

TUGAS I-M3

KNOWLEDGE DISCOVERY

DECISION TREE

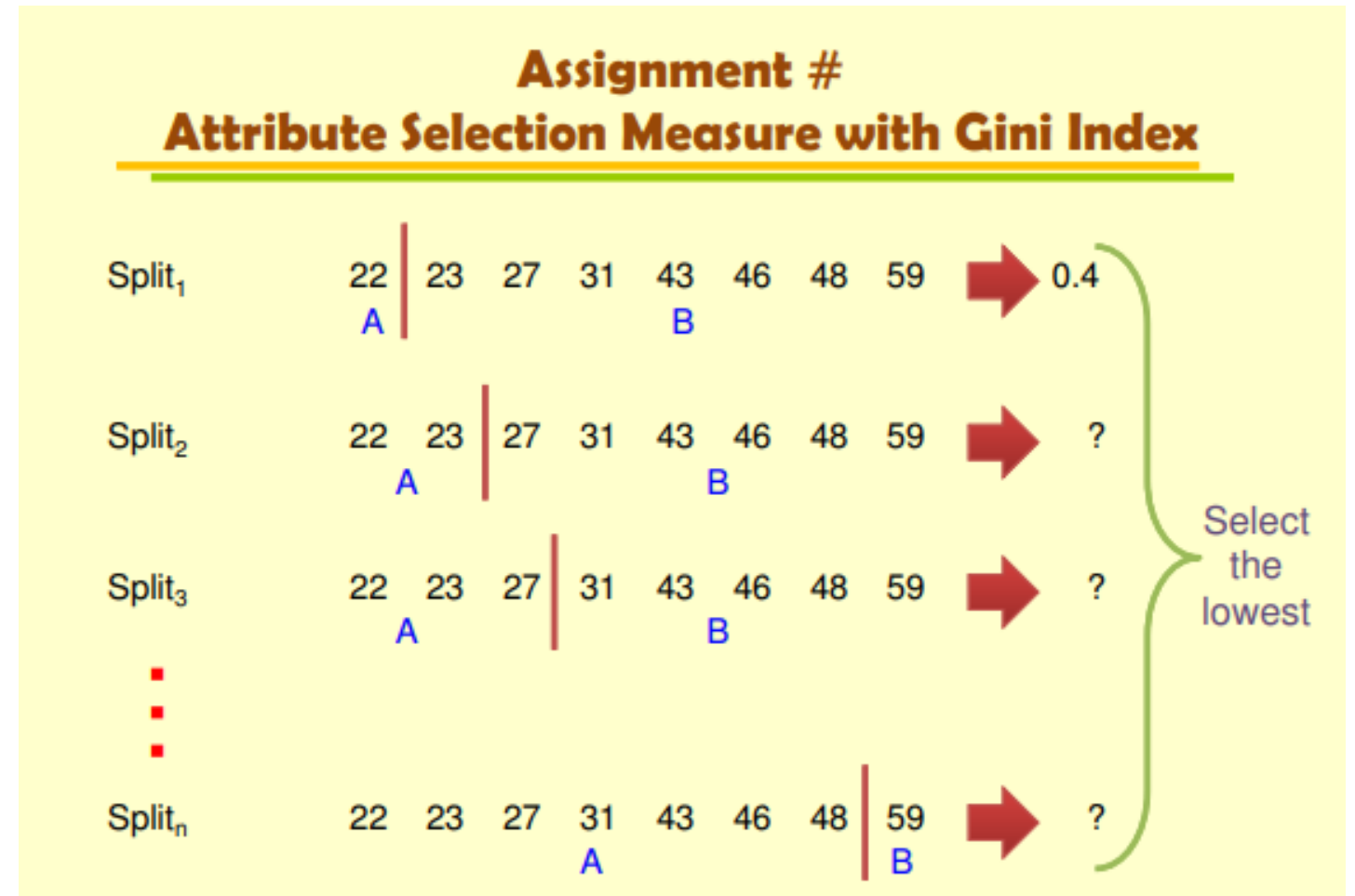
Nama : Ni Putu Devira Ayu Martini
Kelas : S2 Elektro 2020
NRP : 1120800012
Pengampu : Bp. Ali Ridho Barakbah

DATASET

#	Age	Weight	Sex	Hypertension
1	22	overweight	male	yes
2	27	underweight	male	no
3	31	average	female	no
4	46	overweight	male	no
5	59	overweight	male	yes
6	23	underweight	male	no
7	48	overweight	female	yes
8	43	average	male	no

SOAL!

- Konversikan data numerik disamping menjadi data kategorikal dengan Teori Gini Index.



JAWAB!

- Split 1 = $\frac{1}{8} Gini_A + \frac{7}{8} Gini_B$
= $\frac{1}{8} [1 - (\frac{1}{1})^2 - (\frac{0}{1})^2] + \frac{7}{8} [1 - (\frac{2}{7})^2 - (\frac{5}{7})^2] = 0.4$
- Split 2 = $\frac{2}{8} Gini_A + \frac{6}{8} Gini_B$
= $\frac{2}{8} [1 - (\frac{1}{2})^2 - (\frac{1}{2})^2] + \frac{6}{8} [1 - (\frac{4}{6})^2 - (\frac{2}{6})^2] = 0.45$
- Split 3 = $\frac{3}{8} Gini_A + \frac{5}{8} Gini_B$
= $\frac{3}{8} [1 - (\frac{1}{3})^2 - (\frac{2}{3})^2] + \frac{5}{8} [1 - (\frac{2}{5})^2 - (\frac{3}{5})^2] = 0.04$

←
Nilai Split paling kecil yang dipilih

- Split 4 = $\frac{4}{8} Gini_A + \frac{4}{8} Gini_B$
 $= \frac{4}{8} [1 - (\frac{1}{4})^2 - (\frac{3}{4})^2] + \frac{4}{8} [1 - (\frac{2}{4})^2 - (\frac{2}{4})^2] = 0.43$
- Split 5 = $\frac{5}{8} Gini_A + \frac{3}{8} Gini_B$
 $= \frac{5}{8} [1 - (\frac{2}{5})^2 - (\frac{3}{5})^2] + \frac{3}{8} [1 - (\frac{1}{3})^2 - (\frac{2}{3})^2] = 0.46$
- Split 6 = $\frac{6}{8} Gini_A + \frac{2}{8} Gini_B$
 $= \frac{6}{8} [1 - (\frac{2}{6})^2 - (\frac{4}{6})^2] + \frac{2}{8} [1 - (\frac{1}{2})^2 - (\frac{1}{2})^2] = 0.45$

- Split 7 = $\frac{7}{8} Gini_A + \frac{1}{8} Gini_B$
 $= \frac{7}{8} \left[1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 \right] + \frac{1}{8} \left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2 \right] = 0.42$

ANALISA

- Untuk mengubah/mengkonversi data numerik menjadi kategorikal pada suatu atribut dengan menggunakan rumus metode Gini Index sebagai berikut:

$$gini_A(D) = \frac{|D_1|}{|D|} gini(D_1) + \frac{|D_2|}{|D|} gini(D_2)$$

- Penentuan konversi data numerik menjadi kategorikan dapat dipilih dengan nilai perhitungan Gini Index yang paling kecil.
- Maka, data kategorikan yang dipilih pada soal dislide sebelumnya ada pada Split-3 yaitu

22	27	31	46	59	23	48	43
↓	↓	↓	↓	↓	↓	↓	↓
A	A	A	B	B	B	B	B