Decision Tree

Q1. Use the following dataset. Create classification model using decision-tree and hence classify the following tuple-

Tid	Income	Age	Own House
1.	very high	young	Yes
2.	high	medium	Yes
3.	low	young	Rented
4.	high	medium	Yes
5.	very high	medium	Yes
6.	Medium	young	Yes
7.	High	old	Yes

Tid.	Income	Age	Own House
8.	medium	medium	rented
9.	low	medium	rented
10.	low	old	rented
11.	High	young	Yes
12.	Medium	old	rented

Ans. Class P :- own house = "Yes"

class N :- own house = "rented"

total no. of records = 12.

no. of records with "Yes" = 7
[Class P]

no. of records with "No" = 5
[class N]

Calculating information gain

$$I(P, N) = \frac{-P}{P+N} \log_2 \frac{P}{P+N} - \frac{N}{P+N} \log_2 \frac{N}{P+N}$$

$$I(7, 5) = \frac{-\cancel{7}}{5+7} \log_2 \frac{7}{7+5} - \frac{5}{7+5} \log_2 \frac{5}{7+5}$$

$$= \left(\frac{-7}{12}\right)\left(\log_2 \frac{7}{12}\right) - \left(\frac{5}{12}\right)\left(\log_2 \frac{5}{12}\right)$$

$$= +0.4536 + 0.5262$$

$$= \boxed{0.9798}$$

Entropy :- Income

Income	P	N	I(P, N)
very high	2	0	0
High	4	0	0
Medium	1	2	0.9182
Low	0	3	0

$$I(2,0) = \frac{-2}{2+0} \log_2 \frac{2}{2+0} - 0 = -1\log_2 1 = \boxed{0}$$

$$I(1,2) = \frac{-1}{1+2} \log_2 \frac{1}{1+2} - \frac{2}{1+2} \log_2 \frac{2}{1+2}$$

$$= \frac{-1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}$$

$$= 0.5283 + 0.3899$$

$$= \boxed{0.9182}$$

$$E(\text{Income}) = \sum_{i=1}^{v} \frac{P_i + N_i}{P+N} . I(P_i, N_i)$$

$$= \frac{2+0}{7+5} . I(2,0) + \frac{4+0}{7+5} . I(4,0) + \frac{1+2}{7+5} . I(1,2) + \frac{0+3}{7+5} I($$

$$= \frac{2}{12} I(2,0) + \frac{4}{12} I(4,0) + \frac{3}{12} I(1,2) + \frac{3}{12} I(0,3)$$

$$= 0 + 0 + \frac{3}{12} \times 0.9182 + 0$$

$$= \boxed{0.2295}$$

Gain = I(P, N) − E(Income)

$$= 0.9798 - 0.2295$$

$$= \boxed{0.7503}$$

Entropy :- Age

Age	P_i	N_i	$I(P_i, N_i)$
Young	3	1	0.8112
Medium	3	2	0.9712
Old	1	2	0.9182

$$I(3,1) = \frac{-3}{3+1}\log_2\frac{3}{3+1} - \frac{1}{3+1}\log_2\frac{1}{3+1}$$

$$= \frac{-3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 0.3112 + 0.5$$

$$= \boxed{0.8112}$$

$$I(1,2) = \frac{-1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$$

$$= \cancel{0.5 +} \boxed{0.9182}$$

$$E(Age) = \sum_{i=1}^{v} \frac{P_i + N_i}{P+N} \cdot (I(P_i, N_i))$$

$$= \frac{4}{12}(0.8112) + \frac{5}{12}(0.9712) + \frac{3}{12}(0.9182)$$

$$= \boxed{0.9044}$$

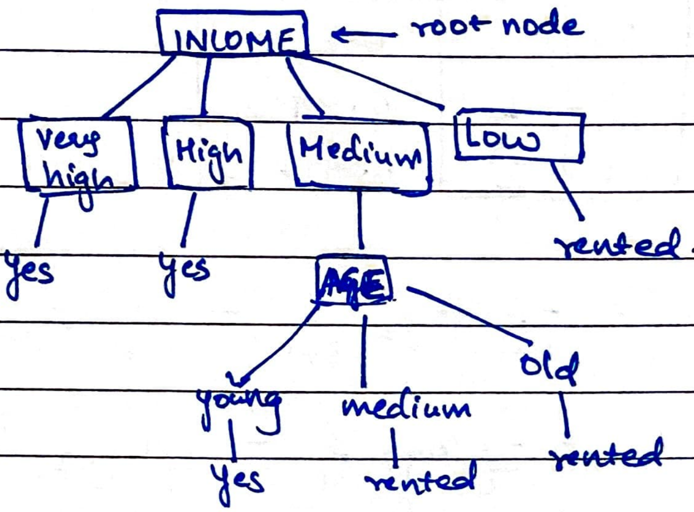
$$Gain = I(P,N) - E(Age)$$

$$= 0.9798 - 0.9044$$

$$= \boxed{0.0754}$$

Ans-

Income attribute has highest gain, therefore it is used as root node.



Q2. A sample training dataset for stock market is given. Profit is the class attribute and value is based on age, contest and type.

Age	Contest	Type	Profit.
old	Yes	Swr	down
old	No	Swr	down
old	No	Hwr	down
mid	Yes	Swr	down
mid	Yes	Hwr	down
mid	No	Hwr	up
mid	No	Swr	up
new	Yes	Swr	up
new	No	Hwr	up
new	No	Swr	up

Class P = up

Class N = down

total no. of records = 10

class P = 5

class N = 5

$$I(P,N) = -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} =$$

$$= -0.5 - 0.5 = \boxed{+1}$$

Entropy :- age -

Age	P	N	I(P,N)
old	0	3	0
mid	2	2	1
new	3	0	0

$$\cancel{gain} \; E(age) = \frac{2+2}{10} \times 1 = \boxed{0.4}$$

$$Gain = 1 - 0.4 = \boxed{0.6}$$

Entropy - contest -

contest	P	N	I(P,N)
Yes	1	3	0.8112.
No	4	2	0.9182

$$I(1,3) = -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{3}{4}\log_2\left(\frac{3}{4}\right)$$

$$= 0.5 + 0.3112 = \boxed{0.8112}$$

$$I(4,2) = -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)$$

$$= 0.3899 + 0.5283 =$$

$$= \boxed{0.9182}$$

$$E(contest) = \frac{4}{10} \times 0.8112 + \frac{6}{10} \times 0.9182 =$$

$$= \boxed{0.8753}$$

0.3244 + 0.5509

$\boxed{Gain = 0.1247}$

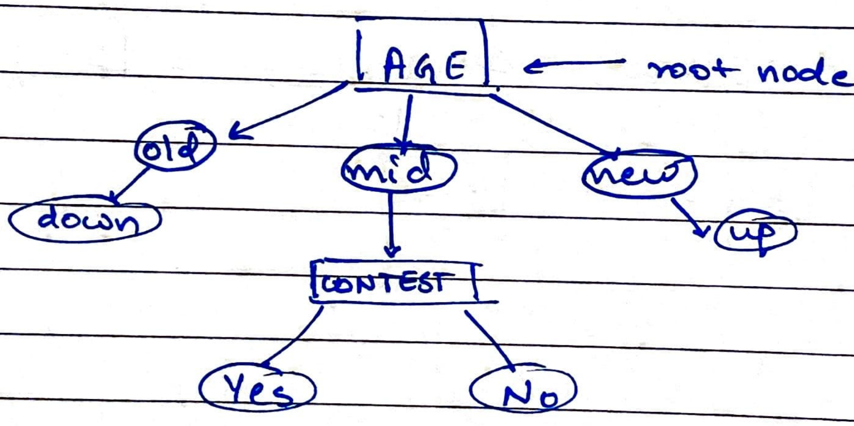
~~Typ~~

Type	P	N	I(P,N)
Swr	3	3	1
Hwr	2	2	1

$$E(typ) = \frac{6}{10} \times 1 + \frac{4}{10} \times 1 = \boxed{1}$$

$\boxed{\text{gain} = 0}$

gain(age) = 0.6 → highest = node.
gain(contest) = 0.1247
gain(type) = 0.



(Contest because gain of contest is higher)

Q3 Training set for play-tennis-

Outlook	Temperature	Humidity	Windy	Case
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	NO

~~Sunny~~ class P - ~~sunny~~ - 'Yes'

class N - ~~sunny~~ - 'No'

total no. = ~~8~~ 14

no. of Yes = ~~2~~ 9

no. of No = ~~3~~ 5

$I(P,N) = I(9,5) = \cancel{-\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}}$

$= -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}$

$0.4097 + 0.5305$

$= \boxed{0.9402}$

Entropy - outlook -

Type	P	N	I(P,N)
sunny	2	3	0.9708
overcast	4	0	0
rain	3	2	0.9708

$I(2,3) = 0.5287 + 0.4421 = \boxed{0.9708}$

$E(\text{outlook}) = \frac{2+3}{14} \times 0.9708 + \frac{3+2}{14} \times 0.9708$

$= \boxed{0.6934}$

$\text{Gain}(\text{outlook}) = 0.9402 - 0.6934$

$= \boxed{0.2468}$

Entropy - temperature

Type	P	N	I(P, N)
hot	2	2	1
mild	4	2	0.3899 + 0.5283 = 0.9182
cool	3	1	0.3112 + 0.5 = 0.8112

$E(\text{temperature}) = \frac{2+2}{14} \times 1 + \frac{4+2}{14} \times 0.9182 + \frac{3+1}{14} \times 0.8112$

$= 0.2857 + 0.3935 + 0.2317$

$= \boxed{0.9109}$

$\text{Gain}(\text{temperature}) = 0.9402 - 0.9109 =$

$= \boxed{0.0293}$

Entropy - humidity

type	P	N	I(P,N)	
high	3	4	0.5238 + 0.4618	≈ 0.9851
normal	6	1	0.1906 + 0.2857	= 0.4763

$$E(\text{humidity}) = \frac{7}{14} \times 0.9941 + \frac{7}{14} \times 0.4763$$

$$\approx 0.4970 + 0.2381 = \boxed{0.7351}$$

$$\text{Gain}(\text{humidity}) = 0.9402 - 0.7351$$

$$= \boxed{\cancel{0.2051}} \quad \boxed{0.151}$$

0.0481

Entropy - windy

type	P	N	I(P,N)
True	3	3	1
False	6	2	0.3112 + 0.5 = 0.8112

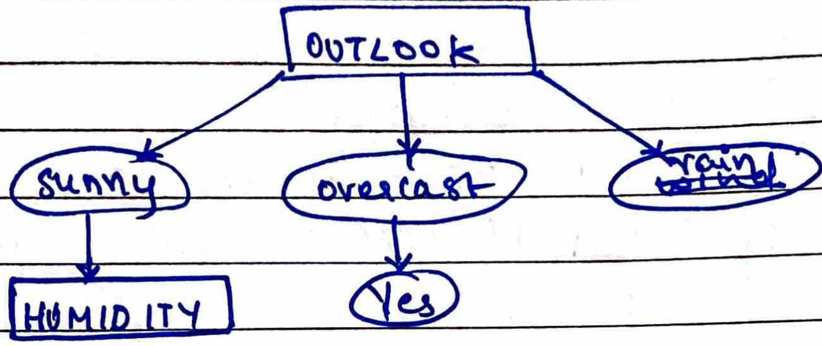
$$E(\text{windy}) = \frac{3+3}{14} \times 1 + \frac{6+2}{14} \times 0.8112$$

$$\approx 0.4285 + 0.4635 = \boxed{0.892}$$

$$\text{Gain}(\text{windy}) = 0.9402 - 0.8920$$

$$\approx \boxed{0.0482}$$

Gain	value
outlook	0.2468 ✓
temperature	0.0293
humidity	0.151
wind	0.0482



outlook (sunny) ~~sunny~~	temp.	humidity	windy	case
1	hot	high	false	no
2	hot	high	true	no
3	mild ~~I (P, N high~~	~~false~~ ~~(3, 2)~~		no
4	cool	normal	false	yes
5	mild	normal	true	yes

$I(P,N) = (2,3) = 0.5287 + 0.4421 = \underline{0.9708}$

~~Gain =~~

Entropy (temp) =

	P	N	I(P, N)
hot	0	2	0
mild	1	1	1
cool	1	0	0

$= \frac{1+1}{5} \times 1 = \boxed{0.4}$

Gain $= 0.9708 - 0.4 = \underline{0.5708}$ ~~= 0.8281~~

Entropy (humidity)

high	0	3	0
normal	2	0	0

entropy = 0

gain = $\boxed{0.9708}$ ~~0.7142~~

Entropy (windy)

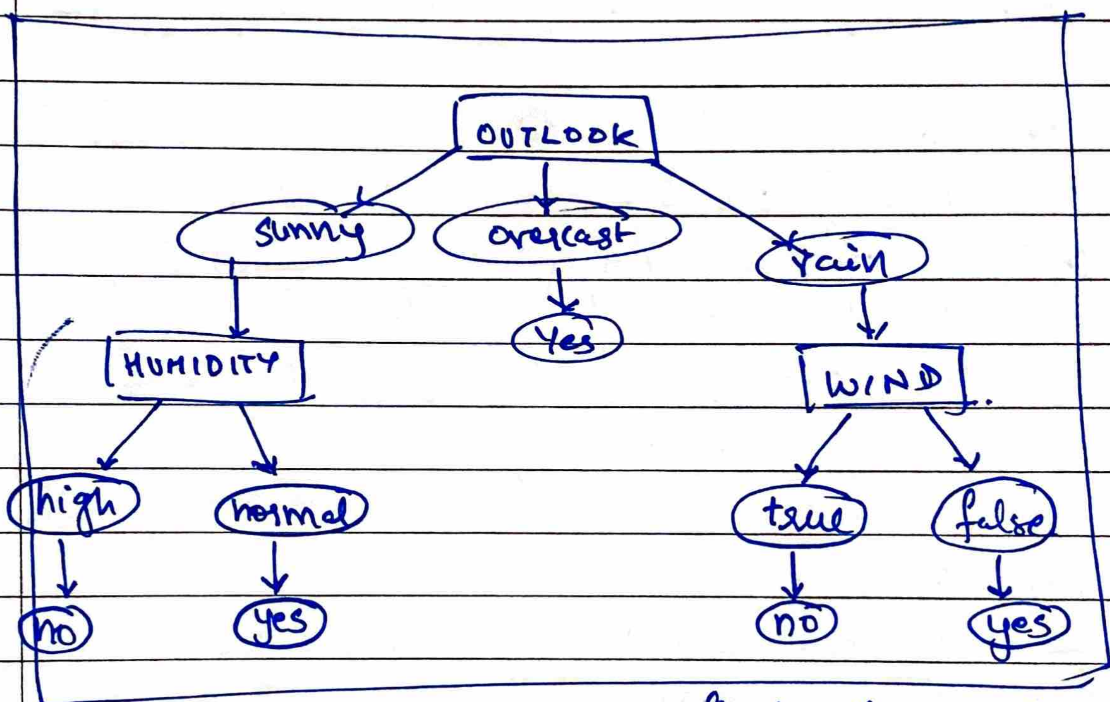
True	1	1	1
False	1	2	$0.5283 + 0.3899 = 0.9182$

$Ent = \frac{2}{5} \times 1 + \frac{3}{5} \times 0.9182 =$

$0.4 + 0.5509 = \underline{0.9509}$

$Gain = 0.9708 - 0.9509 = \boxed{0.0199}$

Gain	value
temp	0.5708
humidity	0.9708 ✓
wind	0.0199



∴ temperature does not affect the outcome.
If outlook is sunny and humidity is high / outlook is rain and wind is true, decision is always NO, rest is YES.

Rain	temp	wind	case
1	mild	f	y
2	cool	f	y
3	cool	t	n
4	mild	f	y
5	mild	t	n

Entropy = $I(3,2) = \boxed{0.9708}$

Entropy temp

mild	2	1	0.9182
cool	1	1	1

$E = 0.5509 + 0.4 = \underline{0.9509}$

$G = \boxed{0.0199}$

entropy wind ✓

false	3	0	0
true	0	2	0

$E = 0$

$\boxed{G = 0.9708}$

Q4 Apply ID3 on the following training dataset from all electronic customer database and extract the classification rule from the tree.

~~Age Income Student~~

Class P = buys computer = "Yes"

Class N = buys computer = "No"

no. of P = 9

no. of N = 5

$I(P,N) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$

$= 0.40977 + 0.5305$

$= \boxed{0.94028}$

Entropy - Age.

Value	P	N	I(P,N)
<=30	2	3	0.9709
31---40	4	0	0
>40	3	2	0.9709

$I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.52877 + 0.44217$

$= 0.9709$

$E(Age) = \frac{2+3}{14} \times 0.9709 + \frac{3+2}{14} \times 0.9709$

$= \boxed{0.6935}$

$Gain(Age) = 0.9403 - 0.6935$

$= \boxed{0.2468}$

__/__/__

Entropy - Income

Value	P	N	I(P,N)
low	3	1	0.8113
medium	4	2	0.9183
high	2	2	1

$I(3,1) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 0.3113 + 0.5$
$= \boxed{0.8113}$

$I(4,2) = -\frac{4}{6}\log_2\left(\frac{4}{8}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) = 0.389975 + 0.528$
$= \boxed{0.9183}$

$E(\text{Income}) = \left[\frac{3+1}{14} \times 0.813 + \frac{4+2}{14} \times 0.9183 + \frac{2+2}{14} \times 1\right]$

$= 0.2318 + 0.39360 + 0.2857$

$= \boxed{0.9111}$

$\text{Gain(income)} = 0.9403 - 0.9111 = \boxed{0.0292}$

Entropy - Student.

Value	P	N	I(P,N)
Yes	6	1	0.59165
No	3	4	0.9852

$I(6,1) = -\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right) = 0.1906 + 0.40105$
$= \boxed{0.59165}$

$I(3,4) = 0.5239 + 0.4613 = \boxed{0.9852}$

$E(\text{student}) = \frac{6+1}{14} \times 0.59165 + \frac{3+4}{14} \times 0.9852$

$= 0.2958 + 0.4926 = \boxed{0.7884}$

$\text{Gain(student)} = \boxed{0.1519}$

Entropy - rating

value	P	N	I(P,N)
Fair	6	2	0.8113
Excellent	3	3	1

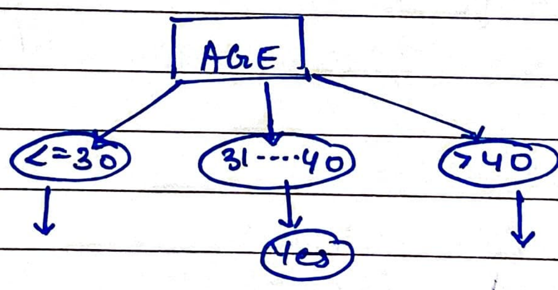
I (6,2) = 0.3113 + 0.5

$E(\text{rating}) = \frac{6+2}{14} \times 0.8113 + \frac{3+3}{14} \times 1$

$= 0.4636 + 0.4286 = \boxed{0.8922}$

$\text{Gain}(\text{rating}) = 0.9403 - 0.8922 = \boxed{0.0481}$

∴ Gain (Age) = highest. = 0.2468



Age <=30.

	income	student	rating.	class
1	high	No	fair	No
2	high	No	excellent	No
3	medium	No	fair	No
4	low	Yes	fair	Yes
5	medium	Yes	excellent	Yes

Entropy - income

value	P	N	I(P,N)
high	0	2	0
medium	1	1	1
low	1	0	0

$\therefore E(\text{Income}) = \frac{1+1}{14} \times 1 = \boxed{0.14285}$

$\text{Gain}(\text{income}) = \boxed{0.7975}$

Entropy(student)

value	P	N	I(P, N)
Yes	2	0	0
No	0	3	0

E(student) = 0.

Gain (student) = $\boxed{0.9403}$

Entropy(rating)

value	P	N	I(P, N)
fair	1	2	0.9183
excellent	1	1	1

$I(1,2) = \cancel{\frac{1}{3}\log}\ 0.5283 + 0.38997 = 0.9183$

$E(\text{rating}) = \frac{1+2}{14} \times 0.9183 + \frac{2}{14} \times 1$

$= 0.1968 + 0.14286$

$= \boxed{0.3397}$

Gain (student) = highest

> 40

	income	rating	class
1	~~high~~ ~~low~~ medium	fair	Yes
2	~~medium~~ low	fair	Yes
3	low	excellent	No
4	medium	fair	Yes
5	medium	excellent	No

Entropy (income)

	P	N	I(P, N)
low	1	1	1
medium	2	1	0.9183

0.38997 + 0.5283

=

$\boxed{\text{Gain} = 0.6006}$

$E(\text{income}) = \frac{2}{14} \times 1 + \frac{3}{14} \times 0.9183$

$= 0.1429 + 0.1968$

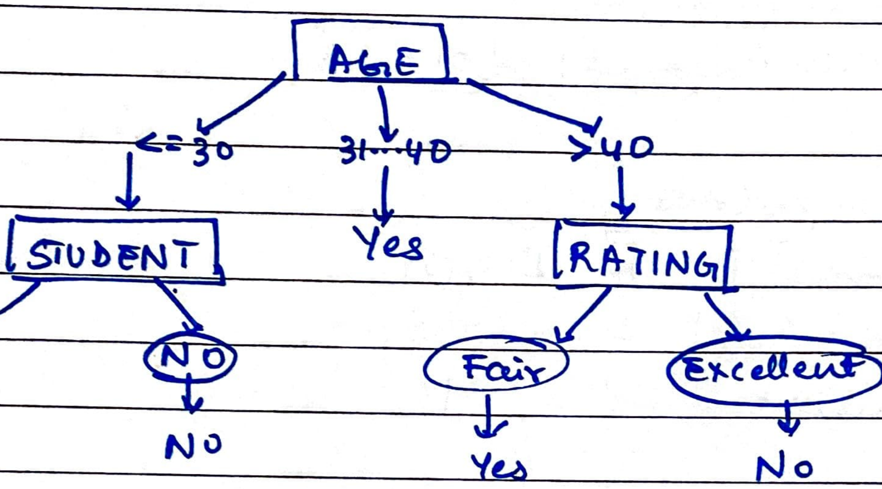
$= \boxed{0.3397}$

Entropy rating

	P	N	I(P,N)
fair	3	0	0
excellent	0	2	0

Gain = 0.9403

∴ Gain(rating) = highest.



95

Class 'N' = shape = Triangle = no.

Class 'P' = shape = Square = yes

total no. = 14

$I(P,N) = I(9,5) = 0.4098 + 0.5305 = \boxed{0.9403}$

Entropy - colour.

value	P	N	I(P,N)
Green	2	3	0.9709
Yellow	4	0	0
Red	3	2	0.9709

$\therefore E(colour) = \frac{2+3}{14} \times 0.9709 + \frac{3+2}{14} \times 0.9$

$= 0.6935$

$Gain(colour) = \boxed{0.2468}$

Entropy - outline

value	P	N	I(P,N)
dashed	3	4	0.9852
solid	6	1	0.5917

$\therefore E(outline) = \frac{7}{14} \times 0.9852 + \frac{7}{14} \times 0.59$

$= 0.78845$

$Gain(outline) = \boxed{0.15185}$

Entropy - dot

value	P	N	I(P,N)
Yes	3	3	1
No	6	2	0.8113

$E(dot) = \frac{6}{14} \times 1 + \frac{8}{14} \times 0.8113$

$= 0.4286 + 0.4636$

$= 0.8922$

$\therefore$ highest gain = colour = 0.2468

$Gain(dot) = \boxed{0.6481}$

	outline	dot	shape
1.	dash	no	T
2	dash	yes	T
3	solid	no	S
4	dash	no	T
5	solid	yes	S

Entropy (outline) -

value	P	N	I(P,N)
dash	0	3	0
solid	2	0	0

$\boxed{Gain = 0.9403}$

highest gain = ~~dot~~ outline.

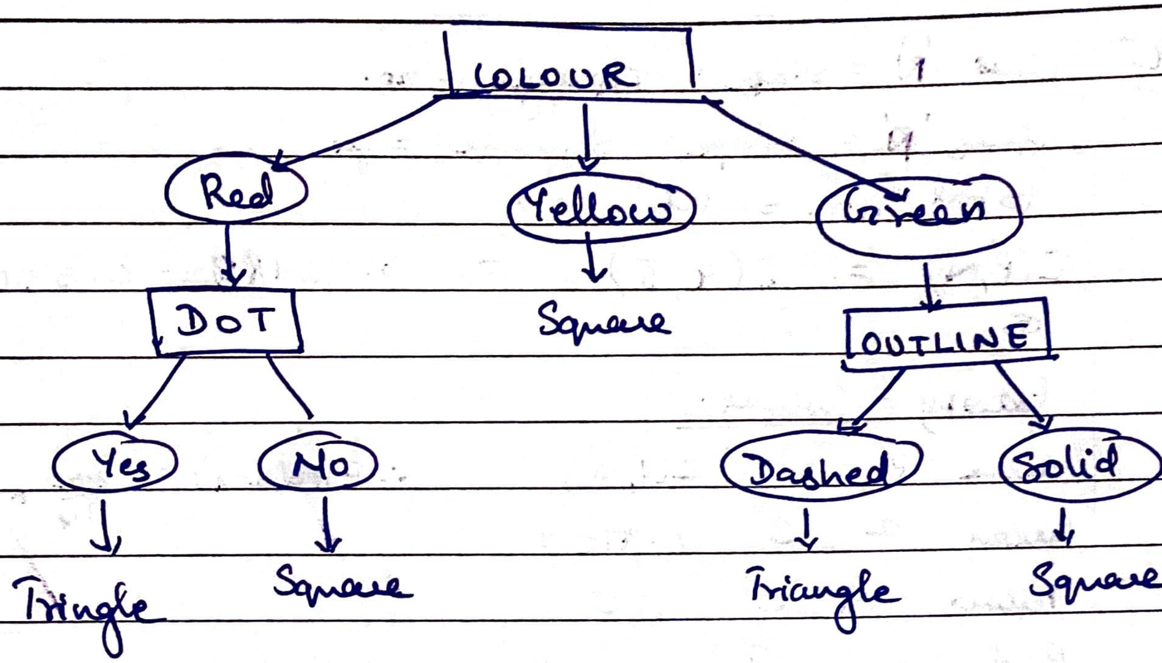
Entropy (dot) -

value	P	N	I(P,N)
Yes	1	1	1
no	1	2	0.9183

$E(dot) = \frac{2}{5} \times 1 + \frac{3}{5} \times 0.9183$

$= 0.4 + 0.1968$

$= \boxed{0.6006}$


COLOUR
Red
Yellow
Green
DOT
Square
OUTLINE
Yes
No
Dashed
Solid
Tringle
Square
Triangle
Square

14/9/20

Data Mining Using Frequent Item Sets (Apriori Algorithm)

* Five transactions are given below-

1. Minimum support = 30%
2. Minimum confidence = 80%

Determine frequent item-sets

Transaction	Items
T1	bread, jelly, butter
T2	bread, butter
T3	bread, milk, butter
T4	coke, bread
T5	coke, milk

Soln-

item	count	
bread	4	
jelly	1	×
butter	3	
milk	2	
coke	2	

minimum support count $= \frac{30}{100} \times 5 = 1.5 = \boxed{2}$.

∵ count of jelly < minimum support count - it is eliminated.

Item sets (pairs)	count
bread, butter	3
bread, milk	1
bread, coke	1
butter, milk	1
butter, coke	0
milk, coke	1

minimum support count ≥ 2
= bread, butter (3).

Frequent - 2 items	count
bread, butter	3

Association rule-	support	confidence	confidence %
bread → butter	3	$3/4 = 0.75$	75%
butter → bread	3	$3/3 = 1$	100%

∴ final rule = butter → bread. (100% possibility of buying bread, if you have bought butter).

Q2. A database has 4 transactions.

Minimum support = 60%.

Minimum confidence = 80%.

TID	items bought
T_1	K, A, B, D
T_2	D, A, C, E, B
T_3	C, A, B, E
T_4	B, A, D

frequency

Items	frequency (count)
A	4
B	4
C	2
D	3
E	2
K	1

minimum support count $= \frac{60}{100} \times 4$

$= \cancel{36 \approx 4.} = 2.4 \approx 3$

Only taking items with count ≥ 3.

~~∴ Only A & B are considered.~~ ∴ Only A, B, D are considered

Item sets (pairs)	count
A, B	4

frequent - 2 items	count
A, B	4

item sets (pairs)	count
A, B	4
A, D	3
B, D	3

≥ 3

frequent - 2 items	count
A, B	4
A, D	3
B, D	3

~~Association rule support confidence confidence %.~~

3 items - item set	count
A, B, D	3

- final ~~table~~

confidence = $\frac{\text{support}}{\text{no. of occurrences of LHS items}}$

Association rule	support	confidence	confidence %
$A \rightarrow B \cap D$	3	$3/4 = 0.75$	75 %
$B \rightarrow A \cap D$	3	$3/4 = 0.75$	75 %
$D \rightarrow A \cap B$	3	$3/3 = 1$	100 %
$B \cap D \rightarrow A$	3	$3/3 = 1$	100 %
$A \cap D \rightarrow B$	3	$3/3 = 1$	100 %
$A \cap B \rightarrow D$	3	$3/4 = 0.75$	75 %

ii)

minimum confidence = 80 %.

∴ These three rules hold true [$D \rightarrow A \cap B$, $B \cap D \rightarrow A$, $A \cap D \rightarrow B$].

Q3

Q1 Consider the transaction database given below. Use Apriori algorithm with minimum support count 2. Generate the association rules along with its confidence. (Minimum confidence = 75%)

Ans.

TID	List of item IDs
T100	I_1, I_2, I_5
T200	I_2, I_4
T300	I_2, I_3
T400	I_1, I_2, I_4
T500	I_1, I_3
T600	I_2, I_3
T700	I_1, I_3
T800	I_1, I_2, I_3, I_5
T900	I_1, I_2, I_3

Items	count
I_1	6
I_2	7
I_3	6
I_4	2
I_5	2

minimum support count = 2

∴ only taking items with count ≥ 2

∴ All are considered.

item sets (pairs)	count
I_1, I_2	4
I_1, I_3	4
I_1, I_4	1
I_1, I_5	2
I_2, I_3	4
I_2, I_4	2
I_2, I_5	2
I_3, I_4	0
I_3, I_5	1
I_4, I_5	0

with count ≥ 2

frequent-2items	count
I_1, I_2	4
I_1, I_3	4
I_1, I_5	2
I_2, I_3	4
I_2, I_4	2
I_2, I_5	2

eliminating pairs with count ≥ 2.

~~I_1, I_2, I_3~~

~~I_1, I_2, I_4~~

~~I_1, I_2, I_5~~

~~I_2, I_3, I_4~~

~~I_2, I_3, I_5~~

~~I_3, I_4, I_5~~

with count ≥ 2

Item sets (3)	count
✓ I_1, I_2, I_3	2
✓ I_1, I_2, I_4	1
✓ I_1, I_2, I_5	2
✓ ~~I_2, I_3, I_5~~	~~1~~
✓ I_2, I_3, I_4	0
✓ I_2, I_4, I_5	0
~~I_1, I_3, I_4~~	~~0~~
~~I_3, I_4, I_5~~	~~0~~
~~I_2, I_3, I_5~~	~~1~~
~~I_1, I_4, I_5~~	~~0~~

frequent sets	count
I_1, I_2, I_3	2
I_1, I_2, I_5	2

$\therefore$ only I_1, I_2, I_3 and $I_1, I_2\ I_5$ are considered

Item sets (4)	count
I_1, I_2, I_3, I_5	1

$\because$ count < 2 $\therefore$ it is not considered.

Association rules	support	confidence	confidence %
$I_1 \rightarrow I_2 \cap I_3$	2	$2/6 = 0.3333$	33.33 %
$I_2 \rightarrow I_1 \cap I_3$	2	$2/7 = 0.2857$	28.57 %
$I_3 \rightarrow I_1 \cap I_2$	2	$2/6 = 0.3333$	33.33 %
$I_2 \cap I_3 \rightarrow I_1$	2	$2/4 = 0.5$	50 %
$I_1 \cap I_3 \rightarrow I_2$	2	$2/4 = 0.5$	50 %
$I_2 \cap I_1 \rightarrow I_3$	2	$2/4 = 0.5$	50 %

Association rule	support	confidence	confidence %
$I_1 \rightarrow I_2 \cap I_5$	2	$2/6 = 0.3333$	33.33%
$I_2 \rightarrow I_1 \cap I_5$	2	$2/7 = 0.2857$	28.57%
$I_5 \rightarrow I_1 \cap I_2$	2	$2/2 = 1$	100%
$I_2 \cap I_5 \rightarrow I_1$	2	$2/2 = 1$	100%
$I_1 \cap I_5 \rightarrow I_2$	2	$2/2 = 1$	100%
$I_1 \cap I_2 \rightarrow I_5$	2	$2/4 = 0.5$	50%

$\therefore$ minimum confidence = 75%

The rules hold true for - $I_5 \rightarrow I_1 \cap I_2$, $I_2 \cap I_5 \rightarrow I_1$, $I_1 \cap I_5 \rightarrow I_2$

Q1

Transaction ID	Items
01	A, B, C, D
02	A, B, C, D, E, G
03	A, C, G, H, K
04	B, C, D, E, K
05	D, E, F, H, L
06	A, B, C, D, L
07	B, I, E, K, L
08	A, B, D, E, K
09	A, E, F, H, L
10	B, C, D, F

minimum support count = 30% = $\frac{30}{100} \times$ no. of transactions

$= \frac{30}{100} \times 10 = \boxed{3}$

minimum confidence = ~~80%~~ 70%

Item set -1	count
A	6
B	7
C	6
D	7
E	6
F	3
G	2
H	3
I	1
K	4
L	4

b

item set-2	count
AB	4
AC	4
AD	4
AE	3
AF	1
AH	2
AK	2
AL	2
BC	5
BD	6
BE	4
BF	1
BH	0
BK	3
BL	2
CD	5
CE	2
CF	1
CH	1
CK	2
CL	1
DE	4
DF	2
DH	1
DK	2
DL	2
EF	2
EH	2
EK	3
EL	3
FH	2

item set-2	count
FK	0
FL	2
HK	1
HL	2
KL	1

minimum support count = 3

frequent item sets (2)	count
AB	4
AC	4
AD	4
AE	3
BC	5
BD	6
BE	4
BK	3
CD	5
DE	4
EK	3
EL	3

after eliminating sets with count < 3

Item sets - 3	count
ABC	3
ABD	4
ABE	2
ABK	1
ABL	1
ACD	3
ACE	1
ACK	1
ACL	1
ADE	1
ADK	1
ADL	1
AEK	1
AEL	

Item sets - 3	count
AKL	0
BCD	5
BCE	2
BCK	1
BCL	1
BDE	3
BDK	2
BDL	1
BEK	3
BEL	1
BKL	1
CDE	2
CDK	1
CDL	1

__/__/____

item sets - 3	count
CKL	0
DEK	2
DKL	0
EKL	1

~~freq~~

∴ minimum support count = 3 ∴ rejecting all < 3.

∴

frequent item sets - 3	count
ABC	3
ABD	4
ACD	3
BCD	5
BDE	5
BEK	3

item sets - 4	count
ABCD	3
ABDE	2
BDCE	2
BEDK	1

frequent itemsets - 4	count
ABCD	3

↓

only selecting ones with count ≥ 3

~~Association rule~~ ~~support count~~ ~~conditions~~ confid

Association rule	~~confidence~~ support count	confidence	% confidence
A → BCD	3	3/6	50 %
B → ACD	3	3/7	42.85 %
C → ABD	3	3/6	50 %
D → ABC	3	3/7	42.85 %
BCD → A	3	3/5	60 %
ACD → B	3	3/3	100 %
ABD → C	3	3/4	75 %
ABC → D	3	3/3	100 %
AB → CD	3	3/4	75 %
AC → BD	3	3/4	75 %
AD → BC	3	3/4	75 %
BC → AD	3	3/5	60 %
BD → AC	3	3/6	50 %
DC → AC	3	3/6	60 %

Since minimum confidence = ~~[illegible]~~ 70 %.

final selections =

ACB → B	100 %
ABD → C	75 %
ABC → D	100 %
AB → CD	75 %
AC → BD	75 %
AD → BC	75 %