The column that has the highest information gain, will be the root of the decision tree.

Information Gain = Entropy of class − Entropy of Attributes

To calculate the entropy of class:

1. Find out how many levels (n) the class has.
2. calculate the probability of each level.
3. Then, entropy of class = $-\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$

$x_i \rightarrow$ i-th level

From the given table/matrix, "Buy Computer" is the class, which has two levels: Yes, No.

$\therefore P(BC = Yes) = \frac{12}{20} = 0.6$

$P(BC = No) = 0.4$

$\therefore$ Entropy, $H(BC) = -(0.6 \times \log_2 0.6 + 0.4 \times \log_2 0.4)$

$= 0.97095$

To calculate the entropy of an attribute:

- Find out the number of levels (n) in that attribute
- Calculate the probability of each level
- calculate the probability of class for every level of the attribute
- use these to find out entropy

Entropy of attribute "Age"

step-1:

n = number of levels = 4

step-2:

$P(Age < 18) = \frac{3}{20} = 0.15$

$P(Age: 18-35) = \frac{4}{20} = 0.2$

$P(Age: 36-55) = \frac{8}{20} = 0.4$

$P(Age > 55) = \frac{5}{20} = 0.25$

step-3:

Probability of "Buy Computer" when "Age < 18"

$= P(BC \mid Age<18)$

$= -\left(\frac{2}{3} \times \log_2 \frac{2}{3} + \frac{1}{3} \times \log_2 \frac{1}{3}\right)$

$= 0.9183$

here, we have 3 rows with Age<18.

- 2 of them buy's computer
- 1 doesnot

when Age<18, $P(BC = Yes) = \frac{2}{3}$

when Age<18, $P(BC = No) = \frac{1}{3}$

then, we have used the formula of entropy of class.

[why? Don't ask. I don't know.]

Entropy of Attribute "Age"

= ∑ Probability of each level × Probability of Buy Computer (class) for that level

- $P(BC \mid Age: 18-35) = -\left(0 \times \cancel{\log_2 0} + 1 \times \log_2 1\right)$

$= 0$

$(BC \mid Age: 36-55) = -\left(\frac{5}{8} \times \log_2 \frac{5}{8} + \frac{3}{8} \times \log_2 \frac{3}{8}\right)$

$= 0.9544$

$P(BC \mid Age > 55) = -(1 \times \log_2 1 + 0) = 0$

$\therefore$ Entropy (Age) $= 0.15 \times 0.9183 + 0.2 \times 0 + 0.4 \times 0.9544 + 0.25 \times 0$

$= 0.51951$

Information Gain(Age) $= 0.97095 - 0.51951$

$= 0.45144$

Attribute → Education

$P(Ed.: Masters) = \frac{7}{20} = 0.35$

$P(Ed.: High\ School) = \frac{7}{20} = 0.35$

$P(Ed.: Bachelors) = \frac{6}{20} = 0.3$

$P(BC \mid Ed.: Masters) = -\left(\frac{6}{7} \times \log_2 \frac{6}{7} + \frac{1}{7} \times \log_2 \frac{1}{7}\right)$

$= 0.59167$

$P(BC \mid Ed.: High\ School) = -\left(\frac{4}{7} \times \log_2 \frac{4}{7} + \frac{3}{7} \times \log_2 \frac{3}{7}\right)$

$= 0.98523$

$P(BC \mid Ed.: Bachelors) = -\left(\frac{2}{6} \times \log_2 \frac{2}{6} + \frac{4}{6} \times \log_2 \frac{4}{6}\right)$

$= 0.9183$

$\therefore$ Entropy, $H(BC \mid Education)$

$= 0.35 \times 0.59167 + 0.35 \times 0.98523 + 0.3 \times 0.9183$

$= 0.82741$

$IG(Education) = 0.97095 - 0.82741 = 0.14354$

Income:

$P(In: High) = \frac{10}{20} = 0.5$

$P(In: Low) = 0.5$

$P(BC \mid In: High) = -(0.5 \times \log_2 0.5 + 0.5 \times \log_2 0.5)$

$= 1$

$P(BC \mid In: Low) = -(0.7 \times \log_2 0.7 + 0.3 \times \log_2 0.3)$

$= 0.88129$

$H(BC \mid Income) = 0.5 \times 1 + 0.5 \times 0.88129$

$= 0.94065$

$IG(Income) = 0.97095 - 0.94065$

$= 0.0303$

Marital Status:

$P(MS: Married) = \frac{8}{20} = 0.4$

$P(MS: single) = 0.6$

$P(BC \mid MS: Married) = -(\frac{3}{8} \times \log_2 \frac{3}{8} + \frac{5}{8} \times \log_2 \frac{5}{8})$

$= 0.95443$

$P(BC \mid MS: Single) = -(\frac{9}{12} \times \log_2 \frac{9}{12} + \frac{3}{12} \times \log_2 \frac{3}{12})$

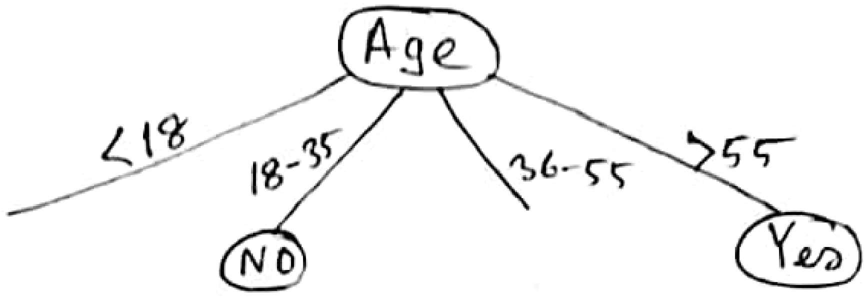
$= 0.81128$

$H(BC \mid MS) = 0.4 \times 0.95443 + 0.6 \times 0.81128$

$= 0.86854$

$IG(MS) = 0.97095 - 0.86854 = 0.10241$

can see that, Age has the highest Information Gain. So Age will be the root of the decision tree.



from previous calculation, we see that, when Age : 18-35, we are sure that nobody will buy computere. Alternatively, when Age > 55, everybody will buy computere.

Now, fore the remaining two levels, we do not enough information to ~~expand~~ make decision. So, we will have to expand the tree.

When Age < 18, oure dataset is:

Education	Income	Marital Status	Buy computere
High School	Low	Single	Yes
High School	High	Single	No
High School	Low	Married	Yes

Now, we will have to do the same process as before on this table.

ntropy of Class $= -\left(\frac{2}{3}\times\log_2\frac{2}{3}+\frac{1}{3}\times\log_2\frac{1}{3}\right)$

$= \cancel{0.58496}\ 0.9183$

<u>Education</u>

$P(\text{Ed}: \text{High School}) = 1$

$P(\text{BC}|\text{Ed}: \text{High School}) = \cancel{(1\times\log_2 1) = 0}\ -\left(\frac{2}{3}\times\log_2\frac{2}{3}+\frac{1}{3}\times\log_2\frac{1}{3}\right)$

$= 0.9183$

$H(\text{BC}|\text{Education}) = 1\times 0.9183 = 0.9183$

$IG(\text{Education}) = \cancel{0.58496 - 0 = 0.58496}$

$= 0.9183 - 0.9183 = 0$

<u>Income</u>

$P(\text{In}: \text{High}) = \frac{1}{3}$

$P(\text{In}: \text{Low}) = \frac{2}{3}$

$P(\text{BC}|\text{In}: \text{High}) = 0$

$P(\text{BC}|\text{In}: \text{Low}) = -(1\times\log_2 1) = 0$

$H(\text{BC}|\text{Income}) = \frac{1}{3}\times 0 + \frac{2}{3}\times 0 = 0$

$IG(\text{Income}) = 0.9183 - 0 = 0.9183$

<u>Marital Status</u>

$P(\text{MS}: \text{Married}) = \frac{1}{3}$

$P(\text{MS}: \text{Single}) = \frac{2}{3}$

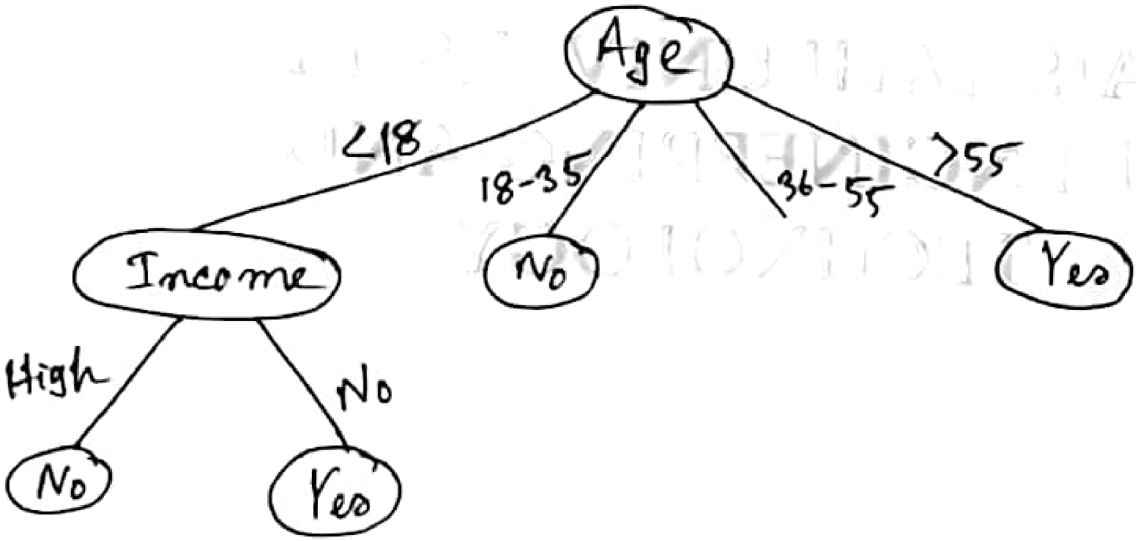
$P(\text{BC}|\text{MS}: \text{Married}) = -(1\times\log_2 1) = 0$

$P(\text{BC}|\text{MS}: \text{Single}) = -(0.5\times\log_2 0.5 + 0.5\times\log_2 0.5)$

$= 1$

$H(\text{BC}|\text{MS}) = \frac{1}{3}\times 0 + \frac{2}{3}\times 1 = 0.6667$

Income has the highest Information Gain.
So, the decision tree now -



For Age: 36-55, the dataset is:

Education	Income	Marital Status	Buy Computer
~~36-55~~ Master's	High	Single	Yes
Master's	Low	Single	Yes
Bachelor's	Low	Married	No
Master's	Low	Married	No
Master's	High	Single	Yes
High School	Low	Single	Yes
Master's	Low	Single	Yes
High School	High	Married	No

$$H(BC) = -\left(\frac{5}{8}\times\log_2\frac{5}{8} + \frac{3}{8}\times\log_2\frac{3}{8}\right) = 0.95443$$

Education

$P(Ed: Master's) = \frac{5}{8}$

$P(Ed: High\ School) = \frac{2}{8} = \frac{1}{4}$

$P(Ed: Bachelor's) = \frac{1}{8}$

$P(BC|Ed: Master's) = -\left(\frac{4}{5} \times \log_2 \frac{4}{5} + \frac{1}{5} \times \log_2 \frac{1}{5}\right)$

$= 0.72193$

$P(BC|Ed: Bachelor's) = 0$

$P(BC|Ed: High\ School) = -(0.5 \times \log_2 0.5 + 0.5 \times \log_2 0.5)$

$= 1$

$H(BC|Ed) = \frac{5}{8} \times 0.72193 + \frac{1}{8} \times 0 + \frac{1}{4} \times 1$

$= 0.70121$

$IG(Education) = 0.95443 - 0.70121$

$= 0.25322$

<u>Income</u>

$P(In: High) = \frac{3}{8}$

$P(In: Low) = \frac{5}{8}$

$P(BC|In: High) = -\left(\frac{2}{3} \times \log_2 \frac{2}{3} + \frac{1}{3} \times \log_2 \frac{1}{3}\right)$

$= 0.9183$

$P(BC|In: Low) = -\left(\frac{3}{5} \times \log_2 \frac{3}{5} + \frac{2}{5} \times \log_2 \frac{2}{5}\right)$

$= 0.97095$

$H(BC|Income) = \frac{3}{8} \times 0.9183 + \frac{5}{8} \times 0.97095$

$= 0.95121$

$IG(Income) = 0.95443 - 0.95121$

$= 0.00322$

arrital Status:

$P(MS: Married) = \frac{3}{8}$

$P(MS: Single) = \frac{5}{8}$

$P(BC \mid MS: Married) = 0$

$P(BC \mid MS: Single) = -(1 \times \log_2 1) = 0$

$H(BC \mid MS) = \frac{3}{8} \times 0 + \frac{5}{8} \times 0 = 0$

$IG(MS) = 0.95443 - 0 = 0.95443$

Hence, the decision tree is,