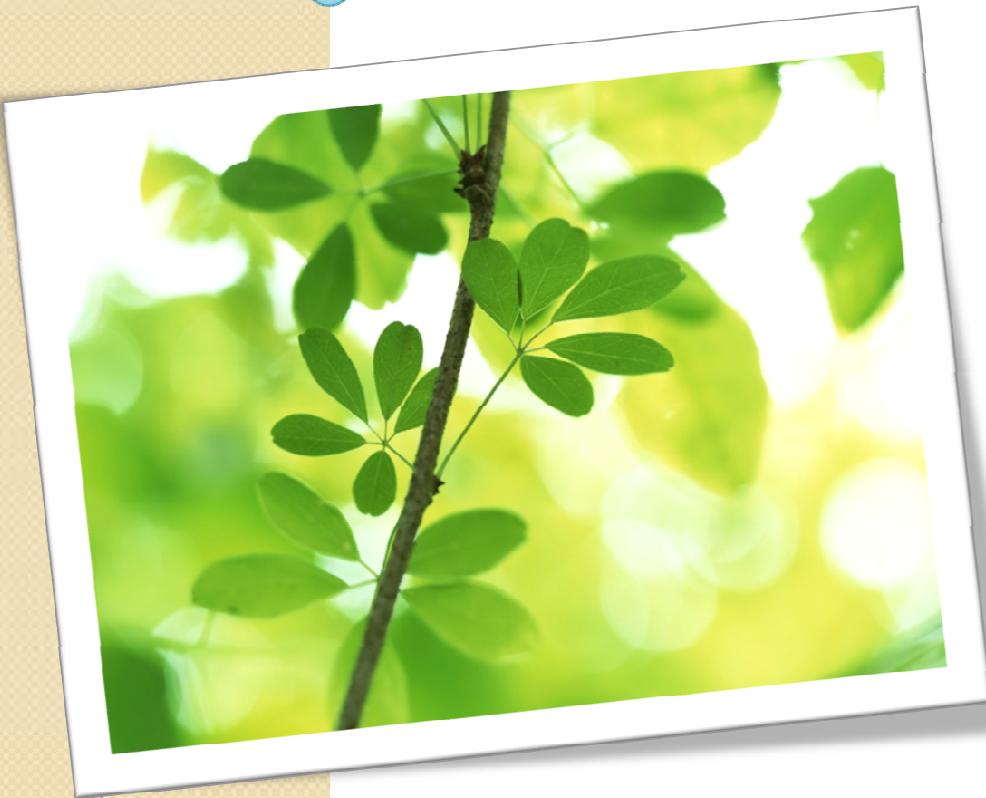


DATA MINING
METODE KLASIFIKASI

DECISION TREE (POHON KEPUTUSAN)





Pendahuluan

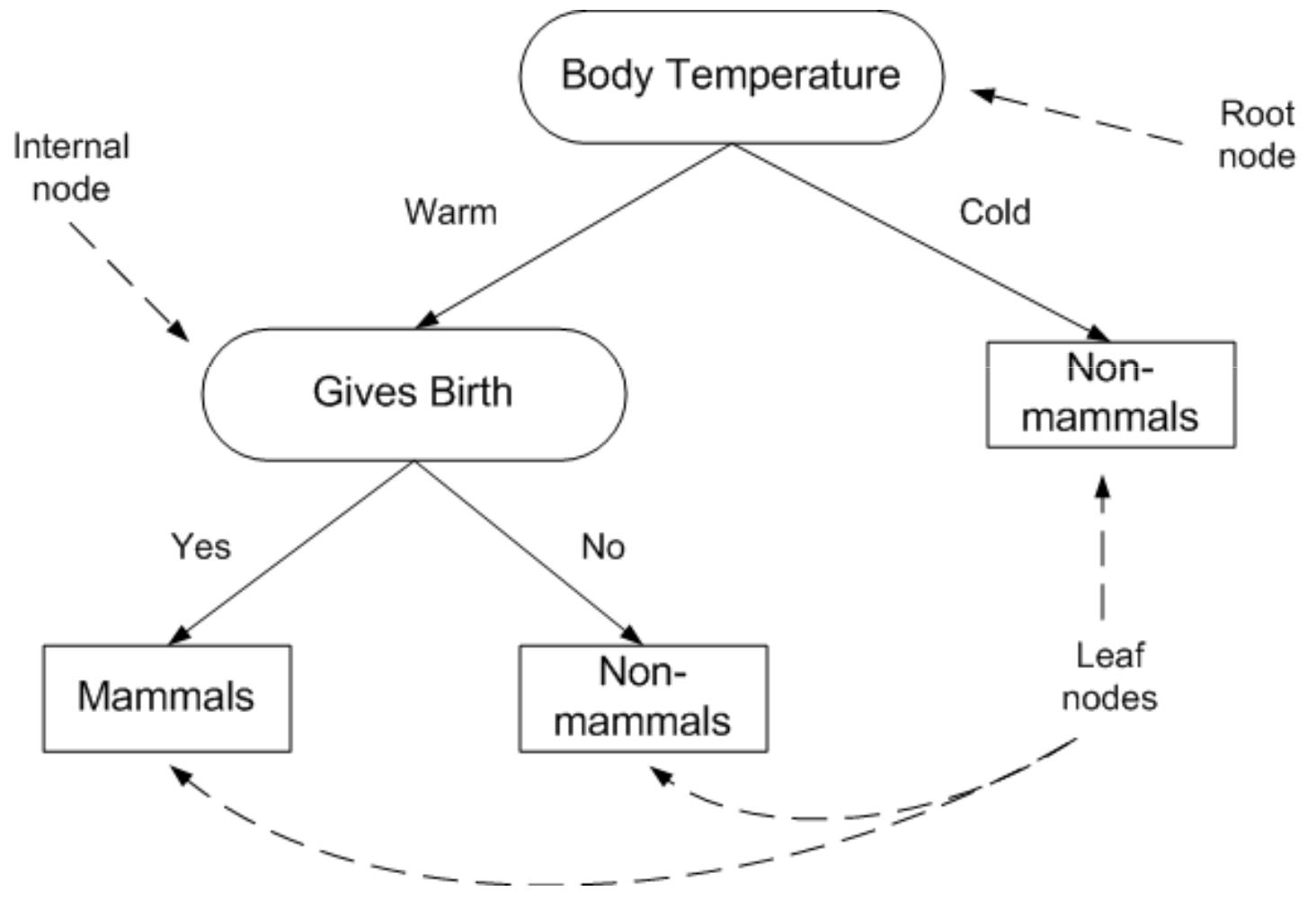
- Pohon (Tree) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*).
- Simpul pada sebuah pohon terdiri dari 3:
 - Simpul akar (*root node*)
 - Simpul percabangan/internal (*branch/internal node*)
 - Simpul daun (*leaf node*)



Pendahuluan

- Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga
 - Simpul internal maupun simpul akar ditandai dengan nama atribut,
 - Rusuk diberi label nilai atribut yang mungkin
 - Simpul daun ditandai dengan kelas-kelas yang berbeda

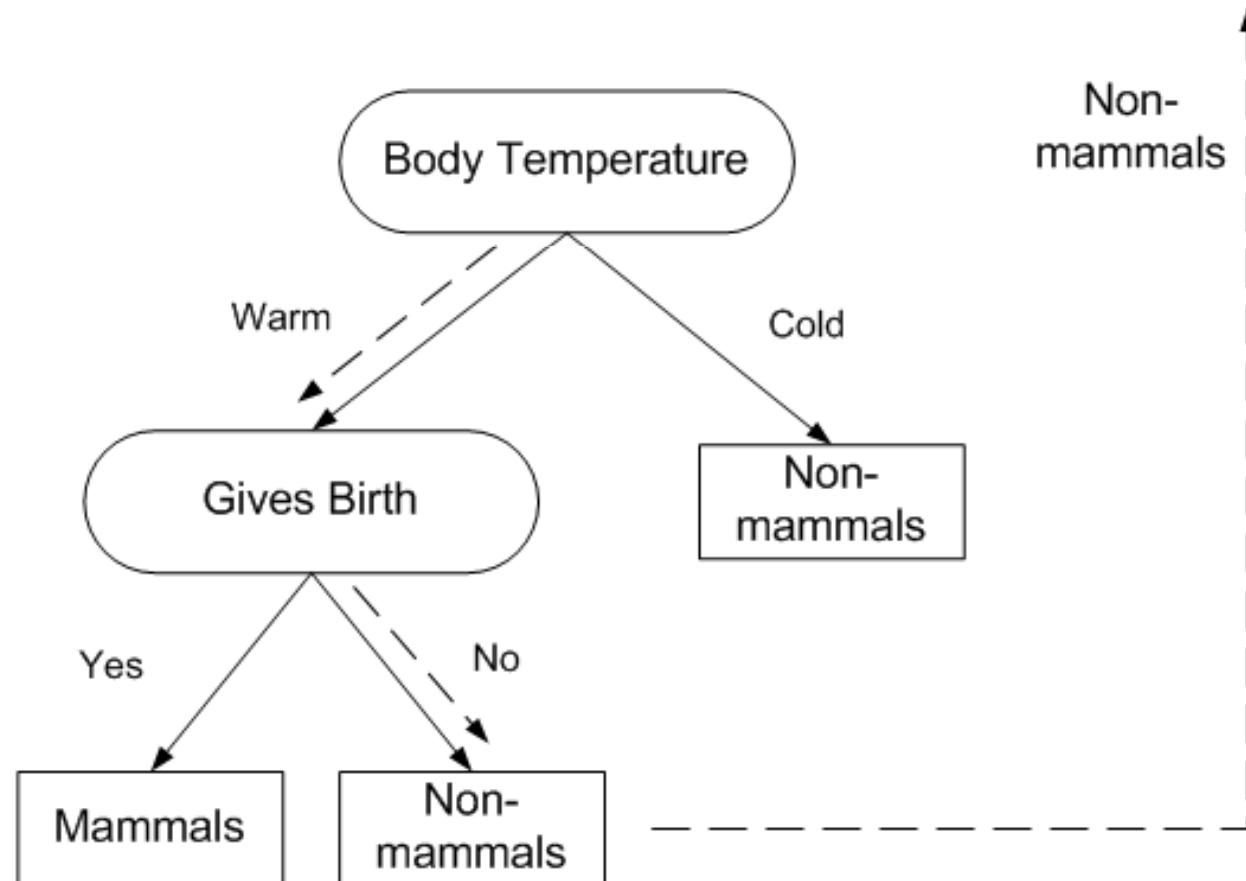
Contoh Klasifikasi Mamalia



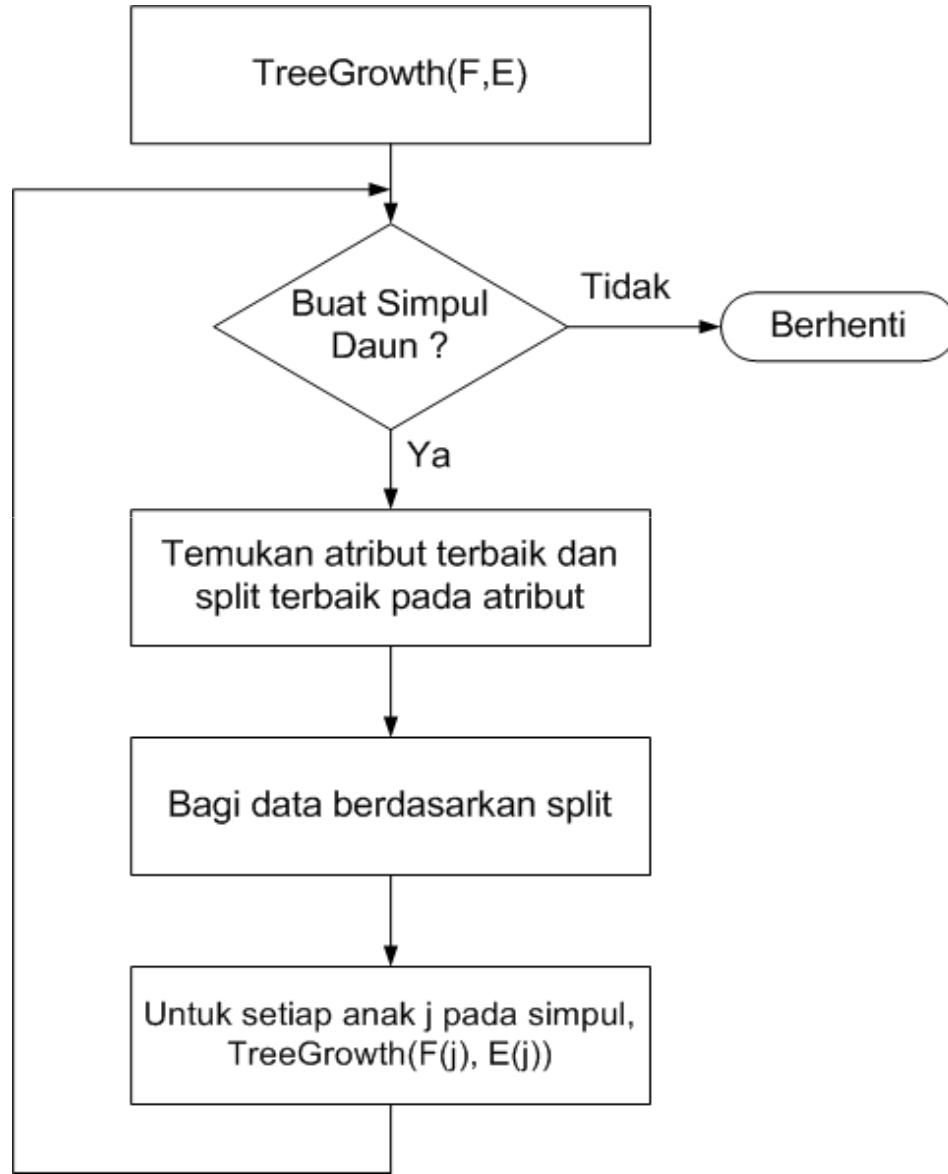
- Objek (*record*) diklasifikasikan dengan mengikuti suatu jalur (*path*) yang dimulai dari simpul akar sesuai dengan nilai atribut dalam *record* tersebut.

Contoh: Penelusuran Pohon Keputusan

Unlabeled data	Name	Body Temperature	Gives Birth	...	Class
	Flamingo	Warm	No	...	?



Algoritma Induksi Pohon Keputusan





Algoritma Hunt

- Algoritma Hunt merupakan salah satu algoritma yang digunakan untuk membangun pohon keputusan yang berbasis algoritma induksi pohon keputusan seperti ID3, C4.5 dan CART.



Prosedur Algoritma Hunt

Misalkan D_t adalah himpunan record pembelajaran yang mencapai suatu simpul t dan $y=\{y_1, y_2, \dots, y_c\}$ adalah label kelas.

1. Jika semua record dalam D_t menempati kelas y_t yang sama maka t adalah simpul daun yang diberi label y_c
2. Jika D_t adalah himpunan kosong maka t merupakan sebuah simpul daun yang diberi label kelas default, y_d



Prosedur Algoritma Hunt

3. Jika D_t terdiri dari record yang menempati lebih dari satu kelas, gunakan sebuah atribut tes untuk membagi data ke dalam himpunan bagian yang lebih kecil.
4. Secara rekursif terapkan prosedur ini untuk tiap himpunan bagian

Contoh Algoritma Hunt

Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Proses Algoritma Hunt

Simpul $D_t = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
dikategorikan dalam 2 kelas $y = \{\text{yes}, \text{no}\}$

Cek apakah semua record dalam D_t terletak pada satu kelas

D_t tidak terletak pada satu kelas,
maka dilakukan proses splitting
dengan menggunakan salah satu
atribut. Misalkan = “Pembayaran”

Contoh Algoritma Hunt

Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Contoh Algoritma Hunt

Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Contoh Algoritma Hunt

Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Proses Algoritma Hunt

Atribut pembayaran memiliki dua nilai yaitu ‘Yes’ dan ‘No’, maka Dt menjadi:

$$Dt(\text{pembayaran}=\text{Yes}) = \{1, 4, 7\}$$

$$Dt(\text{pembayaran}=\text{No}) = \{2, 3, 5, 6, 8, 9, 10\}$$

$$Dt(\text{pembayaran}=\text{Yes}) \cup
Dt(\text{pembayaran}=\text{No}) = Dt_{\text{mula-mula}}$$

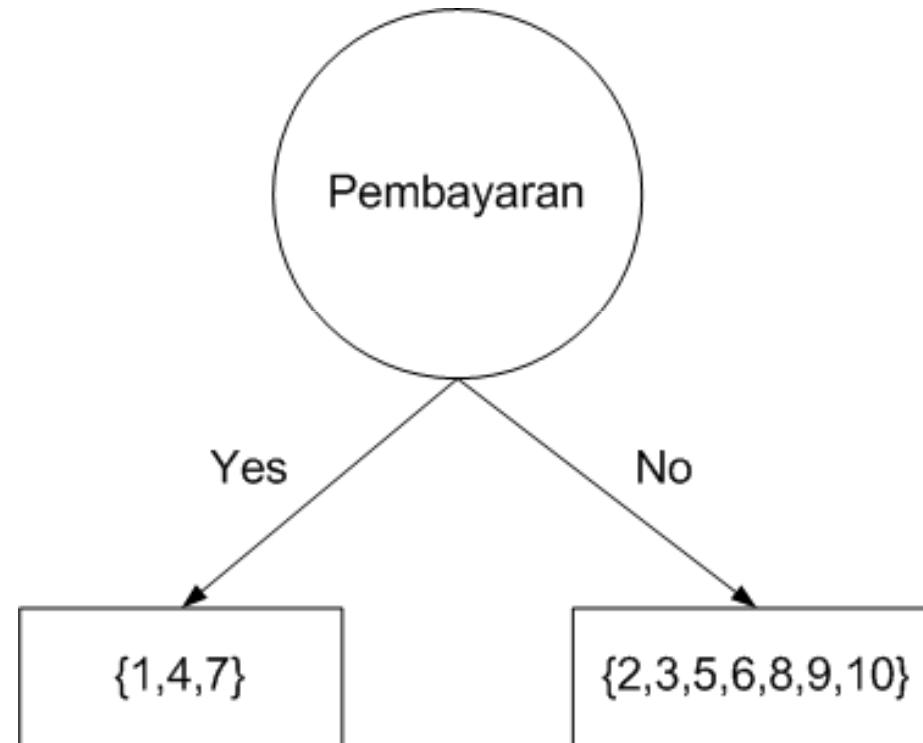
$$Dt(\text{pembayaran}=\text{Yes}) \cap
Dt(\text{pembayaran}=\text{No}) = \{ \}$$

Contoh Algoritma Hunt

Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Proses Algoritma Hunt



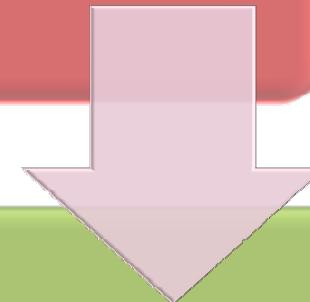
Proses Algoritma Hunt

Ulangi pengecekan pada langkah 2 untuk setiap Dt.

Misalkan

$Dt = Dt(\text{pembayaran} = \text{'Yes'}) = \{1, 4, 7\}$

Karena Dt terletak pada satu kelas yang sama ‘No’, maka simpul t dimana Dt berada menjadi simpul daun dengan label ‘No’”

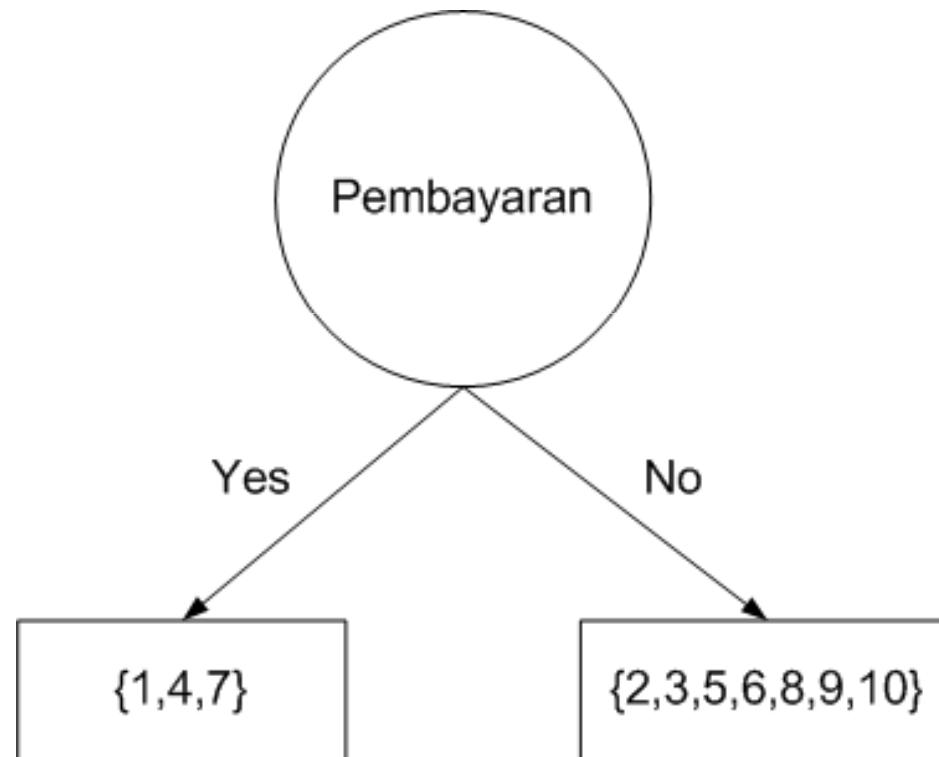


Contoh Algoritma Hunt

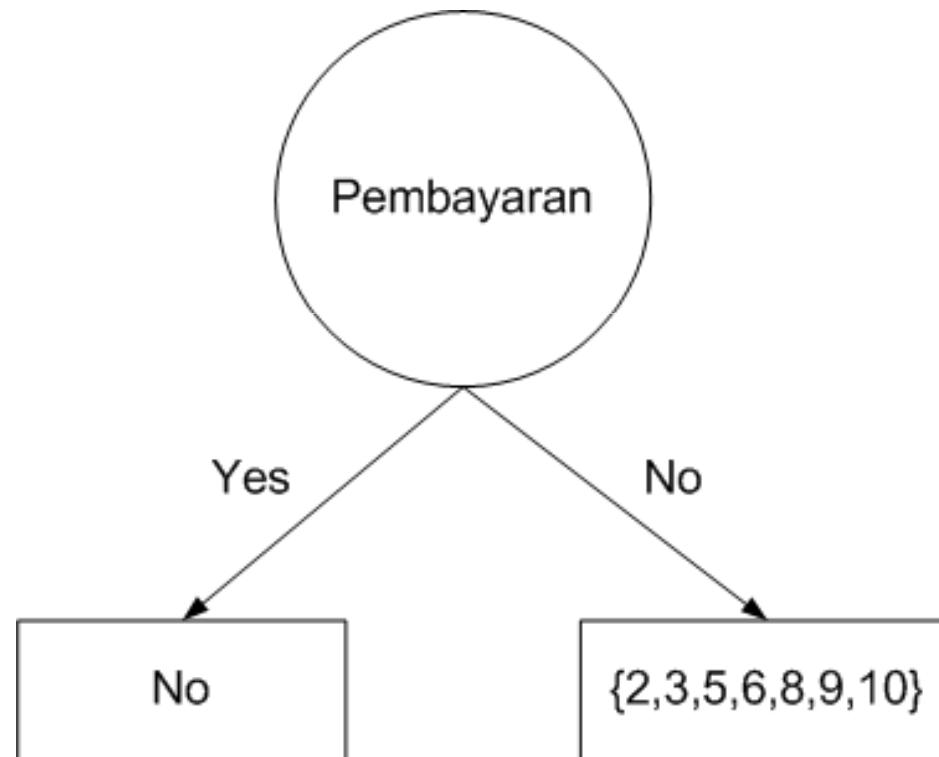
Id	Pembayaran	Status	Pendapatan (x10.000)	Kelas
1	Yes	Single	125	No
2	No	Married	100	No
3	No	Single	70	No
4	Yes	Married	120	No
5	No	Divorced	95	Yes
6	No	Married	60	No
7	Yes	Divorced	220	No
8	No	Single	85	Yes
9	No	Married	75	No
10	No	Single	90	Yes

Sumber : Hermawati, Fajar Astuti, 2013. Data Mining. CV.Andi Offset,Yogyakarta

Proses Algoritma Hunt



Proses Algoritma Hunt



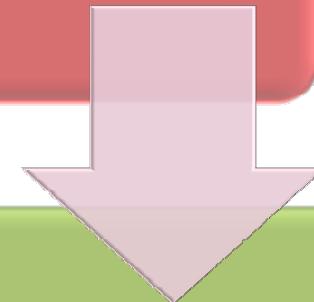
Proses Algoritma Hunt

Ulangi pengecekan pada langkah 2 untuk setiap Dt.

Misalkan

$Dt = Dt(\text{pembayaran} = \text{'No'}) = \{2, 3, 5, 6, 8, 9, 10\}$

Dt tidak terletak pada satu kelas, maka dilakukan proses splitting menggunakan salah satu atribut sisa, misal: Status



Proses Algoritma Hunt

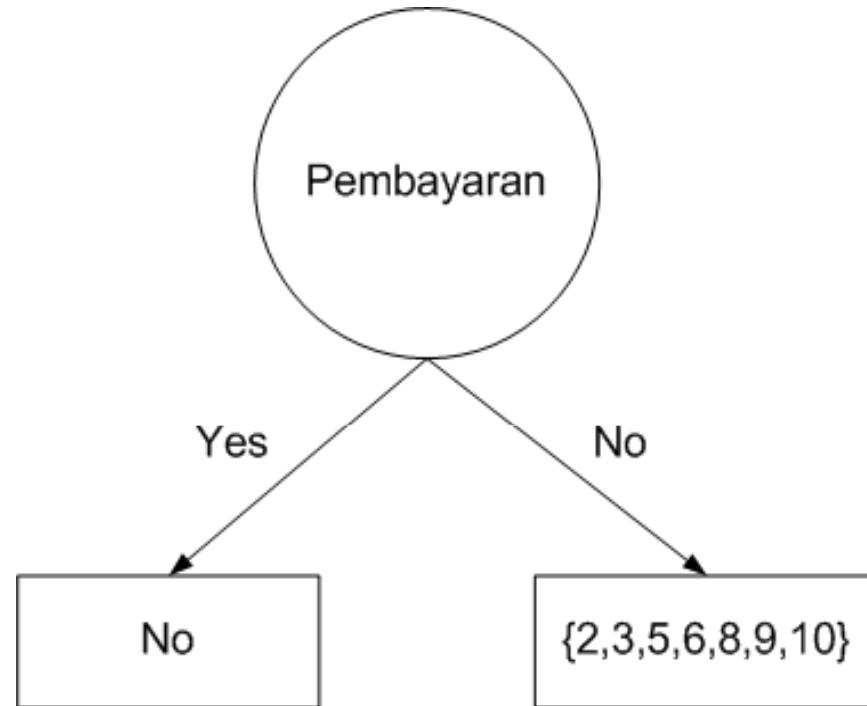
Atribut ‘Status’ memiliki 3 nilai, sehingga:

$Dt(\text{status}=\text{'Single'})=\{3,8,10\}$

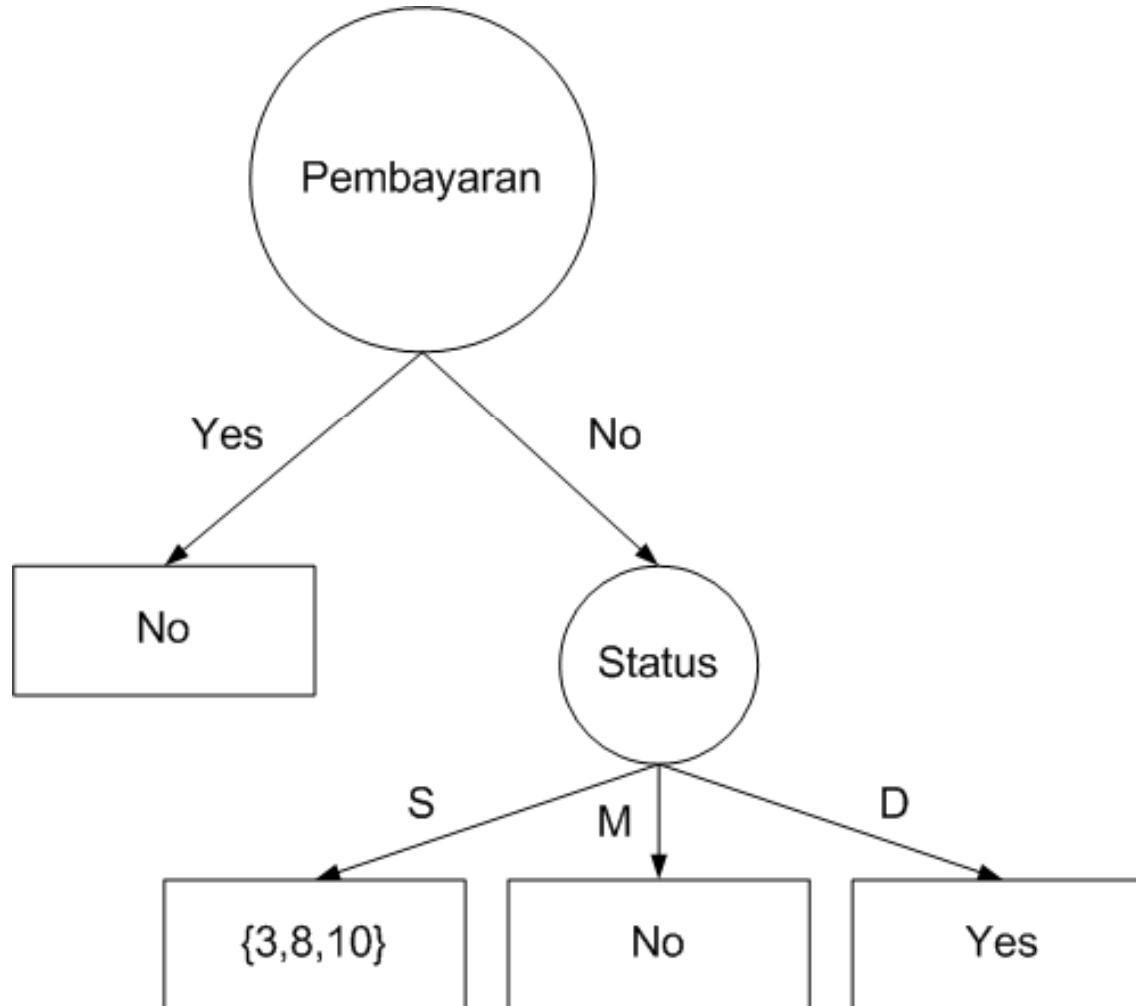
$Dt(\text{status}=\text{'Married'})=\{2,6,9\}$

$Dt(\text{status}=\text{'Divorced'})=\{5\}$

Proses Algoritma Hunt



Proses Algoritma Hunt



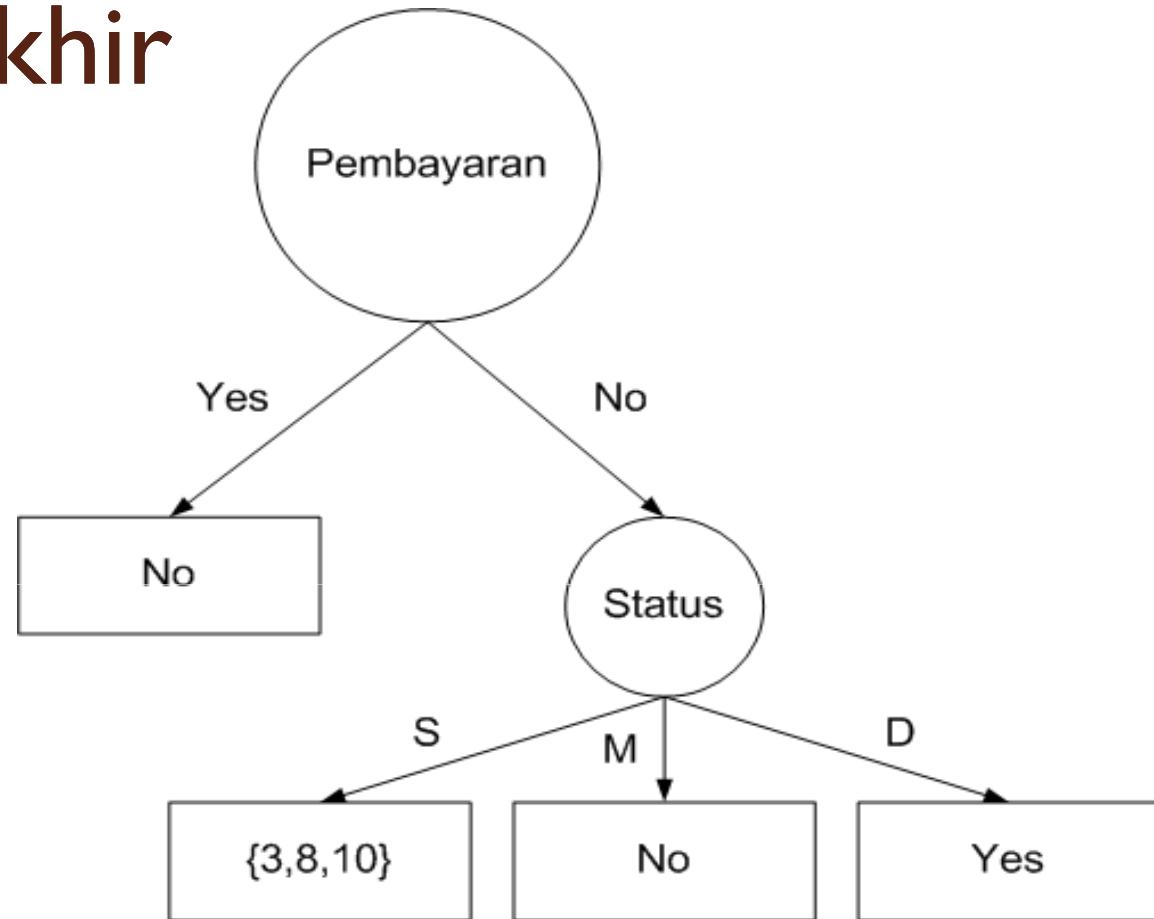
Proses Algoritma Hunt

$Dt=\{3,8,10\}$ tidak terletak pada satu kelas, maka dilakukan proses splitting dengan atribut “Pendapatan” misalkan dibagi dalam ‘ ≤ 80 ’ dan ‘ > 80 ’

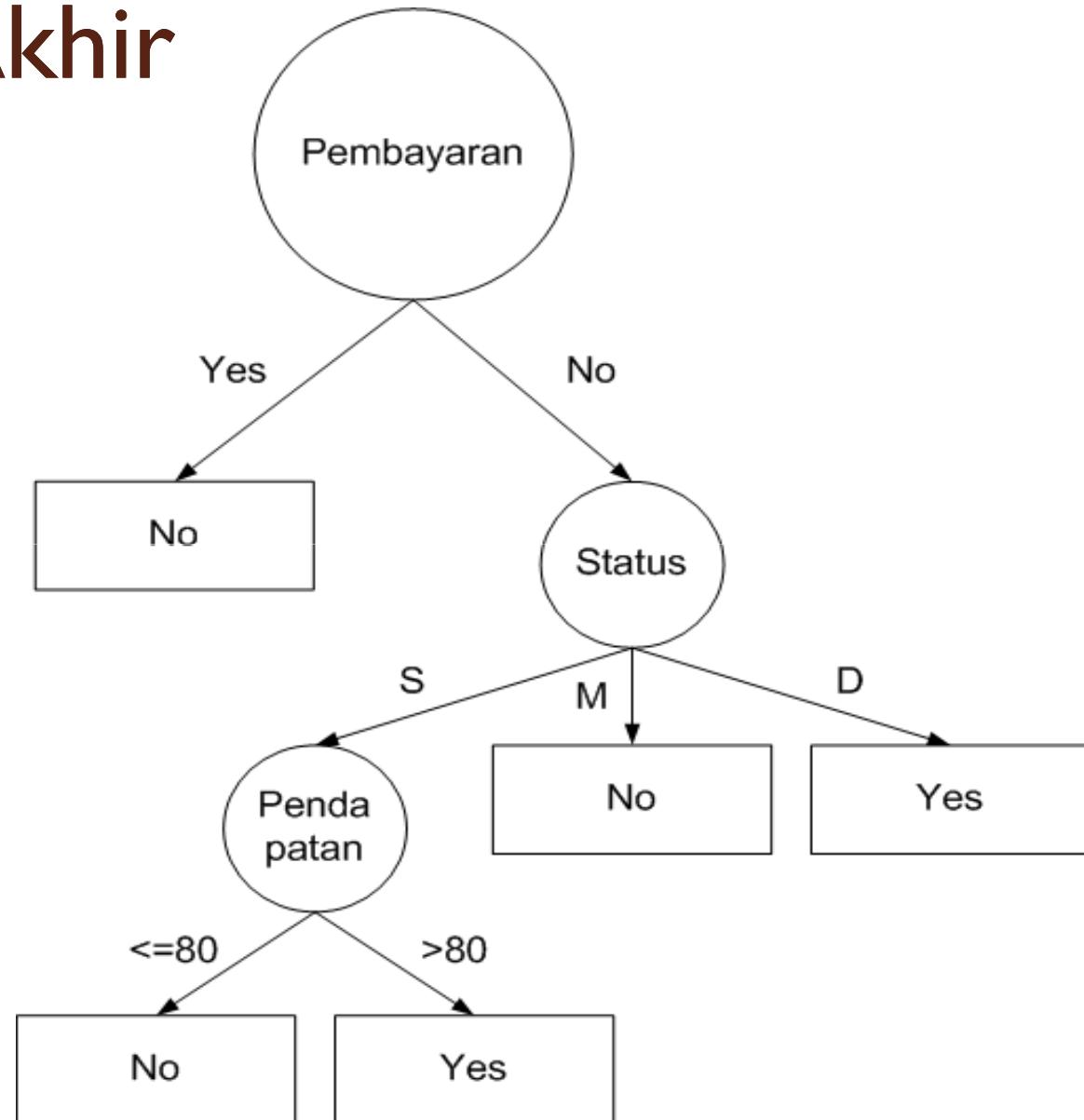
$Dt(\text{pendapatan} \leq 80) = \{3\}$

$Dt(\text{pendapatan} > 80) = \{8, 10\}$

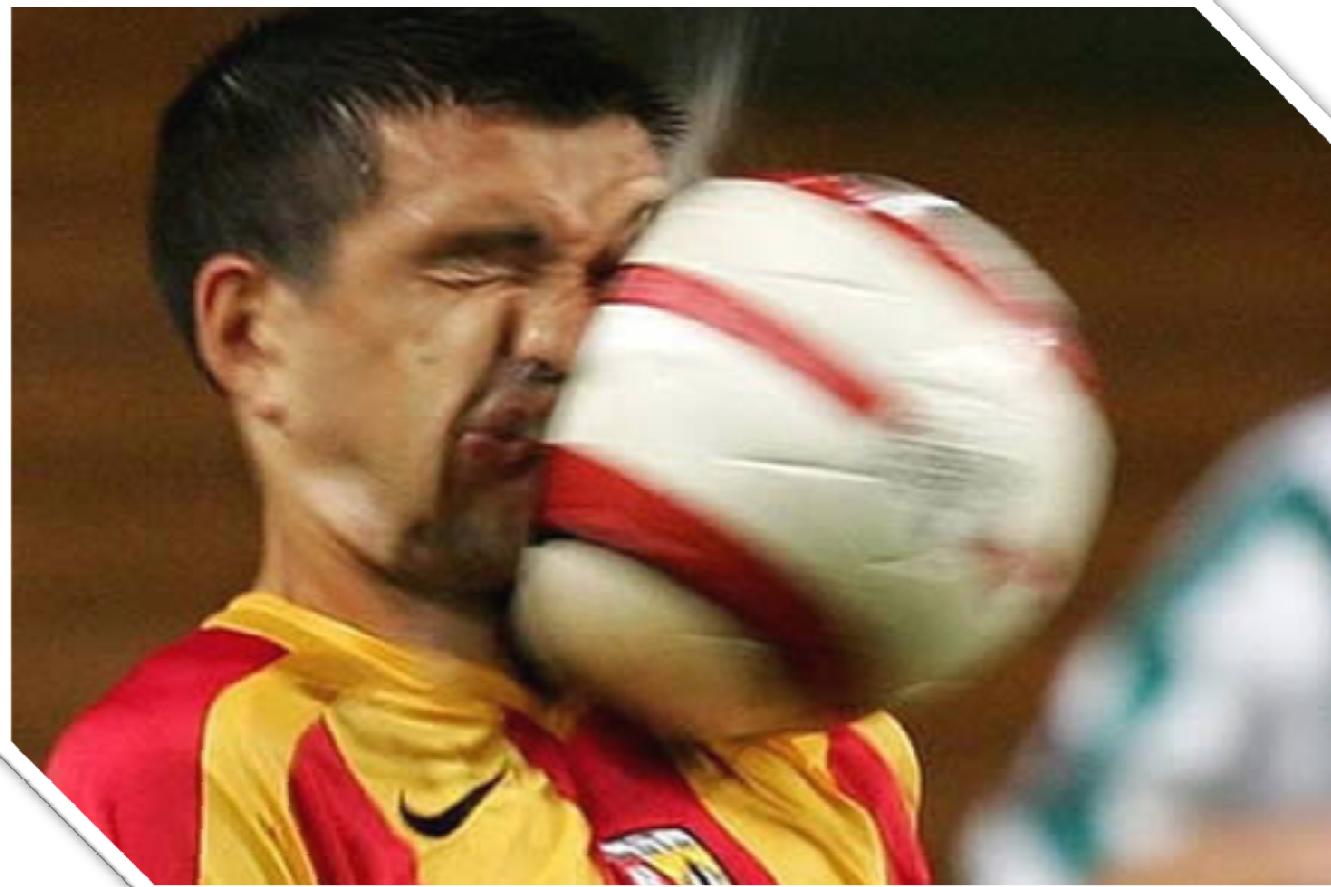
Hasil Akhir



Hasil Akhir



Sesederhana itukah ?



TIDAK !!!!

Bagaimana Memilih Atribut ?

