

Total number of printed pages = 5

19/6th Sem/DCSE 612

2022

**DATA MINING**

Full Marks – 100

Time – Three hours

The figures in the margin indicate full marks for the questions.

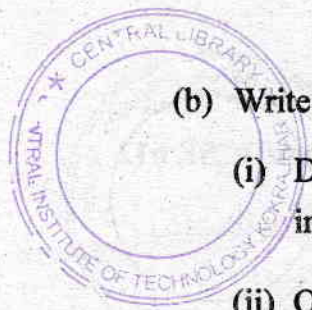
Answer any *five* questions.

1. Answer the following questions :  $1 \times 8 = 8$

(a) Match the following “Clustering Approach” and “Clustering Methods” :

Clustering Approach	Clustering Methods
(a) Hierarchical approach	(i) BIRCH
(b) Hierarchical approach	(ii) k-medoids
(c) Density approach	(iii) CLARANS
(d) Density approach	(iv) CLIQUE
(e) Grid-based approach	(v) STING
(f) Grid-based approach	(vi) Agnes
(g) Partitioning approach	(vii) DBSACN
(h) Partitioning approach	(viii) OPTICS

[Turn over



(b) Write True or False :  $1 \times 12 = 12$

- (i) Data mining uses historical data to improve the decision.
- (ii) OLAP is a major task of traditional relational DBMS.
- (iii) OLTP is a major task of data warehouse system.
- (iv) Data analysis and decision making are processed in OLTP.
- (v) The operation of moving from finer-granularity data to a coarser granularity is called a drill down.
- (vi) The Roll-up operation navigates from less detailed data to more detailed data.
- (vii) The class labels of training data are unknown in supervised learning.
- (viii) In Unsupervised learning, the training data are accompanied by labels indicating the class of the observations.

(ix) Nominal variables take more than 2 states.

(x) A good clustering method will produce high inter-class similarity.

(xi) Ordinal variables can be discrete only.

(xii) The most detailed part of the cube is called a base cuboid.

2. (a) Describe the Numerosity reduction in data reduction and its related Parametric and Non-parametric methods ? 6

(b) What are the four major features of data warehousing ? 6

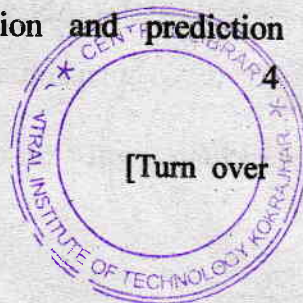
(c) Given temperature values (in Celsius) such as 50, 44, 22, 55, 48, 12, 15, 11, 8, 42, 35 and 14. Apply the binning methods to partition the data. 8

3. (a) Describe the Hierarchical clustering approach and also the algorithm. 6

(b) Explain the classification and prediction methods. 4

43/19/6th Sem/DCSE 612 (3)

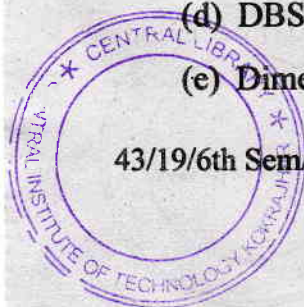
[Turn over



- (c) Explain Naive Bayesian Classification (NBC) with the steps. 5
- (d) Explain the multi-layer feed forward neural network classification with a diagram. 5
4. (a) What is the role of support and confidence in Association Rule Mining? 10
- (b) A database has five transactions. Let  $\text{min\_sup} = 60\%$  and  $\text{min\_conf} = 75\%$ . Find all frequent item sets using Apriori. 10

TID	Item sets
T100	{D, E, B, O, R, G, A, O, N}
T200	{B, A, S, U, G, A, O, N }
T300	{K, A, R, I, G, A, O, N }
T400	{J, A, I, G, A, O, N }

5. Write short notes on any *four* of the following :  
5×4=20
- (a) Covariance (Numeric Data)
- (b) Back propagation based neural network
- (c) Data Mining Query Language
- (d) DBSCAN algorithm
- (e) Dimension table, Fact table and Data cube.



6. Differentiate between the following any *four* :

5×4=20

- (a) K-means and K-medoids
- (b) Roll up and drill down
- (c) CURE and OPTICS
- (d) Partitioning algorithm and Sampling algorithm
- (e) OLAP and OLTP.

