Roll No. ..... Total Printed Pages - 8

F - 830

M.Sc. (IT) (Third Semester)
EXAMINATION, Dec. - Jan., 2021-22
(Elective IV - II)
DATA MINING AND DATA WAREHOUSING

[Time : Three Hours] *[Maximum Marks : 100]*
[Minimum Pass marks : 40]

Note : Attempt all sections as directed

Section - A
(Objective / Multiple Choice Questions)
(1 mark each)

Note : Attempt all questions.

Choose the correct answers:

1. A priori algorithm operates in method
   (A) Bottom-up search method
   (B) Breadth - first search method
   (C) None of above
   (D) Both (A) & (B)

2. A bi-directional search takes advantage of process -
   (A) Bottom-up process
   (B) Top-down process
   (C) None
   (D) Both (A) & (B)
3. DIC algorithm stands for -
   (A) Dynamic itemset counting algorithm
   (B) Dynamic itself counting algorithm
   (C) Dynamic item set countless algorithms
   (D) None of above
4. If the item set is in a dashed circle while completing a full pass, it moves towards -
   (A) Dashed circle
   (B) Dashed box
   (C) Solid Box
   (D) Solid circle
5. Frequent set properties are -
   (A) Downward closure property
   (B) Upward closure property
   (C) A & B
   (D) None of these

6. Periodic maintenance of a data mart means-
   (A) Loading
   (B) Refreshing
   (C) Purging
   (D) All are true
7. The Fp-tree Growth algorithm was proposed by-
   (A) Srikant
   (B) Aggrawal
   (C) Hanetal
   (D) None of these
8. The main idea of the algorithm is to maintain a frequent pattern tree of the date set. An extended prefix tree structure starting crucial and quantitative information about frequent sets.
   (A) Priori Algorithm
   (B) Pinchers Algorithm
   (C) FP-Tree Growth Algorithm
   (D) All of these
9. The data warehousing and data mining technologies have extensive potential applications in the govt . in various central govt. sectors such as :
   (A) Agriculture
   (B) Rural Development
   (C) Health and Energy
   (D) All are true

10. ODS stands for -
    (A) External operational data sources
    (B) Operational data source
    (C) Output data source
    (D) None of the above
11. Good performance can be achieved in a data mart environment by extensive use of -
    (A) Indexes
    (B) Creating profile records
    (C) Volumes of data
    (D) All of the above
12. For a list T, we donote head_t as its first element and body-t as the remaining part of the list (the portion of the list T often removal of head_t) thus t is -
    (A) {head} {body}
    (B) {head_t} {body_t}
    (C) {t_head} {t_body}
    (D) None of these
13. Partition Algorithm executes in -
    (A) One phase
    (B) Two phase
    (C) Three phase
    (D) None of these

14. In the first phase of Partition Algorithm-
    (A) Logically divides into a number of non-overlapping partitions
    (B) Logically divides into a number of overlapping Parti tions.
    (C) Not divides into partitions
    (D) Divides into non-logically and non-overlapping Partitions.
15. Functions of the second phase of the partition Algorithm are
    (A) Actual support of item sets are generated
    (B) Frequent item sets are identified
    (C) Both (A) & (B)
    (D) None of these
16. Partition Algorithm is based on the
    (A) Size of the global Candidate set
    (B) Size of the local Candidate set
    (C) Size of frequent item sets
    (D) No. of item sets
17. Pincer search Algorithm based on the principle of-
    (A) Bottom-up
    (B) Top-Down
    (C) Directional
    (D) Bi- Directional

18. Is a full- breadth search, where no background knowledge of frequent item sets is used for pruning?
    (A) Level-crises filtering by the single item
    (B) Level-by-level independent
    (C) Multi-level mining with uniform support
    (D) Multi-level mining with reduced support
19. Disadvantage of uniform support is
    (A) Items at lower levels of abstraction will occur as frequently.
    (B) If minimum supoort threshold is set too high, I could miss several meaningful associations.
    (C) Both (A) & (B)
    (D) None of these
20. Warehouse administrator responsible for
    (A) Administrator
    (B) Maintenance
    (C) Both (A) and (B)
    (D) None of the above

Section - B

(Very short answer type question)

(2 marks each)

Note : Attempt all questions. Write answer in 2 - 3 sentences.

1. Define: Data Mining.
2. What is frequent item set?
3. List any four reasons to perform data preprocessing.
4. What is a data reduction technique?
5. Differentiate classification and prediction.
6. What is back propagation?
7. What is data warehouse?
8. Define: Data Cube.
9. List any two applications of data mining.
10. What are inductive databases?

Section - C

(Short Answer Type Questions)

(3 marks each)

Note: Attempt all questions. Answer precisely using < 75 words.

1. What factors lead to the mining of data?
2. What are the various forms of visualizing the discovered patterns in data mining?
3. What is data discretization? Give an example.
4. Define the terms support and confidence.

5. Differentiate agglomerative and divisive hierarchical clustering.
6. Compare clustering and classification.
7. What is metadata?
8. Define data cube.
9. Write a short note about text mining.
10. What is web usage mining ?

Section - D

(Long Answer Type Question)

(6 marks each)

Note: Attempt any five questions. Answer precisely using150 words.

1. What is data mining functionality? Expalin different types of data mining functionality with examples.
2. With a neat diagram explain the architecture of data mining.
3. Explain how data mining system can be integrated with database/data warehouse system.
4. Describe the different methods for data cleaning.
5. Explain FP tree algorithm with an example.
6. Explain the algorithm for constructing a decision tree from training samples.
7. Discuss the following clustering algorithm using examples:
   (i) K - means.
   (ii) K - medoid.