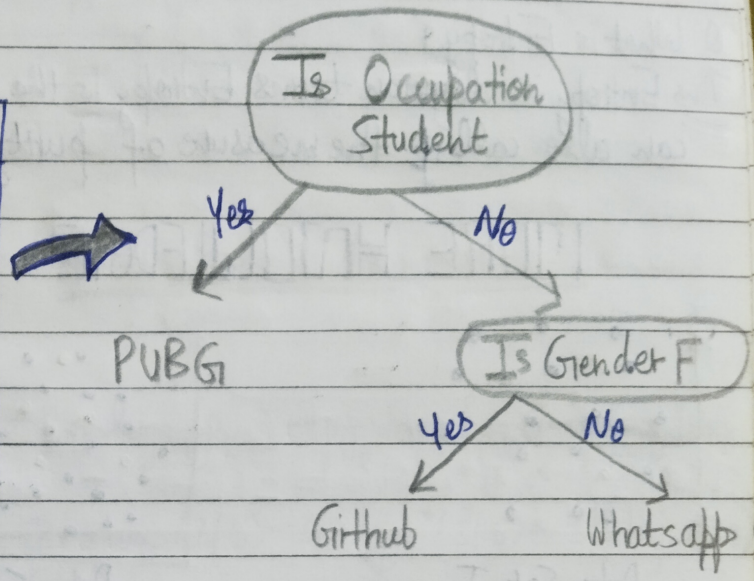
Date.....

Decision Tree

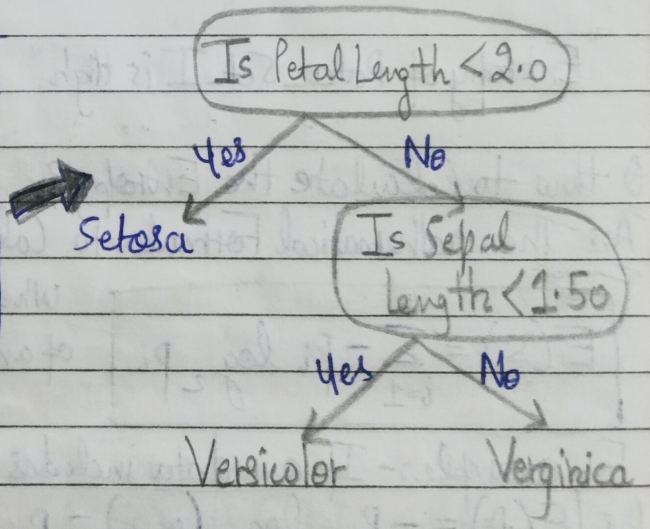
Example 1:-

Gender	Occupation	Suggestion
F	Student	PUBG
F	Programmer	Github
M	Programmer	Whatsapp
F	Programmer	Github
M	Student	PUBG
M	Student	PUBG

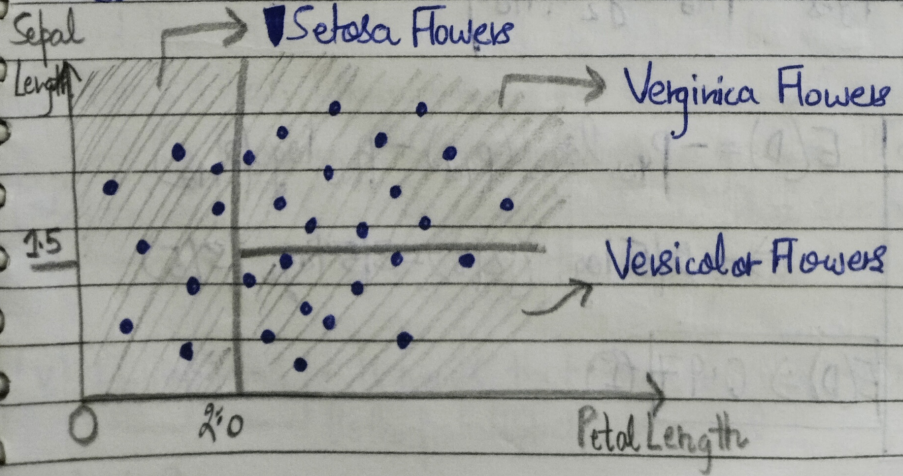


Example 2:-

Petal Length	Sepal Length	Type
1.34	0.34	Setosa
3.45	1.45	Versicolor
1.69	0.98	Setosa
2.56	1.79	Virginica
3.00	1.13	Versicolor
1.3	0.88	Setosa



Geometric Intuition



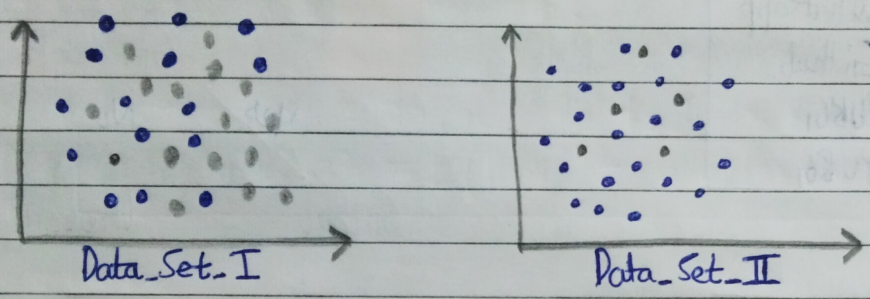
Date.....

Entropy in Decision Trees

Q What is Entropy?

Ans Entropy in layman terms, Entropy is the measure of disorder. Or you can also call it the measure of purity/impurity. Let's see an example:-

MORE KNOWLEDGE LESS ENTROPY



"Entropy of Data_Set_I is High."

Q How to Calculate the Entropy?

Ans The mathematical Formula to Calculate Entropy is:-

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

Where 'pi' is simply the frequent probability of an element/class 'i' in our data.

For example:- If our data includes 2 classes only. "Yes" and "No".

Ans $$E(D) = -p_{yes} \log_2(p_{yes}) - p_{no} \log_2(p_{no})$$

Example:-(1)

Salary	Age	Purchase
20000	21	Yes
10000	45	No
60000	27	Yes
15000	31	No
12000	18	No

Data_Set_I

$$E(D) = -p_{Yes} \log_2(p_{Yes}) - p_{No} \log_2(p_{No})$$

$$\Rightarrow -2/5 \log_2(2/5) - 3/5 \log_2(3/5)$$

$$E(D) \Rightarrow 0.97 \quad (I)$$

Spiral

Date.....

Data_Set_II

Salary	Age	Purchase
34000	31	No
15000	25	No
69000	57	Yes
25000	21	No
32000	28	No

$\Rightarrow E(D) = -P_{Yes}\log_2(P_{Yes}) - P_{No}\log_2(P_{No})$

$\Rightarrow -1/5\log_2(1/5) - 4/5\log_2(4/5)$

$E(D) \Rightarrow 0.72$ (II)

From (I) and (II), Entropy of (I) is High....

Example:- (2)

Salary	Age	Purchase
20000	21	Yes
10000	45	No
60000	47	Yes
15000	31	No
30000	19	Maybe
12000	40	No
40000	35	Maybe
20000	29	Maybe

$\Rightarrow E(D) = -P_{Yes}\log_2(P_{Yes}) - P_{No}\log_2(P_{No}) - P_{Maybe}\log_2(P_{Maybe})$

$\Rightarrow -2/8\log_2(2/8) - 3/8\log_2(3/8) - 3/8\log_2(3/8)$

$E(D) \Rightarrow 1.56$

Data_Set_III

Observations.....

(i) As, the Knowledge decreases, the Entropy Increases.

(ii) For 2 Class Classification, the minimum Value of Entropy will be = "0" and the Maximum Value of Entropy will be = "1".

(iii) For 3 Class Classification, the minimum Value of Entropy will be = "0" and the Maximum Value of Entropy is greater than > 1.

(iv) **Entropy = 0** → It means that the Data Set has too much Knowledge to predict the new input...

(v) **Entropy = 1** → It means that the Data Set values let's say "Yes" and "No" Both probabilities are equal....

Spiral