Assignment - 3

Manish. B. Shrijan
4AL16CS131

① Decision tree for new store option.

→ Considering city size as big:

City size	Decision
Big	Yes
Big	Yes
Big	No
Big	

∴ City-size : Big → yes.
2/3 is correct 2 1/3 is error.

→ Considering city-size as Med

City size	Decision
Med	No
Med	Yes
Med	No
Med	No
Med	No
Med	

∴ city-size : Med → No
4/5 is correct 2 1/5 is error

→ Considering city size as small

City-size	Decision
Small	No
Small	No
Small	Yes
Small	No

∴ City size : Small → No
3/4 is correct 2 1/4 is error.

Attribute rules:

		Error	Total error
City-size	Big → Yes	1/3	3/12
	Med → No	1/5	
	Small → No	1/4	

→ Considering Avg income as high

Avg-income	Decision
High	Yes
High	Yes
High	Yes
High	No
High	Yes
High	No

∴ Avg-income : High → Yes
4/6 is correct 2 2/6 is error.

→ Considering Avg income as Med

Avg income	Decision
Med	No
Med	No
Med	No
Med	No
Med	No

∴ Avg-income : Med → No
5/5 is correct, 0/5 is error.

→ Considering Avg-Income as low:

Avg Income	Decision
low	No
low	No

∴ Avg income : Low → No

1 is correct, 0 is error.

Attribute	Rule	error	total errors
Avg-income	High → Yes	2/6	2/12
	Mid → No	0/5	
	Low → No	0/1	

→ Considering local inventors as Yes

local-inventors	Decision
Yes	Yes
Yes	No
Yes	No
Yes	Yes
Yes	No
Yes	No

∴ local-inventors : Yes → No

4/6 is correct, 2/6 is error.

→ Considering local inventors as No.

local Inventors	Decision
No	No
No	Yes
No	No
No	Yes
No	No
No	No

∴ local-inventors : No → No

4/6 is correct, 2/6 is error.

Attribute	rule	error	total errors.
local-inventor	Yes → No	2/6	4/12
	No → No	2/6	

→ Considering Local awareness as High.

Local awareness	Decision
High	Yes
High	Yes
High	No
High	Yes
High	No

∴ Local awareness : High → Yes

3/5 is correct, 2/5 is error.

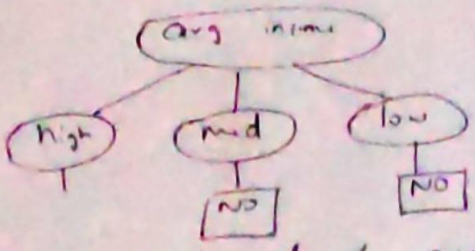
→ Considering Local awareness as low

Local awareness	Decision
Low	No
low	No

∴ Local awareness : low → No

2/2 is correct, 0 is error

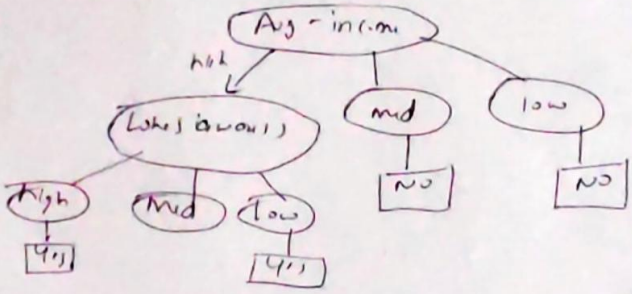
Attribute	rules	error	total error
Local awareness	High → Yes	2/5	3/12
	Mid → No	1/1	
	low → No	0/2	



-> Now consider Avg income as root node & avg income as high.

Avg income	city size	local - investor	Labour-as	Decision
high	Big	Yes	high	Yes
high	Big	No	high	Yes
high	mid	Yes	mid	Yes
high	mid	Yes	Low	NO
high	small	No	high	Yes
high	mid	No	mid	NO

Attributes	Rules	error	total
City-size	Big -> Yes	0/2	1/6
	Mid -> No	1/3	
	Small -> Yes	0/1	
local-investor	Yes -> Yes	1/3	2/6
	No -> Yes	1/3	
Labours	high -> yes	0/3	1/6 ✓
	Mid -> Yes	1/3	
	Small -> No	0/1	

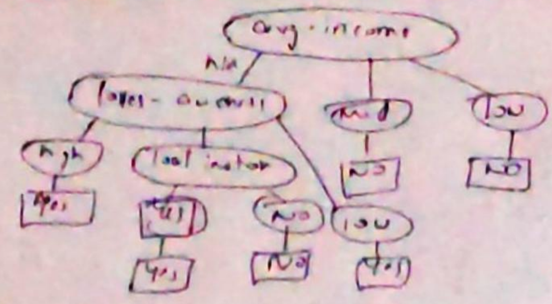


-> Considering Labours - avg as the next node & mid.

Labs - avg	city-size	local investors	Decision
mid	mid	Yes	Yes
mid	mid	No	NO

Attributes	rules	error	total
City - size	mid -> Yes	1/2	1/2
local investor	Yes -> Yes	0/1	0/2 ✓
	No -> No	0/1	

(4)



For city size med, avg-income med, local investor No, taxes amount med, decision No : decision is No.

Step 1: the centroid C1 and C2 will be first 2 values (2,4) as C1 & (4,6).

$Ed = \sqrt{(x_0 - x_c)^2 + (y_0 - y_c)^2}$

Step 2: to calculate ED b/w Rs & Centroid of cluster C1 & C2

Step 3:

Ed for R3 & C1 = $\sqrt{(5-2)^2 + (6-4)^2} = \sqrt{3^2+2^2} = \sqrt{13} = 3.605$

Ed for R3 & C2 = $\sqrt{(5-2)^2 + (6-6)^2} = 3$

R3 belongs to cluster C2 as the distance is less. ∴ Centroid will be.

C1 = (2,4) & C2 = $\left[\frac{5+2}{2}, \frac{6+6}{2}\right]$ = (3.5, 6).

ED for R4 & C1 = $\sqrt{(4-2)^2 + (4-4)^2} = 3.605$

ED for R4 & C2 = $\sqrt{(4-3.5)^2 + (7-6)^2} = 1.118$

∴ Centroid will be C1 = (2,4) & C2 = (3.75, 6.5)

ED for R5 & C1 = $\sqrt{(8-2)^2 + (3-4)^2} = 6.0827$

ED for R5 & C2 = $\sqrt{(8-3.75)^2 + (3-6.5)^2} = 5.5056$

∴ Centroid will be C1 = (2,4) & C2 = (5.875, 4.75)

ED for R6 & C1 = $\sqrt{(6-2)^2 + (6-4)^2} = 4.4721$

ED for R6 & C2 = $\sqrt{(6-5.875)^2 + (6-4.75)^2} = 1.2562$

∴ Centroid value is C1 = (2,4) & C2 = $\left(\frac{6+5.875}{2}, \frac{6+4.75}{2}\right)$ = (5.892, 5.375)

ED for R7 & C1 = $\sqrt{(5-2)^2 + (2-4)^2} = 3.605$

ED for R7 & C2 = $\sqrt{(5-5.892)^2 + (2-5.375)^2} = 3.491$

∴ Centroid C1 = (2,4) & C2 = (5.446, 3.687)

ED for R8 & C1 = $\sqrt{(7-2)^2 + (7-4)^2}$ = 4.242

ED for R8 & C2 = $\sqrt{(8.5-5.446)^2 + (7-3.687)^2}$ = 2.612

∴ centroid C1 = (2,4) & C2 = (5.223, 5.343)

ED for R9 & C1 = $\sqrt{(6-2)^2 + (3-4)^2}$ = 4.123

ED for R9 & C2 = $\sqrt{(6-5.223)^2 + (3-5.343)^2}$ = 2.468

∴ Centroid = C1 = (2,4) & C2 = (5.611, 4.171)

ED for R10 & C1 = $\sqrt{(4-2)^2 + (4-4)^2}$ = 2

ED for R10 & C2 = $\sqrt{(4-5.611)^2 + (4-4.171)^2}$ = 1.17

∴ C1 = (2,4) & C2 = (4.505, 4.085).

C1 = [1] C2 = [2,3,4,5,6,7,8,9,10]

③ S = 33.34% C = 60%

Transaction ID	items
T1	Hotdogs, Buns, ketchup
T2	Hotdogs, Buns
T3	Hotdogs, coke, chips
T4	chips, coke
T5	chips, ketchup
T6	Hotdogs, coke, chips.

1-item set	freq	Support freq.
Hotdogs	4	66.6% ✓
Buns	2	33.33%
Ketchup	2	33.33%
Coke	3	50% ✓
chips	4	66.66% ✓

2-item set	freq.	Support freq.
Hotdog, coke	2	33.33%
Hotdog, chips	2	33.33%
Coke, chips	3	50%

Rule generated is

[coke, chips]

		Confidence	Support
coke → chips	= 3/3	= 100%	3/6 = 50%
chips → coke	= 3/4	= 75%	3/6 = 50%

(1) S = 30% C = 70%

ID	items
1	Bread, cheese, egg, juice.
2	Bread, cheese, juice.
3	Bread, milk, yogurt
4	Bread, juice, milk.
5	cheese, juice, milk.

1-item set

1-item set	freq.	S-ppd freq.
Bread	4	80% ✓
cheese	3	60% ✓
egg	1	20% ✗
juice	4	80% ✓
milk	3	60% ✓
yogurt	1	20% ✗

2-item set	freq.	Support - freq
Bread - cheese	2	40%
Bread, juice	3	60% ✓
Bread, milk.	2	40%
cheese, juice	3	60% ✓
cheese, milk	1	20%
juice, milk	2	40%

3-item set	freq	Suppd-freq
Bread, juice cheese	2	40% ✗

No 3-item set will generate rule, hence.

Rule 1: (Bread, juice)

		Confidence	S-ppd
Bread → juice	=	75%	60%
juice → Bread	=	75%	60%

Rule 2: (cheese, juice)

			S-ppd
cheese → juice	3/3	= 100%	60%
juice → cheese	3/4	= 75%	60%

(5) The probability of stmt is $p(s) = 4/7 = p(h)$

Probability of quest is $p(q) = 4/7 = p(\neg h)$

No of words in stmt are 17

No of words in quest are 19

Unique words are 22

No of data = 7

No of class = 2

for class 'h' stmt ..

$\alpha = 1$

$d = 22$

$o_i = \frac{n_i + \alpha}{N + \alpha d}$

$p(\text{what}/h) = \frac{0+1}{17+22} = \frac{1}{39}$

$p(\text{what}/\neg h) = \left(\frac{2+1}{17+22}\right) = 3/40$

$p(\text{is}/h) = \frac{3}{39}$

$p(\text{is}/\neg h) = 2/40$

$p(\text{the}/h) = \frac{2}{39}$

$p(\text{the}/\neg h) = 3/40$

$p(\text{price}/h) = 1/39$

$p(\text{price}/\neg h) = \frac{1}{40}$

$p(\text{of}/h) = 1/39$

$p(\text{of}/\neg h) = 1/40$

$p(\text{work}/h) = 3/39$

$p(\text{work}/\neg h) = 3/40$

$p(\text{this data}/h) = 4/7 \times 1/39 \times 3/39 \times (2/39)^2 * 1/39 \times 1/39 \times 3/35 = 1.49703655 e^{-11}$

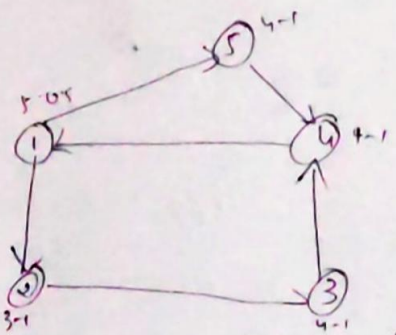
$= 0.000000000014$

$p(\text{test data}/\neg h) = 4/7 \times 3/40 \times 2/40 \times (3/40)^2 \times (1/40) \times 1/40 \times 3/40$

$= 5.650111616 e^{-10} = 0.000000000565$

$\therefore$ the test data " what is the price of Bath belongs to the class question.

(6)



Step 1:

Variable	initial value
R_1	0.2
R_2	0.2
R_3	0.2
R_4	0.2
R_5	0.2

Step 2: only inlink equation.

$R_1 = R_4$

$R_2 = 0.5 R_1$

$R_3 = R_2$

$R_4 = R_5 + R_6$

$R_5 = 0.5 R_1$

Initial matrix based on R_1, R_2, R_3, R_4, R_5 equation.

	R_1	R_2	R_3	R_4	R_5
R_1	0	0	0	1	0
R_2	0.5	0	0	0	0
R_3	0	1	0	0	0
R_4	0	0	1	0	1
R_5	0.5	0	0	0	0

Step 3:

Var	init val	I_1	I_2	I_3	I_4	I_5
R_1	0.2	0.2	0.4	0.3	0.2	0.3
R_2	0.2	0.1	0.1	0.2	0.15	0.1
R_3	0.2	0.2	0.1	0.1	0.2	0.15
R_4	0.2	0.4	0.3	0.2	0.3	0.35
R_5	0.2	0.1	0.1	0.2	0.15	0.1

	I_6	I_7	I_8	I_9	I_{10}	I_{11}
R_1	0.25	0.26	0.26	0.325	0.3	0.275	0.287
R_2	0.15	0.175	0.196	0.125	0.162	0.15	0.12
R_3	0.1	0.15	0.17	0.125	0.162	0.162	0.11
R_4	0.25	0.26	0.326	0.3	0.25	0.287	0.31
R_5	0.1	0.125	0.125	0.125	0.1625	0.15	0.12

	I_{12}	I_{13}	I_{14}
R_1	0.31	0.27	0.26
R_2	0.14	0.15	0.13
R_3	0.12	0.14	0.15
R_4	0.27	0.26	0.29
R_5	0.14	0.15	0.13

Ranking is follows

{4, 1, 3, 2, 5} or

{4, 1, 3, 5, 2}

∴ Highest rank value is 0.29 of R_4.