

Roll No.....

D-1018(A)

M.Sc. (Fourth Semester) (ATKT)  
Examination-May-June 2020  
COMPUTER SCIENCE  
(Paper Third)  
(Data mining and Data warehousing )

Time : 3 hrs]

[Maximum Marks: 100

Note : Attempt all questions as directed.

Section A

1 each

(Objective/ MCQ)

1. Classification is

- A.A subdivision of a set of examples into a number of classes
- B.A measure of the accuracy, of the classification of a concept that is given by a certain theory
- C.The task of assigning a classification to a set of examples
- D.None of these

2. Cluster is

- A.Group of similar objects that differ significantly from other objects
- B.Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
- C.Symbolic representation of facts or ideas from which information can potentially be extracted
- D.None of these

3. Data mining is

- A. The stage of selecting the right data for a KDD process
- B. The actual discovery phase of a knowledge discovery process
- C. A subject-oriented integrated time variant non-volatile collection of data in support of management
- D. None of these

4. Data selection is

- A. The actual discovery phase of a knowledge discovery process
- B. The stage of selecting the right data for a KDD process
- C. A subject-oriented integrated time variant non-volatile collection of data in support of management
- D. None of these

5: Classification task referred to

- A. A subdivision of a set of examples into a number of classes
- B. A measure of the accuracy, of the classification of a concept that is given by a certain theory
- C. The task of assigning a classification to a set of examples
- D. None of these

6: Euclidean distance measure is

- A. A stage of the KDD process in which new data is added to the existing selection.
- B. The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them
- C. The distance between two points as calculated using the Pythagoras theorem
- D. None of these

7: Kohonen self-organizing map referred to

- A. The process of finding the right formal representation of a certain body of knowledge in order to represent it in a knowledge-based system
- B. It automatically maps an external signal space into a system's internal representational

space. They are useful in the performance of classification tasks

C. A process where an individual learns how to carry out a certain task when making a transition from a situation in which the task cannot be carried out to a situation in which the same task under the same circumstances can be carried out.

D. None of these

8: Incremental learning referred to

A. Machine-learning involving different techniques

B. The learning algorithmic analyzes the examples on a systematic basis and makes incremental adjustments to the theory that is learned

C. Learning by generalizing from examples

D. None of these

9: KDD (Knowledge Discovery in Databases) is referred to

A. Non-trivial extraction of implicit previously unknown and potentially useful information from data

B. Set of columns in a database table that can be used to identify each record within this table uniquely.

C. collection of interesting and useful patterns in a database

D. none of these

10: Machine learning is

A. An algorithm that can learn

B. A sub-discipline of computer science that deals with the design and implementation of learning algorithms

C. An approach that abstracts from the actual strategy of an individual algorithm and can therefore be applied to any other form of machine learning.

D. None of these

11: Multi-dimensional knowledge is

A. A class of learning algorithms that try to derive a Prolog program from examples

- B. A table with  $n$  independent attributes can be seen as an  $n$ -dimensional space
- C. A prediction made using an extremely simple method, such as always predicting the same output.
- D. None of these

12: Data dictionary is

- A. Large collection of data mostly stored in a computer system
- B. The removal of noise errors and incorrect input from a database
- C. The systematic description of the syntactic structure of a specific database. It describes the structure of the attributes the tables and foreign key relationships.
- D. None of these

13: Data warehouse is

- A. The actual discovery phase of a knowledge discovery process
- B. The stage of selecting the right data for a KDD process
- C. A subject-oriented integrated time variant non-volatile collection of data in support of management
- D. None of these

14: Data cleaning is

- A. Large collection of data mostly stored in a computer system
- B. The removal of noise errors and incorrect input from a database
- C. The systematic description of the syntactic structure of a specific database. It describes the structure of the attributes the tables and foreign key relationships.
- D. None of these

15. In contrast to statistics, data mining is \_\_\_\_\_ driven.

- A. Assumption
- B. Knowledge
- C. Human
- D. Database

16. Which of the following is required by K-means clustering?

- a) defined distance metric
- b) number of clusters
- c) initial guess as to cluster centroids
- d) all of the mentioned

View Answer

17: Data can be store , retrieve and updated in ...

- a) SMTOP
- b) OLTP
- c) FTP
- d) OLAP

18. We define a \_\_\_\_\_ as a subdivision of a set of examples into a number of classes.

- a) kingdom
- b) tree
- c) classification
- d) array

19. What is inductive learning?

- a) learning by hypothesis
- b) learning by analyzing
- c) learning by generalizing
- d) none of these

20: Patterns that can be discovered from a given database are which type...

- a) More than one type
- b) Multiple type always
- c) One type only
- d) No specific type

**Section – B**

**2 each**

**(Very Short Answer Type Question)**

**Q-1 What is data mining**

**Q2. What is data cleaning**

**Q-3 what is transactional databases**

**Q-4 Define is classification**

**Q-5 What is prediction**

**Q-6 What is cluster analysis**

**Q-7 What is spatial databases**

**Q-8 Give any two data mining applications**

**Q-9 What is data warehouse**

**Q-10 Define metadata**

**Section C**

**3 each**

**(Short Answer Type Questions )**

**Give Answer in less than 75 words.**

**Q1. Give Classification of data mining system**

**Q2. Explain concept hierarchy generation.**

**Q-3 What are the types of data in cluster analysis**

**Q-4 What is outlier analysis**

**Q-5 What is multidimensional data model**

**Q-6 Explain mining text databases**

**Q-7 What is commit protocol**

**Q-8 Explain data cube technology**

**Q-9 What is data integration**

**Q-10 What are the steps involved in KDD process.**

**Section D**

**6 each**

**(Long answer question)**

**Q1.Explain Data warehouse architecture in detail.**

**Q2 Discuss data mining query languages.**

**Q-3 Give detailed architecture of data mining system**

**Q-4 Explain classification based on concept from association rule mining**

**Q-5 Discuss multidimensional analysis and descriptive mining of complex data object.**