



WEST BENGAL STATE UNIVERSITY
B.Sc. Honours 5th Semester Examination, 2021-22

CMSADSE02T-COMPUTER SCIENCE (DSE1/2)

Time Allotted: 2 Hours

Full Marks: 40

*The figures in the margin indicate marks of question.
Candidates should answer in their own words and adhere to the word limit as practicable.
All symbols are of usual significance.*

GROUP-A

1. Answer any **four** questions from the following: 2×4 = 8
- (a) What do you mean by Clustering?
 - (b) Differentiate between OLTP and OLAP.
 - (c) What do you mean by outlier detection?
 - (d) What is Knowledge Discovery in Databases?
 - (e) Why data preprocessing is required?
 - (f) How can you check the efficiency of a classifier model?
 - (g) Explain the difference between data mining and data warehousing.

GROUP-B

Answer any four questions from the following

8×4 = 32

2. (a) What are the steps of KDD? 3+2+3
- (b) What do you understand by nominal attribute? Give a suitable example.
 - (c) What are the major tasks in data preprocessing?
3. (a) Execute Apriori algorithm in the below given dataset: 6+2

Tid	Items
1	ACD
2	ABCE
3	BCE
4	BE
5	ABCE
6	BCE

- (b) Define Gini impurity measure.

4. (a) Explain the confusion matrix for a 2-class problem. 3+3+2
 (b) Explain information gain in Decision Tree based classification.
 (c) What do you mean by feature selection? State with a suitable example.

5. Consider these 10 points (2, 3, 5, 6, 8, 9, 11, 13, 15, 16). Perform K-means clustering with $K = 2$ considering the first cluster centers 2 and 15. Now find the following: 3+3+2
 (i) SSE after first iteration.
 (ii) Final clusters.
 (iii) SSE after the termination of the K-means algorithm.

6. (a) Explain the following terms in the context of association rule mining: 3×2 = 6
 (i) Support of an itemset.
 (ii) Frequent closed itemset.
 (iii) Lift of a rule.

- (b) What is the time complexity of computing the supports for m number of itemsets in a database of n transactions? 2

7. (a) Consider the following confusion matrix and compute the values of precision, recall, false positive rate from it. 2×3 = 6

	Actually malignant	Not actually malignant
Predicted malignant	100	10
Predicted not malignant	5	50

- (b) What are the limitations of K-means clustering? 2

8. Write short notes on any **two** of the following: 4+4
 (a) Divisive hierarchical clustering
 (b) Logistic regression
 (c) K-nn Classification
 (d) Apriori algorithm.

N.B. : Students have to complete submission of their Answer Scripts through E-mail / Whatsapp to their own respective colleges on the same day / date of examination within 1 hour after end of exam. University / College authorities will not be held responsible for wrong submission (at in proper address). Students are strongly advised not to submit multiple copies of the same answer script.

—x—