Reg. No. : | | | | | | | | | | |

Question Paper Code: 45205A

B.E. / B.Tech. DEGREE EXAMINATION, APRIL 2019

Fifth Semester

Computer Science and Engineering

14UCS505 - DATA WAREHOUSING AND DATA MINING

(Regulation 2014)

Duration: Three hours   Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 1 = 10 Marks)

1. The process of removing the deficiencies and loopholes in the data is called as

    (a) Aggregation of data   (b) Extracting of data
    (c) Cleaning up of data   (d) Loading of data

2. Data warehouse bus matrix is a combination of

    (a) Dimensions and data marts   (b) Dimensions and facts
    (c) Facts and data marts   (d) Dimensions and detailed facts

3. Data warehouse and OLAP are based on

    (a) multidimensional Data Model   (b) multidimensional Database
    (c) data cube   (d) relational data

4. Which of the following statements are true?

    (a) support and confidence are same
    (b) frequent itemset and candidate itemset are same
    (c) apriori algorithm may be used for supervised classification
    (d) Association mining cannot be used for medical applications

5. The technique that does not use candidate generation in association data mining is

    (a) Apriori   (b) FP Growth   (c) Depth first   (d) Breadth first

6. Spot the preprocessing technique that converts the data into appropriate forms of mining

     (a) Data cleaning   (b) Data transformation

     (c) Data reduction   (d) Data integration

7. _____ consists of using/developing data mining algorithms to discover interesting unexpected and useful patterns in databases.

     (a) Pattern mining   (b) Data mining

     (c) Data warehousing   (d) Data cleaning

8. _____ attempts to form patterns that permit to to predict the next events given the available input data.

     (a) Classification   (b) Prediction

     (c) Correlation   (d) Association

9. _____ is a collection of data objects.

     (a) Cluster   (b) Supervised learning

     (c) Unsupervised learning   (d) Learning by observation

10. _____ technique is used to identify the most crucial factor that may influence a customer's decision regarding banking.

     (a) Clustering   (b) Classification

     (c) Prediction   (d) Association

PART - B (5 x 2 = 10 Marks)

11. How is a data warehouse different from a database? How are they similar?

12. What is dimensionality reduction?

13. Use data smoothing techniques to preprocess the following data. 3 5 4 12 8 9 9 9 7 7 5 6.

14. State the characteristics of classification in terms of Decision tree algorithm?

15. Differentiate between row scalability and column scalability issues.

PART - C (5 x 16 = 80 Marks)

16. (a) Explain in detail about Meta data and data marts.   (16)

Or

    (b) Explain about database architectures for parallel processing.   (16)

45205A

17. (a) Describe OLAP operations in multidimensional data model. (16)

Or

(b) Discuss Multidimensional Online Analytical Processing (MOLAP) and Multi Relational Online Analytical Processing (ROLAP) with relevant example. (16)

18. (a) (i) Classify and explain data mining functionalities. (8)

(ii) Classify and explain the classification of data mining system. (8)

Or

(b) Explain about data preprocessing. (16)

19. (a) Explain in detail about constraint based association mining. (16)

Or

(b) Explain in detail about support vector machine. (16)

20. (a) (i) Explain with an example density based clustering methods. (8)

(ii) Explain in detail about K-means partitioning algorithm. (8)

Or

(b) Explain outlier analysis in detail with an example. Discuss the use of outlier analysis. (16)

———————

45205A