

Roll No.

E-790

M. Sc. (Third Semester)
EXAMINATION, Dec.-Jan., 2020-21

COMPUTER SCIENCE

(Elective—II)

(Data Mining and Data Warehousing)

Time : Three Hours]

[Maximum Marks : 100

Note : Attempt all Sections as directed.

Section—A

1 each

(Objective/Multiple Choice Questions)

Note : Attempt all questions.

Choose the correct answer :

1. In a data mining task when it is not clear about what type of patterns could be interesting, the data mining system should :
 - (a) Perform all possible data mining tasks
 - (b) Handle different granularities of data and patterns
 - (c) Perform both descriptive and predictive tasks
 - (d) Allow interaction with the uses to guide the mining process

P. T. O.

2. Which of the following is used to find inherent regularities in data ?
 - (a) Clustering
 - (b) Frequent pattern analysis
 - (c) Regression analysis
 - (d) Outlier analysis
3. The goal of clustering analysis is to :
 - (a) Maximize the inter-cluster similarity
 - (b) Maximize the intra-cluster similarity
 - (c) Maximize the number of clusters
 - (d) Minimize the intra-cluster similarity
4. Pruning a decision tree always :
 - (a) Increases the error rate
 - (b) Reduces the size of the tree
 - (c) Provides the partitions with lower entropy
 - (d) Reduces classification accuracy
5. Which of the following classifiers fall in the category of lazy learners ?
 - (a) Decision trees
 - (b) Bayesian classifiers
 - (c) K-NN classifiers
 - (d) Rule-based classifiers

6. In which of the following, data are stored, retrieved and updated ?
 - (a) OLAP
 - (b) MOLAP
 - (c) HTTP
 - (d) OLTP
7. Data warehouse deals with which type of data that is never found in the operational environment ?
 - (a) Normalized
 - (b) Informal
 - (c) Summarized
 - (d) Denormalized
8. Classification is a data mining task that maps the data into
 - (a) predefined group
 - (b) real valued prediction variable
 - (c) time series
 - (d) clusters
9. Data transformation includes which of the following ?
 - (a) A process to change data from a detailed level to a summary level

- (b) A process to change data from a summary level to a detailed level
 - (c) Joining data from one source into various sources of data
 - (d) Separating data from one source into various sources of data
10. A multifield transformation does which of the following ?
- (a) Converts data from one field into multiple fields
 - (b) Converts data from multiple fields into one field
 - (c) Converts data from multiple fields into multiple fields
 - (d) All of the above
11. How many categories of functions are involved in data mining ?
- (a) 2
 - (b) 3
 - (c) 4
 - (d) 5
12. “Handling of relational and complex types of data” issue comes under :
- (a) Mining methodology and user interaction issues
 - (b) Performance issues
 - (c) Diverse data types issues
 - (d) None of these

13. DMQL stands for :

- (a) Data Mining Query Language
- (b) Dataset Mining Query Language
- (c) DBMiner Query Language
- (d) Data Marts Query Language

14. A bi-directional search takes advantages of process.

- (a) Bottom-up
- (b) Top-down
- (c) Both (a) and (b)
- (d) None of these

15. Partition Algorithm executes in :

- (a) One phase
- (b) Two phase
- (c) Three phase
- (d) None of these

16. Pincer search algorithm based on the principle of :

- (a) Bottom-up
- (b) Top-down
- (c) Directional
- (d) Bi-directional

17. Warehouse administrator responsible for :
- (a) Administrator
 - (b) Maintenance
 - (c) Both (a) and (b)
 - (d) None of these
18. What are the common approaches to tree pruning ?
- (a) Pre-pruning and post-pruning approach
 - (b) Pre-pruning
 - (c) Post-pruning
 - (d) None of these
19. Classification and prediction are two forms of :
- (a) Data analysis
 - (b) Decision tree
 - (c) Both (a) and (b)
 - (d) None of these
20. What is the full form of CLARA ?
- (a) Clustering Large Applicant
 - (b) Close Large Applicant
 - (c) Clustering Large Applications
 - (d) None of these

Section—B

2 each

(Very Short Answer Type Questions)

Note : Attempt all questions.

1. What is data cleaning ?
2. What is data reduction ?

3. What is classification ?
4. What is data pre-processing ?
5. What is multimedia mining ?
6. What is temporal mining ?
7. Define OLTP.
8. What do you mean by cluster ?
9. What is Metadata ?
10. What are the types of dimensional modelling ?

Section—C

3 each

(Short Answer Type Questions)

Note : Attempt all questions.

1. What are the stages of data warehousing ?
2. What is the difference between OLTP and OLAP ?
3. What are aggregate tables ?
4. What is Dimensional Modelling and its types ?
5. Explain Association rule in mathematical notations.
6. What is the purpose of Apriori algorithm ?
7. Define wave cluster.
8. Explain Outlier Analysis.
9. Explain data reduction.
10. Define ROLAP and MOLAP.

Section—D

6 each

(Long Answer Type Questions)

Note : Attempt any *five* questions.

1. Briefly explain various steps of data mining process. What are the major classifications of data mining systems ?
2. Define association rule mining and explain how Apriori algorithm works.

3. Explain the classification of major clustering methods.
4. What is data warehouse ? Explain the data warehouse architecture .
5. Explain the basic dimensional modelling techniques.
6. Explain mining of spatial databases and mining of text databases.
7. Explain in detail Bayesian classification.
8. Explain Density-based and Grid-based methods.