Register Number.....

M.E. DEGREE EXAMINATIONS: OCTOBER / NOVEMBER-2008

Third Semester

COMPUTER SCIENCE AND ENGINEERING

P07CSE14: Data Warehousing And Data Mining

Time: Three Hours **Maximum Marks: 100**

Answer ALL Questions: -

PART A (20 × 1 = 20 Marks)

1. A_____ stores sequences of values or events obtained over repeated measurements of time
   A. Temporal Database B. Time-series database
   C. Sequence Database D. Spatial Database

2. _____ describes and models regularities or trends for objects whose behavior changes over time
   A. Classification B. Prediction
   C. Data evolution analysis D. Correlation analysis

3. _____means that besides linking a DM system to a DB/DW system, efficient implementations of a few essential data mining primitives (identified by the analysis of frequently encountered data mining functions) can be provided in the DB/DW system
   A. Loose Coupling B. Semitight Coupling C. Tight Coupling D. No Coupling

4. _____is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes.
   A. Data discrimination B. Data characterization
   C. Classification D. Pattern Evaluation

5. Suppose that the data for analysis includes the attribute age. The values for the data tuples are (in increasing order) 14,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,35,36. What is the midrange of the data?
   A. 25 B. 23.4 C. 22 D. 21

6. Find ODD one out of the following
   A. sum() B. count() C. avg() D. min()

7. A_____ technique can be used that requires just two database scans to mine the frequent itemsets
   A. transaction reduction B. hash-based C. partitioning D. dynamic itemset counting

8. Which one of the following is not correct statement?
   A. Entropy is one of the dimensionality reduction technique
   B. Attribute subset selection reduces the data set size by removing irrelevant or relevant attributes
   C. Smoothing works to remove noise from the data
   D. Histograms use binning to approximate data distribution and are a popular form of data reduction

9. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8). Compute the Euclidean distance
   A. 139 B. 6.7 C. 11 D. 7

10. _____is designed for clustering a large amount of numerical data by integration of hierarchical clustering and other clustering methods such as iterative partitioning.
A. BIRCH B.ROCK C. Chamelon D. DBSCAN

11. In_____, the learned model is represented as a set of IF-THEN rules
A. Rule-based classifiers B. Bayesian classification
C. Rough Set Approach D. Neural network

12. _____is a special case of *k*-fold cross-validation where *k* is set to the number of initial tuples
A. Bootstrap B. Bagging
C. Leave-one-out D. Boosting

13. Which one of the statement is not true about OLAP System?
A. Users of the OLAP system are manager, executive and analysts
B. It makes use of E-R based database design
C. Millions of records can be accessed in this system
D. This system is mostly used for reading the data

14. The_____ operation performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension or by dimension reduction.
A. Drill-up B. Drill-down C. Slice and Dice D. Pivot

15. _____converts data from legacy or host format to warehouse format
A. Data Transformation B. Load
C. Data Extraction D. Data Cleaning

16. _____is considered a good choice for data warehouse development, especially for data marts, because the turnaround time is short, modifications can be done quickly, and new designs and technologies can be adapted in a timely manner.
A. Spiral method B. Waterfall method
C. Prototyping method D. Incremental method

17. EnterpriseMiner was developed by_____
A. SPSS B.SAS Institute
C. Oracle D. Microsoft

18. _____systems usually handle vector data that consist of points, lines, and polygons.
A. Temporal database B. Spatial database
C. Relational database D. Distributed database

19. Document Clustering Analysis is one of the_____ approaches
A. Text mining B. WWW mining
B. Spatial data mining D. Multimedia data mining

20. In _____ approach, the signature of an image includes a composition of multiple features: color histogram, shape, image topology, and texture.
A. Wavelet-based signature B. Multi-feature composed signature
C. Color histogram-based signature
D. Wavelet-based signature with region-based granularity

PART B (5 ×16 = 80 Marks)

21. (a) (i) Describe the various steps involved in KDD (8)

(ii)Discuss about the classification of Data mining system (8)

(OR)

21. (b) Explain various techniques of Data mining in detail (16)

22 (a). With a relevant example explain the single dimensional Boolean Association rule mining algorithm from transactional databases? (16)

(OR)

22 (b). Explain the need and steps involved in data preprocessing (16)

23 (a). Write short notes on
(i) Bayesian Classification (8)
(ii) Partitioning methods (8)

(OR)

23(b). The following table shows the midterm and final exam grades obtained for students in a database course.

x Midterm exam	y Final exam
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

(i) Plot the data. Do x and y seem to have a linear relationship? (6)

(ii) Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course and Predict the final exam grade of a student who received an 86 in the midterm exam. (10)

24(a). In a data warehousing technology, a multiple multidimensional view can be implemented by ROLAP or by a MOLAP or by a MOLAP or by a HOLAP. Briefly describe each implementation techniques and the various optimization techniques used for cube computation. (16)

(OR)

24(b) (i) Compare OLTP with OLAP System (8)

(ii) Explain the operations performed on data warehouse with examples. (8)

25(a). Discuss in detail the Application of Data mining for financial Data Analysis? Give suitable data flow diagram. (16)

(OR)

25. (i) Discuss how data mining is done in Spatial databases. (8)
(ii) Explain about mining WWW. (8)

****