

3.3.3.3 Pemilahan ke-tiga

Pemilah Optimal dipilih berdasarkan nilai goodness of split terbesar, perbandingan nilai goodness of split dapat dilihat pada tabel 3.15.

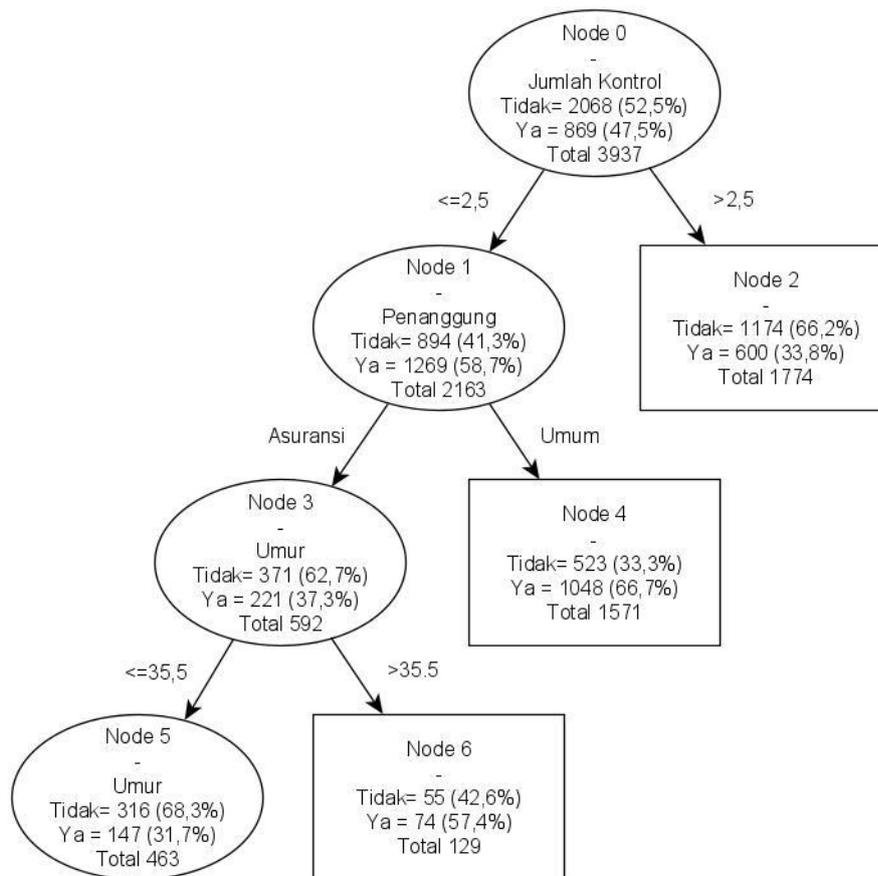
Tabel 3.15 Perbandingan *goodness of split* pemilahan ketiga

Nama Variabel	Nama Threshold	Goodness of Split $\phi(s,t)$
Jumlah Kontrol		0,000000876
Umur	16,5	0,004001
	17,5	0,004334

19	0,005051
20,5	0,001901
21,5	0,000715
22,5	0,000524
23,5	0,002345
24,5	0,002396
25,5	0,001519
26,5	0,001608
27,5	0,002525
28,5	0,002867
29,5	0,003854
30,5	0,005102
31,5	0,008546
32,5	0,010132
33,5	0,012931
34,5	0,019512
35,5	0,022364
36,5	0,020797
37,5	0,014259
38,5	0,021606
39,5	0,015211
40,5	0,015187
41,5	0,021806
42,5	0,020541
43,5	0,018559
44,5	0,01589
45,5	0,017684
47	0,014936
50	0,013496
52,5	0,012126

	53,5	0,01076
	55,5	0,009399
	57,5	0,006691
	58,5	0,005343
	60	0,004001
	62	0,002663
	63,5	0,001329

Berdasarkan tabel perhitungan tersebut maka yang akan menjadi Pemilah adalah threshold 35,5 dari variabel umur dengan nilai goodness of split sebesar 0,022364. Pohon klasifikasi hasil pemilahan dapat dilihat pada **gambar 3.5**



Gambar 3.5 Klasifikasi pemilahan ke-tiga

Simpul kiri dalam pemilahan ini adalah noktah keputusan dan masih akan dilakukan pemilahan lagi. Sedangkan simpul kanan adalah simpul terminal karena pada node tersebut apabila dilakukan pemilahan lagi tidak terjadi penurunan keheterogenan yang signifikan.

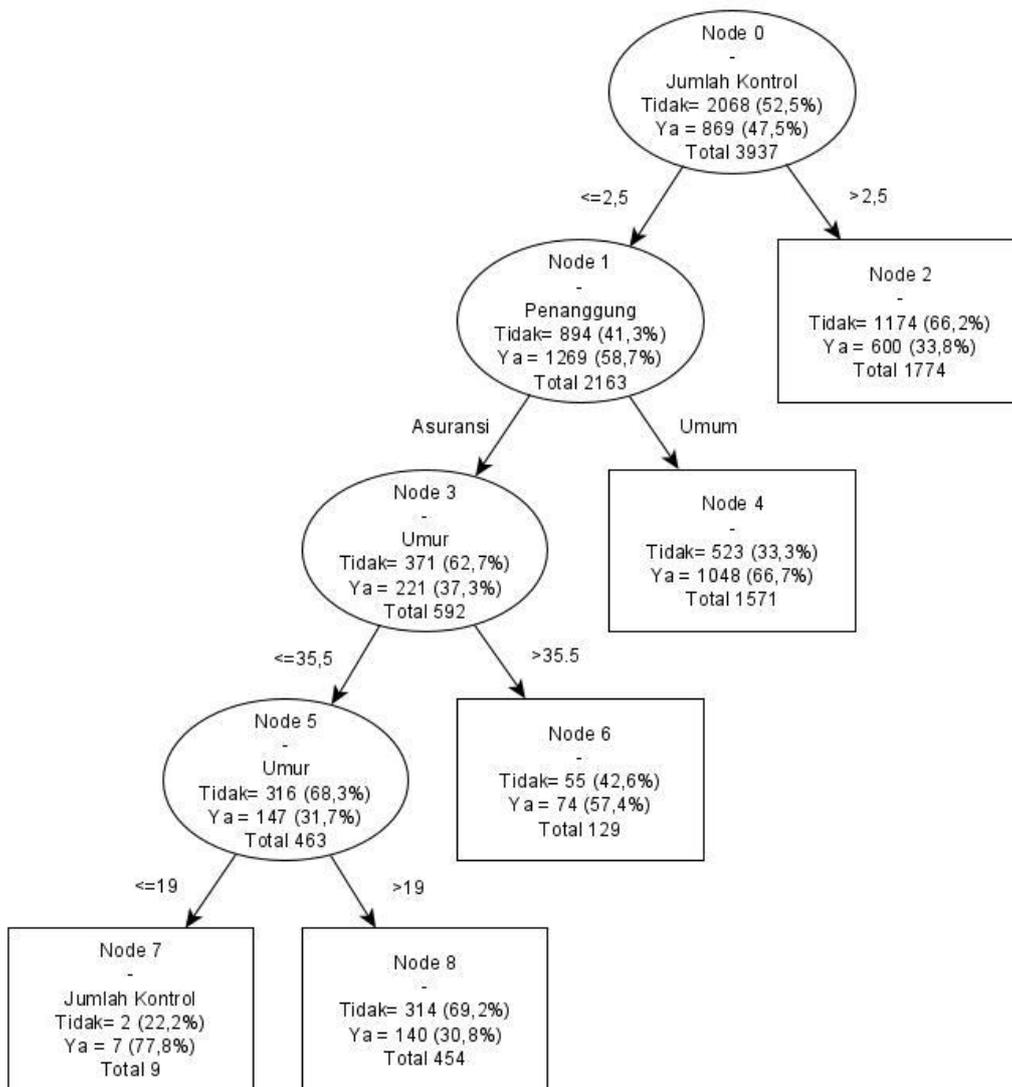
3.3.3.4 Pemilahan Ke-empat

Pemilah Optimal dipilih berdasarkan nilai goodness of split terbesar, perbandingan nilai goodness of split dapat dilihat pada tabel 3.15.

Tabel 3.15 Perbandingan *goodness of split* pemilahan ke-empat

Nama Variabel	Nama Threshold	Goodness of Split $\phi(s,t)$
Jumlah Kontrol	1,5	0,000174
Umur	20,5	0,0002616
	21,5	0,0000158
	22,5	0,0015677
	23,5	0,0020330
	24,5	0,0009711
	25,5	0,0000223
	26,5	0,0000692
	27,5	0,0000513
	28,5	0,0001937
	29,5	0,0002162
	30,5	0,0003944
	31,5	0,0000575
	32,5	0,0003695
	33,5	0,0003189
34,5	0,0001597	

Berdasarkan **tabel 3.15** maka yang akan menjadi Pemilah adalah threshold 23,5 dari variabel umur dengan nilai goodness of split sebesar 0,0020330. Pohon klasifikasi hasil pemilahan dapat dilihat pada **gambar 3.6**



Gambar 3.6 Klasifikasi pemilahan ke-empat

Kedua Simpul dalam pemilahan ini adalah noktah terminal. Karena data dalam cabang kiri berisi pengamatan dengan nilai $n \leq 5$ sedangkan pada simpul kanan apabila dilakukan pemilahan lagi

tidak terjadi penurunan keheterogenan yang signifikan. Dengan begitu maka penyusunan pohon keputusan dihentikan.

3.3.4 Penandaan Label Kelas

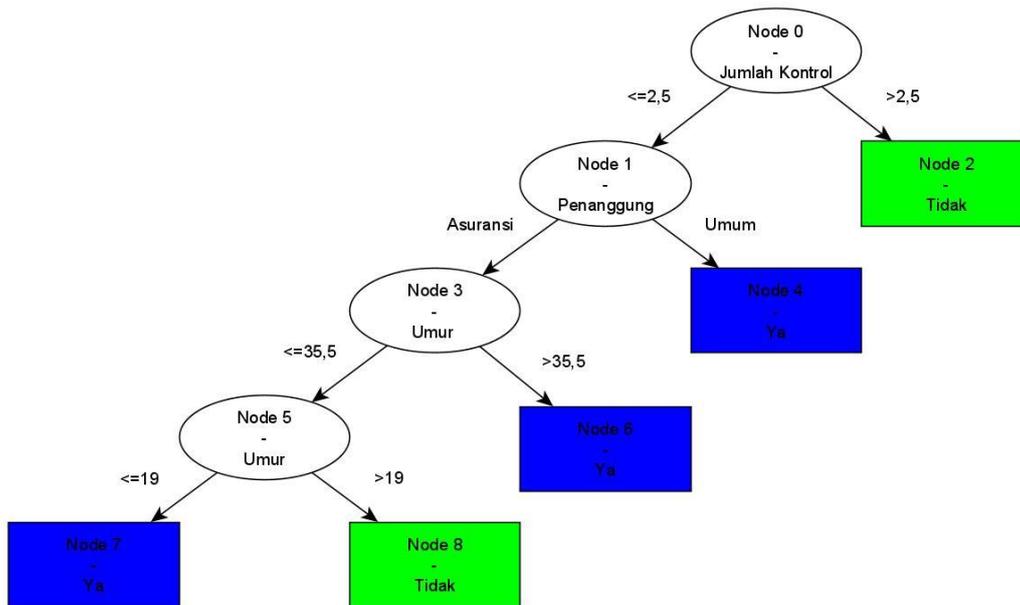
Langkah selanjutnya adalah pelabelan kelas pada setiap terminal nodes. Pemberian label kelas yaitu label kelas Ya untuk pasien yang churn dan label kelas tidak untuk pasien yang tidak churn, ditentukan berdasarkan persamaan (2.4) dengan mencari nilai peluang kelas $j=1$ yang maksimum diantara kedua simpul anak.

Tabel 3.16 Label kelas

<i>Nodes</i>	<i>Terminal Nodes</i>	$N(1 t)$	$N(t)$	$p(1 t)$	Nama Label Kelas
2	1	1174	1774	0,66	Tidak
4	2	523	1571	0,33	Ya
6	3	55	129	0,42	Ya
7	4	2	9	0,22	Ya
8	5	314	454	0,69	Tidak

Berdasarkan Tabel 2.13 diketahui bahwa hasil pemilahan pertama diperoleh nilai $p(j=1|t)$ untuk node 2 sebesar 0,66. Karena nilai Node 2 memiliki nilai 0,66, maka Node 2 diberi label kelas Tidak. Akan tetapi Node 1 belum menjadi terminal node dan akan dilakukan pemilahan selanjutnya. Begitu juga dengan node 4 yang memiliki nilai $p(j=1|t)$ sebesar 0,33, simpul ini dilabeli kelas Tidak. Sedangkan node 3 dipilah menjadi simpul keputusan dilanjutkan ke pemilahan selanjutnya. Node 5 bukan merupakan noktah terminal dan akan dilakukan pemilahan lagi, sedangkan node 6 merupakan noktah terminal dan diperoleh nilai $p(j=1|t)$ sebesar 0,42 dan diberi label kelas ya. Node 7 dan node 8 adalah terminal node, diperoleh nilai $p(j=1|t)$ untuk node 7 sebesar 0,22 dan dilabeli ya dan nilai $p(j=1|t)$ untuk node 8 sebesar 0,69 maka node 6 dilabeli tidak. Pada Gambar 2.5,

simpul terminal dengan label kelas tidak (pasien yang tidak churn) ditandai dengan simpul yang berwarna hijau dan label kelas ya (pasien yang churn) ditandai dengan simpul yang berwarna biru. Pohono klasifikasi yang telah diberi label kelas dapat dilihat pada **gambar 3.7**



Gambar 3.7 Pohon Klasifikasi Setelah Dilakukan Pelabelan Kelas

3.3.5 Rule

Berdasarkan pohon klasifikasi yang telah terbentuk, didapatkan 4 noktah keputusan, oleh karena itu maka akan ada 4 rule yang didapat dari pohon klasifikasi tersebut. Rule yang diperoleh dapat dilihat pada **tabel 3.17**

Tabel 3.17 Rule

No	Kondisi	Churn
1	Jika Jumlah Kontrol $>2,5$ Maka	Tidak
2	Jika Jumlah Kontrol $\leq 2,5$ dan Penanggung = Umum Maka	Ya
3	Jika Jumlah Kontrol $\leq 2,5$ dan Penanggung = Asuransi dan Umur $>35,5$ Maka	Ya

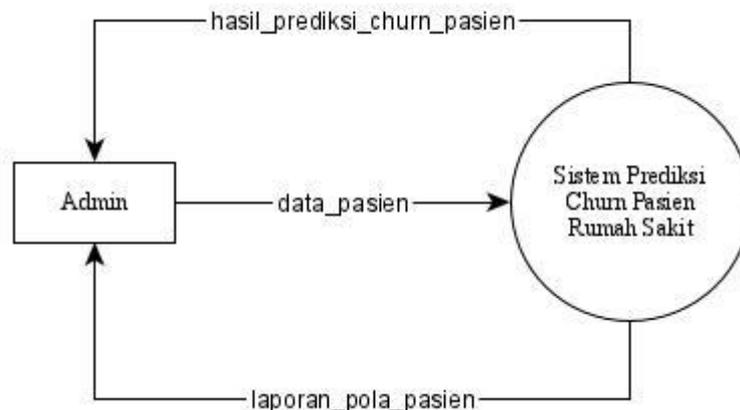
4	Jika Jumlah Kontrol $\leq 2,5$ dan Penanggung = Asuransi dan Umur $\leq 35,5$ dan Umur ≤ 19 Maka	Ya
5	Jika Jumlah Kontrol $\leq 2,5$ dan Penanggung = Asuransi dan Umur $\leq 35,5$ dan Umur > 19 Maka	Tidak

3.4 Perancangan Sistem

Perancangan sistem merupakan merancang atau men-desain sebuah sistem yang baik, dimana isinya adalah langkah-langkah operasi dalam proses pengolahan data dan prosedur untuk mendukung operasi sistem.

3.4.1 Context Diagram

Context diagram adalah sebuah diagram yang didalamnya berisi dokumen-dokumen suatu sistem dari beberapa level diagram. Diagram konteks ini terdiri dari *data flow diagram* yang berfungsi memetakan model lingkungan yang akan dipresentasikan dengan lingkaran tunggal yang mewakili seluruh sistem. Aliran dalam diagram konteks menggambarkan masukan dan keluaran sistem seperti pada **gambar 3.3**.



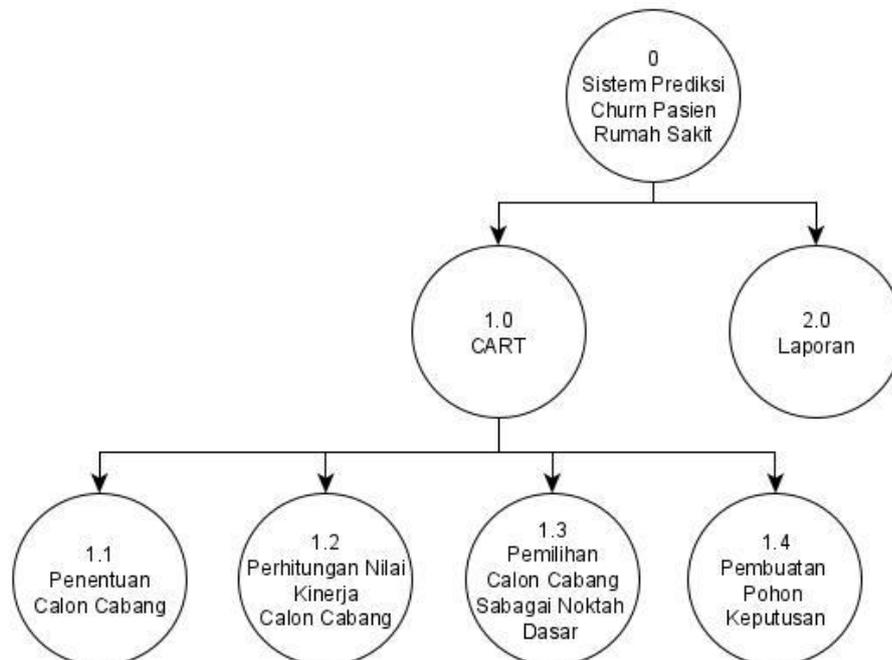
Gambar 3.8 *Context Diagram*

Keterangan pada **gambar 3.8** adalah sebagai berikut :

1. Admin memasukkan data pasien ke dalam sistem. Data yang dimasukkan merupakan data *training* dan data *testing*.
2. Dalam hal ini admin bertanggung jawab penuh atas kerja dari sistem prediksi churn pasien. Mulai dari menambahkan data, mengedit data dan menghapus data. Karena admin merupakan entitas utama.

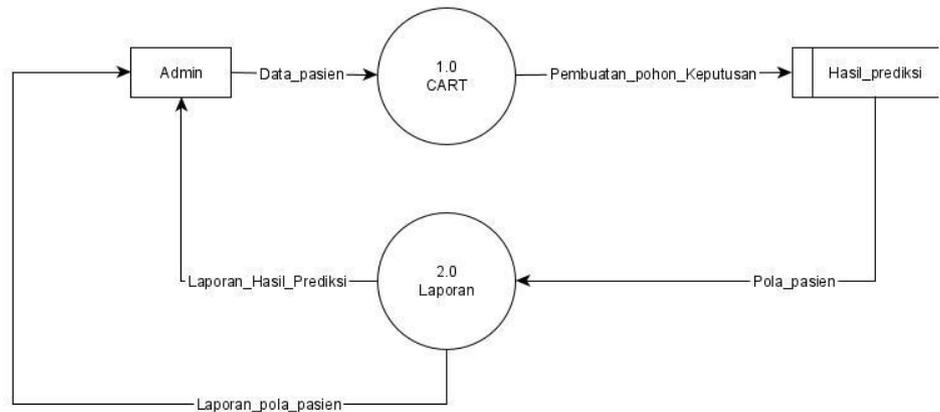
3.4.2 Diagram Berjenjang

Diagram berjenjang sangat diperlukan dalam perancangan semua proses yang ada. Diagram berjenjang merupakan penggunaan awal dalam menggambarkan *Data Flow Diagram* ke level-level lebih bawah lagi. Diagram berjenjang Sistem prediksi *churn* pasien dapat dilihat pada **gambar 3.10**



Gambar 3.9 Diagram Berjenjang

partitioning. Di bawah ini pada **gambar 3.10** dapat dilihat DFD Level 0 dari Sistem Prediksi Churn Pasien Rumah Sakit sebagai berikut :



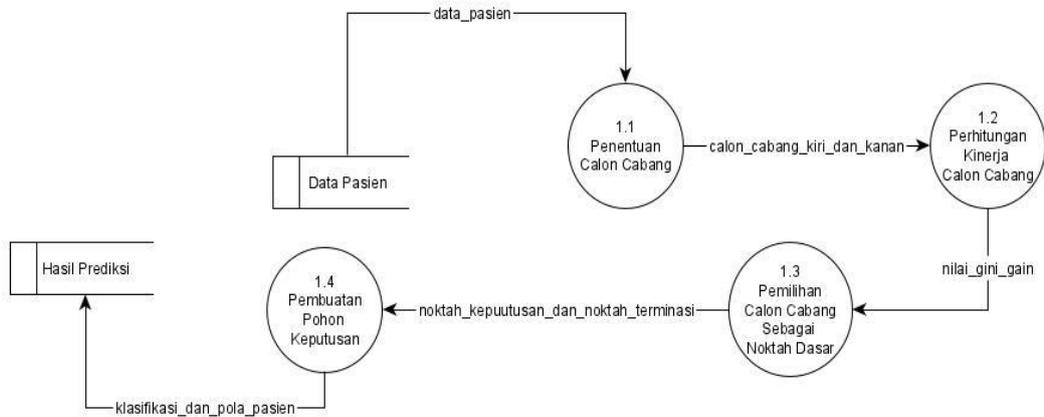
Gambar 3.10 DFD Level 0

Keterangan pada **gambar 3.10** adalah sebagai berikut :

DFD level 0 yang ditunjukkan pada gambar di atas menjelaskan beberapa proses yang terjadi pada sistem pendukung keputusan pemilihan asisten yang merupakan hasil dari *break down* dari *context diagram* awal untuk mendapatkan perilaku sistem yang lebih detail. Beberapa proses yang ada pada DFD level 0, antara lain:

1. CART : Perhitungan CART, merupakan proses mengolah data pasien menggunakan metode CART.
2. Laporan : Pembuatan laporan hasil prediksi dan laporan pola pasien menggunakan metode CART.

3.4.3.2 DFD Level 1



Gambar 3.11 DFD Level 1

DFD level 1 proses 1 yang ditunjukkan pada **gambar 3.6** menjelaskan beberapa proses yang terjadi pada sistem pendukung keputusan pemilihan asisten praktikan yang merupakan hasil *break down* dari DFD level 0 untuk mendapatkan perilaku sistem yang lebih detail. Beberapa proses yang ada pada DFD level 1 proses 1 antara lain :

1. Proses 1.1 : Menentukan calon cabang kiri dan calon cabang kanan.
2. Proses 1.2 : Menghitung kinerja calon cabang dengan cara menghitung *gini gain*.
3. Proses 1.3 : Memilih calon cabang dengan nilai *gini gain* tertinggi sebagai noktah dasar.
4. Proses 1.4 : Membuat pohon keputusan yang berisi noktah keputusan dan noktah terminasi.

3.5.1 Struktur Tabel

Struktur tabel merupakan susunan dari tabel yang akan digunakan atau diimplementasikan ke dalam basis data, di mana struktur tabel ini memuat detail data tipe tabel dan *primary key* serta *foreign key* dari tabel tersebut.

1. Tabel User

Tabel User digunakan untuk menyimpan data pengguna yang mengolah data sistem prediksi churn pasien.

Tabel 3.18 Struktur Tabel User

Nama Field	Tipe Data	Ukuran	Keterangan
Id_user	integer	15	<i>Primary Key</i>
Nama_user	Varchar	90	
Password	Varchar	90	

2. Tabel data_pasien

Tabel data pasien digunakan untuk menyimpan data pasien

Tabel 3.19 Struktur Tabel data_pasien

Nama Field	Tipe Data	Ukuran	Keterangan
Id_pasien	integer	15	<i>Primary Key</i>
Nama_pasien	Varchar	90	
Umur	integer	15	
J_kontrol	integer	15	
penanggung	varchar	90	
churn	varchar	90	
Jenis	integer	15	

3. Tabel Pemilah

Tabel Pemilah digunakan untuk menyimpan data pemilahan

Tabel 3.20 Struktur Tabel Pemilah

Nama Field	Tipe Data	Ukuran	Keterangan
Id_pemilah	integer	15	<i>Primary Key</i>
Nama_pemilah	Varchar	90	
Nama_cabang_kiri	Varchar	90	

Nama_cabang_kanan	Varchar	90	
gini	integer	15	

4. Tabel Node

Tabel Node digunakan untuk menyimpan data Node

Tabel 3.21 Struktur Tabel Node

Nama Field	Tipe Data	Ukuran	Keterangan
Id_node	integer	15	<i>Primary Key</i>
Nama_node	Varchar	90	
kedalaman	integer	15	
J_ya	integer	15	
J_tidak	integer	15	
Total	varchar	90	
Label	varchar	90	

5. Tabel Goodness

Tabel Goodnes digunakan untuk menyimpan data perhitungan Goodness of split.

Tabel 3.22 Struktur Tabel Goodness

Nama Field	Tipe Data	Ukuran	Keterangan
Id_treshold	integer	15	<i>Primary Key</i>
Nama_treshold	Varchar	90	
NI	Decimal	15	
NI_ya	Decimal	15	
NI_tidak	Decimal	15	
PI	Decimal	15	
PI_ya	Decimal	15	
PI_tidak	Decimal	15	

Nr	Decimal	15	
Nr_ya	Decimal	15	
Nr_tidak	Decimal	15	
Pr	Decimal	15	
Pr_ya	Decimal	15	
Pr_tidak	Decimal	15	
goodness	Decimal	15	
Id_pemilah	integer	15	
Id_node	integer	15	

6. Tabel Rule

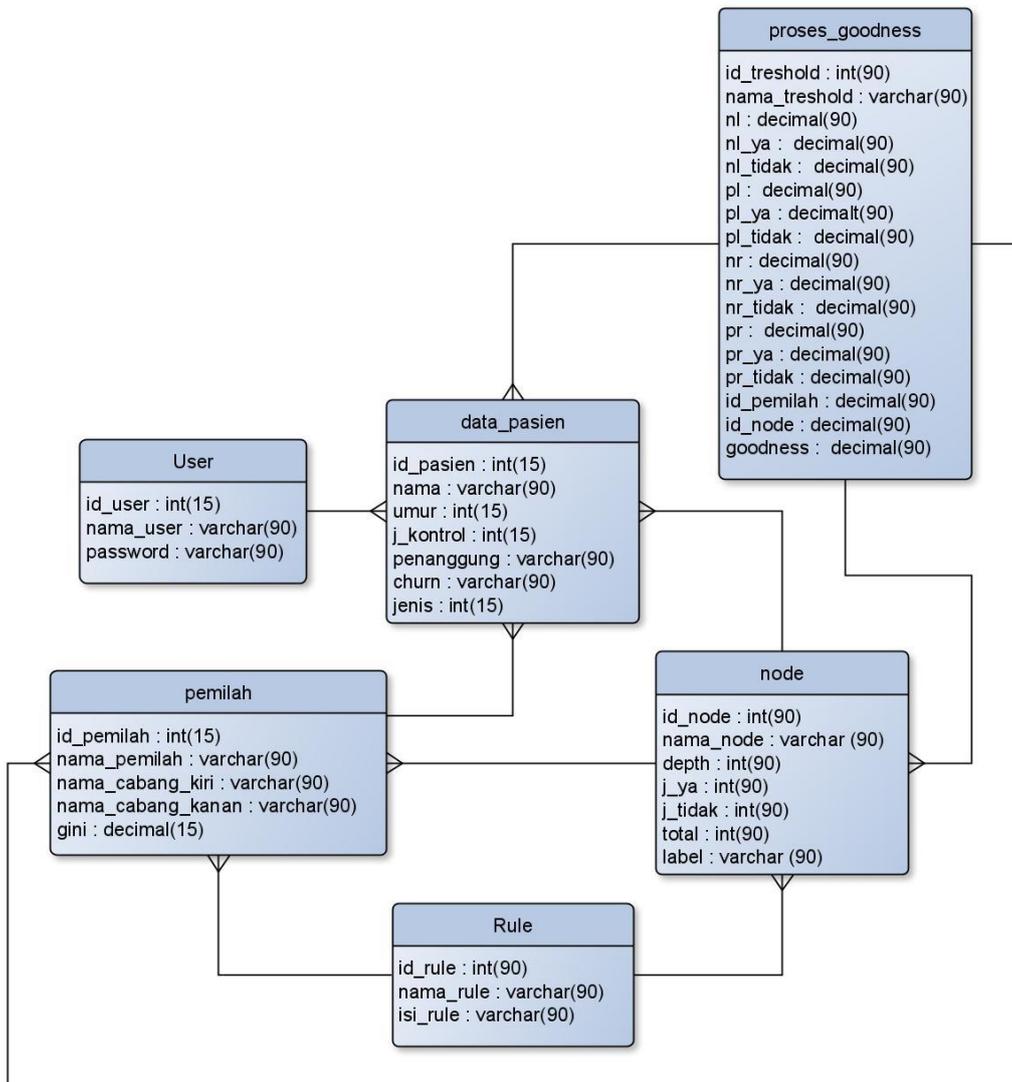
Tabel Rule digunakan untuk menyimpan data Rule.

Tabel 3.23 Struktur Tabel Rule

Nama Field	Tipe Data	Ukuran	Keterangan
Id_rule	integer	15	<i>Primary Key</i>
Nama_rule	Varchar	90	
Isi_rule	Varchar	90	

3.5.2 Entity Relationship Diagram (ERD)

Berikut adalah gambar *entity relationship diagram* (ERD) dari sistem prediksi churn pasien.



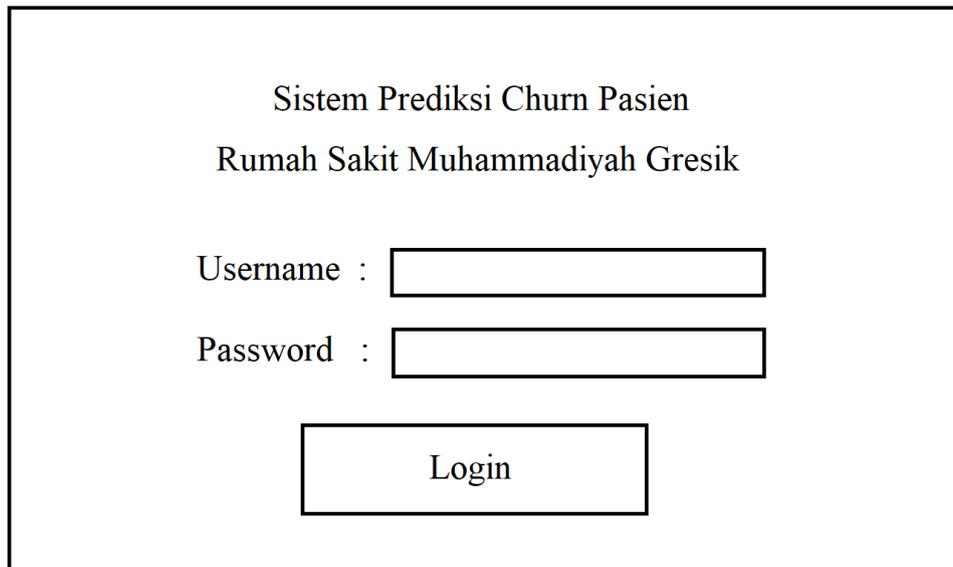
Gambar 3.12 Entity Relationship Diagram (ERD)

3.6 Perancangan Antarmuka

Merancang antarmuka merupakan bagian yang paling penting dari merancang sistem. Biasanya hal tersebut juga merupakan bagian yang paling sulit, karena dalam merancang antarmuka harus memenuhi tiga persyaratan: sebuah antarmuka harus sederhana, sebuah antarmuka harus lengkap, dan sebuah antarmuka harus memiliki kinerja yang cepat.

3.6.1 Antarmuka Halaman *Login*

Halaman *login* merupakan halaman awal pada sistem. Sebelum masuk ke halaman utama terlebih masukkan *username* dan *password* . Tampilan halaman login dapat dilihat pada **gambar 3.13**



Sistem Prediksi Churn Pasien
Rumah Sakit Muhammadiyah Gresik

Username :

Password :

Login

Gambar 3.13 Rancangan Halaman *Login*

3.6.2 Antarmuka Halaman Utama

Di halaman utama ini merupakan halaman pertama kali muncul ketika sistem dijalankan. Terdapat menu yang berisi halaman sitem prediksi churn pasien, diantaranya halaman prediksi, halaman data latih dan halaman rule. Rancangan tampilan halaman utama dapat dilihat pada **gambar 3.14**

	Sistem Prediksi Churn Pasien RSMG
Home	Selamat Datang User
Prediksi	
Data Latih	
Rule	

Gambar 3.14 Rancangan tampilan halaman utama

3.6.2 Antarmuka Halaman prediksi

Halaman prediksi digunakan untuk melakukan input data pasien yang akan diprediksi apakah pasien tersebut akan melakukan churn atau tidak. Rancangan tampilan halaman prediksi dapat dilihat pada **gambar 3.15**.

	Sistem Prediksi Churn Pasien RSMG
Home	Masukkan Data Pasien
Prediksi	Nama <input type="text"/>
Data Latih	Penanggung <input type="text"/>
Rule	Jumlah Kontrol <input type="text"/>
	Umur <input type="text"/>
	<input type="button" value="Prediksi"/>

Gambar 3.15 Rancangan tampilan halaman Prediksi

Setelah user melakukan input data pasien, maka selanjutnya akan ditampilkan hasil prediksi pasien tersebut. Rancangan tampilan halaman hasil prediksi dapat dilihat pada **gambar 3.16**.

	Sistem Prediksi Churn Pasien RSMG
Home	Hasil Prediksi :
Prediksi	
Data Latih	
Rule	

Gambar 3.16 Rancangan tampilan halaman Hasil Prediksi

3.6.3 Antarmuka Halaman Data latih

Halaman data latih digunakan untuk melakukan input data latih yang akan digunakan untuk pada sistem prediksi churn pasien. Rancangan tampilan halaman Data latih dapat dilihat pada **gambar 3.17**

	Sistem Prediksi Churn Pasien RSMG
Home	Masukkan Data Pasien
Prediksi	Nama <input type="text"/>
Data Latih	Penanggung <input type="text"/>
Rule	Jumlah Kontrol <input type="text"/>
	Umur <input type="text"/>
	<input type="button" value="Input Data"/>

Gambar 3.17 Rancangan tampilan halaman Data Latih

3.6.4 Antarmuka Halaman Rule

Halaman Rule digunakan untuk melihat rule yang diperoleh dari pengolahan data latih. Rancangan tampilan halaman Rule dapat dilihat pada **gambar 3.18**

	Sistem Prediksi Churn Pasien RSMG
Home	Rule atau Karakteristik Pasien :
Prediksi	1.
Data Latih	2.
	3.
Rule	4.

Gambar 3.18 Rancangan tampilan halaman Rule

3.3.6 Perancangan Pengujian

Untuk menilai tingkat akurasi, maka dilakukan pengetesan model dengan menggunakan Metode confusion matrix. Metode ini akan membandingkan hasil klasifikasi model dengan hasil klasifikasi yang sebenarnya, lalu dari kedua hasil akurasi tersebut akan dihitung rata-rata akurasi. Gambaran mengenai confusion matrix dapat dilihat pada **tabel 3.24**

Tabel 3.24 Confusion Matrix

	<i>Actual Values</i>	
<i>Predicted Values</i>	TP	FP
	FN	FI

Keterangan **Tabel 3.24** Adalah sebagai berikut:

1. *TP(True Positive)* : Merupakan Data positif yang diprediksi Benar.
2. *FP(False Positive)* : Merupakan Data positif yang diprediksi negatif.
3. *FN(False Negative)* : Merupakan Data negatif yang diprediksi Positif.
4. *TN(True Negative)* : Merupakan Data negatif yang diprediksi Benar.