E 6236

Reg. No.................................

Name....................................

B.Sc./B.C.A. DEGREE (C.B.C.S.S.) EXAMINATION, SEPTEMBER 2024

Sixth Semester

Choice Based Core Course—DATA MINING

(Common for B.Sc. Computer Science, B.C.A., and B.Sc. Computer Applications]

(Triple Main)

{Prior to 2013 Admissions}

Time : Three Hours   Maximum Weight : 25

Section A (Objective Type Questions)

Answer **all** *questions.*
A bunch of **four** *questions carries a weight of* 1.

I   1   KDD stands for ———————.

    (a)   Knowledge discovery of data.

    (b)   Knowledge discovery of decisions.

    (c)   Knowledge discovery of databases.

    (d)   Knowledge discovery of dimensions.

  2   A confidence of 50 % confidence and 1 % support if someone buys a computer and is likely to buy a printer means ———————.

    (a)   1 % chance that he will buy software as well and 50 % of all the transactions under analysis show that computer and printer are purchased together.

    (b)   50 % chance that he will buy software as well and 1 % of all the transactions under analysis show that computer and printer are purchased together.

    (c)   51 % chance that he will not buy software as well and there is no way to looks at the transactions under analysis to show that computer and printer are purchased together.

    (d)   None of the above.

Turn over

3   PCA stand for ——————.

    (a)   Parametric Component Analysis.

    (b)   Problem Component Analysis.

    (c)   Principal Component Analysis.

    (d)   Preprocessing Component Analysis.

4   Sampling is a —————— process.

    (a)   Data integration.   (b)   Data cleaning.

    (c)   Data discretization.   (d)   Data Reduction.

II.   5   Binning is a method for data ——————.

     6   Correlation analysis is a method for data

     7   OLAP stands for ——————.

     8   Market basket analysis is an example of ——————.

III   9   DWT stands for ——————.

    (a)   Database Wavelet Transform.

    (b)   Discrete Wavelet Transform.

    (c)   Dimensional Wavelet Transform.

    (d)   None of the above.

10   OLTP stands for ——————.

    (a)   Online transaction processing.

    (b)   Online table processing.

    (c)   Online text processing.

    (d)   None of the above.

11   k-medoid is a —————— method.

    (a)   Classification.   (b)   Hierarchical.

    (c)   Density.   (d)   Partitioning.

12   Agglomerative is a ——————— method.

   (a) Classification.   (b) Hierarchical.

   (c) Density.   (d) Partitioning.

IV.   13 k-means is a ——————— method.

14   SVM stands for ———————.

15   Decision tree is a ——————— method.

16   BIRCH is a ——————— method.

$(4 \times 1 = 4)$

Section B (Short Answer Questions)

Answer any **five** *questions.*
Each question carries 1 *weight.*

17   What is an ER model ? Give examples.

18   What is a data cube ? Give examples.

19   What is numerosity reduction ?

20   What is a Snowflake ?

21   What is a maximal frequent itemset ?

22   What is hierarchical clustering ?

23   What is DBSCAN ?

24   What is spatial data mining ?

$(5 \times 1 = 5)$

Section C (Short Essays/Problems)

Answer any **four** *questions.*
Each question carries 2 *weight.*

25   Outline the various steps in the data mining process.

26   Briefly, explain Attribute Subset Selection with an example.

27   Briefly, explain various OLAP operations.

Turn over

28   Briefly, explain Apriori algorithm.

29   Briefly, explain Bayes classification method.

30   Briefly, explain multimedia data mining.

$(4 \times 2 = 8)$

Section D (Essay Questions)

Answer any **two** *questions.*
Each question carries 4 weight.

31   Describe in detail the various methods for data transformation.

32   Describe in detail the CART method and various parameters arising from the method.

33   Describe in detail various partitioning methods for cluster analysis.

$(2 \times 4 = 8)$