B. Tech Odd Semester Examination, February, 2023

Computer Science & Engineering

(5th Semester)

Course No.: CSE-505(6)

(Data Mining)

Full Marks: 50

Pass Marks: 25

Time: 2 hours

Note:
1. Attempt any five questions.
2. Begin each answer in a new page.
3. Answer parts of a question at a place.
4. Assume reasonable data wherever required.
5. The figures in the right margin indicate full marks for the question.
6. All the mathematical symbols and abbreviations have their usual meanings.

1. (a) What is Clustering? Show the flowchart of K-means clustering. 2+2

(b) Apply the K-Means clustering for the following given dataset for two clusters. Tabulate all the assignments of clusters and get the final centroid values.

Sample	X	Y
A	1	1
B	2	1
C	4	3
D	5	4

Consider Sample A and C as initial cluster centres in Clusters C1 and C2 respectively. 6

2. (a) Explain the different types of attributes in data mining. 5

(b) Explain the major tasks in data preprocessing. 5

3. (a) Apply the K-Medoid clustering for the following given dataset for two clusters. (Show one iteration)

{(8,7), (3,7), (4,9), (9,6), (8,5), (5,8), (7,3), (8,4), (7,5), (4,5)}

Consider C1 = (8,5) and C2 = (4,5) and find the total cost. 6

(b) Find the interquartile range for the following series of numbers:

80, 45, 62, 47, 52, 52, 53, 55, 56, 58. 2

(c) Give the five number summary for the following data set:

79, 53, 82, 91, 87, 98, 80, 93 2

4. (a) Write briefly on Logistic Regression. 5

(b) Find the least square Regression line y = ax + b for the following data. Also estimate the value of y when x = 8. 5

x	y
2	4
3	5
5	7
7	10
9	15

5. Explain briefly about mining of time series data. And also explain the major components to characterize time series data. How can the trend of the data be determined? Explain. 4+4+2=10

6. (a) What is Back Propagation. Explain the working of a Back Propagation with the help of a diagram to calculate the error factor. 1+5

(b) Mention the advantages and disadvantages of backpropagation. 4

7. (a) Consider the given dataset and apply Naive bayes algorithm and using these probabilities estimate the probability values for the new instance – (Color= Green, legs=2, Height=Tall, and Smelly=No) and classify it. 6

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

(b) Explain Bayesian Classification 4

8. (a) Explain the different types of clustering methods. 5

(b) Give a short note on the following:

i) Association Analysis 3

ii) Artificial Neural Network 2

****