1. For the following data set, apply ID3 separately, and show all steps of derivation (computation, reasoning, developing / final decision trees, and rules).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | color | shape | size | Pattern (shirt) | class |
| 1 | red | square | big | Checked | + |
| 2 | blue | square | big | striped | + |
| 3 | red | round | small | dotted | - |
| 4 | green | square | small | checked | - |
| 5 | red | round | big | striped | + |
| 6 | green | round | big | dotted | - |

𝐸𝑛𝑡𝑟𝑜𝑝𝑦(𝑡) = − ∑ p(𝑗|𝑡) log2 p(𝑗|𝑡)

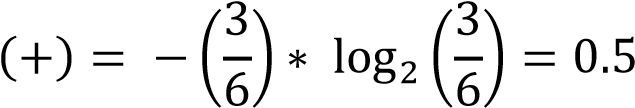
𝑗

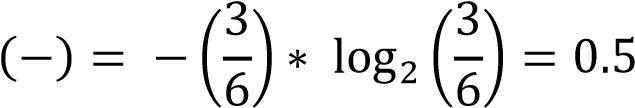
Here class is the target attribute and has two values ( + and -). So it is a binary classification problem.

For a binary classification problem

* If all examples are positive or all are negative, then entropy will be ***zero***i.e. low.
* If half of the examples are of positive class and half are of negative class, then entropy is **one** i.e. high.
  1. Calculating Initial Entropy

Out of 6 instances, 3 are + and 3 are –

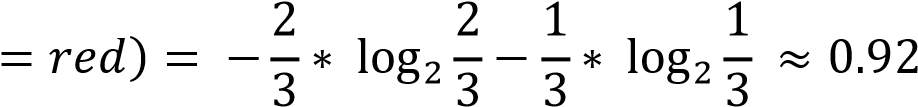
𝑃

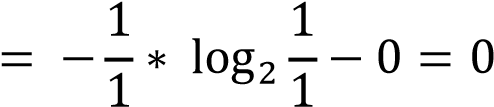
𝑃

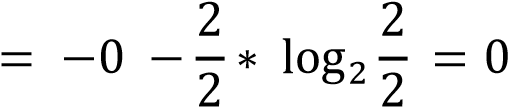
𝐸𝑛𝑡𝑟𝑜𝑝𝑦(𝑡) = 𝐸(𝑡) = 0.5 + 0.5 = 1

Note: 1 indicates that the classes are highly impure. It is true in our case as there are equal number of observations with target class + and –

**For attribute color**

𝐸(𝐶𝑜𝑙𝑜r 

𝐸(𝐶𝑜𝑙𝑜𝑟 = 𝑏𝑙𝑢𝑒) 

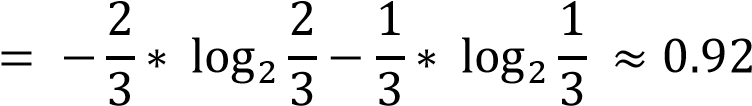
𝐸(𝐶𝑜𝑙𝑜𝑟 = 𝑔𝑟𝑒𝑒𝑛) 

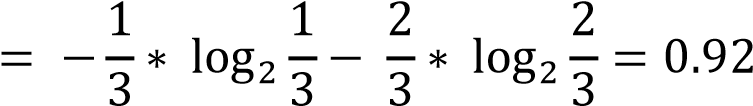
Average Entropy = (0.92) + (0) +  (0) = 0.46

Gain(Outlook) = 1 − 0.46 = 𝟎. 𝟓𝟒



**For attribute shape**

𝐸(𝑠ℎ𝑎𝑝𝑒 = 𝑠𝑞𝑢𝑎𝑟𝑒) 

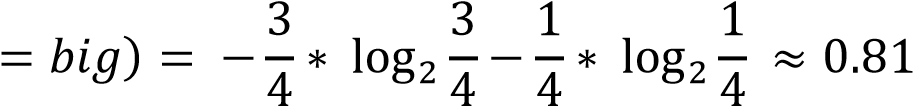
𝐸(𝑠ℎ𝑎𝑝𝑒 = 𝑟𝑜𝑢𝑛𝑑) 

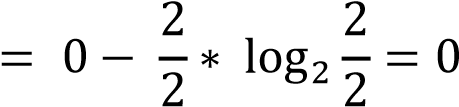
Average Entropy = (0.92) + (0.92) = 0.92

Gain(Outlook) = 1 − 0.92 = 𝟎. 𝟎𝟖



**For attribute size**

𝐸(𝑠𝑖𝑧𝑒 

𝐸(𝑠𝑖𝑧𝑒 = 𝑠𝑚𝑎𝑙𝑙) 

Average Entropy = (0.81) + (0) = 0.54

Gain(Outlook) = 1 − 0.54 = 𝟎. 𝟒𝟔



Feature ‘color’ provides more information on the ‘class’ as it has the highest information gain and hence will be chosen as the first splitting attribute

Diagram

Description automatically generated

Likewise, we create the entire tree by selecting the splitting attribute as the attribute that gives the most information.



**For attribute pattern of Shirt**

E(Pattern=checked) = -1/2 \* log2 (1/2)- 1/2 \* log2(1/2) = 1

E (Pattern = Striped) = -2/2 \* log2 (2/2)- 0/2 \* log2(0/2) = 0

E (Pattern = dotted) = -0/2 \* log2 (0/2)- 2/2 \* log2(2/2) = 0

Average Entropy = 2/6 (1) + 2/6(0) +2/6(0) = 0.33

Gain (Outlook) = 1 – 0.33= 0.67

Recalculation of Color Shape and Size after adding new CHECKED PATTERN attribute

E(color = red) = -(1/1) log2 (1/1) -0 =0

E(color = green) = 0 – (1/1) log2(1/1) =0

Average Entropy = 0

Gain(Outlook)= 1 -0 = 1

E(Shape = square) = -(1/2) log2(1/2) -(1/2) log2(1/2) =1

Average Entropy = 1

Gain(Outlook)= 1 -1 = 0

E(Size = big)= – (1/1) log2(1/1) – 0 =0

E(Size = small)= 0 – (1/1) log2(1/1) =0

Average Entropy = 0

Gain(Outlook)= 1 -0 = 1



After added attribute

Diagram

Description automatically generated

Part 2:

1. Understand what impact may happen to your created tree, if you later add a new missing attribute after creating the tree?

After adding the new attribute, the pattern of shirts got the highest attribute of o.67 thus the new tree started with the pattern of shirts. Before adding the new attribute color had the highest gain of 0.54 which gave us the information if the class will be positive or negative.

1. What are some of the different possible changes you may expect to see on the classification decision tree you just created?

Adding a new attribute, helped us to verify our result against the original tree. By adding a new attribute, the result of the tree came out to be equal to the original tree

1. What if a data scientist provided his or her results with high confidence, by missing this attribute altogether?

If a data scientist provides a result with a missing attribute, the result will be uncertain. For example, if the scientist has the attribute shape missing with only color and size the result might come out to be different than the original one. The result of class to be positive or negative is a combination of all the attributes.

1. What if his or her results are used for decision making on how many million more shirts to produce for the next year?

Any missing attribute will result in an uncertainty. Especially when a large number of shirts are being produced it is very important to have all the possible sicarios of the customer.

1. Do you think the data scientist surprises and makes an impact on the manager and CEO in case he or she discovers the new attribute and it's influence in getting more reliable results valuable to the company?

Having a new attribute helps the data scientist to split the data in multiple pathways which will create more certainty in determining the result. The scientist will easily be able to identify if the class will result in a positive or negative direction. As more attributes are added the level of certainty increases.