Department of Computer Science

National University of Computer and Emerging Sciences

Lahore Campus

Data Mining

Assignment 1

Due Date: 12 Feb 2023, 11:59 PM

Submit a handwritten PDF document in Google Classroom.

No late submissions are allowed.

Show your complete work to get points.

It's your responsibility to submit clear readable images. If we will not be able to read it, we will not grade it.

Question 1: 3 marks

Compute the entropy for the random variable x

a. P(x=TRUE) = 0.5 P(x=False) = 0.5

b. P(x=TRUE) = 1. P(x=False) = 0

c. P(x=TRUE) = 0.2. P(x=False) = 0.8

Question 2: 18 points

Example	Pointed	Threaded	Width	Class
Example #1	No	Yes	Slim	Nail
Example #2	No	Yes	Slim	Nail
Example #3	No	No	Medium	Nail
Example #4	Yes	Yes	Fat	Nail
Example #5	Yes	Yes	Medium	Bolt
Example #6	No	Yes	Fat	Bolt
Example #7	No	Yes	Medium	Bolt
Example #8	No	No	Fat	Bolt

Let Class be the class label attribute.

(a) Calculate the Entropy of each feature. 3 points

(b) Calculate the Gini Index of each feature. 3 points

(c) Calculate the Information Gain of each feature and then report the best feature to split the examples using the information gain metric. 6 points

(d) Calculate the gain ratio for each feature and then report the best feature to split the examples using the gain ratio metric. 6 points

Question3: 9 points

(a) Find P (A = t)

(b) P (B = f)

(c) P (C = t)

(d) P (B = t/ C = t)

(e) P (A = f/ C = t)

(f) P (A= t/C = f)

(g) P (A = f, C = t)

(h) P (A = t, C = t)

(i) P (A = t, b = f)

A	B	C	P(a,b,c)
t	t	t	0.03
t	t	f	0.12
t	f	t	0.17
t	f	f	0.18
f	t	t	0.03
f	t	f	0.12
f	f	t	0.24
f	f	f	0.11