Assignment 1 – Data Mining (2023)

Q1. Describe the following terms:
a) Nominal Attributes b) Outlier Analysis (2)

Q2. What are the advantages of data transformations? Describe two strategies used for data transformation. (3)

Q3. What is meant by asymmetric attributes? Give an example of a dataset having asymmetric binary attributes and an example having asymmetric discrete or continuous attributes. (3)

Q4.Let x denote the nominal feature having values {Good, Better, Best}. How will this data be binarized? How many bits will be required? (2)

Q5. List down an advantage and a disadvantage of leave-one-out approach used in cross-validation for evaluating the performance of the classifier.
(2)

Q6. Given the points p1 (0,2), p2 (2,0), p3(3,1) and p4(5,1), find the euclidean distance between points p1 and p2, p3 and p2.
(2)

Q7. Let confusion matrix for a 2-class problem is given as follows: (3)

		Predicted Class	
		Class=1	Class=0
Actual Class	Class=1	45	10
	Class=0	25	20

Calculate the accuracy, recall, precision, sensitivity, specificity, F_1 measure.

Q8. Consider the following data set for a binary class problem: (3)

A	B	Class Label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

Calculate the gain in the Gini Index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?