Q5. Apply the Naive Bayes classification and decision tree algorithm for boys-computers classification and classify the tuple.
X - (age - 'youth', income = 'medium', student = 'yes' & credit_rating = "fair')

→ Baye's classification.

① Step1 :- ~~Eta~~ calculate prior probability.

C1 : No C2 :- Yes

$P(C1) = \frac{5}{14}$ $P(C2) = \frac{9}{14}$

② Step2 :- Calculate $X/C1, X/C2$ ... (posterior prob)

Attribute	values	Count C1	Count C2	Probability.	
1. Age	Young (5)	3	2	3/5	2/9
	Middle-age (4)	0	4	0/5	4/9
	Old (5)	2	3	2/5	3/9
2. Income	High (4)	2	2	2/5	2/9
	Medium (6)	2	4	2/5	4/9
	Low (4)	1	3	1/5	3/9
3. Student	No (7)	4	3	4/5	3/9
	Yes (7)	1	6	1/5	6/9
4. Credit rating	fair (8)	2	6	2/5	6/9
	Good (6)	3	3	3/5	3/9

Step 3: (age = "youth", income = "medium", student = "yes", credit-rating = "fair")

young (5)	3	2	3/5	2/9
Medium (6)	2	4	2/5	4/9
yes (7)	1	6	1/5	6/9
fair (8)	2	6	2/5	6/9

$$P(X/C1) = \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} = \frac{12}{625}$$

$$P(X/C2) = \frac{2}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{6}{9} = \frac{32}{729}$$

$$P(C1/X) = \frac{P(X/C1) \cdot P(C1)}{P(X)} = \frac{12}{625} \times \frac{5}{14} = \frac{6}{875}$$

$$P(C2/X) = \frac{P(X/C2) \cdot P(C2)}{P(X)} = \frac{32}{729} \times \frac{9}{14} = \frac{16}{567}$$

Decision Tree Algorithm.

① Step 1: Calculate prior probability

C1: No C2: yes

$P(C1) = \frac{5}{14}$, $P(C2) = \frac{9}{14}$

② Step 2: Calculate entropy of entire dataset.

$$\sum_{i=1}^{n} P(Ci) \log \frac{1}{P(Ci)}$$

$$= \frac{5}{14} \log \frac{14}{5} + \frac{9}{14} \log \frac{14}{9}$$

$$= 0.2830.$$

Step 3: Calculate entrophy of each of the attribute from given set and also its gain.

3a(i) a) H (Age).

$= \frac{5}{14} \times [H(\text{young})] + \frac{4}{14} \times [H(\text{middle age})] + \frac{5}{14} \times [H(\text{old})]$

$= \frac{5}{14} \times \left[\frac{3}{5}\log\frac{5}{3} + \frac{2}{5}\log\frac{5}{2}\right] + \frac{4}{14}\left[0 + \frac{4}{4}\log\frac{4}{4}\right]$

$+ \frac{5}{14}\left[\frac{2}{5}\log\frac{5}{2} + \frac{3}{5}\log\frac{5}{3}\right]$

$= \frac{5}{14} \times 0.2922 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.2922$

$= 0.2087.$

Gain (Age) = H (Data) − H (Age)

= 0.2830 − 0.2087

= 0.0749

3b H(Income).

$= \frac{4}{14} \times [H(\text{High})] + \frac{6}{14} \times [H(\text{medium})] + \frac{4}{14} \times [H(\text{LOW})]$

$= \frac{4}{14} \times \left[\frac{2}{4}\log\frac{4}{2} + \frac{2}{4}\log\frac{4}{2}\right] + \frac{6}{14} \times \left[\frac{2}{6}\log\frac{6}{2} + \frac{4}{2}\log\frac{6}{4}\right]$

$+ \frac{4}{14} \times \left[\frac{1}{4}\log\frac{4}{1} + \frac{3}{4}\log\frac{4}{3}\right]$

$= \frac{4}{14} \times (0.3010) + \frac{6}{14} \times 0.276 + \frac{4}{14}(0.2442)$

$= 0.2740.$

Gain (Income) = H(Data) - H(Income).
= 0.2830 - 0.2740
= 0.009.

3(c). c) H (student).
$= \frac{7}{14} \times [H(NO)] + \frac{7}{14} \times [H(yes)]$

$= \frac{7}{14}\left[\frac{4}{7}\log\frac{7}{4} + \frac{3}{7}\log\frac{7}{3}\right] + \frac{7}{14}\left[\frac{1}{7}\log\frac{7}{1} + \frac{6}{7}\log\frac{7}{6}\right]$

$= \frac{7}{14} \times (0.2965) + \frac{7}{14} \times (0.1781)$

= 0.2373.

Gain (student) = H (Data) - H (student.)
= 0.2830 - 0.2373
= 0.0457.

3(d). d) H(Credit).
$= \frac{8}{14}[H(fair)] + \frac{6}{14} \times [H(Good)]$

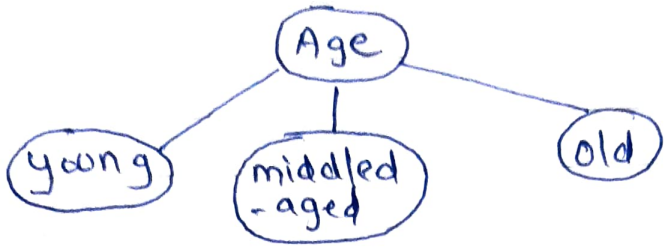
$= \frac{8}{14}\left[\frac{2}{8}\log\frac{8}{2} + \frac{6}{8}\log\frac{8}{6}\right] + \frac{6}{14}\left[\frac{3}{6}\log\frac{6}{3} + \frac{3}{6}\log\frac{6}{3}\right]$

$= \frac{8}{14} \times (0.2442) + \frac{6}{14} \times (0.3010)$

= 0.2685

Gain (Credit) = H (Data) - H (Credit).
= 0.2830 - 0.2685.
= 0.0145.

Step 4:- Decision Tree.



income	stu	credit	Buys
high	No	fair	No
high	NO	good	NO
medium	No	fair	No
~~high~~	Yes	fair	yes
medium	yes	good.	yes

income	stu	credit	Buys
high	No	fair	yes
low	Yes	good	yes
medium	No	good	yes
high	Yes	fair	yes

Income	stud	credit	Buys
Medium	No	fair	yes
Low	yes	fair	yes
Low	yes	Good	No
medium	No	fair	yes
medium	No	Good	No.