

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 41255

B.E. / B.Tech. DEGREE EXAMINATION, MAY 2017

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. Data mining is viewed as a process that
 - (a) Facilitate efficient storage and retrieval
 - (b) Uncover hidden patterns in the data
 - (c) Offers multidimensional view of data
 - (d) Provides summary 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. Any subset of a frequent set is a frequent set. This is
 - (a) Upward closure property
 - (b) Downward closure property
 - (c) Maximal frequent set
 - (d) Border set
8. In Bayesian classification, using hypothesis H for observed data tuple X, P(H/X) is determined for classification problem, which is known as
 - (a) Posterior Probability
 - (b) Prior Probability
 - (c) Conditional Probability
 - (d) Joint Conditional Probability
9. What are the requirements of cluster analysis?
 - (a) Scalability
 - (b) High dimensionality
 - (c) Both (a) and (b)
 - (d) None of these
10. The characteristics of an outlier in DENCLUE is
 - (a) Density attractor has a density higher than certain threshold
 - (b) Density attractor has a density lower than certain threshold
 - (c) To form clusters
 - (d) Computed to detect cluster

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. Distinguish between Partitioning and hierarchical methods.

PART - C (5 x 16 = 80 Marks)

16. (a) Explain the Data Warehouse architecture and components with a neat sketch. (16)
- Or
- (b) Describe the DBMS schemas for decision support system. (16)

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

Or

(b) Explain in detail about the various OLAP Query and Reporting tools. (16)

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

Or

(b) List and explain the primitive for specifying a data mining task. Also explain the various data pre-processing methods with suitable example. (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) (i) Discuss the strength and weakness of k-means algorithm and How to overcome those limitations. (8)

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

Or

(b) Explain the Density based data clustering and state the DBSCAN algorithm in detail and state a real time application that uses this algorithm. (16)
