

PERHITUNGAN ALGORITMA C4.5



www.tyas-tamimy.com



tyas_fa@stmikmpb.ac.id



fitriayuningtyas@stmikmpb.ac.id



081804767700 (Telegram only)



Tyas Tamimy

Definisi

- Algoritma C4.5 merupakan salah satu algoritma klasifikasi yang digunakan untuk melakukan klasifikasi dan bersifat prediktif, C4.5 merupakan evolusi dari ID3, disajikan oleh penulis yang sama [Quinlan (1993)], ID3 menggunakan *Information Gain* sedangkan C4.5 menggunakan *Gain Ratio* sebagai kriteria pemisahan (*splitting criteria*) [1, 2].
- Dasar algoritma C4.5 adalah pembentukan pohon keputusan (*decision tree*). Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelasnya.

Tahapan Algoritma C4.5

1. Siapkan Dataset

2. Pilih atribut sebagai akar

2.1. Menghitung nilai Entropy Total

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

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

2.3. Menghitung Split Information

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

2.4. Menghitung nilai Gain Ratio

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{Split\ Information(S, A)}$$

3. Buat cabang untuk tiap-tiap nilai

4. Bagi kasus dalam cabang

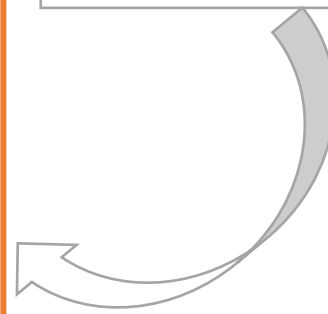
5. Ulangi proses 2 – 4 untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

1. Siapkan Dataset

age	spectacle-prescrip	astigmatism	tear-prod-rate	contact-lenses
young	myope	no	reduced	none
young	myope	no	normal	soft
young	myope	yes	reduced	none
young	myope	yes	normal	hard
young	hypermetrope	no	reduced	none
young	hypermetrope	no	normal	soft
young	hypermetrope	yes	reduced	none
young	hypermetrope	yes	normal	hard
pre-presbyopic	myope	no	reduced	none
pre-presbyopic	myope	no	normal	soft
pre-presbyopic	myope	yes	reduced	none
pre-presbyopic	myope	yes	normal	hard
pre-presbyopic	hypermetrope	no	reduced	none
pre-presbyopic	hypermetrope	no	normal	soft
pre-presbyopic	hypermetrope	yes	reduced	none
pre-presbyopic	hypermetrope	yes	normal	none
presbyopic	myope	no	reduced	none
presbyopic	myope	no	normal	none
presbyopic	myope	yes	reduced	none
presbyopic	myope	yes	normal	hard
presbyopic	hypermetrope	no	reduced	none
presbyopic	hypermetrope	no	normal	soft
presbyopic	hypermetrope	yes	reduced	none
presbyopic	hypermetrope	yes	normal	none

Menentukan
kesesuaian lensa
kontak dengan
mata pasien

contact-lenses.xls



Label

2. Pilih Atribut sebagai akar

- Menghitung nilai *gain ratio* dari masing-masing atribut, **nilai gain ratio yang paling tinggi yang akan menjadi akar pertama.**
- Hitung Nilai *Entropy*, Nilai *Information Gain*, dan Nilai *Split Information* sebelum menghitung nilai *Gain Ratio*.
- *Entropy* didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data [3].
- *Information Gain* didefinisikan sebagai ukuran efektivitas suatu *attribute* dalam mengklasifikasikan data [3].

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.1.: Menghitung nilai Entropy Total

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

Keterangan:

- S: Himpunan Kasus
- n: Jumlah Partisi dalam S
- p_i : Proporsi dari S_i terhadap S

Jumlah Kasus	none (S1)	soft (S2)	hard (S3)
24	15	5	4

$$Entropy\ Total = \left(-\frac{15}{24} * \log_2 \left(\frac{15}{24}\right)\right) + \left(-\frac{5}{24} * \log_2 \left(\frac{5}{24}\right)\right) + \left(-\frac{4}{24} * \log_2 \left(\frac{4}{24}\right)\right)$$

$$Entropy\ Total = 1,326087525$$

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.2.: Menghitung nilai Information Gain

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

Keterangan:

- S: Himpunan kasus
- A: Atribut
- n: Jumlah partisi dalam A
- $|S_i|$: Jumlah kasus pada partisi ke-i
- $|S|$: Jumlah kasus dalam S
- *Hitung Entropy (Si) dengan cara seperti menentukan Entropy (S)*

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.2.: Menghitung nilai Information Gain

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain
Total		24	15	5	4	1,32609	
age	Young	8	4	2	2	1,5	0,039396504
	Pre-presbyopic	8	5	2	1	1,29879	
	Presbyopic	8	6	1	1	1,06128	
spectacle-prescrip	myope	12	7	2	3	1,38443	0,039510835
	hypermetrope	12	8	3	1	1,18872	
astigmatism	yes	12	8	0	4	0,9183	0,37700523
	no	12	7	5	0	0,97987	
tear-prod-rate	normal	12	3	5	4	1,55459	0,548794941
	reduced	12	12	0	0		

$$\text{Gain (Total, Age)} = 1,32609 - \left(\left(\frac{8}{24} * 1,5 \right) + \left(\frac{8}{24} * 1,29879 \right) + \left(\frac{8}{24} * 1,06128 \right) \right)$$

- Gain (Total, spectacle-prescript) =
- Gain (Total, astigmatism) =
- Gain (Total, tear-prod-rate) =

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.3.: Menghitung Split Information

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

Keterangan:

- S: Himpunan kasus
- A: Atribut
- n: Jumlah partisi dalam A
- Si: Jumlah kasus pada partisi ke-i

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.3.: Menghitung Split Information

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain	Split Information
Total		24	15	5	4	1,32609		
age	Young	8	4	2	2	1,5	0,039396504	1,58496
	Pre-presbyopic	8	5	2	1	1,29879		
	Presbyopic	8	6	1	1	1,06128		
spectacle-prescrip	myope	12	7	2	3	1,38443	0,039510835	1
	hypermetrope	12	8	3	1	1,18872		
astigmatism	yes	12	8	0	4	0,9183	0,37700523	1
	no	12	7	5	0	0,97987		
tear-prod-rate	normal	12	3	5	4	1,55459	0,548794941	1
	reduced	12	12	0	0			

$$\text{SplitInfo}(\text{Total}, \text{Age}) = \left(-\frac{8}{24} * \log_2 \left(\frac{8}{24} \right) \right) + \left(-\frac{8}{24} * \log_2 \left(\frac{8}{24} \right) \right) + \left(-\frac{8}{24} * \log_2 \left(\frac{8}{24} \right) \right)$$

- SplitInfo (Total, spectacle-prescript) =
- SplitInfo (Total, astigmatism) =
- SplitInfo (Total, tear-prod-rate) =

2. Pilih Atribut sebagai akar ... (Lanjutan)

- Langkah 2.4. : Menghitung nilai Gain Ratio

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)}$$

Keterangan:

- S: Himpunan Kasus
- A: Atribut

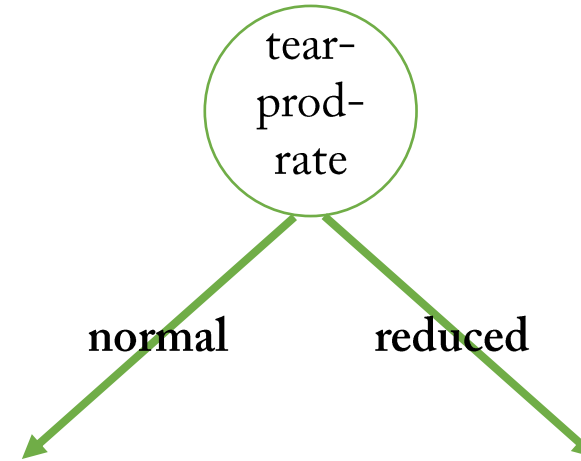
2. Pilih Atribut sebagai akar ... (Lanjutan)

tear-prod-rate memiliki nilai Gain Ratio tertinggi sehingga ter-prod-rate dipilih sebagai Root

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain	Split Information	Gain Ratio
Total		24	15	5	4	1,32609			
age							0,039396504	1,58496	0,024856426
	Young	8	4	2	2	1,5			
	Pre-presbyopic	8	5	2	1	1,29879			
	Presbyopic	8	6	1	1	1,06128			
spectacle-prescrip							0,039510835	1	0,039510835
	myope	12	7	2	3	1,38443			
	hypermetrope	12	8	3	1	1,18872			
astigmatism							0,37700523	1	0,37700523
	yes	12	8	0	4	0,9183			
	no	12	7	5	0	0,97987			
tear-prod-rate							0,548794941	1	0,548794941
	normal	12	3	5	4	1,55459			
	reduced	12	12	0	0				

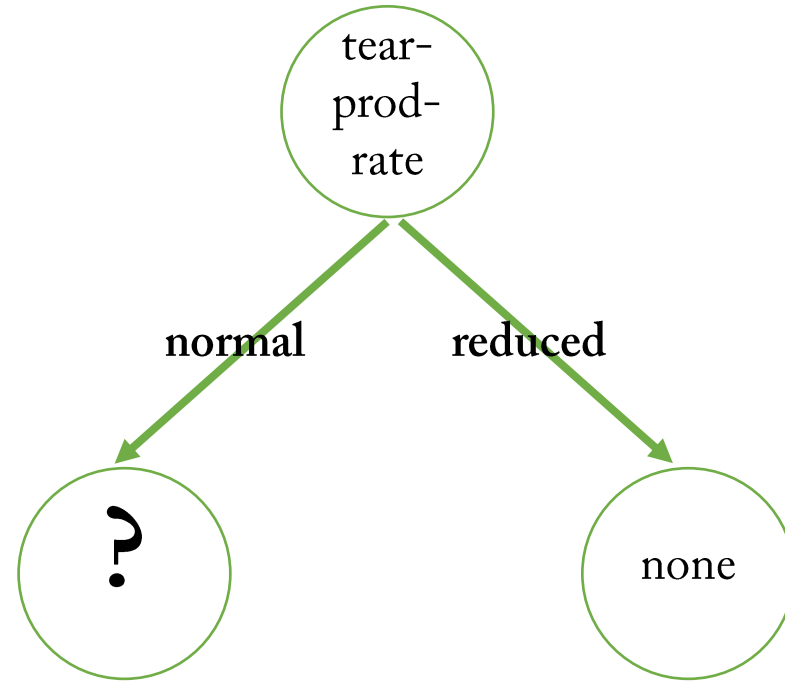
3. Buat cabang untuk tiap-tiap nilai

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)
tear-prod-rate					
	normal	12	3	5	4
	reduced	12	12	0	0



4. Bagi kasus dalam cabang

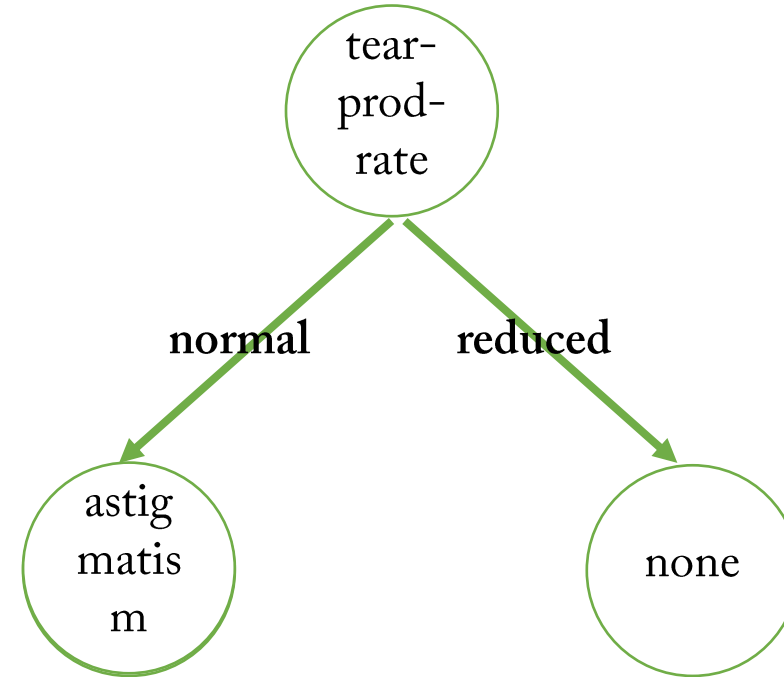
Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)
tear-prod-rate					
	normal	12	3	5	4
	reduced	12	12	0	0



5. Ulangi proses 2-4 untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

5.1.

age	spectacle-prescrip	astigmatism	tear-prod-rate	contact-lenses
young	myope	no	normal	soft
young	myope	yes	normal	hard
young	hypermetrope	no	normal	soft
young	hypermetrope	yes	normal	hard
pre-presbyopic	myope	no	normal	soft
pre-presbyopic	myope	yes	normal	hard
pre-presbyopic	hypermetrope	no	normal	soft
pre-presbyopic	hypermetrope	yes	normal	none
presbyopic	myope	no	normal	none
presbyopic	myope	yes	normal	hard
presbyopic	hypermetrope	no	normal	soft
presbyopic	hypermetrope	yes	normal	none



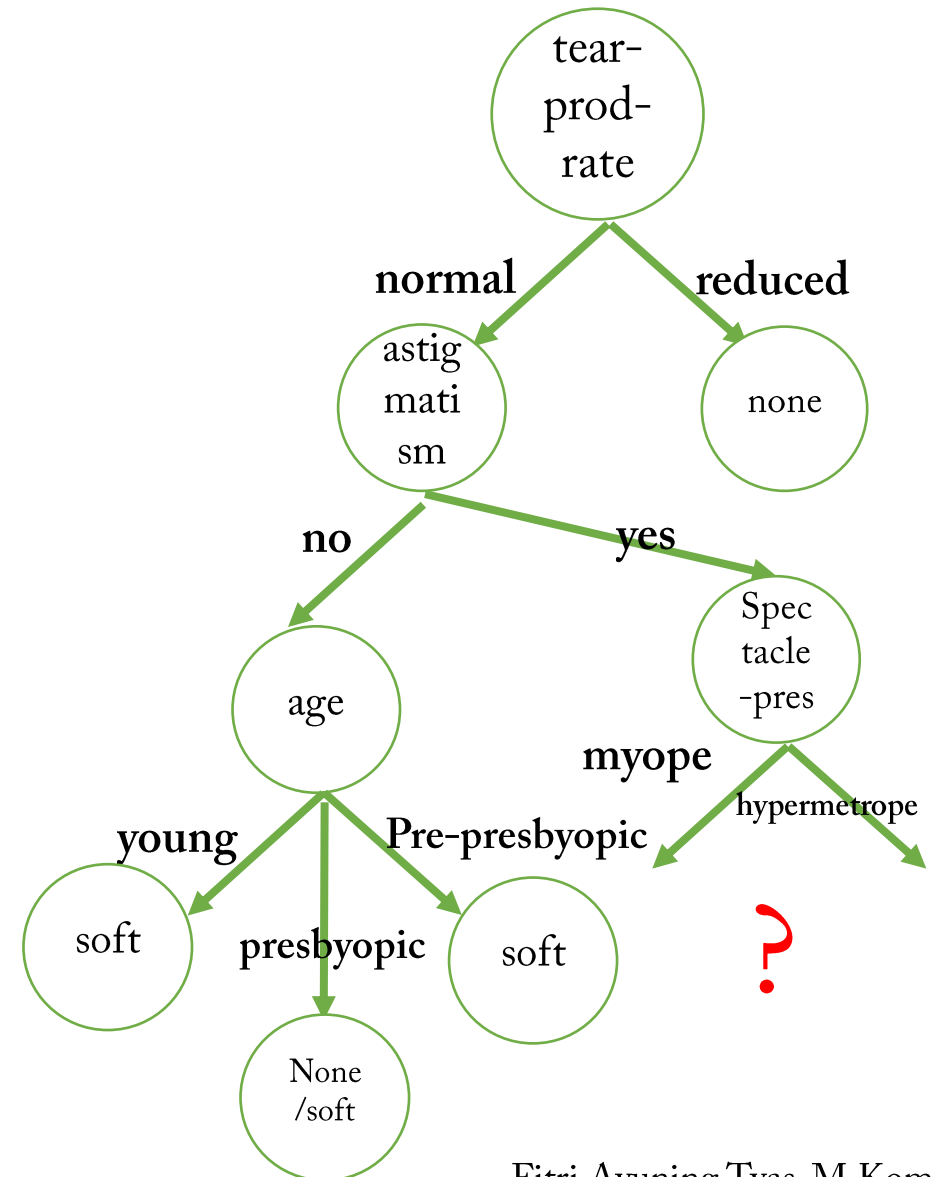
Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain	Split Information	Gain Ratio
Total		12	3	5	4	1,55459			
age							0,22125	1,58496	0,1395944
	Young	4	0	2	2	1			
	Pre-presbyopic	4	1	2	1	1,5			
	Presbyopic	4	2	1	1	1,5			
spectacle-prescrip							0,09544	1	0,0954373
	myope	6	1	2	3	1,45915			
	hypermetrope	6	2	3	1	1,45915			
astigmatism							0,77043	1	0,770426
	yes	6	2	0	4	0,9183			
	no	6	1	5	0	0,65002			

5. Ulangi proses 2-4 untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

5.2.

age	spectacle-prescrip	astigmatism	tear-prod-rate	contact-lenses
young	myope	no	normal	soft
young	hypermetrope	no	normal	soft
pre-presbyopic	myope	no	normal	soft
pre-presbyopic	hypermetrope	no	normal	soft
presbyopic	myope	no	normal	none
presbyopic	hypermetrope	no	normal	soft

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain	Split Information	Gain Ratio
Total		6	1	5	0	0,65002			
age							0,31669	1,58496	0,1998086
	Young	2	0	2	0	0			
	Pre-presbyopic	2	0	2	0	0			
	Presbyopic	2	1	1	0	1			
spectacle-prescrip							0,19087	1	0,1908745
	myope	3	1	2	0	0,9183			
	hypermetrope	3	0	3	0	0			

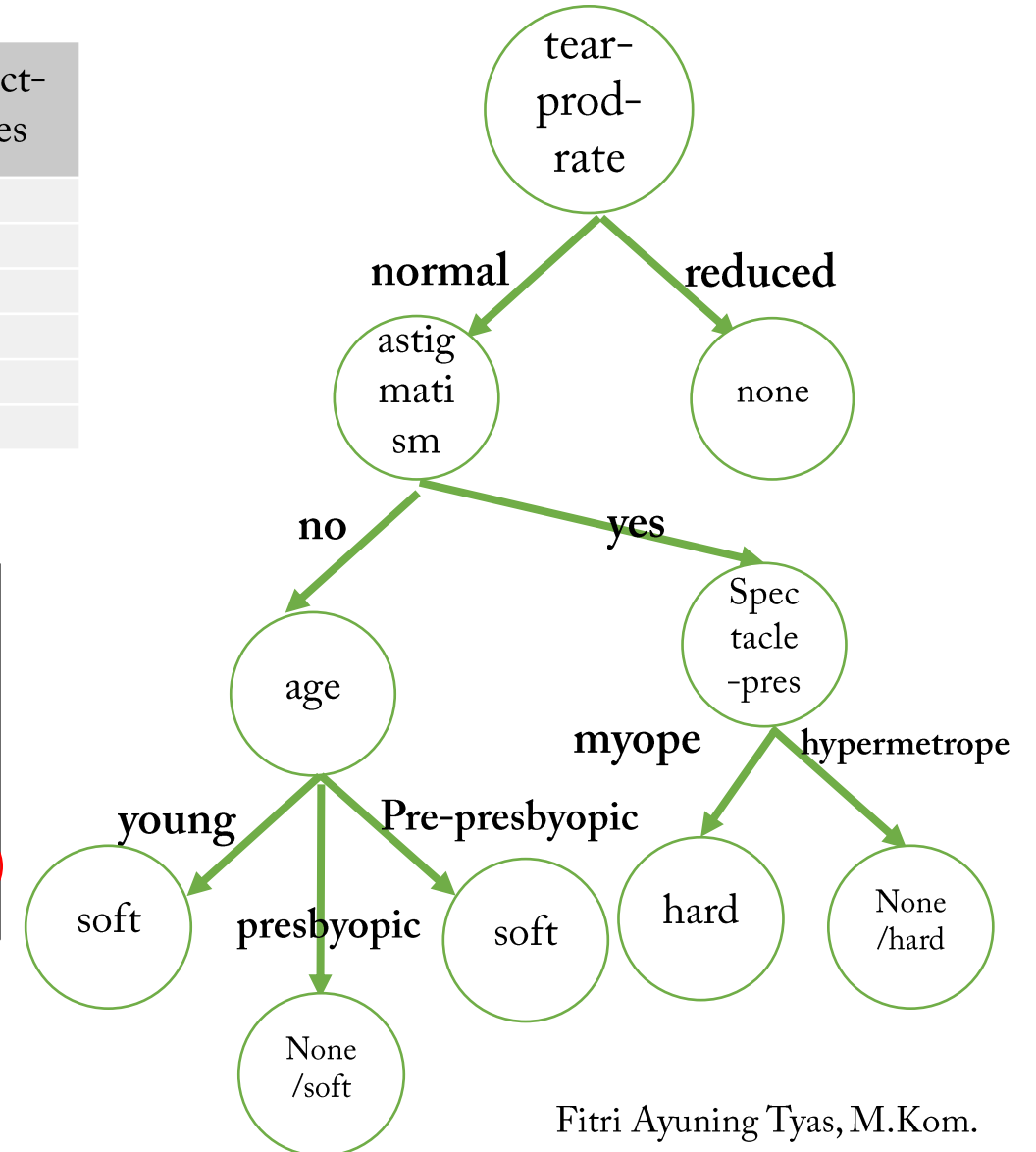


5. Ulangi proses 2-4 untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

5.3.

age	spectacle-prescrip	astigmatism	tear-prod-rate	contact-lenses
young	myope	yes	normal	hard
young	hypermetrope	yes	normal	hard
pre-presbyopic	myope	yes	normal	hard
pre-presbyopic	hypermetrope	yes	normal	none
presbyopic	myope	yes	normal	hard
presbyopic	hypermetrope	yes	normal	none

Atribut	Value	Jumlah Kasus	none (S1)	soft (S2)	hard (S3)	Entropy	Information Gain	Split Information	Gain Ratio
Total		6	2	0	4	0,9183			
age							0,25163	1,58496	0,1587603
	Young	2	0	0	2	0			
	Pre-presbyopic	2	1	0	1	1			
	Presbyopic	2	1	0	1	1			
spectacle-prescrip							0,45915	1	0,4591479
	myope	3	0	0	3	0			
	hypermetrope	3	2	0	1	0,9183			



Daftar Pustaka

- [1] L. Rokach and O. Maimon, Data Mining with Decision Trees Theory and Applications 2nd Editions. World Scientific Publishing Co. Pte. Ltd., 2015.
- [2] C. J. Mantas and J. Abellán, “Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data,” Expert Syst. Appl., vol. 41, no. 10, pp. 4625–4637, 2014.
- [3] Dr. Suyanto, S. M, Data Mining Untuk Klasifikasi Data. Bandung : Informatika , 2018.