

Reg. No. :

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

Question Paper Code: 45205

B.E. / B.Tech. DEGREE EXAMINATION, NOV 2018

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) Transformations details
 - (c) Location and description of warehouse systems and data components
 - (d) Cleanup information

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. _____ 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. Define slice and dice operation.
12. List the various access types to the data stored in the data warehouse.
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 in detail about the architecture and implementation of the data warehouse with neat sketch. (16)

Or

(b) Explain in detail about metadata. Classify metadata and explain the same. (16)

17. (a) (i) List and discuss the basic features that are provided by reporting and query tools used for business analysis. (10)

(ii) Describe multidimensional data model in detail. (6)

Or

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

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

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

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) Explain in detail about hierarchical method. (16)

Or

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

