

Bab3



COURSES
KMMI
Kredensial Mikro Mahasiswa Indonesia

Tugas Mandiri C4.5

2 PILIHAN KURSUS
- PEMROGRAMAN ANDROID
- KOTLIN
- DATA MINING WITH
PYTHON

NIM : 2000019
NAMA : Muhammad Ajriel Rahayu
NAMA KAMPUS : Universitas Pendidikan Indonesia

Tugas

no	Nama	Asal	organisasi	hobi	prioritas
1	umi	Malang	tidak	ym	2
2	vita	Malang	ya	facebook	1
3	vitis	Surabaya	tidak	shooping	3
4	waskita	Kediri	ya	facebook	1
5	wika	Surabaya	ya	facebook	1
6	wiratno	Kediri	tidak	ym	2
7	yudi	Surabaya	tidak	ym	2
8	yianto	Kediri	tidak	ym	2
9	yunita	Surabaya	ya	facebook	1
10	zida	Kediri	ya	shooping	2
11	desi	Surabaya	tidak	shooping	3
12	sita	Kediri	tidak	facebook	2
13	novia	Malang	ya	facebook	1
14	ika	Malang	tidak	shooping	3
15	enggar	Kediri	tidak	facebook	2

1. Buat decision tree untuk melakukan klasifikasi tersebut,
2. Langkah :
 1. Seleksi dulu atribut mana saja yang dipergunakan dan mana yang dibuat
 2. Gunakan algoritma c4.5 untuk membuat decision tree nya dengan kelasnya adalah prioritas Dari decision tree tersebut, buat rulanya

Jawaban :

1. Langkah 1 : Menyeleksi atribut yang dipakai
Untuk atribut yang dipakai adalah Asal, Organisasi, Hobi dan Prioritas sebagai label. Untuk atribut Nomor dan juga Nama tidak dipakai karena hanya sebagai identitas saja, berikut atribut yang dipakai



COURSES KMMI

Kredensial Mikro Mahasiswa Indonesia

Tugas Mandiri C4.5

2 PILIHAN KURSUS

- PEMROGRAMAN ANDROID
- KOTLIN
- DATA MINING WITH PYTHON

NO	Asal	Organisasi	Hobi	Prioritas
1	Malang	tidak	ym	2
2	Malang	ya	facebook	1
3	Surabaya	tidak	shooping	3
4	Kediri	ya	facebook	1
5	Surabaya	ya	facebook	1
6	Kediri	tidak	ym	2
7	Surabaya	tidak	ym	2
8	Kediri	tidak	ym	2
9	Surabaya	ya	facebook	1
10	Kediri	ya	shooping	2
11	Surabaya	tidak	shooping	3
12	Kediri	tidak	facebook	2
13	Malang	ya	facebook	1
14	Malang	tidak	shooping	3
15	Kediri	tidak	facebook	2

2. Langkah 2 : Membuat decision tree c4.5

- a. Untuk membangun sebuah decision tree, diawali dengan mencari nilai *entropy*, *Gain*, *Splitinfo* dan *Gain Ratio* pada masing- masing atribut yang diformulasikan dengan

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

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

$$Split_{info}(S) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

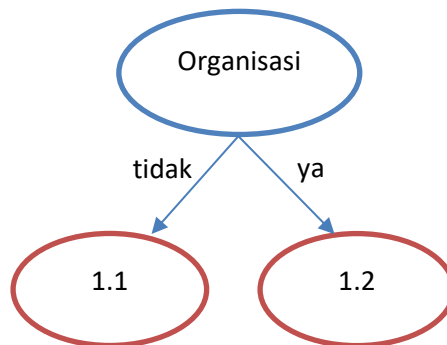
$$GainRatio(A) = \frac{Gain(A)}{Split_{info}(A)}$$

Sehingga jika data dilakukan perhitungan diperoleh hasil sebagai berikut:

		Jumlah(S)	1(Si)	2(Si)	3(Si)	Entropy	Gain	splitinfo	Gain Ratio
Total		15	5	7	3	1,5058231			
Asal							0,598513735	1,56559623	0,382291247
	Malang	4	2	1	1	1,5			
	Surabaya	5	2	1	2	1,521928095			
	Kediri	6	1	5	0	0			
Organisasi							1,5058231	0,970950594	1,5508751
	ya	6	5	1	0	0			
	tidak	9	0	6	3	0			
Hobi							1,5058231	1,530124965	0,984117725
	ym	4	0	4	0	0			
	facebook	7	5	2	0	0			
	Shooping	4	0	1	3	0			



Untuk menentukan *Root node* dari Decision Tree, dipilih *gain ratio* yang paling besar, dalam hal ini atribut organisasi memiliki *gain ratio* yang lebih besar dari atribut lainnya, sehingga *Root node* dipilih organisasi dengan cabang nilai ya dan tidak



Karena *leaf node*(label) belum ditemukan maka harus membuat *internal node*, kemudian langkah selanjutnya mencari atribut apa yang cocok untuk mengisi *internal node* pada 1.1 dan 1.2, pada percobaan ini saya mengambil nilai ya pada atribut organisasi untuk mengisi internal node 1.2.

No	Nama	Asal	Organisasi	Hobi	Prioritas
1	Vita	Malang	ya	facebook	1
2	Waskita	Kediri	ya	facebook	1
3	Wika	Surabaya	ya	facebook	1
4	Yunita	Surabaya	ya	facebook	1
5	Zida	Kediri	ya	shooping	2
6	Novia	Malang	ya	facebook	1

Kemudian mencari nilai *entropy*, *Gain*, *Splitinfo* dan *Gain Ratio* pada masing- masing atribut, diperoleh hasil sebagai berikut

		Jumlah(S)	1(Si)	2(Si)	Entropy	Gain	splitinfo	Gain Ratio
Total		6	5	1	0,650022422			
Asal						0,316689088	1,584962501	0,199808568
	Malang	2	2	0	0			
	Surabaya	2	2	0	0			
	Kediri	2	1	1	1			
Hobi						0,650022422	0,650022422	1
	ym	0	0	0	0			
	facebook	5	5	0	0			
	Shooping	1	0	1	0			

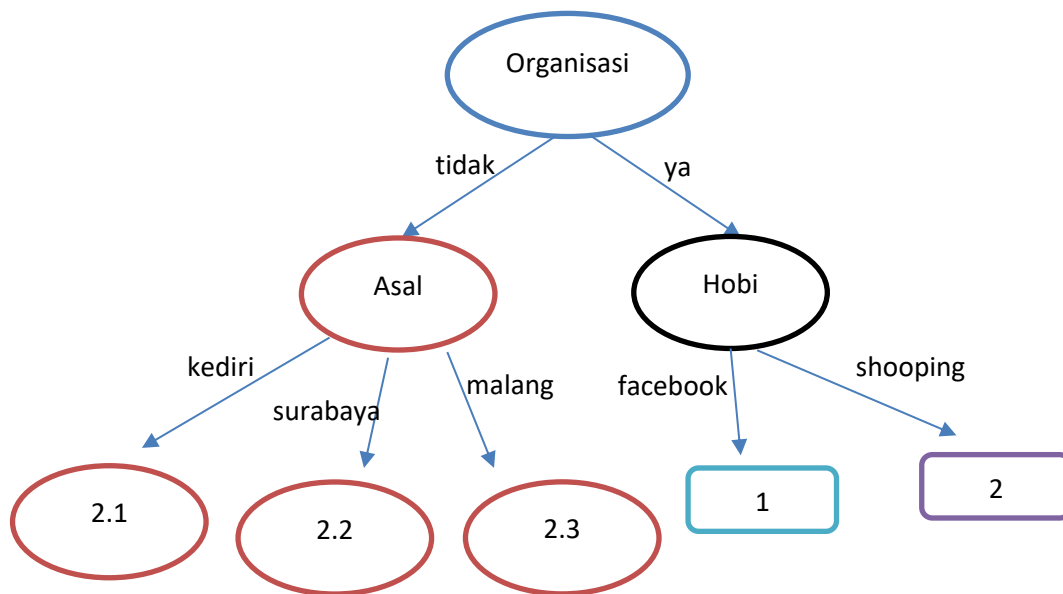


**COURSES
KMMI**
Kredensial Mikro Mahasiswa Indonesia

Tugas Mandiri C4.5

2 PILIHAN KURSUS
- PEMROGRAMAN ANDROID
- KOTLIN
- DATA MINING WITH PYTHON

Untuk mengisi internal node 1.2 yang kosong dipilih Gain ratio terbesar sehingga atribut hobi menjadi pengisi untuk node 1.2 karena memiliki nilai gain ratio terbesar, dan asal dipilih untuk node 1.1. Kemudian pada node hobi, seperti pada tabel diatas semua nilai entropy nya 0 dan juga semua nilai facebook ada pada label 1(Si) dan shooping ada pada label 2(Si) maka bisa langsung diberi label untuk node hobi.



Langkah selanjutnya mencari nilai node 2.1, 2.2, 2.3 , dilakukan dengan memfilter data yang atribut organisasinya bernilai tidak saja seperti pada tabel berikut

No	Nama	Asal	Organisasi	Hobi	Prioritas
1	Umi	Malang	tidak	ym	2
2	Vitis	Surabaya	tidak	shooping	3
3	Wiranto	Kediri	tidak	ym	2
4	Yudi	Surabaya	tidak	ym	2
5	Yianto	Kediri	tidak	ym	2
6	Desi	Surabaya	tidak	shooping	3
7	Sita	Kediri	tidak	facebook	2
8	Ika	Malang	tidak	shooping	3
9	Enggar	Kediri	tidak	facebook	2

Disini melihat atribut asal dan juga label prioritas, ditemukan hasil bahwa semua nilai kediri berlabel 2 sehingga dapat langsung diberikan label pada node 2.1, sedangkan nilai malang dan Surabaya masih memiliki label yang berbeda beda, lebih jelasnya dapat dilihat pada tabel di bawah ini



COURSES KMMI

Kredensial Mikro Mahasiswa Indonesia

Tugas Mandiri C4.5

2 PILIHAN KURSUS
- PEMROGRAMAN ANDROID
- KOTLIN
- DATA MINING WITH
PYTHON

No	Nama	Asal	Organisasi	Hobi	Prioritas
1	Vitis	Surabaya	tidak	shooping	3
2	Yudi	Surabaya	tidak	ym	2
3	Desi	Surabaya	tidak	shooping	3
4	Umi	Malang	tidak	ym	2
5	Ika	Malang	tidak	shooping	3

Karena malang dan Surabaya masih memiliki label yang berbeda maka dibuat cabang node kembali dan atribut yang tersisa hanya hobi jadi node 2.2 dan 2.3 bisa diisi dengan atribut hobi dan juga dapat langsung diberi label karena semua nilai atribut hobi yang sesuai dengan asal sudah sesuai dengan label, dan gambar pola decision tree dari percobaan ini dapat disimpulkan sebagai berikut:

