

**Ph.D. IN COMPUTER SCIENCE
(PHDCS)**

Term-End Examination

December, 2017

00440

RCSE-001 : DATA MINING

Time : 3 hours

Maximum Marks : 100

(Weightage : 50%)

Note : *Question no. 1 is compulsory. Answer any three questions from the rest.*

1. (a) What is a Transactional database ? How are transactional databases used to support online and real-time systems ? 10
- (b) Define a Data Warehouse (DW). What are the various components of a DW system ? Draw and explain its architecture. 10
- (c) Discuss ID3, C4.5 and CART algorithms in detail. Give suitable example for each. 10
- (d) Consider the following data for the attribute "age" in increasing order :
- 13, 15, 16, 19, 20, 21, 22, 22, 25, 25, 25, 30, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70
- Use "Smoothing by bin" means to smooth these data, using a bin depth of 3. Illustrate your steps. Also, comment on the effect of this technique for the given data. 10

- 2. (a) Compare and contrast between the following :**
- (i) Classification and Prediction models 3
 - (ii) Supervised and Unsupervised learning 3
 - (iii) Lazy learner and Eager learner algorithms 4
- (b) Write the Back Propagation algorithm. Explain how the algorithm leads to the learning of a neural network. 10**
- 3. (a) Suppose you are the Vice President of an e-commerce company. Describe any three different ways you will tend to analyze the sales during a festive season. What will be your business dimensions ? Also, find out the hierarchies and categories. Draw an Information Package Diagram (IPD). 10**
- (b) Convert the above IPD (Q. 3(a)) to STAR schema. 10**
- 4. (a) What is Data Mining ? Explain the areas of applications of data mining. Describe the stages in KDD. 10**
- (b) Explain the Apriori algorithm for finding frequent item sets by confined candidate generation. 10**

5. What are the value ranges of the following normalization methods ?

- (a) Min-max Normalization 5
 - (b) Z-score Normalization 5
 - (c) Z-score Normalization using the Mean Absolute Deviation instead of Standard Deviation 5
 - (d) Normalization by Decimal Scaling 5
-