Entropy Calculations

Calculate Entropy for the data set provided in the Machine Learning Lecture, we proceed as follows :

Demand	Strategic	Campaign	Conversion
heavy	yes	aggressive	high
moderate	no	aggressive	high
heavy	yes	aggressive	medium
low	no	lowkey	medium
heavy	yes	aggressive	low
low	yes	lowkey	low
moderate	yes	aggressive	medium
low	no	aggressive	medium
heavy	yes	lowkey	low
moderate	no	lowkey	low
heavy	yes	aggressive	high
moderate	no	lowkey	high
low	no	lowkey	low
heavy	yes	aggressive	high
heavy	yes	aggressive	medium
low	no	lowkey	low

Solution :

We have 3 independent variables here : Demand, Strategic and Campaign.
We deduce their occurrences with respect to the dependant variable as below :

Demand : Heavy – Moderate - Low

Conversion →	**High**	**Medium**	**Low**	**Occurrences**
Heavy	3	2	2	7
Moderate	2	1	1	4
Low	0	2	3	5
			Total Occurrence	**16**

Strategic : Yes - No

Conversion →	**High**	**Medium**	**Low**	**Occurrences**
Yes	3	3	3	9
No	2	2	3	7
			Total Occurrence	**16**

Campaign : Aggressive –Low Key

Conversion →	**High**	**Medium**	**Low**	**Occurrences**
Aggressive	4	4	1	9
LowKey	1	1	5	7
			Total Occurrence	**16**

Step 2 : Calculate Probability and Entropy of the Master Data

Conversion	Occurrence
High	5
Medium	5
Low	6
Total	**16**

(a)

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Entropy = P_{high} x $\log_2 P_{high}$ **+** P_{medium} x $\log_2 P_{medium}$ **+** P_{low} x $\log_2 P_{low}$

$P_{high = 5/16}$

$P_{medium = 5/16}$

$P_{low= 6/16}$

Hence **Entropy E(S) = 1.577**

Entropy for each of the independent variables will be calculated using the below formula :

(b)

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

Entropy of Demand :

From the table, we can calculate the probability as below :

P_{heavy} = 7/16

$P_{moderate}$ = 4/16

P_{low} = 5/16

Entropy(Conversion, Demand) = P_{heavy} x E_{heavy} + $P_{moderate}$ x $E_{moderate}$ +P_{low} x E_{low}

Entropy for single variable of Demand is calculated using formula (a) which is as below :

E_{heavy} = 1.527, $E_{moderate}$ = 1.5, E_{low} = 0.972

Using formula(b), we get Entropy(Conversion, Demand) = 1.34

Entropy of Strategy :

From the table, we can calculate the probability as below :

P_{yes} = 9/16

P_{no} = 7/16

Entropy(Conversion, Strategic) = $P_{yes} \times E_{yes} + P_{no} \times E_{no}$

Entropy for single variable of Strategic is calculated using formula (a) which is as below :

E_{yes} = 1.599, E_{no} = 1.555

Using formula(b), we get Entropy(Conversion, Strategic) = 1.571

Entropy of Campaign :

From the table, we can calculate the probability as below :

$P_{aggressive}$ = 9/16

P_{lowkey} = 7/16

Entropy(Conversion, Campaign) = $P_{aggressive} \times E_{aggressive} + P_{lowkey} \times E_{lowkey}$

Entropy for single variable of Campaign is calculated using formula (a) which is as below :

E_{yes} = 1.393, E_{no} = 1.149

Using formula(b), we get Entropy(Conversion, Campaign) = 1.286

Calculate Gain for each independent variables :

$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

$Gain_{Demand}$ = E(s) - Entropy(Conversion, Demand) = 1.577-1.34 =0.237

$Gain_{Campaign}$ = E(s) - Entropy(Conversion, Campaign) =1.577-1.286= 0.291

$Gain_{Strategic}$ = E(s) - Entropy(Conversion, Strategic)=1.577-1.571=0.006

Gain for Campaign is the maximum, hence the root node is Campaign. We are required to keep splitting each of the child nodes till we get entropy 0, which is then called Leaf Node. If Entropy >0, then we need to further split recursively till we get entropy =0.