

PSG COLLEGE OF ARTS & SCIENCE  
(AUTONOMOUS)

BVoc DEGREE EXAMINATION DECEMBER 2022  
(Fifth Semester)

Branch – NETWORKING AND MOBILE APPLICATION

**DISCIPLINE SPECIFIC ELECTIVE – I : DATA MINING**

Time: Three Hours

Maximum: 75 Marks

**SECTION-A (10 Marks)**

Answer ALL questions

ALL questions carry EQUAL marks (10 x 1 = 10)

1. \_\_\_\_\_ is an essential process where intelligent methods are applied to extract data patterns.  
(i) Data Warehousing (ii) Data mining  
(iii) Text Mining (iv) Data Selection
2. \_\_\_\_\_ is a summarization of the general characteristics or features of a target class of data.  
(i) Data Characterization (ii) Data classification  
(iii) Data Discrimination (iv) Data Selection
3. Histograms use \_\_\_\_\_ to approximate data distribution.  
(i) Binning (ii) Attributes  
(iii) Values (iv) Variables
4. Min-Max normalization performs a linear \_\_\_\_\_.  
(i) Integration (ii) Reduction  
(iii) Transformation (iv) Clustering
5. Data warehouses provide \_\_\_\_\_ tools for the interactive analysis of multidimensional data of varied granularities.  
(i) OLTP (ii) Data warehouse (iii) ROLTP (iv) OLAP
6. \_\_\_\_\_ is a variant of the star schema model where the dimension tables are normalized.  
(i) Star (ii) Snow flakes  
(iii) Fact constellation (iv) All of these
7. What is a heuristic for selecting the splitting criterion?  
(i) Array selection measure (ii) Index selection measure  
(iii) Attribute selection measure (iv) None of the above
8. Association rule mining finds interesting association or correlation relationships among \_\_\_\_\_.  
(i) A large set of data items (ii) Huge amount of data items  
(iii) A small set of data items (iv) None of these
9. \_\_\_\_\_ and \_\_\_\_\_ methods are commonly used in Partitioning.  
(i) Agglomerative and Divisive (ii) Data and object  
(iii) K-means and K-Medoids (iv) None of the above
10. A \_\_\_\_\_ consist of sequence of ordered elements or event or without a concrete notion of time.  
(i) Text Data Base (ii) Time Serious Data B  
(iii) Spatial Data Base (iv) Sequence Data Base

**SECTION - B (25 Marks)**

Answer ALL questions

ALL questions carry EQUAL Marks (5 x 5 = 25)

11. a) Illustrate the architecture of typical data mining system.  
OR  
b) Summarize the major issues of data mining.
12. a) Explain about data transformation by normalization.  
OR  
b) Describe about attribute subset selection.
13. a) Illustrate about the concept hierarchies.  
OR  
b) Summarize the operations of OLAP.
14. a) Explain market basket analysis.  
OR  
b) Describe the general approaches to classification.
15. a) Explain the requirements for cluster analysis.  
OR  
b) Classify the various clustering methods.

**SECTION -C (40 Marks)**

Answer ALL questions

ALL questions carry EQUAL Marks (5 x 8 = 40)

16. a) Categorize the different kinds of data.  
OR  
b) Summarize the data mining functionalities.
17. a) Elucidate about data integration.  
OR  
b) Outline about Data cleaning with example.
18. a) Discuss the architecture of dataware house.  
OR  
b) Enumerate the different types of schemas in multidimensional databases.
19. a) Discuss the Apriori algorithm with suitable illustrations.  
OR  
b) Discuss the Decision Tree Induction with example.
20. a) Examine the k-means clustering algorithm.  
OR  
b) Discuss about mining graphs and networks.

Z-Z-Z

END