

Example of Data Table: ID3 Algorithm

Data:

ID	AGE	INCOME	EDUCATION	Credit Score	Subscribed
1	Young	Low	Primary	Fair	NO
2	Young	Medium	Secondary	Good	NO
3	Mid-Age	High	Tertiary	Excellent	Yes
4	Mid-aged	Medium	Secondary	Good	Yes
5	Senior	Low	Primary	Fair	NO
6	Senior	Medium	Tertiary	Excellent	Yes
7	Young	High	Secondary	Good	NO
8	Mid-aged	Low	Primary	Fair	NO
9	Mid-aged	High	Tertiary	Excellent	Yes
10	Senior	Medium	Primary	Good	NO

Question:- You are a data-Scientist at a bank, and your team is developing a model to predict whether a Customer will subscribe to a new term deposit offer based on their past interactions and demographics. Your goal is to build a decision tree using the ID3 algorithm to understand the key factors influencing a Subscription decision.

Entropy of AGE: Entire data Set

(43, -4)

$$(+4, -6) \Rightarrow -\frac{4}{10} \log_2 \frac{4}{10} - \frac{6}{10} \log_2 \frac{6}{10} \Rightarrow 1.331$$

Entropy of Attribute:-

$$\text{Ent: Young} = (+0, -3) = 0$$

$$\text{Ent: Mid-age} = (+3, -1) \Rightarrow -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \Rightarrow -0.188$$

$$\text{Ent: Senior} = (+1, -2) \Rightarrow -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \Rightarrow -0.167$$

$$\text{I.G. of AGE: } 1.331 - \frac{4}{10} \times 0 - \frac{3}{4} \times -0.188 - \frac{2}{3} \times -0.167$$

$$\Rightarrow 1.096$$

Ent of Incom

$$\text{Total Data Ent} = 1.331$$

Ent of Att

$$\text{Ent of High: } (+2, -1) \Rightarrow -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \Rightarrow -0.167$$

$$\text{Ent of Medium: } (+2, -2) \Rightarrow -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \Rightarrow -0.150$$

$$\text{Ent of Low: } (+0, -3) \Rightarrow 0$$

$$\text{Ent of Incom} \Rightarrow 1.331 - \frac{3}{10} \times -0.167 - \frac{4}{10} \times -0.150$$

$$\Rightarrow 1.081$$

Ent of Education r

Total later Ent :- -1.131

Ent of Primary : (+0, -4) \Rightarrow -0.301

Ent of Secondary : (+1, -2) \Rightarrow -0.167

Ent of Tertiary : (+3, -0) \Rightarrow -0.301

Ent of Education = ~~0.971~~ $\frac{4}{10} \times -0.301 - \frac{3}{10} \times -0.167$

\Rightarrow ~~0.8702~~ $\frac{1}{10} \times 1.232$

Ent of Credit-Score

Total Ent of Data: -1.131

$$\text{Ent of Fair} : (+0, -3) \Rightarrow -\frac{0}{9} \log_2 \frac{0}{9} - \frac{3}{9} \log_2 \frac{3}{9} \Rightarrow -0.301$$

$$\text{Entropy of Good} : (+1, -3) \Rightarrow -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} \Rightarrow -0.188$$

$$\text{Ent of Excellent} : (+3, -0) \Rightarrow -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} \Rightarrow -0.301$$

56

Ent of Credit-Score \Rightarrow

0.971

$$\cancel{1.131} - \frac{3}{10} \times -0.301 - \frac{4}{10} \times -0.188 - \frac{3}{10} \times -0.301$$

$$\Rightarrow \cancel{0.8452} \quad 1.227$$

Step 2:- I.G

$$\text{IG of Age} : - \cancel{1.0057} \quad 1.096$$

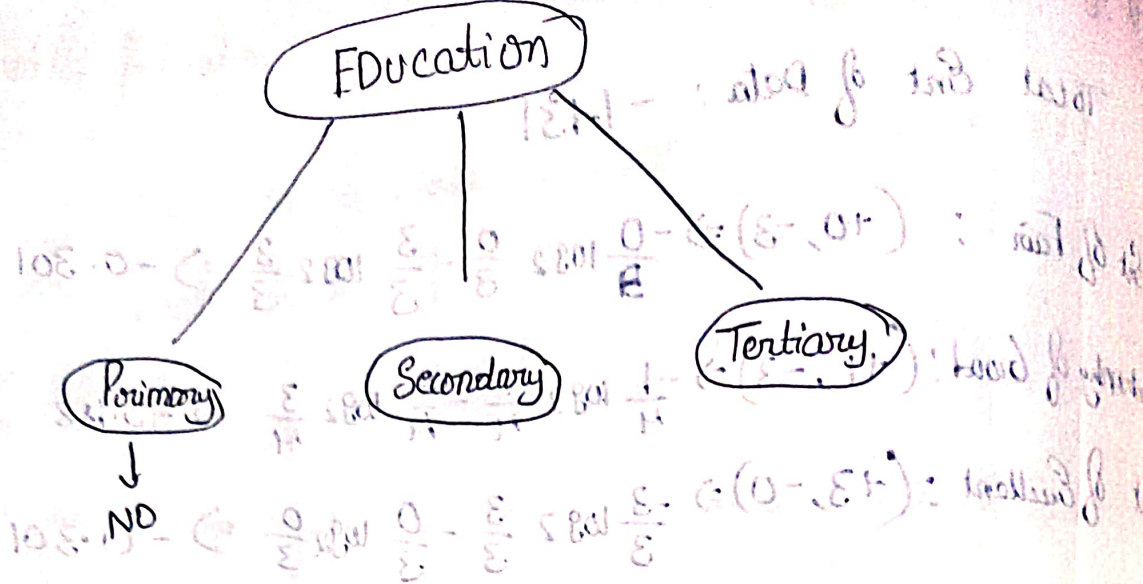
$$\text{IG of Income} : - \cancel{1.0209} \quad 1.081$$

$$\text{IG of Education} : - \cancel{0.8302} \quad 1.232$$

$$\text{IG of Credit-Score} : - \cancel{0.8452} \quad 1.227$$

$$\text{Max is} : - \cancel{0.8452} \quad 1.232$$

\therefore The Root Node is ~~Education~~
Education.



Entropy of Primary.

Education	Age	Income	Credit Score	Subscribed
Primary	Young	Low	Fair	NO
Primary	Seniour	Low	Fair	NO
Primary	Mid	Low	Fair	NO
Primary Seniour	Seniour	Medium	Good	NO

Total Ent of Primary : $(+0, -4) \Rightarrow \frac{-0}{4} \log_2 \frac{0}{4} - \frac{-4}{4} \log_2 \frac{4}{4}$

$\Rightarrow 0$

Ent(Age)

Total IG of Primar : 0

#Entropy of Secondary

Total Ent of Secondary:

Data:

Educa	Age	Income	Credit Score	Subscribed
Secondary	Young	Medium	Good	NO
Secondary	Mid-age	Medium	Good	Yes
Secondary	Young	High	Good	NO

Ent of total data of Secondary :-

$$(+1, -2) \Rightarrow -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}$$

$$\Rightarrow 0.918$$

~~Ent of Age $\Rightarrow (+1, -2) \Rightarrow -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \Rightarrow$~~

Ent of Income \rightarrow

Ent of Age: (Young) $= (+0, -2) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} \Rightarrow 0$

(Mid-age) $= (+1, -0) = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} \Rightarrow 0$

Ent of Income:-

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

(High) $\Rightarrow 0$

Ent of Credit Score:- (Good) $\Rightarrow (+1, -2)$

$$-\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \Rightarrow 0.918$$

Total I.G of Secondary: $0.918 - \frac{2}{3} \times 0 - \frac{1}{3} \times 0$

$$- \frac{2}{3} \times 1 = 0$$

$$\Rightarrow 0.251$$

Entropy of Secondary Variables

Education	Age	Income	Credit	Subscribed
Primary	Mid	High	Excellent	Yes
Secondary	Senior	Med	Excellent	Yes
Tertiary	Mid	High	Excellent	Yes
Tertiary				

Total Ent of Tertiary: $\frac{1}{8} \log \frac{1}{8} = (-1 + 3) = 2$

$$(+3, -0) = 0 \quad \text{SIP} = 0$$

IG of Education:

Ent of Primary: $0 \cdot (\text{prob}) = 0$

Ent of Secondary: 0.251

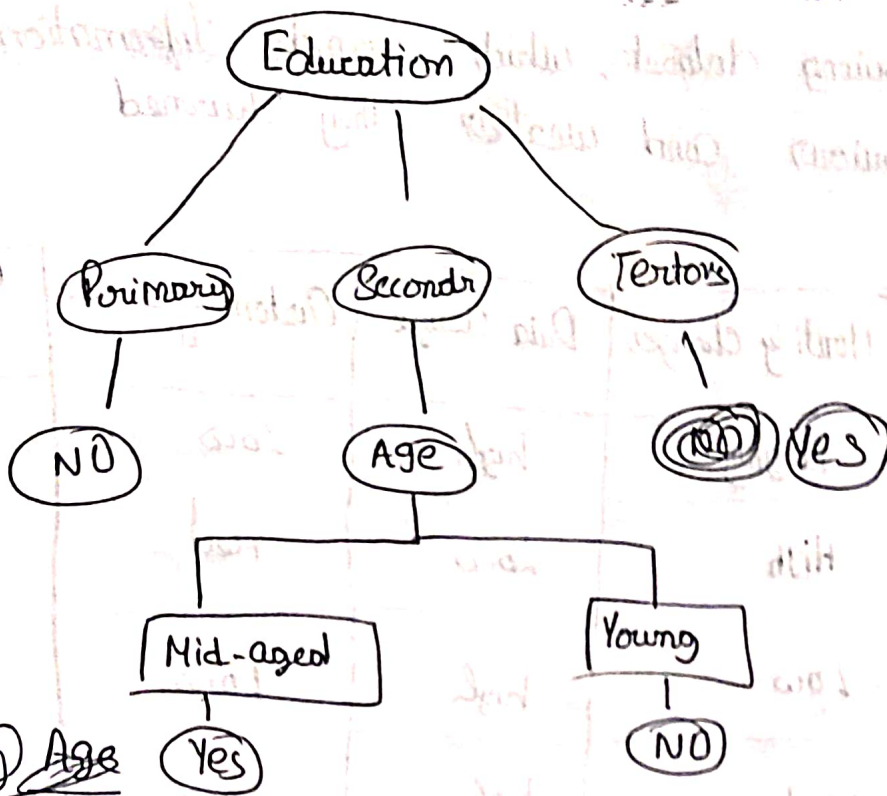
Ent of Tertiary: 0

Secondary

$(=) 0.251$ Ent of age: $\frac{1}{8}$

$(=) \text{Secondary} \cdot \frac{1}{8} \log \frac{1}{8}$

$0 \cdot \frac{1}{8} = 0$ SIP = 0.251



Ent of Mid-aged

$$\begin{aligned}
 & 0 - (0 - 5+) : \text{exit} \\
 & 1 < \frac{2}{4} \text{ loss } \frac{5}{4} - \frac{2}{4} \text{ loss } \frac{5}{4} \cdot (0 - 5+) : \text{another} \\
 & \quad (2 - 1+) : \text{exit} \\
 & 0 < \frac{2}{4} \text{ loss } \frac{5}{4} - \frac{2}{4} \text{ loss } \frac{5}{4} \cdot (0 - 5+)
 \end{aligned}$$