

PSG COLLEGE OF ARTS & SCIENCE  
(AUTONOMOUS)

BVoc DEGREE EXAMINATION MAY 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. Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabeled data?  
a) Supervised learning  
b) Unsupervised learning  
c) Hybrid learning  
d) Reinforcement learning
2. Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?  
a) Warehousing  
b) Data Mining  
c) Text Mining  
d) Data Selection
3. Which of the following also used as the first step in the knowledge discovery process?  
a) Data selection  
b) Data cleaning  
c) Data transformation  
d) Data integration
4. Which of the following refers to the steps of the knowledge discovery process, in which these several data sources are combined?  
a) Data selection  
b) Data cleaning  
c) Data transformation  
d) Data integration
5. .... is a good alternative to the star schema.  
a) Star Schema  
b) Snowflake schema  
c) Fact constellation  
d) Star-snowflake schema
6. The full form of OLAP is  
a) Online Analytical Processing  
b) Online Advanced Processing  
c) Online Advanced Preparation  
d) Online Analytical Performance
7. Which of the following is direct application of frequent item set mining?  
a) Social Network Analysis  
b) Market Basket Analysis  
c) Outlier Detection  
d) Intrusion Detection
8. Which of these is not a frequent pattern mining algorithm?  
a) Apriori  
b) growth  
c) Decision Trees  
d) Eclat
9. Which is needed by K-means clustering?  
a) Defined Distance Metric  
b) Number of Clusters  
c) Initial guesses as to cluster centroids  
d) All of these
10. Which function is used for k-means clustering?  
a) k-means  
b) k-mean  
c) Heatmap  
d) None of the mentioned

Cont...

**SECTION - B (35 Marks)**

Answer **ALL** Questions

**ALL** Questions Carry **EQUAL** Marks (5 x 7 = 35)

11. a) Narrate the importance of Cluster Analysis.  
(OR)  
b) Describe Scalability.
12. a) Explain Data Quality for a Data Processing.  
(OR)  
b) Outline the attribute Selection measures in detail.
13. a) Explain Multitier Architecture with Diagram.  
(OR)  
b) Show the Snowflake schema in Data Warehouse model.
14. a) Sketch the Association Rules in Data Mining.  
(OR)  
b) State the Basic concepts in Classification.
15. a) Outline the requirements for an Cluster Analysis.  
(OR)  
b) Analyze the concept of Time Series Clustering Analysis.

**SECTION - C (30 Marks)**

Answer any **THREE** Questions

**ALL** Questions Carry **EQUAL** Marks (3 x 10 = 30)

16. Discuss in detail about the major issues in Data Mining.
17. Elucidate Redundancy and Correlation Analysis.
18. Summarize the purpose of Virtual Warehouse used in Data Warehousing.
19. Explain Apriori Algorithm with example.
20. Justify K-Means Algorithm with example.

Z-Z-Z

END