Decision Trees

What are Decision Trees ?

Decision Tree is an algorithm used for building classification and regression models by representing it in the form of a tree structure.

Types of Decision Trees

DATA FOLKZ®
#CATAPULT DATA LEADERS

Categorical Variable Decision Tree: Decision Tree which has categorical target variable then it called as categorical variable decision tree.

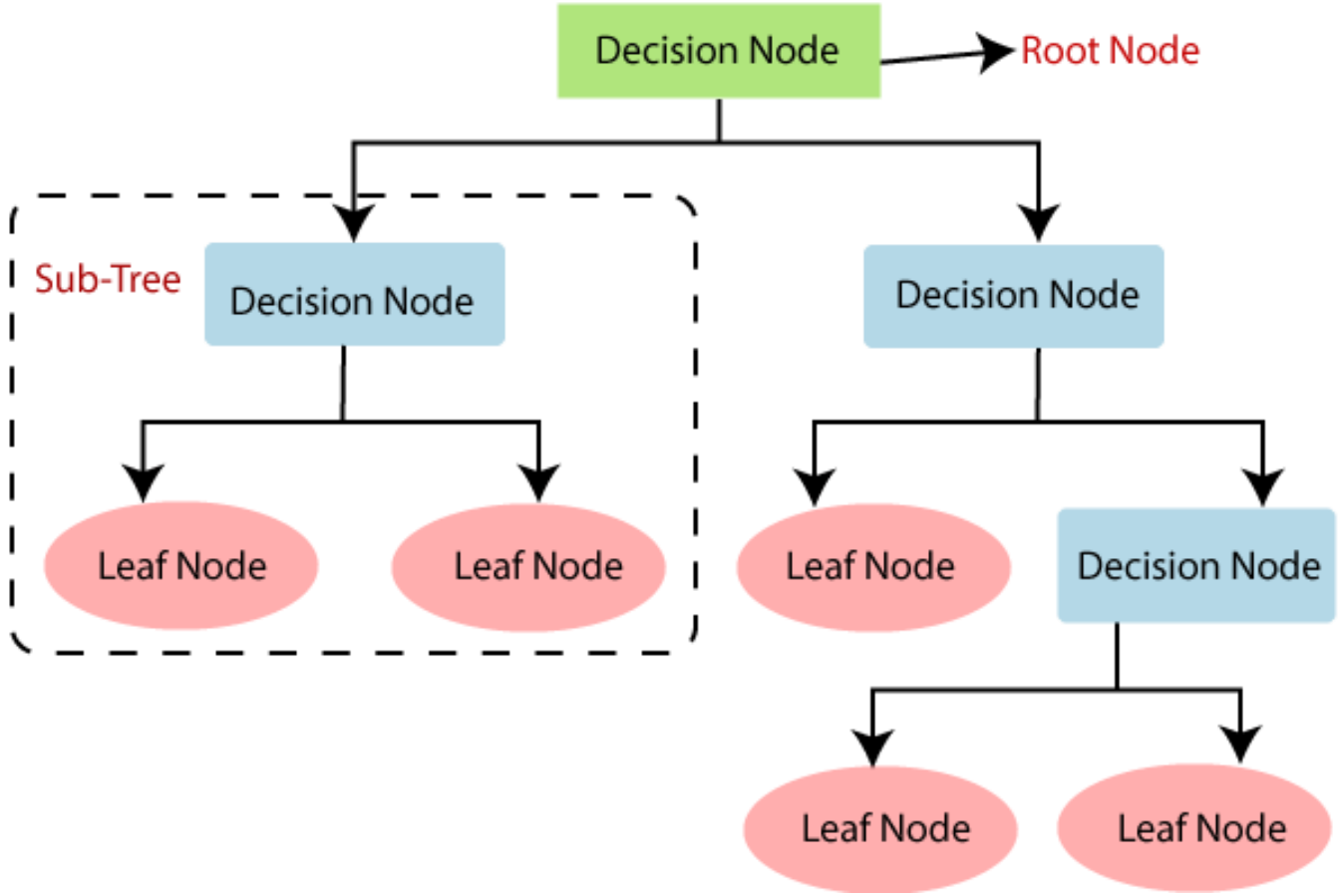
Continuous Variable Decision Tree: Decision Tree has continuous target variable then it is called as Continuous Variable Decision Tree.

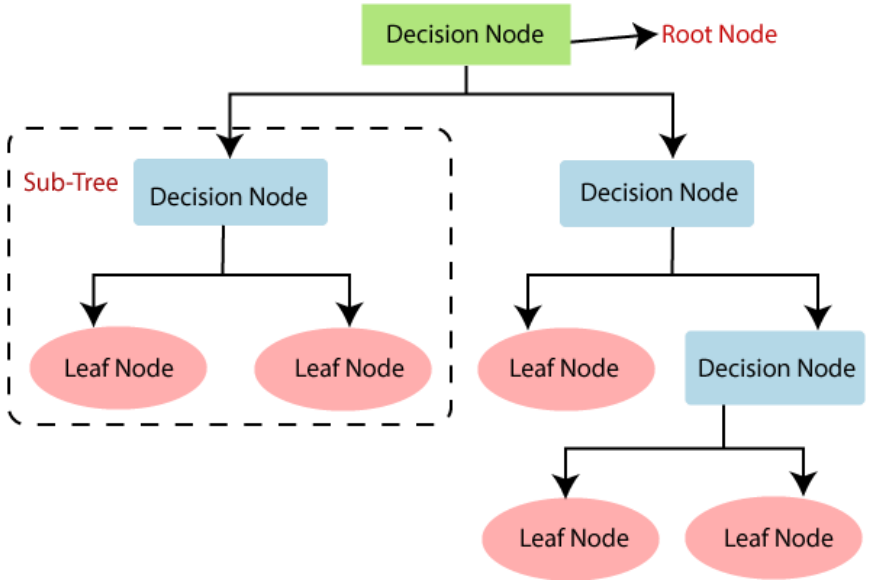
How does a Decision Tree look like ?

Root Node - The root node is the highest node in the tree structure and has no parent. This node is a global element and represents the entire message. It may have one or more child nodes but can never have sibling nodes or be repeating.

Internal Node - An internal node (also known as an inner node, inode for short, or branch node) is any node of a tree that has child nodes.

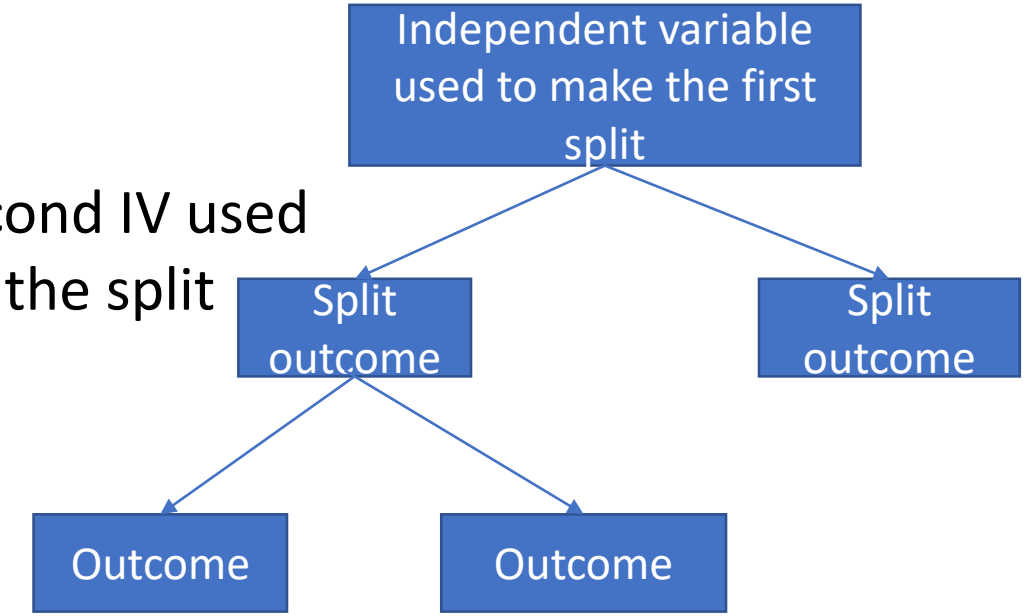
Leaf Nodes - An external node (also known as an outer node, leaf node, or terminal node) is any node that does not have child nodes.



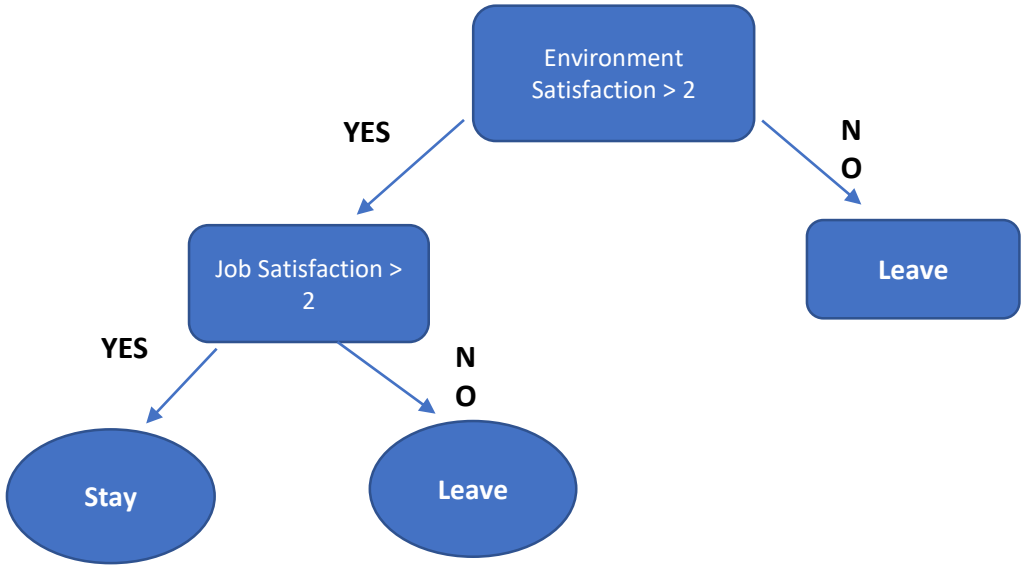


DATA FOLKZ®
#CATAPULT DATA LEADERS

Second IV used for the split



Will the employee stay with the company or leave ?



Example of Decision Tree

Example of Decision tree

- In the above slide , we discussed whether employee will stay or leave the organisation based on the information we have regarding Environment Satisfaction and Job Satisfaction .

- We used Environment Satisfaction as the first criteria to decide whether he will stay or not .

But how do we decide this ?
How do we decide which variable to use first for the decision making ?

How does the attribute for the split is decided ?

The popular attribute selection measures are :

- Entropy / Information Gain
- Gini index

If dataset consists of **“n”** attributes like our dataset consists 35 attributes , then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a n essential step .

For solving this attribute selection problem, we have some criterion like **entropy/ information gain, Gini index,** etc.

Entropy

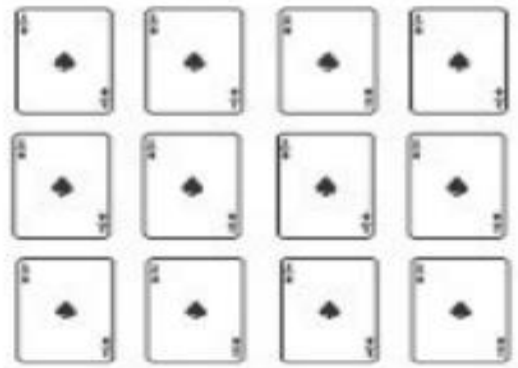
DATA FOLKZ®
#CATAPULT DATA LEADERS

Entropy is the measurement of impurity or randomness in the data . In other words, how much variance the data has or unpredictability of a dataset.

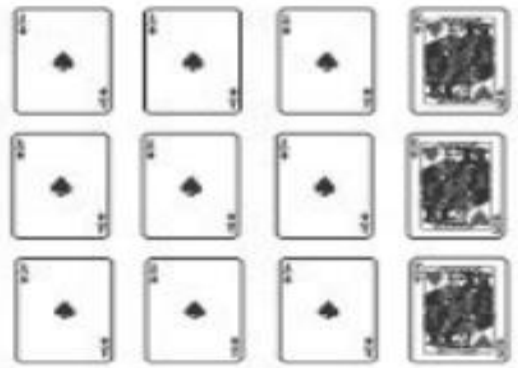
The value of entropy lies between 0 to 1 ,
Where entropy = 0 means no impurity
And entropy = 1 means high impurity .

The variable/ Descriptive feature with the lower entropy is used for split .

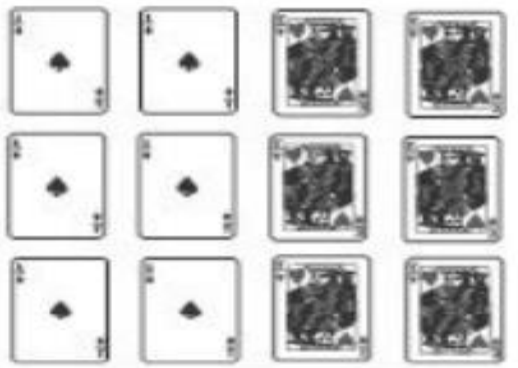
The entropy of different sets of playing cards measured in bits.



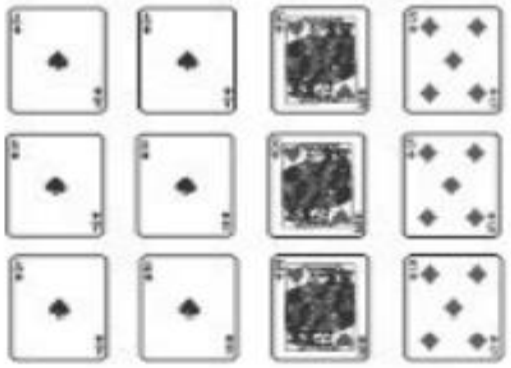
(a) $H(card) = 0.00$



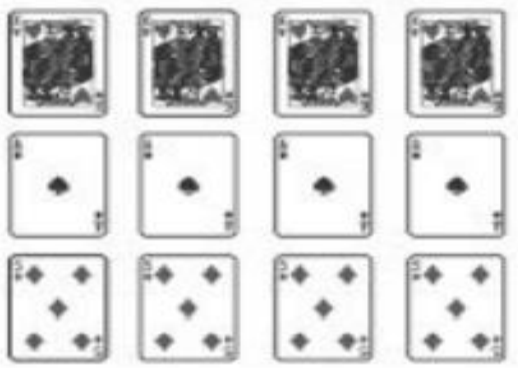
(b) $H(card) = 0.81$



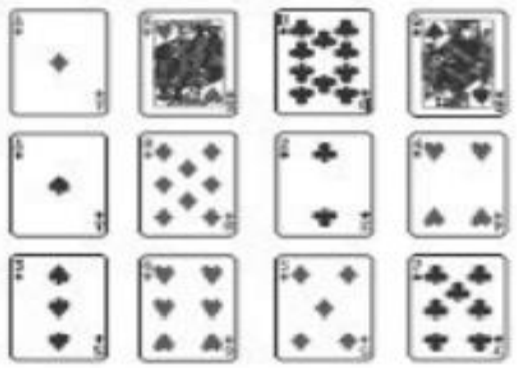
(c) $H(card) = 1.00$



(d) $H(card) = 1.50$



(e) $H(card) = 1.58$



(f) $H(card) = 3.58$

Probability to Entropy

We can transform the probabilitiy of randomly selecting an element from a set to entropy values.

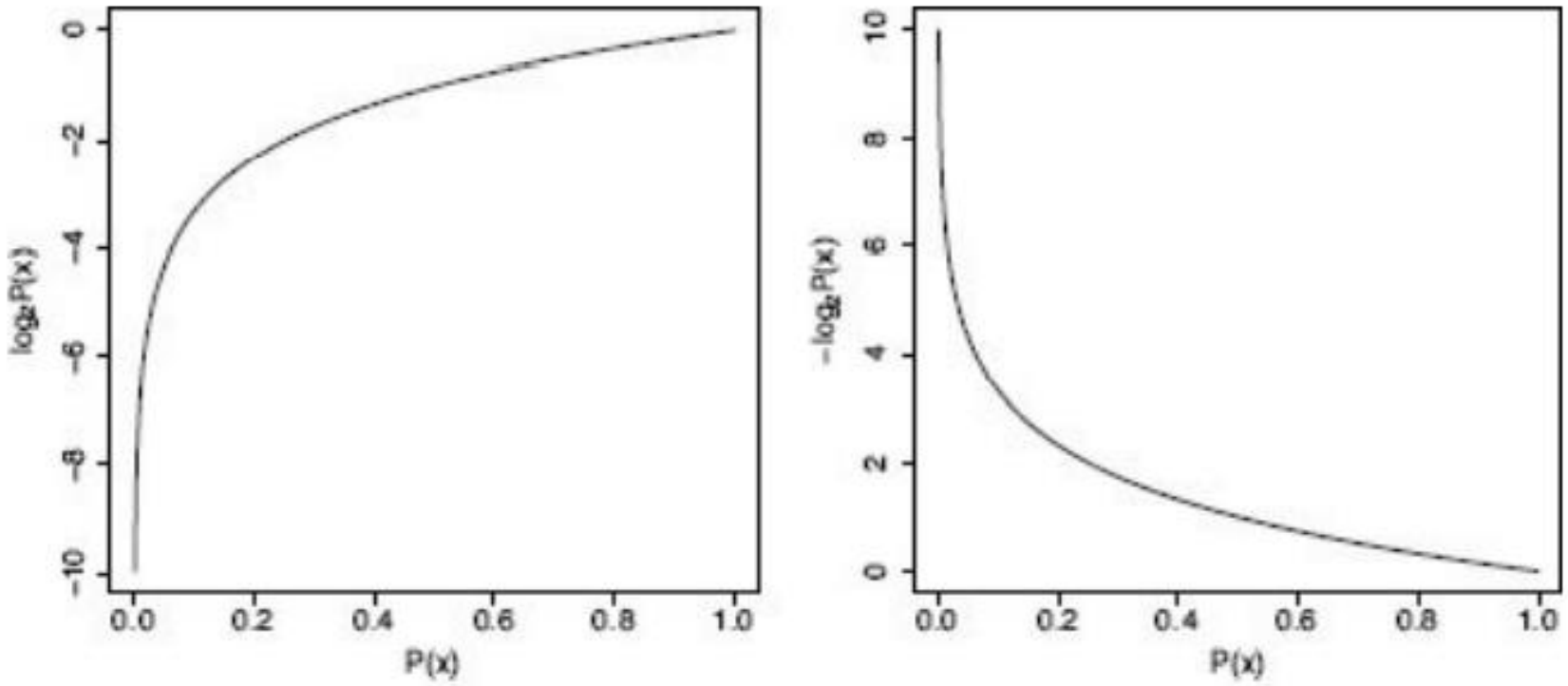
An outcome with a large probability should map to a low entropy value, while an outcome with a small probability should map to a large entropy value.

The binary logarithm (a logarithm to the base 2) of probabilities ranging from 0 to 1 does almost exactly the transformation that we need.

The logarithm function returns large negative numbers for low probabilities, and small negative numbers for high probabilities

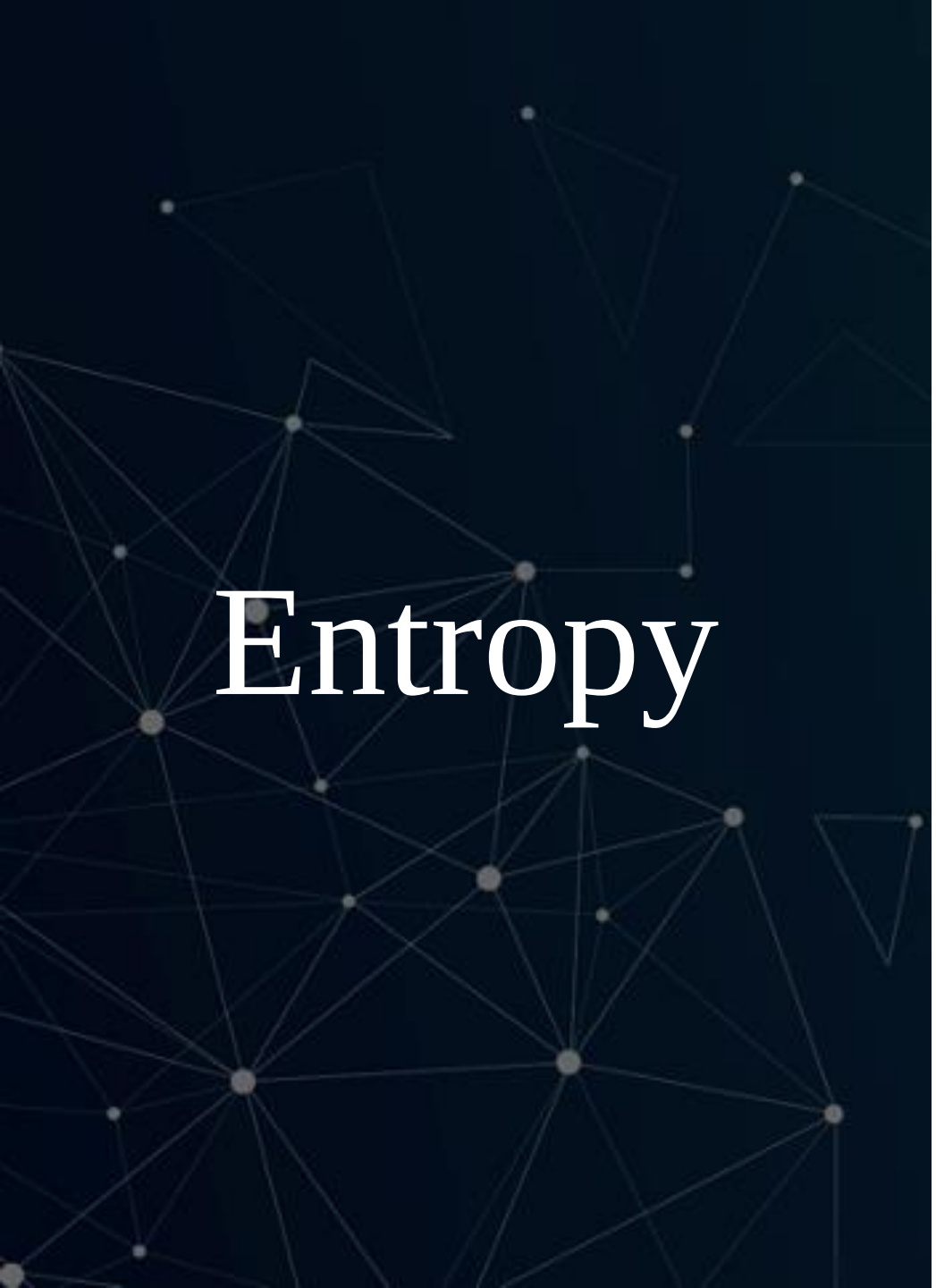
The entropy of different sets of playing cards measured in bits.

We see that the logarithm function returns large negative numbers for low probabilities, and small negative numbers for high probabilities



(a) $P(x)$ and $log_2 P(x)$

(b) $P(x)$ and $-log_2 P(x)$



Entropy

DATA FOLKZ®
#CATAPULT DATA LEADERS

Mathematical equation for entropy:

$$H = -\sum p(x)\ log\ p(x)$$

It is negative summation of probability times the log of probability of item x.

where p(x) is the probability of randomly picking an element of class.

Employee count has single class : 1
It will have no randomness which means it has lowest entropy (almost 0).

Education has multiple classes : 2,1,4,3
It will have high randomness which means it has higher entropy.

Example of Entropy calculation :

DATA FOLKZ®
#CATAPULT DATA LEADERS

$$
\begin{aligned}
H(card) &= -\sum_{i=1}^{52} P(card = i) \times log_2(P(card = i)) \\
&= -\sum_{i=1}^{52} 0.019 \times log_2(0.019) \\
&= -\sum_{i=1}^{52} -0.1096 \\
&= 5.700\ bits
\end{aligned}
$$

Example of Entropy calculation :

DATA FOLKZ®
#CATAPULT DATA LEADERS

$$
\begin{aligned}
H(suit) &= -\sum_{l \in \{\heartsuit,\clubsuit,\diamondsuit,\spadesuit\}} P(suit = l) \times log_2(P(suit = l)) \\
&= -\big((P(\heartsuit) \times log_2(P(\heartsuit))) + (P(\clubsuit) \times log_2(P(\clubsuit))) \\
&\quad + (P(\diamondsuit) \times log_2(P(\diamondsuit))) + (P(\spadesuit) \times log_2(P(\spadesuit)))\big) \\
&= -\Big(\Big(\tfrac{13}{52} \times log_2\big(\tfrac{13}{52}\big)\Big) + \Big(\tfrac{13}{52} \times log_2\big(\tfrac{13}{52}\big)\Big) \\
&\quad + \Big(\tfrac{13}{52} \times log_2\big(\tfrac{13}{52}\big)\Big) + \Big(\tfrac{13}{52} \times log_2\big(\tfrac{13}{52}\big)\Big)\Big) \\
&= -((0.25 \times -2) + (0.25 \times -2) + (0.25 \times -2) + (0.25 \times -2)) \\
&= 2\ bits
\end{aligned}
$$

Example of Entropy calculation :

DATA FOLKZ®
#CATAPULT DATA LEADERS

An email spam prediction dataset.

ID	Suspicious Words	Unknown Sender	Contains Images	Class
376	true	false	true	spam
489	true	true	false	spam
541	true	true	false	spam
693	false	true	true	ham
782	false	false	false	ham
976	false	false	false	ham

Example of Entropy calculation :

DATA FOLKZ®
#CATAPULT DATA LEADERS

$$
\begin{aligned}
H(t,\mathcal{D}) &= -\sum_{l \in \{spam,ham\}} (P(t=l) \times log_2(P(t=l))) \\
&= -\big((P(t=spam) \times log_2(P(t=spam)) \\
&\qquad + (P(t=ham) \times log_2(P(t=ham))\big) \\
&= -\left(\left({}^{3}/_{6} \times log_2({}^{3}/_{6}) \right) + \left({}^{3}/_{6} \times log_2({}^{3}/_{6}) \right) \right) \\
&= 1\ bit
\end{aligned}
$$