

PE-516
(564) M. Sc. COMPUTER SCIENCE (**FOURTH** SEMESTER) Examination
JUNE 2022

Compulsory/Optional

Group-

Paper-II

Name/Title of the paper – DATA MINING AND DATA WARE HOUSING

Time: Three Hours

Maximum Marks-080

Minimum passing Marks-...

Note: Answer From Both the section as directed. The figures in the right hand margin indicate marks.

Section A

1 Answer the objective type question.

10X1

- I. Which of the following can be considered as the correct process of Data Mining?
 - a. Infrastructure, Exploration, Analysis, Interpretation, Exploitation
 - b. Exploration, Infrastructure, Analysis, Interpretation, Exploitation
 - c. Exploration, Infrastructure, Interpretation, Analysis, Exploitation
 - d. Exploration, Infrastructure, Analysis, Exploitation, Interpretation
- II. The adaptive system management refers to:
 - a. Science of making machine performs the task that would require intelligence when performed by humans.
 - b. A computational procedure that takes some values as input and produces some values as the output.
 - c. It uses machine learning techniques, in which programs learn from their past experience and adapt themselves to new conditions or situations.
 - d. All of the above.
- III. Which of the following statements is incorrect about the hierarchical clustering?
 - a. The hierarchical type of clustering is also known as the HCA
 - b. The choice of an appropriate metric can influence the shape of the cluster
 - c. In general, the splits and merges both are determined in a greedy manner
 - d. All of the above
- IV. Which of the following statement is true about the classification?
 - a. It is a measure of accuracy
 - b. It is a subdivision of a set
 - c. It is the task of assigning a classification

- d. None of the above
- V. In order to integrate heterogeneous databases, how many types of approaches are there in the data warehousing?
 - a. 3
 - b. 4
 - c. 5
 - d. 2
- VI. Identify the options below that a data warehouse can include.
 - a. Database table
 - b. Online data
 - c. Flat files
 - d. All of the above
- VII. Small logical units where data warehouses hold large amounts of data is known as _____.
 - a. Access layers
 - b. Data marts
 - c. Data storage
 - d. Data miners
- VIII. DSS in data warehouse stands for _____.
 - a. Decision single system
 - b. Decision support system
 - c. Data support system
 - d. Data storable system
- IX. Who is responsible for running queries and reports against data warehouse tables?
 - a. Software
 - b. Hardware
 - c. Middle ware
 - d. End-users
- X. ETL stands for _____.
 - a. Effect, transfer, and load
 - b. Explain, transfer and load
 - c. Extract, transfer, and load
 - d. Extract, transform, and load

2 Answer the short type question.

5X2

- a. Write importance of data mining.
- b. Define OLTP and OLAP.
- c. What is classification?
- d. Define association rule.

- e. Write different data mining tool.

Section B

Answer the Long type question.

12X5

UNIT-I

3 Define data mining and describe different functionality of data mining.

OR

Explain the following.

- a. Knowledge discovery
- b. Steps of data mining

UNIT-II

4 Explain the concept data integration and transformation ^M respect of data pre-processing.

OR

Write notes on.

- a. Delaware house architecture
- b. Data cleaning and data Reduction

UNIT-III

5 Define classification and explain different decision tree induction with suitable example.

OR

Explain the following

- a. Classification by back propagation
- b. Prediction using linear and multiple regression

UNIT-IV

6 describe different association rules with suitable example in respect of mining.

OR

Explain frequent item set mining using apriori algorithm.

UNIT-V

7 what do you mean by cluster analysis? Describe K-means clustering with suitable example.

OR

Write notes on:

- a. WEKA data mining tool
- b. Clustering and association rule mining