

PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

^{BSc}
~~BSc~~ DEGREE EXAMINATION MAY 2024
(Fifth Semester)

Branch - NETWORKING AND MOBILE APPLICATIONS

DISCIPLINE SPECIFIC ELECTIVE-I: DATA MINING

Time: Three Hours

Maximum: 50 Marks

SECTION-A (5 Marks)

Answer ALL questions

ALL questions carry EQUAL marks

(5 x 1 = 5)

- 1 Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?
 - (i) Data Ware housing
 - (ii) Data mining
 - (iii) Text Mining
 - (iv) Data selection
- 2 How to handle the missing values in the dataset?
 - (i) Dropping the missing rows or columns
 - (ii) Imputation with mean/median/mode value
 - (iii) Taking missing values into a new row or column
 - (iv) All of the above
- 3 What does OLTP stand for?
 - (i) online transaction processing
 - (ii) outline transaction processing
 - (iii) offline transaction processing
 - (iv) none of these
- 4 From where are classification rules extracted?
 - (i) Branch
 - (ii) Decision tree
 - (iii) Root node
 - (iv) Intermediate node
- 5 _____ partitions the objects into different groups.
 - (i) Classification
 - (ii) Mapping
 - (iii) Clustering
 - (iv) Prediction

SECTION - B (15 Marks)

Answer ALL Questions

ALL Questions Carry EQUAL Marks

(5 x 3 = 15)

- 6 a Explain about different kinds of data.
OR
b How is a data warehouse different from a database?
- 7 a Why Preprocess the Data?
OR
b Illustrate the normalization methods for data transformation.
- 8 a Compare OLTP with OLAP systems
OR
b Outline about metadata repository.
- 9 a Explain about the two step process of Association rule mining.
OR
b How does classification work?
- 10 a Discuss about k -Medoids algorithm.
OR
b Outline about Graph Pattern Mining.

Cont...

SECTION -C (30 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks

(5 x 6 = 30)

- 11 a Summarize the various aspects of mining methodologies.
OR
b Outline the major issues in data mining in terms of efficiency and scalability.
- 12 a Discuss about the major steps involved in data preprocessing.
OR
b Explain about the heuristic methods of attribute subset selection.
- 13 a Discuss about the Multidimensional model of data warehousing.
OR
b Categorize the measures used in computing a data cube.
- 14 a Explain apriori algorithm for finding frequent itemsets.
OR
b Explain about how to calculate gini index in decision tree.
- 15 a Classify the clustering methods.
OR
b Discuss about the Statistical modeling of networks.

Z-Z-Z

END