

Latihan DTL #2

13523149@std.stei.itb.ac.id [Switch account](#)



The name, email, and photo associated with your Google account will be recorded when you upload files and submit this form

* Indicates required question

Nama *

Your answer

NIM *

Your answer

Data

Untuk soal Decision Tree Learning, diberikan sebuah dataset berisi 7 data (dengan satu data memiliki missing value pada salah satu atribut) dengan 3 fitur (Atr1, Atr2, Atr3) dan Kelas Target dengan label "Yes" dan "No" sebagai berikut:

Atr1	Atr2	Atr3	Class
A	70	0	Yes
A	90	1	No
B	90	1	Yes
B	95	0	Yes
C	75	0	Yes
C	85	1	No
?	88	1	No

Untuk semua soal di bawah ini, gunakan prosedur yang telah diajarkan di kelas untuk menghitung entropy, information gain dan gain ratio untuk data dengan missing value.



1. hitunglah nilai entropy data tanpa instance dengan missing value
2. hitunglah Information Gain untuk atribut Atr1 atau ditulis dengan Gain(S,Atr1) sesuai soal pada poin 1 dan dataset di atas
3. hitunglah Gain Ratio untuk atribut Atr1 atau ditulis dengan GainRatio(S,Atr1) dan gunakan nilai-nilai yang diperoleh pada poin 1 dan 2 soal DTL ini

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

S: set dari data
 c: jumlah kelas target
 p_i: proporsi data pada S dengan kelas i

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in \mathcal{V}_{A(A)}} \frac{|S_v|}{|S|} Entropy(S_v)$$

S: set dari data parent
 A: atribut yg dihitung information gain-nya
 v: nilai dari atribut A
 S_v: subset dari data S yg memiliki nilai v untuk atribut A

Rumus entropy mengikuti soal no. 1, baik untuk data parent (S) maupun subset data dengan nilai v (S_v)

Karena ada data dengan missing value, maka nilai Gain(S,A) dg missing value = proporsi known value * Gain(S,A) keseluruhan

$$GainRatio(S, A) \equiv \frac{Gain(S, A)}{SplitInformation(S, A)}$$

$$SplitInformation(S, A) \equiv - \sum_{v \in \mathcal{V}_A} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|}$$

where S_v is subset of S for which A has value v_i
 S: set dari data parent
 A: atribut yg dihitung information gain-nya
 v_i: nilai dari atribut A
 S_v: subset dari data S yg memiliki nilai v_i untuk atribut A

Rumus entropy mengikuti soal no. 1, baik untuk data parent (S) maupun subset data dengan nilai v (S_v)

Karena ada data dengan missing value, maka nilai Gain(S,A) dg missing value = proporsi known value * Gain(S,A) keseluruhan

Upload 1 supported file. Max 10 MB.

 Add file

Submit

Clear form

Never submit passwords through Google Forms.

This form was created inside of School of Electrical Engineering and Informatics. - [Contact form owner](#)

Does this form look suspicious? [Report](#)

Google Forms

