PUSL2076 Data Programming in R

In Class Assessment 1 – Entropy
30 Minutes

Consider the following table to answer subsequent questions under Q4. The table shows a dataset of COVID-19 test results of 10 individuals against three attributes.

#	Cold	Temperature	Fatigue	Result
1	No	Normal	No	- Negative
2	No	High	Yes	- Negative
3	No	High	Yes	**+ Positive**
4	No	Very High	Yes	**+ Positive**
5	Yes	Normal	No	- Negative
6	Yes	Normal	Yes	**+ Positive**
7	Yes	High	No	- Negative
8	Yes	High	Yes	**+ Positive**
9	Yes	Very High	No	**+ Positive**
10	Yes	Very High	Yes	**+ Positive**

Question 1

Calculate the ***Entropy*** for the dataset given above considering the target variable ***Result***. Show your calculations clearly.

Question 2

In order to decide the first split of the decision tree, it has been decided to measure the *Entropy* for three possible cases. It has been found that the ***Entropy*** for two cases as given below.

- *Entropy(Result | Cold) = 0.93*
- *Entropy(Result | Fatigue) =0.70*

Find the third case ***Entropy(Result | Temperature)*** by showing your calculation steps clearly.