

Printed pages: 02

Sub Code: ECS 075

Paper Id:

1	0	1	4
---	---	---	---

Roll

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

 No.

B.TECH.
(SEM VII) THEORY EXAMINATION 2017-18
DATA MINING & DATA WAREHOUSEING

Time: 3 Hours

Total Marks: 100

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

- 1. Attempt all questions in brief. 2 x10 = 20**
- a. Discuss the term Data Mining & Data Warehousing.
 - b. What is role of KDD? Discuss KDD process with suitable steps.
 - c. Explain the terms Data, Information, Knowledge and Wisdom?
 - d. How can we fill up the Missing Values from large amount of data?
 - e. Explain the methods of generating the Concept Hierarchies for categorical data with suitable example.
 - f. Differentiate OLAP with OLTP.
 - g. What is Clustering? Discuss various issues related with Clustering.
 - h. What is Histogram? Discuss various types of Histogram.
 - i. How can Noise be removed from data?
 - j. What is Z-Score Normalization?

SECTION B

- 2. Attempt any three of the following: 10 x 3 = 30**
- a. Explain the architecture of Data Mining System. Discuss various components & major issues related to Data Mining.
 - b. What is Multidimensional Database Schema? Discuss Star, Snow Flack and Fact Constellation schema in detail with suitable example
 - c. What is Association Rule mining? Explain the Apriori Algorithm with an example to find the frequent item sets. How the efficiency of the Apriori algorithm can be improved?
 - d. What is a Multidimensional Data Model? How we convert tables and spreadsheets to Data Cubes? Convert 2-D tables into 3-D Data cube with an example.
 - e. Define and describe the basic similarities and differences among ROLAP, MOLAP, and HOLAP.

SECTION C

- 3. Attempt any two parts of the following: 5 x 2 = 10**
- (a) Describe the role of Genetic Algorithm in Data Mining.
 - (b) Explain mining multilevel association rules from Transactional databases.
 - (c) Discuss various OLAP operations. Explain how query performance can be improved by cascading the operations.
- 4. Attempt any two parts of the following: 5 x 2 = 10**
- (a) Explain the Market Basket Analysis with suitable example.
 - (b) What are the main purposes of Statistics used in Data Mining? Give the Statistical measures in large databases.
 - (c) Explain Attribute Subset Selection method for Data Deduction with examples.

- 5. Attempt any *two* parts of the following: **5 x 2 = 10****
- (a) Describe classification. Briefly outline the major ideas of Basiyan classification.
 - (b) Explain the types of data that often occur in Cluster Analysis and briefly explain how to preprocess that data for clustering.
 - (c) What are Outliers? How outliers analysis can be done ?
- 6. Attempt any *two* parts of the following: **5 x 2 = 10****
- (a) Explain Principal Component Analysis (PCA) in detail.
 - (b) Distinguish between Dimensionality Reduction and Numerosity Reduction.
 - (c) Describe Classification and Prediction. Explain the Classification by Decision Tree Induction method.
- 7. Attempt any *one* part of the following: **10 x 1 = 10****
- (a) Describe the difference between the following approaches for the integration of data mining system with database or data warehouse systems : no coupling
loose coupling, semi tight coupling, tight coupling
 - (b) Write Short notes on following:
 - i. Testing of Data Warehouse
 - ii. CLIQUE
 - iii. STING
 - iv. Backup and Recovery