



B.TECH DEGREE EXAMINATIONS: APRIL/MAY 2016

(Regulation 2013)

Sixth Semester

INFORMATION TECHNOLOGY

U13ITT602 : Data Warehousing and Data Mining

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. _____ holds the highest level of summarization.
 - a) Apex cuboid
 - b) n-D cuboid
 - c) Base cuboid
 - d) 1-D data cube
2. _____ tools use domain knowledge to detect errors and make corrections in the data.
3. Which is not a measure of Data dispersion?
 - a) Quartile
 - b) midrange
 - c) Outlier
 - d) Variance
4. A _____ hierarchy is a total / partial order among attributes in the database schema.
5. Positive tuples that were mislabeled as negative is termed as _____
 - a) true positives
 - b) true negatives
 - c) false positives
 - d) false negative
6. Zero probability problem can be solved using _____
7. A tree structure called a _____ is commonly used to represent the process of Hierarchical Clustering
 - a) Histogram
 - b) Pictogram
 - c) Dendrogram
 - d) Monograms
8. _____ is the property that if a set cannot pass a test, all of its supersets will fail the same test as well.
9. Which is not a type of dimensions in a spatial data cube?
 - a) Nonspatial
 - b) Spatial-to-nonspatial
 - c) Spatial to spatial
 - d) Nonspatial-to - spatial
10. _____ mines Web log records to discover user access patterns of Web pages.

PART B (10 x 2 = 20 Marks)

(Answer not more than 40 words)

11. List the various OLAP Operations on multidimensional data.
12. With an example define concept hierarchy
13. Mention the ways to represent the discovered patterns?
14. Write the formula for calculating the weighted arithmetic mean?
15. Differentiate Classification and Prediction
16. What is the process of Lazy Learner model?
17. Expand DIANA and AGNES.
18. Mention some applications of Outlier mining.
19. Define Spatial data mining?
20. What is meant by a frequent graph?

PART C (5 x 14 = 70 Marks)

(Answer not more than 400 words)

Q.No. 21 is Compulsory

21. List and elaborately describe the five primitives for specifying a Data mining task?

22. (a) Illustrate with a neat diagram, explain all the necessary components of the data warehouse architecture.

(OR)

- (b) (i) Explain about the different schemas used to represent the multidimensional model. (8)
- (ii) Summarize how missing values are handled in data cleaning process. (6)

23. (a) Outline the steps for classifying the tuples based on the Decision tree induction algorithm.

(OR)

- (b) (i) The following table contains a set of paired data where x is the number of years of work experience of a graduate and Y is the salary of the graduate. (8)

Salary data	
x	salary (in \$1000s) y
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

Using Linear Regression Technique, Predict the salary of the graduate with 10 years of experience.

- (ii) Write short notes on Classifier Accuracy Measures. (6)

24. (a) Explain k -Means & k -Medoids techniques with algorithm that organizes the objects of a set into several exclusive groups or clusters.

(OR)

- (b) Consider the following transactional database.

Given: $\text{min_sup} = 60\%$ and $\text{min_conf} = 80\%$. Find all frequent item sets using Apriori algorithm for the given database. Also generate the strong association rules from the frequent item sets generated.

TID	Items bought
T1	{M,O,N,K,E,Y}
T2	{D,O,N,K,E,Y}
T3	{M,A,K,E}
T4	{M,U,C,K,Y}
T5	{C,O,O,K,I,E}

25. (a) Explain how similarity search and subsequence matching is done in time-Series data.

(OR)

- (b) (i) List the challenges faced during web mining and explain the purpose of Web usage mining. (6)
- (ii) Write notes on Mining Spatial Databases. (8)
