KUSHAGRA AGARWAL 2018113012 (1)

FIITJEE

Dataset assigned → 3.

Rows 2 and 4 are flipped,

Angle	Distance	Speed	Kill
1.5	450	220	N
4.5	520	-120	N
3	490	120	Y
5.5	530	117	Y
3.2	470	-170	N
5.2	505	-90	Y
1.85	465	120	Y
4.8	517	147	Y
1.7	430	-100.	Y

Decision Number 1

First, we calculate starting entropy

$$I(3,6) = \frac{3}{9}\log_2(9/3) + \frac{6}{9}\log(9/6)$$

$$= \boxed{0.918295}.$$

As the columns have continuous data, we opted for a trial and error method. The choices were as follows →

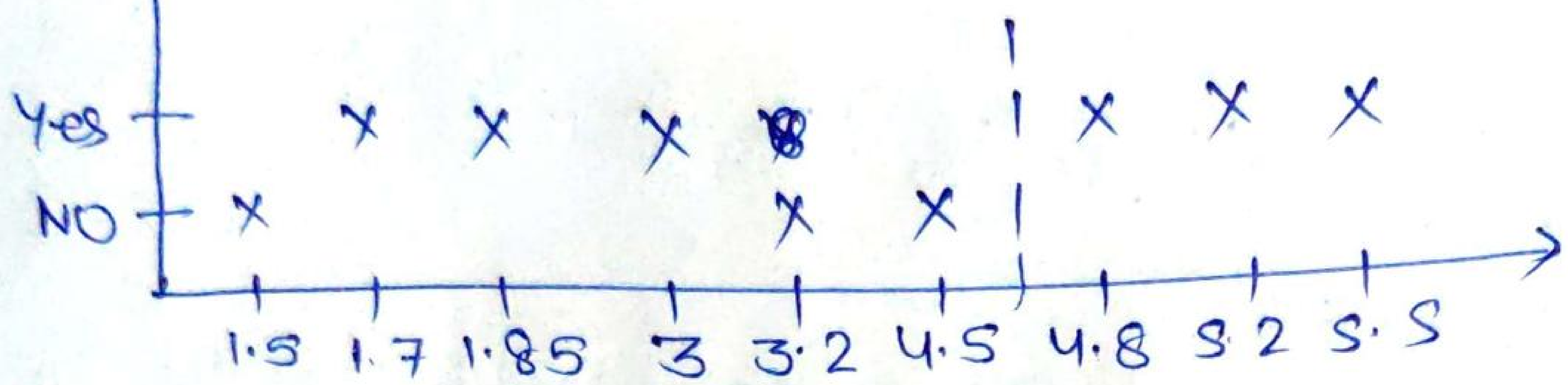
i> Mean of Angle ii> Mean of distance iii> Mean of speed.

iv> Angle ≥ 4.8 v> ~~Speed~~ Speed < -100.

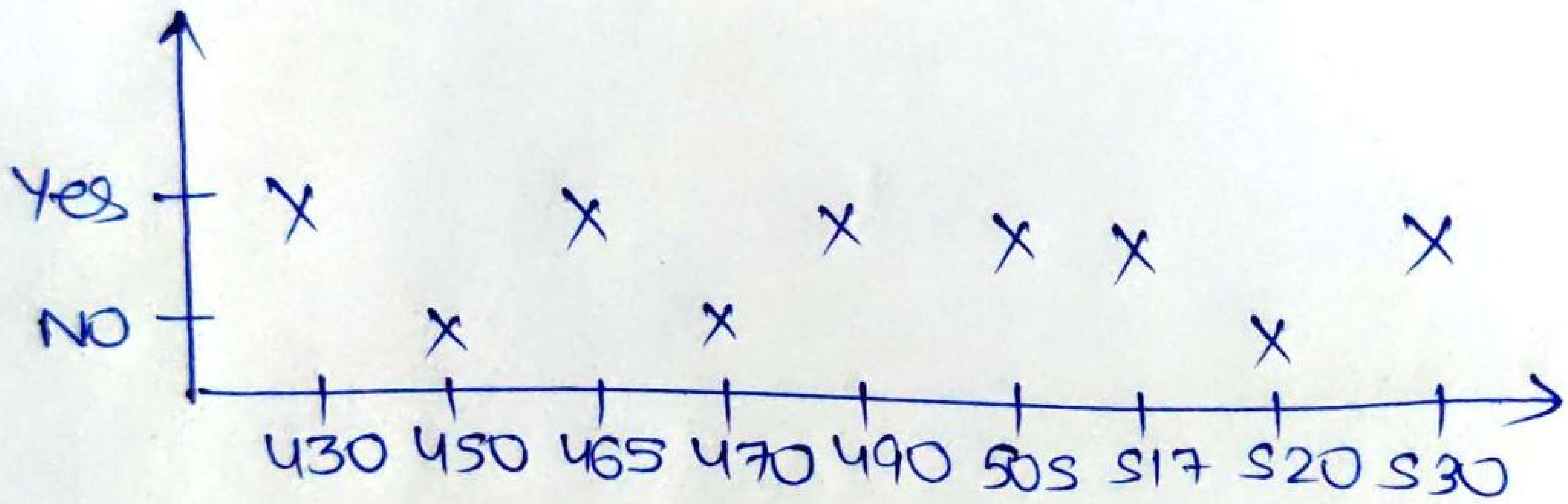
The last 2 were chosen on account of patterns observed ~~as~~ in the data.

for patterns in data we create the following graphs.

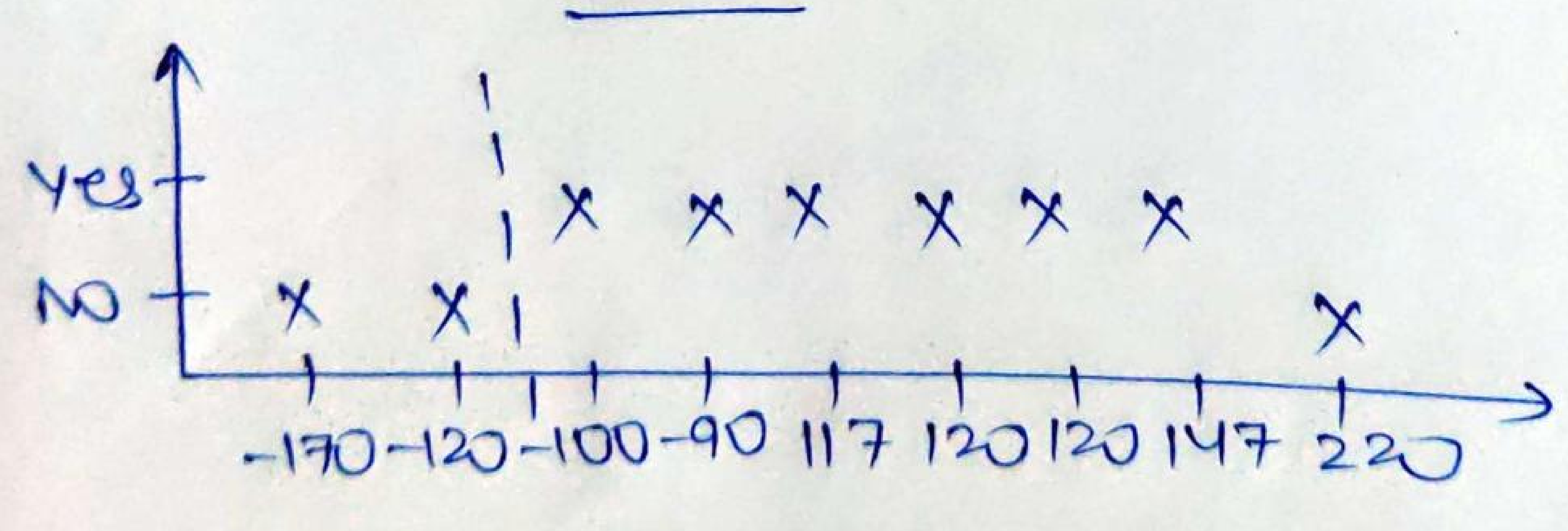
ANGLE



DISTANCE



SPEED



From these graphs the 2 patterns
Angle $\geq$ 4.8 and Speed $<$ -100 were observed

FIITJEE

Case 1 Mean of Angle

Mean = $(1.9 + 4.5 + 3 + 5.5 + 3.2 + 5.2 + 1.85 + 4.8 + 1.7)/9$

$= 3.4722$

so let cond be $\boxed{\geq 3.47}$

True

4.5	5.5	5.2	4.8
N	Y	Y	Y

false

1.9	3	3.2	1.85	1.7
N	Y	N	Y	Y

$E_1 = I(3,1)$

$= \frac{3}{4}\log_2(4/3) + \frac{1}{4}\log_2(4)$

$= 0.3127 + 0.5 = \underline{0.81127}$

$E_2 = I(3,2)$

$= \frac{3}{5}\log_2(5/3) + \frac{2}{5}\log_2(5/2)$

$= \underline{0.97095}$

so $E = \frac{4}{9}(E_1) + \frac{5}{9}(E_2)$

$= 0.36056 + 0.5394 = 0.899964$

$\therefore$ Gain $= 0.918295 - 0.899964 = \boxed{0.018331}$

Case 2 Mean of distance.

Mean $= (450 + 520 + 490 + 530 + 470 + 505 + 465 + 517 + 430)/9$

$= 486.33$

cond. $\boxed{\geq 486.33}$

True

520	490	530	505	517
N	Y	Y	Y	Y

false

450	470	465	430
N	N	Y	Y

$E_1 = I(4,1)$

$= \frac{4}{5}\log_2(5/4) + \frac{1}{5}\log_2(5) = \underline{0.7219}$

$E_2 = I(2,2)$

$= \underline{1}$

$E = \frac{5}{9} \times E_1 + \frac{4}{9} \times E_2 = 0.84551$

Gain $= 0.918295 - 0.84551 = \boxed{0.0727}$

case 3 Mean of Speed

Mean = (220 + (−120) + 120 + 117 + (−170) + (−90) + 120 + 147 + (−100))/9

= 27.11

cond. ≥ 27.11

True False

220	120	117	120	147
N	y	y	y	y

−120	−170	−90	−100
N	N	y	y

$E_1 = I(4,1)$

$= 0.7219$ (calc before)

$E_2 = I(2,2)$

$= \underline{1}$

so $E = \frac{5}{9} \times E_1 + \frac{4}{9} \times E_2 = 0.84551$

Gain = 0.918295 − 0.84551 = $\boxed{0.07277}$

case 4 Angle ≥ 4.8.

True False

5.5	5.2	4.8
y	y	y

1.5	4.5	3	3.2	1.85	1.7
N	N	y	N	y	y

$E_1 = I(3,0)$

$= \underline{0}$

$E_2 = I(3,3)$

$= 1.$

so $E = \frac{3}{9} \times E_1 + \frac{6}{9} \times E_2 = 0.66666$

Gain = 0.918295 − 0.66666 = $\boxed{0.2516284}$

case 5 Speed < −100

True false

−120	−170
N	N

220	120	117	−90	120	147	−100
N	y	y	y	y	y	y

$E_2 = I(6,1) = \frac{6}{7}\log_2(7/6) + \frac{1}{7}\log_2(7)$

$= 0.190622 + 0.40105$

$= 0.59167$

$E_1 = I(0,2)$

$= \underline{0}$

$E = \frac{2}{9} \times 0 + \frac{7}{9} \times E_2 = 0.460189$

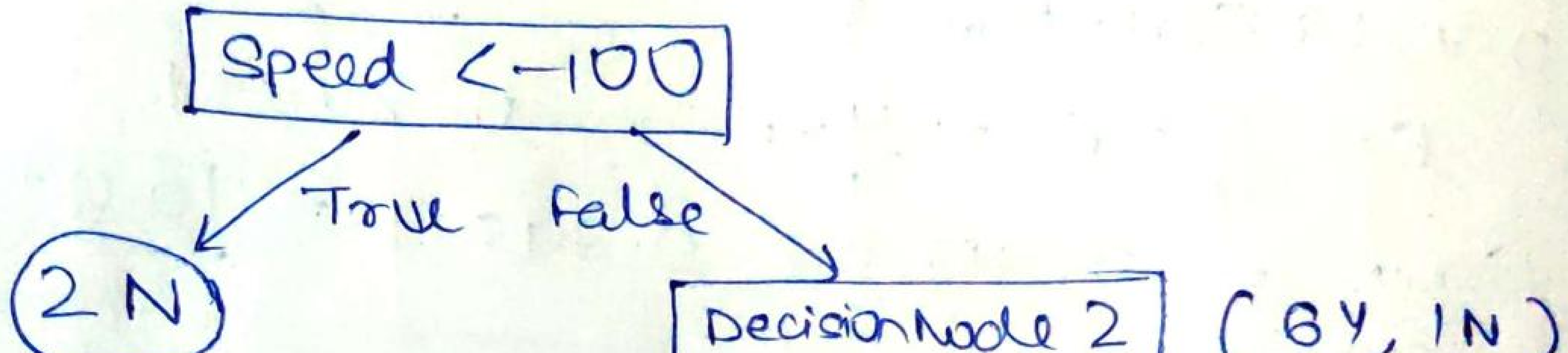
Gain = 0.918295 − 0.460189 = $\boxed{0.458105}$

Comparing all the Gains.

Mean Angle	0.018331
Mean distance	0.07277
Mean speed	0.07277
Angle ≥ 4.8	0.2516284
Speed < -100	0.458105 → Highest

hence our first decision node is Speed < -100

Tree till now



Decision Node 2

First let's calculate the starting entropy. Our new dataset is as follows.

Angle	Distance	Speed	Kill
1.5	450	220	N
3	490	120	Y
5.5	530	117	Y
5.2	505	-90	Y
1.85	465	120	Y
4.8	517	147	Y
1.7	430	-100	Y

so $E = I(6,1) = \frac{6}{7}\log_2(7/6) + \frac{1}{7}\log_2(7)$

$= 0.190622 + 0.401050$

$\boxed{= 0.591672}$

i> Speed ≥ 220 ii> Angle ≤ 1.5

Case 1 Speed ≥ 220

True / False

220
N

120	117	-90	120	147	-100
Y	Y	Y	Y	Y	Y

$E_1 = I(0,1) = \underline{0}$ $E_2 = I(6,0) = 0$.

$\therefore E = \frac{1}{7} \times E_1 + \frac{6}{7} \times E_2 = 0$

Gain $= 0.591672 - 0 = \boxed{0.591672}$

Case 2 Angle ≤ 1.5

True / False

1.5
N

3	5.5	5.2	1.85	4.8	1.7
~~N~~	~~N~~	~~N~~	~~N~~	~~N~~	Y
Y	Y	Y	Y	Y	

$E_1 = I(0,1) = \underline{0}$

$E_2 = I(6,0) = \underline{0}$.

$E = \frac{1}{7} \times E_1 + \frac{6}{7} \times E_2 = 0$ $\boxed{\text{Gain} = 0.591672}$

Gain for both the cases are same, but for better generalisation taking a new attribute we construct the final tree as follows.

Speed < -100

True → 2N

False → Angle ≤ 1.5

True → N

False → 6Y

Entropy final $= 0 + 0 = \boxed{0}$