

DECISION TREE BERBASIS ALGORITMA UNTUK PENGAMBILAN KEPUTUSAN

Zulfian Azmi^{#1}, Muhammad Dahria^{#2}

^{#1} Program Studi Sistem Komputer, ^{#2} Program Studi Sistem Informasi

STMIK Triguna Dharma - Jl. A.H. Nasution No. 73 F - Medan

E-mail : ^{#1}zulfianazmi@gmail.com

Abstrak

Decision Tree merupakan representasi sederhana dari teknik klasifikasi yang merupakan proses pembelajaran suatu fungsi tujuan yang memetakan tiap himpunan atribut ke satu dari kelas yang didefinisikan sebelumnya. Pohon keputusan dapat menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Dan pohon keputusan dapat memadukan antara eksplorasi data dan pemodelan, sehingga sangat baik sebagai langkah awal dalam proses pemodelan. Dengan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple. Dan pengambilan keputusan merupakan suatu proses pemikiran dalam rangka pemecahan suatu masalah untuk memperoleh hasil akhir untuk dilaksanakan.

Kata Kunci: *Data Mining, Decision Tree, Pengambilan Keputusan.*

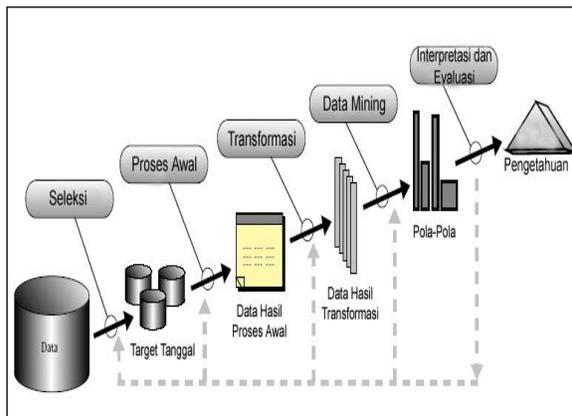
Abstract

Decision Tree technique is a simple representation of the classification is the process of learning an objective function that maps each set of attributes to one of the previously defined classes. Decision tree can find hidden relationships between a number of potential input variables to a target variable. And the decision tree can combine data exploration and modeling, so it is good as a first step in the modeling process. With its ability to break down a complex decision making process to be more simple. And decision-making is a process of thinking in a problem solving framework for obtaining the final result to be implemented.

Keywords: *Data Mining, Decision Tree, Decision Making.*

PENDAHULUAN

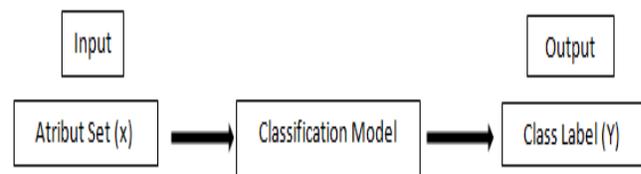
Data Mining merupakan proses yang diulang dan memerlukan interaksi manusia dalam prosesnya untuk menemukan pola atau model baru yang dapat digeneralisasi untuk masa yang akan datang, dan bermanfaat jika digunakan untuk melakukan suatu tindakan. Data mining berisi pencarian pola yang diinginkan dalam database besar untuk membantu dalam pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang kemudian dapat dipelajari dengan teliti dan bermanfaat. Dan salah satu teknik dari bagian data mining adalah klasifikasi, yaitu menentukan sebuah record data baru ke salah satu dari beberapa kategori atau klas yang telah didefinisikan sebelumnya. Dan salah satu bagian teknik dari klasifikasi adalah pohon keputusan.



Gambar 1. Tahapan Proses Knowledge Discovery in Databases

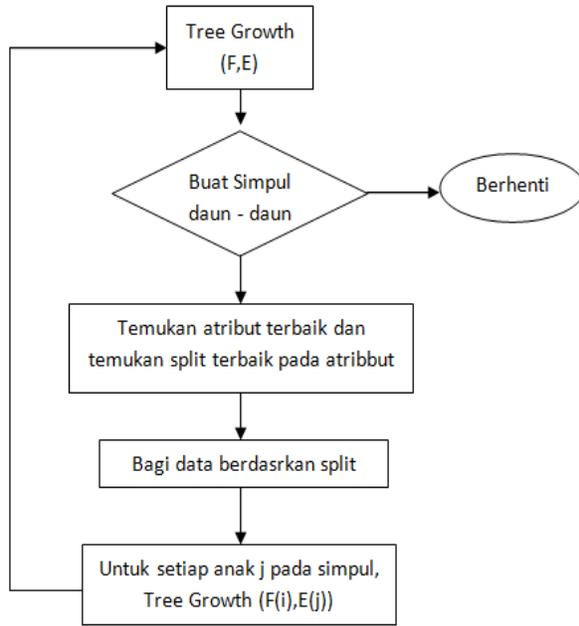
Pohon Keputusan merupakan representasi sederhana dari teknik klasifikasi yang merupakan proses pembelajaran suatu fungsi tujuan yang

memetakan tiap himpunan atribut ke satu dari kelas yang didefinisikan sebelumnya. Pohon keputusan merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Dengan kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simple. Pohon keputusan juga dapat menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Selain itu pohon keputusan dapat memadukan antara eksplorasi data dan pemodelan, sehingga sangat baik sebagai langkah awal dalam proses pemodelan.



Gambar 2. Blok Diagram Model Klasifikasi

Dengan menggunakan metode pohon keputusan dapat menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan. Dan kekurangan dari pohon keputusan diantaranya adalah terjadi *overlap* terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak dan kualitas hasil keputusan yang didapatkan sangat tergantung pada bagaimana pohon tersebut didesain.



Gambar 3. Algoritma Induksi Pohon Keputusan

Pengambilan Keputusan adalah suatu proses pemikiran dalam rangka pemecahan suatu masalah untuk memperoleh hasil akhir untuk dilaksanakan. Kesalahan dalam pengambilan keputusan bisa berdampak kepada kerugian. Misal dalam perusahaan, keputusan yang diambil oleh pimpinan perusahaan merupakan hasil pemikiran yang harus dilaksanakan oleh bawahannya atau mereka yang harus dilaksanakan oleh bawahannya atau mereka yang bersangkutan dengan organisasi yang dia pimpin.

PEMBAHASAN

Salah satu algoritma yang digunakan untuk membangun pohon keputusan yang berbasis algoritma induksi pohon keputusan yaitu C4.5. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan *Algoritma Decision*.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam Rumus 1 (Craw, S., ---).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Dengan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, Larose yaitu :

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data *histori* yang pernah terjadisebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai

entropy. Untuk menghitung nilai *entropy* digunakan rumus :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Dengan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

Misal kita ambil sebuah kasus menentukan data UKM yang terdiri dari 15 UKM. Data yang akan digunakan dalam bentuk pohon keputusan untuk menganalisa UKM survive. Data UKM tersebut selanjutnya akan dilakukan pra proses untuk menghasilkan data khusus yang siap untuk dibentuk menjadi sebuah pohon keputusan.

Data UKM tersebut terdapat pada table 1 berikut, yaitu:

NO	NAMA	JLH. PINJAMAN	REALISASI	AWAL ANGS	J.TEMPO	KUM.	HRS. NYA	Hutang	SISA
1	UD.A	50,000,000	27/12/02	10/02/03	10/01/05	3	24	21	21
2	UD.B	10,000,000	27/12/02	10/02/03	10/01/05	17	24	7	7
3	UD.C	10,000,000	27/12/02	10/02/03	10/01/05	5	24	19	19
4	UD.D	20,000,000	27/12/02	10/02/03	10/01/05	23	24	1	1
5	UD.E	100,000,000	27/12/02	10/02/03	10/01/05	2	24	22	22
6	UD.F	40,000,000	27/12/02	10/02/03	10/01/05	5	24	19	19
7	UD.G	10,000,000	01/02/03	10/04/03	10/03/05	8	24	16	16
8	UD.F	50,000,000	31/12/07	31/01/08	31/12/09	19	24	5	5
9	UD.H	40,000,000	01/12/07	10/09/03	10/09/05	20	24	4	4
10	UD.I	15,000,000	01/12/07	10/11/03	10/10/05	0	24	24	244

Data pada table 1 sangat banyak dan tidak lengkap dan *inkonsisten* umumnya terjadi pada setiap database. Data yang tidak lengkap disebabkan karena adanya data yang kosong atau atribut yang

salah, sehingga proses data *preprocessing* perlu dilakukan sehingga *database* sesuai dengan ketentuan yang diperlukan.

Data *preprocessing* merupakan hal yang penting dalam proses data mining, hal yang termasuk antara lain

B.1. Data Selection

Data UKM tersebut nantinya akan menjadi kasus dalam proses operasional *data mining*. Dari data yang ada, kolom yang diambil sebagai atribut/*variable* keputusan adalah *survive*.

sedangkan kolom yang diambil variabel penentuan dalam pembentukan pohon keputusan adalah:

1. Nama UKM
2. Manajemen
 - Ditentukan dari angsuran komulatif UKM dalam periode tersebut, yaitu dengan devenisi:
 - a. Manajemen baik jika jumlah angsuran komulatifnya 18 s/d 24 bulan
 - b. Manajemen sedang jika jumlah angsuran komulatifnya 11 s/d 17 bulan
 - c. Manajemen buruk jika jumlah angsuran komulatifnya 0 s/d 10 bulan
3. Income (pendapatan)
 - Ditentukan dari jumlah pinjaman, karena melalui jumlah pinjaman dapat diketahui besar kecilnya income (pendapatan) suatu UKM
4. Hutang
 - Dapat diketahui dari besar kecilnya tunggakan yang dimiliki UKM

B.2. Data Preprocessing/Data Cleaning

Data *cleaning* diterapkan untuk menambah isi atribut yang hilang atau kosong, dan merubah data yang tidak konsisten

1. Data Transformasi

Dalam proses ini, data ditranspormasikan ke dalam bentuk yang sesuai untuk proses *data* mining.

2. Data Reduction

Reduksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan sehingga ukuran dari database menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses *data mining*, karena akan lebih efisien terhadap data yang lebih kecil.

Masalah klasifikasi berakhir dengan dihasilkan sebuah pengetahuan yang dipresentasikan dalam bentuk diagram yang biasa disebut pohon keputusan (*decision tree*) untuk menentukan UKM *survive*, kriteria yang diperhatikan adalah mampunya suatu UKM dalam melunasi angsuran .

Reduksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan sehingga ukuran dari database menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses *data mining*, karena akan lebih efisien terhadap data yang lebih kecil.

Masalah klasifikasi berakhir dengan dihasilkan sebuah pengetahuan yang dipresentasikan dalam bentuk diagram pohon keputusan (*decision tree*) untuk menentukan UKM *survive*, kriteria yang diperhatikan adalah mampunya suatu UKM dalam melunasi angsuran. Berikut ini data UKM yang dipergunakan untuk menentukan *survive* tidaknya suatu UKM, data selengkapnya tampak pada tabel 2 berikut ini :

Tabel 2. Tabel data hasil survey

NO	NAMA UKM	MANAJEMEN	INCOME	HUTANG	SURVIVE
1	UD.A	BURUK	SEDANG	TINGGI	TIDAK
2	UD.B	SEDANG	RENDAH	RENDAH	TIDAK
3	UD.C	BURUK	RENDAH	TINGGI	TIDAK
4	UD.D	BAIK	RENDAH	RENDAH	TIDAK
5	UD.E	BURUK	RENDAH	RENDAH	TIDAK
6	UD.F	BURUK	RENDAH	RENDAH	TIDAK
7	UD.G	BAIK	TINGGI	RENDAH	YA
8	UD.H	BURUK	RENDAH	RENDAH	TIDAK
9	UD.I	BAIK	TINGGI	RENDAH	YA
10	UD.J	BURUK	SEDANG	TINGGI	TIDAK

B.3. Mengubah Data Menjadi Tree

Dalam mengubah data menjadi tree terlebih dahulu data dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*.

B.4. Menentukan Node Terpilih / Menentukan Nilai Atribut

Dalam data sampel tentukan dulu *node* terpilih, yaitu dengan menghitung nilai informasi gain masing-masing atribut untuk menentukan node terpilih, gunakan nilai informasi *gain* yang paling besar. dengan menggunakan persamaan $-P_{(+)} \log_2 P_{(+)} - P_{(-)} \log_2 P_{(-)}$ Dapat dihitung nilai (i) dari seluruh data training:

$$I = -(11/10) * \text{LOG}((11/10), 2) - (89/10) * \text{LOG}((89/10), 2) = 0,2173$$

Tabel 3. Tabel nilai atribut

Solusi	J.(+)	JR	J.(-)
YA	-2	10	-0,2
	2	10	0,2
TIDAK	-8	10	-0,8
	8	10	0,8

1. Menghitung Nilai Atribut Manajemen

Adapun untuk menghitung nilai informasi dari atribut manajemen adalah sebagai berikut :

Tabel 4. Tabel nilai atribut manajemen

Hutang	SURVIVE	jumlah
Tinggi	Ya	0
Tinggi	Tidak	2
Rendah	Ya	2
Rendah	Tidak	5
Sedang	Tidak	1
TINGGI	Ya	2
TINGGI	Tidak	0

Tabel 5. Parameter dari atribut manajemen

MANAJMN	Paramter
Baik	q1
SEDANG	q2
BURUK	q3

$$q1 = -(2/3)*\text{LOG}(2/3)-(1/3)*\text{LOG}(1/3) \quad \mathbf{0.042}$$

$$q2 = -(0/1)*\text{LOG}(0/1)-(1/1)*\text{LOG}(1/1) \quad \mathbf{0}$$

$$q3 = -(0/6)*\text{LOG}(0/6)-(6/6)*\text{LOG}(6/6) \quad \mathbf{0}$$

2. Menghitung Nilai Informasi Atribut Income (Pendapatan)

Adapun untuk menghitung nilai informasi income (pendapatan) adalah sebagai berikut :

Tabel 6. Nilai informasi atribut income (pendapatan)

INCOME	SURVIVE	Jumlah
Rendah	Ya	0
Rendah	Tidak	6
Sedang	Ya	0

Hutang	SURVIVE	jumlah
Tinggi	Ya	0
Tinggi	Tidak	2
Rendah	Ya	2
Rendah	Tidak	5
Sedang	Tidak	1
TINGGI	Ya	2
TINGGI	Tidak	0

$$q1 = -(0/6)*\text{LOG}((0/6)-(6/6)*\text{LOG}(6/6) \quad \mathbf{0}$$

$$q2 = -(0/1)*\text{LOG}(0/1)-(1/1)*\text{LOG}(1/1) \quad \mathbf{0}$$

$$q3 = -(2/2)*\text{LOG}(2/2,2)-(0/2)*\text{LOG}(0/2) \quad \mathbf{0}$$

3. Selanjutnya Menghitung Nilai Informasi Atribut Hutang

Adapun untuk menghitung nilai informasi dari atribut hutang adalah sebagai berikut:

$$q1 = -(0/2)*\text{LOG}((0/2)-(2/2)*\text{LOG}2/2) \quad \mathbf{0}$$

$$q2 = -(2/7)*\text{LOG}(2/7)-(5/7)*\text{LOG}(5/7) \quad \mathbf{0,0510228}$$

B.5. Menghitung Nilai Entropy Tiap Atribut

Selanjutnya menghitung nilai entropy tiap atribut, baik *entropy* atribut manajemen, atribut hutang.

Selanjutnya Menghitung Nilai Informasi Gain Atribut

1. Menghitung Nilai Informasi *Gain* Atribut Manajemen

Adapun menghitung nilai informasi *gain* dari atribut manajemen adalah sebagai berikut:

$$Gain(\text{Manajemen}) = \text{Nilai Informasi Seluruh Data Training} - Entropy \text{ Atribut Manajemen}$$

2. Menghitung nilai informasi *gain* atribut *income* (pendapatan)

Adapun menghitung nilai informasi *gain* dari atribut *income* (pendapatan) adalah sebagai berikut:

$$Gain(\text{Income}) = \text{Nilai Informasi Seluruh Data Training} - Entropy \text{ Atribut Income}$$

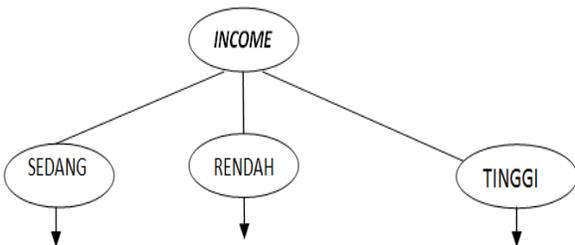
3. Menghitung nilai informasi *gain* atribut hutang

Adapun menghitung nilai informasi *gain* dari atribut hutang adalah sebagai berikut:

$$Gain(\text{Hutang}) = \text{Nilai Informasi Seluruh Data Training} - \text{Entropy Atribut Hutang}$$

B.6. Menyusun *Tree* Awal

Setelah nilai informasi selesai dihitung, maka ambilah nilai informasi *gain* terbesar dari beberapa atribut, kemudian dijadikan sebagai *node* awal.



Gambar 4. Pohon Note awal

Node berikutnya dapat dipilih pada bagian yang mempunyai nilai + dan -, pada kasus diatas hanya *income* = sedang dan tinggi yang memiliki nilai + dan nilai -, Maka semua pasti mempunyai internal *node*. Untuk menyusun *internal node*, ini dilakukan satu persatu.

B.7. Penentuan *Internal Node* Untuk *income*, tinggi dan sedang

Tabel 7. Penentuan *Internal Node*

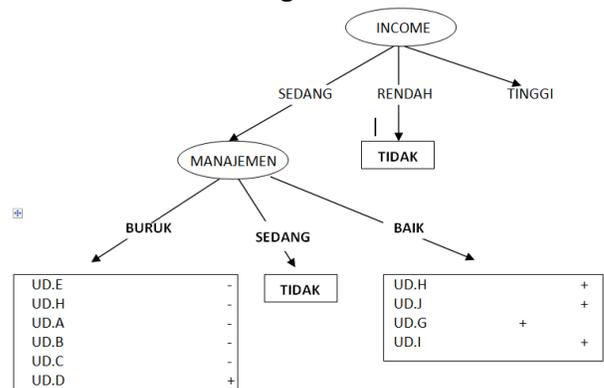
NAMA UKM	MANAJEMEN	HUTAN	SURVIVE
UD.E	BURUK	TINGGI	TIDAK
UD.H	BAIK	RENDAH	YA
UD.A	BURUK	TINGGI	TIDAK
UD.B	BURUK	RENDAH	YA
UD.C	BURUK	RENDAH	TIDAK
UD.D	BURUK	TINGGI	TIDAK
UD.F	BURUK	RENDAH	TIDAK
UD.I	BAIK	RENDAH	YA
UD.J	BAIK	RENDAH	YA
UD.G	BURUK	TINGGI	TIDAK

Untuk menentukan *node* untuk *income* sedang dan tinggi dapat digunakan persamaan: $-P_{(+)} \log_2 P_{(+)} - P_{(-)} \log_2 P_{(-)}$ dapat dihitung nilai informasi (*i*) dari seluruh data training :

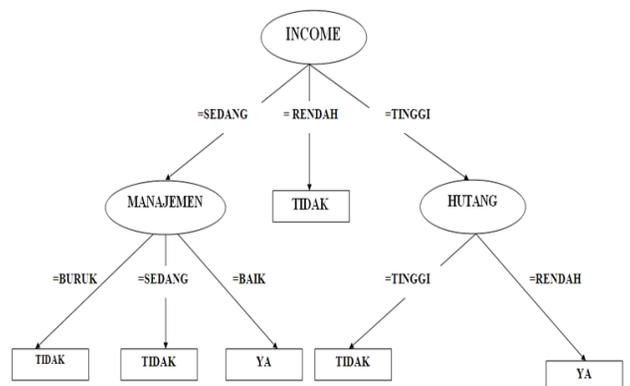
Kemudian setelah mendapat hasil nilai informasi (*i*) untuk *income*, selanjutnya diperlukan perhitungan untuk atribut yang tersisa yaitu atribut manajemen dan atribut hutang.

B.8. Menyusun *Tree* Lanjutan

dari tabel 3.10 dapat diketahui bahwa atribut dan *entropy* rata-rata terkecil adalah atribut manajemen dengan nilai *gain* -2.49, sehingga atribut *survive* di pilih sebagai *node* selanjutnya atau *node* kedua. Dengan demikian nilai dapat menjadi *node* cabang dari nilai atribut hutang = <= 0.60.



Gambar 5. Pohon Keputusan Hasil Perhitungan *Node* Tabel 3.10



Gambar 6 Pohon Keputusan Hasil *Tree* Terakhir

Dengan memperhatikan pohon keputusan pada gambar 3.5, diketahui bahwa semua kasus sudah masuk dalam kelas. Dengan demikian, pohon keputusan pada gambar 3.5 merupakan pohon keputusan terakhir yang terbentuk.

B.10. Mengubah Tree Menjadi Rule

Dengan memperhatikan pohon keputusan pada gambar 3.5 diketahui bahwa pohon keputusan telah terbentuk. Dan setelah didapatkan *tree* akhirnya kemudian diubah menjadi *rule*. Berikut ini adalah bentuk *tree* yang diubah menjadi *rule*:

- R1 :if *income* = rendah THEN *Survive* = tidak
- R2 :if *income* = Sedang ^ manajemen =buruk ^ THEN *Survive* = Tidak
- R3 :if *income* = sedang ^ manajemen = sedang ^ THEN *Survive* = tidak
- R4 :if *income* = sedang ^ manajemen = baik ^ THEN *Survive* = ya
- R5 :if *income* = tinggi ^ hutang = tinggi ^ THEN *Survive* = tidak
- R6 :if *income* = tinggi ^ hutang = rendah ^ THEN *Survive* = ya

Berikut ini adalah bentuk keterangan umumnya dari *rule* yang sudah disederhanakan adalah sebagai berikut:

- R1= Jika *income* = sedang dan manajemen = buruk maka *survive* tidak
- R2 =Jika *income* = sedang dan manajemen = sedang maka *survive* tidak
- R3 = jika *income* = sedang dan manajemen = baik maka *survive* ya
- R4 = jika *income* rendah maka *survive* tidak
- R5 = jika *income* = tinggi dan hutang = tinggi maka *survive* tidak
- R6 = jika *income* = tinggi dan hutang = rendah maka *survive* ya

SIMPULAN

Pohon keputusan merupakan sistem yang dikembangkan untuk membantu mencari dan membuat keputusan untuk suatu permasalahan dan dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah melihat mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan pohon keputusan ini sebagai alat bantu dalam mengambil keputusan yang telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan pohon keputusan yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

D. DAFTAR PUSTAKA

- Hermawati. 2009. Data Mining. ANDI <http://www.ilmukomputer.com>
- Kusrini. 2006. Sistem Pakar Teori dan Aplikasi. Yogyakarta: Andi Offset.
- Pramudiono, Iko. Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data.
- Tata Sutabri, S, Kom., MKom. 2003. Sistem Informasi Manajemen :ANDI.