Reg. No. :

Question Paper Code: 45205

B.E. / B.Tech. DEGREE EXAMINATION, NOV 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. _____,which contains information about warehouse data for use by warehouse designers and administrators when carrying the warehouse development and management tasks.

   (a) Business metadata   (b) Metadata
   (c) Technical metadata   (d) Data sourcing

2. Metadata contains

   (a) Migration tools
   (b) Tram formations details
   (c) Location and description of warehouse systems and data components
   (d) Cleanup information

3. Which one of the tool is not a decision supporting tool?

   (a) Reporting and OLAP   (b) Data mining and managed query
   (c) Executive information systems (d) Data warehousing

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. Any subset of a frequent set is a frequent set. This is

   (a) Upward closure property   (b) Download closure property

   (c) Maximal frequent set   (d) Border set

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. What are the requirements of cluster analysis?

   (a) Scalability   (b) High dimensionality
   (c) Both (a) and (b)   (d) None of these

10. Which of the following is/are applications of data mining?

   (a) Financial Data Analysis   (b) Retail Industry
   (c) Telecommunication Industry   (d) All the above

PART - B (5 x 2 = 10 Marks)

11. Define slice and dice operation.

12. What is dimensionality reduction?

13. List the techniques to improve the efficiency of Apriori algorithm.

14. What is tree pruning?

15. Differentiate between row scalability and column scalability issues.

45205

PART - C (5 x 16 = 80 Marks)

16. (a) Explain in detail about the architecture and implementation of the data warehouse with neat sketch. (16)

Or

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

17. (a) List and discuss the basic features that are provided by reporting and query tools used for business analysis. (16)

Or

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

18. (a) Describe the data mining functionality and examine. What kinds of patterns can be mined. (16)

Or

(b) Explain about data preprocessing. (16)

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

Or

(b) Explain the Naïve Bayes algorithm for solving classification problems. State the possible steps and conditions for effective analysis. (16)

20. (a) Explain in detail about hierarchical method. (16)

Or

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

_____

45205