./Roll No. .....

E–833

M. Sc. (IT) (Third Semester)

EXAMINATION, Dec.-Jan., 2020-21

(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 1 each

(Objective/Multiple Choice Questions)

Note : Attempt all questions.

Choose the correct answer :

1. Which of the following applied on warehouse ?

   (a) write only

   (b) read only

   (c) Both (a) and (b)

   (d) None of the above

2. Data can be store, retrieve and updated in ..... .

   (a) SMTOP

   (b) OLTP

   (c) FTP

   (d) OLAP

3. Which of the following is a good alternative to the star schema ?

   (a) snow flake schema

   (b) star schema

   (c) star snow flake schema

   (d) fact constellation

4. 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

5. Background knowledge is ..... .

   (a) It is a form of automatic learning

   (b) A neural network that makes use of a hidden layer

   (c) The additional acquaintance used by a learning algorithm to facilitate the learning process

   (d) None of the above

6. Which of the following is true for Classification ?

   (a) A subdivision of a set

   (b) A measure of the accuracy

   (c) The task of assigning a classification

   (d) All of these

7. Data mining is :

   (a) Time variant non-volatile collection of data

   (b) The actual discovery phase of a knowledge

   (c) The state of selecting the right data

   (d) None of the above

8. ..... is not a data mining functionality.

   (a) Clustering and analysis

   (b) Selection and interpretation

   (c) Classification and regression

   (d) Characterization and discrimination

9. Which of the following can also applied to other forms ?

   (a) Data streams and Sequence data

   (b) Networked data

   (c) Text and Spatial data

   (d) All of the above

10. Which of the following is general characteristics or features of a target class of data ?

    (a) Data selection

    (b) Data discrimination

    (c) Data classification

    (c) Data characterization

11. ..... is the output of KDD.

    (a) Query

    (b) Useful Information

    (c) Data

    (d) Information

12. What is noise ?

    (a) Component of a network

    (b) Context of KDD and data mining

    (c) Aspects of a data warehouse

    (d) None of these

13. Data mining system classification consists of :

    (a) Database Technology

    (b) Machine Learning

    (c) Information Science

    (d) All of the above

14. Firms that are engaged in sentiment mining are analyzing data collected from :

    (a) social media sites

    (b) in-depth interviews

    (c) focus groups

    (d) experiments

15. Which of the following forms of data mining assigns records to one of a predefined set of classes ?

    (a) Classification

    (b) Clustering

    (c) Both (a) and (b)

    (d) None of the above

16. What is the use of data cleaning ?

    (a) To remove the noisy data

    (b) Correct the inconsistencies in data

    (c) Transformations to correct the wrong data

    (d) All of the above

17. The learning which is used to find the hidden pattern in unlabelled data is called :

    (a) Unsupervised learning

    (b) Supervised learning

    (c) Reinforcement learning

    (d) Mixed learning

18. The learning which is the example of Self-organizing maps ?

    (a) Reinforcement learning

    (b) Supervised learning

    (c) Unsupervised learning

    (d) Missing data imputation

19. According to Storks' population size, find the total number of babies from the following examples of predicting the number of babies :

    (a) feature

    (b) outcome

    (c) attribute

    (d) observation

20. Which of the following is not belong to data mining ?

    (a) Knowledge extraction

    (b) Data transformation

    (c) Data exploration

    (d) Data archaeology

Section—B 2 each

(Very Short Answer Type Questions)

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

1. Define pattern evaluation.
2. Define knowledge representation.

3. List the *five* primitives for specification of a data mining task.
4. Define Data integration.
5. Why do we need Data transformation ?
6. What is meant by Data discretization ?
7. List some data mining tools.
8. Describe the use of DB Miner.
9. Define Relational databases.
10. What are Time-series databases ?

Section—C 3 each

(Short Answer Type Questions)

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

1. What are the classifications of data mining system ?
2. What are the issues of data mining ?
3. What are OLAP and OLTP ?
4. What is discrete and continuous data in data mining world ?
5. What is a decision tree algorithm ?
6. Highlight clustering algorithm.
7. What is time series algorithm in data mining ?
8. What is association algorithm in data mining ?
9. What is sequence clustering algorithm ?
10. Specify the concepts and capabilities of Data Mining.

Section—D 6 each

(Long Answer Type Questions)

Note : Attempt any *five* questions. Answer precisely using 150 words.

1. What are the Foundations of Data Mining ? Explain unique Index.

2. Define Wave Cluster. Explain Statistical Perspective in Data Mining.

3. What is Smoothing ? Explain the advantages of data mining over traditional approaches.

4. What are the different ways of moving data/databases between servers and databases in SQL server ?

5. What is Legacy database ? Explain various steps in the data mining process.

6. What do you understand by classification ? Also explain classification of data mining system.

7. What is data preprocessing ? Explain preprocessing technique in detail.