K nearest neighbor algorithm:-

rating = 4.2, audience = 4.5

movie ID	rating	audience (in millions)	Genre
1	4.5	5	Action
2	3.8	2	Comedy
3	4.9	6	Action
4	2.9	1	Drama
5	3.2	3	Comedy
6	4.0	4	Drama
7	2.9	1.5	Comedy
8	3.7	2.5	Action
		2.8	

For k=3

$\sqrt{(4.2-5)^2+(4.5-5)^2} = 0.943$

~~$\sqrt{(4.2-3.8)^2+($~~

$\sqrt{(4.2-3.8)^2+(4.5-2)^2} = 2.532$

$\sqrt{(4.2-4.9)^2+(4.5-6)^2} = 1.655$

$\sqrt{(4.2-2.5)^2+(4.5-1)^2} = 3.891$

Scanned with CamScanner

$\sqrt{(4.2-3.2)^2+(4.5-3)^2} = 1.803$

$\sqrt{(4.2-4)^2+(4.5-4)^2} = 0.539$

$\sqrt{(4.2-2.9)^2+(4.5-1.5)^2} = 3.269$

$\sqrt{(4.2-3.7)^2+(4.5-2.5)^2} = 2.062$

For $k=3$

$\Rightarrow$ MovieID = 6, 1, 3

Genre = Comedy, Action, Action

$\therefore$ Genre for the new rating and audience at $k=3$ is Action

Scanned with CamScanner

K means algorithm:-

Customer ID	Annual Income (in $1000s)	Spending Score (1-100)
1	15	39
2	16	81
3	17	6
4	18	77
5	19	40
6	20	76
7	21	6
8.	22	94
9	23	3
10	24	72

$k1 = 15, 39 \quad k2 = 16, 81$

for ③ $= \sqrt{(17-15)^2 + (6-39)^2} = 33.06$

$\sqrt{(17-16)^2 + (6-81)^2} = 75.0$

$k_1 = \frac{15+17}{2}, \frac{39+6}{2} = 16, 22.5$

$k_1 = 16, 22.5 \quad k_2 = 16, 81$

④ $= \sqrt{(18-16)^2 + (77-22.5)^2} = 54.54$

$\sqrt{(18-16)^2 + (77-81)^2} = 4.47$

Scanned with CamScanner

$$k_2 = \frac{16+18}{2}, \frac{81+77}{2}$$

$$k_2 = 17, 79$$

$$k_1 = 16, 22.9, \quad k_2 = 17, 79$$

for ③ $= \sqrt{(19-16)^2 + (40-22.9)^2} = 17.76$

$\sqrt{(19-17)^2 + (40-79)^2} = 39.09$

$$k_1 = \frac{16+19}{2}, \frac{40+22.9}{2}$$

$$k_1 = 17.9, 31.25$$

for ⑥ $= \sqrt{(20-17.9)^2 + (76-31.25)^2} = 44.82$

$\sqrt{(20-17)^2 + (76-79)^2} = 4.24$

$$k_2 = \frac{20+17}{2}, \frac{76+76.79}{2}$$

$$k_2 = 18.9, 77.5$$

$k_1 = 17.9, 31.25, \quad k_2 = 18.9, 77.5$

for ⑦ $= \sqrt{(21-17.9)^2 + (6-31.25)^2} = 25.49$

$\sqrt{(18.9)(21-18.5)^2 + (76-77.9)^2} = 11.54$

$$k_1 = \frac{21+17.5}{2}, \frac{6+31.25}{2}$$

$$k_1 = 19.25, 18.63$$

PAPERWORK

Scanned with CamScanner

$k_1 = 19.29, 18.63 \quad k_2 = 18.5, 77.5$

for (8) $= \sqrt{(22-19.29)^2+(94-18.63)^2} = 79.42$

$\sqrt{(22-18.5)^2+(94-77.5)^2} = 16.87$

$k_2 = \frac{22+18.5}{2}, \frac{94+77.5}{2}$

$k_2 = 20.25, 85.75$

$k_1 = 19.29, 18.63 \quad k_2 = 20.25, 85.75$

for (9) $= \sqrt{(23-19.25)^2+(3-18.63)^2} = 16.07$

$\sqrt{(23-20.25)^2+(3-85.75)^2} = 82.80$

$k_1 = \frac{23+19.25}{2}, \frac{3+18.63}{2}$

$k_1 = 21.13, 10.82$

$k_1 = 21.13, 10.82 \quad k_2 = 20.25, 85.75$

for (10) $= \sqrt{(24-21.13)^2+(72-10.82)^2} = 61.25$

$\sqrt{(24-20.25)^2+(72-85.75)^2} = 14.25$

$k_1 = \{1, 3, 4, 9, 7, 9\}$

$k_2 = \{2, 6, 8, 10\}$

Scanned with CamScanner

Decision Trees:-

Dataset:

PersonID	Age	Income	Student	Bought Car
1	<30	High	No	No
2	<30	High	No	No
3	31-40	High	No	Yes
4	>40	Medium	No	Yes
5	>40	Low	Yes	No
6	>40	Low	Yes	Yes
7	31-40	Low	Yes	Yes
8	<30	Medium	No	No
9	<30	Low	Yes	Yes
10	>40	Medium	Yes	Yes
11	<30	Medium	Yes	Yes
12	31-40	Medium	No	Yes
13	31-40	High	Yes	Yes
14	>40	Medium	No	Yes

$P_+ = \frac{9}{14}$

Bought Car:-

$P_+ = \frac{9}{14}$ $\quad P_- = \frac{5}{14}$

$$\text{Entropy} = -\left(\frac{9}{14}\right)\log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right)\log_2\left(\frac{5}{14}\right)$$

$$= 0.940$$

Scanned with CamScanner

Dye :- ~~P1 = 3/5 P2 = 2/5~~

$$<30 = -\frac{2}{5} * \log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right).$$

$$\Rightarrow 0.971$$

$$31-40 = -\frac{0}{4}\log_2\left(\frac{0}{4}\right) - \frac{4}{4}\log_2\left(\frac{4}{4}\right)$$

$$\Rightarrow 0$$

$$>40 = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$$

$$\Rightarrow 0.971$$

Weighted Entropy :-

$$\frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971) = 0.693$$

IG for Dye $= 0.94 - 0.693 = 0.247$

Income :

High :-

$$-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

Medium :-

$$-\frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{4}{5}\log_2\left(\frac{4}{5}\right) = 0.722$$

Low :- $-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$

Scanned with CamScanner

Weighted Entropy :-

$$\frac{5}{14}(0.971) + \frac{5}{14}(0.722) + \frac{4}{14}(1) = 0.836$$

IG for Income :-

$$0.940 - 0.836 = 0.10$$

Student :-

Yes :-

$$-\frac{2}{8}\log_2\left(\frac{2}{8}\right) - \frac{6}{8}\log_2\left(\frac{6}{8}\right) = 0.811$$

No :-

$$-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) = 1$$

$$\frac{8}{14}(0.811) + \frac{6}{14}(1) = 0.892$$

$$0.940 - 0.892 = 0.048$$

Highest IG for age for Dgl

Age :-

Age<30 :

No : 2

Yes : 3 impure ; next attribute

PAPERWORK

Scanned with CamScanner

Age 31-40:

NO = 0
Yes = 4 — consists of only one class

Age >40:

No: 3
yes: 2 — impure; next attribute

Age < 30 & Income = High

~~Boi~~ No: 2
Yes: 3

<30
~~31-40~~
Age ~~<30~~ & Income = Medium

Yes: 2 ∴ Pure

Age ~~>40~~ <30 & Income = Low

Yes: 1 Pure

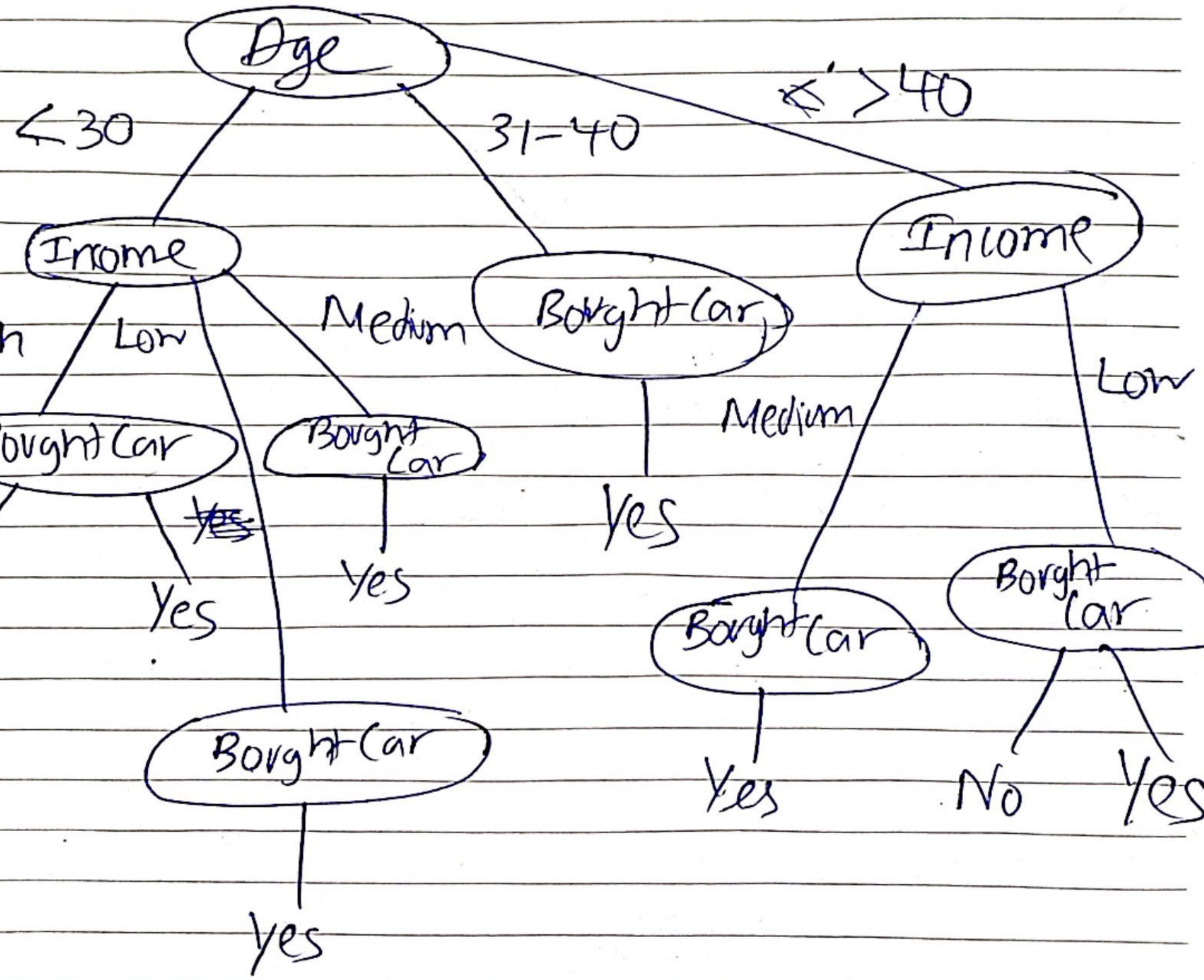
Scanned with CamScanner

Dye > 40 & Income = Medium

Yes : 2 ∴ Pure

Dye > 40 & Income = Low

Yes = 2
No = 3 ∴ Not Pure



PAPERWORK

Scanned with CamScanner