WEST BENGAL STATE UNIVERSITY

B.Sc. Honours 5th Semester Examination, 2020, held in 2021

CMSADSE02T-COMPUTER SCIENCE (DSE1/2)

DATA MINING

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.

GROUP-A

1. Answer any *four* questions   2×4 = 8

    (a) Why text mining is required?

    (b) What do you mean by an outlier?

    (c) Explain the difference between data mining and data warehousing.

    (d) What is an output of Apriori Algorithm?

    (e) Differentiate between Clustering and Classification.

    (f) What do you understand by predictive data mining?

    (g) What is Knowledge Discovery in Databases?

GROUP-B

Answer any *four* questions from the following   8×4 = 32

2. Differentiate between OLAP and OLTP. Explain different OLAP operations with   8
    the help of examples.

3. Explain the different methods of Data Cleaning and Data Transformation.   8

4. Write the algorithms for K-means clustering. Compare it with k-nearest neighbor   8
    algorithm.

5. How Decision Trees assist in the process of classification? Explain with the help   6+2
    of an example.

6. What is associative classification in data mining? Why is associative   2+4+2
    classification being able to achieve higher classification accuracy than a classical
    decision tree method?

7. Define information gain. How can we measure information gain? Define the concept of Splitting attribute and Splitting criterion. | 2+3+3

8. Write short notes on any *two* of the following: | 4+4
    (a) Hierarchical clustering
    (b) Support Vector Machine
    (c) Linear Regression
    (d) k-means clustering.

    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.*

——✕——